Monte Carlo algorithms for Bayesian social network models

Alberto Caimo
alberto.caimo@usi.ch

Social Network Analysis Research Centre
InterDisciplinary Institute of Data Science
University of Lugano, Switzerland

AMS Sectional Meeting, Loyola University Chicago
October 2015

Outline

- Exponential random graph models (ERGMs)
- Bayesian exponential random graph models (BERGMs)
- Computational approaches for BERGMs
- Examples

Exponential random graph models (ERGMs)

- The relational structure of an observed network y can be explained by the relative prevalence of a set of overlapping sub-graph configurations $s(y)$ called network statistics

Exponential random graph models (ERGMs)

- The relational structure of an observed network y can be explained by the relative prevalence of a set of overlapping sub-graph configurations $s(y)$ called network statistics
- The likelihood of an ERGM represents the probability distribution of y given a parameter θ:

$$p(y|\theta) = \exp\{\theta^t s(y) - \gamma(\theta)\}$$

Exponential random graph models (ERGMs)

- The relational structure of an observed network y can be explained by the relative prevalence of a set of overlapping sub-graph configurations $s(y)$ called network statistics
- The likelihood of an ERGM represents the probability distribution of y given a parameter θ:

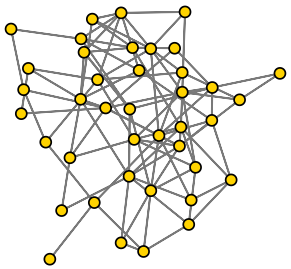
$$p(y|\theta) = \exp\{\theta^t s(y) - \gamma(\theta)\}$$

- From a computational point of view we have an **intractable likelihood problem**

Bayesian exponential random graph models (BERGMs)

(1)

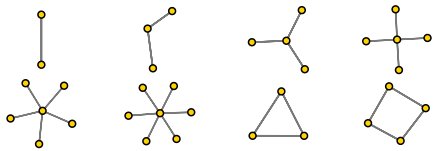
Observed network $\mathbf{y}$



(2)

Model specification

$p(\mathbf{y}|\boldsymbol{\theta}, m) \propto \exp\{\boldsymbol{\theta}^t s(\mathbf{y})\}$



(3)

Doubly-intractable posterior

$p(\boldsymbol{\theta}, m|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, m)\ p(\boldsymbol{\theta}|m)\ p(m)$

(4)

Parameter inference

$p(\boldsymbol{\theta}|\mathbf{y}, m)$

(5)

Model choice

across-model

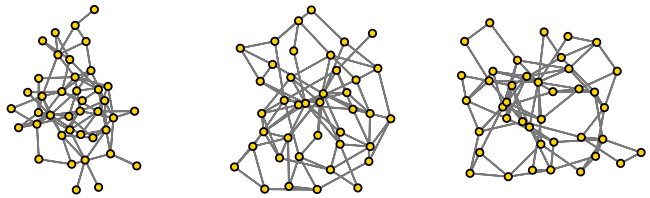
$p(m|\boldsymbol{\theta}, \mathbf{y})$

within-model

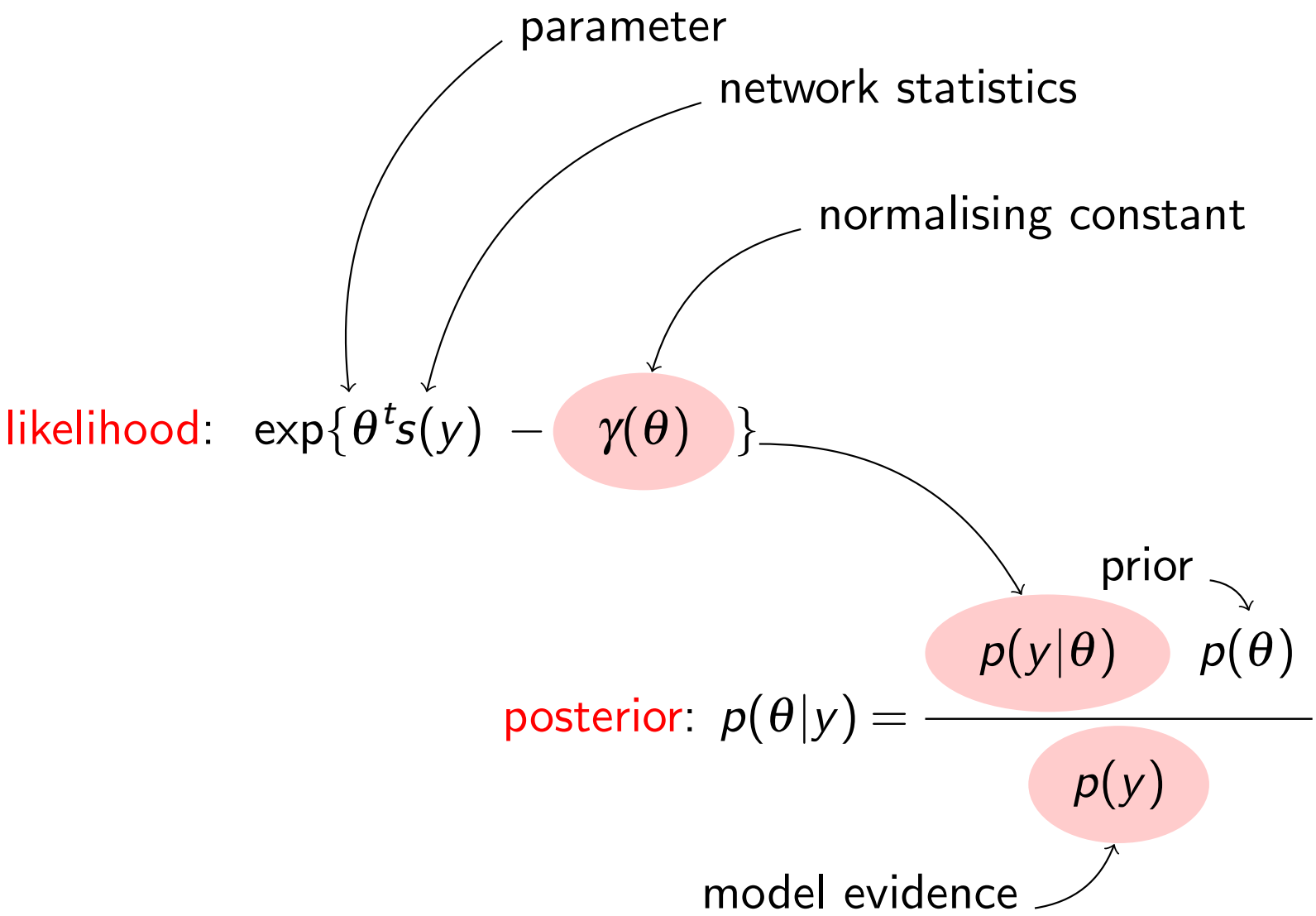
$p(\mathbf{y}|m)$

(6)

Goodness of fit



Bayesian exponential random graph models (BERGMs)



Computational approaches

Let's define:

- $q_\theta(y) = \exp\{\theta^t s(y)\}$ unnormalised likelihood
- $z(\theta) = \exp\{\gamma(\theta)\}$ normalising constant

Metropolis-Hastings

1 Gibbs update of (θ')
   i Draw $\theta' \sim h(\cdot|\theta)$

2 Accept move from θ to θ' with probability

$$1 \wedge \frac{q_{\theta'}(y)}{q_\theta(y)} \frac{p(\theta')}{p(\theta)} \frac{h(\theta|\theta')}{h(\theta'|\theta)} \times \frac{z(\theta)}{z(\theta')}$$

Approximate exchange algorithm (AEA)

(Murray et al., 2006; Caimo and Friel, 2011)

1 GIBBS UPDATE OF (θ', y')

(i) Draw $\theta' \sim h(\cdot|\theta)$

(ii) Draw $y' \sim p(\cdot|\theta')$ via MCMC

Approximate exchange algorithm (AEA)

(Murray et al., 2006; Caimo and Friel, 2011)

1 Gibbs update of (θ', y')

(i) Draw $\theta' \sim h(\cdot|\theta)$

(ii) Draw $y' \sim p(\cdot|\theta')$ via MCMC

2 Exchange move from θ to θ' with probability

$$1 \wedge \frac{q_{\theta'}(y)}{q_{\theta}(y)} \frac{p(\theta')}{p(\theta)} \frac{h(\theta|\theta')}{h(\theta'|\theta)} \frac{q_{\theta}(y')}{q_{\theta'}(y')} \times \underbrace{\frac{z(\theta)z(\theta')}{z(\theta)z(\theta')}}_{1}$$

Computational approaches

Computational challenges

- Posterior distribution $p(\theta|y)$ is difficult to sample efficiently from as ERGM parameters is typically very thin and highly correlated

Computational approaches

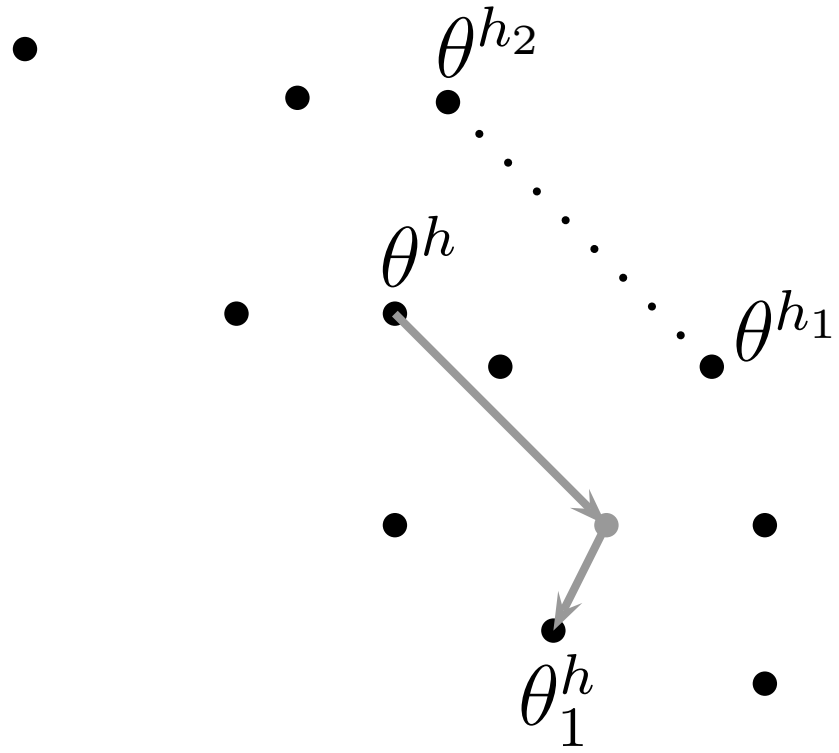
Improving chain mixing and convergence

(Caimo and Friel, 2011)

1(i) Parallel adaptive direction sampling (ADS) for Gibbs update of θ'

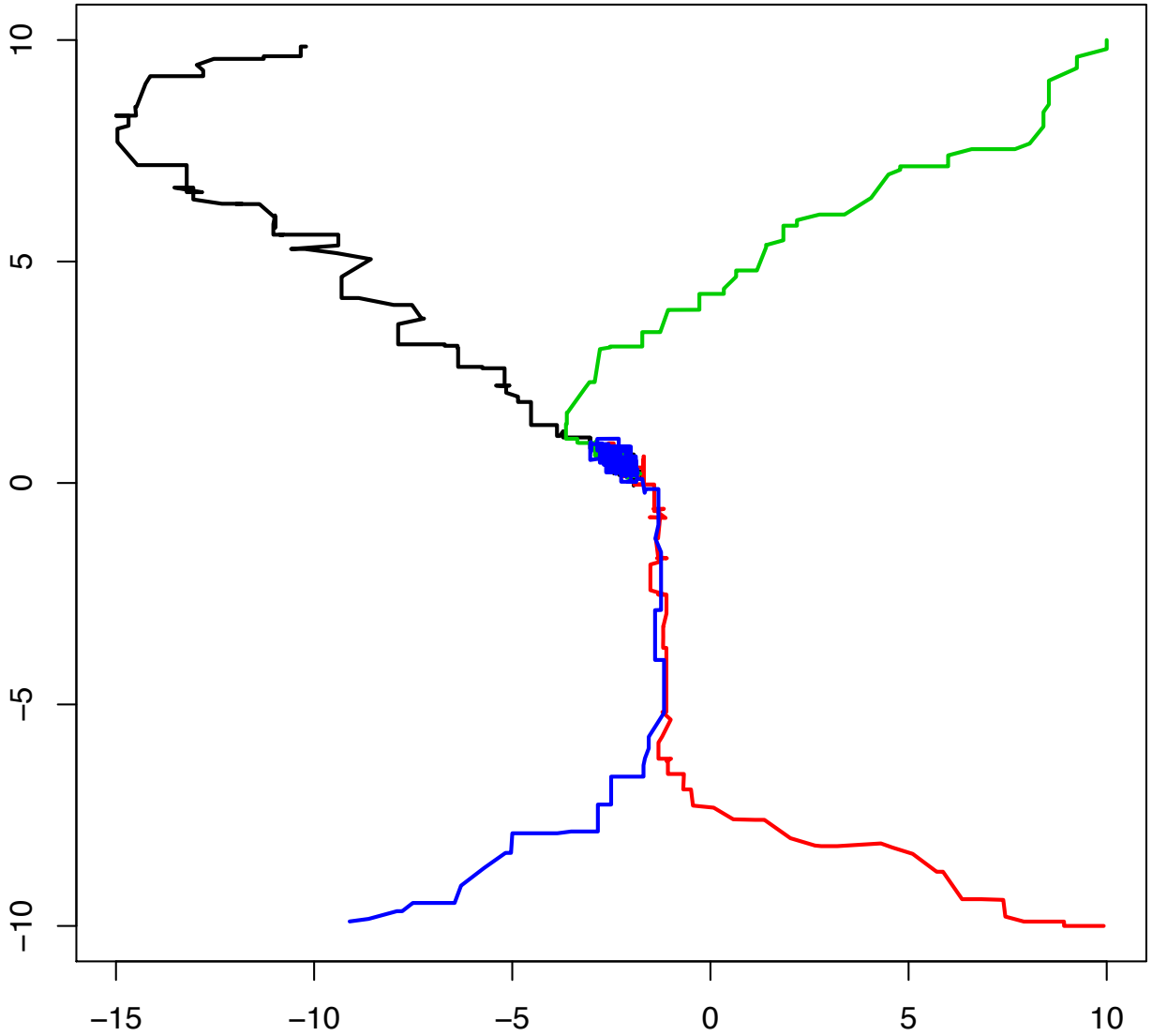
1(ii) Tie/no tie (TNT) sampler (as in the `ergm` package for `R`)

Computational approaches



ADS sampler: "snooker move"

Bayesian exponential random graph models (BERGMs)



Computational approaches

Adaptive approximate exchange algorithm with delayed rejection (Caimo and Mira, 2015)

1(i) Adaptive strategies

Computational approaches

Adaptive approximate exchange algorithm with delayed rejection (Caimo and Mira, 2015)

1(i) Adaptive strategies

2 Approximate exchange algorithm with delayed rejection

Computational approaches

Adaptive strategies for Gibbs update of θ'

(Roberts and Rosenthal, 2007; Haario et al., 2001)

- **vertical** adaptation: all past particles along the same chain (AAEA-1+DR)
- **horizontal** adaptation: all particles at the current time for all chains (AAEA-2+DR)
- **rectangular** adaptation: particles from all chains and all past simulations (AAEA-3+DR)

Adaptive exchange algorithm with delayed rejection

- First stage move:

$$\alpha_1(\theta, \theta') = 1 \wedge \frac{q_{\theta'}(y)}{q_\theta(y)} \frac{p(\theta')}{p(\theta)} \frac{h_1(\theta|\theta')}{h_1(\theta'|\theta)} \frac{q_\theta(y')}{q_{\theta'}(y')}$$

Adaptive exchange algorithm with delayed rejection

- First stage move:

$$\alpha_1(\theta, \theta') = 1 \wedge \frac{q_{\theta'}(y)}{q_\theta(y)} \frac{p(\theta')}{p(\theta)} \frac{h_1(\theta|\theta')}{h_1(\theta'|\theta)} \frac{q_\theta(y')}{q_{\theta'}(y')}$$

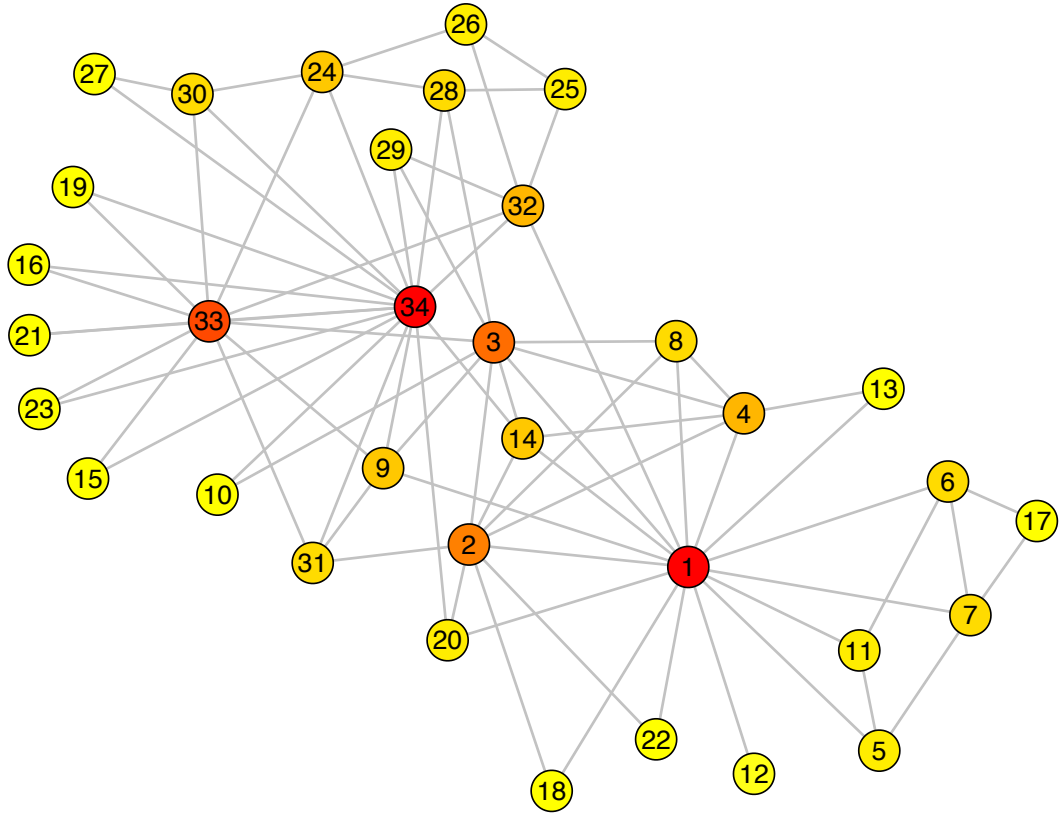
- If θ' rejected, try a second stage move to θ'' with probability

$$\alpha_2(\theta, \theta', \theta'') =$$
$$1 \wedge \frac{q_\theta(y'')\, p(\theta'')\, h_1(\theta'|\theta'')\, [1 - \alpha_1(\theta'', \theta')]\, h_2(\theta|\theta'', \theta')\, q_{\theta''}(y)}{q_\theta(y)\, p(\theta)\, h_1(\theta'|\theta)\, [1 - \alpha_1(\theta, \theta')]\, h_2(\theta''|\theta, \theta')\, q_{\theta''}(y')}$$

Computational approaches

- A hierarchy of proposal distributions can be exploited
- Moves (tennis-service strategy): "first bold" $h_1(\cdot)$ proposal versus "second timid" $h_2(\cdot)$

Examples



Zachary Karate Club Network: Friendship relations between 34 members of a karate club at a US university in the 1970

Examples

ERGM: $p(y|\theta) \propto \exp\{\theta_1\, s_1(y) + \theta_2\, s_2(y, \phi_u) + \theta_3\, s_3(y, \phi_v)\}$

Examples

ERGM: $p(y|\theta) \propto \exp\{\theta_1\, s_1(y) + \theta_2\, s_2(y,\phi_u) + \theta_3\, s_3(y,\phi_v)\}$
where:

$s_1(y) = \sum_{i<j} y_{ij}$ = number of edges

$s_2(y,\phi_v) = e^{\phi_v} \sum_{i=1}^{n-2} \left\{1 - \left(1 - e^{-\phi_v}\right)^i\right\} EP_i(y)$

geometrically weighted edgewise shared partners (gwesp)

$s_3(y,\phi_u) = e^{\phi_u} \sum_{i=1}^{n-1} \left\{1 - \left(1 - e^{-\phi_u}\right)^i\right\} D_i(y)$

geometrically weighted degrees (gwdegree)

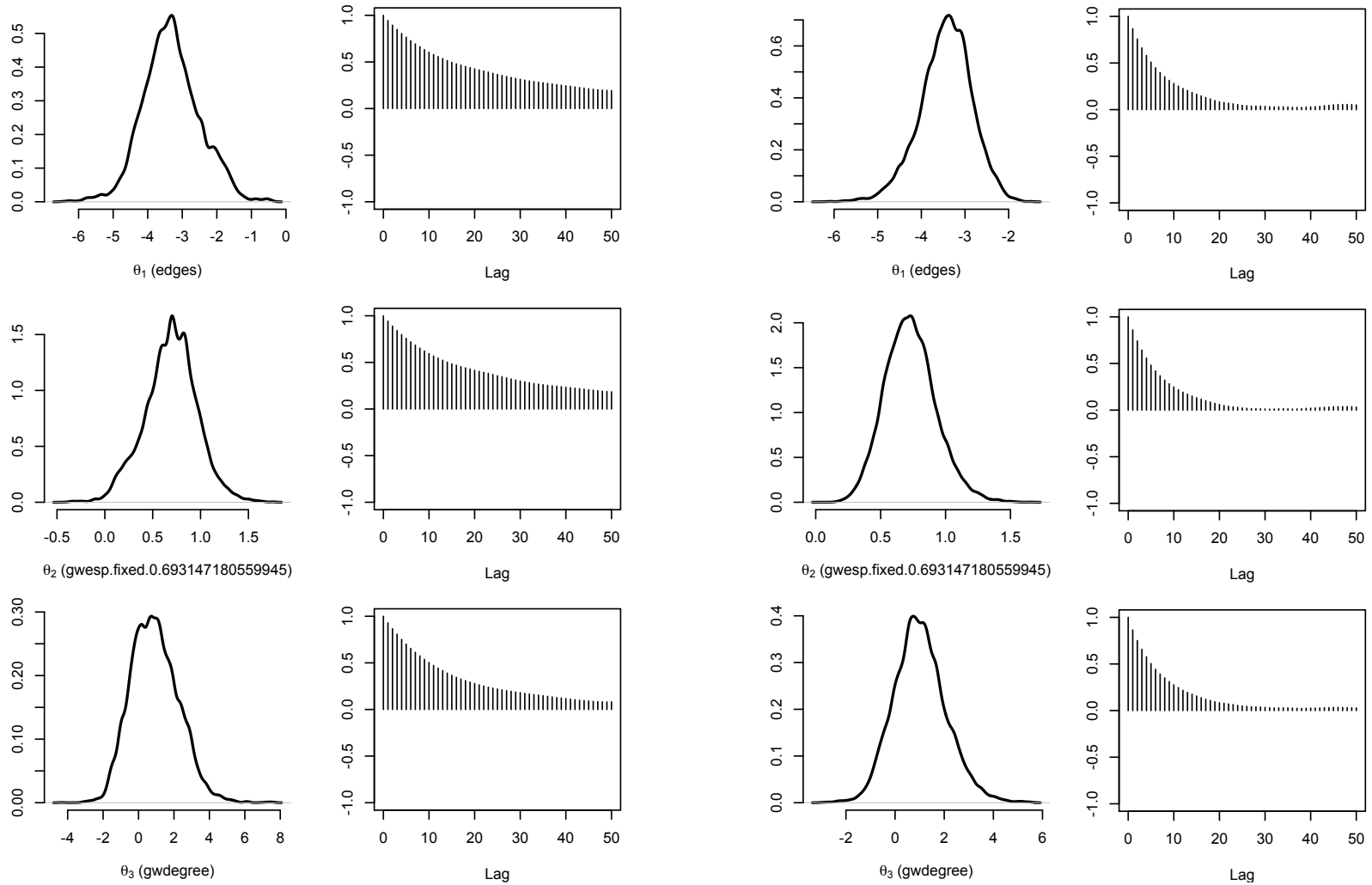
$EP_i(y)$ = distribution of the number of unordered pairs of connected nodes having exactly k common neighbours
$D_i(y)$ = degree distribution

Examples

	$\theta^{(1)}$ (edges)	$\theta^{(2)}$ (gwesp)	$\theta^{(3)}$ (gwdegree)
ADS-AEA			
Post. mean	−3.51	0.74	1.18
Post. sd	0.62	0.21	1.12
AAEA-2+DR (horizontal adaptation + DR)			
Post. mean	−3.44	0.72	1.01
Post. sd	0.59	0.21	1.07

Estimated posterior means and standard deviations

Examples



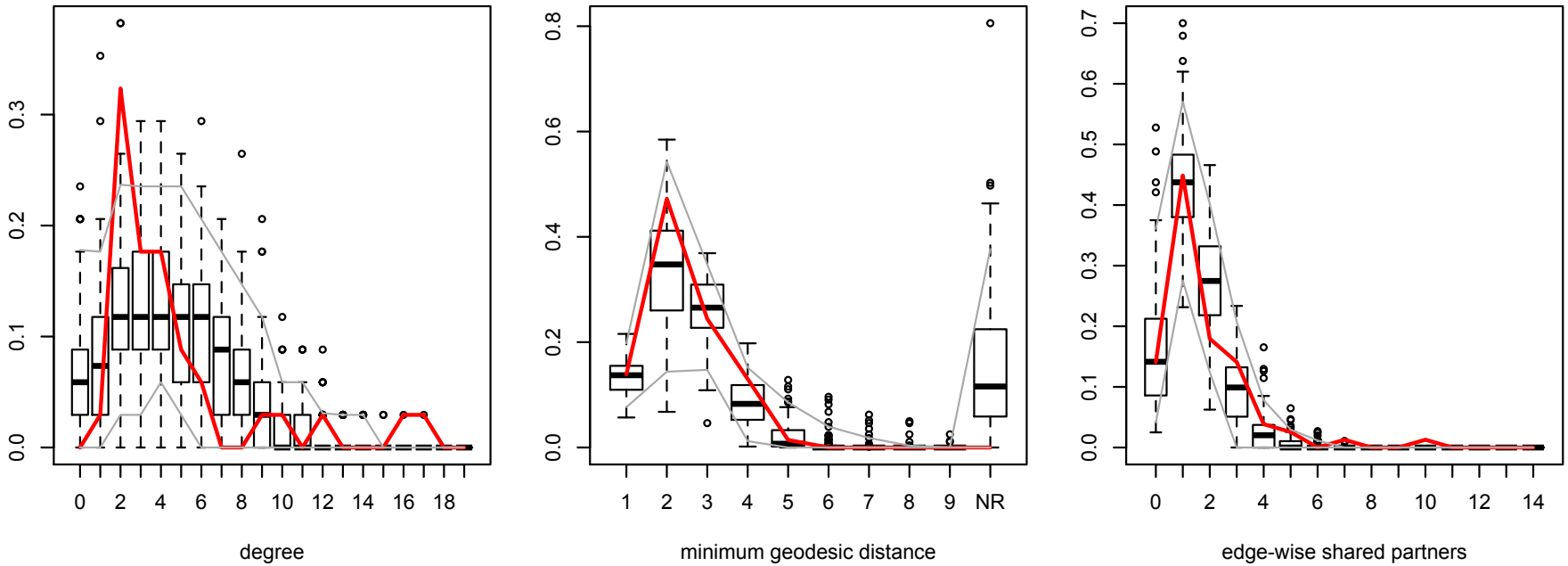
Posterior density estimates for AEA (left) and AAEA+DR (right)

Examples

- **Effective sample size** (ESS):
  AAEA-2+DR +70%

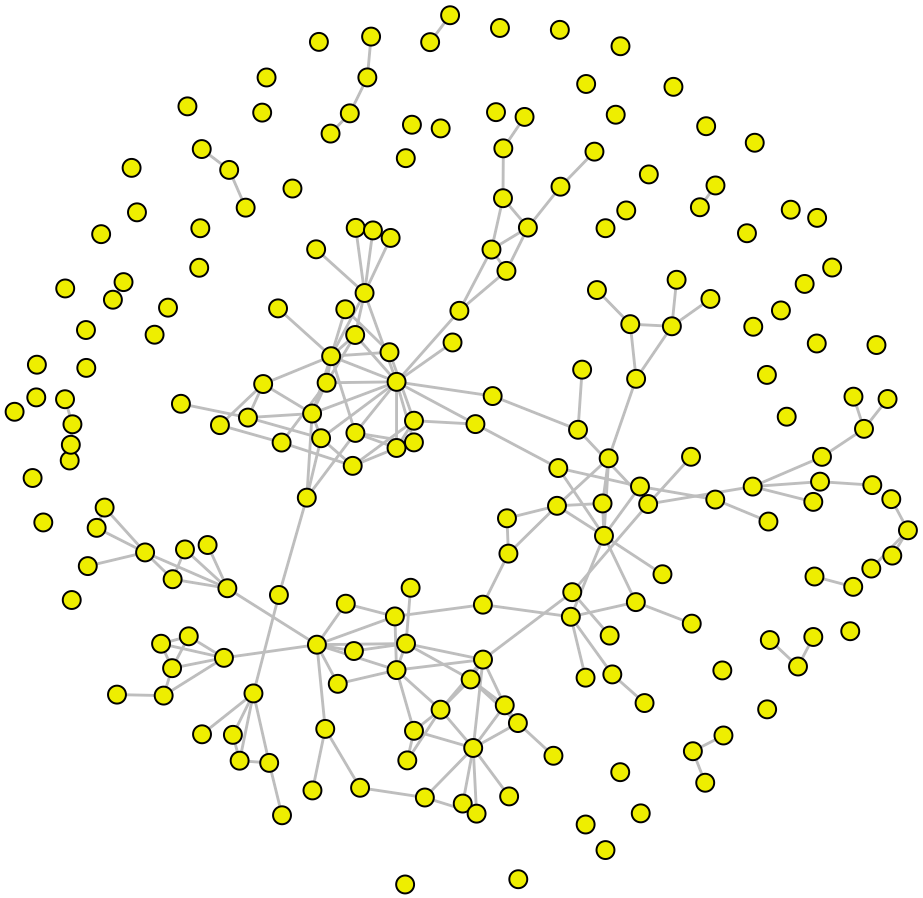
- **Performance** (= ESS/CPU time):
  AAEA-2+DR +40%

Examples



Bayesian goodness-of-fit diagnostics: the observed network y is compared with a set of networks simulated from independent realisations of $p(\theta|y)$ in terms of high-level network statistics

Examples



Faux Mesa High School friendship 203-node network graph.

Examples

ERGM specification:

$$s_1(y) = \sum_{i<j} y_{ij} \text{ number of edges}$$

$$s_2(y,x) = \sum_{i<j} y_{ij}(\mathbb{1}_{(grade_i=8)} + \mathbb{1}_{(grade_j=8)})$$
node factor for "grade" = 8

$$s_3(y,x) = \sum_{i<j} y_{ij}(\mathbb{1}_{(grade_i=9)} + \mathbb{1}_{(grade_j=9)})$$
node factor for "grade" = 9

$$s_4(y,x) = \sum_{i<j} y_{ij}(\mathbb{1}_{(grede_i=10)} + \mathbb{1}_{(grade_j=10)})$$
node factor for "grade" = 10

$$s_5(y,x) = \sum_{i<j} y_{ij}(\mathbb{1}_{(grade_i=11)} + \mathbb{1}_{(grade_j=11)})$$
node factor for "grade" = 11

$$s_6(y,x) = \sum_{i<j} y_{ij}(\mathbb{1}_{(grade_i=12)} + \mathbb{1}_{(grade_j=12)})$$
node factor for "grade" = 12

$$s_7(y,x) = \sum_{i<j} y_{ij}(\mathbb{1}_{(sex_i=M)} + \mathbb{1}_{(sex_j=M)})$$
node factor for "sex = male"

$$s_8(y) = v(y,\phi_v) \text{ GWESP}$$

$$s_9(y) = u(y,\phi_u) \text{ GWD}$$

Examples

	ADS-AEA	AAEA-1	AAEA-2	AAEA-3
ESS	667	1041	1008	1094
Performance (per sec)	1.8	2.3	2.1	2.2
	ADS-AEA+DR	AAEA-1+DR	AAEA-2+DR	AAEA-3+DR
ESS	873	1376	1320	1440
Performance (per sec)	1.4	2.6	2.6	2.6

Variance reduction of AAEA-based algorithms varies between 55% and 98% relative to the ADS-AEA.

This translates into a better performance varying from 25% to 40%.