Bayesian inference for the exponential random graph model

Nial Friel and Alberto Caimo

University College Dublin
nial.friel@ucd.ie

August, 2009

Social networks

A Social Network consists of individuals (or organisations) represented by nodes of a graph.

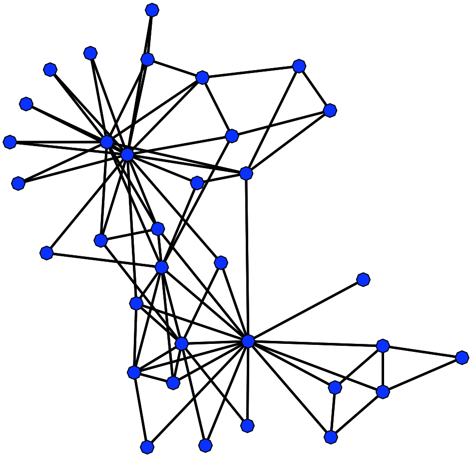
Individuals which are connected by some sort of dependency eg friendship, business relationship, common interest etc, are joined by an edge.

The resultant graphs can be very complex. The statistical challenge is to model such networks probabilistically, predict behaviour of the networks, model networks across time etc etc...

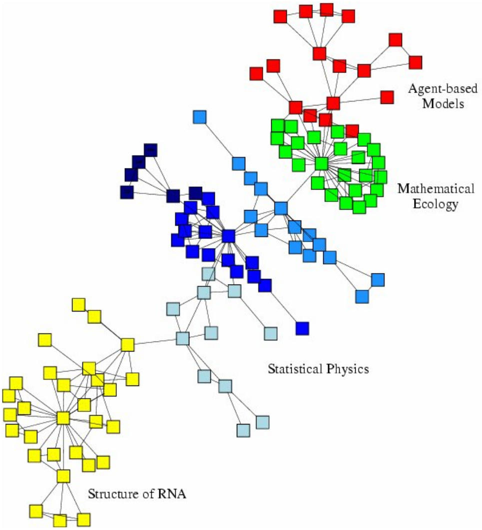
Networks are everywhere!

An example: Zachary karate club

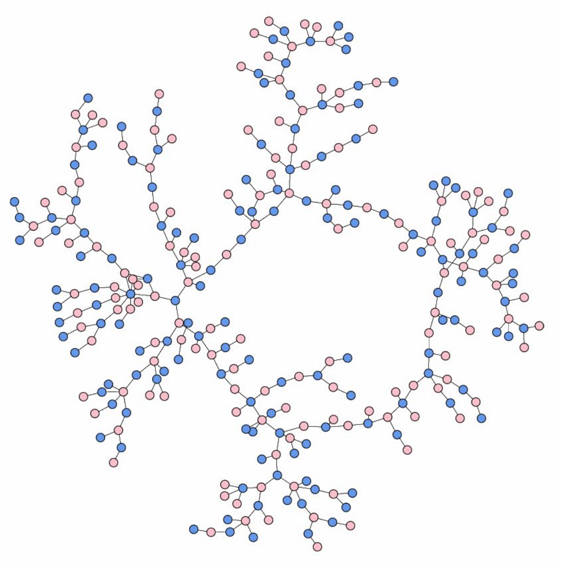
34 members of a university karate club



Scientific collaborations



High school dating



Random graph models

Random graph models go back at least to the 1950's.

The Erdös-Rényi model allows edges to form independently of one another with equal probability.

Let $y_{ij} = 1$ denote an edge connecting nodes i and j.

$$y_{ij} \sim \text{Bernoulli}\,(p) \;\; \text{with log odds} \;\; \theta.$$

$$\pi(\mathbf{y}|\theta) \propto \exp\left(\theta \sum_{i,j} y_{ij}\right).$$

Random graph models

Let $G(n, p)$ denote an Erdös-Rényi model with n nodes and Bernoulli probability p.

It is clearly a simple description of what happens in reality, yet it does have some nice properties.

- The degree distribution for any vertex is binomial

$$\pi(x_i = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

- If $np < 1$, then a $G(n, p)$ graph will have no connected components of size larger than $O(\log n)$, a.s.
- If $np = 1$, then a $G(n, p)$ graph will have a largest component whose size is of order $n^{2/3}$, a.s.
- This model has some nice properties which are respected by subgraphs, eg connectedness.

The exponential random graph model

First proposed by Frank and Strauss (JASA, 1986).

The exponential random graph model

First proposed by Frank and Strauss (JASA, 1986).

1. Edges y_{ij} and y_{kl} are neighbours of one another, if they share a common node.

The exponential random graph model

First proposed by Frank and Strauss (JASA, 1986).

1. Edges y_{ij} and y_{kl} are neighbours of one another, if they share a common node.
2. If y_{ij} and y_{kl} are not neighbours, then y_{ij} and y_{ij} are conditionally independent, given the rest of the graph.

The exponential random graph model

First proposed by Frank and Strauss (JASA, 1986).

1. Edges y_{ij} and y_{kl} are neighbours of one another, if they share a common node.
2. If y_{ij} and y_{kl} are not neighbours, then y_{ij} and y_{ij} are conditionally independent, given the rest of the graph.

The p^* model

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^t s(\mathbf{y})\}}{z(\boldsymbol{\theta})} = \frac{q(\mathbf{y}|\boldsymbol{\theta})}{z(\boldsymbol{\theta})}$$

- $\mathbf{y}$ observed graph
- $s(\mathbf{y})$ known vector of sufficient statistics
- $\boldsymbol{\theta}$ vector of parameters
- $z(\boldsymbol{\theta})$ normalizing constant

The p^* model

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^t s(\mathbf{y})\}}{z(\boldsymbol{\theta})} = \frac{q(\mathbf{y}|\boldsymbol{\theta})}{z(\boldsymbol{\theta})}$$

- $\mathbf{y}$ observed graph
- $s(\mathbf{y})$ known vector of sufficient statistics
- $\boldsymbol{\theta}$ vector of parameters
- $z(\boldsymbol{\theta})$ normalizing constant

$$z(\boldsymbol{\theta}) = \sum_{\textit{all possible graphs}} \exp\{\boldsymbol{\theta}^t s(\mathbf{y})\}$$

- $2^{\binom{n}{2}}$ possible undirected graphs of n nodes
- Calculation of $z(\boldsymbol{\theta})$ is infeasible for non-trivially small graphs

4-dimensional model

$$\pi(\mathbf{y}|\boldsymbol{\theta}) \propto \frac{1}{z(\boldsymbol{\theta})} \exp\left\{\sum_{i=1}^{4} \theta_i s_i(\mathbf{y})\right\} \pi(\boldsymbol{\theta})$$

$s_1(\mathbf{y}) = \sum_{i<j} y_{ij}$	number of edges
$s_2(\mathbf{y}) = \sum_{i<j<k} y_{ik} y_{jk}$	number of two-stars
$s_3(\mathbf{y}) = \sum_{i<j<k<l} y_{il} y_{jl} y_{kl}$	number of three-stars
$s_4(\mathbf{y}) = \sum_{i<j<k} y_{ik} y_{ik} y_{ij}$	number of triangles

Pseudolikelihood

(Besag 1974, Strauss & Ikeda 1990)

$$\pi(\mathbf{y}|\boldsymbol{\theta}) \approx \pi_{pseudo}(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i \neq j} \pi(y_{ij}|\mathbf{y}_{-ij}, \boldsymbol{\theta})$$

- $\mathbf{y}_{-ij}$ all the graph excluding y_{ij}

- Assumption of weak dependence between the variables
- Generally inadequate since it only uses local information whereas the graph is affected by global interaction

Monte Carlo maximum likelihood

(Geyer & Thompson 1992)

$$\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}_0}\left[\frac{q(\mathbf{y}|\boldsymbol{\theta})}{q(\mathbf{y}|\boldsymbol{\theta}_0)}\right] = \sum_{\mathbf{y}} \frac{q(\mathbf{y}|\boldsymbol{\theta})}{q(\mathbf{y}|\boldsymbol{\theta}_0)} \frac{q(\mathbf{y}|\boldsymbol{\theta}_0)}{z(\boldsymbol{\theta}_0)} = \frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\theta}_0)}$$

- $\boldsymbol{\theta}_0$ is fixed vector of parameters
- $\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}_0}$ expectation with respect to $\pi(\mathbf{y}|\boldsymbol{\theta}_0)$

Monte Carlo maximum likelihood

(Geyer & Thompson 1992)

$$\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}_0}\left[\frac{q(\mathbf{y}|\boldsymbol{\theta})}{q(\mathbf{y}|\boldsymbol{\theta}_0)}\right] = \sum_{\mathbf{y}} \frac{q(\mathbf{y}|\boldsymbol{\theta})}{q(\mathbf{y}|\boldsymbol{\theta}_0)} \frac{q(\mathbf{y}|\boldsymbol{\theta}_0)}{z(\boldsymbol{\theta}_0)} = \frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\theta}_0)}$$

- $\boldsymbol{\theta}_0$ is fixed vector of parameters
- $\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}_0}$ expectation with respect to $\pi(\mathbf{y}|\boldsymbol{\theta}_0)$

$$\frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\theta}_0)} = \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}_0}\left[\frac{q(\mathbf{y}|\boldsymbol{\theta})}{q(\mathbf{y}|\boldsymbol{\theta}_0)}\right] = \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t s(\mathbf{y})\right\}\right]$$

$$\approx \frac{1}{m}\sum_{i=1}^{m} \exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t s(\mathbf{y}_i)\right\}$$

Monte Carlo maximum likelihood (cont'd)

Therefore

$$\log\left\{\frac{\pi(\mathbf{y}|\boldsymbol{\theta}_0)}{\pi(\mathbf{y}|\boldsymbol{\theta})}\right\} \approx (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t s(\mathbf{y}) - \log\left\{\frac{1}{m}\sum_{i=1}^{m}\exp\left[(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t s(\mathbf{y}_i)\right]\right\}$$

- Very sensitive to the choice of $\boldsymbol{\theta}_0$ that should be close to the MLE of $\boldsymbol{\theta}$
- A poorly chosen value of $\boldsymbol{\theta}_0$ may lead to a function that may not even have a maximum
- Often $\boldsymbol{\theta}_0$ is chosen as the maximiser of the pseudolikelihood function

Model degeneracy

(Handcock 2003, Rinaldo, Fienberg & Zhou 2009)

- C convex hull of the set $\{s(\mathbf{y}) : \mathbf{y} \in \mathbf{Y}\}$
- $ri(C)$ relative interior
- $rbd(C)$ relative boundary
- Mean value parametrisation: $\mu(\boldsymbol{\theta}) = \mathbb{E}[s(\mathbf{y})]$

The model is near degenerate if $\mu(\boldsymbol{\theta})$ is close to $rbd(C)$

- In practice most of $\pi(\mathbf{y}|\boldsymbol{\theta})$ is placed on a few configurations (eg empty graphs, full graphs,...)
- In such instances, inference is problematic

Model degeneracy (cont'd)

Results

- the MLE exists and it is unique $\Leftrightarrow s(\mathbf{y}) \in ri(C)$
- if $s(\mathbf{y}) \in rbd(C) \Rightarrow$ the MLE does not exist

Model degeneracy (cont'd)

Results

- the MLE exists and it is unique $\Leftrightarrow s(\mathbf{y}) \in ri(C)$
- if $s(\mathbf{y}) \in rbd(C) \Rightarrow$ the MLE does not exist

MC-MLE may fail

- It may be difficult to choose $\boldsymbol{\theta}_0$ far from degeneracy and close to MLE
- Parameter values in the near degenerate region can hinder the convergence of common MCMC algorithms
- Simulating from $\mathbf{y}|\boldsymbol{\theta}_0$ may yield graphs which are full or empty thereby leading to a poor estimate of $z(\boldsymbol{\theta})/z(\boldsymbol{\theta}_0)$
- MC-MLE may have high variance and may not exist

Bayesian inference

Doubly-intractable posterior

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

- Naïve Metropolis-Hastings algorithm proposes the move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$ with probability:

$$\alpha = \min\left(1, \frac{q(\mathbf{y}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)}{q(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \times \underbrace{\frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\theta}^*)}}_{\textit{intractable}}\right)$$

Exchange algorithm

(Murray, Ghahramani & MacKay 2006)

Sample from an augmented distribution

$$\pi(\boldsymbol{\theta}', \mathbf{y}', \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})h(\boldsymbol{\theta}'|\boldsymbol{\theta})\pi(\mathbf{y}'|\boldsymbol{\theta}')$$

whose marginal distribution for $\boldsymbol{\theta}$ is the posterior of interest

- $\pi(\mathbf{y}'|\boldsymbol{\theta}')$ same distribution as the original one on which $\mathbf{y}$ is defined
- $h(\boldsymbol{\theta}'|\boldsymbol{\theta})$ arbitrary distribution for the augmented variable $\boldsymbol{\theta}'$ which might depend on $\boldsymbol{\theta}$ (eg random walk distribution centred at $\boldsymbol{\theta}$)

Exchange algorithm (cont'd)

How it works

1 GIBBS UPDATE OF $(\boldsymbol{\theta}', \mathbf{y}')$

- *i* Draw $\boldsymbol{\theta}' \sim h(\cdot|\boldsymbol{\theta})$
- *ii* Draw $\mathbf{y}' \sim \pi(\cdot|\boldsymbol{\theta}')$

Exchange algorithm (cont'd)

How it works

1. Gibbs update of $(\boldsymbol{\theta}', \mathbf{y}')$
   - *i* Draw $\boldsymbol{\theta}' \sim h(\cdot|\boldsymbol{\theta})$
   - *ii* Draw $\mathbf{y}' \sim \pi(\cdot|\boldsymbol{\theta}')$
2. Exchange move from $(\boldsymbol{\theta}, \mathbf{y})$, $(\boldsymbol{\theta}', \mathbf{y}')$ to $(\boldsymbol{\theta}', \mathbf{y})$, $(\boldsymbol{\theta}, \mathbf{y}')$ with probability

$$\alpha = \min\left(1, \underbrace{\frac{q(\mathbf{y}'|\boldsymbol{\theta})}{q(\mathbf{y}|\boldsymbol{\theta})}}_{*} \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})} \frac{h(\boldsymbol{\theta}|\boldsymbol{\theta}')}{h(\boldsymbol{\theta}'|\boldsymbol{\theta})} \underbrace{\frac{q(\mathbf{y}|\boldsymbol{\theta}')}{q(\mathbf{y}'|\boldsymbol{\theta}')}}_{**} \times \underbrace{\frac{z(\boldsymbol{\theta})z(\boldsymbol{\theta}')}{z(\boldsymbol{\theta})z(\boldsymbol{\theta}')}}_{1}\right)$$

- Exchange move proposes to "offer" the data $\mathbf{y}$ the auxiliary $\boldsymbol{\theta}'$ and similarly to "offer" the auxiliary data $\mathbf{y}'$ the parameter $\boldsymbol{\theta}$
- The affinity between $\boldsymbol{\theta}'$ and $\mathbf{y}$ is measured by (**) and the affinity between $\boldsymbol{\theta}$ and $\mathbf{y}'$ by (*)

MCMC sample from the p^* model

- The main difficulty is the need to draw an exact sample $\mathbf{y}' \sim \pi(\cdot|\boldsymbol{\theta}')$
- Perfect sampling is an obvious approach, if this is possible
- A pragmatic alternative is to take a realisation from a long MCMC run with stationary distribution $\pi(\mathbf{y}'|\boldsymbol{\theta}')$ as an approximate draw

Importance sampling

MH ratio in the Exchange algorithm ($h(\cdot|\boldsymbol{\theta})$ symmetric)

$$\frac{q(\mathbf{y}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')}{q(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}\,\frac{q(\mathbf{y}'|\boldsymbol{\theta})}{q(\mathbf{y}'|\boldsymbol{\theta}')}$$

Standard MH ratio:

$$\frac{q(\mathbf{y}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')}{q(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}\,\frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\theta}')}$$

$q(\mathbf{y}'|\boldsymbol{\theta})/q(\mathbf{y}'|\boldsymbol{\theta}')$ importance sampling estimate of $z(\boldsymbol{\theta})/z(\boldsymbol{\theta}')$ since

$$\mathbb{E}_{\mathbf{y}'|\boldsymbol{\theta}'}\frac{q(\mathbf{y}'|\boldsymbol{\theta})}{q(\mathbf{y}'|\boldsymbol{\theta}')} = \sum_{y}\frac{q(\mathbf{y}'|\boldsymbol{\theta})}{q(\mathbf{y}'|\boldsymbol{\theta}')}\,\frac{q(\mathbf{y}'|\boldsymbol{\theta}')}{z(\boldsymbol{\theta}')} = \frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\theta}')}$$

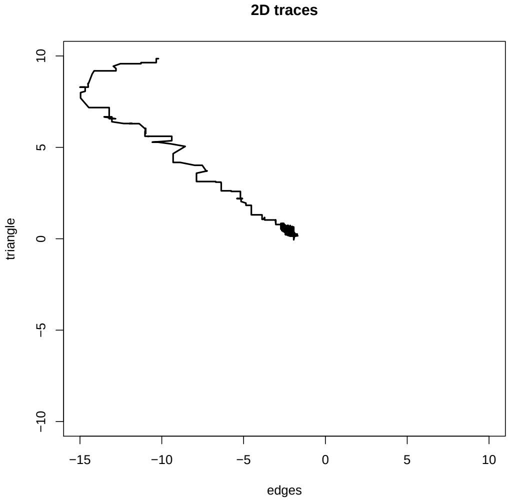
Mixing of the Markov chain

Assuming $\pi(\cdot)$ very flat and $h(\cdot)$ symmetric

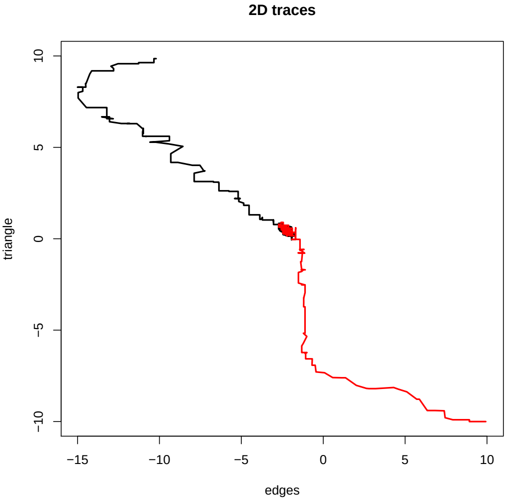
$$\log \alpha \approx \min \left\{ 0, (\boldsymbol{\theta} - \boldsymbol{\theta}')^t \underbrace{\left[s(\mathbf{y}') - s(\mathbf{y}) \right]}_{\textit{disparity measure}} \right\}$$

- Acceptance probability is high when $||s(\mathbf{y}') - s(\mathbf{y})||$ is close to 0
- A move of $\boldsymbol{\theta}'$ to the degenerate region usually produces a disproportionate increase of $||s(\mathbf{y}') - s(\mathbf{y})||$ thus reducing the probability of accepting the move
- If we reach high probability region, the algorithm would not tend to allow excursions into degenerate regions

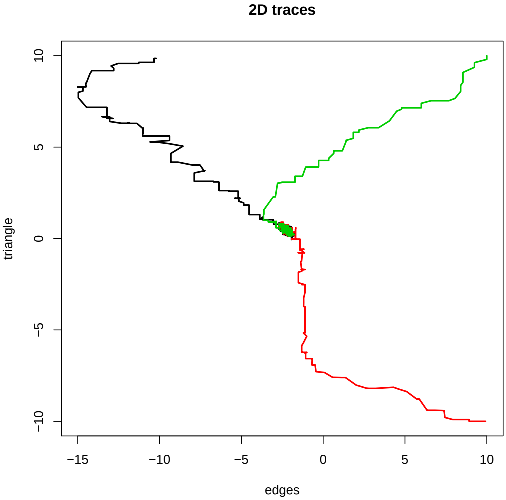
Mixing of the Markov chain (cont'd)



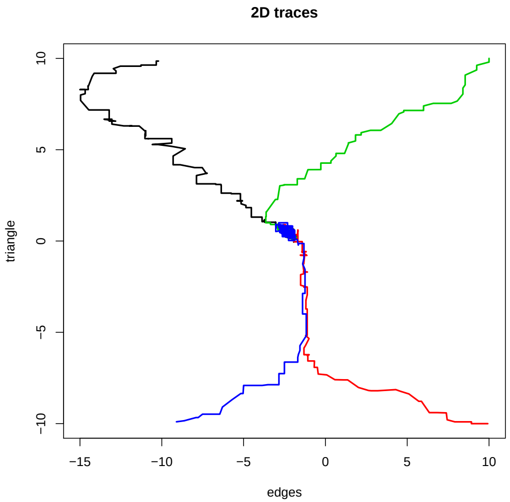
Mixing of the Markov chain (cont'd)



Mixing of the Markov chain (cont'd)



Mixing of the Markov chain (cont'd)



Connection with Approximate Bayesian Computation (ABC)

- Likelihood-free method handling distributions with intractable normalising constants
- Both rely on proposing new $\boldsymbol{\theta}'$ and simulating $\mathbf{y}'|\boldsymbol{\theta}'$
- Proposed move to $\boldsymbol{\theta}'$ is accepted if there is good agreement between auxiliary data and observed data in terms of summary statistics
- In ABC, $\boldsymbol{\theta}'$ is accepted if this distance is sufficiently small

Connection with Approximate Bayesian Computation (ABC)

- Likelihood-free method handling distributions with intractable normalising constants
- Both rely on proposing new $\boldsymbol{\theta}'$ and simulating $\mathbf{y}'|\boldsymbol{\theta}'$
- Proposed move to $\boldsymbol{\theta}'$ is accepted if there is good agreement between auxiliary data and observed data in terms of summary statistics
- In ABC, $\boldsymbol{\theta}'$ is accepted if this distance is sufficiently small
- Good approximation to the true posterior is guaranteed by the fact that the summary statistics are sufficient statistics of the probability model

Implementing the algorithm

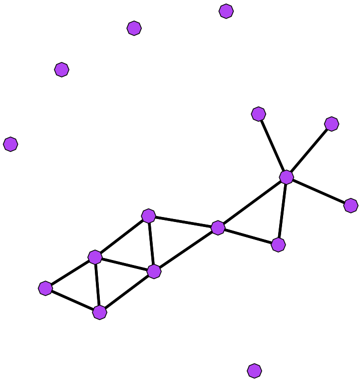
Implementation by existing software (eg **statnet** package for **R**)

```
Initialise:
y.obs       # observed graph
s(y.obs)    # observed statistics
theta       # initial value of the chain
loop        # no. of iterations
h           # (symmetric) proposal
prior       # prior distribution

for i in 0:loop
Single site update:
   draw theta' from h(theta[i])
   pr <- prior(theta') / prior(theta[i])
   simulate y' conditional on theta'
   delta <- s(y') - s(y.obs)
   alpha <- (theta[i] - theta') * delta + log(pr)
   u <- log RandomU(0,1)
        if(alpha >= u)
           set theta[i+1] <- theta'
        else
           set theta[i+1] <- theta[i]
```

Example 1: Florentine family business

Business relations between 16 families in around 1430



Two-star model

MC-MLE failed

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{z(\boldsymbol{\theta})} \exp\left\{ \theta_1 \sum_{i<j} y_{ij} + \theta_2 \sum_{i<j<k} y_{ik} y_{jk} \right\}$$

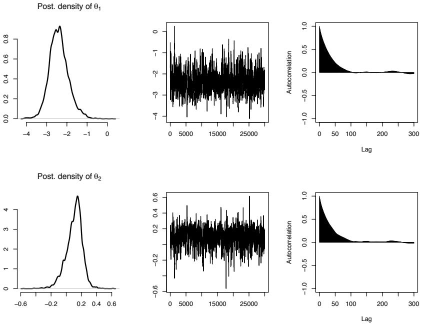
Exchange algorithm

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \exp\left\{ \theta_1 \sum_{i<j} y_{ij} + \theta_2 \sum_{i<j<k} y_{ik} y_{jk} \right\} \pi(\boldsymbol{\theta})$$

$$h(\cdot|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta}, \Psi) \quad \text{and} \quad \pi(\boldsymbol{\theta}) \sim \mathcal{N}(0, \sigma)$$

$$\Psi = \begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix} \quad \text{and} \quad \sigma = 30$$

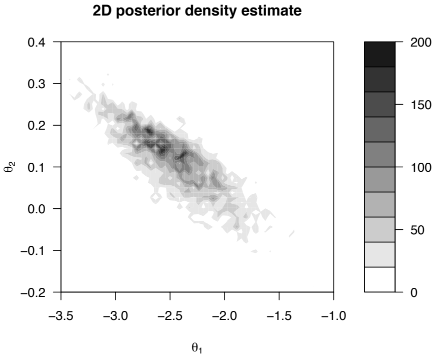
MCMC Output



Parameters	post. mean	post. sd
θ_1	2.40	0.47
θ_2	0.10	0.11

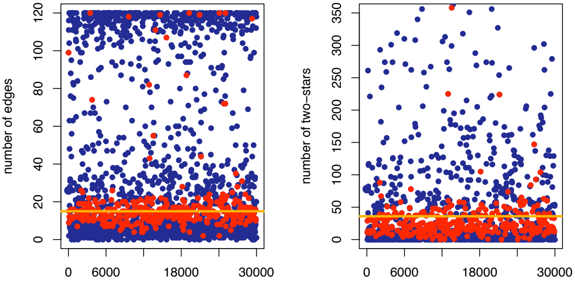
- Single-site Gibbs update, acceptance rates: 19% and 15%
- 30,000 iterations for the main chain, 5,000 iterations for the auxiliary chain

Remarks



- High correlation between the parameters
- Long time to explore the entire posterior distribution
- Slow mixing of the chain

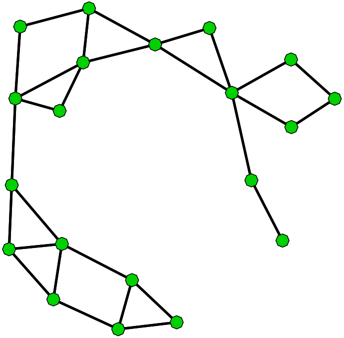
Remarks (cont'd)



- dots: $2,000$ proposed graphs (thinned by a factor of 15 from $30,000$)
- red dots: graphs whose parameters were accepted
- orange line: observed graph

Example 2: Synthetic network

Elongated-shaped graph of 20 nodes (**ergm** package for **R**)



4-dimensional model

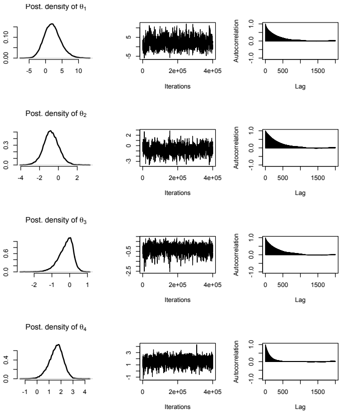
$$\pi(\mathbf{y}|\boldsymbol{\theta}) \propto \frac{1}{z(\boldsymbol{\theta})} \exp\left\{\sum_{i=1}^{4} \theta_i s_i(\mathbf{y})\right\} \pi(\boldsymbol{\theta})$$

$s_1(\mathbf{y}) = \sum_{i<j} y_{ij}$	number of edges
$s_2(\mathbf{y}) = \sum_{i<j<k} y_{ik} y_{jk}$	number of two-stars
$s_3(\mathbf{y}) = \sum_{i<j<k<l} y_{il} y_{jl} y_{kl}$	number of three-stars
$s_4(\mathbf{y}) = \sum_{i<j<k} y_{ik} y_{ik} y_{ij}$	number of triangles

$$h(\cdot|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta}, \Psi), \quad \pi(\boldsymbol{\theta}) \sim \mathcal{N}(0, \sigma)$$

$$\Psi = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.3 \end{bmatrix} \quad \text{and} \quad \sigma = 30$$

MCMC output



Parameters	post. mean	post. sd
θ_1	2.12	2.61
θ_2	-0.77	0.86
θ_3	-0.22	0.43
θ_4	1.59	0.59

- Single-site Gibbs update, acceptance rates around 10%

Population MCMC can improve mixing

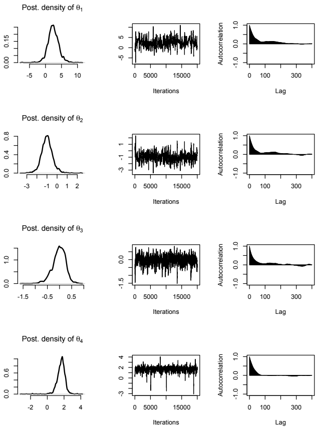
Here we consider a collection of chains which interact with one another.

State space: $\{(\theta^1, \theta^2, \ldots, \theta^n)\}$ with target distribution $\pi(\theta^1|\mathbf{y}) \otimes \cdots \otimes \pi(\theta^n|\mathbf{y})$.

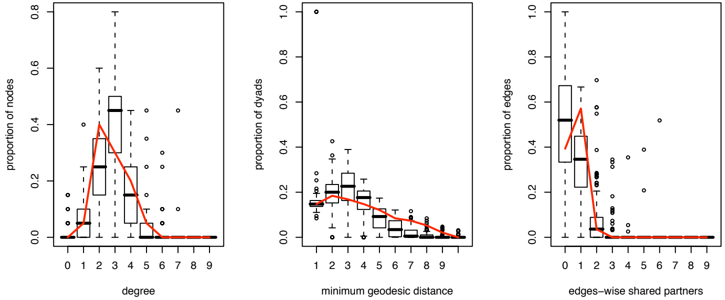
"snooker move" (at iteration i):

$$\theta^h_{i+1} = \theta^h_i + \gamma\left(\theta^h_i - \theta^{h+1}_i\right) + \epsilon \quad \gamma \sim N(0, \sigma_\gamma) \quad \epsilon \sim N(0, \sigma_\epsilon)$$

Output from population MCMC



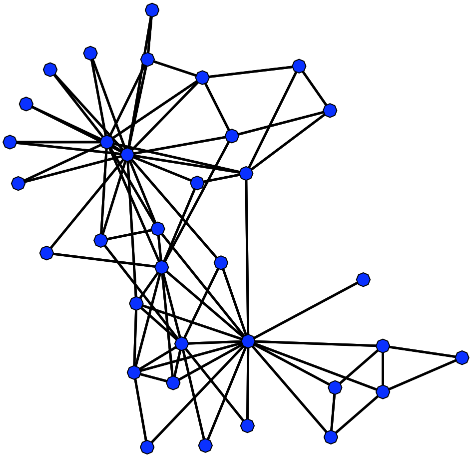
Goodness-of-fit test



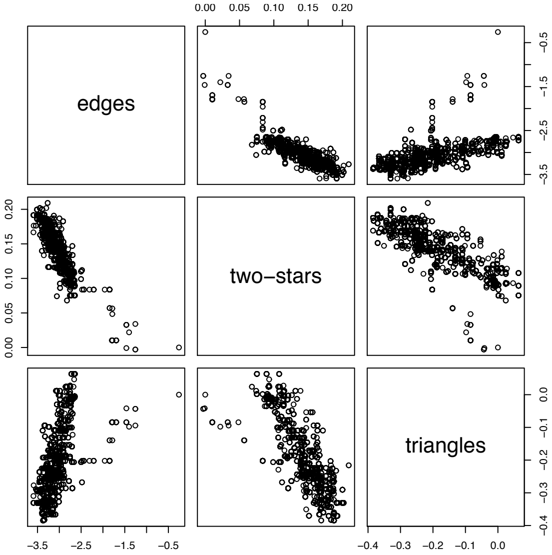
100 graphs are simulated from 100 realisations taken from the estimated posterior distribution and compared to the observed graph in terms of high-level characteristics

Example 3: Zachary karate club

34 members of a university karate club



High correlation, slow mixing



Model with new specifications statistics

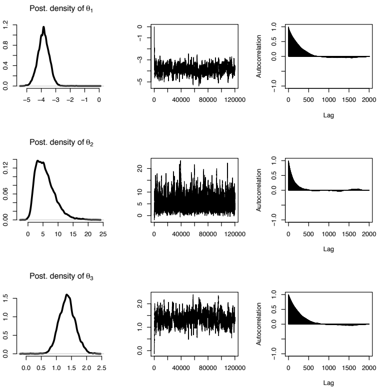
$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{1}{z(\boldsymbol{\theta})} \exp\left\{\theta_1 s_1(\mathbf{y}) + \theta_2 u(\mathbf{y},\phi) + \theta_3 v(\mathbf{y},\phi)\right\} \pi(\boldsymbol{\theta})$$

$s_1(\mathbf{y}) = \sum_{i<j} y_{ij}$	number of edges
$u(\mathbf{y},\phi) = e^{\phi} \sum_{i=1}^{n-1} \left\{1 - \left(1 - e^{-\phi}\right)^{i}\right\} D_i(\mathbf{y})$	GWD
$v(\mathbf{y},\phi) = e^{\phi} \sum_{i=1}^{n-2} \left\{1 - \left(1 - e^{-\phi}\right)^{i}\right\} EP_i(\mathbf{y})$	GWESP

$$h(\cdot|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta},\psi), \quad \pi(\boldsymbol{\theta}) \sim \mathcal{N}(0,\sigma)$$

$$\Psi = \begin{bmatrix} 0.7 & 0 & 0 \\ 0 & 1.2 & 0 \\ 0 & 0 & 0.07 \end{bmatrix}, \quad \sigma = 30, \quad \text{and} \quad \phi = 0.2$$

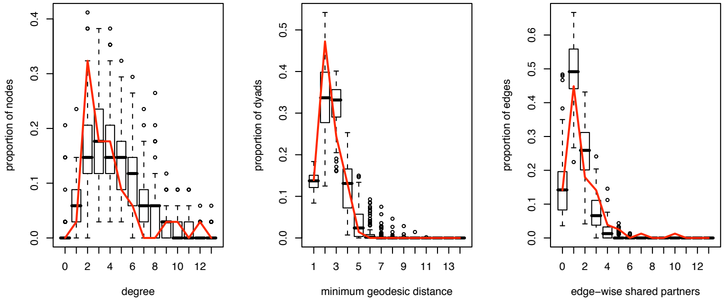
MCMC output



Parameters	post. mean	post. sd
θ_1	-3.84	0.38
θ_2	5.77	3.22
θ_3	1.35	0.26

- Single-site Gibbs update, acceptance rates: 14%, 16%, and 14%

Goodness-of-fit test



Summary

- A crucial aspect of MC-MLE is the choice of $\boldsymbol{\theta}_0$
- MC-MLE may fail due to a poorly choosen $\boldsymbol{\theta}_0$
- Model degeneracy is an important obstacle to estimation

Summary

- A crucial aspect of MC-MLE is the choice of $\boldsymbol{\theta}_0$
- MC-MLE may fail due to a poorly choosen $\boldsymbol{\theta}_0$
- Model degeneracy is an important obstacle to estimation

- Exchange algorithm overcomes the problem of the choice of $\boldsymbol{\theta}_0$
- Good approximation is guaranteed by the agreement between simulated and observed graphs in terms of sufficient statistics
- A thin and correlated support of the posterior can cause slow mixing of the chain (an appropriate design of the MCMC procedure can overcome this)
- Computationally intensive but it can be easily developed by existing software (eg **statnet** package for **R**)

References

- Besag J. (1974) *Spatial Interaction and the Statistical Analysis of Lattice Systems*. Journal of the Royal Statistical Society, B 36, 192 – 236
- Strauss D., and Ikeda M. (1990) *Pseudolikelihood estimation for social networks*. Journal of the American Statistical Association, 85, 204 – 212
- Geyer C.J., and Thompson E.A. (1992) *Constrained Monte Carlo maximum likelihood for dependent data*. Journal of the Royal Statistical Society, B 54, 657 – 699
- Handcock, Mark S. (2003) *Assessing Degeneracy in Statistical Models of Social Networks Center for Statistics and the Social Sciences*, University of Washington, Working Paper No. 39
- Murray I., Ghahramani, Z., and MacKay, D. (2006) *MCMC for doubly-intractable distributions*. In Proceedings of the 22nd annual conference on uncertainty in artificial intelligence
- Snijders, T.A.B., Pattison P., Robins G.L., Handcock M.S. (2006) *New Specifications for Exponential Random Graph Models*. Sociological Methodology, 36, 99 – 153
- Rinaldo A., Fienberg S.E., Zhou Y. (2009) *On the geometry of discrete exponential families with application to exponential random graph models*. Electronic Journal of Statistics, 3, 446 – 484
- Beaumont, M.A., Zhang, W., and Balding, D.J. (2002) *Approximate Bayesian Computation in Population Genetics*. Genetics 162 (4), 2025 â- 2035