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
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 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 $\theta$:

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

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

From a computational point of view we have an intractable likelihood problem.
Bayesian exponential random graph models (BERGMs)

(1) Observed network $y$

(2) Model specification

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

(3) Doubly-intractable posterior

$$p(\theta, m|y) \propto p(y|\theta, m) p(\theta|m) p(m)$$

(4) Parameter inference

$$p(\theta|y, m)$$

(5) Model choice

- across-model
  $$p(m|\theta, y)$$
- within-model
  $$p(y|m)$$

(6) Goodness of fit
Bayesian exponential random graph models (BERGMs)

- Parameter
- Network statistics
- Normalising constant

Likelihood: \( \exp\left\{ \theta^t s(y) - \gamma(\theta) \right\} \)

Posterior: \( p(\theta | y) = \frac{p(y | \theta) \cdot p(\theta)}{p(y)} \)

Model evidence
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 \((\theta')\)**
   - Draw \( \theta' \sim h(\cdot | \theta) \)

2. **Accept move from \( \theta \) to \( \theta' \) with probability**

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

Alberto Caimo

Bayesian computation for ERGMs
Approximate exchange algorithm (AEA)
(Murray et al., 2006; Caimo and Friel, 2011)

1 GIBBS UPDATE OF \((\theta', y')\)
   (i) Draw \(\theta' \sim h(\cdot|\theta)\)
   (ii) Draw \(y' \sim p(\cdot|\theta')\) via MCMC
Computational approaches

Approximate exchange algorithm (AEA)
(Murray et al., 2006; Caimo and Friel, 2011)

1 **Gibbs update of** $(\theta', y')$
   
   (i) Draw $\theta' \sim h(\cdot|\theta)$
   (ii) Draw $y' \sim p(\cdot|\theta')$ via MCMC

2 **Exchange move from** $\theta$ **to** $\theta'$ **with probability**

$$1 \wedge \frac{q_{\theta'}(y)}{q_\theta(y)} \frac{p(\theta')}{p(\theta)} \frac{h(\theta|\theta')}{h(\theta'|\theta)} \frac{q_\theta(y')}{q_{\theta'}(y')} \times \frac{z(\theta)z(\theta')}{z(\theta')z(\theta') \times 1}$$
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 $\theta'$
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)
Adaptive approximate exchange algorithm with delayed rejection (Caimo and Mira, 2015)

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

1(i) Adaptive strategies
2 Approximate exchange algorithm with delayed rejection
Adaptive strategies for Gibbs update of $\theta'$
(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:

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

Adaptive exchange algorithm with delayed rejection

- First stage move:

\[ \alpha_1(\theta, \theta') = 1 \land \frac{q_{\theta'}(y)}{q_{\theta}(y)} \frac{p(\theta')}{p(\theta)} \frac{h_1(\theta|\theta')}{h_1(\theta'|\theta)} \frac{q_{\theta}(y')}{q_{\theta'}(y')} \]

- If \( \theta' \) rejected, try a second stage move to \( \theta'' \) with probability

\[ \alpha_2(\theta, \theta', \theta'') = \]

\[ 1 \land \frac{q_{\theta}(y'')}{q_{\theta}(y)} \frac{p(\theta'')}{p(\theta)} \frac{h_1(\theta'|\theta'')}{h_1(\theta|\theta)} \frac{[1 - \alpha_1(\theta'', \theta')]}{[1 - \alpha_1(\theta, \theta')]} \frac{h_2(\theta|\theta'', \theta')}{h_2(\theta''|\theta, \theta')} \frac{q_{\theta''}(y)}{q_{\theta'}(y')} \]
A hierarchy of proposal distributions can be exploited

- Moves (tennis-service strategy): “first bold” $h_1(\cdot)$ proposal versus “second timid” $h_2(\cdot)$
Zachary Karate Club Network: Friendship relations between 34 members of a karate club at a US university in the 1970s.
**Examples**

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

- \( s_1(y) = \sum_{i<j} y_{ij} \) - number of edges
- \( s_2(y, \phi_u) = e_{\phi_u} \sum_{i=1}^{n-2} \{ 1 - (1 - e - \phi_u)^i \} \) - geometrically weighted edgewise shared partners (gwesp)
- \( s_3(y, \phi_v) = e_{\phi_v} \sum_{i=1}^{n-1} \{ 1 - (1 - e - \phi_v)^i \} \) - geometrically weighted degrees (gwdegree)

**Alberto Caimo**

Bayesian computation for ERGMs
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:

\[
\begin{align*}
\text{s}_1(y) &= \sum_{i<j} y_{ij} = \text{number of edges} \\
\text{s}_2(y, \phi_v) &= e^{\phi_v} \sum_{i=1}^{n-2} \left\{ 1 - (1 - e^{-\phi_v})^i \right\} EP_i(y) \\
&\quad \text{geometrically weighted edgewise shared partners (gwesp)} \\
\text{s}_3(y, \phi_u) &= e^{\phi_u} \sum_{i=1}^{n-1} \left\{ 1 - (1 - e^{-\phi_u})^i \right\} D_i(y) \\
&\quad \text{geometrically weighted degrees (gwdegree)}
\end{align*}
\]

\( EP_i(y) = \text{distribution of the number of unordered pairs of connected nodes having exactly } k \text{ common neighbours} \)

\( D_i(y) = \text{degree distribution} \)
### Estimated posterior means and standard deviations

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$\theta^{(1)}$ (edges)</th>
<th>$\theta^{(2)}$ (gwesp)</th>
<th>$\theta^{(3)}$ (gwdegree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS-AEA</td>
<td>-3.51</td>
<td>0.74</td>
<td>1.18</td>
</tr>
<tr>
<td>Post. mean</td>
<td>-3.44</td>
<td>0.72</td>
<td>1.01</td>
</tr>
<tr>
<td>Post. sd</td>
<td>0.62</td>
<td>0.21</td>
<td>1.12</td>
</tr>
<tr>
<td>AAEA-2+DR (horizontal adaptation + DR)</td>
<td>-3.44</td>
<td>0.72</td>
<td>1.01</td>
</tr>
<tr>
<td>Post. mean</td>
<td>-3.44</td>
<td>0.72</td>
<td>1.01</td>
</tr>
<tr>
<td>Post. sd</td>
<td>0.59</td>
<td>0.21</td>
<td>1.07</td>
</tr>
</tbody>
</table>
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%
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.
ERGM specification:

\[ s_1(y) = \sum_{i<j} y_{ij} \text{ number of edges} \]
\[ s_2(y, x) = \sum_{i<j} y_{ij} \left( \mathbb{1}_{(\text{grade}_i=8)} + \mathbb{1}_{(\text{grade}_j=8)} \right) \]
node factor for “grade” = 8
\[ s_3(y, x) = \sum_{i<j} y_{ij} \left( \mathbb{1}_{(\text{grade}_i=9)} + \mathbb{1}_{(\text{grade}_j=9)} \right) \]
node factor for “grade” = 9
\[ s_4(y, x) = \sum_{i<j} y_{ij} \left( \mathbb{1}_{(\text{grade}_i=10)} + \mathbb{1}_{(\text{grade}_j=10)} \right) \]
node factor for “grade” = 10
\[ s_5(y, x) = \sum_{i<j} y_{ij} \left( \mathbb{1}_{(\text{grade}_i=11)} + \mathbb{1}_{(\text{grade}_j=11)} \right) \]
node factor for “grade” = 11
\[ s_6(y, x) = \sum_{i<j} y_{ij} \left( \mathbb{1}_{(\text{grade}_i=12)} + \mathbb{1}_{(\text{grade}_j=12)} \right) \]
node factor for “grade” = 12
\[ s_7(y, x) = \sum_{i<j} y_{ij} \left( \mathbb{1}_{(\text{sex}_i=M)} + \mathbb{1}_{(\text{sex}_j=M)} \right) \]
node factor for “sex = male”
\[ s_8(y) = v(y, \phi_v) \text{ GWESP} \]
\[ s_9(y) = u(y, \phi_u) \text{ GWD} \]
### Table 8
**Faux Mesa High School network - Posterior parameter estimates.**

<table>
<thead>
<tr>
<th></th>
<th>ADS-AEA</th>
<th>AAEA-1</th>
<th>AAEA-2</th>
<th>AAEA-3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Post. mean</strong></td>
<td>-5.53</td>
<td>-0.15</td>
<td>-0.09</td>
<td>-0.04</td>
</tr>
<tr>
<td><strong>Post. sd</strong></td>
<td>0.33</td>
<td>0.15</td>
<td>0.17</td>
<td>0.21</td>
</tr>
</tbody>
</table>

### Table 9
**Faux Mesa High School network - ESS and performance for each algorithm based on 10 simulations.**

<table>
<thead>
<tr>
<th></th>
<th>ADS-AEA</th>
<th>AAEA-1</th>
<th>AAEA-2</th>
<th>AAEA-3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ESS</strong></td>
<td>667</td>
<td>1041</td>
<td>1008</td>
<td>1094</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>1.8</td>
<td>2.3</td>
<td>2.1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ADS-AEA+DR</th>
<th>AAEA-1+DR</th>
<th>AAEA-2+DR</th>
<th>AAEA-3+DR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ESS</strong></td>
<td>873</td>
<td>1376</td>
<td>1320</td>
<td>1440</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>1.4</td>
<td>2.6</td>
<td>2.6</td>
<td>2.6</td>
</tr>
</tbody>
</table>

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%.

### 12 Alberto Caimo, Antonietta Mira

### Table 10
**Faux Mesa High School network - Posterior correlation matrix between the parameters.**

<table>
<thead>
<tr>
<th></th>
<th>ADS-AEA</th>
<th>AAEA-1</th>
<th>AAEA-2</th>
<th>AAEA-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓(1)</td>
<td>✓(2)</td>
<td>✓(3)</td>
<td>✓(4)</td>
</tr>
<tr>
<td></td>
<td>✓(5)</td>
<td>✓(6)</td>
<td>✓(7)</td>
<td>✓(8)</td>
</tr>
<tr>
<td></td>
<td>✓(9)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the results displayed in Table 8 we can conclude that the network is very sparse (✓) and that students having the same gender seem to create friendship connections (✓). The transitivity effect expressed by (8) and the popularity effect expressed by (9) are important features of the network.

### 9 Conclusions
The exchange algorithm of Murray et al (2006) makes the computation of the MH acceptance probability feasible even for target distributions whose normalizing constant depends on the parameter of interest (doubly intractable problems).

The approximate exchange algorithm, due to Caimo and Friel (2011), modifies the original exchange algorithm and makes it applicable also in settings where sampling from the assumed data generating process is not feasible. This is the case for exponential random graphs the model we focus on in this paper.

The delayed rejection strategy allows to locally adapt the proposal distribution within each sweep of a MH algorithm at the cost of additional computational time.

The adaptive random walk proposal of Haario et al (2001) revised by Roberts and Rosenthal (2009) allows for global adaptation between MH iterations. This learning from the past process is also expensive from a computational point of view.

These three ingredients are combined in different ways within the approximate exchange algorithm (AEA) to avoid the computation of intractable normalising constant that appears in exponential random graph models. This gives rise to the AEA+DR: a new methodology to sample doubly intractable target distributions which gives rise to the AEA+DR: a new methodology to sample doubly intractable target distributions which

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%.