Markov chain Monte Carlo exact inference for social networks

John W. McDonald, Peter W. F. Smith and Jonathan J. Forster

Southampton Statistical Sciences Research Institute, University of Southampton,
Southampton, SO17 1BJ, United Kingdom

Abstract

We propose modifications to existing Markov chain Monte Carlo algorithms to generate from the conditional distribution of an adjacency matrix, given the in-degrees, the out-degrees and the number of mutual dyads. We compare our results with those obtained by using various approximations.

Keywords: Adjacency matrices; Exact conditional test; Metropolis-Hastings algorithm; Reciprocity; Triad census
Markov chain Monte Carlo exact inference for social networks

Abstract

We propose modifications to existing Markov chain Monte Carlo algorithms to generate from the conditional distribution of an adjacency matrix, given the in-degrees, the out-degrees and the number of mutual dyads. We compare our results with those obtained by using various approximations.

Keywords: Adjacency matrices; Exact conditional test; Metropolis-Hastings algorithm; Reciprocity; Triad census

1 Introduction

Inference concerning a social network of $g$ actors is often based on $U\{X_{ij}\}$, the conditional uniform distribution of the corresponding adjacency matrix $X = \{X_{ij}\}$ given the in-degrees and the out-degrees. For example, an exact conditional test of reciprocity in the $p_1$ model (Holland and Leinhardt, 1981) requires the conditional distribution of $M$, the number of mutual dyads, given the in-degrees and the out-degrees. Snijders (1991) enumerated this distribution for $g$ equal to 7. For larger $g$ where enumeration is infeasible, he proposed an importance sampling algorithm. Rao, Jana and Bandyopadhyay (1996) and Roberts (2000) proposed Markov chain Monte Carlo algorithms for generating adjacency matrices with given in-degrees and out-degrees, and we present what is essentially Roberts’s algorithm in Section 2.1.

The conditional uniform distribution given the in-degrees, the out-degrees and the number of mutual dyads, $U\{M, \{X_{ij}\}\}$, can be used to assess the group structure of the network, by examining the distribution of the triad census. Enumeration of this distribution is also typically infeasible, and Wasserman (1977) uses various approximations; see Section 3 for details. In Section 2.2 we describe how the Markov chain Monte Carlo algorithm presented in Section 2.1 can be modified to generate realizations from $U\{M, \{X_{ij}\}\}$. In Section 3 we present two examples of its use. Further modifications to the algorithm are briefly discussed in Section 4.

2 Metropolis-Hastings sampling

Metropolis-Hastings sampling (Hastings, 1970) is a Markov chain Monte Carlo method for generating samples from arbitrary multivariate distributions. The procedure is as follows.
1. Given current value \( \mathbf{x} \), generate a new value \( \mathbf{x}' \) from some probability distribution \( q(\mathbf{x}, \mathbf{x}') \).

2. Accept \( \mathbf{x}' \) as the next realization of the chain with probability

\[
a(\mathbf{x}, \mathbf{x}') = \min\left\{ \frac{f(\mathbf{x}')q(\mathbf{x}', \mathbf{x})}{f(\mathbf{x})q(\mathbf{x}, \mathbf{x}')}, 1 \right\};
\]

otherwise, retain \( \mathbf{x} \).

If the proposal distribution \( q \) is chosen appropriately, then \( f \) is the stationary distribution for this Markov chain. See Smith and Roberts (1993) and references therein for details.

2.1 \( U|\{X_{1i}\}, \{X_{+j}\} \)

For \( U|\{X_{1i}\}, \{X_{+j}\} \), it is sensible to choose \( q \) to ensure that the margins of \( \mathbf{x}' \) are the same as those for \( \mathbf{x} \). Also it is desirable to ensure that it is possible to obtain every possible \( \mathbf{x} \) from every other \( \mathbf{x} \) by a sequence of updates, i.e., to ensure that the Markov chain is irreducible. Other important considerations are that \( q \) should be easy to generate from and the chain should mix well.

A simple algorithm, which ensures that the margins of \( \mathbf{x}' \) are the same as those for \( \mathbf{x} \), updates \( \mathbf{x} \) as follows.

1. Select a random rectangle defined by the intersection of two distinct rows (\( i_1 \) and \( i_2 \))
   and two distinct columns (\( j_1 \) and \( j_2 \)) of \( \mathbf{x} \), which does not contain a diagonal element from \( \mathbf{x} \). Smith, Forster and McDonald (1996) refer to such a rectangle as a tetrad, but here we adopt the terminology of Rao et al. (1996) to avoid confusion with a group of four actors.

2. If the selected rectangle is \[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\]
   replace it with \[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]
   to yield \( \mathbf{x}' \);

if the selected rectangle is \[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]
   replace it with \[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\]
   to yield \( \mathbf{x}' \);

otherwise \( \mathbf{x}' = \mathbf{x} \).

Unfortunately, as proved by Rao et al. (1996), updating rectangles in this manner is not sufficient to ensure that the Markov chain is irreducible. Hence, a Metropolis-Hasting algorithm based on updating rectangles will not necessarily produced realizations from the correct conditional distribution. Note that Rao et al. (1996) called the two rectangles that can be replaced in Step 2 of the algorithm above alternating rectangles. Rao et al. (1996) also proved that the chain is irreducible if updates are made using what they called compact
alternating hexagons, as well as alternating rectangles. They defined compact alternating hexagons to be an ordered set of six cells of the type \( \{i_1 i_2, i_1 i_3, i_2 i_3, i_2 i_1, i_3 i_1, i_3 i_2\} \) with entries 1, 0, 1, 0, 1, 0 respectively, where \( i_1, i_2, i_3 \) are distinct.

For our Markov chain we use a richer set of hexagonal moves which also update six elements in the adjacency matrix. Here a hexagon is an ordered set of six distinct non-diagonal cells of the type \( \{i_1 j_2, i_1 j_3, i_2 j_3, i_2 j_1, i_3 j_1, i_3 j_2\} \), i.e., we allow distinct row and column indices. If \( i_1 = j_1, i_2 = j_2 \) and \( i_3 = j_3 \), then we have a compact hexagon; otherwise, we have a non-compact hexagon. Note that a compact alternating hexagon is a compact hexagon with entries 1, 0, 1, 0, 1, 0.

The following algorithm updates hexagons while ensuring that the margins of \( x' \) are the same as those for \( x \).

1. Select a hexagon at random.
2. If the selected hexagon has entries 1, 0, 1, 0, 1, 0 or 0, 1, 0, 1, 0, 1, then replace the 1s by 0s and the 0s by 1s; otherwise \( x' = x \).

By combining hexagon updates with rectangle updates, we have an algorithm where it is possible to obtain every possible \( x \) from every other \( x \). If we select at random whether to update a rectangle or a hexagon and then select which rectangle or hexagon to update as described above, we have \( q(x, x') = q(x', x) \). Furthermore, if the selected rectangle or hexagon can be replaced, then \( f(x') = f(x) \propto 1 \). Hence, \( a(x, x') = 1 \) in (1) and such proposals are automatically accepted. A sample from the required conditional distribution of \( M \) is easily obtained by calculating \( m \) for each generated \( x \). Note that the difference between our algorithm and that proposed by Roberts (2000) is that he only used rectangle and compact hexagon updates.

### 2.2 \( U|M, \{X_{i+}\}, \{X_{+j}\} \)

In order to generate from \( U|M, \{X_{i+}\}, \{X_{+j}\} \), Roberts (2000) modified his algorithm so that only moves between adjacency matrices with the same number of mutual dyads were permitted. To do this he introduced an update that moved four edges between four actors by simultaneously updating two symmetrically opposite rectangles. However, whether or not the resulting Markov chain was irreducible was left unanswered. Unfortunately, there is a simple counterexample to demonstrate that the chain is reducible. Consider the networks in Figures 1 (a) and (b), where every actor has in- and out-degree one and there is one mutual. The move between these networks requires the relationships between all five actors.
Figure 1: Two networks, (a) and (b), which are not connected by Roberts’s algorithm for $U|M, \{X_{i+}\}, \{X_{+j}\}$, but are connected via network (c) to be changed, which is not possible using any of the updates used in the algorithm proposed by Roberts (2000). Furthermore, all the networks with the same in-degrees, out-degrees and number of mutual dyads as networks (a) and (b) are isomorphic, i.e., if the actor labels are ignored they are the same. Hence, there are no other networks in the sample space that can be used as an intermediary between networks (a) and (b). Therefore, these networks are not connected by the updates used in Roberts’s algorithm and the chain is reducible.

By updating the non-compact hexagon $\{21, 24, 34, 32, 52, 51\}$, where the vertices are labelled clockwise from the top, we can move between networks (a) and (b) in Figure 1 in one step. Hence, an algorithm which combines the updates of Roberts (2000) and our non-compact hexagon updates, rejecting any proposals that do not maintain the number of mutual dyads, may yield an irreducible Markov chain. However, one way to ensure that the chain is irreducible is to generate from $U|\{X_{i+}\}, \{X_{+j}\}$ using the Metropolis-Hastings algorithm described in Section 2.1 and discard from the final sample all the matrices with the incorrect number of mutual dyads. With such an algorithm we can move between networks (a) and (b) in Figure 1 via network (c) in two steps and network (c) is discarded from our final sample. Provided a reasonable proportion of the generated matrices are retained, say greater than 10%, this method works well; see Section 3 for an example. However, if the proportion of the generated matrices retained is small, then this simple ‘generate and reject’ approach may be impracticable.

If we permit acceptance of $x'$ with a different number of mutual dyads than required, but with probability less than that under $U|\{X_{i+}\}, \{X_{+j}\}$, and acceptance of $x'$ with the required number of mutual dyads with probability more than under $U|\{X_{i+}\}, \{X_{+j}\}$, then we can increase the proportion of retained matrices while ensuring that the Markov chain is
irreducible. Formally, we accept $x'$ with probability

$$\min \left\{ \exp \left\{ -\beta (|m_0 - m'| - |m_0 - m|) \right\}, 1 \right\} ,$$

(2)

where $m_0$ is the required number of mutual dyads, $m$ and $m'$ are the number of mutual dyads in $x$ and $x'$ respectively, and $\beta > 0$ is a parameter which controls the rejection rate. Hence, if $m'$ is closer to $m_0$ than $m$ is, then $x'$ is always accepted. If $m$ is closer to $m_0$ than $m'$ is, then $x'$ is accepted with probability less than one. The larger $\beta$ the greater the proportion of matrices retained, but the greater the correlation between successive iterations of the chain and the greater the chance that some important part of the sample space is not visited by the chain. We advocate choosing $\beta$ so that the proportion of matrices retained is greater than 10%.

3 Examples

We use two examples to assess our proposed algorithm for generating from $U|M, \{X_{i+}\}, \{X_{+j}\}$.

Wasserman (1977) analyzed an adjacency matrix taken from McKinney (1948), derived from relationships in a ninth-grade classroom. Wasserman presented two approximations to the expected value of the triad census, as he was unable to calculate it exactly. These approximations use the distributions he denoted $U|M\ AN, B_{in}, B_{out}$ and $U|\{X_{i+}\}, B_{in}, m$. We reproduce his results in Table 1 along with our estimate of the expected value of the triad census. As the simple generate and reject algorithm retained approximately 20% of the matrices, this algorithm was run until 100,000 matrices from $U|M, \{X_{i+}\}, \{X_{+j}\}$ were obtained. We also present 99% confidence intervals for our estimates, calculated using batch means (see Geyer, 1992). From Table 1, we find that of the two approximations, the expected values under $U|M\ AN, B_{in}, B_{out}$ are closer to those under $U|M, \{X_{i+}\}, \{X_{+j}\}$.

For Krackhardt’s friendship relation (Krackhardt, 1987), Wasserman and Faust (1994) presented the adjacency matrix and the expected value, standard deviations and covariance matrix for the triad census under $U|M\ AN$. As calculation of the expected values and covariance matrix for the triad census under $U|M, \{X_{i+}\}, \{X_{+j}\}$ has not been possible, $U|M\ AN$ has become the reference distribution of choice when assessing the triad census (see, for example, Wasserman and Faust, 1994, Section 13.8.2). We reproduce the expected values and standard deviations in Table 2, along with our estimate of the expected values and standard deviations of the triad census. As the simple generate and reject algorithm retained only 0.05% of the matrices, $\beta$ was increased until a reasonable proportion of
Table 1: Triad census and expected values

| Triad | Observed | $U|M_{AN}, B_{in}, B_{out}$ | $U|M, \{X_{i+}\}, \{X_{+j}\}$ |
|-------|----------|-----------------------------|--------------------------------|
| 003   | 8        | 7.00                        | $5.23 \pm 0.18$                |
| 012   | 19       | 20.50                       | $19.12 \pm 0.30$               |
| 102   | 7        | 8.77                        | $10.23 \pm 0.16$               |
| 021D  | 5        | 4.91                        | $5.83 \pm 0.16$                |
| 021U  | 5        | 7.95                        | $5.71 \pm 0.22$                |
| 021C  | 11       | 7.07                        | $9.20 \pm 0.14$                |
| 111D  | 18       | 14.94                       | $13.07 \pm 0.24$               |
| 111U  | 7        | 7.07                        | $12.70 \pm 0.14$               |
| 030T  | 4        | 6.62                        | $5.29 \pm 0.15$                |
| 030C  | 0        | 0.28                        | $0.75 \pm 0.14$                |
| 201   | 6        | 6.52                        | $7.08 \pm 0.12$                |
| 120D  | 5        | 5.79                        | $4.00 \pm 0.14$                |
| 120U  | 7        | 3.70                        | $4.26 \pm 0.14$                |
| 120C  | 4        | 5.51                        | $6.94 \pm 3.22$                |
| 210   | 6        | 10.92                       | $9.08 \pm 0.12$                |
| 300   | 4        | 1.41                        | $1.49 \pm 0.12$                |

matrices were retained. With $\beta = 2$, 18% of the matrices were retained and the algorithm was run until 500,000 matrices from $U|M, \{X_{i+}\}, \{X_{+j}\}$ were obtained.

As a result of the extra conditioning, the expected values under $U|M, \{X_{i+}\}, \{X_{+j}\}$ are closer to the observed triad census than the expected values under $U|M_{AN}$, especially where under $U|M_{AN}$ triads of different orientations are equally probable, whereas given the in- and out-degrees they are not. Again as a result of the extra conditioning, most of the standard deviations under $U|M, \{X_{i+}\}, \{X_{+j}\}$ are smaller than those under $U|M_{AN}$.

For this example, Roberts (2000) estimated the probability of observing exactly six 300 triads, under $U|M, \{X_{i+}\}, \{X_{+j}\}$, as 10.17% and the probability of observing seven or more such triads as 8.67%. The corresponding estimates from the Markov chain used to produce our estimates in Table 2 are $11.04 \pm 2.95\%$ and $7.62 \pm 3.22\%$. These values suggest that, for this example, Roberts was obtaining realizations from the correct distribution of the triad census under $U|M, \{X_{i+}\}, \{X_{+j}\}$. However, given that we know that his algorithm is not necessarily irreducible we cannot be certain that this will always be the case.
Table 2: Triad census, expected values and standard deviations

| Triad | Observed | Expected values under $U|M$ | Expected values under $U|M, \{X_i^+, X_{i+j}\}$ | Standard deviations under $U|M$ | Standard deviations under $U|M, \{X_i^+, X_{i+j}\}$ |
|-------|----------|-----------------------------|-----------------------------------------------|-------------------------------|-----------------------------------------------|
| 003   | 376      | 320.06                      | 381.94 ± 1.26                                | 9.39                         | 7.02                                          |
| 012   | 366      | 416.82                      | 361.21 ± 2.44                                | 14.56                        | 15.44                                         |
| 102   | 143      | 171.19                      | 131.45 ± 0.90                                | 9.43                         | 6.55                                          |
| 021D  | 114      | 44.09                       | 109.90 ± 1.00                                | 6.22                         | 6.21                                          |
| 021U  | 34       | 44.09                       | 38.33 ± 0.93                                 | 6.22                         | 5.64                                          |
| 021C  | 35       | 88.17                       | 32.55 ± 0.69                                 | 8.17                         | 4.82                                          |
| 111D  | 39       | 73.74                       | 41.46 ± 0.88                                 | 7.78                         | 5.74                                          |
| 111U  | 101      | 73.74                       | 107.57 ± 0.99                                | 7.78                         | 7.01                                          |
| 030T  | 23       | 18.17                       | 26.20 ± 0.77                                 | 3.86                         | 4.65                                          |
| 030C  | 0        | 6.06                        | 0.41 ± 0.08                                  | 2.39                         | 0.67                                          |
| 201   | 20       | 28.97                       | 28.06 ± 0.67                                 | 4.52                         | 3.86                                          |
| 120D  | 16       | 7.74                        | 11.55 ± 0.330                                | 2.71                         | 3.24                                          |
| 120U  | 25       | 7.74                        | 24.87 ± 0.47                                 | 2.71                         | 4.46                                          |
| 120C  | 9        | 15.48                       | 9.07 ± 0.40                                  | 3.74                         | 2.92                                          |
| 210   | 23       | 12.38                       | 21.42 ± 0.47                                 | 3.25                         | 3.41                                          |
| 300   | 6        | 1.55                        | 4.03 ± 0.26                                  | 1.20                         | 1.75                                          |

4 Discussion

Roberts (2000) also presented two further modifications to his algorithm in order to generate from $U|\{M_i, X_i^+, X_{i+j}\}$ and $U|\{X_i^+, X_{i+j}\}$, $B$. The first maintained, for each actor $i$, the in- and out-degree and $M_i$, the number of mutual relationships the actor was involved in. The second maintained the in- and out degrees, the number of ties from one block of actors to a second block of actors and the number of ties in the opposite direction. Note that given the in- and out-degrees this only imposes one extra conditioning constraint, say the number of ties from the first to the second block, $B$. Again he left as an open question whether or not the resulting Markov chains were irreducible.

Networks (a) and (b) in Figure 2 have the same $\{M_i, X_i^+, X_{i+j}\}$, but it is not possible to move between them using rectangle, pair of symmetrically opposite rectangles and compact hexagon updates, as proposed by Roberts (2000), and maintain these constraints. If we use the updates that only maintain the in- and out-degrees and only
retain adjacency matrices that also maintain $\{M_i\}$ in the final sample, then the resulting chain will be irreducible. To reduce the rejection rate we could modify expression (2), although the algorithm may still be very inefficient because there will usually be a large number of additional constraints which must be satisfied, one for each actor. Note that in this example it is possible to move from network (a) to (b) in one step using a non-compact hexagon update.

![Figure 2: Two networks which are not connected by Roberts’s algorithm for $U|M_i, \{X_{i+}\}, \{X_{+j}\}$](image)

Networks (a) and (b) in Figure 3 have two blocks where all the actors have in- and out-degree one and there is one tie from the upper block to the lower block, but again it is not possible to move from network (a) using rectangle, pair of symmetrically opposite rectangles and compact hexagon updates, as proposed by Roberts (2000), and maintain these constraints. However, by relaxing the constraint on $B$ we can move from network (a) to (b) via (c) in two steps using rectangle updates. Alternatively, we can move from network (a) to (b) in one step using a non-compact hexagon update.

As these examples demonstrate, it often possible to construct an algorithm that maintains the constraints, but it is not easy to prove that the resulting Markov chain is irreducible. All three of our counterexamples, which demonstrate that the algorithms proposed by Roberts (2000) are not irreducible, are not counterexamples to the extensions

![Figure 3: Two networks, (a) and (b), which are not connected by Roberts’s algorithm for $U|M, \{X_{i+}\}, \{X_{+j}\}, B$, but are connected via network (c)](image)
to his algorithms that include non-compact hexagon updates. However, it is still an open question whether these modifications produce irreducible chains. Non-compact hexagon updates can change the ties between four, five or six actors, greatly increasing the number of possible adjacency matrices that can be reached in one step from any other matrix, and hence increasing the likelihood of the chain being irreducible. Alternatively, if we use the updates that only maintain the in- and out-degrees and only retain adjacency matrices that also maintain the other constraints in the final sample, possibly reducing the rejection rate using an expression similar to expression (2), then the resulting chain will be irreducible.

References