Received: 29 June 2020 | Accepted: 27 February 2022

DOI: 10.1111/rssa.12844

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Royal Statistical Society
Series A
Statistics in Society

ORIGINAL ARTICLE

Bayesian analysis of social influence

Johan Koskinen[1,2] | **Galina Daraganova**[3]

[1]Faculty of Medicine, Dentistry and Health Sciences, Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Victoria, Australia

[2]Institute of Analytical Sociology, University of Linköping, Linköping, Sweden

[3]Business Intelligence, South Eastern Melbourne Primary Health Network, Heatherton, Victoria, Australia

Correspondence
Johan Koskinen, Faculty of Medicine, Dentistry, and Health Sciences, Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, VIC, Australia.
Email: jkoskinen@unimelb.edu.au

Abstract

The network influence model is a model for binary outcome variables that accounts for dependencies between outcomes for units that are relationally tied. The basic influence model was previously extended to afford a suite of new dependence assumptions and because of its relation to traditional Markov random field models it is often referred to as the auto logistic actor-attribute model (ALAAM). We extend on current approaches for fitting ALAAMs by presenting a comprehensive Bayesian inference scheme that supports testing of dependencies across subsets of data and the presence of missing data. We illustrate different aspects of the procedures through three empirical examples: masculinity attitudes in an all-male Australian school class, educational progression in Swedish schools, and unemployment among adults in a community sample in Australia.

KEYWORDS

auto-logistic actor attribute model, contagion, exponential family models, ising model, peer effects, social influence

1 | INTRODUCTION

In social statistics it has become commonplace to take dependencies between outcomes into account using multilevel models (e.g. Goldstein, 1995; Snijders & Bosker, 2011). Thus, when we consider educational outcomes, we may account for compositional or contextual factors with random effects for school classes or neighbourhoods. If we acknowledge the possibility of our

---

observational units being connected with each other through social networks, we can account for some of the dependence this induces using multilevel models (Tranmer et al., 2014) but we cannot capture the detail and diversity of what the social networks literature has termed *social influence* (Robins, 2015). The notion of attitudes and information spreading through friendship networks was already a premise in Moreno's (1934) seminal work explaining a runaway epidemic in a reformatory. Coleman et al.'s (1957) study of the diffusion of the prescription of a novel drug among a network of physicians has been followed by numerous empirical studies of spread on different types of networks (e.g. Strang, 1991; Strang & Tuma, 1993; Valente, 1995). Here, one of our examples aims to investigate if a young, male student's masculinity attitude, coded as high (1) or low (0), depends on whether their friends' attitudes are high or low.

Social influence in social network analysis can broadly be seen as representing processes whereby people tend to be, or become, similar to their friends (or contacts) in their behaviours, attitudes, or beliefs. Social influence is sometimes referred to as *social contagion* (e.g. Burt, 1987; Robins et al., 2012) by analogy to how diseases spread through contact between individuals. Network models are indeed frequently used to model disease spread and epidemics (Jenness et al., 2016; Krivitsky & Morris, 2017; Morris, 2004; Rolls et al., 2012) even though the mechanisms of social contagion may differ. The current canonical empirical framework for investigating social influence is stochastic actor-oriented models, SAOMs (Steglich et al., 2010). While a powerful tool, SAOMs require longitudinal network data and researchers do not always have the resources or opportunity to collect network data at multiple points in time. For cross-sectional network data, even if you have to assert the existence of contagion or influence, controlling for dependencies is a statistical reality (see e.g. Bailey & Hoff, 2015) and neglecting these dependencies may have adverse effects (Doreian et al., 1984; Lubbers & Snijders, 2007). Consequently, we defer to other work for discussions and analysis of identification of peer effects (An, 2011; Bramoullé et al., 2009; Manski, 1993) and focus here on the inferential aspects of a well-defined framework for accounting for network dependence in individual outcomes.

We consider a class of models for investigating social influence for cross-sectional data called auto-logistic actor-attribute models (ALAAMs) (Daraganova & Robins, 2013; Robins et al., 2001) where the outcome of interest is binary. A number of continuous models for social influence exist (Agneessens & Koskinen, 2016; Doreian, 1982; Leenders, 2002; Marsden & Friedkin, 1994; Sewell, 2017; Vitale et al., 2016) that can easily be modified to suit binary outcome variables (Koskinen & Stenberg, 2012; Zhang et al., 2013) but these do not allow specifying the types of dependencies that the ALAAM does.

Gibbs random fields, such as the auto-logistic Ising model (Besag, 1972), have been studied in great detail in statistics and employed in various forms in spatial statistics for modelling binary outcomes with neighbourhood dependencies. To accommodate interpretations in terms of the Behavioural and Social Sciences, Robins et al. (2001) elaborated on these Gibbs-distributions and derived a class of 'social influence' models from a set of specific dependence assumptions. These were later extended by Daraganova (2009) to form a family of actor-attribute auto-logistic models for inferring contagion in cross-sectional data. Exponential random graph models (ERGMs) is a related class of models aimed at modelling the network ties conditional on fixed actor covariates (see Lusher et al., 2013, for an introduction). There are known problems with ERGMs (Handcock, 2003; Schweinberger, 2011) and it is well-known that simple model-specifications do not work, for example, in the sense that the maximum likelihood does not exist or predictive distributions place most of their mass on empty or complete graphs (Schweinberger, 2011; Snijders et al., 2006; see also section 3.1 of Schweinberger et al., 2020). For ALAAMs this is less of an issue but inhomogeneous ALAAMs still present considerable challenges relative to the simpler Ising model.

Maximum likelihood estimation for the elaborated model as described in Daraganova and Robins (2013) is implemented in the statistical software package MPnet (Wang et al., 2014) and is becoming increasingly more popular (some recent studies include acquisition of norms through networks, Kashima et al., 2013; network effects on performance, Letina, 2016; and 'contagion' of depression and PTSD, Bryant et al., 2017; see Parker et al., 2021, for a review). Maximum likelihood estimation follows Snijders' (2002) implementation of the Robbins and Monro (1951) algorithm and the square roots of the diagonal elements of the inverse Fisher information matrix are used as standard errors. These standard errors are motivated by the usual (large n) asymptotics for exponential families, asymptotics that do not apply for ERGMs (Schweinberger et al., 2020). There is no reason to assume that the required asymptotics apply for ALAAMs either which means that a Bayesian inference procedure offers considerable advantages over the maximum likelihood approach, the latter not having well defined measures of uncertainty. Similarly, whereas the asymptotics required for, for example, Akaike's Information are not available (nor are the number of observations defined), Bayesian model selection criteria are well-defined. In addition, the Bayesian approach offers a flexible framework for handling missing data and lends itself to extensions to hierarchical modelling.

Møller et al. (2006) proposed an auxiliary variable MCMC for Bayesian inference for auto logistic models. While this works well for the Ising model, it fails for the inhomogeneous ALAAM of the more elaborate model of Robins et al. (2001), and modified MCMC samplers are required (Koskinen, 2008). To accommodate the challenges presented by realistic ALAAM specifications with multiple covariates, we draw on an adoption of the exchange algorithm (Murray et al., 2006) that has previously been applied by Caimo and Friel (2011) to exponential family random graph models. This is an improvement on the previous Bayesian inference approach for ALAAMs (Koskinen, 2008) and provides a straightforward and flexible inference scheme. We demonstrate how this inference procedure caters to the practical issues often encountered when working with complex empirical network data, such as handling missing data, performing goodness-of-fit, and choosing between competing models. We introduce the model by describing it in some detail. We then proceed to outline various aspect of inference for the model, something that we then illustrate in three empirical examples.

2 | NOTATIONAL PRELIMINARIES

We consider networks represented as graphs or digraphs $G(V, E)$ on a fixed set of nodes $V = \{1, 2, \dots, n\}$, with an arc-set $E \subset V^{(2)} = \{(i,j) \in V \times V : i \neq j\}$, for digraphs, and $E \subset \binom{V}{2}$ for graphs. In social network research, the nodes typically represent n individuals and E the set of connections amongst them (Robins, 2015). We further assume a stochastic binary vertex labelling $\phi : V \to \{0, 1\}^V$, that corresponds to the binary outcome variable of interest for the nodes of the graph. We represent $G(V, E, \phi)$ by its $n \times n$ binary adjacency matrix $X = (X_{ij} : (i, j) \in V^2)$, where the tie-indicators

$$X_{ij} = \begin{cases} 1, & \text{if there is a tie from } i \text{ to } j \text{ in } G \\ 0, & \text{else} \end{cases},$$

and the attribute indicators $Y = (Y_i)_{i=1}^{n}$

$$Y_i = \begin{cases} 1, & \text{if } \phi(i) = 1 \\ 0, & \text{else} \end{cases}.$$

We denote the space of all adjacency matrices by $\mathcal{X} = \{0,1\}^{V^{(2)}}$ and the support of the attribute vector by $\mathcal{Y} = \{0,1\}^V$. We allow for binary and continuous, fixed and exogenous covariates, but suppress the notational dependency on these for the sake of exposition.

In the examples to follow in Section 5, V consists of 108 males in a Year 10 level Australian secondary school; 403 sixth grade students across 19 school classes in Sweden; and 551 adult individuals in Australia. For the first two cases, the network ties X are friendship nominations (both directed) and for the third, nominations of whom you are close to and/or with whom you discuss employment matters (treated as undirected). The outcome variables (Y) are a binary masculine attitudes index, progression to higher education (intention), and employment status respectively.

3 | THE AUTO-LOGISTIC ACTOR ATTRIBUTE MODEL

The general form of the log-linear model used here is

$$p_\theta(y|x) = \Pr(Y = y|X = x, \theta) = \exp\{\theta^\top z(y,x) - \psi(\theta)\}, \tag{1}$$

where $z(\cdot)$ is a $p \times 1$ vector-valued function on G, $\theta \in \mathbb{R}^p$ are the natural parameters, and

$$\psi(\theta) = \log \sum_{y \in \mathcal{Y}} \exp\{\theta^\top z(y,x)\},$$

is a normalising constant. In the next section we proceed to define dependence assumptions from which the statistics $z(\cdot)$ may be derived.

3.1 | Dependence

The simplest form of an ALAAM is a model in which Y_i and Y_j are independent conditional on X and a fixed set of exogenous covariates, for all $i, j \in V$. In this case the ALAAM reduces to a logistic regression model. Auto-logistic models relax the assumption of independence by allowing the state of sites to depend on the states of their *neighbours* in, for example, a lattice like in the Ising model (Besag, 1972). Besag (1974) elaborate auto-logistic models for different types of lattice systems and define dependencies of the first as well as the second order. The neighbourhood for lattice systems is straightforwardly given by the index set of the site variables. In a rectangular lattice, the variable $y_{i,j}$ has neighbours $y_{i-1,j}$,$y_{i+1,j}$,$y_{i,j-1}$, and $y_{i,j+1}$. As social networks rarely are regular graphs, nodes will differ both in the number of neighbours they have as well as their structural position within a graph. This makes for a possibly rich dependence structure but it is not self-evident that just taking the observed network as representing the neighbourhood structure makes for a coherent probability model (especially in the case of directed graphs that are not chordal and that have cycles). Besag (1974) briefly discuss how to define neighbourhoods for non-lattice systems such as points distributed in the plane but this offers little advice for general structures such as networks.

Frank and Strauss (1986) derived a class of exponential family models for the network ties X from dependence assumptions represented by a dependence graph. Robins et al. (2001) similarly specified a dependence graph for the variables Y conditionally on $X = x$. In the dependence graph, the absence of a tie between two variables means that the two variables are conditionally independent. Throughout, we will aim to model the conditional probability structure of Y given $X = x$ and make no statement about the marginal probability $\Pr(X = x)$. To capture the fact that X are assumed to be exogenous and Y endogenous, we define a two-block *chain graph* (Wermuth & Lauritzen, 1990) $\mathcal{D}$, with a block consisting of parent variables X and a block consisting of child variables Y, and where there may be directed ties from nodes in the parent block to nodes in the child block, and possibly undirected ties between variables within the same block. For the purposes of describing the dependence structure, we will use a relabelled index set $T = \{n+1, \ldots, n+t\}$ for the tie-variables whenever there is no ambiguity, where $t = n(n-1)$ or $t = n(n-1)/2$ according to whether the network is directed or not respectively. It is convenient to denote the whole set of variables by $M = (M_i)_{i \in S}$, where $M_i = Y_i$ for $i \in V$ and $M_i = X_i$ for $i \in T$, $S = V \cup T$. Lower case m is taken to be the corresponding realisation of M.

There is a directed tie from a parent node X_i, $i \in T$, to a child node Y_h, if X_j is a parent of Y_h, denoted $X_i \in \text{pa}(Y_h)$. We say that $X_i \in \text{pa}(Y_h)$ if the functional form $\Pr(Y_h = y_h \mid X = x, Y_{-h} = y_{-h})$ depends on x_i and define mutual conditional dependence among child variables Y_i and Y_j as occurring if the functional form of $\Pr(Y_i = y_i \mid Y_{-i} = y_{-i}, X = x)$ depends on y_j and if the functional form of $\Pr(Y_j = y_j \mid Y_{-j} = y_{-j}, X = x)$ depends on y_i.

The moral graph $\mathcal{Q}$ of a chain graph $\mathcal{D}$ is obtained by adding undirected edges between parents of the same child and tuning all directed ties in $\mathcal{D}$ into undirected. Writing $Q(m) = \log \Pr(Y = y \mid X = X) - \log \Pr(Y = 0 \mid X = X)$, following our definition of $\mathcal{D}$ and Besag's (1974) application of the Hammersley-Clifford theorem, $Q(y) = \sum_A \gamma_{A \subseteq S} \prod_{i \in A} M_i$, where γ_A are non-zero if and only if A is a clique in $\mathcal{Q}$ (see section A of the Supplementary Material for further details). This is of the form of Equation (1) with statistics that have elements $z_A = \prod_{i \in A} Y_i \prod_{jk \in A} X_{jk}$.

We proceed to describe three basic classes of dependence assumptions that imply three basic types of models. These models are characterised by their own specific set of statistics.

3.1.1 | Network activity dependence

The simplest form of ALAAM that still accounts for some dependence on the network ties leads to a model where outcomes are conditionally independent, conditionally on the network. Robins et al. (2001) followed the method of Frank and Strauss (1986) in defining a dependence graph based on the elements of the index set V of Y and X. They define the 'Network Activity' [NA] dependence assumption.

Assumption [NA]: An attribute variable Y_i is conditionally dependent on the network tie-variable X_{kh} if and only if $\{i\} \cap \{k,h\} \neq \emptyset$.

The [NA] dependence assumption defines the chain graph (a) in Figure 1 whose moral graph is (b) in Figure 1. Cliques in $\mathcal{Q}$ are the singletons of type Y_i, two-cliques of the type (X_{ij}, Y_i), (X_{ik}, Y_i), and (X_{ih}, Y_i), as well as three-cliques corresponding to labelled network 2-stars (X_{ij}, X_{ik}, Y_i) and four-cliques corresponding to labelled network 3-stars $(X_{ij}, X_{ik}, X_{ih}, Y_i)$, as well as higher order cliques, all corresponding to the association of different stars with the attribute value $y_i = 1$. The model (1) thus has statistics of the form $y_i x_{i,k_1} \cdots x_{i,k_s}$ with parameters $\theta_{i,k_1,\ldots,k_s}$, for $i \in V$

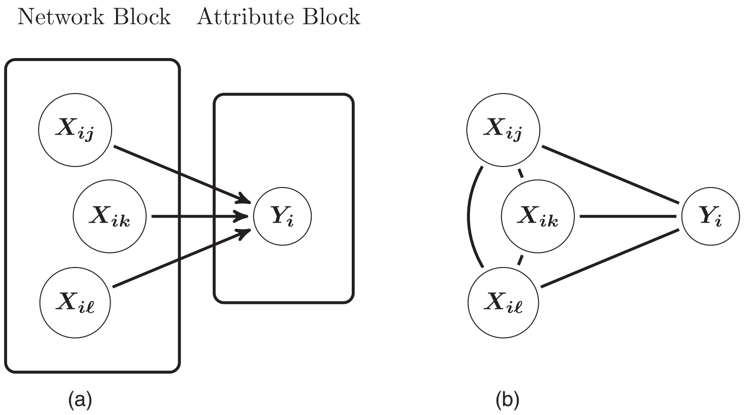


FIGURE 1 Dependence graph (a) and Moral graph (b) of network activity dependence model (Robins et al., 2001), where X_{ij} are tie variables and Y_i binary nodal outcomes, for $i, j \in V$

and s-element subsets $(k_1, \dots, k_s) \in \binom{V_{-i}}{s}$. To reduce the number of parameters, the following homogeneity assumption may be imposed:

$$\theta_{i,k_1,\dots,k_s} = \theta_s$$

for all $i \in V$ and $(k_1, \dots, k_s) \in \binom{V_{-i}}{s}$. Thus, for example the interaction terms $X_{ik}X_{ij}Y_i$ and $X_{jk}X_{ji}Y_j$ both contribute to the same sufficient statistic (see Koskinen et al., 2018, on homogeneity constraints in ERGMs).

While [NA] takes the dependence on network ties into account, the nodal outcomes are conditionally independent for all $i \in V$ conditional on $X = x$. Thus, while the dependence assumption of Frank and Strauss (1986) for the conditional model of X given Y induces dependencies, [NA] does not induce dependence for the elements of Y given X. In the network activity model, the statistic $\sum_i Y_i$ acts as regular intercept term. The *activity* statistic, $\sum_i Y_i X_{i+}$ informs us of the extent to which nodes that have a high degree (are popular) are more or less likely to have a non-zero outcome on Y. The two-stars statistic $\sum_i Y_i \binom{X_{i+}}{2}$ effectively acts as a quadratic degree effect. Higher order star statistics may similarly be interpreted as various forms of accounting for the effect on the outcome of heterogeneity in the degree distribution. For directed networks, the network activity dependence model allows for a richer description of the dependence of the outcome on the network structure. In addition to the baseline effect of degree, the model includes a reciprocity statistic $\sum_i Y_i \sum_{j \neq i} X_{ij}X_{ji}$ which may capture the extent to which nodes that have many reciprocated ties are more or less likely to have the outcome. Among directed stars, we may have out-2-stars $Y_i X_{ij} X_{ik}$; in-2-stars $Y_i X_{ji} X_{ki}$; as well two-paths $Y_i X_{ij} X_{ki}$. The two-stars have relevance from an influence perspective to the extent that they relate to nodes that are in-between other nodes, acting as brokers (Burt, 1987).

3.1.2 | Network contagion dependence

Modelling social influence requires us to define a model that relaxes independence of outcomes for actors. A number of different network social influence mechanisms have been specified (Burt, 1987; Friedkin, 1984; Marsden & Friedkin, 1994). Here we define network contagion in terms of dependence between Y_i and Y_j for pairs $(i,j) \in E$. We cannot simply represent this in the chain graph by adding edges between child variables. For modelling network ties, Pattison and Robins (2002) defined partial conditional dependence in order to extend the local dependence assumptions of Frank and Strauss (1986) by allowing the dependence between variables to be contingent on the states of a third variable. Daraganova (2009) elaborated a number of partial conditional dependence assumptions for the ALAAM that define a number of contagion parameters. We call the first dependence assumption 'Direct Contagion' denoted as [DC].

Assumption [DC]: Any two attribute variables Y_i and Y_j are conditionally dependent if and only if they are connected by a tie $x_{ij} = 1$.

This dependence assumption can be represented by a series of partial dependence structures Q_B for subsets of variables $B \subseteq S$. The node set of Q_B is $S \backslash B$ and $\{i,j\} \in Q_B$ if M_i and M_j are conditionally dependent, given that $M_h = m_h$ for $h \in S \backslash \{i,j\}$ and $M_h = 0$ for $h \in B$. Consequently, for [DC], $\{Y_i.Y_j\}$ is not an edge of Q_B for $B = S \backslash \{Y_i, Y_j\}$, but $\{Y_i, Y_j\}$ may be an edge of Q_B for $B = S \backslash \{Y_i, Y_j, X_{ij}\}$. The partial dependence structures prescribe what interactions among variables S are zero in (1) (see section B of the Supplementary Material for further details):

If $A \subseteq S \backslash B$ and A is not a clique in Q_B for some $B \subset S \backslash A$, then the parameter θ_A corresponding to the statistic $\prod_{i \in A} M_i$ is 0. It follows that the parameter θ_A is non-zero if and only if A is a clique in Q and in all Q_B for which $A \cap B = \emptyset$.

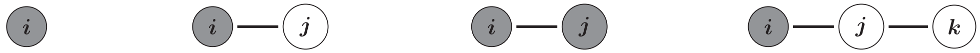
Thus, every possible outcome $x \in \mathcal{X}$ determines a dependence structure described by the moral graph Q and the associated sequence of partial dependence structures Q_B. Non-zero interactions in (1) under assumption [DC] include the non-zero interactions of the network activity model as well as the contagion statistic $\sum_{i<j} Y_i Y_j X_{ij}$. A positive parameter for the contagion statistic means that a node is more likely to have the outcome if it is connected to another node that has the outcome. Note that the cliques corresponding to non-zero interactions are not necessarily hierarchical. For example, while $Y_i Y_j X_{ij}$ may be a non-zero interaction, $Y_i Y_j$ is not for $\{i,j\} \subset B$. Further, a model with the activity star statistics and the contagion term satisfies [DC] but so does a model with additional interactions. The choice of what parameters to set to zero thus involves decisions that, albeit arbitrary, can be motivated from the perspective of parsimony. For the direct contagion model, it would be wise to set all interactions involving more than two outcome variables to zero. Directed ties in the network allow for more elaborate network effects and network contagion effects but care has to be taken to respect the symmetry of the mutual dependencies among child variables (see section E of the Supplementary Material for further details).

3.1.3 | Indirect network and contagion dependencies

Dependencies of outcomes can be further elaborated to incorporate nodes at distances greater than one. We can allow for dependencies of outcome variables on indirect ties. A first assumption, 'Indirect Structural Influence' [ISI], elaborates on the structural effects of [NA].

Assumption [ISI]: An attribute variable Y_i is conditionally dependent on any network tie-variable X_{hk}. $h, k \neq i$, if and only if $x_{ik} = 1$ or $x_{ih} = 1$.

The assumption [ISI] affords statistics that capture how being connected to nodes that themselves have many ties (are popular) may affect the outcome as well as the effect of triadic closure (see Figure 6a in the Supplementary Materials). We may also allow for dependence of the outcome



(a) Attribute (b) Activity (c) Contagion (d) Indirect influence

(e) Indirect contagion (f) Partner activity (g) Partner activity

FIGURE 2 Statistics associated with the [NA] dependence assumption (a and b); the [DC] dependence assumption (c); the indirect structural dependence assumption (d); and the indirect dependent attribute assumption (e, f, and g). Filled nodes indicate $y_i = 1$, and unfilled nodes represent $y_i = 0$ or $y_i = 1$

of Y_i, on outcomes Y_j, of those nodes j to whom node i is only indirectly connected, in order to capture a form of indirect contagion (Brock & Durlauf, 2002; Marsden & Friedkin, 1993) through the 'Indirect Dependent Attribute' [IDA] assumption.

Assumption [IDA]: Any two attribute variables Y_i and Y_k ($i \neq k$) are conditionally dependent if and only if they are directly connected or connected by a path of length two, $x_{ij} = x_{jk} = 1$, for some $j \neq i, k$.

In a regular lattice, [IDA] yields the second-order neighbourhoods of Besag's (1974) auto-logistic model for Plantago Ianceolata but [DC], [ISI], and [IDA] taken together and applied to general network structures yield an extensive set of possible non-zero interactions. Non-zero interactions of two tie-variables and three outcome variables that are derived out of the partial dependence structures (see section D of the Supplementary Material) are depicted in Figure 2. Indirect contagion (Figure 2e) reflects processes where someone for example is likely to believe in something if the friends of their friends believe in it. Partner activity (Figure 2f) reflects processes where someone for example is likely to believe in something if the friends that believe in the same things are also very popular.

4 | INFERENCE

4.1 | Simulating from the model

The expression of Equation (1) cannot be evaluated analytically as the normalising constant $\kappa(\theta)$ is a sum over all of $\mathcal{Y}$. Simulation for Markov random fields is however straightforward and has a long history. Many algorithms have been proposed and they typically draw on the conditional independence that implies that

$$\text{logit}\,\{\Pr_\theta(Y_i = 1 \mid Y_{-i} = y_{-i}, X = x)\} = \theta^\top\{z(\Delta_i^+ y, x) - z(\Delta_i^- y, x)\}$$

where $z(\cdot)$ is the vector of sufficient statistics, $\Delta_i^+ y$ is the vector y with element i set to one, and $\Delta_i^- y$ is the vector y with element i set to zero. For a nearest neighbour algorithm we can update y iteratively by selecting i at random from V and either update it using a Gibbs-update, or through a Metropolis updating step by proposing to change the value from y_i to $1 - y_i$. It is also possible to update multiple variables in parallel using Besag (1974) coding scheme approach for the Ising mode. For example, outcomes for nodes that are isolated in a graph can be updated independently of all other values. Blocks of variables may also be updated independently of each other if they are well separated in the sense of Pattison et al. (2013).

Snijders (2002) discusses a number of sampling schemes for improving mixing for ERGMs and Butts (2018) has shown that perfect sampling is possible for ERGMs and, by implication, ALAAMs. Here we prefer the standard Metropolis algorithm since it is robust and has low memory overhead. Snijders (2002) developed the rule that the burnin for sampling from the ERGM should be a multiple of $n^2 d(1-d)$, with the rationale that each variable is given an opportunity to change but the more ties there are (in the distribution), the longer the burning needs to be. The tie-no-tie algorithm, used for example by Caimo and Friel (2013), addresses the latter problem by having different proposal probabilities for null-ties and ties. Here we set the burnin to γn which, for the same multiplication factor γ, allows for longer burnin than that of Snijders (2002).

4.2 | Estimation

The main obstacle to Bayesian inference for the model of Equation (1), is that the posterior is doubly intractable in the sense that both the normalising constant of the posterior and the likelihood are intractable. With prior distribution $\pi(\theta)$, the posterior distribution is

$$\pi(\theta \mid y, x) = \frac{\exp\{\theta^\top z(y,x) - \psi(\theta)\}\pi(\theta)}{\int \exp\{\theta^\top z(y,x) - \psi(\theta)\}\pi(\theta)\mathrm{d}\theta},$$

where we note that the numerator contains the intractable normalising constant $\psi(\theta)$ and the denominator involves an intractable integral (of an intractable expression). The auxiliary variable MCMC elegantly avoids having to evaluate $\psi(\theta)$ by drawing variables from an auxiliary distribution with the same set of parameters (Møller et al., 2006). The performance of the auxiliary-variable MCMC relies critically on how well the values of the parameters in the auxiliary distribution represent the true but unknown posterior. The linked importance sample auxiliary variable MCMC (Koskinen, 2008) alleviates this issue by introducing bridging distributions, linking the reference distribution to the likelihood. The performance of the linked importance sample auxiliary variable MCMC is however dependent on a good choice of auxiliary variable parameters. The exchange algorithm (Murray et al., 2006) removes the need for the parameters of the auxiliary variable to be fixed and in the process not only reduces computational overheads but also automatically tunes the auxiliary parameters in the course of the MCMC. Caimo and Friel (2011) adopted the exchange algorithm to ERGMs approximating the Gibbs updating step by a Metropolis MCMC. Here we adopt this approximate exchange algorithm to ALAAMs.

For the ALAAM, the exchange MCMC has as its target distribution the joint distribution

$$\pi(\theta, y^*, \theta^* | y, x,) \propto \exp\{\theta^\top z(y,x) - \psi(\theta)\}\pi(\theta)h(\theta^*|\theta)\exp\{\theta^{*\top} z(y^*,x) - \psi(\theta^*)\}$$

where y^* is a variable with the same distributional form as y but with a parameter θ^*, the prior of which is $h(\theta^*|\theta)$. In other words, y^* follows the same ALAAM as data y but with a different parameter. Marginalising this joint posterior with respect to y^* and θ^*, we obtain our desired posterior for θ given y.

A sample $\{\theta_t, y_t^*, \theta_t^*\}_{t=1}^T$, is generated through a two-step updating procedure in each iteration t. Given the current values θ_{t-1}, y_{t-1}^*, and θ_{t-1}^*, we first propose θ^* from $h(\theta^*|\theta_{t-1})$, and conditional on the proposed value, draw y^* from $p_{\theta^*}(\cdot|x)$. Given these proposed values, we propose to swap the parameters with probability $\min\{1, H\}$

$$H = \frac{p_{\theta^*}(y|x)p_{\theta_{t-1}}(y^*|x)}{p_{\theta_{t-1}}(y|x)p_{\theta^*}(y^*|x)} \frac{h(\theta_{t-1}|\theta^*)}{h(\theta^*|\theta_{t-1})} \frac{\pi(\theta^*)}{\pi(\theta_{t-1})},$$

setting $\theta_t = \theta^*$ and $\theta_t^* = \theta_{t-1}$. Similar to the case of (non-curved) ERGMs (Caimo & Friel, 2011), the acceptance probability $\alpha(\theta_{t-1}, \theta^*|y^*)$ simplifies to

$$\alpha(\theta, \theta^*|y^*) = \min\left[1, \exp\{(\theta - \theta^*)^\top(z(y^*, x) - z(y, x))\}\pi(\theta^*)/\pi(\theta)\right]. \tag{2}$$

It is convenient to use a symmetric proposal distribution for $h(\cdot|\cdot)$. In particular, we propose a simplistic multivariate normal with mean vector θ_{t-1} and a variance-covariance matrix that is set to $cp^{-1/2}$ times the inverse of $Cov_{\theta_0}(z(Y,x), z(Y,x))$, approximated from a short initial sample $(y^{(t)})$ from the model defined by the initial value θ_0. For exponential family models in canonical form we have that $I(\theta) = Cov_{\theta_0}(z(Y,x), z(Y,x))^{-1}$ (this procedure was also used by Koskinen et al., 2013, for tuning the algorithm for ERGMs). We set θ_0 to the MLE under a model where contagion parameters are set to zero. As this model is equivalent to logistic regression, the MLE is readily available using standard estimation techniques.

We draw y^* as described in Section 4.1 and hence the algorithm is an *approximate* exchange sampler. Everitt (2012) discusses the implications of the approximation for the properties of the sampler, but ultimately the performance of the algorithm will depend on how well the draws of y^* mixes for different θ, something which will have to be decided on a case by case basis.

4.3 | Missing data

Assume that we observe data y only for a subset of actors given by the missing data indicator $(I_i)_{i=1}^n$, where $I_i = 1$ if the response y_i is unobserved for i and $I_i = 0$ if the response y_i is observed for i. Following Rubin (1976) and Little and Rubin (1987) we define a missing data mechanism $f(I|y, \phi)$ conditional on the response variables where the parameter ϕ is distinct from the model parameters θ. Initialising y by assigning initial values to missing entries, with a prior $\pi(\phi)$, the estimation is carried out as above with two additional updating steps in each iteration. The first consists of updating the missing values and is done by, for each $i \in \{i \in V : I_i = 1\}$, proposing to set $y_i = 1 - y_i$, and accepting this with probability

$$\min\left[1, \exp\{\theta^\top(z(\Delta_i y, x) - z(y, x))\}\frac{f(I|\Delta_i y, \phi)}{f(I|y, \phi)}\right]$$

where $\Delta_i y$ is y with element i toggled and set to $1 - y_i$. To update ϕ, propose a move to ϕ^* drawn from a proposal distribution $q(\cdot|\phi)$, and accept this with probability

$$\min\left\{1, \frac{f(I|y, \phi^*)\pi(\phi^*)}{f(I|y, \phi)\pi(\phi)}\right\}.$$

If data are missing not at random (MNAR) we can define a missing data generating mechanism to test the sensitivity of our inference for θ to deviations from data being missing at random (MAR).

4.4 | Goodness of fit

For ERGMs it has become standard practice to evaluate model fit by considering the predictive distributions for a range of different features of the network (Hunter et al., 2008; Robins & Lusher, 2013). This is partly because the high-dimensional network space admits a number of projections. The outcomes in an influence model have a range-space $\mathcal{Y}$ that is considerably more straightforward to summarise. Given the suite of different statistics that the different dependence assumptions of Daraganova (2009) imply, it is still however necessary to consider a number of functions of $\mathcal{Y} \times \mathcal{X}$ as these may inform us of dependencies in data that we have not captured. Similar to the Bayesian goodness-of-fit (GOF) for ERGMs (Koskinen et al., 2010, 2013), the GOF distribution is the posterior predictive distribution, marginalised over the parameters. The predictive distribution $\{y^{(t)}\}$ is obtained from drawing $y^{(t)}$ from the ALAAM defined by the posterior draw θ_t. In the MCMC that generates the posterior draws, whenever θ is updated we set $\theta = \theta^*$. As the auxiliary variable y^* is drawn from the distribution $p_{\theta^*}(\cdot|x) = p_\theta(\cdot|x)$, the draw of y^* is also a draw from the posterior predictive distribution. Thus, if we let $y^{(t)} = y^*$ for every t such that $\theta_t \neq \theta_{t-1}$, and $y^{(t)} = y^{(t-1)}$, otherwise, we have a draw $\{y^{(t)}\}$ from the posterior predictive distribution at the termination of the estimation algorithm. Note that the Bayesian GOF is based on draws of replicate data from the predictive distribution $\pi(\cdot|y)$ and as such accounts for uncertainty in parameters. Assuming that we have models $M_1, \dots, M_K$ with posterior distributions $\pi(M_k|y)$, we can average the predictive distributions over models.

4.5 | Model selection

Caimo and Friel (2013) propose an across-model procedure to evaluate model evidence for ERGM. They note that within-model estimation of evidence that relies on density estimation of the posterior breaks down for high-dimensional parameter vectors (greater than 5). Friel (2013) proposes an elegant method for estimation of Bayes factors of pair-wise nested models based on the MCMC updating in the exchange algorithm and demonstrates their application to two simple Markov random field models (the Ising model and a Markov two-star ERGM) (Everitt et al., 2017, propose direct estimation of the marginal likelihood using an importance sampling scheme that circumvents the need to evaluate $\psi(\cdot)$ using the trick of Møller et al., 2006). Here we aim to provide a within-model estimation scheme that works for the types of complex models that you would expect when modelling outcomes in the social and behavioural sciences. We follow an adoption of Chib and Jeliazkov (2001) that has previously been used for ERGMs (Koskinen, 2004). First we note from the so-called basic marginal likelihood identity that

$$m(y) = \frac{p_\theta(y|x)\pi(\theta)}{\pi(\theta|y)}$$

where $m(y)$ is the marginal likelihood or equivalently the normalising constant of the posterior distribution of θ given y. This equality holds for any choice of θ and thus we can calculate the marginal likelihood by picking any value θ' and evaluate the basic marginal likelihood for $\theta = \theta'$. We can use the path sampler to evaluate the likelihood ordinate (as in Caimo & Friel, 2013; Hunter & Handcock, 2006; for details see e.g. Gelman & Meng, 1998) but obtaining a good numerical approximation of the posterior ordinate $\pi(\theta'|y)$ is hard.

Proceeding by the method of Chib and Jeliazkov (2001), we define the subkernel of the M-H update as

$$p(\theta, \theta^*, y^*) = \alpha(\theta, \theta^* \,|\, y^*) h(\theta^* \,|\, \theta) p_{\theta^*}(y^* \,|\, x),$$

in which $\alpha(\theta, \theta^* | y^*)$ is defined as (2). By construction

$$\begin{aligned} \pi(\theta \,|\, y) p(\theta, \theta^*, y^*) &= \pi(\theta^* \,|\, y) p(\theta^*, \theta, y^*) \\ \pi(\theta \,|\, y) p_{\theta^*}(y^* \,|\, x) \alpha(\theta, \theta^* \,|\, y^*) h(\theta^* \,|\, \theta) &= \pi(\theta^* \,|\, y) p_\theta(y^* \mid x) \alpha(\theta^*, \theta \,|\, y^*) h(\theta \,|\, \theta^*), \end{aligned} \tag{3}$$

which can be verified by direct calculation. Upon integrating both sides of (3) with respect to θ over Θ and taking the sum over y^* we obtain

$$\int \left[\sum_{y^* \in \mathcal{Y}} p_{\theta^*}(y^* | x) \alpha(\theta, \theta^* | y^*) h(\theta^* | \theta) \right] \pi(\theta | y) \mathrm{d}\theta = \pi(\theta^* | y) \int \left[\sum_{y^* \in \mathcal{Y}} p_\theta(y^* | x) \alpha(\theta^*, \theta | y^*) \right] h(\theta | \theta^*) \mathrm{d}\theta.$$

We recognise the summands as expectations with respect to the likelihood, and the integrands as expectations with respect to the posterior (LHS) and the proposal distribution (RHS). Consequently, solving for $\pi(\theta|y)$, for any parameter value θ', the posterior ordinate can be written as a ratio of expectations

$$\pi(\theta' | y) = \frac{E_{\pi(\theta|y)} \left[E_{p_{\theta'}(y^*|x)} \left\{ \alpha(\theta, \theta' | y^*) h(\theta' | \theta) \right\} \right]}{E_{h(\theta|\theta')} \left[E_{p_\theta(y^*|x)} \left\{ \alpha(\theta', \theta | y^*) \right\} \right]}. \tag{4}$$

We can evaluate the numerator using the Monte-Carlo estimate, taking θ from our posterior draws $\{\theta_t\}_{t=1}^T$ and for the inner expectation we can take a sample of auxiliary variables for each θ'. We may also change the order of the expectations in the numerator, meaning that we draw one large sample $\{y^{(g)}\}_{g=1}^G$ from the distribution defined by $\theta = \theta'$ and average $\alpha(\theta^{(t)}, \theta' | y^{(g)}) h(\theta' | \theta^{(t)})$. For the denominator we draw a number of $\{\theta_j\}$ from the proposal distribution $h(\cdot|\theta')$ and similarly calculate Monte Carlo averages of the conditional acceptance probability $\alpha(\cdot)$ across samples from $\mathcal{Y}$.

With missing data in Y, the likelihood is given by $\exp\{\psi(\theta, \phi, I) - \psi(\theta)\}$ where $\psi(\theta, \phi, I) = \log \left[\sum_{y_i : I_i = 1} \exp\{\theta^\top z(y, x)\} f(I | y, \phi) \right]$. We can evaluate $\psi(\theta, \phi, I)$ using the path sampler with the restriction that $Y_i = y_i$ are fixed for i such that $I_i = 0$. With missing data Equation (4) needs to be modified to account for the uncertainty in the missing outcomes. When evaluating the acceptance probability in the numerator of Equation (4), we may take the Monte Carlo average over the joint posterior of θ and $\{y_i | I_i = 1\}$ for the corresponding draws from the joint posterior. For the denominator of Equation (4), the Monte Carlo average will be taken with respect to draws from $h(\theta|\theta')$ and draws of $\{y_i | I_i = 1\}$ from the conditional distribution $p_{\theta'}(y|x)$. The Monte Carlo estimate of (4) with missing data is written as follows

$$\hat{\pi}(\theta' \,|\, \{y_i : I_i = 0\}) = \frac{\sum_{g=1}^G G^{-1} \sum_{t=1}^T T^{-1} \alpha(\theta_t, \theta' | y^{(g)}, v_t) h(\theta' | \theta_t)}{\sum_{j=1}^J J^{-1} \sum_{m=1}^M M^{-1} \alpha(\theta', \theta_j | y^{(j,m)}, u_j)}, \tag{5}$$

where $\{v_t\}$ are posterior draws of $\{y_i : I_i = 1\}$, $\{y^{(g)}\}$ are draws from $p_{\theta'}(\cdot|x)$, $\{\theta_j\}$ are independent draws from $h(\cdot|\theta')$, $\{u_j\}$ are draws of $\{y_i : I_i = 1\}$ from the model $p_{\theta_j}(\cdot|x)$, and finally $\{y^{(j,m)}\}$ are M

draws from $p(\cdot|\theta_j)$. The numerator in Equation (5) is computationally cheap to evaluate as we only need one large sample $\{y^{(g)}\}$ from the distribution defined by θ'. The denominator in Equation (5) does however require a sample of size M from the model defined by θ_j for all $j = 1, \ldots, J$. The variance of the estimator is not very sensitive to the size of M and G and setting both to about 100 appears sufficient. As discussed in Chib and Jeliazkov (2001), the estimator requires a well mixing sample from the posterior for the numerator of a length T of the order 10,000 to 20,000. Here we have to factor in the variation in the evaluation of the likelihood (as well as the acceptance probability) and a precise estimate is likely to require T in excess of 20,000. (A brief illustration of the effect of different sample sizes T on the estimator is provided in section F of the Supplementary Material.)

4.5.1 | Prior distributions

There are good reasons for performing inference for ALAAMs with prior distributions that are proper. With an improper prior distribution for θ, the posterior distribution is proper if the observed vector of statistics is in the relative interior of the convex hull on $\mathcal{Z}$, where $\mathcal{Z}$ is the image of $\mathcal{Y}$ under $z(\cdot)$. Since there are instances where $z(y, x)$ does not fall in the (relative interior of the) convex hull on $\mathcal{Z}$ (Handcock, 2003), a proper prior distribution formally is a safeguard against the risk of the posterior not being defined. The Bayes factor $\pi(y \mid M_i)/\pi(y \mid M_j)$ is only properly defined if the prior distributions for the parameters of both models are proper. A convenient choice for prior distribution for the canonical parameters for an exponential family model is a multivariate normal distribution $N_p(\mu, \Sigma)$. While one can motivate setting $\mu = 0$ a priori to reflect no bias on the parameters, setting the scale through Σ is less straightforward. For related binomial models, Chen et al. (2008) argue the merits of using Jeffreys' prior (Jeffreys, 1946). Here, this would translate to the prior being $\pi(\theta) \propto |I(\mu_0)|^{1/2}$, for $\mu = 0$ which motivates the scalable normal prior with variance covariance matrix $\lambda I(\mu_0)^{-1}$. The information matrix is straightforward to obtain as the Monte Carlo estimate of the variance covariance of the model sufficient statistics under $\theta = \mu_0$. For some data sets where $\sum_i y_i$ is small, one can motivate using a data-dependent prior with $(\mu_0)_1 = \hat{\theta}_{MLE}$, where $\hat{\theta}_{MLE} = -\log(\bar{y} - 1)$. As λ will shrink the prior distribution and pull parameters towards the origin, setting $(\mu_0)_1 = \hat{\theta}_{MLE}$ will reduce the influence of the intercept which is largely a nuisance.

4.5.2 | Posterior deviance

To evaluate model fit with constant or reference priors, posterior predictive p-values (Meng, 1994) may be applied for any function of the network and attributes that are typically used in GOF (Hunter et al., 2008; Robins & Lusher, 2013). For single value summaries of model fit we may also consider functions of the deviance. In the context of complex network models, Aitkin et al. (2017) considered evidence in terms of the posterior distribution of the deviance (a full discussion of this approach is given in Aitkin, 2010). This provides a useful graphical representation of relative fit of a model that can be summarised using the deviance information criterion (Gelman et al., 2004; Spiegelhalter et al., 2002). As the likelihood of Equation (1) is intractable, we need to evaluate the log-likelihood for each draw θ_t numerically using the path-sampler (Gelman & Meng, 1998; Hunter & Handcock, 2006). With missing data defined as in Section 4.3, the likelihood is estimated as $\hat{\ell}(\theta_t) = \hat{\psi}(\theta_t, \phi, I) - \hat{\psi}(\theta_t)$. Here the likelihood is estimated using the path sampler relative to the MLE for a nested independent model.

5 | APPLICATIONS

We demonstrate the proposed inference procedures using three data sets, two sociocentric school networks and one snowball-sampled data set on unemployment status.

5.1 | Masculine attitudes in a school class

Lusher and Dudgeon (2007) developed a scale, MAI, for measuring male dominance attitudes. In school classes it may be of interest to know if a (male) pupil's attitudes to masculinity is contingent on those of his friends. MAI scores as well as friendship nominations were collected for 106 pupils in a Year 10 level in a single-sex, religious secondary school in Australia (Lusher, 2011). The response variable y_i is the MAI dichotomised at the mean. Controls are: 'dominant culture' (indicates if i has an Anglo-Australian ethno-cultural background (1) or not (0)); the socio-economic status of the pupil's household (as measured by standardised SES based on postcode); the occupational score for the father of the pupil ; the equivalently defined occupational score for the mother of the pupil (see Lusher, 2011, for further details of the network data). The network is depicted in Figure 3.

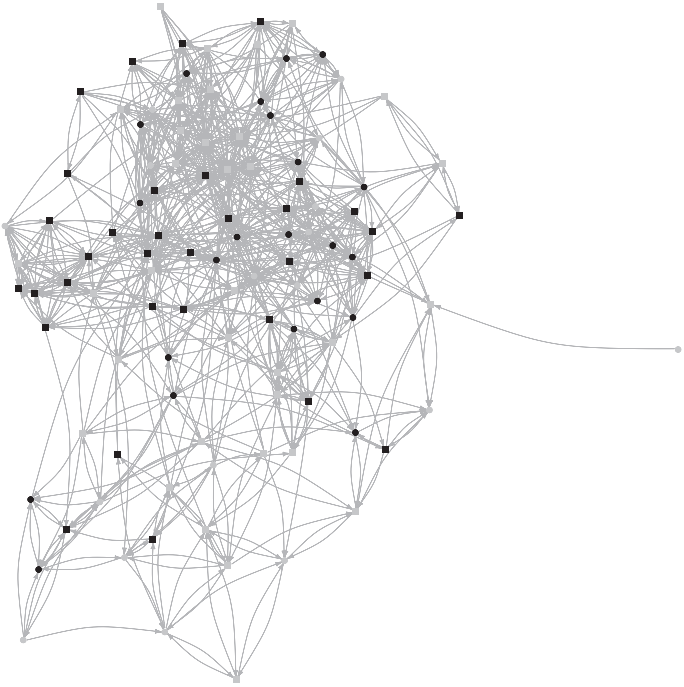


FIGURE 3 Friendship network among 106 pupils in an all-male school. Dominant culture indicated by squares (1) and circles (0), and outcome black ($y_i = 1$), grey ($y_i = 0$), for high and low masculinity index, respectively

5.1.1 | Direct contagion

Figure 4 provides the MCMC output for a model under assumption [DC] with 20,000 draws using the standard settings of Section 4.2, namely with the proposal variance-covariance matrix based on $cI(\tilde{\theta})$, where $I(\theta)$ is the inverse of the covariance matrix of the statistics under θ; $\tilde{\theta}$ estimated from a simulation of statistics under the MLE for a logistic regression with contagion parameters set to 0; and the tuning constant is set to $c = 1$, and the multiplication factor for drawing from the likelihood is set to $\gamma = 30$. These settings will be used also for the rest of the examples unless otherwise specified. The auto-correlation for the contagion parameter is fairly large even at large lags. This can be improved upon by setting the proposal covariance matrix equal to the covariance of the posteriors (this reduces the sample autocorrelation function, SACF, greatly). According to the posterior summaries provided in Table 1, there is evidence for a positive contagion parameter.

5.1.2 | Indirect contagion

To infer whether there is evidence of influence on MAI being transmitted through indirect ties under assumption [IDA], we add the statistic $y_i \sum_j y_j \sum_{k,k\neq i,j} x_{ik}x_{kj}$. In addition we include a statistic for the number of nodes that are reachable from an individual, the number of indirect ties $\sum_i y_i \sum_{j,k} x_{ik}x_{kj}$. The potential for a brokerage effect on Y is controlled for by the mixed 2-path effect $\sum_i y_i \sum_{j,k} x_{ij}x_{ik}$. The results for the elaborated model are provided in Table 2. The introduction of the additional contagion effect reduces the direct contagion (posterior correlation of −0.69), making interpretation less conclusive than in the simpler model. The number of indicted ties is positive with a large posterior probability suggesting that pupils that are indirectly connected to many others are likely to have masculine attitudes.

5.1.3 | GOF

The posterior predictive distributions for some functions of (Y, X) are provided in Table 3. For reference, predictive distributions for a latent network effects model (LNAM) are provided (with estimates in Table 1). For this model we assume that there exists an $n \times 1$ vector u that follows the standard network effects model $u = \alpha Wu + \beta^{\top}B + \epsilon$, where W is the row-normalised adjacency matrix, $\alpha \in (-1, 1)$ is the network effects parameter (Marsden & Friedkin, 1994), B is a matrix with the same fixed covariates as for the ALAAM in Table 1, and ϵ are i.i.d. standard normal variates. As Y is binary, we use u as the latent variable for a probit link-function by letting $y_i = \mathbb{1}\{u_i \geq 0\}$. Estimation of α and β largely follows Koskinen and Stenberg (2012). We assume the same form of prior, as described in Section 4.5, for the β in the network effects model and θ in the complex contagion model with the exception that the α in the former is not included in the regression parameters.

For this relatively limited set of attribute and network interactions, the ALAAM marginally outperforms the LNAM judging by the posterior predictive p-values of Table 3. However, for this dataset there is no clear evidence of the LNAM completely failing to reproduce any of the observed statistics. The goodness-of-fit does also illustrate that the simpler specification of the ALAAM is sufficient for explaining higher-order dependencies such as indirect contagion.

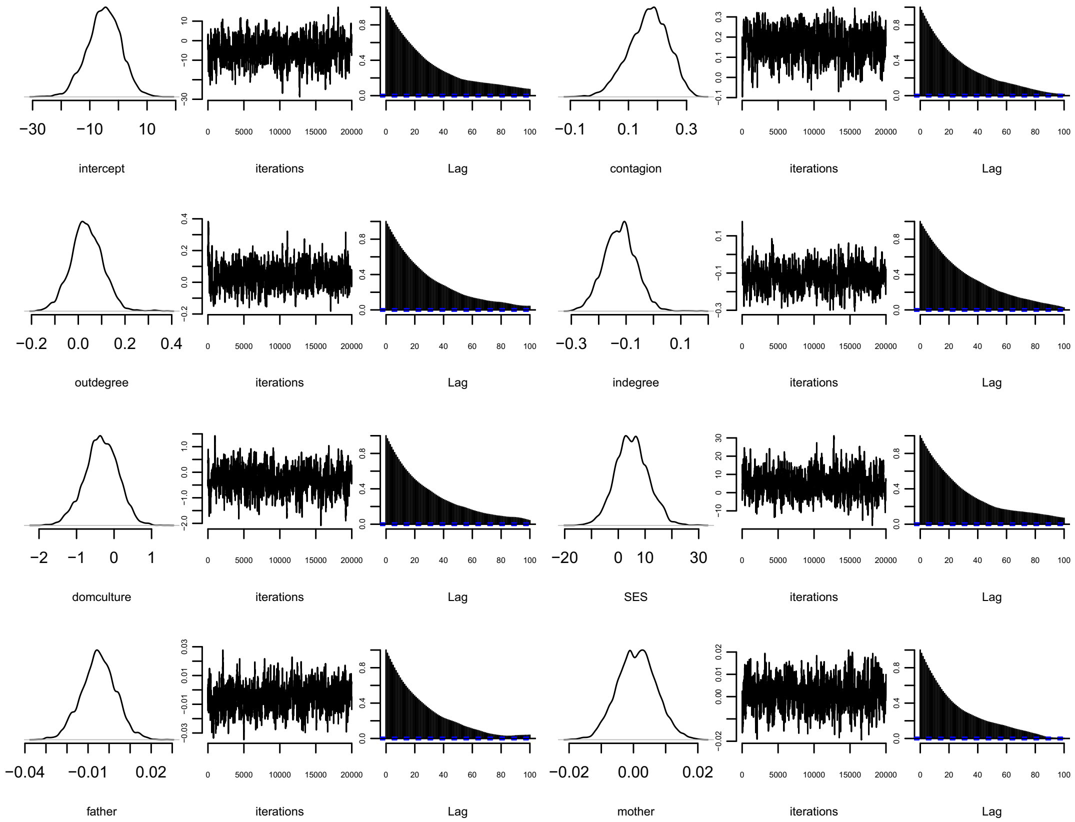


FIGURE 4 MCMC output for ALAAM contagion-model for masculine attitudes in an Australian school. Density estimate, trace plot, and SACF for each of the parameters reported in Table 1

TABLE 1 Posterior summaries for an ALAAM and an LNAM of contagion of masculine attitudes in a school in Australia

	ALAAM							LNAM	
	Mean	**SD**	**ESS**	**SACF 10**	**SACF 30**	**2.5 perc**	**97.5 perc**	**Mean**	**SD**
Intercept	−4.94	5.78	269.73	0.72	0.37	−16.23	6.33	−7.52	3.91
Contagion	0.17	0.07	251.07	0.70	0.35	0.03	0.29		
Outdegree	0.03	0.07	292.16	0.72	0.37	−0.10	0.16	−0.04	0.03
Indegree	−0.12	0.06	293.03	0.74	0.42	−0.23	−0.00	0.13	0.04
Domculture	−0.32	0.43	235.78	0.70	0.38	−1.28	0.47	−0.61	0.31
SES	5.04	6.14	271.11	0.73	0.38	−7.90	17.11	7.85	4.11
Father	−0.00	0.01	281.46	0.68	0.34	−0.02	0.01	−0.01	0.01
Mother	0.00	0.01	285.92	0.68	0.31	−0.01	0.01	0.01	0.00
Alpha								0.49	0.18

TABLE 2 Posterior summaries for an ALAAM with indirect contagion of masculine attitudes in a school in Australia

	Mean	SD	ESS	SACF 10	SACF 30	2.5 perc	97.5 perc
Intercept	−6.83	6.53	314.11	0.78	0.44	−20.14	5.31
Contagion	0.21	0.13	281.64	0.79	0.51	−0.05	0.48
Indirect cont	−0.02	0.02	275.48	0.79	0.52	−0.06	0.01
Outdegree	−0.48	0.27	242.09	0.80	0.54	−1.02	0.02
Indegree	0.07	0.16	324.60	0.79	0.50	−0.27	0.38
Brokerage	−0.01	0.02	323.54	0.79	0.51	−0.05	0.02
Indirect ties	0.07	0.03	211.01	0.80	0.53	0.02	0.13
Domculture	−0.41	0.49	292.16	0.79	0.51	−1.30	0.59
SES	7.28	6.64	305.18	0.77	0.43	−6.32	20.32
Father	−0.01	0.01	1256.67	0.78	0.46	−0.03	0.01
Mother	−0.00	0.01	305.47	0.79	0.49	−0.02	0.01

TABLE 3 Posterior predictive *p*-values for ALAAM and latent network autocorrelation models of Table 1

		ALAAM		LNAM	
Statistic	**Observed**	**Mean**	***p*-Value**	**Mean**	***p*-Value**
Intercept	55	55.87	0.21	53.02	0.20
Direct contagion	272	277.94	0.25	250.54	0.17
Reciprochal contagion	75	78.17	0.24	74.39	0.22
Indirect contagion	2069	2132.74	0.26	1880.53	0.17
Closure contagion	763	790.04	0.26	695.57	0.17
Transitive contagion	456	534.64	0.24	456.68	0.28
Indegree	428	432.53	0.24	393.71	0.14
Outdegree	478	488.87	0.21	481.77	0.23
Two-paths	3859	3972.61	0.22	3786.91	0.22
Out-2-star	2362	2404.59	0.22	2452.69	0.17
In-2-star	1780	2082.39	0.19	1701.24	0.18
Out-triangles	1368	1404.57	0.21	1385.58	0.22
In-triangles	1210	1191.38	0.23	1048.03	0.11
Transitive triangles	1004	1033.12	0.23	951.75	0.19
Indirecct ties	3957	3914.89	0.25	3895.86	0.24

5.2 | Stockholm Birth Cohort

The Stockholm Birth Cohort is a large cohort study in the Stockholm Metropolitan area that includes detailed surveys and school-class network data (Stenberg, 2018; Stenberg & Vågerö, 2006). The networks are the best-friend nominations of school children and for each pupil there are a range of sociological, psychological, and educational variables. The survey was carried out in May 1966 when the pupils were nearing the end of the sixth grade. This is when they would have started considering whether they were going to proceed to higher secondary education (grades 10 and above) and been talking about this with their peers. We chose for our example 19 school classes out of the 1966 survey.

We let X be the directed best-friend network (this had a cap of three nominations), and y be indicators $y_i = 1$ of whether pupil i said that they intended to proceed to higher secondary school, and $y_i = 0$ otherwise (in accordance with the model of Koskinen & Stenberg, 2012). By design there are no ties between pupils in different school classes. The proportion of missing entries range from 0 to 0.286 with an average of 0.079. We apply the ALAAM specified by assumption [DC] but set the parameter for out-stars (of the form $y_i x_{i,+}$) to zero as the nominations were capped at three and there is little variance in the out-degree distribution. In addition to this structural part, we control for: sex (female:1); family support (an 11-point scale measuring the family's attitude toward school ranging from least positive, 0 to most positive 10); average school marks (scaled to range from 0 to 10); an indicator of whether the father belongs to the top social class or not. The network is visualised in Figure 5.

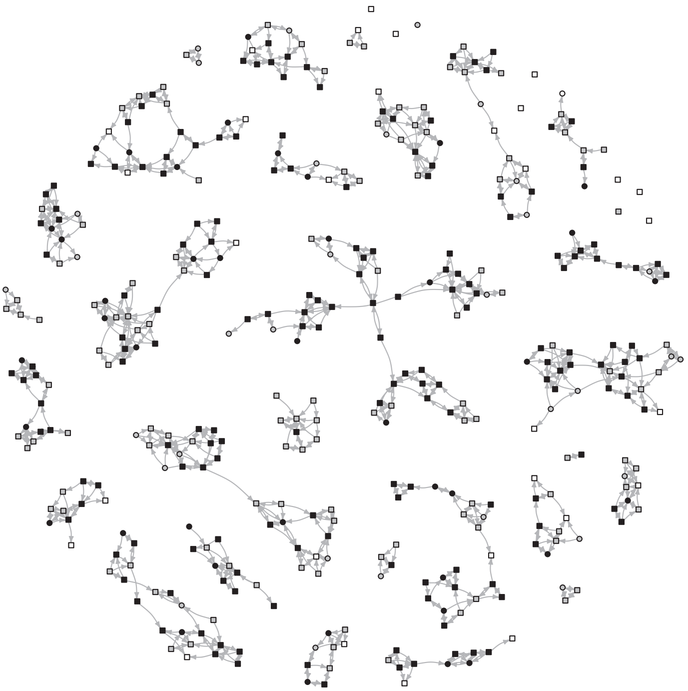


FIGURE 5 Best friend network in four schools in the Stockholm Birth Cohort. Sex indicated by squares (girl) and circles (boys), and outcome – intention to proceed to higher secondary education – black ($y_i = 1$), grey ($y_i = 0$), and white for missing

TABLE 4 Model 1 posterior summaries for contagion-model for progression to upper-secondary school in SBC (posterior means, SD, and probability interval based on a thinned sample of 10,000 iterations, taking every 20th iteration, with burnin of 1000; SACF and ESS based on un-thinned sample)

	Mean	SD	ESS	SACF 10	SACF 30	2.5 perc	97.5 perc
Intercept	−9.67	1.11	178.03	0.68	0.32	−11.83	−7.51
Contagion	0.16	0.10	183.10	0.68	0.32	−0.04	0.35
Indegree	−0.07	0.11	183.55	0.67	0.32	−0.29	0.13
Sex	−0.09	0.29	134.35	0.70	0.39	−0.66	0.47
Family attitude	0.48	0.09	164.22	0.70	0.32	0.33	0.65
Marks	0.99	0.15	168.66	0.68	0.32	0.69	1.28
Social class 1	0.59	0.32	198.40	0.66	0.24	−0.06	1.19

The results from the MCMC with 10,000 iterations with constant priors $\pi(\theta) \propto c$ are summarised in Table 4 (the table is based on default settings with a burnin of 1000 and thinning of 20 iterations, and the same proposal as in the previous example; $\gamma = 7.5$). Mixing of the MCMC can be said to be satisfactory with default settings. There is strong evidence for a positive family attitude to school and high grades to increase the likelihood of the intention to proceed to higher education. The evidence is inconclusive for other effects. In particular, the contagion parameter is positive with posterior probability 0.93.

5.2.1 | Testing difference in contagion

The classes come from 4 schools that differ in socio economic status of uptake area as reflected in the composition of social class of pupils. We divide the schools into one subset with less than 15% of students (across school classes) from the highest social class and a subset with more than 15% of students from the highest social class. Table 5 presents the results for a model (M_2) that includes an interaction of the contagion parameter and an indicator for the type of school (=1 for schools with low proportion of pupils from the highest social class) as well as the main effect. There is stronger evidence than for model 1 for a contagion effect (the contagion parameter is positive with 0.988 posterior probability). There is weak evidence for contagion being absent in schools with a lower proportion of pupils from the highest social class (the posterior distribution for $\theta_2 + \theta_9$ has a mean of 0.0328 and a standard deviation of 0.144 and is negative with 0.435 posterior probability).

Consider first evaluating the evidence for M_1 against M_2 based on the results in Tables 4 and 5 that are based on improper priors. We estimate the likelihood as in Section 4.5.2, relative to the MLE for a model with the contagion parameter, composition, and contagion interaction set to zero. We estimate $\hat{\ell}(\theta_t)$ for a thinned sample of 226 posterior draws, using 20 bridges and 100 samples for each. In fact, using half of these posterior draws and only 5 sampled networks for each give virtually identical results. Figure 6 (left panel) shows that the deviance distributions are stochastically ordered (Aitkin et al., 2017) and that model 2 is the preferred model. Based on the posterior deviances of Figure 6 (left panel), we provide two versions of the DIC measure (Gelman et al., 2004; Spiegelhalter et al., 2002) in Table 6, both of which suggest that M_2 is preferred over M_1.

TABLE 5 Model 2 posterior summaries for contagion-model for progression to upper-secondary school in SBC with an interaction between social contagion and social class (Posterior Means, SD, and probability interval based on thinned sample of 10,000 iterations, taking every 20th iteration, with burnin of 1000; SACF and ESS based on un-thinned sample)

	Mean	SD	ESS	SACF 10	SACF 30	2.5 perc	97.5 perc
Intercept	−10.13	1.19	168.32	0.76	0.44	−12.81	−8.04
Contagion	0.24	0.12	143.31	0.72	0.39	0.02	0.48
Indegree	−0.08	0.12	122.80	0.75	0.41	−0.33	0.13
Sex	−0.09	0.28	126.04	0.76	0.45	−0.69	0.47
Family attitude	0.48	0.08	140.26	0.72	0.38	0.34	0.65
Marks	1.01	0.14	265.08	0.72	0.40	0.76	1.31
Composition	0.91	0.55	137.33	0.74	0.39	−0.25	1.97
Social class 1	0.57	0.34	143.59	0.73	0.37	−0.07	1.21
Contagion × social class 1	−0.21	0.16	152.15	0.72	0.37	−0.51	0.11

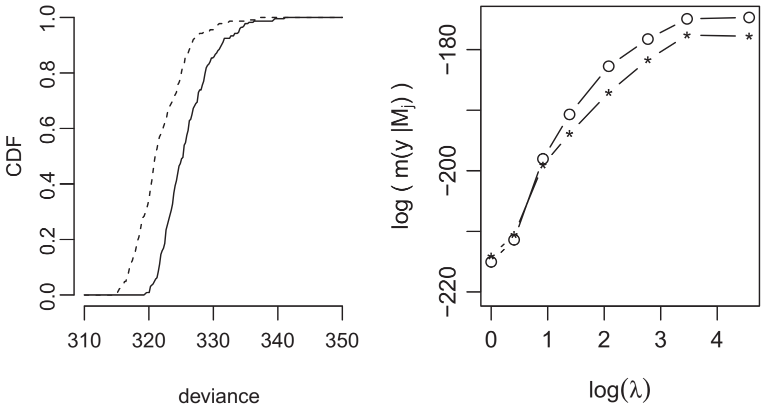


FIGURE 6 Left panel: The posterior deviance for model 1 (Table 4) (solid) and model 2 (Table 5) (dashed). Right panel, model evidence for model 1 (o−) and model 2 ($*$−) for prior $N_p(\mu_0, \lambda I(\mu_0)^{-1})$ with scale λ and $\mu_0 = 0$

TABLE 6 Deviance information criteria (DIC) evaluated for for Models 1 and 2 fitted to the SBC

	$\overline{D} + p_D$	$\overline{D} + p_V$
Model 1	332.94	333.68
Model 2	330.32	329.73

Examining the evidence for the two models in Figure 6 (right panel), the interaction model, Model 2, is preferred for λ between 1 and 4. As λ gets larger, the prior variance increases, penalising model complexity and thus favouring the more parsimonious model (c.p. Bartlett, 1957). The figure is meant to illustrate the dependence on λ and the precision of the estimates of the evidence is not sufficient to draw firm conclusions (estimates in the range $\lambda \in (1, 3)$ are the average of three estimates with $T = J = 2152$, the rest are based on $T = 19{,}000$; a

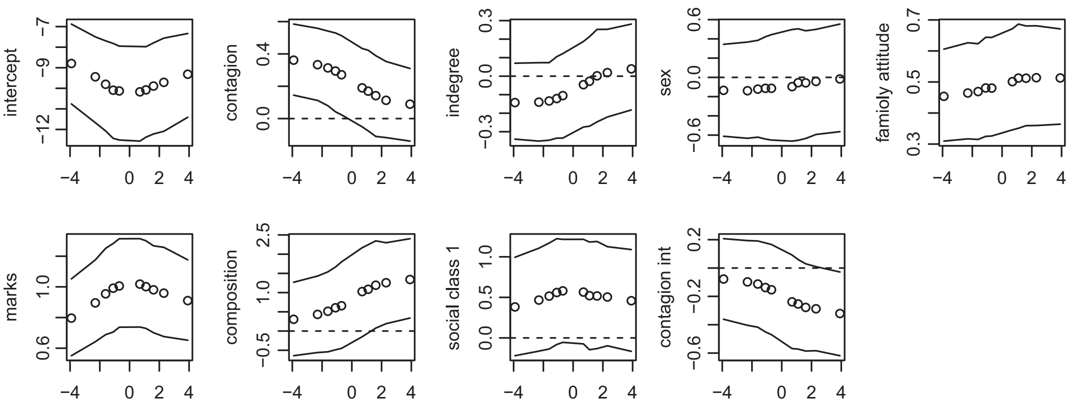


FIGURE 7 Posterior summaries (95% CI) for parameters of model 2 (Table 5) for SBC against missing not at random parameter ϕ_1

brief illustration of the effect of the sample size is provided in section F of the Supplementary Material).

5.2.2 | Sensitivity to MAR assumption

To test the sensitivity of the posteriors to violations of the MAR assumption, we posit the MNAR missing data mechanism assuming the logistic form

$$\text{logit}\,\{\Pr(I_i = 1|y, x)\} = \phi_0 + \phi_1 y_i + \phi_2 x_{+i}$$

independently for all $i \in V$ conditional on y. With $\phi_1 < 0$ the interpretation would be that pupils that do not intend to proceed to higher secondary education are less likely to respond. Assuming that receiving few best-friend nominations is associated with social isolation, a negative ϕ_2 would mean that socially relatively isolated pupils are more likely to be missing. Fixing ϕ, only ϕ_1 will affect inference as the covariate dependent ϕ_2 and the intercept ϕ_0 cancel out in simulating Y_i for missing cases. Figure 7 plots the change in credibility intervals for some of the parameters of model 2. If missingness is strongly predicted by an intention to proceed to higher secondary education (ϕ_1 positive), the contagion effect is weakened. If pupils not intending to proceed are more likely to be missing, the contagion effect is strengthened. The strength of the MNAR mechanism also affects the composition parameter and the interaction with composition and contagion. The bias, as represented by ϕ_1, does however need to be strong to have an effect (at $\phi_1 = 4$, $y_i = 1$ almost with probability 1 for missing cases).

5.3 | Unemployment in a large network

When the node-set of a network is not unambiguously defined or the population size is too big to allow for a complete census of the network, we may still want to estimate network-related effects from a sample of the population network. We consider a dataset analysed previously

TABLE 7 Posterior summaries for a model for employment status for a sample from Victoria, Australia, estimated conditional on outcomes in waves 3 and greater

	Mean	SD	2.5 perc	97.5 perc
Intercept	−1.582	0.246	−2.119	−1.13
Contagion	0.324	0.131	0.03	0.548
Degree	−0.106	0.048	−0.204	−0.005
Sex	0.541	0.264	0.01	1.099
Age	0.001	0.128	−0.264	0.242

by Daraganova and Pattison (2013) that consists of 551 individuals recruited via a three-wave snowball sample (Frank, 2005; Goodman, 1961) in Australia. Drawing on Besag's (1974) coding scheme, Pattison et al. (2013) demonstrated how the dependence assumptions of an ERGM can be used to define a conditional inference scheme. For ALAAMs this translates to estimating the model as described above with the condition that y_i remains fixed at their observed values for $i \in A \subset V$, where A is a set that *separates* (Pattison et al., 2013) data in $\mathcal{Y} \times \mathcal{X}$. For the three-wave snowball sample this means conditioning on the outcomes of nodes in wave 3 (184 nodes), and conditionally on these nodes, modelling only the outcomes of the seed nodes, and outcomes of nodes in waves 1 and 2 (367 nodes). The outcome variable of interest is employment status, with 'employed individuals' being defined as those individuals who worked full or part time, and students who worked part time ($Y = 0$); and 'unemployed individuals' was defined as those individuals who did not work at the time of the interview ($Y = 1$). In addition we use a reduced set of other variables, namely the number of network partners (degree); sex (male: 0; female: 1); and age (ranging from 19 to 67 with a mean of 37).

The results of Table 7 largely agree with the analysis of Daraganova and Pattison (2013), and there is clear evidence of a positive association between people that are relationally tied (the posterior mean is 0.322) and a lower risk of being unemployed the more people that you are connected to. Of course, for a sample of a community network we cannot discount the possibility that the network and outcomes are spatially clustered (Butts, 2003; Daraganova et al., 2012) or that there are other geographical network effects (Sohn et al., 2019).

6 | SUMMARY

Building on previous work on ALAAMs (Daraganova, 2009; Daraganova & Robins, 2013; Robins et al., 2001) we draw on advances in modelling Markov random fields (Caimo & Friel, 2011; Friel, 2013) to improve on previous Bayesian estimation schemes (Koskinen, 2008) for the social influence model.[1]

We illustrated various aspects of fitting the model using three example datasets. We found that pupils that have friends that have male-dominance attitudes also tend to have male-dominance attitudes themselves. Posterior predictive *p*-values show that a simple model with direct contagion is sufficient for explaining more complex interactions and in addition shows that the ALAAM compares favourably with an alternative network dependence model. For a Swedish dataset we found that pupils that have friends that intend to proceed to higher education are more likely to

---

[1] Code is available at https://github.com/johankoskinen/ALAAM

have the same intention themselves. We also found tentative evidence for this ‘contagion’ effect to be present in schools with pupils of higher social class than in schools with a lower proportion of pupils from a high social class. The estimates for the contagion effect was demonstrated to be robust to violations of the MAR assumption. Finally, a dataset collected using snowball sampling in Australia showed that people that have unemployed friends are more likely to be unemployed themselves.

A benefit of the Bayesian estimation approach for ALAAMs is that the coherent treatment of uncertainty allows greater flexibility in handling missing data and performing model evaluation relative to the maximum likelihood approach. This likelihood-based framework is also readily extended to hierarchical modelling so that we for example can analyse social influence jointly for multiple datasets (c.p. the continuous case, Agneessens & Koskinen, 2016).

Cross-sectional network data does not allow us to distinguishing social influence and social contagion from *social selection* (Steglich et al., 2010), but when only cross-sectional network data are available it is still necessary to account for peer-dependence through network ties. A Bayesian ALAAM framework allows us to take a number of different types of network dependencies into account.

For ERGMs a number of alternatives to the approximate exchange algorithm have been proposed, such as those in Alquier et al. (2016). Given that the dependencies in ALAAMs are weaker than those in ERGMs, approximate algorithms that do not sample from the model hold some promise for ALAAMs.

ACKNOWLEDGEMENTS

The work by Koskinen was partially supported by NSF-CMMI-2005661 and the Department of Defense under Grant ARO W911NF-21-1-0335. We are grateful for comments on early drafts by a number of people at IC^{2}S^2 and ARS’19, as well as researchers that have patiently tested the inference procedure: Bella Vartanyan, the Science of Networks in Communities (SONIC) research group, and the research group on bushfire data whose work was supported by the grants APP1073041 and the Melbourne climate futures accelerator grant.

ORCID

Johan Koskinen https://orcid.org/0000-0002-6860-325X

REFERENCES

Agneessens, F. & Koskinen J. (2016) Modelling individual outcomes using a multilevel social influence (MSI) model. In: Lazega, E. & Snijders, T. (Eds.) *Multilevel network analysis for the social sciences*. London: Springer. pp 81–105.

Aitkin, M. (2010) *Statistical inference: an integrated Bayesian/likelihood approach*. Boca Raton: Chapman and Hall–CRC.

Aitkin, M., Vu, D. & Francis, B. (2017) Statistical modelling of a terrorist network. *Journal of the Royal Statistical Association (Series A)*, 180, 751–768.

Alquier, P., Friel, N., Everitt, R. & Boland, A. (2016) Noisy Monte Carlo: convergence of Markov chains with approximate transition kernels. *Statistics and Computing*, 26(1–2), 29–47.

An, W. (2011) Models and methods to identify peer effects. *The Sage handbook of social network analysis*. London: Sage, pp. 515–532.

Bailey, K.F & Hoff, P.D. (2015) Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association*, 110(511), 1047–1056.

Bartlett, M. (1957) A comment on D. V. Lindley’s statistical paradox. *Biometrika*, 44(3/4), 533–534.

Besag, J.E. (1972) Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society Series B (Methodological)*, 34(1), 75–83.

Besag, J.E. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society Series B (Methodological)*, 36, 96–127.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009) Identification of peer effects through social networks. *Journal of Econometrics*, 150(1), 41–55.

Brock, W.A. & Durlauf, S.N. (2002) A multinomial-choice model of neighborhood effects. *American Economic Review*, 92(2), 298–303.

Bryant, R.A., Gallagher, H.C., Gibbs, L., Pattison, P., MacDougall, C., Harms, L. et al. (2017) Mental health and social networks after disaster. *American Journal of Psychiatry*, 174(3), 277–285.

Burt, R.S. (1987) Social contagion and innovation: cohesion versus structural equivalence. *American Journal of Sociology*, 92(6), 1287–1335.

Butts, C.T. (2003) Predictability of large-scale spatially embedded networks. In: Butts, C.T., Breiger, R., Carley, K. & Pattison, P. (Eds.) *Dynamic social network modeling and analysis*, Washington, DC: National Academies Press, pp. 313–323.

Butts, C.T. (2018) A perfect sampling method for exponential family random graph models. *The Journal of Mathematical Sociology*, 42(1), 17–36.

Caimo, A. & Friel, N. (2011) Bayesian inference for exponential random graph models. *Social Networks*, 33, 41–55.

Caimo, A. & Friel, N. (2013) Bayesian model selection for exponential random graph models. *Social Networks*, 35(1), 11–24.

Chen, M.-H., Ibrahim, J.G. & Kim, S. (2008) Properties and implementation of Jeffreys's prior in binomial regression models. *Journal of the American Statistical Society*, 103(484), 1659–1664.

Chib, S. & Jeliazkov, I. (2001) Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Society*, 96(453), 270–281.

Coleman, J., Katz, E. & Menzel, H. (1957) The diffusion of an innovation among physicians. *Sociometry*, 20(4), 253–270.

Daraganova, G. (2009) Statistical models for social networks and network-mediated social influence processes: Theory and Applications, University of Melbourne, unpublished PhD thesis.

Daraganova, G. & Pattison, P. (2013) Autologistic actor attribute model analysis of unemployment: dual importance of who you know and where you live. In: Lusher, D., Koskinen, J. & Robins, G. (Eds.) *Exponential random graph models for social networks: theory, methods and applications*. New York: Cambridge University Press, pp. 237–247.

Daraganova, G. & Robins, G. (2013) Autologistic actor attribute model. In: Lusher, D., Koskinen, J. & Robins, G. (Eds.) *Exponential random graph models for social networks: theory, methods and applications*. New York: Cambridge University Press. pp. 102–114.

Daraganova, G., Pattison, P., Koskinen, J., Mitchell, B., Bill, A., Watts, M. et al. (2012) Networks and geography: modelling community network structures as the outcome of both spatial and network processes. *Social Networks*, 34(1), 6–17.

Doreian, P. (1982) Maximum likelihood methods for linear models. *Sociological Methods and Research*, 10, 243–269.

Doreian, P., Teuter, K. & Wang, C. (1984) Network autocorrelation models: some Monte Carlo evidence. *Sociological Methods and Research*, 13, 155–200.

Everitt, R.G. (2012) Bayesian parameter estimation for latent markov random fields and social networks. *Journal of Computational and Graphical Statistics*, 21(4), 940–960.

Everitt, R.G., Johansen, A.M., Rowing, E. & Evdemon-Hogan, M. (2017) Bayesian model comparison with un-normalised likelihoods. *Statistics and Computing*, 27(2), 403–422.

Frank, O. (2005) Network sampling and model fitting. In: Carrington, P.J., Scott, J. & Wasserman, S. (Eds.) *Models and methods in social network analysis*. Vol. 28. New York: Cambridge University Press. pp. 31–56.

Frank, O. & Strauss, D. (1986) Markov graphs. *Journal of the American Statistical Association*, 81, 832–842.

Friedkin, N.E. (1984) Structural cohesion and equivalence explanations of social homogeneity. *Sociological Methods and Research*, 12, 235–61.

Friel, N. (2013) Evidence and Bayes factor estimation for Gibbs random fields. *Journal of Computational and Graphical Statistics*, 22(3), 518–532.

Gelman, A. & Meng, X.L. (1998) Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, 13, 163–185.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian data analysis*, 2nd Edition. Texts in Statistical Science. Boca Raton: CRC Press.

Goldstein, H. (1995) *Multilevel statistical models*. London: Edward Arnold.

Goodman, L.A. (1961) Snowball sampling. *The Annals of Mathematical Statistics*, 32(1), 148–170.

Handcock, M.S. (2003) Assessing degeneracy in statistical models of social networks. Working Paper no. 39, Center for Statistics and the Social Sciences, University of Washington, Available from: http://www.csss.washington.edu/Papers/wp39.pdf

Hunter, D.R. & Handcock, M.S. (2006) Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15, 565–583.

Hunter, D.R., Goodreau, S.M. & Handcock, M.S. (2008) Goodness of fit of social network models. *Journal of the American Statistical Association*, 103, 248–258.

Jeffreys H. (1946) An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London*, 196, 453–46. Ser. A.

Jenness S.M., Goodreau S.M. & Morris, M. (2016) EpiModel: mathematical modeling of infectious disease. R Package Version 1.2.6. Available from: http://epimodel.org/, 10.5281/zenodo.16767

Kashima, Y., Wilson, S., Lusher, D., Pearson, L.J. & Pearson, C. (2013) The acquisition of perceived descriptive norms as social category learning in social networks. *Social Networks*, 35(4), 711–719.

Koskinen, J.H. (2004) Bayesian analysis of exponential random graphs - estimation of parameters and model selection. Research Report 2004, 2, Department of Statistics, Stockholm University.

Koskinen, J.H. (2008) The linked importance sampler auxiliary variable metropolis hastings algorithm for distributions with intractable normalising constants. MelNet Social Networks Laboratory Technical Report 08-01, Department of Psychology, School of Behavioural Science, University of Melbourne, Australia.

Koskinen, J.H. & Stenberg, S.-Å. (2012) Bayesian analysis of multilevel probit models for data with friendship dependencies. *Journal of Educational and Behavioural Statistics*, 37(2), 203–230.

Koskinen, J. H., Robins, G.L. & Pattison, P.E. (2010) Analysing exponential random graph (p-star) models with missing data using bayesian data augmentation. *Statistical Methodology*, 7(3), 366–384.

Koskinen, J.H., Robins, G.L., Wang, P. & Pattison, P.E. (2013) Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks*, 35(4), 514–527.

Koskinen, J.H., Wang, P., Robins, G.L. & Pattison, P.E. (2018) Outliers and influential observations in exponential random graph models. *Psychometrika*, 83(4), 809–830.

Krivitsky, P.N. & Morris, M. (2017) Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *The Annals of Applied Statistics*, 11(1), 427–455.

Leenders, R.T.A.J. (2002) Modelling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24, 21–47.

Letina, S. (2016) Network and actor attribute effects on the performance of researchers in two fields of social science in a small peripheral community. *Journal of Informetrics*, 10(2), 571–595.

Little, R.J.A. & Rubin, D.B. (1987) *Statistical analysis with missing data*. New York: Wiley.

Lubbers, M.J. & Snijders, T.A. (2007) A comparison of various approaches to the exponential random graph model: a reanalysis of 102 student networks in school classes. *Social Networks*, 29, 489–507.

Lusher, D. (2011) Masculinity, educational achievement and social status: a social network analysis. *Gender and Education*, 23(6), 655–675.

Lusher, D. & Dudgeon, P. (2007) The Masculine Attitudes Index (MAI). Working Paper, The University of Melbourne.

Lusher, D., Koskinen, J. & Robins, G. (2013) *Exponential random graph models for social networks: theory, methods and applications*, New York: Cambridge University Press.

Manski, C.F. (1993) Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies*, 60, 531–542.

Marsden, P.V. & Friedkin, N.E. (1993) Network studies of social influence. *Sociological Methods & Research*, 22(1), 127–151.

Marsden, E.V. & Friedkin, N.E. (1994) Network studies of social influence. In: Wasserman, S. & Galaskiewicz, J. (Eds.) *Advances in social network analysis*. Thousand Oaks, CA: Sage. pp. 3–25.

Meng, X.L. (1994) Posterior predictive *p*-values. *The Annals of Statistics*, 22(3), 1142–160.

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

Moreno, J.L. (1934) *Who shall survive?: a new approach to the problem of human interrelations*. Washington, DC, US: Nervous and Mental Disease Publishing Co.

Morris, M. (2004) *Network epidemiology: a handbook for survey design and data collection*. Oxford: Oxford University Press.

Murray, I., Ghahramani, Z. & MacKay, D. (2006) MCMC for doubly-intractable distributions. In: *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*. AUAI Press, Arlington, Virginia.

Parker, A., Pallotti, F. & Lomi, A. (2021) New network models for the analysis of social contagion in organizations: an introduction to autologistic actor attribute models. *Organizational Research Methods* doi:10.1177/10944281211005167.

Pattison, P. & Robins, G. (2002) Neighborhood based models for social networks. *Sociological Methodology* 32, 301–337.

Pattison, P.E., Robins, G.L., Snijders, T.A.B. & Wang, P. (2013) Conditional estimation of exponential random graph models from snowball sampling designs. *Journal of Mathematical Psychology*, 57, 284–296.

Robbins, H. & Monro, S. (1951) A stochastic approximation method. *The annals of Mathematical Statistics*, 22(3), 400–407.

Robins, G. (2015) *Doing social network research: network-based research design for social scientists*. London: Sage.

Robins, G.L. & Lusher, D. (2013) Illustrations: simulation, estimation, and goodness of fit. In: Lusher, L., Koskinen, J. & Robins, G. (Eds.) *Exponential random graph models for social networks: theory, methods, and applications*. Cambridge: Cambridge University Press. pp. 167–185.

Robins, G., Pattison, P. & Elliott, P. (2001) Network models for social influence processes. *Psychometrika*, 66, 161–190.

Robins, G., Lewis, J.M. & Wang, P. (2012) Statistical network analysis for analyzing policy networks. *Policy Studies Journal*, 40(3), 375–401.

Rolls, D.A., Daraganova, G., Sacks-Davis, R., Hellard, M., Jenkinson, R., McBryde, E. et al. (2012) Modelling hepatitis C transmission over a social network of injecting drug users. *Journal of Theoretical Biology*, 297, 73–87.

Rubin, D.B. (1976) Inference and missing data (with discussion). *Biometrika*, 63, 581–592.

Schweinberger, M. (2011) Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496), 1361–1370.

Schweinberger, M., Krivitsky, P.N., Butts, C.T. & Stewart, J.R. (2020) Exponential-family models of random graphs: inference in finite, super and infinite population scenarios. *Statistical Science*, 35(4), 627–662.

Sewell, D.K. (2017) Network autocorrelation models with egocentric data. *Social Networks*, 49, 113–123.

Snijders, T.A.B. (2002) Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1–40.

Snijders, T.A.B. & Bosker, R.J. (2011) *Multilevel analysis: an introduction to basic and advanced multilevel modeling*, 2nd edn. London: Sage.

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

Sohn, C., Christopoulos, D. & Koskinen, J. (2019) Borders moderating distance: a social network analysis of spatial effects on policy interactions. *Geographical Analysis* (forthcoming).

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4), 583–639.

Steglich, C.E.G., Snijders, T.A.B. & Pearson, M. (2010) Dynamic networks and behavior: separating selection from influence. *Sociological Methodology*, 40, 329–393.

Stenberg, S.-Å. (2018) *Born in 1953*. Stockholm: Stockholm University Press.

Stenberg, S.-Å. & Vågerö, D. (2006) Cohort profile: the Stockholm birth cohort of 1953. *International Journal of Epidemiology*, 35, 546–548.

Strang, D. (1991) Adding social structure to diffusion models: an event history framework. *Sociological Methods and Research*, 19, 324–353.

Strang, D. & Tuma, N.B. (1993) Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology*, 99(3), 614–639.

Tranmer, M., Steel, D. & Browne, W.J. (2014) Multiple-membership multiple-classification models for social network and group dependences. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(2), 439–455.

Valente, T.W. (1995) *Network models of the diffusion of innovations*. New York: Hampton Press.

Vitale, M.P., Porzio, G.C. & Doreian, P. (2016) Examining the effect of social influence on student performance through network autocorrelation models. *Journal of Applied Statistics*, 43(1), 115–127.

Wang, P., Robins, G., Pattison, P. & Koskinen, J. (2014) MPNet, Program for the Simulation and Estimation of (p*) Exponential Random Graph Models for Multilevel Networks: USER MANUAL. Melbourne School of Psychological Sciences, The University of Melbourne Australia. Available from: http://sna.unimelb.edu.au/__data/assets/pdf_file/0004/1185745/MPNetManual.pdf

Wermuth, N. & Lauritzen, S.L. (1990) On substantive research hypotheses, conditional independence graphs and graphical chain models. *Journal of the Royal Statistical Society: Series B (methodological)*, 52(1), 21–50.

Zhang, B., Thomas, A.C., Doreian, P., Krackhardt, D. & Krishnan, R. (2013) Contrasting multiple social network autocorrelations for binary outcomes, with applications to technology adoption. *ACM Transactions on Management Information Systems (TMIS)*, 3(4), 18.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Koskinen, J. & Daraganova, G. (2022) Bayesian analysis of social influence. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–27. Available from: https://doi.org/10.1111/rssa.12844