UCLA
UCLA Previously Published Works

Title
Local dependence in random graph models: characterization, properties and statistical inference

Permalink
https://escholarship.org/uc/item/0vs1d542

Journal
Journal of the American Statistical Association, 77(3)

ISSN
0162-1459

Authors
Schweinberger, Michael
Handcock, Mark S

Publication Date
2015-06-01

DOI
10.1111/rssb.12081

Peer reviewed

Local dependence in random graph models: characterization, properties and statistical inference

Michael Schweinberger

Rice University, Houston, USA

and Mark S. Handcock

University of California, Los Angeles, USA

Summary. Dependent phenomena, such as relational, spatial and temporal phenomena, tend to be characterized by local dependence in the sense that units which are close in a well-defined sense are dependent. In contrast with spatial and temporal phenomena, though, relational phenomena tend to lack a natural neighbourhood structure in the sense that it is unknown which units are close and thus dependent. Owing to the challenge of characterizing local dependence and constructing random graph models with local dependence, many conventional exponential family random graph models induce strong dependence and are not amenable to statistical inference. We take first steps to characterize local dependence in random graph models, inspired by the notion of finite neighbourhoods in spatial statistics and *M*-dependence in time series, and we show that local dependence endows random graph models with desirable properties which make them amenable to statistical inference. We show that random graph models with local dependence satisfy a natural domain consistency condition which every model should satisfy, but conventional exponential family random graph models do not satisfy. In addition, we establish a central limit theorem for random graph models with local dependence, which suggests that random graph models with local dependence are amenable to statistical inference. We discuss how random graph models with local dependence can be constructed by exploiting either observed or unobserved neighbourhood structure. In the absence of observed neighbourhood structure, we take a Bayesian view and express the uncertainty about the neighbourhood structure by specifying a prior on a set of suitable neighbourhood structures. We present simulation results and applications to two real world networks with 'ground truth'.

Keywords: Exponential families; Local dependence; *M*-dependence; Model degeneracy; Social networks; Weak dependence

1. Introduction

Network data arise in many fields, including biology, the health sciences, economics, political science, sociology, machine learning and engineering. In these fields there are many applications with important societal implications, such as protein–protein interactions, the spread of infectious diseases, contagion in financial markets, insurgencies, terrorist networks, criminal networks, social networks, the Internet and power grids (e.g. Kolaczyk (2009)).

We consider a single observation of a network (e.g. a social network) with n nodes and $N = n(n-1)$ directed or $N = n(n-1)/2$ undirected edge variables. The statistical analysis of a single observation of a network is more challenging than the statistical analysis of multiple

Address for correspondence: Michael Schweinberger, Department of Statistics, Rice University, 6100 Main Street, MS-138, Houston, TX 77005, USA.
E-mail: michael.schweinberger@rice.edu

independent networks, because such data are both dependent and high dimensional and give rise to unique conceptual, computational and statistical challenges.

We are concerned with problems of *specification* in the sense of Fisher (1922), page 313, i.e. with the problem of identifying families of distributions which are capable of modelling a wide range of dependences of substantive interest and are amenable to statistical inference. In the past decade, exponential random graph models (ERGMs) have attracted much attention (e.g. Frank and Strauss (1986), Wasserman and Pattison (1996) and Lusher *et al.* (2013)). Despite attractive finite sample properties (e.g. Barndorff-Nielsen (1978)), ERGMs have turned out to be problematic models of real world networks (e.g. Snijders (2002), Handcock (2003), Hunter *et al.* (2008) and Chatterjee and Diaconis *et al.* (2013)).

One of the most striking observations is that some of the most interesting ERGMs do not place much probability mass on graphs which resemble real world networks (e.g. Handcock (2003) and Hunter *et al.* (2008)). Let $\mathbf{Y} \in \mathbb{Y}$ be a random graph on a finite set of nodes, corresponding to edge variables $Y_{i,j}$ between pairs of nodes (i, j). To simplify the discussion, suppose that the edge variables $Y_{i,j}$ are binary, i.e. $Y_{i,j} \in \{0, 1\}$. A convenient representation of a distribution $\mathbb{P}$ with support $\mathbb{Y}$ is as an exponential family of the form

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}) = \exp\{\langle \boldsymbol{\theta}, s(\mathbf{y}) \rangle - \psi(\boldsymbol{\theta})\}, \qquad \mathbf{y} \in \mathbb{Y}, \tag{1}$$

where $\langle \boldsymbol{\theta}, s(\mathbf{y}) \rangle$ denotes the inner product of a d-vector of natural parameters $\boldsymbol{\theta}$ and a d-vector of sufficient statistics $s(\mathbf{y})$, and $\psi(\boldsymbol{\theta})$ is a log-normalizing constant. The d-vector of sufficient statistics may include statistics of interest, such as the number of transitive triples of nodes (e.g., in friendship networks, 'a friend of my friend is my friend'). Such statistics induce dependence between edge variables and are of great interest in the fast growing field of network science (e.g. Wasserman and Faust (1994), Kolaczyk (2009) and Lusher *et al.* (2013)). If $\mathbf{S} = s(\mathbf{Y})$ denotes the vector of sufficient statistics and $\mathbb{S}$ denotes the convex hull of $\{s(\mathbf{y}) : \mathbf{y} \in \mathbb{Y}\}$, then the induced distribution of $\mathbf{S}$ is given by

$$\mathbb{Q}_{\boldsymbol{\theta}}(S \in \mathcal{S}) = \mathbb{P}_{\boldsymbol{\theta}}\{\mathbf{Y} \in S^{-1}(\mathcal{S})\} = \sum_{\mathbf{y} \in S^{-1}(\mathcal{S})} \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}), \tag{2}$$

where $S^{-1}(\mathcal{S})$ denotes the subset of $\mathbb{Y}$ mapping into $\mathcal{S} \subset \mathbb{S}$. If the sufficient statistics include counts of the number of transitive triples and other subgraph configurations, then the induced distribution of $\mathbf{S}$ tends to place much probability mass on extreme graphs which do not resemble real world networks. There is both theoretical and empirical evidence which suggests that many families of distributions with such count statistics place much mass on the relative boundary rather than in the relative interior of $\mathbb{S}$ (e.g. Snijders (2002), Handcock (2003) and Hunter *et al.* (2008)). Worse, theoretical results indicate that the behaviour of conventional ERGMs does not improve as the number of nodes n increases, but deteriorates (Strauss, 1986; Jonasson, 1999; Schweinberger, 2011; Butts, 2011). The best-known examples are Markov random graph models (Frank and Strauss, 1986), though other interesting models are problematic as well. The flawed nature of conventional ERGMs is demonstrated by Fig. 1. It relates to an ERGM of the form (1) with the number of edges and triangles as sufficient statistics with $n = 100$ nodes and $N = 4950$ undirected edge variables. Fig. 1 shows the prior predictive distributions of the sufficient statistics from the model, which is described in detail in Section 5.1. It demonstrates that the prior predictive distribution places most of its mass on graphs which are extreme in terms of the number of edges and triangles.

In this paper, we address the root of the problem by characterizing and constructing well-behaved random graph models which are amenable to statistical inference. The point of
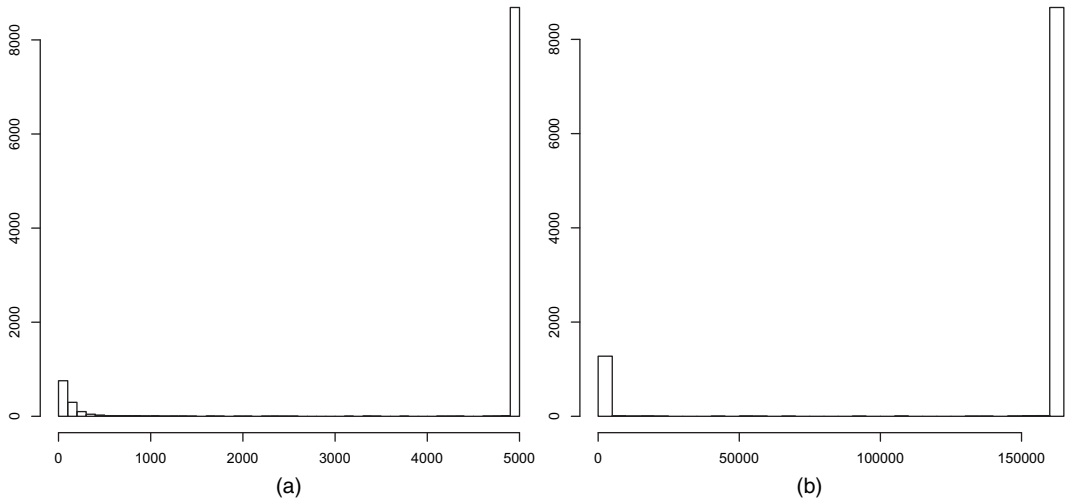
Fig. 1. Prior predictions of the number of (a) edges and (b) triangles under the global triangle model with $N = 4950$ variables: note the extreme polarization

departure is the observation that many statistics of interest, **S**, are sums of random variables, e.g. counts of the number of edges and transitive triples. The distribution of (normed) sums of independent or weakly dependent random variables tends to be Gaussian by virtue of some version of the central limit theorem (e.g. Billingsley (1995)). If the expected value of **S** is in the relative interior of $\mathbb{S}$, then the model should place significant mass around the expected value of **S** by virtue of the approximate Gaussian distribution of **S**. Therefore, for all expected values of **S** in the relative interior of $\mathbb{S}$, the model should place much mass on the relative interior of $\mathbb{S}$. The difficulty is that, in the absence of spatial, temporal and other structure, it is not evident which edge variables should be dependent. Therefore, the specification of random graph models with weak dependence is challenging and conventional ERGMs induce either no dependence and are simplistic (e.g. Bernoulli random graph models) or strong dependence and are near degenerate (e.g. Markov random graph models; Frank and Strauss (1986)).

We take first steps to characterize local dependence in random graph models in Section 2. We demonstrate that local dependence endows random graph models with desirable properties which make them amenable to statistical inference. One property is a natural domain consistency condition that any probability model should satisfy, but many parameterizations of ERGMs do not satisfy (Shalizi and Rinaldo, 2013). A second and more important property is asymptotic Gaussian behaviour of statistics, which suggests that random graph models with local dependence place much probability mass around the expected value of statistics of interest. We discuss the construction of random graph models with local dependence in Section 3 and Bayesian inference given complete as well as incomplete data in Section 4. If suitable neighbourhood structure is observed, at least two approaches to statistical inference are possible, depending on whether the observed neighbourhood structure is regarded as fixed or random. If no suitable neighbourhood structure is observed, we take a Bayesian view and express the uncertainty about the neighbourhood structure by specifying a prior on a set of suitable neighbourhood structures, using hierarchical parametric and non-parametric priors and auxiliary variable Markov chain Monte Carlo methods. We present simulation results and applications to two real world networks with ground truth in Section 5.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

1.1. Other related work

Snijders *et al.* (2006) and Hunter and Handcock (2006) considered non-linear constraints on the parameter space of ERGMs. Such curved ERGMs have been applied with some success (Hunter *et al.*, 2008) but do not admit simple representations of dependences and the interpretation of parameters is challenging, as noted by Snijders *et al.* (2006), page 149. An alternative is latent variable models, which we discuss in Section 2.1. Selected special cases and other related work are discussed in Section 3.4.

2. Dependence

We discuss in Section 2.1 two broad approaches to modelling dependence, one based on latent variable models and the other based on ERGMs, and we argue that ERGMs are attractive when dependence is of substantive interest. We discuss in Section 2.2 the challenges that are encountered in modelling dependence of substantive interest by ERGMs. In Section 2.3, we introduce a notion of local dependence and in Section 2.4 we show that that local dependence endows models with desirable properties which make them amenable to statistical inference.

2.1. Dependence of substantive interest

Most relational phenomena are dependent phenomena, and dependence is often of substantive interest. Examples can be found in the social sciences (e.g. Wasserman and Faust (1994) and Lusher *et al.* (2013)), economics (e.g. Jackson (2008)), the health sciences (e.g. Welch *et al.* (2011)) and physics (Newman *et al.*, 2002). As an example, if (i, j, k) is a triple of nodes and the edge variables $Y_{i,j}, Y_{j,k}, Y_{i,k} \in \{0, 1\}$ are binary and undirected, then the triple is called *transitive* if $Y_{i,j} Y_{j,k} Y_{i,k} = 1$, i.e. there are edges between i and j as well as j and k and i and k. Other examples were discussed by Wasserman and Faust (1994) and Lusher *et al.* (2013).

When modelling transitivity and other dependences, it is not attractive to assume conditional independence of edge variables, e.g.

$$Y_{i,j} | p_{i,j} \overset{\text{ind}}{\sim} \text{Bernoulli}(p_{i,j}), \qquad i < j, \tag{3}$$

where $p_{i,j}$ denotes the probability of a binary undirected edge between nodes i and j, which may depend on observed and unobserved latent variables. Examples of models of the form (3) are stochastic block models (e.g. Nowicki and Snijders (2001)) and mixed membership models (e.g. Airoldi *et al.* (2008)), random-effects and mixed effects models (e.g. van Duijn *et al.* (2004) and Hoff (2005)) and latent space models (Hoff *et al.*, 2002; Schweinberger and Snijders, 2003; Handcock *et al.*, 2007; Krivitsky *et al.*, 2009). Although models of the form (3) can capture transitive closure by introducing latent structure, such models induce dependence indirectly through latent variables rather than directly. In situations where dependence is of substantive interest, scientists tend to prefer models which allow us to specify dependences directly. Examples are the spatial covariance and variogram functions in spatial random-field models (Cressie, 1993), interaction functions in spatial point processes (Møller and Waagepetersen, 2004) and covariance terms in time series (Granger and Morris, 1976). An additional, well-known example is the Ising model in physics (e.g. Georgii (2011)). The Ising model allows the explicit specification

of the nature of interactions between particles. Physicists would hesitate to exchange the Ising model for a latent variable model which assumed that particles are independent conditionally on observed structure (e.g. observed locations on a lattice) or unobserved latent structure.

In the realm of networks, Frank and Strauss (1986) and Wasserman and Pattison (1996) introduced exponential family models which resemble the Ising model in physics and lattice models in spatial statistics and which allow the modelling of a wide range of dependences of substantive interest, including transitive closure. Such models have attracted much attention in the social sciences and health sciences and elsewhere, as Lusher *et al.* (2013) testifies.

2.2. Modelling dependence: challenges

Despite the fact that ERGMs are the natural relatives of well-established models in physics, spatial statistics, machine learning and artificial intelligence, ERGMs lack something that most other areas have: neighbourhood structure. The lack of neighbourhood structure makes modelling dependence challenging.

In spatial statistics (e.g. Cressie (1993) and Stein (1999)) and time series (e.g. Granger and Morris (1976)) and the related work on mixing conditions in probability theory (Billingsley (1995), pages 363–370, and Dedecker *et al.* (2007)), it is assumed that dependence decreases as the distance between random variables in the spatial or temporal domain increases. Thus, events may be dependent as long as the distance between the locations of the events is small, whereas distant events are almost dependent. In the realm of networks, though, it is not evident what distance between subgraphs means and how the dependence between subgraphs should decrease. A possible approach is to assume uniform weak dependence in the sense that all edges and subgraphs in large graphs are almost independent. Such dependence assumptions are not appealing, however, because network science would expect strong local dependence between some of the edges and subgraphs.

In fact, even when spatial or temporal structure is available, there is often more local structure than would be expected on the basis of the location of nodes in space and time: for example, some subsets of nodes may be close in geographical space, but the members of the subset may be distant in 'network space', whereas other subsets of nodes may be distant in geographical space, but close in network space, where network space is understood as other structure that is not captured by geographical space; for example, researchers in the same building, on the same floor and in the same department may not collaborate, but may engage in transitive collaborations with other researchers who are distant in geographical space.

A well-known example of the challenge of modelling dependence is Markov random graph models (Frank and Strauss, 1986). Suppose that the random graph $\mathbf{Y}$ is undirected and binary. Motivated by the nearest neighbour definition in physics (Georgii, 2011) and spatial statistics (Besag, 1974), Frank and Strauss (1986) called two dyads $\{i, j\}$ and $\{k, l\}$ neighbours if $\{i, j\}$ and $\{k, l\}$ share a node and assumed that, if $\{i, j\}$ and $\{k, l\}$ are not neighbours, then $Y_{i,j}$ and $Y_{k,l}$ are independent conditionally on the rest of random graph $\mathbf{Y}$. Markov random graph models can be represented in exponential family form (1) with the number of edges $s_1(\mathbf{y}) = \Sigma_{i<j} y_{i,j}$, the number of k-stars $s_k(\mathbf{y}) = \Sigma_i \Sigma_{j_1 < \ldots < j_k} y_{i,j_1} \ldots y_{i,j_k}$ and the number of triangles $s_n(\mathbf{y}) = \Sigma_{i<j<k} y_{i,j} y_{j,k} y_{i,k}$ as sufficient statistics. Markov random graph models and generalizations to ERGMs (Wasserman and Pattison, 1996) allow scientists to model transitive closure and other dependences along with covariate-related similarity, which scientists have long considered to be of great interest (e.g. Wasserman and Faust (1994) and Lusher *et al.* (2013)). Despite the underlying nearest neighbour assumption and its scientific appeal, however, Markov random graph models are problematic models of real world networks. A simple observation by Strauss (1986) that demonstrates the fundamental flaws of Markov random graph models is that, for

any given pair of nodes $\{i, j\}$, the number of neighbours is $2(n-2)$ and thus increases with the number of nodes n. The large and growing neighbourhoods indicate that Markov random graph models induce increasingly stronger dependence as n increases and are problematic when n is large. This has been confirmed by a growing body of theoretical and empirical results (Strauss, 1986; Jonasson, 1999; Snijders, 2002; Handcock, 2003; Hunter *et al*., 2008; Rinaldo *et al*., 2009; Schweinberger, 2011; Butts, 2011; Shalizi and Rinaldo, 2013; Chatterjee and Diaconis, 2013).

2.3. *Characterizing local dependence*
We take first steps to characterize local dependence, drawing inspiration from two sources: network science and probability theory. On the one hand, network science (e.g. Homans (1950), Wasserman and Faust (1994) and Pattison and Robins (2002)) suggests that interactions in networks are local. On the other hand, weak dependence conditions in probability theory, such as mixing conditions (e.g. Billingsley (1995) and Dedecker *et al*. (2007)), suggest that dependence should be local to ensure weak dependence and thus desirable behaviour, such as central limit theorems. The study of probability measures on infinite domains in physics (e.g. Georgii (2011)), spatial statistics (e.g. Cressie (1993), section 7.3.1, and Stein (1999), chapter 3) machine learning and artificial intelligence (e.g. Singla and Domingos (2007) and Xiang and Neville (2011)) as well as the notion of M-dependence in time series (e.g. Billingsley (1995), pages 363–370) suggest that each random variable should depend on a finite subset of other random variables. In other words, a natural starting point is to assume that each edge variable depends on a finite subset of other edge variables. We adapt the idea of finite neighbourhoods and M-dependence to random graphs as follows.

Definition 1 (local dependence). Let **Y** be a random graph with domain $\mathcal{D} = \mathcal{N} \times \mathcal{N}$ and sample space $\mathbb{Y}$, where $\mathcal{N}$ is a finite set of nodes. The dependence that is induced by a probability measure $\mathbb{P}$ on $\mathbb{Y}$ is called *local* if there is a partition of the set of nodes $\mathcal{A}$ into $K \geqslant 2$ non-empty finite subsets $\mathcal{A}_1, \ldots, \mathcal{A}_K$, called neighbourhoods, such that the within- and between-neighbourhood subgraphs $\mathbf{Y}_{k,l}$ with domains $\mathcal{A}_k \times \mathcal{A}_l$ and sample spaces $\mathbb{Y}_{k,l}$ satisfy, for all $\mathcal{Y} \subseteq \mathbb{Y}$ and $\mathcal{Y}_{k,l} \subseteq \mathbb{Y}_{k,l}$,

$$\mathbb{P}_K(\mathbf{Y} \in \mathcal{Y}) = \prod_{k=1}^{K} \mathbb{P}_{k,k}(\mathbf{Y}_{k,k} \in \mathcal{Y}_{k,k}) \prod_{l=1}^{k-1} \mathbb{P}_{k,l}(\mathbf{Y}_{k,l} \in \mathcal{Y}_{k,l}, \mathbf{Y}_{l,k} \in \mathcal{Y}_{l,k}), \qquad (4)$$

where within-neighbourhood probability measures $\mathbb{P}_{k,k}$ induce dependence within subgraphs $\mathbf{Y}_{k,k}$, whereas between-neighbourhood probability measures $\mathbb{P}_{k,l}$ induce independence between subgraphs, i.e., for all $\mathcal{Y}_{k,l} \subseteq \mathbb{Y}_{k,l}$, $\mathcal{Y}_{i,k} \subseteq \mathbb{Y}_{i,k,j}$, $\mathcal{Y}_{\mathrm{B},i,j} \subseteq \mathbb{Y}_{\mathrm{B},i,j}$ and $\mathcal{Y}_{\mathrm{B},j,i} \subseteq \mathbb{Y}_{\mathrm{B},j,i}$,

$$\mathbb{P}_{k,l}(\mathbf{Y}_{k,l} \in \mathcal{Y}_{k,l}, \mathbf{Y}_{l,k} \in \mathcal{Y}_{l,k}) = \prod_{i \in \mathcal{A}_k, j \in \mathcal{A}_l} \mathbb{P}_{\mathrm{B},i,j}(Y_{\mathrm{B},i,j} \in \mathcal{Y}_{\mathrm{B},i,j}, Y_{\mathrm{B},j,i} \in \mathcal{Y}_{\mathrm{B},j,i}), \qquad l < k, \quad (5)$$

where $Y_{\mathrm{B},i,j}$ denote between-neighbourhood edge variables corresponding to nodes i and j who are not members of the same neighbourhood.

Thus, local dependence breaks down the dependence of the random graph **Y** into dependence within subgraphs $\mathbf{Y}_{k,k}$. The construction of random graph models with local dependence is discussed in Section 3.

The first and foremost advantage of local dependence is that it makes no assumptions about the form and strength of dependence within subgraphs. Scientists are free to incorporate dependences of interest, such as transitive closure within subgraphs. In contrast, conventional ERGMs (e.g. Frank and Strauss (1986)) induce unbounded neighbourhoods and global dependence, as discussed in Section 2.2.

A second advantage is that local dependence endows models with desirable properties, which we discuss in Section 2.4.

2.4. *Properties of local dependence*

We show that random graphs with local dependence have two natural properties. The first property is a domain consistency property that any probability model should have. The second property is asymptotic Gaussian behaviour of statistics of interest. The two properties help to address both problems of *specification* and *distribution* in the sense of Fisher (1922), page 313, by allowing us to model a wide range of dependences within subgraphs and facilitating the derivation of the distributions of estimators and goodness-of-fit statistics.

Since we are considering a single observation of a graph, we follow a domain increasing approach to asymptotics that resembles the domain increasing approach in spatial statistics (e.g. Cressie (1993), section 7.3.1, and Stein (1999), chapter 3). Suppose that the domain of the random graph increases as follows. Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be a sequence of non-empty finite sets of nodes and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be a sequence of random graphs with increasing domain $\mathcal{N}_1 \times \mathcal{N}_1, \mathcal{N}_2 \times \mathcal{N}_2, \ldots$, where the set of nodes $\mathcal{N}_K = \cup_{k=1}^{K} \mathcal{A}_k$ is the union of the sets of nodes $\mathcal{A}_1, \ldots, \mathcal{A}_K$.

The first property is a domain consistency property that should be satisfied by any probability model (e.g. Billingsley (1995), section 36), but which many parameterizations of ERGMs do not satisfy (Shalizi and Rinaldo, 2013).

Theorem 1. Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be a sequence of non-empty, finite sets of nodes and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be a sequence of random graphs with increasing domain $\mathcal{N}_1 \times \mathcal{N}_1, \mathcal{N}_2 \times \mathcal{N}_2, \ldots$, where $\mathcal{N}_K = \cup_{k=1}^{K} \mathcal{A}_k$. Let $\mathbf{Y}_{K+1 \backslash K}$ be the random graph $\mathbf{Y}_{K+1}$ excluding $\mathbf{Y}_K$, i.e. $\mathbf{Y}_{K+1 \backslash K}$ corresponds to the within-neighbourhood subgraph $\mathbf{Y}_{K+1, K+1}$ and the between-neighbourhood subgraphs $\mathbf{Y}_{k, K+1}$ and $\mathbf{Y}_{K+1, k}, k = 1, \ldots, K$; and let $\mathbb{Y}_{K+1 \backslash K}$ be the sample space of $\mathbf{Y}_{K+1 \backslash K}$. If a sequence of random graphs $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ satisfies local dependence, then it is *domain consistent* in the sense that, for all $K > 0$ and $\mathcal{Y}_K \subseteq \mathbb{Y}_K$,

$$\mathbb{P}_K(\mathbf{Y}_K \in \mathcal{Y}_K) = \mathbb{P}_{K+1}(\mathbf{Y}_K \in \mathcal{Y}_K, \mathbf{Y}_{K+1 \backslash K} \in \mathbb{Y}_{K+1 \backslash K}). \tag{6}$$

In other words, the probability measure $\mathbb{P}_K$ of random graph $\mathbf{Y}_K$ with domain $\mathcal{N}_K \times \mathcal{N}_K$ can be recovered from the probability measure $\mathbb{P}_{K+1}$ of random graph $\mathbf{Y}_{K+1}$ with domain $\mathcal{N}_{K+1} \times \mathcal{N}_{K+1}$ by marginalizing with respect to $\mathbf{Y}_{K+1 \backslash K}$. It is worth noting that the domain consistency condition that is considered here is weaker than the domain consistency condition that was considered by Shalizi and Rinaldo (2013) and is motivated by the way that the domain of local random graphs increases.

In addition to domain consistency, it is desirable that random graphs with increasing domain satisfy sparsity. A random graph can be called *sparse* if the expected degrees $\mathbb{E}(\Sigma_j Y_{i,j})$ of nodes i are bounded, suggesting that $\mathbb{E}(Y_{i,j}) \to 0$ as the number of nodes increases. The importance of sparsity has been recognized by social scientists, computer scientists, mathematicians (e.g. Jonasson (1999) and Lovász (2012), page viii and page 4) and statisticians (e.g. Krivitsky *et al.* (2011) and Vu *et al.* (2013)). Sparsity embodies the notion that, in the real world, resources are bounded—animals and humans face real world constraints such as limited time and therefore cannot maintain arbitrarily many relationships. If random graphs with local dependence do not satisfy sparsity, then the expected degrees of nodes would be dominated by edges to nodes in other neighbourhoods, which would be in conflict with the notion of local interaction in network science. We therefore focus on graphs where within-neighbourhood subgraphs may be dense, but between-neighbourhood subgraphs are sparse.

Definition 2 (sparsity). Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be a sequence of non-empty, finite sets of nodes and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be a sequence of random graphs with increasing domain $\mathcal{N}_1 \times \mathcal{N}_1, \mathcal{N}_2 \times \mathcal{N}_2, \ldots$, where $\mathcal{N}_K = \cup_{k=1}^{K} \mathcal{A}_k$. A sequence of random graphs $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ satisfying local dependence is called δ *sparse* if there exist constants $A > 0$ and $\delta > 0$ such that

$$\mathbb{E}(|Y_{\mathrm{B},i,j}|^p) \leqslant An^\delta, \qquad p = 1, 2. \tag{7}$$

The second, and more important, property is the fact that the asymptotic distribution of statistics of interest is Gaussian, which helps to address problems of estimation and goodness of fit. In practice, most statistics of interest are sums of subgraph configurations. Let $\mathcal{S}_K \subseteq \times_{i=1}^{d} \mathcal{N}_K$ be a subset of the d-dimensional Cartesian product of $\mathcal{N}_K$ with itself and $S_K : \mathcal{S}_K \mapsto \mathbb{R}$ be a real-valued function with domain $\mathcal{S}_K$, e.g. the number of edges or transitive triples. Such sums of subgraph configurations can be written as

$$S_K = \sum_{i \in \mathcal{S}_K} S_{K,i}, \tag{8}$$

where $S_{K,i} = \Pi_{k=1}^{q} Y_{i,a_k,b_k}$ are interactions of q distinct edge variables Y_{i,a_k,b_k}, which resemble the interactions in undirected graphical models and related models in physics (e.g. the Ising model) and spatial statistics (e.g. random fields). If, for example, edge variables are binary and $\mathcal{S}_K \subseteq \mathcal{N}^3$, then $S_{K,i} = Y_{a,b} Y_{b,c} Y_{a,c}$ is an indicator of whether the triple of nodes (a, b, c) is transitive. The sum S_K can be decomposed into within- and between-neighbourhood sums W_K and B_K:

$$S_K = W_K + B_K, \tag{9}$$

where $W_K = \Sigma_{k=1}^{K} W_{K,k}$ is the total within-neighbourhood sum, composed of the within-neighbourhood sums $W_{K,k} = \Sigma_{i \in \mathcal{S}_K} \mathbf{1}_{W,k,i} S_{K,i}$ in neighbourhoods k, and $B_K = \Sigma_{i \in \mathcal{S}_K} \mathbf{1}_{B,i} S_{K,i}$ is the total between-neighbourhood sum. The indicator function $\mathbf{1}_{W,k,i}$ is 1 if subgraph configuration i involves nodes in neighbourhood k and neighbourhood k only and is 0 otherwise, whereas $\mathbf{1}_{B,i}$ is 1 if $S_{K,i}$ involves nodes of more than one neighbourhood and is 0 otherwise.

It turns out that sequences of local and sparse random graphs with increasing domain are well behaved in the sense of satisfying a central limit theorem for weakly dependent random variables.

Theorem 2. Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be a sequence of non-empty, finite sets of nodes and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be a sequence of random graphs with increasing domain $\mathcal{N}_1 \times \mathcal{N}_1, \mathcal{N}_2 \times \mathcal{N}_2, \ldots$, where $\mathcal{N}_K = \cup_{k=1}^{K} \mathcal{A}_k$. Consider sums of the form $S_K = \Sigma_{i \in \mathcal{S}_K} S_{K,i}$, where $\mathcal{S}_K \subseteq \times_{i=1}^{d} \mathcal{N}_K$ and $S_{K,i} = \Pi_{k=1}^{q} Y_{i,a_k,b_k}$. Suppose that the edge variables $Y_{a,b}$ satisfy uniform boundedness in the sense that there is a constant $C > 0$ such that, for all $K > 0$, $a \in \mathcal{N}_K$ and $b \in \mathcal{N}_K$, $\mathbb{P}(|Y_{a,b}| \leqslant C) = 1$. Without loss of generality, assume that, for all $K > 0$ and $i \in \mathcal{S}_K$, $\mathbb{E}(S_{K,i}) = 0$. If the sequence of random graphs $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ is local $\delta > d$-sparse and $\mathbb{V}(W_K) \to \infty$ as $K \to \infty$, then

$$\lim_{K \to \infty} \max_{1 \leqslant k \leqslant K} \mathbb{P}\{|W_{K,k}| > \varepsilon \sqrt{\mathbb{V}(W_K)}\} = 0 \tag{10}$$

and

$$\frac{S_K}{\sqrt{\mathbb{V}(S_K)}} \xrightarrow{\mathrm{d}} N(0, 1) \qquad \text{as } K \to \infty, \tag{11}$$

where $\mathbb{V}(W_K)$ and $\mathbb{V}(S_K)$ denote the variance of W_K and S_K respectively.

We discuss implications, starting with the most important: theorem 2 respects the *desiderata* that random graphs be local and sparse and imposes no constraints on the form and shape of within-neighbourhood probability distributions, granting scientists complete freedom to specify arbitrary dependences of interest within neighbourhoods, such as transitive closure. At the

same time, local and sparse random graphs tend to be well behaved in the sense that neighbourhoods cannot dominate the whole graph by equation (10) and the distribution of statistics, e.g. the number of transitive triples, tends to be Gaussian by expression (11) provided that the number of neighbourhoods K is large. As a result, random graph models with local dependence can be expected to place much probability mass around the expected values of statistics. If a graph is observed and the method of estimation (e.g. the method of maximum likelihood in exponential families) matches the expected and observed values of selected statistics, then the goodness of fit of the model with respect to the selected statistics can be expected to be acceptable.

Some additional remarks are in order. The uniform boundedness condition covers the most common cases, including the case of binary edge variables $Y_{i,j} \in \{0, 1\}$. Multivariate extensions of the central limit theorem may be obtained by the Cramér–Wold theorem (e.g. Billingsley (1995), page 383). If suitable parameterizations of random graph models with local dependence are chosen, then the δ-method can be used to establish asymptotic normality of maximum likelihood estimators and test statistics along the lines of, for example, DasGupta (2008), section 16.3.

3. Model construction and parameterizations

To construct random graph models with local dependence, a suitable neighbourhood structure is needed. In practice, suitable neighbourhood structure may or may not be observed.

Let $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ be membership indicators, where $\mathbf{Z}_i$ is the vector of membership indicators Z_{ik} of node i, where $Z_{ik} = 1$ if node i is a member of neighbourhood $\mathcal{A}_k$ and $Z_{ik} = 0$ otherwise. Since most network data are discrete, we consider throughout discrete network data and probability densities with respect to counting measure. We assume that the conditional probability mass function (PMF) of a random graph $\mathbf{Y}$ given a neighbourhood structure $\mathbf{Z} = \mathbf{z}$ can be written as

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}) = \prod_{k=1}^{K} \mathbb{P}(\mathbf{Y}_{k,k} = \mathbf{y}_{k,k} | \mathbf{Z} = \mathbf{z}) \prod_{l=1}^{k-1} \mathbb{P}(\mathbf{Y}_{k,l} = \mathbf{y}_{k,l}, \mathbf{Y}_{l,k} = \mathbf{y}_{l,k} | \mathbf{Z} = \mathbf{z}), \quad (12)$$

where between-neighbourhood PMFs can be factorized into dyad-bound PMFs:

$$\mathbb{P}(\mathbf{Y}_{k,l} = \mathbf{y}_{k,l}, \mathbf{Y}_{l,k} = \mathbf{y}_{l,k} | \mathbf{Z} = \mathbf{z}) = \prod_{i \in \mathcal{A}_k, j \in \mathcal{A}_l} \mathbb{P}(Y_{\mathrm{B},i,j} = y_{\mathrm{B},i,j}, Y_{\mathrm{B},j,i} = y_{\mathrm{B},j,i} | \mathbf{Z} = \mathbf{z}), \quad (13)$$

whereas the within-neighbourhood PMFs are not assumed to be factorizable.

In practice, the question is the source of the neighbourhood structure $\mathbf{Z}$. If suitable neighbourhood structure were observed, then it should be used. We discuss model construction and two approaches to statistical inference given observed neighbourhood structure in Section 3.1. An important practical problem is that in most applications no suitable neighbourhood structure is observed. We discuss model construction and statistical inference in the absence of observed neighbourhood structure in Section 3.2. Parameterizations are discussed in Section 3.3 and selected special cases in Section 3.4.

3.1. Model construction with observed neighbourhoods
Consider the situation where a suitable neighbourhood structure is observed. It is worth noting

that the theoretical results of Section 2.4 suggest that, to be suitable, none of the neighbourhoods should dominate the whole graph.

We consider two approaches to statistical inference given observed neighbourhood structure $\mathbf{Z} = \mathbf{z}_{\text{obs}}$ and assume that the conditional PMF (12) of random graph $\mathbf{Y}$ given observed neighbourhood structure $\mathbf{Z} = \mathbf{z}_{\text{obs}}$ is parameterized by $\boldsymbol{\theta}$.

The first approach regards the observed neighbourhood structure $\mathbf{z}_{\text{obs}}$ as fixed and bases statistical inference on the likelihood function

$$L(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}_{\text{obs}}). \tag{14}$$

The second approach considers the observed neighbourhood $\mathbf{z}_{\text{obs}}$ as the outcome of a random variable $\mathbf{Z}$ with a PMF $\mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z}_{\text{obs}})$ parameterized by $\boldsymbol{\pi}$ and bases statistical inference on the likelihood function

$$L(\boldsymbol{\theta}, \boldsymbol{\pi}) = \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}_{\text{obs}}) \, \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z}_{\text{obs}}). \tag{15}$$

If the two parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are variation independent in the sense that the parameter space $\Omega_{\boldsymbol{\theta}, \boldsymbol{\pi}}$ is a product space of the form $\Omega_{\boldsymbol{\theta}, \boldsymbol{\pi}} = \Omega_{\boldsymbol{\theta}} \times \Omega_{\boldsymbol{\pi}}$, where $\Omega_{\boldsymbol{\theta}}$ is the parameter space of $\boldsymbol{\theta}$ and $\Omega_{\boldsymbol{\pi}}$ is the parameter space of $\boldsymbol{\pi}$, then the likelihood function is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\pi}) = \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}_{\text{obs}}) \, \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z}_{\text{obs}}) = L(\boldsymbol{\theta}) \, L(\boldsymbol{\pi}), \tag{16}$$

where $L(\boldsymbol{\theta})$ is given by equation (14) and $L(\boldsymbol{\pi})$ is given by $L(\boldsymbol{\pi}) = \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z}_{\text{obs}})$. Thus, if the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are variation independent, then the likelihood function factorizes and statistical inference for $\boldsymbol{\theta}$ can be based on $L(\boldsymbol{\theta})$. In other words, the two approaches are equivalent as long as the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are variation independent. In general, the random-neighbourhood approach may be more suitable than the fixed neighbourhood approach when it is believed that the neighbourhoods are generated by a stochastic mechanism and that mechanism is itself of interest. The likelihood function $L(\boldsymbol{\theta})$ can be maximized by the maximum likelihood methods of Hunter and Handcock (2006) by using $\mathbf{z}_{\text{obs}}$ as a covariate and Bayesian inference can be conducted by using the Bayesian methods of Koskinen *et al.* (2010) and Caimo and Friel (2011).

There are multiple data structures which could be exploited to construct random graph models with local dependence. Strauss and Ikeda (1990) suggested constructing neighbourhoods by exploiting the categories of categorical covariates; for example the two categories of the categorical covariate gender could be used to form two neighbourhoods, corresponding to females and males. However, many categorical covariates are collected by surveys and have a small number of categories. The theoretical results in Section 2.4 suggest that the number of neighbourhoods should be large and none of the neighbourhoods should dominate the graph, making categorical covariates with a small number of categories problematic when the number of nodes is large. A second approach exploits multilevel structure. In the health sciences and social sciences, many network data have a multilevel structure in the sense that subgraphs are nested in graphs; for example, researchers are located in departments, departments are nested in buildings, and buildings belong to a campus. Such multilevel structure could be exploited to form the neighbourhood structure. A third approach exploits spatial structure provided that it is available, though spatial structure may not capture the whole dependence, as discussed in Section 2.2.

3.2. Model construction without observed neighbourhoods
It is common that no suitable neighbourhood structure is observed. In such cases, we follow a

Bayesian approach and express the uncertainty about the neighbourhood structure by a prior on a set of suitable neighbourhood structures. We consider both hierarchical parametric and non-parametric priors.

In principle, statistical inference could be based on the likelihood function

$$L(\boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{\mathbf{z} \in \mathcal{Z}} \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}) \, \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z}), \qquad (17)$$

where $\mathcal{Z}$ is the sample space of $\mathbf{Z}$. The difficulty is that $\mathcal{Z}$ is either a finite but large set with $\exp\{n \log(K)\}$ elements—provided that the number of neighbourhoods K is fixed and known— or a countably infinite set, and in general the sum cannot be computed by complete enumeration. To facilitate statistical inference, we augment the observed data $\mathbf{Y}$ by the unobserved data $\mathbf{Z}$ and exploit hierarchical parametric and non-parametric priors.

To be specific, assume that

$$\mathbf{Z}_i | \pi_1, \ldots, \pi_K \overset{\text{IID}}{\sim} \text{multinomial}(1; \pi_1, \ldots, \pi_K), \qquad i = 1, \ldots, n, \qquad (18)$$

is the distribution of membership vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. We note that one could incorporate predictors of memberships by using multinomial logit or probit link functions along the lines of Tallberg (2005).

A parametric approach could be based on Dirichlet priors:

$$\pi_1, \ldots, \pi_K \sim \text{Dirichlet}(\omega_1, \ldots, \omega_K). \qquad (19)$$

A potential problem with Dirichlet priors is that the number of neighbourhoods K must either be known or selected by model selection methods, which is not straightforward. An alternative is to express the uncertainty about K by specifying a prior for K (e.g. Richardson and Green (1997)), which leads to complicated Markov chain Monte Carlo algorithms. We follow a non-parametric approach based on stick breaking priors (Ishwaran and James, 2001), which sidesteps these difficulties. It allows the number of non-empty neighbourhoods *a posteriori* to be large, while encouraging it *a priori* to be small. Suppose that there is an infinite number of neighbourhoods and that nodes belong to neighbourhood $k = 1, 2, \ldots$ with probability π_k, $k = 1, 2, \ldots$, where

$$\pi_1 = V_1, \qquad (20)$$

$$\pi_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \qquad k = 2, 3, \ldots, \qquad (21)$$

where

$$V_k | \alpha \overset{\text{IID}}{\sim} \text{beta}(1, \alpha), \qquad k = 1, 2, \ldots. \qquad (22)$$

Here $\alpha > 0$ is a parameter and $\Sigma_{k=1}^{\infty} \pi_k = 1$ with probability 1 (Ishwaran and James, 2001).

3.3. Parameterizations
Exponential parameterizations of the conditional PMF (12) are convenient, though other parameterizations may be used as well.

The dyad-bound between-neighbourhood PMFs can be written as

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_{\text{B},i,j} = y_{\text{B},i,j}, Y_{\text{B},j,i} = y_{\text{B},j,i} | \mathbf{Z} = \mathbf{z}) = \exp\{\langle \boldsymbol{\theta}_{\text{B}}, s_{\text{B},i,j}(y_{\text{B},i,j}, y_{\text{B},j,i}) \rangle - \psi_{\text{B},i,j}(\boldsymbol{\theta}_{\text{B}})\}, \qquad (23)$$

where $s_{\text{B}, i, j}(y_{\text{B},i,j}, y_{\text{B},j,i})$ is a vector of between-neighbourhood sufficient statistics, $\boldsymbol{\theta}_{\text{B}}$ is a vector

of between-neighbourhood natural parameters and $\psi_{\mathrm{B},i,j}(\theta_{\mathrm{B}})$ is the between-neighbourhood log-normalizing constant,

$$\psi_{\mathrm{B},i,j}(\theta_{\mathrm{B}}) = \log\left[\sum_{y'_{\mathrm{B},i,j} \in \mathbb{Y}_{\mathrm{B},i,j}} \sum_{y'_{\mathrm{B},j,i} \in \mathbb{Y}_{\mathrm{B},j,i}} \exp\{\langle \theta_{\mathrm{B}}, s_{\mathrm{B},i,j}(y'_{\mathrm{B},i,j}, y'_{\mathrm{B},j,i})\rangle\}\right]. \qquad (24)$$

The between-neighbourhood sufficient statistics $s_{\mathrm{B},i,j}(y_{\mathrm{B},i,j}, s_{\mathrm{B},j,i})$ may be functions of edges $y_{\mathrm{B},i,j}$ and $y_{\mathrm{B},j,i}$ and covariates. In the interest of model parsimony, we assume that the between-neighbourhood parameter vector θ_{B} is constant across dyads.

The within-neighbourhood PMFs can be written as

$$\mathbb{P}_{\theta}(\mathbf{Y}_{k,k} = \mathbf{y}_{k,k}|\mathbf{Z} = \mathbf{z}) = \exp\{\langle \theta_{\mathrm{W},k}, s_{\mathrm{W},k}(\mathbf{y}_{k,k})\rangle - \psi_{\mathrm{W},k}(\theta_{\mathrm{W},k})\}, \qquad (25)$$

where $s_{\mathrm{W},k}(\mathbf{y}_{k,k})$ is a vector of within-neighbourhood sufficient statistics, $\theta_{\mathrm{W},k}$ is a vector of within-neighbourhood natural parameters and $\psi_{\mathrm{W},k}(\theta_{\mathrm{W},k})$ is the within-neighbourhood log-normalizing constant,

$$\psi_{\mathrm{W},k}(\theta_{\mathrm{W},k}) = \log\left[\sum_{\mathbf{y}'_{k,k} \in \mathbb{Y}_{k,k}} \exp\{\langle \theta_{\mathrm{W},k}, s_{\mathrm{W},k}(\mathbf{y}'_{k,k})\rangle\}\right]. \qquad (26)$$

The within-neighbourhood sufficient statistics $s_{\mathrm{W},k}(\mathbf{y}_{k,k})$ may include interactions, such as the number of triangles within neighbourhood k, which induce dependence within neighbourhoods. In addition, covariates can be used.

The exponential parameterization of the between- and within-neighbourhood PMFs implies that the conditional PMF of $\mathbf{Y}$ given $\mathbf{Z}$ can be written as

$$\mathbb{P}_{\theta}(\mathbf{Y} = \mathbf{y}|\mathbf{Z} = \mathbf{z}) = \exp\{\langle \eta(\theta, \mathbf{z}), s(\mathbf{y})\rangle - \psi(\theta, \mathbf{z})\}, \qquad (27)$$

where the vector of parameters $\eta(\theta, \mathbf{z})$ is a linear function of the vectors of between- and within-neighbourhood parameters, the vector of sufficient statistics $s(\mathbf{y})$ is a linear function of between- and within-neighbourhood vectors of sufficient statistics and the log-normalizing constant $\psi(\theta, \mathbf{z})$ is given by

$$\psi(\theta, \mathbf{z}) = \sum_{k=1}^{K} \sum_{l=1}^{k-1} \sum_{i \in \mathcal{A}_k, j \in \mathcal{A}_l} \psi_{\mathrm{B},i,j}(\theta_{\mathrm{B}}) + \sum_{k=1}^{K} \psi_{\mathrm{W},k}(\theta_{\mathrm{W},k}). \qquad (28)$$

The between- and within-neighbourhood parameter vectors θ_{B} and $\theta_{\mathrm{W},k}$ index exponential families and therefore conjugate priors exist, though direct sampling from the resulting full conditional distributions is infeasible. In the absence of computational advantages, multivariate Gaussian priors are convenient alternatives:

$$\begin{aligned} \theta_{\mathrm{B}}|\mu_{\mathrm{B}}, \Sigma_{\mathrm{B}}^{-1} &\sim \mathrm{MVN}(\mu_{\mathrm{B}}, \Sigma_{\mathrm{B}}^{-1}), \\ \theta_{\mathrm{W},k}|\mu_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1} &\overset{\mathrm{IID}}{\sim} \mathrm{MVN}(\mu_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}), \qquad k = 1, 2, \ldots, \end{aligned} \qquad (29)$$

where μ_{B} and μ_{W} are mean parameter vectors and Σ_{B}^{-1} and Σ_{W}^{-1} are precision matrices of suitable order.

To acknowledge the uncertainty about the hyperparameters α, μ_{W} and Σ_{W}^{-1}, we assign conjugate gamma, multivariate Gaussian and Wishart hyperpriors to α, μ_{W} and Σ_{W}^{-1} respectively.

3.4. Special cases and related models

Special cases of interest are the block models of Wang and Wong (1987), the stochastic block models of Nowicki and Snijders (2001) and the mixed membership models of Airoldi *et al.*

(2008). These models assume that edge variables are independent conditionally on an observed or unobserved partition of the set of nodes into subsets, which are called blocks and correspond to neighbourhoods. These models satisfy local dependence but do not allow scientists to specify directly the nature of interactions of interest, as discussed in Section 2.1.

The models of Strauss and Ikeda (1990) were discussed in Section 3.1. The usefulness of the models is limited, because neighbourhood structure may either not be observed, in which case the models cannot be used, or neighbourhood structure is observed but is unsuitable in the sense that the observed number of neighbourhoods is small and the number of nodes is large.

Last, the model of Koskinen (2009) does not restrict dependence to blocks and therefore does not satisfy local dependence.

4. Bayesian inference

We focus on Bayesian inference without observed neighbourhood structure, which is the most common and most challenging case. A Bayesian approach must overcome multiple obstacles. The most serious obstacle is the fact that with positive probability one or more neighbourhoods k contains $n_k \gg 5$ nodes and thus one or more within-neighbourhood log-normalizing constants, which are log-sums of $\exp\{\binom{n_k}{2}\log(2)\}$ terms (see equation (26)), is intractable. To facilitate posterior computations, we approximate the prior and augment the posterior.

We describe the approximation of the prior in Section 4.1, discuss the augmentation of the posterior and sampling from the augmented posterior in Section 4.2, with additional details in the on-line supplements A and B, and address the non-identifiability of within-neighbourhood parameter vectors and membership indicators in supplement C.

4.1. Prior truncation

The stick breaking prior of Section 3.2 can be approximated by a truncated stick breaking prior along the lines of Ishwaran and James (2001), which facilitates posterior computations.

We choose a maximum number of neighbourhoods, which is denoted by $K_{\max}$. Some general advice concerning the choice of $K_{\max}$ was given by Ishwaran and James (2001). We are here more concerned with the goodness of fit of the model than the approximation of the stick breaking prior and choose $K_{\max}$ in accordance. In practice, we choose $K_{\max}$ by either

  (a) trying out multiple values of $K_{\max}$ and comparing the goodness of fit of the model,
  (b) exploiting on-the-ground knowledge or
  (c) setting $K_{\max} = n$.

Strategy (a) is motivated by the fact that model estimation is time consuming and the computing time increases with $K_{\max}$; thus there is an incentive to choose $K_{\max}$ as small as possible. We demonstrate strategies (a) and (b) in Section 5.3.

Given $K_{\max}$, the membership probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{K_{\max}})$ are constructed by truncated stick breaking (Ishwaran and James, 2001):

$$\pi_1 = V_1, \tag{30}$$

$$\pi_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \qquad k = 2, \ldots, K_{\max}, \tag{31}$$

where

$$V_k | \alpha \stackrel{\text{IID}}{\sim} \text{beta}(1, \alpha), \qquad k = 1, \ldots, K_{\max} - 1,$$
$$V_{K_{\max}} = 1, \tag{32}$$

where $\alpha > 0$ is a parameter and $V_{K_{\max}} = 1$ ensures that $\Sigma_{k=1}^{K_{\max}} \pi_k = 1$. The truncated stick breaking construction of $\boldsymbol{\pi}$ implies that $\boldsymbol{\pi}$ is generalized Dirichlet distributed, which is conjugate to multinomial sampling (Ishwaran and James, 2001).

4.2. *Posterior augmentation*

Under the truncated prior that was described in Section 4.1, the posterior is of the form

$$p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z}|\mathbf{y}) \propto p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}) \, \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z}=\mathbf{z}) \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z}), \qquad (33)$$

where the truncated prior is of the form

$$p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}) = p(\alpha) \, p(\boldsymbol{\mu}_{\mathrm{W}}) \, p(\Sigma_{\mathrm{W}}^{-1}) \, p(\boldsymbol{\pi}|\alpha) \, p(\boldsymbol{\theta}_{\mathrm{B}}) \prod_{k=1}^{K_{\max}} p(\boldsymbol{\theta}_{\mathrm{W},k}|\boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}), \qquad (34)$$

where $\boldsymbol{\theta}_{\mathrm{W}} = (\boldsymbol{\theta}_{\mathrm{W},1}, \ldots, \boldsymbol{\theta}_{\mathrm{W},K_{\max}})$ denotes the within-neighbourhood parameter vectors.

Owing to the fact that the conditional PMF of $\mathbf{Y}$ is not, in general, tractable, the posterior is doubly intractable, implying that standard Markov chain Monte Carlo methods (e.g. Metropolis–Hastings algorithms) cannot be used to sample from the posterior. Auxiliary variable Markov chain Monte Carlo methods for sampling from doubly intractable posteriors arising in complete-data problems were introduced by Møller *et al.* (2006) and extended by Murray *et al.* (2006) and Liang (2010); they have been adapted to networks by Koskinen *et al.* (2010), Caimo and Friel (2011) and Wang and Atchade (2014). We extend them from the complete-data problems that were considered there to the incomplete-data problem that is considered here.

To facilitate posterior computations, we augment α, $\boldsymbol{\mu}_{\mathrm{W}}$, Σ_{W}^{-1}, $\boldsymbol{\pi}$, $\boldsymbol{\theta}_{\mathrm{B}}$, $\boldsymbol{\theta}_{\mathrm{W}}$, $\mathbf{Z}$ and $\mathbf{Y}$ by auxiliary variables $\boldsymbol{\theta}_{\mathrm{W}}^*$, $\mathbf{Z}^*$ and $\mathbf{Y}^*$. The auxiliary variable $\mathbf{Y}^*$ can be interpreted as an auxiliary random graph, $\mathbf{Z}^*$ can be interpreted as an auxiliary neighbourhood structure and $\boldsymbol{\theta}_{\mathrm{W}}^*$ can be interpreted as auxiliary within-neighbourhood parameter vectors. We assume that the joint distribution of α, $\boldsymbol{\mu}_{\mathrm{W}}$, Σ_{W}^{-1}, $\boldsymbol{\pi}$, $\boldsymbol{\theta}_{\mathrm{B}}$, $\boldsymbol{\theta}_{\mathrm{W}}$, $\mathbf{Z}$, $\mathbf{Y}$, $\boldsymbol{\theta}_{\mathrm{W}}^*$, $\mathbf{Z}^*$ and $\mathbf{Y}^*$ is of the form

$$\begin{aligned} p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta}_{\mathrm{W}}^*, \mathbf{z}^*, \mathbf{y}^*) = {} & p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}) \, \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z}=\mathbf{z}) \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z}) \\ & \times q(\boldsymbol{\theta}_{\mathrm{W}}^*, \mathbf{z}^*|\boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z}, \mathbf{y}) \mathbb{P}_{\boldsymbol{\theta}*}(\mathbf{Y}^*=\mathbf{y}^*|\mathbf{Z}^*=\mathbf{z}^*), \end{aligned}$$
$$(35)$$

where $q(\boldsymbol{\theta}_{\mathrm{W}}^*, \mathbf{z}^*|\boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z}, \mathbf{y})$ is a suitable auxiliary distribution, the conditional distributions $\mathbf{Y}$ and $\mathbf{Y}^*$ belong to the same exponential family of distributions and $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}^*)$. The augmented posterior is of the form

$$p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z}, \boldsymbol{\theta}_{\mathrm{W}}^*, \mathbf{z}^*, \mathbf{y}^*|\mathbf{y}) \propto p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \Sigma_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta}_{\mathrm{W}}^*, \mathbf{z}^*, \mathbf{y}^*). \qquad (36)$$

Integrating out the auxiliary variables $\boldsymbol{\theta}_{\mathrm{W}}^*$, $\mathbf{Z}^*$ and $\mathbf{Y}^*$ results in the posterior of α, $\boldsymbol{\mu}_{\mathrm{W}}$, Σ_{W}^{-1}, $\boldsymbol{\pi}$, $\boldsymbol{\theta}_{\mathrm{B}}$, $\boldsymbol{\theta}_{\mathrm{W}}$ and $\mathbf{Z}$. Whereas sampling from the posterior (33) is infeasible, sampling from the augmented posterior (36) and integrating out the auxiliary variables $\boldsymbol{\theta}_{\mathrm{W}}^*$, $\mathbf{Z}^*$ and $\mathbf{Y}^*$ turns out to be feasible. We discuss Markov chain Monte Carlo steps and improved Markov chain Monte Carlo sampling through variational methods in the on-line supplement.

5.   Assessment of local and global models of transitive closure

We compare random graph models with local and global dependence by comparing

  (a) prior predictions of graphs to assess whether models can produce data that resemble real world networks (Section 5.1),

(b) sampling distributions of Bayesian point and interval estimators (Section 5.2) and
(c) posterior predictions of graphs to assess whether models make sense in the light of observed data (Section 5.3).

We consider undirected, binary edge variables, i.e. $Y_{i,j} \in \{0,1\}$ and $Y_{i,j} = Y_{j,i}$ with probability 1, and random graph models capturing transitive closure, because it is one of the most fundamental and problematic forms of dependence. A well-known random graph capturing transitive closure is the triangle model, which is an ERGM of the form (1) with the number of edges $y_{i,j}$ and triangles $y_{i,j} y_{j,h} y_{i,h}$ as sufficient statistics (Jonasson, 1999; Handcock *et al.*, 2008). Its natural relative is the random graph model of the form (23) and (25), where the between-neighbourhood sufficient statistics are the edges $y_{i,j}$ between nodes i and j in neighbourhoods k and l, and the within-neighbourhood sufficient statistics are the number of edges $y_{i,j}$ and triangles $y_{i,j} y_{j,h} y_{i,h}$ within neighbourhoods k. We refer to the two models as the global and local triangle model. The name global triangle model is motivated by the fact that the model is a special case of a Markov random graph model with unbounded neighbourhoods—as explained in Section 2.2—and therefore does not satisfy local dependence.

5.1. Comparison of local and global models of transitive closure

In this subsection, we assess whether models are realistic by considering the prior predictive distributions of the statistics. As discussed in Section 1, the global triangle model places much probability mass on the relative boundary of the convex hull of $\{s(\mathbf{y}) : \mathbf{y} \in \mathbb{Y}\}$. As networks with extreme values on network statistics may be considered odd, one approach to assessing the realism of a model is to consider the distribution of the statistics that it produces.

The prior predictive distribution under the global triangle model can be written as

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \int p(\boldsymbol{\theta}) \, \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}) \, \mathrm{d}\boldsymbol{\theta}, \tag{37}$$

where $p(\boldsymbol{\theta})$ denotes the prior. On the basis of experience, values of θ_1 outside $(-5, 0)$ and values of θ_2 outside $(0, 5)$ index near-degenerate distributions. Therefore, we choose independent, uniform priors given by $\theta_1 \sim \mathrm{uniform}(-5, 0)$ and $\theta_2 \sim \mathrm{uniform}(0, 5)$.

The prior predictive distribution under the local triangle model can be written as

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \int \ldots \int \sum_{\mathbf{z} \in \mathcal{Z}} p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \boldsymbol{\Sigma}_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}})$$
$$\times \mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z}) \, \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}) \, \mathrm{d}\alpha \, \mathrm{d}\boldsymbol{\mu}_{\mathrm{W}} \, \mathrm{d}\boldsymbol{\Sigma}_{\mathrm{W}}^{-1} \, \mathrm{d}\boldsymbol{\pi} \, \mathrm{d}\boldsymbol{\theta}_{\mathrm{B}} \, \mathrm{d}\boldsymbol{\theta}_{\mathrm{W}}, \tag{38}$$

where $p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \boldsymbol{\Sigma}_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}})$ denotes the prior. We assign independent Gaussian priors with means -1 and 1 and standard deviations 0.25 and 0.25 to within-neighbourhood parameters $\theta_{\mathrm{W},k,1}$ and $\theta_{\mathrm{W},k,2}$ respectively. The marginal priors of $\theta_{\mathrm{W},k,1}$ and $\theta_{\mathrm{W},k,2}$ ensure that most of the prior probability mass of $\theta_{\mathrm{W},k,1}$ is concentrated on $(-2, 0)$ and most of the prior probability mass of $\theta_{\mathrm{W},k,2}$ is concentrated on $(0, 2)$, which cover the most reasonable value of $\theta_{\mathrm{W},k,1}$ and $\theta_{\mathrm{W},k,2}$ respectively. To respect the sparse nature of graphs, we assume that the between-neighbourhood parameter θ_{B} is governed by a Gaussian prior with mean μ_{B} and standard deviation 1, where the Gaussian prior is centred at $\mu_{\mathrm{B}} = 3/(n-1)$, the value of θ_{B} under which the expected number of edges of nodes between neighbourhoods is at most 3. The prior of $\boldsymbol{\pi}$ is given by the Dirichlet$(10, \ldots, 10)$ prior.

We generated 1000 model predictions from the local triangle model with $K = 150$ neighbourhoods, $n = 1000$ nodes, and $N = 499500$ edge variables and 1000 realizations from the global
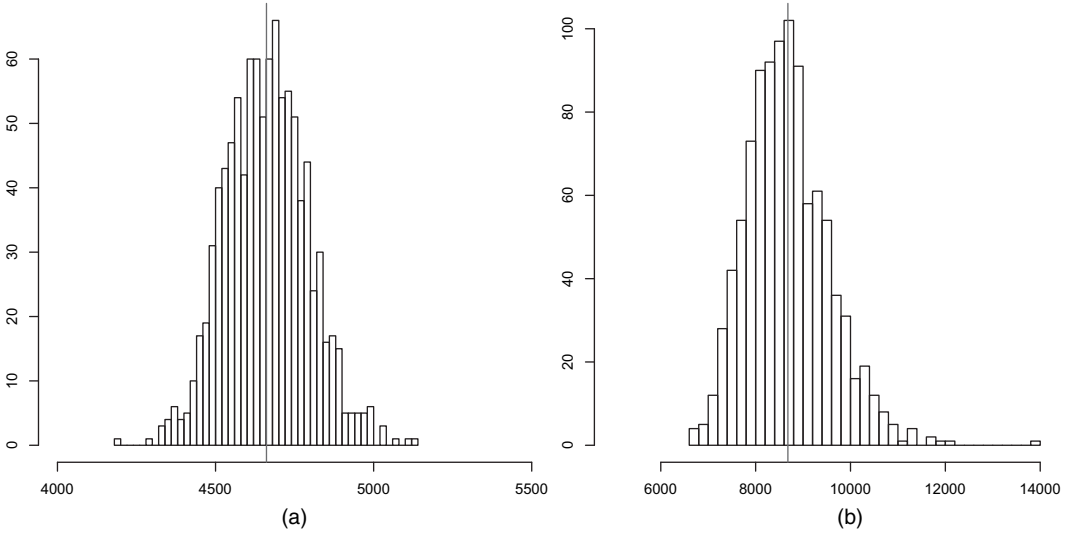
Fig. 2. Prior predictions of the number of (a) edges and (b) triangles under the local triangle model with parametric prior and $K = 150$ neighbourhoods and $N = 499500$ edge variables: most prior predictive mass is concentrated around the means of the prior predictions (|)

triangle model with $n = 100$ nodes and $N = 4950$ edge variables; the difference in the size of the graphs is because the flawed nature of the global triangle model makes it infeasible to sample much larger graphs than $N = 4950$. Monte Carlo samples of size 1000 were generated from the prior and, for every one of the draws from the prior, a prediction was generated by a Markov chain of length 10 million for the local triangle model and 100000 for the global triangle model, accepting the final draw of the Markov chain as a draw from the prior predictive distribution; the sample size is proportional to the size of the graphs.

Fig. 1 shows prior predictions of the number of edges and triangles under the global triangle model. The bulk of the prior predictive mass is placed on extreme graphs with few edges and triangles and graphs with almost all possible edges and triangles. We note that the behaviour of the global triangle model tends to deteriorate as the size of the graph increases, as discussed in Section 1. In contrast, Fig. 2 demonstrates that random graph models with local dependence place much prior predictive mass around the mean and all of its mass on graphs which resemble real world networks, i.e. graphs where the average number of edges of nodes ranges from 4 to 6 and where the number of triangles is a small multiple of the number of edges. It is worth repeating that the number of edge variables is 499500, which demonstrates that random graph models with local dependence are well behaved when the number of edge variables is large, in sharp contrast with conventional ERGMs.

In short, the model predictions confirm what the theoretical results of Section 2.4 suggested: in contrast with conventional ERGMs, random graph models with local dependence are capable of generating graphs which resemble real world networks and can thus be recommended *a priori* as models of real world networks.

5.2. Sampling distributions of Bayesian point and interval estimators
We shed light on the frequentist properties of estimators by simulation. We focus on random graph models with local dependence, because flawed models of the form (1) generate many graphs which fall onto the relative boundary of the convex hull of $\{s(\mathbf{y}) : \mathbf{y} \in \mathbb{Y}\}$ and make

statistical inference problematic (e.g. Barndorff-Nielsen (1978), page 151, Handcock (2003), Rinaldo *et al.* (2009), Koskinen *et al.* (2010) and Bhamidi *et al.* (2011)).

Here, we focus on the frequentist properties of posterior point estimators and interval estimators of the local triangle model with $K = 7$ neighbourhoods, $n = 50$ nodes and $N = 1225$ edge variables. We generated 1000 graphs from the local triangle model by using the same prior as used in Section 5.1. To infer from the simulated graphs to the data-generating values of the parameters, we used two priors: a parametric Dirichlet$(\alpha, \ldots, \alpha)$ prior for π with $K = 7$ with a gamma$(1, 0.1)$ hyperprior for α, and a non-parametric truncated stick breaking prior with $K_{\max} = 7$ with a gamma$(1, 0.1)$ hyperprior for α. In both cases, the marginal priors of $\theta_{W,1}$, $\theta_{W,2}$ and θ_B are independently $N(0, 100)$. We construct 1000 Markov chains with 100000 iterations, discarding the first 20000 iterations as burn-in and recording every 10th post-burn-in iterations.

In practice, within-neighbourhood parameters are of primary interest, because it is the within-neighbourhood models which capture the dependences of interest. Here, we focus on the within-neighbourhood means $\mu_{W,1}$ and $\mu_{W,2}$. The sampling distributions of posterior point estimators for $\mu_{W,1}$ and $\mu_{W,2}$ under parametric and non-parametric priors are shown in Figs 3 and 4. Despite the small number of neighbourhoods $K = 7$, the data-generating parameter of the Gaussian mean μ_W of the $K = 7$ within-neighbourhood parameters $\theta_{W,1}, \ldots, \theta_{W,7}$ can be recovered in the sense that the posterior median clusters around the data-generating parameter value. The distributions are somewhat asymmetric, which is not surprising considering the small number of $K = 7$ neighbourhoods. Table 1 shows that interval estimators, e.g. 95% posterior credibility intervals, have acceptable coverage properties, considering the small number of $K = 7$ neighbourhoods.

The results also suggest that the non-parametric approach seems to outperform the parametric approach, at least in terms of coverage for $\mu_{W,1}$. In addition, we found in the applications in Section 5.3 that the hierarchical non-parametric prior is not overly sensitive to the choice of the hyperparameters of the priors, whereas the hierarchical parametric prior sometimes is sensitive to the choice of hyperparameters, and more so when the specified number of neighbourhoods
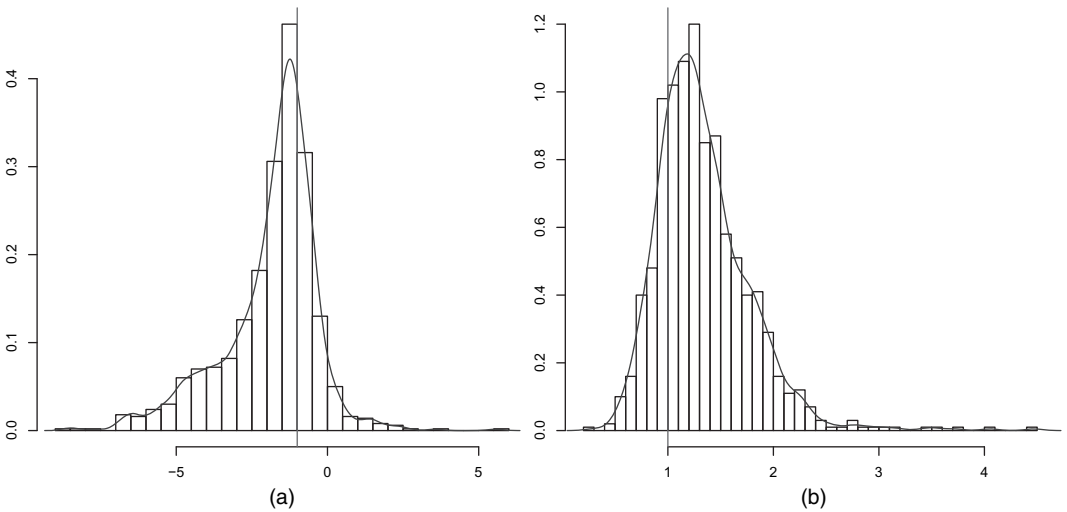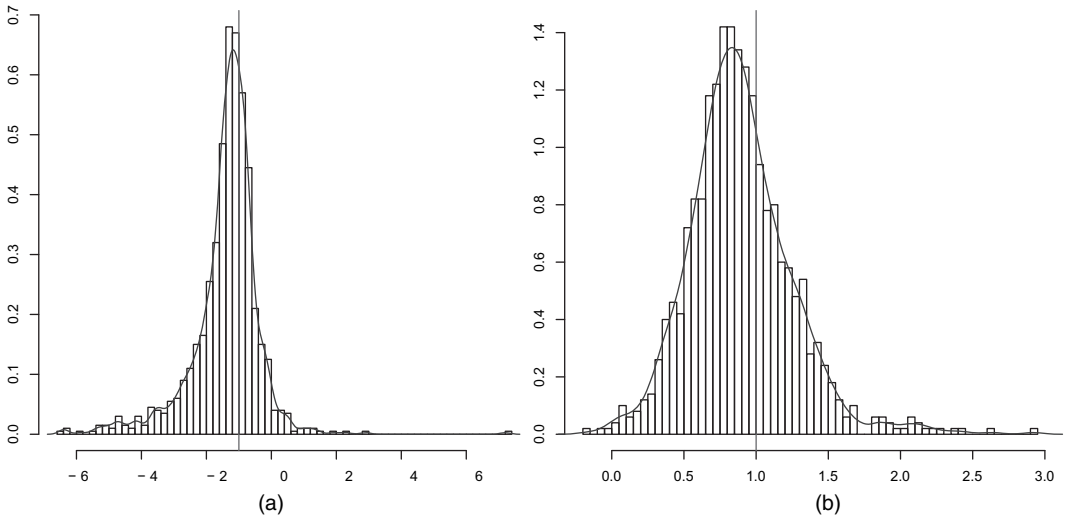

Fig. 3. Sampling distributions of posterior medians of within-neighborhood means (a) $\mu_{W,1}$ and (b) $\mu_{W,2}$ under the local triangle model with parametric prior and $K = 7$ neighbourhoods and $N = 1225$ edge variables (|, data-generating values): despite the small number of $K = 7$ neighbourhoods, the data-generating values of the Gaussian mean μ_W of the $K = 7$ within-neighbourhood parameters $\theta_{W,1}, \ldots, \theta_{W,7}$ can be recovered

Fig. 4. Sampling distributions of posterior medians of within-neighbourhood means (a) $\mu_{W,1}$ and (b) $\mu_{W,2}$ under the local triangle model with non-parametric prior and $K_{max} = 7$ neighbourhoods and $N = 1225$ edge variables (|, data-generating values): despite the small number of neighbourhoods $K_{max} = 7$, the data-generating values of the Gaussian mean μ_W of the $K_{max} = 7$ within-neighbourhood parameters $\theta_{W,1}, \ldots, \theta_{W,7}$ can be recovered

Table 1. Frequentist properties of posterior medians and 95% posterior credibility intervals of within-neighbourhood means $\mu_{W,1}$ and $\mu_{W,2}$†

Prior	Parameter	0.025-quantile	0.50-quantile	0.975-quantile	Coverage (%)
Parametric	$\mu_{W,1} = -1$	-5.75	-1.48	0.48	86
Parametric	$\mu_{W,2} = 1$	0.69	1.26	2.27	98
Non-parametric	$\mu_{W,1} = -1$	-4.19	-1.25	0.26	95
Non-parametric	$\mu_{W,2} = 1$	0.26	0.86	1.68	>99

†Despite the small number of $K_{max} = 7$ neighbourhoods, the data-generating values of the Gaussian mean μ_W of the $K_{max} = 7$ within-neighbourhood parameters $\theta_{W,1}, \ldots, \theta_{W,7}$ can be recovered.

exceeds the true number of neighbourhoods. Thus, the non-parametric approach seems to have advantages over the parametric approach.

5.3. Application to a terrorist network and a social network

A natural approach to comparing ERGMs and random graph models with local dependence is based on their predictions about observable quantities. Hunter *et al.* (2008) argued that, in practice, it is imperative to generate model predictions to assess the goodness of fit of network models.

In this section, we compare ERGMs and random graph models with local dependence in terms of posterior predictions in two real world networks with ground truth: the terrorist network behind the Bali bombing in 2002 as well as social relationships among novices within a novitiate.
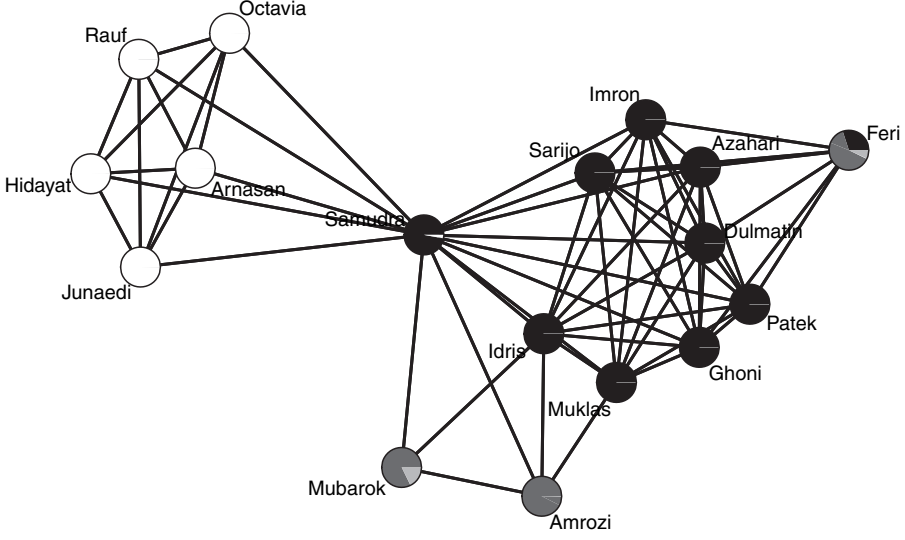
Fig. 5. Terrorist network behind the Bali bombing in 2002 with $N = 136$ edge variables: the shaded pie charts represent posterior membership probabilities; the clustering is a by-product of model estimation and is of secondary interest, but it is comforting that it is consistent with ground truth

The posterior predictive distribution under the global triangle model given data **y** can be written as

$$\mathbb{P}(\dot{\mathbf{Y}} = \dot{\mathbf{y}} | \mathbf{Y} = \mathbf{y}) = \int p(\boldsymbol{\theta} | \mathbf{y}) \, \mathbb{P}_{\boldsymbol{\theta}}(\dot{\mathbf{Y}} = \dot{\mathbf{y}}) \, \mathrm{d}\boldsymbol{\theta}, \tag{39}$$

where $p(\boldsymbol{\theta} | \mathbf{y})$ denotes the posterior. The posterior predictive distribution under the local triangle model can be written as

$$\mathbb{P}(\dot{\mathbf{Y}} = \dot{\mathbf{y}} | \mathbf{Y} = \mathbf{y}) = \int \dots \int \sum_{\mathbf{z} \in \mathcal{Z}} p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \boldsymbol{\Sigma}_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z} | \mathbf{y})$$
$$\times \mathbb{P}_{\boldsymbol{\theta}}(\dot{\mathbf{Y}} = \dot{\mathbf{y}} | \mathbf{Z} = \mathbf{z}) \, \mathrm{d}\alpha \, \mathrm{d}\boldsymbol{\mu}_{\mathrm{W}} \, \mathrm{d}\boldsymbol{\Sigma}_{\mathrm{W}}^{-1} \, \mathrm{d}\boldsymbol{\pi} \, \mathrm{d}\boldsymbol{\theta}_{\mathrm{B}} \, \mathrm{d}\boldsymbol{\theta}_{\mathrm{W}}, \tag{40}$$

where $p(\alpha, \boldsymbol{\mu}_{\mathrm{W}}, \boldsymbol{\Sigma}_{\mathrm{W}}^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_{\mathrm{B}}, \boldsymbol{\theta}_{\mathrm{W}}, \mathbf{z} | \mathbf{y})$ denotes the posterior. Independent priors $\theta_i \sim N(0, 25)$ are used in the case of the ERGM and independent priors $\alpha \sim \mathrm{gamma}(1, 1)$, $\mu_{\mathrm{W},i} \sim N(0, 1)$ and $\sigma_{\mathrm{W},i}^{-2} \sim \mathrm{gamma}(10, 10)$ in the case of the random graph model with local dependence. $120\,000$ draws from the posterior predictive distribution of the ERGM were generated by the Markov chain Monte Carlo algorithm of Caimo and Friel (2011), with a burn-in of $20\,000$ and saving every 10th post-burn-in draw, and $1\,200\,000$ draws from the posterior predictive distribution of the random graph model with local dependence were generated by the Markov chain Monte Carlo algorithm of Section 4, with a burn-in of $200\,000$ and saving every 100th post-burn-in draw.

5.3.1. Terrorist network behind Bali bombing in 2002
The structure of terrorist networks is of interest with a view to understanding how terrorists communicate, to identify cells (i.e. subsets of terrorists), to isolate cells and to dismantle them. We consider here the network of terrorists behind the Bali, Indonesia, bombing in 2002, killing 202 (Koschade, 2006). The 17 terrorists who carried out the bombing were members of the south-east Asian al-Qaeda affiliate Jemaah Islamiyah. The terrorist network can be represented
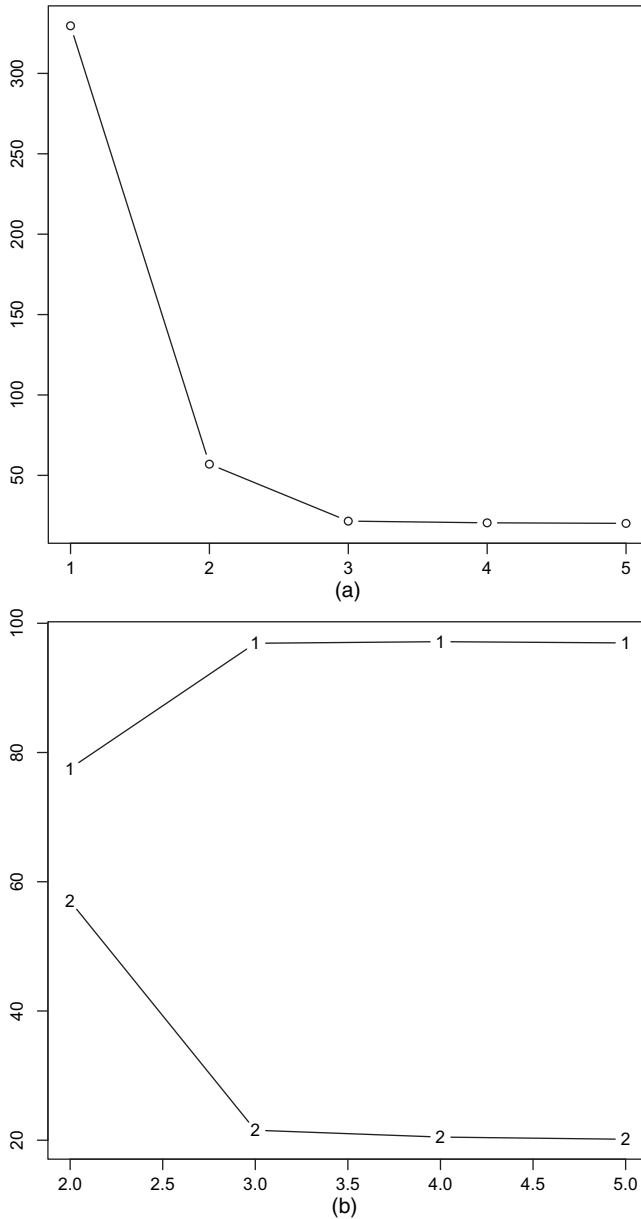
Fig. 6. Terrorist network: root-mean-square deviation of the number of triangles plotted against K_{max} = 2, 3, 4, 5 neighbourhoods: (a) the global triangle model with $K_{max} = 1$ is far inferior to the local triangle model with $K = 2, 3, 4, 5$ neighbourhoods; (b) the stochastic block model (—1—) is likewise far inferior to the local triangle model (—2—)

by a graph with $n = 17$ nodes and $N = 136$ edge variables, where $Y_{i,j} = 1$ if terrorists i and j were in contact before the bombing and $Y_{i,j} = 0$ otherwise. The terrorist network is shown in Fig. 5.

We start by determining the maximum number of neighbourhoods K_{max} to truncate the prior. Using strategy (a) described in Section 4.1, we compare the local triangle model with $K_{max} = 2, 3, 4, 5$ neighbourhoods in terms of predictive power. Predictive power is taken to be

the root-mean-square deviation of the predicted number of triangles. According to Fig. 6(a), the local triangle model with $K_{max} = 2$ neighbourhoods is far superior to the global triangle model, which corresponds to $K_{max} = 1$ neighbourhood. The local triangle model with $K_{max} = 3$ in turn is superior to the local triangle model with $K_{max} = 2$, but increasing K_{max} from 3 to 5 does not increase the predictive power much. Fig. 6(b) compares the local triangle model with stochastic block models. The stochastic block model that is used here is a special case of the local triangle model where the within-neighbourhood sufficient statistics are reduced to the number of edges, which induces conditional independence of edges within neighbourhoods. Stochastic block
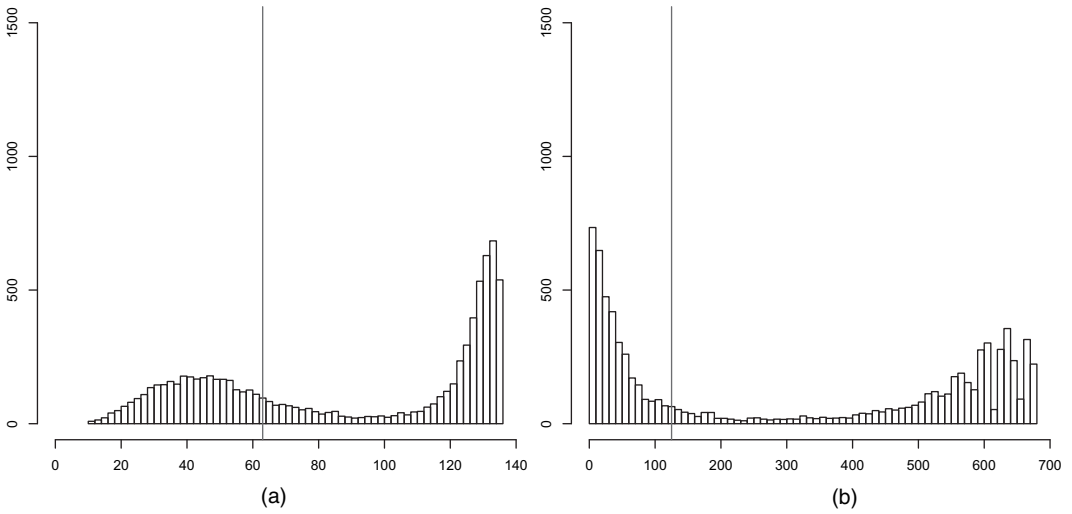
Fig. 7.   Terrorist network: posterior predictions of the number of (a) edges and (b) triangles under the global triangle model (|, observed numbers): although the number of edge variables $N = 136$ is not large, the polarization is evident

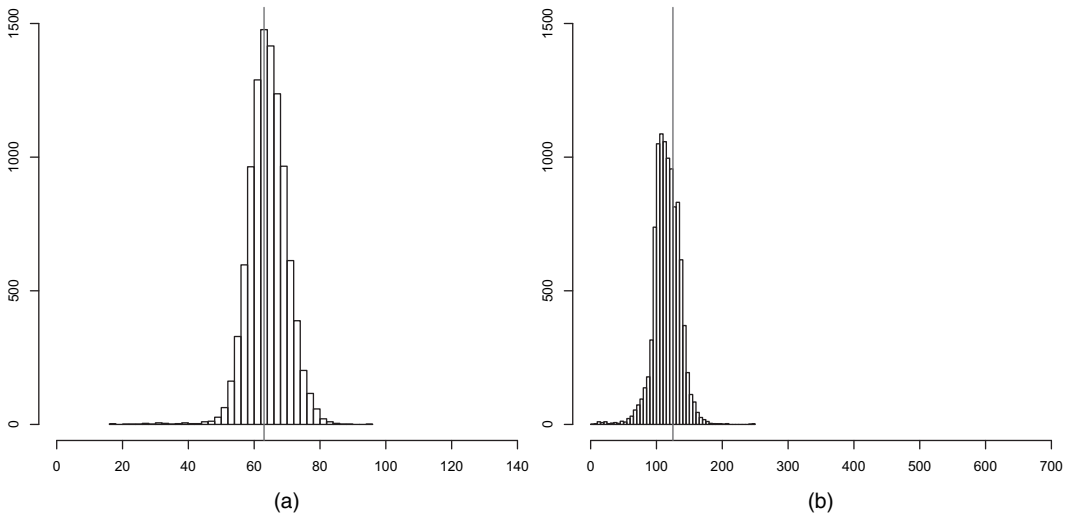**Fig. 8.**   Terrorist network: posterior predictions of the number of (a) edges and (b) triangles under the local triangle model (|, observed numbers): the posterior predictive distributions are unimodal and short tailed, in contrast with Fig. 7

models are special cases of random graph models with local dependence and not appealing when dependence is of substantive interest, because they assume conditional independence within neighbourhoods, as discussed in Sections 2.1 and 3.4. Fig. 6(b) demonstrates that the stochastic block model has much lower predictive power than the local triangle model.

We compare the global and local triangle model with up to $K_{max} = 5$ neighbourhoods in terms of the posterior predictive distribution of the number of edges and triangles, shown in Figs 7 and 8. Under the global triangle model, the posterior predictive distribution is bimodal. In contrast, the posterior predictive distribution under the local triangle model is unimodal and places most mass on graphs which are close to the observed graph in terms of the number of edges and triangles. The fact that the global triangle model places so much mass on dense

Table 2. Terrorist network: posterior of parameters α, $\mu_{W,1}$ and $\mu_{W,2}$

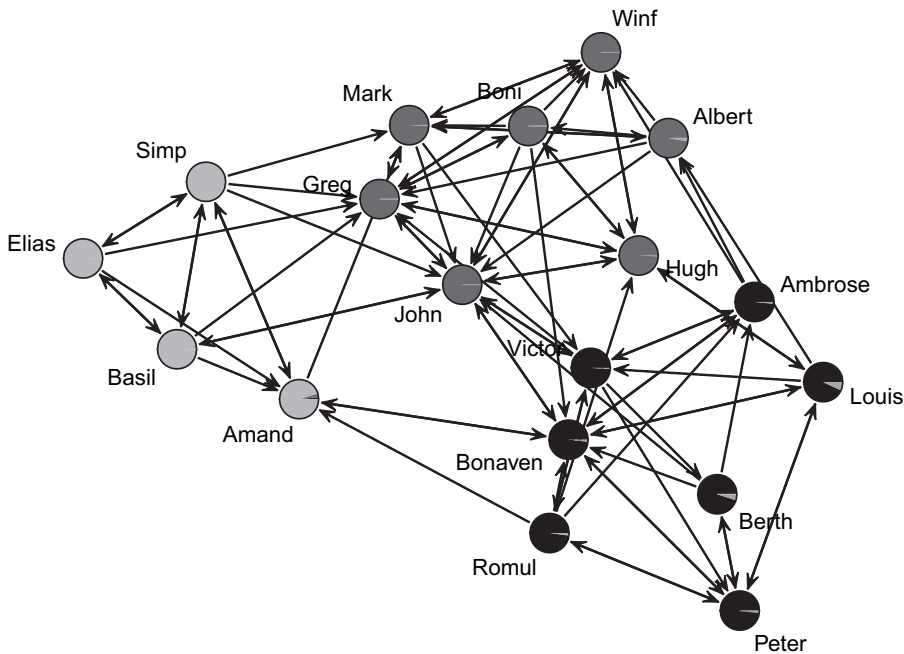| *Parameter*	*0.05-quantile*	*0.50-quantile*	*0.95-quantile*	*Odds of parameter being positive*
α	0.36	1.32	3.43	∞
$\mu_{W,1}$	−1.03	0.45	2.00	2.22
$\mu_{W,2}$	−0.27	0.91	2.22	8.74

Fig. 9.   Sampson network with $N = 306$ edge variables: the shaded pie charts represent posterior membership probabilities; the clustering is a by-product of model estimation and is of secondary interest, but it is comforting that it is consistent with ground truth

graphs with almost all edges and triangles indicates that the global triangle model fits much worse than the random graph model with local dependence, no matter which goodness-of-fit statistics are chosen, because the topology of graphs which are local in nature—such as the observed graph—stands in sharp contrast with the topology of dense graphs in terms of connectivity, centrality, transitivity and other interesting features of graphs (e.g. Kolaczyk (2009)). We note that, although other statistics may be used to compare the two models in terms of goodness of fit, the choice of goodness-of-fit statistics here presents compelling evidence.

The posterior of α, $\mu_{W,1}$ and $\mu_{W,2}$ is shown in Table 2. The mean parameters $\mu_{W,1}$ and $\mu_{W,2}$ governing the within-neighbourhood parameters tend to be both positive—and more so the mean parameter $\mu_{W,2}$ governing the within-neighbourhood triangle parameters—which is not surprising in the light of the large number of edges and triangles within neighbourhoods.
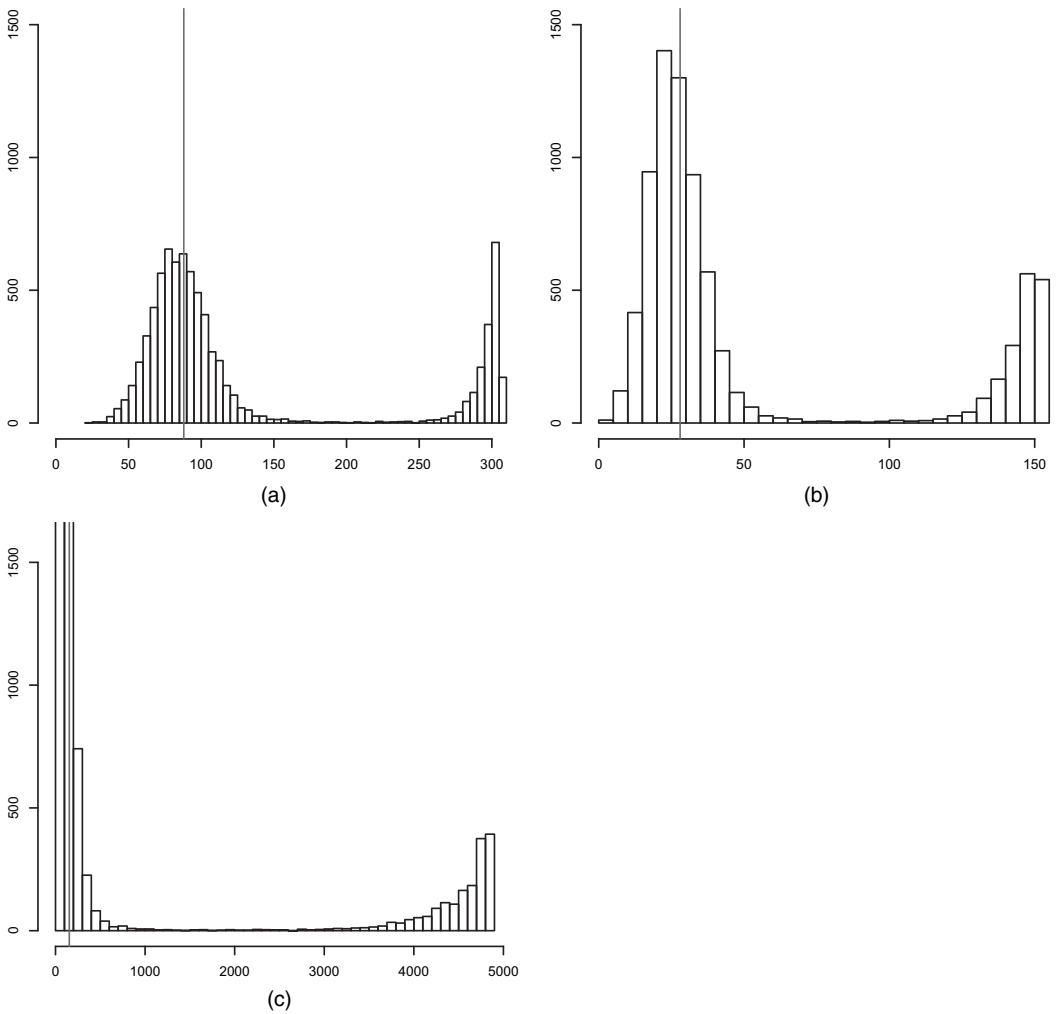

Fig. 10.   Sampson network: posterior predictions of the number of (a) edges, (b) mutual edges and (c) transitive triples under the global triangle model (|, observed numbers); with $N = 306$ edge variables, the polarization is more pronounced than in Fig. 7 with $N = 136$ edge variables

   Last, although the primary purpose of introducing neighbourhoods is the desire to address the model degeneracy and striking lack of fit of ERGMs, predictions of the memberships to neighbourhoods may be of interest as well, e.g. to identify cells. The pie charts in Fig. 5 represent the posterior membership probabilities that were reported by the stochastic relabelling algorithm that is described in the on-line supplement C. The five white-coloured terrorists turn out to be the five members of the so-called support group, which was to supposed to support the so-called main group consisting of all other terrorists. The members of the main group tend to be black coloured, with the exception of Amrozi and Mubarok who are more bright coloured than black coloured. Indeed, although Amrozi and Mubarok belonged to the main group, both resided elsewhere and were almost isolated from the rest of the main group (Koschade, 2006). Most interesting is the membership of Feri. He was a member of the main group and was the
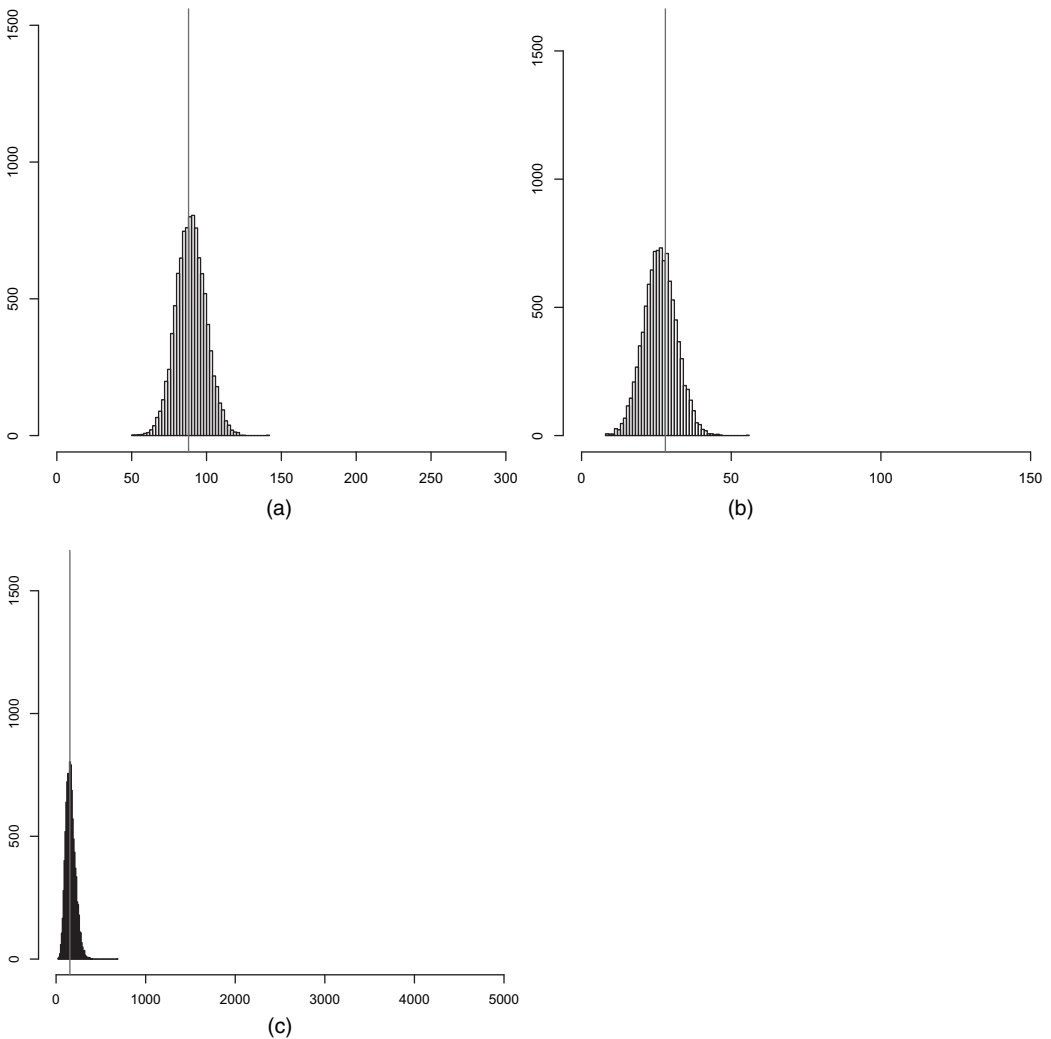

Fig. 11. Sampson network: posterior predictions of the number of (a) edges, (b) mutual edges and (c) transitive triples under the local triangle model (|, observed numbers); the posterior predictive distributions are unimodal and short tailed, in contrast with Fig. 10

suicide bomber who initiated the attack. Feri arrived 2 days before the attack, whereas all other members of the main group had arrived days or weeks earlier and in fact started to leave the night that Feri arrived (Koschade, 2006). As a result, Feri had limited opportunities to communicate with others. In particular, Feri was the one and only member of the main group who did not communicate with the three commanders Muklas (the Jemaah Islamiyah head of operations in Singapore and Malaysia), Samudra (the field commander) and Idris (the logistics commander) (Koschade, 2006). Therefore, the network position of Feri is unique and the uncertainty about his membership is reflected in the posterior membership probability distribution.

In conclusion, random graph models with local dependence capture simple and interesting features of the terrorist network and, under the parameterization that is considered here, posterior membership predictions are consistent with on-the-ground knowledge of the terrorist network.

5.3.2. Social relationships within a novitiate

Sampson (1968) studied social relationships between a group of novices who were preparing to enter a monastic order. The network is a classic data set in social network analysis (White *et al.*, 1976; Handcock *et al.*, 2007) and corresponds to $N = 306$ relationships between the $n = 18$ novices measured at three time points spread out over a 12-month period. We consider here the following directed edge variables $Y_{i,j}$: if novice i liked novice j at any of the three time points, then $Y_{i,j} = 1$; otherwise $Y_{i,j} = 0$. The network is plotted in Fig. 9.

A natural extension of the triangle model to directed graphs is given by a model of the form (1) with the number of edges $y_{i,j}$, mutual edges $y_{i,j}y_{j,i}$ and transitive triples $y_{i,j}y_{j,h}y_{i,h}$ as sufficient statistics. Its local relative is given by the random graph model (23) and (25) with the number of edges $y_{i,j}$ and mutual edges $y_{i,j}y_{j,i}$ between nodes i and j in neighbourhoods k and l as between-neighbourhood sufficient statistics, and the number of edges $y_{i,j}$, mutual edges $y_{i,j}y_{j,i}$
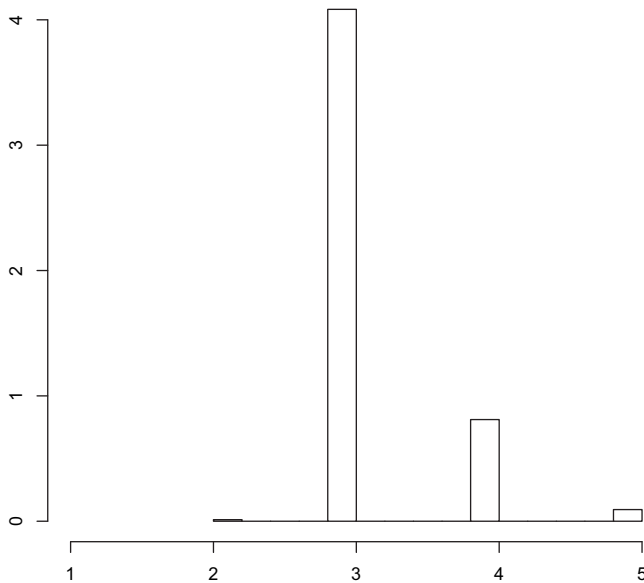

Fig. 12. Sampson network: posterior of the number of non-empty neighbourhoods under the local triangle model; the posterior is consistent with ground truth and confirms that the global triangle model (assuming that all nodes are in one neighbourhood) makes no sense in the light of the observed data

and transitive triples $y_{i,j} y_{j,h} y_{i,h}$ within neighbourhoods k as with-neighbourhood sufficient statistics. Since experts argue that the novices are divided into three or four groups (White *et al.*, 1976; Handcock *et al.*, 2007), we follow strategy (b) that was described in Section 4.1 and set $K_{max} = 5$, which can be considered to be an upper bound on the number of neighbourhoods.

Figs 10 and 11 show posterior predictions of the number of edges, mutual edges and transitive triples. The contrast between the global and local triangle model in terms of goodness of fit is at least as striking as in the case of the terrorist network in Section 5.3.1.

The problematic nature of the global triangle model is underlined by the posterior of the number of non-empty neighbourhoods of the random graph model with local dependence. Fig. 12 shows that the posterior places negligible mass on partitions of the set of nodes where all nodes are assigned to one neighbourhood, which corresponds to the global triangle model. In addition, the posterior mode is 3, which is in line with expert knowledge (White *et al.*, 1976; Handcock *et al.*, 2007).

The neighbourhoods correspond, once again, to physical groups: the posterior membership probabilities that are shown in Fig. 9 agree with the three-group division of novices into 'Loyals', 'Turks' and 'Outcasts' that have been advocated by most experts (White *et al.*, 1976; Handcock *et al.*, 2007).

6. Discussion

We have demonstrated that the notion of local dependence, as introduced here, endows models with desirable properties and makes them amenable to statistical inference. Models with local dependence can be considered to be models of the 'next generation of social network models' (Snijders (2007), page 324), i.e. models which combine latent structure models (e.g. Nowicki and Snijders (2001) and Hoff *et al.* (2002)) and exponential family random graph models (e.g. Frank and Strauss (1986)) in a way that takes advantage of the strengths of ERGMs—i.e. the power of ERGMs to model dependences—while reducing the weaknesses of ERGMs—i.e. the fact that Markov dependence along the lines of Frank and Strauss (1986) is more global than local in nature and are not amenable to statistical inference; note that a partition of the set of nodes can be considered to constitute a latent discrete space.

We believe that random graph models with local dependence constitute a promising and versatile approach to modelling real world networks. Models with small neighbourhoods have been used in physics, machine learning, artificial intelligence and spatial statistics with success, and so have models with M-dependence in time series. We believe that the notion of local dependence that is introduced here is a natural relative of the notions of local dependence in spatial statistics and time series, and as a result can be expected to be useful in applications.

The desirable properties of local dependence suggest that researchers should make every effort to identify and collect information on suitable neighbourhood structures. If suitable neighbourhood structures are not observed, then the auxiliary variable Bayesian methods that are developed here can be used.

We have implemented statistical inference for random graph models with local dependence in the free and open source R package `hergm`, which is publicly available on the Comprehensive R Archive Network.

Acknowledgements

Appendix A: Proofs

A.1. Proof of theorem 1

By local dependence, for all $K > 0$ and $\mathcal{Y}_K \subseteq \mathbb{Y}_K$,

$$
\begin{aligned}
\mathbb{P}_{K+1}(\mathbf{Y}_K \in \mathcal{Y}_K, \mathbf{Y}_{K+1 \setminus K} \in \mathbb{Y}_{K+1 \setminus K}) &= \left\{ \prod_{k=1}^{K} \mathbb{P}_{k,k}(\mathbf{Y}_{k,k} \in \mathcal{Y}_{k,k}) \prod_{l=1}^{k-1} \mathbb{P}_{k,l}(\mathbf{Y}_{k,l} \in \mathcal{Y}_{k,l}, \mathbf{Y}_{l,k} \in \mathcal{Y}_{l,k}) \right\} \\
&\quad \times \left\{ \mathbb{P}_{K+1,K+1}(\mathbf{Y}_{K+1,K+1} \in \mathbb{Y}_{K+1,K+1}) \prod_{l=1}^{K} \mathbb{P}_{K+1,l}(\mathbf{Y}_{k,l} \in \mathbb{Y}_{k,l}, \mathbf{Y}_{l,k} \in \mathbb{Y}_{l,k}) \right\} \\
&= \prod_{k=1}^{K} \mathbb{P}_{k,k}(\mathbf{Y}_{k,k} \in \mathcal{Y}_{k,k}) \prod_{l=1}^{k-1} \mathbb{P}_{k,l}(\mathbf{Y}_{k,l} \in \mathcal{Y}_{k,l}, \mathbf{Y}_{l,k} \in \mathcal{Y}_{l,k}) \\
&= \mathbb{P}_K(\mathbf{Y}_K \in \mathcal{Y}_K). \tag{41}
\end{aligned}
$$

A.2. Proof of theorem 2

By uniform boundedness, $\mathbb{E}(|S_{K,i}|^p) \leqslant C^{pq}$, $p = 1, 2$. Therefore, we can assume that $\mathbb{E}(S_{K,i}) = 0$ without loss of generality, which implies that $\mathbb{E}(W_{K,k}) = 0$, $\mathbb{E}(W_K) = 0$, $\mathbb{E}(B_K) = 0$ and $\mathbb{E}(S_K) = 0$. The variance of S_K can be written as

$$
\begin{aligned}
\mathbb{V}(S_K) &= \sum_{i \in \mathcal{S}_K} \sum_{j \in \mathcal{S}_K} \mathbb{C}(S_{K,i}, S_{K,j}) \leqslant \sum_{i \in \mathcal{S}_K} \sum_{j \in \mathcal{S}_K} |\mathbb{C}(S_{K,i}, S_{K,j})| \\
&= \sum_{k=1}^{K} \sum_{i \in \mathcal{S}_K} \sum_{j \in \mathcal{S}_K} \mathbf{1}_{\mathrm{W},k,i,j} |\mathbb{C}(S_{K,i}, S_{K,j})| + \sum_{i \in \mathcal{S}_K} \sum_{j \in \mathcal{S}_K} \mathbf{1}_{\mathrm{B},i,j} |\mathbb{C}(S_{K,i}, S_{K,j})|, \tag{42}
\end{aligned}
$$

where $\mathbb{C}(S_{K,i}, S_{K,j})$ denotes the covariance of $S_{K,i}$ and $S_{K,j}$ and $\mathbf{1}_{\mathrm{W},k,i,j}$ indicates that both $S_{K,i}$ and $S_{K,j}$ are functions of edge variables in neighbourhood k and neighbourhood k only, whereas $\mathbf{1}_{\mathrm{B},i,j}$ indicates that either $S_{K,i}$ or $S_{K,j}$ or both are functions of between-neighbourhood edge variables. The within-neighbourhood covariances are non-zero but are bounded by uniform boundedness:

$$
\mathbf{1}_{\mathrm{W},k,i,j} |\mathbb{C}(S_{K,i}, S_{K,j})| \leqslant \mathbf{1}_{\mathrm{W},k,i,j} \sqrt{\mathbb{V}(S_{K,i})} \sqrt{\mathbb{V}(S_{K,j})} \leqslant \mathbf{1}_{\mathrm{W},k,i,j} C^{2q}. \tag{43}
$$

Some of the between-neighbourhood covariances may be non-zero as well, because some of the statistics $S_{K,i}$ and $S_{K,j}$ may share edge variables and may therefore be dependent. But $\mathbf{1}_{\mathrm{B},i,j} = 1$ implies that either $S_{K,i}$ or $S_{K,j}$ or both are functions of at least one between-neighbourhood edge variable. Without loss of generality, assume that Y_{i,a_1,b_1} is one of the between-neighbourhood edge variables, and note that both $S_{K,i}$ and $S_{K,j}$ may be functions of Y_{i,a_1,b_1}. The between-neighbourhood covariances can be written as

$$
\mathbf{1}_{\mathrm{B},i,j} |\mathbb{C}(S_{K,i}, S_{K,j})| = \mathbf{1}_{\mathrm{B},i,j} \left| \mathbb{E}\left(\prod_{k=1}^{q} Y_{i,a_k,b_k} Y_{j,a_k,b_k} \right) \right|. \tag{44}
$$

The product $\Pi_{k=1}^{q} Y_{i,a_k,b_k} Y_{j,a_k,b_k}$ can be written as $Y_{i,a_1,b_1}^{p} Y_{-i,a_1,b_1}$, where $p = 1, 2$ because $S_{K,i}$ and $S_{K,j}$ are functions of q distinct edge variables, and Y_{-i,a_1,b_1} is the product of $2q - p$ edge variables distinct from Y_{i,a_1,b_1}. By the independence of the between-neighbourhood edge variable Y_{i,a_1,b_1} and uniform boundedness,

$$
\begin{aligned}
\mathbf{1}_{\mathrm{B},i,j} \left| \mathbb{E}\left(\prod_{k=1}^{q} Y_{i,a_k,b_k} Y_{j,a_k,b_k} \right) \right| &= \mathbf{1}_{\mathrm{B},i,j} |\mathbb{E}(Y_{i,a_1,b_1}^{p}) \mathbb{E}(Y_{-i,a_1,b_1})| \\
&\leqslant \mathbf{1}_{\mathrm{B},i,j} \mathbb{E}(|Y_{i,a_1,b_1}^{p}|) \mathbb{E}(|Y_{-i,a_1,b_1}|) \leqslant \mathbf{1}_{\mathrm{B},i,j} \mathbb{E}(|Y_{i,a_1,b_1}^{p}|) C^{2q-p}. \tag{45}
\end{aligned}
$$

By sparsity, $\mathbb{E}(|Y_{i,a_1,b_1}|^p) \leqslant A n^{-\delta}, \delta > d_j, p = 1, 2$. Thus, by expressions (44) and (45) and sparsity, all covariances vanish in the limit, with the exception of within-neighbourhood covariances. As a result,

$$\lim_{K\to\infty} \mathbb{V}(S_K) = \lim_{K\to\infty} \sum_{k=1}^{K} \sum_{i\in\mathcal{S}_K} \sum_{j\in\mathcal{S}_K} \mathbf{1}_{W,k,i,j}\, \mathbb{C}(S_{K,i}, S_{K,j}) = \lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{V}(W_{K,k}), \qquad (46)$$

$$\lim_{K\to\infty} \mathbb{V}(W_K) = \lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{V}(W_{K,k}), \qquad (47)$$

$$\lim_{K\to\infty} \mathbb{V}(B_K) = 0. \qquad (48)$$

Since the subsets of nodes $\mathcal{A}_k$ contain at most $M < \infty$ nodes and thus subsets $\mathcal{S}_{K,k} \subseteq \times_{k=1}^{d} \mathcal{A}_k$ contain at most $M^d < \infty$ elements, the within-neighbourhood variances $\mathbb{V}(W_{K,k})$ are bounded:

$$\mathbb{V}(W_{K,k}) = \sum_{i\in\mathcal{S}_K} \sum_{j\in\mathcal{S}_K} \mathbf{1}_{W,k,i,j}\, \mathbb{C}(S_{K,i}, S_{K,j}) \leqslant \sum_{i\in\mathcal{S}_K} \sum_{j\in\mathcal{S}_K} \mathbf{1}_{W,k,i,j}\, |\mathbb{C}(S_{K,i}, S_{K,j})|$$
$$\leqslant \sum_{i\in\mathcal{S}_K} \sum_{j\in\mathcal{S}_K} \mathbf{1}_{W,k,i,j}\, C^{2q} \leqslant M^{2d} C^{2q}. \qquad (49)$$

By uniform boundedness and expressions (47) and (49), the within-neighbourhood sums $W_{K,k}$ satisfy Lyaponouv's and thus Lindeberg's condition:

$$\lim_{K\to\infty} \sum_{k=1}^{K} \frac{\mathbb{E}(|W_{K,k}|^4)}{[\mathbb{V}(W_K)]^2} \leqslant \lim_{K\to\infty} \sum_{k=1}^{K} \frac{M^{2d} C^{2q}\, \mathbb{E}(|W_{K,k}|^2)}{[\mathbb{V}(W_K)]^2}$$
$$= \frac{M^{2d} C^{2q} \lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{V}(W_{K,k})}{\lim_{K\to\infty} \mathbb{V}(W_K) \lim_{K\to\infty} \mathbb{V}(W_K)} = \frac{M^{2d} C^{2q} \lim_{K\to\infty} \mathbb{V}(W_K)}{\lim_{K\to\infty} \mathbb{V}(W_K) \lim_{K\to\infty} \mathbb{V}(W_K)} = 0. \qquad (50)$$

By result (50), the within-neighbourhood sums $W_{K,k}$ satisfy the uniform asymptotic negligibility condition (10) (e.g. Resnick (1999), page 315). By expressions (49) and (50) and the Lindeberg–Feller central limit theorem (e.g. Billingsley (1995), page 359, theorem 27.2) applied to the double sequence of random variables $W_K = W_{K,1} + \ldots + W_{K,K}$, $K = 1, 2, \ldots,$

$$\frac{W_K}{\sqrt{\mathbb{V}(W_K)}} \xrightarrow{\mathrm{d}} N(0, 1) \qquad \text{as } K \to \infty. \qquad (51)$$

By Chebyshev's inequality and result (48), for all $\varepsilon > 0$,

$$\lim_{K\to\infty} \mathbb{P}(|B_K| > \varepsilon) \leqslant \lim_{K\to\infty} \frac{1}{\varepsilon^2} \mathbb{V}(B_K) = 0, \qquad (52)$$

implying

$$B_K \xrightarrow{\mathrm{p}} 0 \qquad \text{as } K \to \infty. \qquad (53)$$

By definition, $S_K = W_{K,1} + \ldots + W_{K,K} + B_K$; thus result (11) follows from expressions (51) and (53) and Slutsky's theorem.

References

Airoldi, E., Blei, D., Fienberg, S. and Xing, E. (2008) Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, **9**, 1981–2014.

Barndorff-Nielsen, O. E. (1978) *Information and Exponential Families in Statistical Theory*. New York: Wiley.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* B, **36**, 192–225.

Bhamidi, S., Bresler, G. and Sly, A. (2011) Mixing time of exponential random graphs. *Ann. Appl. Probab.*, **21**, 2146–2170.

Billingsley, P. (1995) *Probability and Measure*, 3rd edn. New York: Wiley.

Butts, C. T. (2011) Bernoulli graph bounds for general random graph models. *Sociol. Methodol.*, **41**, 299–345.

Caimo, A. and Friel, N. (2011) Bayesian inference for exponential random graph models. *Socl Netwrks*, **33**, 41–55.

Chatterjee, S. and Diaconis, P. (2013) Estimating and understanding exponential random graph models. *Ann. Statist.*, **41**, 2428–2460.

Cressie, N. A. C. (1993) *Statistics for Spatial Data*. New York: Wiley.

DasGupta, A. (2008) *Asymptotic Theory of Statistics and Probability*. New York: Springer.

Dedecker, J., Doukhan, P., Lang, G., Leon, J. R., Louhichi, S. and Prieur, C. (eds) (2007) *Weak Dependence: with Examples and Applications*. New York: Springer.

van Duijn, M. A. J., Snijders, T. A. B. and Zijlstra, B. J. H. (2004) P2: a random effects model with covariates for directed graphs. *Statist. Neerland.*, **58**, 234–254.

Fisher, R. A. (1922) On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond.* A, **222**, 309–368.

Frank, O. and Strauss, D. (1986) Markov graphs. *J. Am. Statist. Ass.*, **81**, 832–842.

Georgii, H. (2011) *Gibbs Measures and Phase Transitions*, 2nd edn. Bertin: De Gruyter.

Granger, C. W. J. and Morris, M. J. (1976) Time series modelling and interpretation. *J. R. Statist. Soc.* A, **139**, 246–257.

Handcock, M. (2003) Assessing degeneracy in statistical models of social networks. *Technical Report*. Center for Statistics and the Social Sciences, University of Washington, Seattle. (Available from `http://www.csss.washington.edu/Papers`.)

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M. and Morris, M. (2008) statnet: software tools for the representation, visualization, analysis and simulation of social network data. *J. Statist. Softwr.*, **24**, no. 1.

Handcock, M. S., Raftery, A. E. and Tantrum, J. M. (2007) Model-based clustering for social networks (with discussion). *J. R. Statist. Soc.* A, **170**, 301–354.

Hoff, P. D. (2005) Bilinear mixed-effects models for dyadic data. *J. Am. Statist. Ass.*, **100**, 286–295.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.

Homans, G. C. (1950) *The Human Group*. New York: Harcourt, Brace.

Hunter, D. R., Goodreau, S. M. and Handcock, M. S. (2008) Goodness of fit of social network models. *J. Am. Statist. Ass.*, **103**, 248–258.

Hunter, D. R. and Handcock, M. S. (2006) Inference in curved exponential family models for networks. *J. Computnl Graph. Statist.*, **15**, 565–583.

Ishwaran, H. and James, L. F. (2001) Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Ass.*, **96**, 161–173.

Jackson, M. O. (2008) *Social and Economic Networks*. Princeton: Princeton University Press.

Jonasson, J. (1999) The random triangle model. *J. Appl. Probab.*, **36**, 852–876.

Kolaczyk, E. D. (2009) *Statistical Analysis of Network Data: Methods and Models*. New York: Springer.

Koschade, S. (2006) A social network analysis of Jemaah Islamiyah: the applications to counter-terrorism and intelligence. *Stud. Conflct Terrorsm*, **29**, 559–575.

Koskinen, J. H. (2009) Using latent variables to account for heterogeneity in exponential family random graph models. In *Proc. 6th St Petersburg Wrkshp Simulation*, vol. II (eds S. M. Ermakov, V. B. Melas and A. N. Pepelyshev), pp. 845–849. St Petersburg: St Petersburg State University.

Koskinen, J. H., Robins, G. L. and Pattison, P. E. (2010) Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. *Statist. Methodol.*, **7**, 366–384.

Krivitsky, P. N., Handcock, M. S. and Morris, M. (2011) Adjusting for network size and composition effects in exponential-family random graph models. *Statist. Methodol.*, **8**, 319–339.

Krivitsky, P., Handcock, M. S., Raftery, A. E. and Hoff, P. (2009) Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Socl Netwrks*, **31**, 204–213.

Liang, F. (2010) A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants. *J. Statist. Computn Simuln*, **80**, 1007–1022.

Lovász, L. (2012) *Large Networks and Graph Limits*. Providence: American Mathematical Society.

Lusher, D., Koskinen, J. and Robins, G. (2013) *Exponential Random Graph Models for Social Networks*. Cambridge: Cambridge University Press.

Møller, J., Pettitt, A. N., Reeves, R. and Berthelsen, K. K. (2006) An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, **93**, 451–458.

Møller, J. and Waagepetersen, R. P. (2004) *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall–CRC.

Murray, I., Ghahramani, Z. and MacKay, D. J. (2006) MCMC for doubly-intractable distributions. In *Proc. 22nd A. Conf. Uncertainty in Artificial Intelligence*, pp. 359–366. Corvallis: Association of Uncertainty in Artificial Intelligence Press.

Newman, M. E. J., Watts, D. J. and Strogatz, S. H. (2002) Random graph models of social networks. *Proc. Natn. Acad. Sci. USA*, **99**, 2566–2572.

Nowicki, K. and Snijders, T. A. B. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.

Pattison, P. and Robins, G. (2002) Neighborhood-based models for social networks. In *Sociological Methodology* (ed. R. M. Stolzenberg), vol. 32, ch. 9, pp. 301–337. Boston: Blackwell Publishing.

Resnick, S. I. (1999) *A Probability Path*. Boston: Birkhäuser.

Richardson, S. and Green, P. J. (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. R. Statist. Soc.* B, **59**, 731–792; correction, **60** (1998), 661.

Rinaldo, A., Fienberg, S. E. and Zhou, Y. (2009) On the geometry of discrete exponential families with application to exponential random graph models. *Electron. J. Statist.*, **3**, 446–484.

Sampson, S. F. (1968) A novitiate in a period of change: an experimental and case study of relationships. *PhD Dissertation*. Department of Sociology, Cornell University, Ithaca, New York. Unpublished.

Schweinberger, M. (2011) Instability, sensitivity, and degeneracy of discrete exponential families. *J. Am. Statist. Ass.*, **106**, 1361–1370.

Schweinberger, M. and Snijders, T. A. B. (2003) Settings in social networks: a measurement model. In *Sociological Methodology* (ed. R. M. Stolzenberg), vol. 33, ch. 10, pp. 307–341. Boston: Basil Blackwell.

Shalizi, C. R. and Rinaldo, A. (2013) Consistency under sampling of exponential random graph models. *Ann. Statist.*, **41**, 508–535.

Singla, P. and Domingos, P. (2007) Markov logic in infinite domains. In *Proc. 23rd Conf. Uncertainty in Artificial Intelligence*, pp. 368–375. Corvallis: Association of Uncertainty in Artificial Intelligence Press.

Snijders, T. A. B. (2002) Markov chain Monte Carlo estimation of exponential random graph models. *J. Socl Struct.*, **3**, 1–40.

Snijders, T. A. B. (2007) Discussion on 'Model-based clustering for social networks' (by M. S. Handcock, A. E. Raftery and J. M. Tantrum). *J. R. Statist. Soc.* A, **170**, 322–324.

Snijders, T. A. B., Pattison, P. E., Robins, G. L. and Handcock, M. S. (2006) New specifications for exponential random graph models. *Sociol. Methodol.*, **36**, 99–153.

Stein, M. L. (1999) *Interpolation of Spatial Data*. New York: Springer.

Strauss, D. (1986) On a general class of models for interaction. *SIAM Rev.*, **28**, 513–527.

Strauss, D. and Ikeda, M. (1990) Pseudolikelihood estimation for social networks. *J. Am. Statist. Ass.*, **85**, 204–212.

Tallberg, C. (2005) A Bayesian approach to modeling stochastic blockstructures with covariates. *J. Math. Sociol.*, **29**, 1–23.

Vu, D. Q., Hunter, D. R. and Schweinberger, M. (2013) Model-based clustering of large networks. *Ann. Appl. Statist.*, **7**, 1010–1039.

Wang, J. and Atchade, Y. F. (2014) Bayesian inference of exponential random graph models for large social networks. *Communs Statist. Simuln Computn*, **43**, 359–377.

Wang, Y. J. and Wong, G. Y. (1987) Stochastic blockmodels for directed graphs. *J. Am. Statist. Ass.*, **82**, 8–19.

Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wasserman, S. and Pattison, P. (1996) Logit models and logistic regression for social networks: I, An introduction to Markov graphs and *p**. *Psychometrika*, **61**, 401–425.

Welch, D., Bansal, S. and Hunter, D. R. (2011) Statistical inference to advance network models in epidemiology. *Epidemics*, **3**, 38–45.

White, H. C., Boorman, S. and Breiger, R. L. (1976) Social structure from multiple networks. *Am. J. Sociol.*, **81**, 730–770.

Xiang, R. and Neville, J. (2011) Relational learning with one network: an asymptotic analysis. *J. Mach. Learn. Res.*, **15**, 779–788.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplement: Local dependence in random graph models: characterization, properties, and statistical inference'.