Network and School Variations in Adolescents' Health Behaviour and Educational Attainment: A Multilevel Analysis of US data

George Gerogiannis

Submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy

University of Glasgow

June 2022

Abstract

This thesis develops a statistical methodology for an important area of social network modelling, that of the effects that an individual's social network can have on the individual's propensity to engage in an array of different acts, which has been a public health concern in many societies and is increasingly becoming important in the commercial world as access to such data is becoming increasingly available and can be used to maximise profits. The majority of studies that investigate this phenomenon estimate the fixed effects of network statistics on an individual's propensity to engage in a certain behaviour and are based on network health data. The process thought to generate this phenomenon is typically modelled with a univariate Bernoulli generalised linear model, which simplifies the network component present in the process by summarising it with statistics, a procedure which induces a loss of information. Over the past 20 years, statistical methodology has been developed to remedy this issue with the use of a Bernoulli generalised linear mixed model which explicitly accounts for the network components by modelling them as random effects. The work presented in this thesis provides several novel contributions to these approaches

- The first of which is the development of a multivariate model that extends the multiple membership multiple classification model proposed by Browne et al. (2001).
- The second is the development of a multivariate model that considers a spatio-network interaction involving the sets of spatial and network random effects, as it may be of interest to study whether friendship effects differ depending on the areal unit in which

an individual lives.

- The third concerns the development of a software package that will enable researchers to implement the models developed in this thesis.

These novel contributions are achieved through the use of Bayesian hierarchical models with estimation performed with Markov chain Monte Carlo simulation.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisors Professor Mark Tranmer, Professor Duncan Lee and Professor Laurence Moore for their help, guidance and support over the period of this thesis. I will always appreciate our stimulating discussions during our weekly meetings and their careful and constructive criticism of this thesis. I am grateful to Professor Thomas Valente for providing me with the network data that has motivated the development of models in this thesis. I would also like to thank my family and friends for their unlimited support and guidance. I am indebted to the Lord Kelvin Adam Smith (LKAS) scholarship, for the financial support that it has provided throughout my postgraduate studies, without which this work would not have been possible.

Declaration

With the exception of chapter two, which contain introductory material, all work in this thesis was carried out by the author unless otherwise explicitly stated.

The work presented in chapter five has been accepted for publication in the *Journal of the Royal Statistical Society: Series C* with the title *A Bayesian spatio-network model for multiple adolescent adverse health behaviours*, and is jointly authored with Professor Mark Tranmer, Professor Duncan Lee and Professor Thomas Valente.

The work presented in chapter six has been submitted to the *Journal of Statistical Software* with the title *netcmc: Spatio-Network Generalised Linear Mixed Models for Areal Unit and Network Data*, and is jointly authored with Professor Mark Tranmer and Professor Duncan Lee.

Chapter 1

Introduction

The idea that a relationship exists between an individual's social network and their behaviours dates back over 2000 years to the Classical period in Ancient Greece. Born in Ancient Greece over 470 years before Christ, Socrates would go on to leave his impression on the thinking of mankind and along the way state that, "*everything desires its opposite and not its like*". In contrast to Socrates' line of thought, Plato would write in Phaedrus that, "*similarity begets friendship*", suggesting that an individual's social network is largely made up of people with similar behaviours. During the same period in Ancient Greece, Aristotle, a student of Plato's Academy, mused in Nichomachean Ethics that, "*we love those who are like ourselves*", a notion akin to that of Plato's and widely accepted today. It was Aristotle's belief that friendship was the most important thing that a person could possess and that friends are essential to one's life, writing in Book VIII of Nicomachean Ethics that, "*without friends no one would choose to live, though he had all other goods; even rich men and those in possession of office and of dominating power are thought to need friends most of all*" . The importance of friendship is discussed by many other ancient philosophers. Plautus writes that "*nothing but heaven itself is better than a friend who is really a friend.*", where the use of religious imagery is used to stress the importance of friendship. Similarly, Lucretius uses religious and heavenly imagery to convey the importance of friendship, writing that, "*we are*

each of us angels with only one wing, and we can only fly by embracing one another". Many modern and contemporary philosophers also share the sentiment of ancient philosophers on the topic of friendship. Similar to the way in which Aristotle suggests that friendship is essential to one's life, Francis Bacon implies that friendship is essential in a civil world, writing that, "*without friends the world is but a wilderness*". In contrast to the ways in which Plautus and Lucretius use religious imagery and poetic language to stress the importance of friendship, William James takes a more grounded approach to convey how important he thinks that friendships are, writing that, "*human beings are born into this little span of life of which the best thing is its friendship and intimacies*".

Ancient and modern philosophers both highlight the importance of friendship to one's self. However, there is also a massive interest in understanding how a friend may influence one's behaviour, whether it be in a positive, negative or neutral manner. Episodes of friends impacting one's behaviour are plentiful and are themes often explored in works English literature. In Othello, William Shakespeare illustrates how a friend may influence one's behaviour in a negative manner, with the protagonist succumbing to the influence of friend Iago which leads to him murdering Desdemona. However, in the 21st century a growing interest in understanding how a friend may influence one's behaviour in a more quantitative way has occurred. The spreading of behaviour in social networks through edges that link friends to one another can have a profound real-world impact in a wide array of contexts. The spreading of behaviour within a social network has been studied in the context of voting in an election, with Bond et al. (2012) using 61 million Facebook messages during the 2010 US congressional election to find that messages not only influenced the users who received them but also the users' friends, and friends of friends. The spreading of behaviour within social networks has also been observed in the field of health, with a longitudinal study carried out by Christakis and Fowler (2007) finding that a person's observed chances of becoming obese increased if they had a friend who became obese over the period.

The ways in which friends can impact one's health behaviours are of major concern to

policy makers around the world, particularly when the person in question is an adolescent. The uptake of negative health behaviours often occurs during adolescence and can lead to many disorders in adulthood, which represents a major public health problem. In addition to this, the period of adolescence are formative years for many and is a stage in life in which many may be susceptible to influence by friends. Many studies on this topic have found that the susceptibility of peer influence peaks during the stages of early and middle adolescence. These studies commonly present adolescents with questionnaires centred around a series of hypothetical anti-social scenarios involving their peers and construct metrics to measure an adolescents' sensitivity to peer influence (Berndt, 1979; Krosnick and Judd, 1982). However, there are studies that construct questionnaires that are less focused on anti-social behaviours and seek to be more general. In a study by Steinberg and Monahan (2007), the authors construct a metric called resistance to peer influence (RPI), whose values are derived from a series of less anti-social focused items in a questionnaire. In this approach, which studied 3,600 individuals between the ages of 10-30 years old, the authors found there to be a linear increase in the resistance to peer influence between the ages of 14-18 years old. Whereas, between the ages of 10-14 years old, resistance to peer influence tended to be lowest and there was little evidence of its growth throughout this period for the adolescents in the study.

The three adolescent health behaviours typically of concern in the UK are tobacco, alcohol, and drug consumption. Adolescent tobacco consumption, a negative health behaviour, is associated with various ailments, such as mild airway obstruction and slowed growth of lung function (Gold et al. (1996)). In England, there has been a steady decrease since 1996 in the percentage of 11-15 year-old pupils that have reported ever having smoked, from 49% to 16% (NHS (2019)). This, in part, may be attributed to several key intervention programs. In 2002, the Tobacco Advertising and Promotion Act 2002 placed controls on the advertising and promotion of tobacco products. In July 2007, as a consequence of the Health Act 2006, a ban on smoking in all public areas came into force in England. In October 2007, the Children and Young Persons (Sale of Tobacco) Order 2007 came into force, raising the

minimum purchase age from 16 to 18 years of age.

In comparison, different levels of adolescent alcohol consumption have been associated with poorer cognitive functioning (Lees et al. (2020)), poor school performance (Balsa et al. (2011)), antisocial behaviour (Hammerton et al. (2017)) and violence (Komro et al. (2000)). Similarly to the trend in cigarette consumption, between 2003 and 2014 there has been a decline in the percentage of 11-15 year-old pupils reporting ever having tried alcohol from 61% to 38% (Oldham et al. (2018)). There are a number of key intervention programs that have been put into place across the UK. In 2009, the Chief Medical Officer of England published the first official guidance on alcohol aimed specifically at children and young people (Donaldson (2009)). In May 2018, as a result of The Alcohol (Minimum Price per Unit) (Scotland) Order 2018, Scotland introduced a minimum unit price of 50p.

In this thesis, I build on the models used to estimate the effect that alters, which are defined as individuals who are socially connected with a given individual, in a social network have on the behaviour of the given individuals and compare these novel approaches to those adopted by other researchers. These modelling advancements allow for more insightful inference to be carried out when compared to current methods and highlight some deficiencies in the way in which network data is currently modelled to study its effect on an individual taking up specific behaviours. The research presented in this thesis is centered around several aims, which incorporate covariate, spatial and networks structures in multivariate regression models. The first aim focuses on extending the univariate generalized linear mixed model to a multivariate one, focusing on modelling network effects jointly across more than one health behaviour. The majority of authors who are studying the effect of peers on an individual decision to engage in more than one health behaviour construct as many univariate models as there are health behaviours in order to estimate these effects. The second aim is to modify the way in which the effect of peers in a network is modelled, focusing on approaches that do not use network statistics. On this point, the majority of authors currently use statistics of a friendship network as proxies for concepts of peer influence. The third aim is to

construct a modelling approach which allows for covariate, spatial and network structures to be accounted for in a multivariate model, allowing for inference regarding the effects of an individual's personal covariates, place of residence and friendship network to be made. In contrast, the majority of authors incorporate the friendship network into the personal covariates of an individual and omit the facility for spatial structures within the data set to be included in the model. The fourth aim is to create an `R` package which is free to use and will allow for the adoption of models discussed in this thesis. In addition to this, the software package will also include other models that are not exemplified in this thesis but accommodate slightly different data structures that a researcher may encounter.

The remainder of this thesis is structured into eight chapters. Chapter 2 reviews the statistical methods used in this thesis, focusing on Bayesian methods, generalized linear models, network models and spatial models. The chapter begins by reviewing the Markov chain Monte Carlo methods used for inference in Bayesian inference, as well as model selection and checking techniques used to assess the posterior distribution of parameters in the model. The remainder of the chapter focuses on reviewing generalized linear models, spatial modelling and multiple membership multiple classification models that form the basis for modelling social network structures.

Chapter 3 reviews and critiques the literature on health behaviours and network studies, starting with a focus on studies that are centered around analysing single and multiple health behaviours. The chapter goes on to then review the ways in which the effects of peers on an individual's likelihood of taking up a specific health behaviour has been accounted for in the past, with a focus on non-network, network statistic and random effects approaches.

Chapter 4 analyses multiple health behaviour data that contain a network component with a multivariate multiple membership multiple classification model that extends the standard models that were reviewed in Chapter 2. The multiple health behaviours in the data set can be modelled using multivariate methods, which is in contrast to the common approach

of each health behaviour being modelled separately. This allows for the effects that peers can have on an individual's observed propensity to engage in a particular health behaviour to be modelled jointly. This is appropriate as one would expect for there to potentially be a relationship between how a peer affects an individual's observed uptake of multiple health behaviours. For example, if a peer negatively affects an individual's observed use of alcohol, we may expect them to negatively affect that same individual's observed use of tobacco to a similar degree. Chapter 5 focuses on studying the properties of a range of different prior distributions for univariate and multivariate models. Chapter 6 extends the models presented in Chapter 4 by allowing for spatial structures to be incorporated into the modelling process, allowing for spatial effects to be estimated.

Chapter 7 provides a brief review of the software that is currently available for researchers to use in order to fit models that can be used to estimate the effects of peers in a network. Chapter 8 details the creation of an `R` package and begins by discussing the motivation and novel contributions that the software package makes to the work possible in the field of peer influence and multiple health behaviours. A subsection that details the data likelihoods available in the package, as well as the univariate and multivariate model structures that can be implemented is provided. This is followed by details of how the package can be installed and used. A large part of this chapter focuses on providing a worked example of a problem that a researcher may come across and how the package can be used in their workflow of analysing a data set. The chapter concludes with a general discussion of the package and future work that can enhance the capabilities of the package.

Chapter 9 concludes the thesis with a discussion of the main results and how the novel contributions made fit into the wider literature on the effects that peers can have on an individual's propensity to engage in particular health behaviours. The limitations of this work are also discussed, highlighting ways in which methods could be extended and potential areas for future work.

Chapter 2

Statistical methods

This chapter sets out the statistical frameworks used in this thesis, and provides the foundation on which novel models are created from in Chapters 4 and 6. The remainder of the chapter is structured as follows. Section 2.1 provides a description of what Bayesian methods are and how they can be used to conduct inference for complex statistical models. Section 2.2 provides a description of Bayesian model selection and checking techniques. Section 2.3 presents a general outline of generalized linear models along with how different data likelihoods can be used in this framework to model both continuous and discrete data. Section 2.4 introduces the concept of multiple membership multiple classification (MMMC) models and the role that they play in modelling network data. Section 2.5 concludes with a presentation of spatial conditional autoregressive (CAR) models, and how such models can be used on areal processes.

2.1 Bayesian methods

The Bayesian approach to inference focuses on updating knowledge about unknowns, $\boldsymbol{\theta}$, in a statistical model on the basis of observations $\mathbf{y}$, with updated knowledge expressed in the

posterior density, $p(\boldsymbol{\theta}|\mathbf{y})$. In this approach, the sample of observations $\mathbf{y}$ being studied provides new information about the unknowns $\boldsymbol{\theta}$, while the prior density $p(\boldsymbol{\theta})$ of the unknowns represents a prior belief about these unknowns before observing or analysing the observations $\mathbf{y}$.

In comparison to a frequentist approach, a Bayesian approach can offer a number of advantages in both the estimation and inferential stage of an analysis through the likelihood function. Bayesian estimation allows complex data structures, such as hierarchical nesting, cross classifications, and spatial domains, that often appear in modern research, to be modeled which would otherwise be infeasible or unreliable with a frequentist approach. It also offers a greater amount of flexibility in allowing for prior knowledge about parameters to be included in an analysis. This results in informative priors that can be used to enable evidence synthesis and conduct a sensitivity analysis. These priors can also be adapted to be made uninformative, which is useful when there is prior ignorance about the unknowns $\boldsymbol{\theta}$ and/or a more objective inference is sought. From an inferential standpoint, in a Bayesian approach we compute a posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ for the unknowns, with the posterior density probability statements can be made about the unknowns, that are not possible when a frequentist approach is used.

Consider the joint density $p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(\mathbf{y}|\boldsymbol{\theta})$ is the data likelihood and $p(\boldsymbol{\theta})$ is the prior density imposed on the unknowns $\boldsymbol{\theta}$. The analysis updates knowledge about the unknowns $\boldsymbol{\theta}$ using the observed data $\mathbf{y}$, and so interest lies in the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$. Since it is also true that $p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})p(\mathbf{y})$, where $p(\mathbf{y})$ is the marginal likelihood, we can obtain that

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.1}$$

This can be rearranged to obtain the posterior density that is of interest

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}. \tag{2.2}$$

The marginal likelihood $p(\mathbf{y})$ can be obtained by $\sum_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ if $\boldsymbol{\theta}$ is discrete or $\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\theta$ if $\boldsymbol{\theta}$ is continuous. Thus, in Equation 2.2, the term $p(\mathbf{y})$ performs the role of a normalising constant that ensures that $p(\boldsymbol{\theta}|\mathbf{y})$ integrates or sums to one. So we may write that

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{2.3}$$

where the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ is proportional to the data likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ times the prior $p(\boldsymbol{\theta})$, which encapsulates prior assumptions about the unknowns $\boldsymbol{\theta}$.

2.1.1 Markov chain Monte Carlo methods

Markov chain Monte Carlo sampling is almost as old as Monte Carlo methods themselves. Informally, Anderson (1986) attributes the use of the first Monte Carlo algorithm to Fermi as early as 1934, this instance was supposedly done by hand. One of the earliest computational implementations of the method was carried out on one of the first computers, called ENIAC (*Electronic Numerical Integrator and Computer*), by John Von Neumann. Von Neumann was interested in applying Monte Carlo methods to thermonuclear and fisson problems as early as 1947. Metropolis and Ulam (1949) then went on to publish the very first paper regarding the Monte Carlo method.

Markov chain Monte Carlo techniques have evolved immensely, both in terms of their computational feasibility and applicability, since what is considered to be their inception

in the publication of Metropolis et al. (1953), which bred the Metropolis algorithm. The computational implementation of this Markov chain Monte Carlo algorithm is closely related to another one of the earliest computers, called MANIAC (*Mathematical Analyzer, Numerical Integrator and Computer*), which was constructed under the watchful eye of Metropolis in Los Alamos in 1952. In Metropolis et al. (1953), the Metropolis algorithm was used in conjunction with MANIAC to compute a high-dimensional integral. As stated in the results of the paper, it took between 4-5 hours for the Metropolis algorithm to produce 16 iterations of burn-in and 48-64 subsequent iterations on the MANIAC. Since then, the processing power of computers has increased at an exponential rate with the development of multi-core processors. These advancements have helped to facilitate new MCMC algorithms which are capable of more accurately solving problems science is faced with in this era.

Markov chain Monte Carlo methods are a class of iterative sampling methods that build upon the class of Monte Carlo methods. Markov chain Monte Carlo simulation involves generating a Markov chain $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)}, \ldots\}$ that is a sequence of correlated samples and whose target distribution is that of the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$. The Markov chain is first initialised by a starting value $\boldsymbol{\theta}^{(0)}$, with subsequent samples in the sequence being drawn, and potentially accepted into the sequence, through a proposal distribution. Candidate values of $\boldsymbol{\theta}^{(t)}$ $(t = 1, 2, 3, ...)$ are generated from a proposal distribution,

$$g(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(0)}, ..., \boldsymbol{\theta}^{(t-1)}) = g(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}), \tag{2.4}$$

that is Markovian, as the distribution only depends on the current value of the sequence $\boldsymbol{\theta}^{(t-1)}$ to generate it's next value $\boldsymbol{\theta}^{(t)}$. The proposal distribution and acceptance rule for candidate values of $\boldsymbol{\theta}^{(t)}$ are selected in a way that after a period of burn-in of length B, samples in the Markov chain $\{\boldsymbol{\theta}^{(t)} : t = B + 1, B + 2, ...\}$ are considered samples from the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$. The Markov chain Monte Carlo methods used throughout this

thesis are either Metropolis-Hastings (Hastings (1970); Metropolis et al. (1953)) or Gibbs (Geman and Geman (1984)) methods, so a brief review of the two are given below are given below.

The Metropolis-Hastings algorithm is a member of the Markov chain Monte Carlo class of algorithms that simulate a Markov chain whose target distribution is the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ required to make inferences on the unknowns $\boldsymbol{\theta}$. The algorithm allows for the partitioning of the parameters in the Markov chain into K sub-vectors, such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K)$. Algorithm 1 describes the Metropolis-Hastings procedure

Algorithm 1: *Metropolis-Hastings*

1. *Initialize the Markov chain* $\boldsymbol{\theta}^{(0)}$ *, such that, when the posterior density is evaluated at* $\boldsymbol{\theta}^{(0)}$, $p(\boldsymbol{\theta}^{(0)}|\mathbf{y}) > 0$.

2. *For each iteration of the sampler* $t = 1, 2, 3...$ *repeat steps (a), (b), and (c) for each of the* K *sub-vectors* $\boldsymbol{\theta}_k$ *(k=1, ...,K).*

   (a) *Generate a candidate sample* $\boldsymbol{\theta}_k^*$ *from the* k*th proposal distribution* $g_k(\cdot|\boldsymbol{\theta}_k^{(t-1)})$, *where* $\boldsymbol{\theta}_k^{(t-1)}$ *is the current* $\boldsymbol{\theta}_k$ *in the Markov chain.*

   (b) *Generate an acceptance probability* u *from a uniform distribution with support on [0, 1],* $U \sim$ *Uniform*$(0, 1)$.

   (c) *Accept* $\boldsymbol{\theta}_k^*$ *as the next iteration of the sub-vector in the Markov chain* $\boldsymbol{\theta}_k^{(t)} = \boldsymbol{\theta}_k^*$ *if*

$$u \leq \frac{p(\boldsymbol{\theta}^*|\mathbf{y}) g_k(\boldsymbol{\theta}_k^{(t)}|\boldsymbol{\theta}_k^*)}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y}) g_k(\boldsymbol{\theta}_k^*|\boldsymbol{\theta}_k^{(t-1)})}, \tag{2.5}$$

   otherwise set $\boldsymbol{\theta}_k^{(t)}$ *to the current value* $\boldsymbol{\theta}_k^{(t-1)}$, $\boldsymbol{\theta}_k^{(t)} = \boldsymbol{\theta}_k^{(t-1)}$, *where* $\boldsymbol{\theta}^*$ *is the partially updated Markov chain* $(\boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_{k-1}^{(t)}, \boldsymbol{\theta}_k^*, \boldsymbol{\theta}_{k+1}^{(t-1)}, ..., \boldsymbol{\theta}_K^{(t-1)})$ *and* $g_k(\boldsymbol{\theta}_k^*|\boldsymbol{\theta}_k^{(t)})$ *is the probability of* $\boldsymbol{\theta}^*$ *when the proposal distribution is parameterized by* $\boldsymbol{\theta}^{(t)}$.

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, where the proposal distribution used to generate candidate sample $\boldsymbol{\theta}_k^*$ is the full conditional distribution, and so it follows that the inequality in Equation 2.5 becomes $u \leq 1$. Thus, the acceptance steps (b) and (c) in Algorithm 1 are essentially removed. Markov chain simulation is a complex method that has many components that can be altered in order to improve its efficiency, such as the partitioning of $\boldsymbol{\theta}$ and proposal density.

While it is possible to update the entire vector of parameters in the Markov chain at once, partitioning $\boldsymbol{\theta}$ into blocks or single components can enable better mixing of the parameter space that the Markov chain explores and can lessen the computational expense associated with the algorithm, as the full conditional distribution for some blocks of parameters may allow for the use of a Gibbs step instead of the more computationally intense Metropolis-Hastings step. This does away with the inequality in Equation 2.5 having to be evaluated, which could prove to become very computationally costly if the functions are complex or the number of iterations for the algorithm is set to be very large.

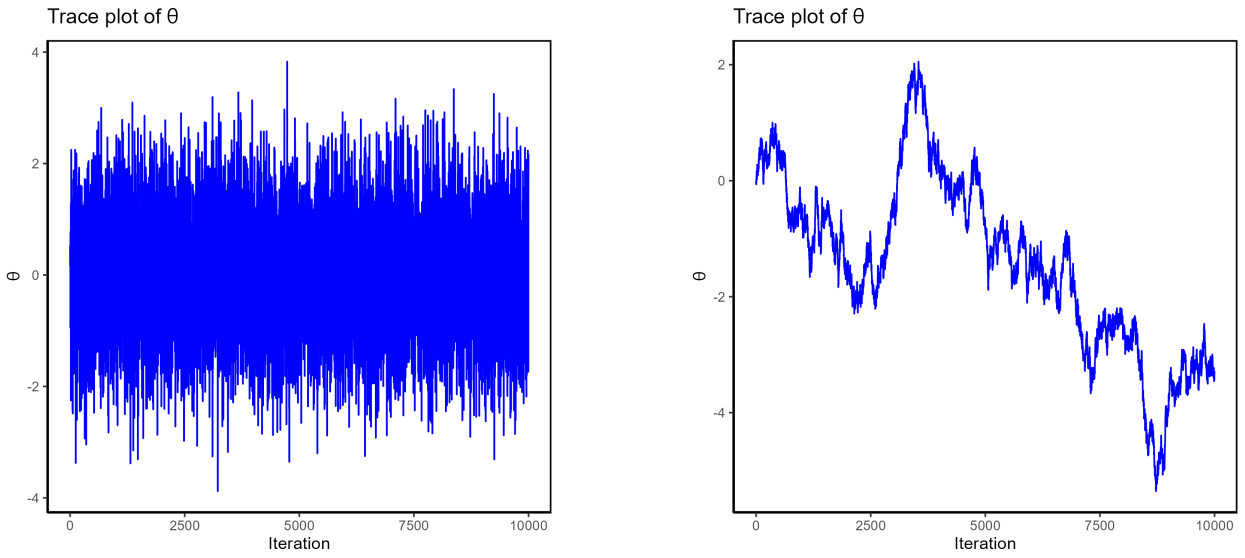
There are many ways in which a proposal density for generating candidate samples for a block of parameters can be chosen. A standard approach is to choose a proposal density with a support that is compatible with the block of parameters and approximates what one thinks the posterior density for that block of parameters looks like. In the case of parameters with infinite support, a Gaussian proposal distribution centred around the current value is often used and has the benefit of symmetry, which means that the evaluation in Equation 2.5 becomes $u \leq p(\boldsymbol{\theta}^*|\mathbf{y})/p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})$ as $g_k(\boldsymbol{\theta}_k^{(t)}|\boldsymbol{\theta}_k^*) = g_k(\boldsymbol{\theta}_k^*|\boldsymbol{\theta}_k^{(t-1)})$. The proposal distribution also plays a vital role in determining the acceptance rate of candidate samples $\boldsymbol{\theta}_k^*$, as the decision to accept or reject a candidate depends on how far $\boldsymbol{\theta}_k^*$ is from $\boldsymbol{\theta}_k^{(t-1)}$ in the parameter space. In the Metropolis-Hastings algorithm, a high acceptance rate is usually associated with a proposal distribution whose variance is small and so tends to choose candidates close to current values, which risks not fully exploring the parameter space. A remedy for this, which allows for the parameter space to be better explored and a more moderate acceptance rate, is

to employ a form of adaptive proposal distribution. Suppose that a proposal distribution has a variance σ_g^2, an adaptive proposal distribution works by scaling this variance after every set amount of iterations in the Metropolis-Hastings algorithm; increasing and updating its value if the acceptance rate is deemed too large or decreasing and updating its value if the acceptance rate is deemed too small.

2.1.2 MCMC diagnostics

2.1.2.1 Trace plots

A common way to inspect sampling behaviour and and assess mixing across chains and convergence is through plotting the trace plot of the Markov chain. Given a single Markov chain for a parameter θ_i $(i = 1, ..., N)$ of length N, we can plot this against the corresponding iteration of the sampler to produce a trace plot for the parameter.



(a) (b)

Figure 2.1: Trace plots of θ for a well behaved chain (a) and a chain that exhibits poor mixing (b).

Figure 2.1a shows a trace plot for the parameter θ that exhibits very good behaviour and

mixing. In comparison, Figure 2.1b shows a trace plot of θ that demonstrates poor mixing. Problems relating to convergence and mixing in MCMC sampling may reflect problems with model identifiability or sub-optimal model parameterization. This method is used throughout the thesis to inspect the mixing of parameter chains produced by models.

2.1.2.2 Autocorrelation

A common way to assess the convergence of a Markov chain is to evaluate the autocorrelation between the elements of the Markov chain. This method is used in later chapters of the thesis to evaluate the autocorrelation between the elements of the Markov chains produced by the models presented. The lag k autocorrelation ρ_k is the correlation between every draw and its k^{th} lag. The autocorrelation, ρ_k, is defined as

$$\rho_k = \frac{\sum_{i=1}^{N-k}(\theta_i - \bar{\theta})(\theta_{i+k} - \bar{\theta})}{\sum_{i=1}^{N}(\theta_i - \bar{\theta})^2}. \tag{2.6}$$

As the value of k increase we would expect the k^{th} lag autocorrelation to decrease, as the elements of the Markov chain used in the calculation of ρ_k become more distant in time. For example, within a generated Markov chain, the correlation between states 1 time unit apart should be greater than those between elements 50 time units apart. If the autocorrelation is relatively high for large values of k, this is indicative of a Markov chain which is characterized as exploring the parameter space slowly (slow mixing). In such cases, it may be of interest to re-parametrize the model or adjust the MCMC algorithm to alleviate the issue. The consequence of high autocorrelation is that we would need to generate a Markov chain of greater length to obtain a given level of precision for our MCMC estimate.

2.1.2.3 Effective sample size

Given the presence of autocorrelation induced by a Markov chain, a method to define an effective sample size (ESS) (Kass et al. (1998); Ripley (1987)) for an MCMC procedure can be produced. The ESS has a natural interpretation as the number of effectively independent draws from a Markov chain of the target distribution, and is given by

$$\text{ESS} = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho_k} \tag{2.7}$$

where N is the total number of Markov chain samples. This method is used in later chapters of the thesis to evaluate the ESS of Markov chains produced by the models presented.

2.1.2.4 Geweke z

The Geweke z diagnostic (Geweke (1992)) is a method that uses spectral analysis and the segmentation of the Markov chain to assess convergence. Given a single Markov chain for a parameter θ_i $(i = 1, ..., N)$ of length N, we can obtain two subsequences of this chain $\{\theta_A\}$ and $\{\theta_B\}$, where $A = \{i;\ 1 \leq i \leq N_A\}$, $B = \{i;\ N^* \leq i \leq N\}$, $1 \leq N_A \leq N^* \leq N$, $|A| = N_A$ and $|B| = N_B$. Let

$$\bar{\theta}_A = \frac{1}{N_A}\sum_{i \in A} \theta_i \ \text{ and } \ \bar{\theta}_B = \frac{1}{N_B}\sum_{i \in B} \theta_i,$$

then the Geweke z statistic has an asymptotically standard normal distribution that is given by

$$z = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{1}{N_A}\hat{S}_\theta^A(0) + \frac{1}{N_B}\hat{S}_\theta^B(0)}} \to N(0, 1), \tag{2.8}$$

where $\hat{S}_\theta(0)$ is a spectral density with no discontinuities at frequency 0 and $\frac{1}{N_A}\hat{S}_\theta^A(0)$ and

$\frac{1}{N_B}\hat{S}_\theta^B(0)$ are asymptotic variances. The null hypothesis of equal location in the parameter space for θ, which indicates convergence, is rejected when z is large (i.e., $|z| > 1.96$). This method is used in later chapters of the thesis to assess the convergence of Markov chains produced by the models presented.

2.1.2.5 Gelman-Rubin statistic

Convergence for multiple Markov chains can be assessed using the Gelman-Rubin scale reduction factor (Gelman et al. (2013)), which evaluates mixing using the between- and within-sequence variances for the simulated posterior distribution of a given parameter. We can label these as $\theta_{ij}(i = 1, ..., N; j = 1, ..., M)$ where N is the length of the chain post-burn-in and $M > 1$ is the number of chains. From this, we can compute the between- and within-sequence variances, B and W:

$$B = \frac{N}{M-1}\sum_{j=1}^{M}(\bar{\theta}_{.j} - \bar{\theta}_{..})^2, \text{ where } \bar{\theta}_{.j} = \frac{1}{N}\sum_{i=1}^{N}\theta_{ij}, \ \bar{\theta}_{..} = \frac{1}{M}\sum_{j=1}^{M}\theta_{.j}$$

$$W = \frac{1}{M}\sum_{j=1}^{M}s_j^2, \text{ where } s_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\theta_{ij} - \bar{\theta}_{.j})^2.$$

The marginal posterior variance of the parameter, $\text{var}(\theta|y)$ can be estimated using a weighted sum of B and W, namely

$$\hat{\text{var}}(\theta|y) = \frac{N-1}{N}W + \frac{1}{N}B.$$

The potential scale reduction factor $\hat{R}$ is then estimated by

$$\hat{R} = \sqrt{\frac{\hat{\text{var}}(\theta|y)}{W}},$$

which declines to 1 as $N \to \infty$. This potential scale reduction factor provides an estimate of how much variance could be reduced by running chains longer. If the potential scale reduction factor is high, then this provides reason to believe that running the simulation for longer may improve inference about the target distribution of the associated parameter θ. The Gelman-Rubin statistic is used in later chapters of the thesis to assess the convergence of Markov chains produced by the models presented.

2.2 Bayesian model selection and checking

2.2.1 DIC

The deviance information criterion (DIC) (Spiegelhalter et al. (2002)) is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo simulation and is used extensively throughout this thesis.

The main idea behind this method is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of $\bar{D}$, which favors a good fit, but also by the effective number of parameters p_D. Since $\bar{D}$ will decrease as the number of parameters in a model increases, the p_D term compensates for this effect by favoring models with a smaller number of parameters. An advantage of DIC over other criteria in the case of Bayesian model selection is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation.

The formulation of the DIC is given as

$$\begin{aligned} \text{DIC} &= D(\bar{\boldsymbol{\theta}}) + 2p_D \\ &= \bar{D} + p_D \end{aligned}$$

where $\boldsymbol{\theta}$ is the parameter space and $p_D = \bar{D} - D(\bar{\boldsymbol{\theta}})$. p_D is the effective number of independent parameters. $\bar{D}$ is the posterior mean deviance. $\bar{D} = \mathbb{E}_{\boldsymbol{\theta}|y}(-2\ \log(p(y|\boldsymbol{\theta})))$. $D(\bar{\boldsymbol{\theta}})$ is the deviance of posterior means. $D(\bar{\boldsymbol{\theta}}) = -2\ \log(p(y|\mathbb{E}[\boldsymbol{\theta}|y]))$. Algorithm 2 describes how the DIC procedure would be computed, given simulated samples from the posterior distribution for each parameter

Algorithm 2: *DIC*

1. *Calculate* $\bar{D}$.
2. *For each iteration of the sampler* $t = 1, 2, 3, ...$ *(post burn-in and thinning), calculate* $-2log(p(y|\boldsymbol{\theta}^t))$.
3. *Compute* $\bar{D}$*, which is given by,*

$$\bar{D} = \frac{\sum_t -2\text{log}(p(y|\boldsymbol{\theta}^t))}{N}, \tag{2.9}$$

   where N is the number of iterations post burn-in and thinning.

4. *Compute* $D(\bar{\boldsymbol{\theta}})$*, which is given by,*

$$D(\bar{\boldsymbol{\theta}}) = -2\text{log}(p(y|\mathbb{E}(\boldsymbol{\theta}|y))). \tag{2.10}$$

2.2.2 Posterior predictive checking

Posterior predictive checking is a Bayesian method used to generate replicated responses $\mathbf{y}^{rep}$ under the model and comparing the replicated data $\mathbf{y}^{rep}$ to that of the observed data $\mathbf{y}$ in order to determine whether it is plausible that the observed data is plausible under the posterior predictive distribution. This method is used in Chapter 6 of the thesis to assess the fit of a novel model presented. The posterior predictive distribution is given by

$$p(\mathbf{y}^{rep}|\mathbf{y}) = \int_{\boldsymbol{\theta}} p(\mathbf{y}^{rep}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \tag{2.11}$$

where $\mathbf{y}$ is the observed data and $\mathbf{y}^{rep}$ is the replicated data from the model. The posterior predictive distribution can be approximated by, sampling $\boldsymbol{\theta}$ from the simulated posterior distribution obtained through the sampling algorithm and generating $\mathbf{y}^{rep}$ by parameterizing the data likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ with the given sample of $\boldsymbol{\theta}$. The posterior predictive distribution can assess model adequacy through the use of a discrepancy measure $T(\mathbf{y}|\boldsymbol{\theta})$, which is a scalar summary of parameters and data that is used to compare observed data to predicted simulations. This discrepancy measure can be used to produce a Bayesian p-value that is defined as the probability that the discrepancy measure evaluated with the replicated data could be more extreme than with the observed data,

$$p_B = \mathbb{P}(T(\mathbf{y}^{rep}|\boldsymbol{\theta}) \geq T(\mathbf{y}|\boldsymbol{\theta})|\mathbf{y}). \tag{2.12}$$

2.2.3 Cross-validation

In Bayesian cross-validation, the data are repeatedly partitioned into a training set $\mathbf{y}_{train}$ and a holdout set $\mathbf{y}_{holdout}$, and then the model is fit to $\mathbf{y}_{train}$, yielding a posterior distribution, $p(\boldsymbol{\theta}|\mathbf{y}_{train})$, with this fit evaluated using an estimate of the log predictive density of the holdout data, $\log(p_{train}(\mathbf{y}_{holdout}))= \log\big(\int p_{pred}(\mathbf{y}_{holdout}|\boldsymbol{\theta})p_{train}(\boldsymbol{\theta})d\boldsymbol{\theta}\big)$. Assuming the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{train})$ is summarized by S simulation draws $\boldsymbol{\theta}$, we calculate the log predictive density as $\log\big(\frac{1}{S}\sum_{s=1}^{S} p(\mathbf{y}_{holdout}|\boldsymbol{\theta}^s)\big)$.

Leave-one-out cross-validation (LOO-CV) is a special case with N partitions in which each holdout set represents a single data point. Performing the analysis for each of the N data points yields N different inferences $p_{post}(-i)$, each summarized by S posterior simulations, $\boldsymbol{\theta}^{is}$.

The Bayesian LOO-CV estimate of out-of-sample predictive fit is

$$lppd_{\text{LOO-CV}} = \sum_{i=1}^{N} \log(p_{post(-i)}(\mathbf{y}_i)), \quad \text{and calculated as } \sum_{i=1}^{N} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(\mathbf{y}_i|\boldsymbol{\theta}^{is})\right). \quad (2.13)$$

Each prediction is conditioned on $N-1$ data points, which causes under-estimation of the predictive fit but for large N the difference is negligible. In this thesis, although popular, the cross-validation method is not used due to the computation complexity of the novel models presented in Chapters 4 and 6. The Watanabe–Akaike information criterion (WAIC) Watanabe (2010) is also explored in this thesis. However, it relies on a partition of the data into N pieces, which can be limiting in some structured-data settings such as spatial and network data, which are key components of models in this thesis.

2.3 Generalized linear models

Generalized linear models (GLMs) were introduced by Nelder and Wedderburn (1972) and constitute a wide class of models encompassing stochastic representations used for the analysis of both quantitative (continuous or discrete) and qualitative response variables. They can be regarded as the natural extension of normal linear regression models and are based on the exponential family of distributions, which includes the most common distributions such as the normal, binomial, and Poisson. Generalized linear models have become very popular because of their generality and wide range of applications. They can be considered one of the most prominent and important components of modern statistical theory. They have provided not only a family of models that are widely used in practice but also a unified, general way of thinking concerning the formulation of statistical models. Chapter 4 utilises and builds upon the GLM framework to compare and contrast a class of univariate and multivariate models for social network data. Chapter 5 uses parts of the GLM framework to explore the properties of different priors. Chapter 6 uses and builds upon the GLM framework to propose a novel multivariate spatio-network model. Chapter 8 introduces an `R` package, `netcmc`, which contains the implementation of GLMs.

The generalized linear model is comprised of three main components:

1. Random component:

$$Y_i \sim \exp\big[y_i\theta_i - b(\theta_i) + c(y_i)\big] \qquad \text{for } i = 1, ..., n, \tag{2.14}$$

   where $\exp\big[y_i\theta_i - b(\theta_i) + c(y_i)\big]$ denotes the exponential family distribution with location parameters θ_i.

2. Systematic component:

   Parametric case:

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \qquad \text{for } i = 1, ..., n. \tag{2.15}$$

3. Link function, the link between the random and systematic components:

$$g(\theta_i) = \eta_i \qquad \text{for } i = 1, ..., n. \tag{2.16}$$

These three components form the foundation upon which many of the novel models in this thesis are built.

2.3.1 The exponential family

A family of probability density functions which can be written in the form

$$f(y_i; \theta_i) = \exp\big[y_i\theta_i - b(\theta_i) + c(y_i)\big] \tag{2.17}$$

is called a *natural exponential family* of distributions. The function $b(\theta_i)$ is called the *cumulant generator*. This representation is called the *canonical parametrization* of the family, and the parameter θ is called the *canonical parameter*.

2.3.2 The exponential dispersion family

The exponential family can be generalized by including a dispersion parameter, say ξ, in the distribution, such that

$$f(y_i; \theta_i, \xi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\xi)} + c(y_i, \xi)\right], \tag{2.18}$$

where θ_i is still the canonical parameter. A family of probability densities which can be written in this form is called a natural exponential dispersion family of distributions. The

mean and variance of Y with distribution in the exponential disperion family with parameters θ and ξ are

$$\mathbb{E}(Y) = \frac{db(\theta)}{d\theta} = b^{'}(\theta) \text{ and } \text{Var}(Y) = \frac{d^2b(\theta)}{d\theta^2}a(\xi) = b^{''}(\theta)a(\xi). \tag{2.19}$$

2.3.3 Common exponential dispersion family distributions

The Gaussian distribution. The probability density function for the Gaussian distribution is given by

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\text{exp}\bigg[-\frac{1}{2\sigma^2}(y-\mu)^2\bigg] = \text{exp}\bigg[\frac{1}{\sigma^2}\bigg(y\mu - \frac{\mu^2}{2}\bigg) - \frac{y^2}{2\sigma^2} + \log\bigg(\frac{1}{\sqrt{2\pi}\sigma}\bigg)\bigg].$$

As a result, the Gaussian distribution belongs to the exponential dispersion family, where $\theta = \mu$, $b(\theta) = \mu^2/2$ and $a(\xi) = \sigma^2$. As expected, $\mathbb{E}(Y) = b^{'}(\theta) = \mu$ and $\text{Var}(Y) = b^{''}(\theta)a(\xi) = \sigma^2$.

The binomial distribution. The probability mass function for the binomial distribution Binomial(N, π) is given by

$$f(y;N,\pi) = \binom{N}{y}\pi^y(1-\pi)^{N-y} = \text{exp}\bigg[y\text{log}\bigg(\frac{\pi}{1-\pi}\bigg) + N\text{log}(1-\pi) + \log\binom{N}{y}\bigg].$$

As a result, the binomial distribution belongs to the exponential dispersion family, where $\theta = \log(\pi/(1-\pi))$ (note: $\pi = \exp(\theta)/(1+\exp(\theta))$), $b(\theta) = N\log(1+\exp(\theta))$ and $a(\xi) = 1$. As expected, $\mathbb{E}(Y) = b^{'}(\theta) = N\exp(\theta)/(1+\exp(\theta)) = N\pi$ and $\text{Var}(Y) = b^{''}(\theta)a(\xi) = N\exp(\theta)/(1+\exp(\theta))^2 = N\pi(1-\pi)$.

The Poisson distribution. The probability mass function for the Poisson distribution Poisson(λ) is given by

$$f(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} = \exp\left[y\log(\lambda) - \lambda + \log\left(\frac{1}{y!}\right)\right].$$

As a result, the Poisson distribution belongs to the exponential dispersion family, where $\theta = \log(\lambda)$ (note: $\lambda = \exp(\theta)$), $b(\theta) = \lambda$ and $a(\xi) = 1$. As expected, $\mathbb{E}(Y) = b^{'}(\theta) = \exp(\theta) = \lambda$ and $\text{Var}(Y) = b^{''}(\theta)a(\xi) = \exp(\theta) = \lambda$.

These common exponential dispersion family of distributions are the main focus of the `netcmc` software package presented and discussed in Chapter 8. The Bernoulli distribution, which is a special case of the binomial distribution, is also used as the likelihood for models presented in Chapters 4 and 6.

2.3.4 Link functions

The link function is a monotonic and differentiable function used to match the parameters of the response variable with the systematic component, namely, the linear predictor and the associated covariates. Usually no restriction lies on the definition of such variables, but often we focus on the mean of the distribution because the measures of central location are usually of main interest. GLM-based extensions in which dispersion or shape parameters are linked with covariates also exist in statistical literature.

Usually the default choice of link function is provided by the canonical link, in which we set the canonical parameter, as an expression of the mean or other parameters of the distribution, equal to the linear predictor.

For certain exponential dispersion family distributions, a wide range of link functions exist. The binomial model is able to utilize a wide variety of link function such as the logit link, which, for a parametric systematic component, is specified to be

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^\top \boldsymbol{\beta}. \tag{2.20}$$

The interpretation of this logit link function for the binomial model is quite straightforward to comprehend. In this parameterization, π_i is the probability of success and $1-\pi_i$ is the probability of failure. Thus $\pi_i/(1-\pi_i)$ are the odds and $\pi_i/(1-\pi_i) = \exp(x_{i1}\beta_1 + ... + x_{ip}\beta_p)$. As a result, holding all other covariates fixed, a unit increase in the p^{th} covariate multiplies the odds by a factor of e^{β_p}. If e^{β_p} is greater than 1, then these odds have increased. If e^{β_p} is less than 1, then these odds have decreased. This logit link function is used in many of the models presented in later chapters of this thesis.

One other very popular link function for the binomial model is the probit link, which is given by

$$g(\pi_i) = \Phi^{-1}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{2.21}$$

where $\Phi(.)$ is the cumulative distribution function (CDF) of the standard normal distribution and is given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{2.22}$$

Another popular link function for the binomial model is the complementary log-log link, which is given by

$$g(\pi_i) = \log(-\log(1-\pi_i)) = \mathbf{x}_i^\top \boldsymbol{\beta}. \tag{2.23}$$

2.4 Multiple membership multiple classification models

Multiple membership multiple classification models build on multilevel models (see Goldstein and Browne (2016)), which themselves are extensions of the GLM discussed in the previous section. Consider a problem that has one response variable and assume that there is a unique response in our dataset for each of N lowest level units. The lowest level units could be individuals, time points, areas, etc. Browne et al. (2001) defines a *classification* as a function, c, that maps from the set $\mathbf{\Theta}$ of N lowest units to a set $\mathbf{\Phi}$ of size M where $M \leq N$, and we define the resulting set $\mathbf{\Phi}$ of M objects as the *classification units*. So we have $c(n_i) = \mathbf{\Phi}_i$, where the lowest level unit $n_i \in \mathbf{\Theta}$ and $\mathbf{\Phi}_i \subset \mathbf{\Phi}$. Browne et al. (2001) considers two types of classifications, single member classification and multiple member classification. A single member classification (see Figure 2.2) is a function c from $\mathbf{\Theta}$ to $\mathbf{\Phi}$ that maps each $n_i \in \Theta$ to a unique $m_j \in \mathbf{\Phi}$. A multiple membership classification (see Figure 2.3) is a map c from $\mathbf{\Theta}$ to $\mathbf{\Phi}$ that maps each $n_i \in \mathbf{\Theta}$ to a subset (possibly of size 1) $\mathbf{\Phi}_i$ of $\mathbf{\Phi}$.

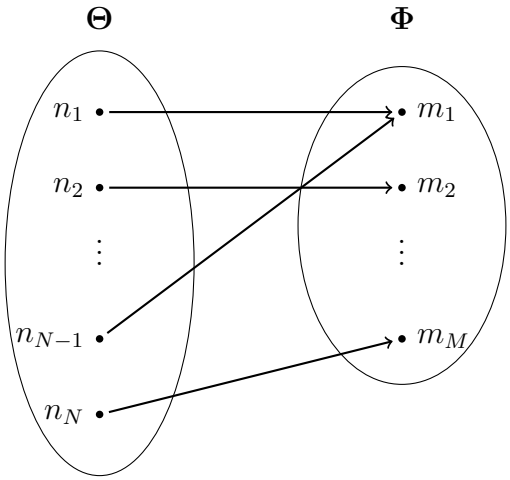


Figure 2.2: An example of a single member classification.

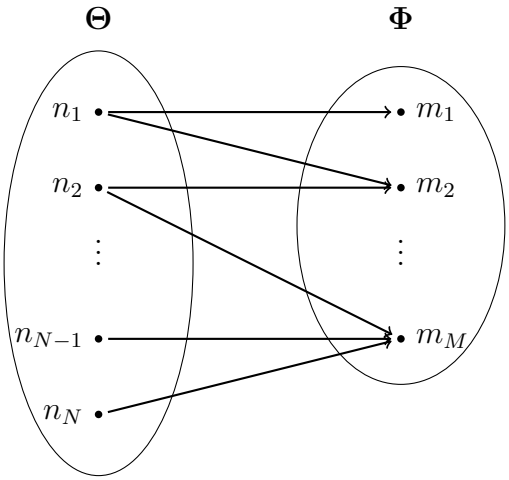


Figure 2.3: An example of a multiple member classification.

We will still maintain that $M \leq N$ to avoid identifiability problems in estimation. Given these definitions we will now see that all the sets of random effects that feature in multilevel

models, cross-classified models, multiple membership models and MMMC models will have an associated classification. 2-level multilevel models are made up of one single membership classification. 2-level cross-classified models are made up of two or more *single membership classifications* (see Figure 2.4).

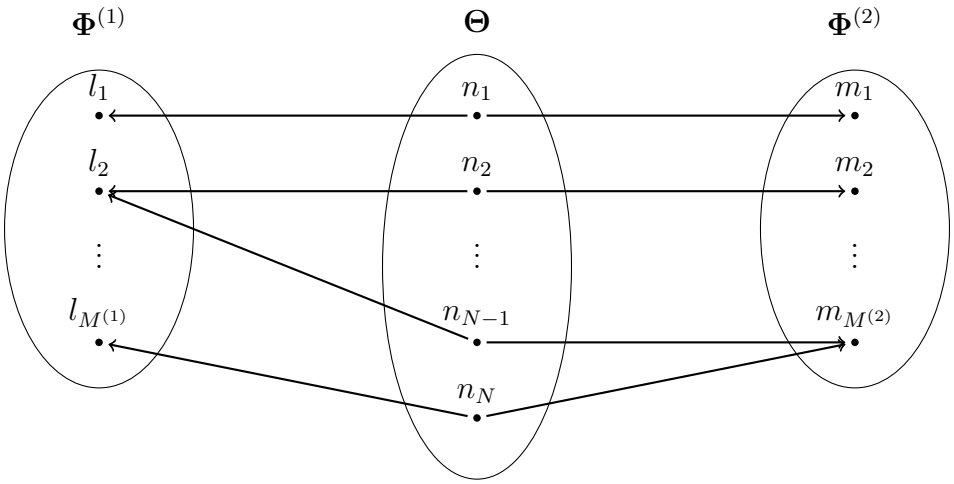


Figure 2.4: An illustrative example of two single member classifications.

2-level *multiple membership multiple classification* models are made up of at least one single membership classification and at least one multiple membership classification (see Figure 2.5).

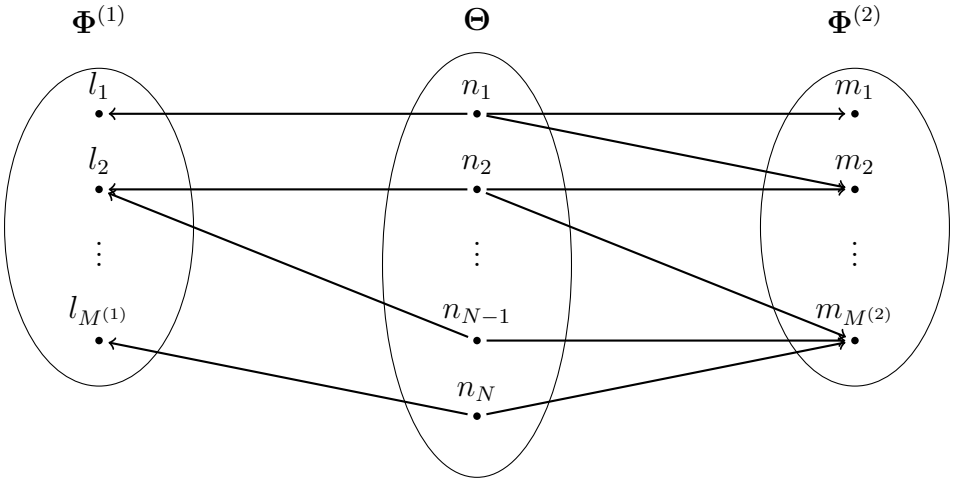


Figure 2.5: An illustrative example of a *multiple membership multiple classification*.

2.4.1 Multilevel model specification

The multilevel model is described in Browne et al. (2001) and is primarily used to model multilevel data. An example of a multilevel process would be whether or not an individual (a level 1 unit) belonging to one rehab center (a level 2 unit) each experience a relapse. Conducting inference with the multilevel model is typically done using a Bayesian approach. The general multilevel model is given by:

$$
\begin{aligned}
Y_i &\sim f(Y_i; \theta_i, \xi) \qquad \text{for } i = 1, ..., N, \\
g(\theta_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{v}_{\boldsymbol{\Phi}(i)}, \\
\boldsymbol{\beta} &\sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \\
\mathbf{v}_{\boldsymbol{\Phi}(i)} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}}), \\
\boldsymbol{\Sigma}_{\mathbf{v}} &\sim \text{Inverse-Wishart}(\nu, \boldsymbol{\Omega}),
\end{aligned}
$$

where $f(Y_i; \theta_i, \xi)$ is a natural exponential dispersion family of distributions, $g(.)$ is an invertible link function, y_i is the response of the ith observation, $\boldsymbol{\Phi}(i)$ is the element in $\boldsymbol{\Phi}$ that the ith individual is a member of in the single membership classification, $\boldsymbol{\beta}$ is a vector of fixed effects parameter, $\mathbf{v}_{\boldsymbol{\Phi}(i)} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}})$ is a vector of random effects relating to the single membership classification. $\mathbf{x}_i$ and $\mathbf{z}_i$ are vectors of predictors.

2.4.2 Multiple membership multiple classification model specification

The multiple membership multiple classification model was introduced in Browne et al. (2001) and is primarily used to model multilevel data in which individuals can belong to more than one higher level unit consisting of at least one single membership classification

and at least one multiple membership classification. An example of such a process would be the grade of individuals (a level 1 unit) that has belonged to one nursery and multiple secondary schools. Conducting inference with the MMMC model is typically done using a Bayesian approach. The general MMMC model is given by:

$$
\begin{aligned}
Y_i &\sim f(Y_i; \theta_i, \xi) \qquad \text{for } i = 1, ..., N, \\
g(\theta_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{v}_{\boldsymbol{\Phi}^{(1)}(i)} + \sum_{j \in \boldsymbol{\Phi}^{(2)}(i)} w_{ij} \mathbf{g}_i^\top \mathbf{u}_j, \\
\boldsymbol{\beta} &\sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \\
\mathbf{v}_{\boldsymbol{\Phi}^{(1)}(i)} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}}), \\
\mathbf{u}_j &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}}), \\
\boldsymbol{\Sigma}_{\mathbf{v}} &\sim \text{Inverse-Wishart}(\nu_1, \boldsymbol{\Omega}_1), \\
\boldsymbol{\Sigma}_{\mathbf{u}} &\sim \text{Inverse-Wishart}(\nu_2, \boldsymbol{\Omega}_2),
\end{aligned}
$$

where $f(Y_i; \theta_i, \xi)$ is a natural exponential dispersion family of distributions, $g(.)$ is an invertible link function, y_i is the response of the ith observation, $\boldsymbol{\Phi}^{(1)}(i)$ is the element in $\boldsymbol{\Phi}^{(1)}$ that the ith individual is a member of through the single membership classification, $\boldsymbol{\Phi}^{(2)}(i)$ is the element(s) in $\boldsymbol{\Phi}^{(2)}$ that the ith individual are a member of through the multiple membership classification, $\boldsymbol{\beta}$ is a vector of fixed effects parameter, $\mathbf{v}_{\boldsymbol{\Phi}^{(1)}(i)} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}})$ is a vector of random effects relating to the single membership classification, $\mathbf{u}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}})$ is a vector of random effects relating to the single membership classification, the random variables $\mathbf{v}_{\boldsymbol{\Phi}^{(1)}(i)}$ and $\mathbf{u}_j$ are mutually independent. $\mathbf{x}_i$, $\mathbf{z}_i$ and $\mathbf{g}_i$ are vectors of predictors. w_{ij} is a scalar weight.

2.5 Spatial modelling

Spatial models extend those previously described in Sections 2.3 and 2.4 of this chapter by accounting for the spatial structure present in the data. Observed spatial data is the realisation from an unknown stochastic process, which if estimated, would allow us to draw inferences about the spatial data we are modelling. A stochastic process is a family of random variables $\{Y(\mathcal{S})\colon \mathcal{S} \in D\}$ indexed by locations $\mathcal{S} \in D$, defined on a probability space, where D is the spatial domain of the process and $Y(\mathcal{S})$ is a random variable representing the quantity that is measured at location $\mathcal{S}$. Spatial data can be observed at a finite number of points over a continuous region (geostatistical) or a set of finite discrete points / sub-regions (areal process). The spatial regression models used in Chapter 6 apply to survey data from adolescents belonging to Zip Codes and thus are spatial data observed at sub-regions, so a brief review of areal spatial models is given below. For more general reviews of spatial models see Cramb et al. (2017). For an areal process the spatial domain D is partitioned into S non-overlapping areal units, which are denoted by

$$D = \{\mathcal{S}_1, ..., \mathcal{S}_S\},$$

where the set of areal units $\{\mathcal{S}_1, ..., \mathcal{S}_S\}$ form a regular grid or are irregularly shaped satisfying two conditions; (1) $\cup_{i=1}^{S} \mathcal{S}_i = D$ and (2) $\mathcal{S}_i \cap \mathcal{S}_j \neq \emptyset$ for all $i \neq j$. This partitioning of the spatial domain results in the areal process having the following stochastic process

$$\mathbf{Y} = \{Y(\mathcal{S}_1), ..., Y(\mathcal{S}_S)\},$$

which is only defined on the n areal units $\{\mathcal{S}_1, ..., \mathcal{S}_S\}$.

2.5.1 Intrinsic CAR

Letting $Y(\mathcal{S}_i)$ denote a spatial observation at the areal unit i, the simplest conditional autoregressive model used for spatial data is the intrinsic conditional autoregressive (ICAR) model, which has an improper multivariate Gaussian distribution given by

$$\mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \tau^2(\mathrm{diag}(\mathbf{A1}) - \mathbf{A})^{-1}), \tag{2.24}$$

where $\mathbf{A}_{S\times S}$ is a non-negative spatial adjacency matrix which defines how spatially close the S areal units are to each other. The elements of $\mathbf{A}_{S\times S}$ can be binary or non-binary. In the more common binary case, $a_{sl} = 1$ if a pair of areal units $(\mathcal{S}_s, \mathcal{S}_l)$ share a common border or are considered neighbours by some other measure, and $a_{sl} = 0$ otherwise. This is not a valid data likelihood model, because the precision matrix $\mathbf{Q} = (\mathrm{diag}(\mathbf{A1}) - \mathbf{A})/\tau^2$ is singular and hence the variance matrix does not exist. Thus, this model does not specify a proper joint distribution for $\mathbf{Y}$. In addition to this, the data model specified above is only appropriate for continuous data with support $y \in \mathbb{R}$, and so is not appropriate for discrete count data or continuous data whose support is not $y \in \mathbb{R}$, such as half-normal data with support $y \in [0, \infty)$. Thus, it is common to specify the intrinsic CAR model as a prior distribution for a set of random effects, rather than directly for the data likelihood model. This approach resolves the two limitations previously mentioned, as in a Bayesian hierarchical model prior distributions can be improper and the data likelihood can take the form of an appropriate distribution for the data. Also, by allowing specifying the intrinsic CAR model as a prior distribution for a set of random effects, this allows the response variable to be something other than the stochastic process within the spatial domain, such as a response generated by a process relating to individuals. A Bayesian hierarchical model approach that specifies the intrinsic CAR model as a prior distribution for a set of random effects is given by

$$
\begin{aligned}
Y_s &\sim f(Y_s;\theta_s,\xi) \quad \text{for } s=1,\ldots,S. \\
g\left(\theta_s\right) &= \mathbf{x}_s^\top \boldsymbol{\beta} + \phi_s, \\
\boldsymbol{\beta} &\sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \\
(\phi_1,\ldots,\phi_S) = \boldsymbol{\phi} &\sim \mathrm{N}(\mathbf{0}, \tau^2(\mathrm{diag}(\mathbf{A1}) - \mathbf{A})^{-1}), \\
\tau^2 &\sim \text{Inverse-Gamma}(a_\tau, b_\tau),
\end{aligned} \tag{2.25}
$$

where $f(Y_s;\theta_s,\xi)$ is a natural exponential dispersion family of distributions, $g(.)$ is an invertible link function, y_s is the response of the sth spatial unit. The covariates are included in a $p \times 1$ vector $\mathbf{x}_s$, and a corresponding $p \times 1$ vector of fixed effect parameters are denoted by $\boldsymbol{\beta}$. The $S \times 1$ vector of joint random effects for the S spatial units is denoted by $\boldsymbol{\phi} = (\phi_1, ..., \phi_S)_{S\times 1}$ and has a joint Gaussian distribution. τ^2 is a measure of the variance relating to the vector of spatial random effects $\boldsymbol{\phi}$. a_τ and b_τ are user-chosen hyperparameters for the Inverse-Gamma distribution imposed on τ^2. Although the joint distribution of the ICAR model is improper, the full conditional distributions $\phi_s|\boldsymbol{\phi}_{-s}$ are proper. Specifically

$$\begin{aligned}
\mathbb{E}(\phi_s|\boldsymbol{\phi}_{-s}) &= \mu_s - Q_{ss}^{-1}\mathbf{Q}_{s,-s}(\boldsymbol{\phi}_{-s} - \boldsymbol{\mu}_{-s}) \\
&= 0 - \left((1/\tau^2)\sum_{l=1}^{S} a_{sl}\right)^{-1}\left(-(1/\tau^2)\sum_{l\neq s}^{S} a_{sl}(\phi_l - 0)\right) \\
&= \frac{\sum_{l\neq s} a_{sl}\phi_l}{\sum_{l=1}^{S} a_{sl}} \\
&= \frac{\sum_{l=1}^{S} a_{sl}\phi_l}{\sum_{l=1}^{S} a_{sl}} \qquad \text{as } a_{ss} = 0 \\
\mathrm{Var}(\phi_s|\boldsymbol{\phi}_{-s}) &= Q_{ss}^{-1} \\
&= \left((1/\tau^2)\sum_{l=1}^{S} a_{sl}\right)^{-1} \\
&= \frac{\tau^2}{\sum_{l=1}^{S} a_{sl}}.
\end{aligned}$$

Thus, the full conditional distribution of $\phi_s|\boldsymbol{\phi}_{-s}$ is

$$\phi_s|\boldsymbol{\phi}_{-s} \sim \mathrm{N}\left(\frac{\sum_{l=1}^{S} a_{sl}\phi_l}{\sum_{l=1}^{S} a_{sl}}, \frac{\tau^2}{\sum_{l=1}^{S} a_{sl}}\right).$$

Under this model, strong spatial correlation is assumed for the stochastic process of the spatial domain, with the distribution of $\phi_s|\boldsymbol{\phi}_{-s}$ parameterized by a mean that corresponds to a weighted sum of neighbouring areal units within the spatial domain. The full conditional distribution of $\phi_s|\boldsymbol{\phi}_{-s}$ shows that $\mathrm{Var}(\phi_s|\boldsymbol{\phi}_{-s}) = \tau^2/\sum_{l=1}^{S} a_{sl}$, and so allows for the conditional variance to differ from areal unit to areal unit, with areal units that have more neighbours having a small level of uncertainty as they share from a pool of more information. However, if the stochastic process of the spatial domain has no spatial correlation, this model may be inappropriate, as we would want for the random effects associated with each areal unit to be independent of one another.

2.5.2 Leroux CAR

The issues that could possibly arise through modelling a stochastic proccess over a spatial domain using an Intrinsic CAR model can be resolved through the introduction of a spatial autocorrelation parameter ρ with support $\rho \in [0, 1]$ and is estimated from the spatial data. Allowing for such a spatial autocorrelation parameter can be done through the use of the Leroux CAR model, which has an improper multivariate Gaussian distribution given by

$$\boldsymbol{\phi} \sim \mathrm{N}(\mathbf{0}, \tau^2(\rho(\mathrm{diag}(\mathbf{A1}) - \mathbf{A}) + (1-\rho)\mathbf{I})^{-1}), \tag{2.26}$$

with the precision matrix $\mathbf{Q} = (\rho(\mathrm{diag}(\mathbf{A1}) - \mathbf{A}) + (1-\rho)\mathbf{I})/\tau^2$ being invertible if $\rho \in [0, 1)$. Thus, for $\rho \in [0, 1)$ this model specifies a proper multivariate Gaussian distribution. The elements of $\mathbf{Q}$ are $\mathbf{Q}_{ii} = (\rho \sum_{j=1}^{n} a_{ij} + 1 - \rho)/\tau^2$ and $\mathbf{Q}_{ij} = -\rho a_{ij}/\tau^2$. A Bayesian hierarchical model approach that specifies the Leroux CAR model as a prior distribution for a set of random effects is given by

$$\begin{aligned}
Y_s &\sim f(Y_s; \theta_s, \xi) \quad \text{for } s = 1, \ldots, S. \\
g(\theta_s) &= \mathbf{x}_s^\top \boldsymbol{\beta} + \phi_s, \\
\boldsymbol{\beta} &\sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \\
(\phi_1, \ldots, \phi_S) = \boldsymbol{\phi} &\sim \mathrm{N}(\mathbf{0}, \tau^2(\rho(\mathrm{diag}(\mathbf{A1}) - \mathbf{A}) + (1-\rho)\mathbf{I})^{-1}), \\
\tau^2 &\sim \text{Inverse-Gamma}(a_\tau, b_\tau), \\
\rho &\sim \mathrm{Uniform}(0, 1),
\end{aligned} \tag{2.27}$$

where $f(Y_s; \theta_s, \xi)$ is a natural exponential dispersion family of distributions, $g(.)$ is an

invertible link function, y_s is the response in the sth spatial unit. The covariates are included in a $p \times 1$ vector $\mathbf{x}_s$, and a corresponding $p \times 1$ vector of fixed effect parameters are denoted by $\boldsymbol{\beta}$. The $S \times 1$ vector of joint random effects for the sth spatial unit is denoted by $\boldsymbol{\phi} = (\phi_1, ..., \phi_S)_{S\times 1}$ and has a joint Gaussian distribution. τ^2 is a measure of the variance relating to the vector of spatial random effects $\boldsymbol{\phi}$. ρ is the single level of spatial autocorrelation everywhere relating to the vector of spatial random effects, with values close to one and zero representing strong autocorrelation and independence respectively. a_τ and b_τ are user-chosen hyper-parameters for the Inverse-Gamma distribution imposed on τ^2. The full conditional distributions $\phi_s|\boldsymbol{\phi}_{-s}$ are proper. Specifically,

$$\begin{aligned}
\mathbb{E}(\phi_s|\boldsymbol{\phi}_{-s}) &= \mu_s - Q_{ss}^{-1}\mathbf{Q}_{s,-s}(\boldsymbol{\phi}_{-s} - \boldsymbol{\mu}_{-s}) \\
&= 0 - \left(\frac{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho}{\tau^2}\right)^{-1}\left(-\frac{\sum_{l\neq s}\rho a_{sl}(\phi_l - 0)}{\tau^2}\right) \\
&= \frac{\rho\sum_{l\neq s} a_{sl}\phi_l}{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho} \\
&= \frac{\rho\sum_{l=1}^{S} a_{sl}\phi_l}{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho} \qquad \text{as } a_{ss} = 0 \\
\mathrm{Var}(\phi_s|\boldsymbol{\phi}_{-s}) &= Q_{ss}^{-1} \\
&= \left(\frac{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho}{\tau^2}\right)^{-1} \\
&= \frac{\tau^2}{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho}
\end{aligned}$$

Thus, the full conditional distribution of $\phi_s|\boldsymbol{\phi}_{-s}$ is

$$\phi_s|\boldsymbol{\phi}_{-s} \sim \mathrm{N}\left(\frac{\rho\sum_{l=1}^{S} a_{sl}\phi_l}{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{l=1}^{S} a_{sl} + 1 - \rho}\right).$$

Under this model, spatial correlation is estimated for the stochastic process of the spatial domain, $\rho = 0$ corresponding to independence in space and $\rho = 1$ corresponding to strong spatial dependence.

2.5.3 Localised autocorrelation model

So far, only spatial models for global spatial autocorrelation have been presented. However, for certain contexts, it may be more appropriate to use a spatial model which allows for local spatial autocorrelation. Local spatial autocorrelation focuses on deviations from the global trend at much more focused levels than the entire map. Local measures of spatial autocorrelation focus on the relationships between each observation and its surroundings, rather than providing a single summary of these relationships across the map. One way to achieve a local smoothing model is to augment the set of spatially smooth random effects with a piecewise constant intercept or cluster model, thus allowing large jumps in the mean surface between adjacent areal units in different clusters. Lee and Sarran (2015) partitions the S areal units into a maximum of G clusters each with their own intercept term (λ_1, ..., λ_G). The model is given by

$$\begin{aligned}
\phi_s &= \psi_s + \lambda_{Z_s}, \\
\psi_s|\boldsymbol{\psi}_{-s} &\sim \mathrm{N}\left(\frac{\sum_{l=1}^{S} a_{sl}\psi_l}{\sum_{l=1}^{S} a_{sl}}, \frac{\sigma^2}{\sum_{l=1}^{S} a_{sl}}\right), \\
\tau^2 &\sim \text{Inverse-Gamma}(a_\tau, b_\tau), \\
\lambda_i &\sim \text{Uniform}(\lambda_{i-1}, \lambda_{i+1}) \quad \text{for } i = 1, ..., G, \\
f(Z_s) &= \frac{\exp(-\delta)(Z_s - G*)^2}{\sum_{r=1}^{G} \exp(-\delta)(r - G*)^2}, \\
\delta &\sim \text{Uniform}(1, M),
\end{aligned} \tag{2.28}$$

where spatial unit s is assigned to one of the G intercepts by $Z_s \in \{1, ..., G\}$, and G is

the maximum number of different intercept terms. Here we penalise Z_s towards the middle intercept value, so that the extreme intercept classes (e.g. 1 or G) may be empty. This is achieved by the penalty term $\delta(Z_s - G^*)^2$ in the prior for Z_s, where $G^* = (G+1)/2$ if G is odd and $G^* = G/2$ if G is even, and is the middle of the intercept terms.

a_{sl} is element sl of the non-negative spatial adjacency matrix $\mathbf{A}_{S\times S}$, which defines how spatially close the S areal units are to each other. The elements of $\mathbf{A}_{S\times S}$ can be binary or non-binary. In the more common binary case, $a_{sl} = 1$ if a pair of areal units ($\mathcal{S}_s$, $\mathcal{S}_l$) share a common border or are considered neighbours by some other measure, and $a_{sl} = 0$ otherwise. τ^2 is a variance term relating to the vector of spatial random effects. a_τ, b_τ and M are user-chosen hyperparameters.

There are a number of other options for spatial models for local spatial autocorrelation, such as the locally adaptive model (Lee and Mitchell (2013)) and the CAR dissimilarity model (Lee and Mitchell (2012)), which are local spatial autocorrelation models that estimate the elements in the spatial adjacency matrix $\mathbf{A}_{S\times S}$. In this thesis, although local spatial autocorrelation models are becoming increasingly popular, the small number of spatial units present in the data of Chapters 4 and 6 limit their use.

2.6 Discussion

chapter has set out a number of statistical methods that are to be used and or built upon throughout the rest of this thesis. Section 2.1 provided a description of what Bayesian methods are and how they can be used to conduct inference for complex statistical models, which will be used in Chapters 4, 5, 6 and 8 of this thesis. Section 2.2 provided an overview of Bayesian model selection and checking methods, some of which will be used in Chapters 4, 6 and 8. Section 2.3 presented a general outline of generalized linear models along with how different data likelihoods can be used in this framework to model both continuous and

discrete data, which will be used in Chapters 4, 5, 6 and 8 of this thesis. Section 2.4 introduced the concept of MMMC models and the role that they play in modelling network data, which will be used in Chapters 4, 5, 6 and 8 of this thesis. Section 2.5 concluded with a presentation of spatial conditional autoregressive models, and how such models can be used on areal processes, which will be used in Chapters 6 and 8 of this thesis.

Chapter 3

Health behaviours and network studies

3.1 Health Behaviours

The survey studies that motivate the theoretical developments described in this thesis are based on data relating to the self-reported negative health behaviours of adolescents. The analyses presented in later chapters are based on data gathered from Los Angeles and the United Kingdom. In order to answer key research questions, interest lies in modelling these negative health behaviours, which are comprised of tobacco, alcohol, and marijuana use. In the wider literature, these negative health behaviours have been treated in both a univariate and multivariate fashion.

3.1.1 Single outcomes

In the literature on negative health behaviours outcomes can be treated singularly, in a univariate manner. This practice is driven by several factors, such as the data available to the researcher(s), the limitation of modelling frameworks available at the time and the severity of the consequences linked with the negative health behaviour at the time.

The epidemiological literature on the negative effects of tobacco has an incredibly long history, with major reports being published in the 1930's, 40's, and 50's. Müller (1939) was the first epidemiological study that sought to establish a link between smoking and lung cancer in 86 patients. This research was further developed in the 40's by other German researchers (Schairer and Schöniger, 1943) in a more ambitious study, leading to a wide array of studies in the 50's. Doll and Hill (1950) found that for those above the age of 45, the risk of developing the disease increases in simple proportion with the amount smoked, and that it may be approximately 50 times as great among those who smoke 25 or more cigarettes a day as among non-smokers. Four years later, Doll and Hill (1954) provided further evidence to support these results. Wynder et al. (1953) sought to establish the relationship using animal experimentation, showing that applying cigarette tar onto mice generated tumours. Future studies such as these contributed greatly to the literature on the harms of smoking, leading to the univariate modelling of tobacco use to better understand the underlying processes which drive individuals to smoke in the hopes of intervening and preventing its uptake. Chassin et al. (1984) is an early example of a longitudinal study which models the univariate outcome of cigarette smoking. The results of the study suggested that smoking prevention schemes might be more effective if targeted at high risk groups rather than at general adolescent populations. In the literature on the single outcome of cigarette smoking, there are numerous examples in which the outcome is treated as binary. Alexander et al. (2001) use a binary outcome for their smoking measure, 1 for students who reported smoking cigarettes on one or more days in the past 30 days and 0 otherwise. The study was based on 2525 adolescents in Grades 7-12 and found that, in comparison to those whose best friend didn't smoke, those with best friends who smoked had greater observed odds of having smoked in the past 30 days (OR = 2.01, 95% CI: 1.66-2.42). Elder et al. (2000) initially had a categorical outcome with five categories, but decided to convert the outcome to binary, 1 for those that reported having ever smoked in the past 30 days and 0 otherwise. This outcome was dichotomized due to most of the respondents (74%) reporting that they had never smoked. The study

involved 600 Hispanic adolescents between the ages of 11-16 years and found that every unit increase in communication with parents, a mean score based on six questions, lowered the odds of smoking (OR = 0.52, 95% CI: 0.31-0.87). Similarly, Valente et al. (2013) and Huang et al. (2014) chose to dichotomize their 5 level categorical outcome due to 69.6% and 71.2% of respondents reporting never having smoked, respectively. In contrast to the previous literature, Ling et al. (2019)) gain their binary outcome by asking *"(1) During the past 30 days, on how many days did you smoke cigarettes (Choice of responses: 0 days, 1-30 days)? (2) During the past 30 days, did you use traditional-hand-rolled cigarette (Choice of responses: Yes or No)? (3) During the past 30 days, did you use roll-your-own cigarette paper (Choice of responses: Yes or No)? Respondents who answered "1 to 30 days" for item (1), and/or "Yes" to either item (2) or (3) were classified as current cigarette smokers"*. Recently, the literature on univariate smoking related outcomes has expanded and embraces advancements in the field of machine learning to predict outcomes (see for example Nam et al. (2019)).

There is a strong consensus in the scientific community that alcohol consumption can cause various forms of cancer, with the National Toxicology Program of the US Department of Health and Human Services listing the consumption of alcohol as a known human carcinogen based on sufficient evidence from studies in humans (NTP, 2016). The epidemiological literature has provided sufficient evidence for alcohol consumption causing oral cavity, pharynx, larynx, oesophagus, liver, and colorectum cancer (Secretan et al., 2009). In a case-control study of Indian men, Znaor et al. (2003) found that after adjusting for covariates, the odds of oral cancer for those that drank alcohol was 98% (95% CI: 68%–133%) times higher than the corresponding odds for those who did not drink. The study also found that the odds of pharynx cancer for those that drank alcohol was 107% (95% CI: 67%–156%) times higher than the corresponding odds for those who did not drink. In another case-control study, Talamini et al. (2002) define one drink as approximately 125 ml of wine/330 ml of beer/30 ml of liquor, finding that, in comparison to abstainers, the risk of laryngeal cancers was greater

for those that who consumed 22-56 drinks per week (OR = 2.6, 95% CI: 1.4–4.7) and more than 56 drinks per week (OR = 5.9, 95% CI: 3.1–11.3). In a meta-analyis of 28 studies on oesophagus cancer, 20 studies on liver cancer, and 22 studies on colorectum cancer conducted by Bagnardi et al. (2001), significantly increased risks were found for ethanol intake of 100 g per day, RR = 4.23 (95% CI: 3.91-4.59), RR = 1.86 (95% CI: 1.53-2.27), and RR = 1.38 (95% CI: 1.29-1.49) respectively. As a result of the negative consequences of alcohol consumption, much of the literature concerning the univariate modelling of this outcome seeks to better understand what makes people drink. In the literature on the single outcome of alcohol consumption, there are examples in which the outcome is treated as binary and categorical. Fujimoto and Valente (2015) use a binary alcohol outcome, 1 for those who have ever drank alcohol and 0 otherwise. The study was based on 1707 adolescents belonging to 5 schools in Los Angeles, California. The study found that for every unit increase in the indegree based on popularity, i.e. summing the total number of popularity nominations received, increased the observed odds of drinking (OR = 1.35, which was statistically significant at the 0.1% level). Jacobs et al. (2016a) initially had a categorical outcome with seven categories, but decided to convert the outcome to binary, 1 for those that reported having ever consumed alcohol in the past 30 days and 0 otherwise. This outcome was dichotomized due to most of the respondents (64.5%) reporting that they had not had at least one drink of alcohol in the past 30 days. The study was based on 1523 high school students in Los Angeles, California. The authors found that, in comparison to those who had no parents that consumed alcohol, those with both parents that consumed alcohol had a greater observed odds of drinking (OR = 2.88, 95% CI: 1.82-4.55). Similarly, Jacobs et al. (2016b) and Jacobs et al. (2017) chose to dichotomize their categorical outcome, 1 for those that reported having ever consumed alcohol in the past 30 days and 0 otherwise. Jacobs et al. (2016b) was based on 1523 10th Grade students in Los Angeles, California. The authors found that, in comparison to those who achieved mostly A's and B's last year, those who achieved mostly D's and F's had a greater observed odds of drinking (OR = 4.496, 95% CI: 2.086-9.689). In contrast, Fujimoto et al.

(2013) uses a five level categorical alcohol outcome to understand what drives adolescents to use alcohol and be susceptible to using it.

In the epidemiological literature, one of the many adverse health effects that have been studied in relation to marijuana is schizophrenia, a psychotic illness that causes a variety of different psychological symptoms, such as delusions and hallucinations. Zammit et al. (2002) use data from a 1969 historical cohort study over 50,000 subjects in Sweden to establish a causal relation between cannabis use and developing schizophrenia. The analysis found that, in comparison to those that didn't use drugs, those that used cannabis over 50 times and no other drug were more likely to develop schizophrenia (OR = 6.7, 95% CI: 2.1-21.7). Fergusson et al. (2003) show that, in comparison to those without a cannabis dependence, those with the dependence had increased psychotic symptoms (RR = 1.8, 95% CI: 1.2-2.6). The epidemiological literature on the negative effects of marijuana is a very active field of research. The modelling of the univariate use of marijuana is also of great importance to understand the driving factors that may increase the likelihood of one consuming marijuana. In the literature on the single outcome of marijuana consumption, there are examples in which the outcome is treated as continuous and binary. In one of the earlier instance of the literature, Dishion and Loeber (1985) uses a continuous response, which was derived by asking subjects to recall the exact number of times they used marijuana over the past year. The authors also opted to take the log of this continuous outcome to better meet the assumption of the normal linear regression model. In contrast, Brook et al. (2001) use a binary outcome in their study, a 0 indicates no marijuana use during the designated time period, and a 1 indicated marijuana use. Schepis et al. (2011) extend this by using binary outcomes for lifetime marijuana use and past 30-day marijuana use. The binary outcome of past 30-day marijuana use was also utilized in Choo et al. (2014). Wilson et al. (2005) and Lorant and Tranmer (2019) also shares the characteristic of using a binary outcome for marijuana use.

3.1.2 Multiple outcomes

The epidemiological literature on the negative effects of consuming alcohol, tobacco, and marijuana also includes the consideration of their joint impact on health. Talamini et al. (2002) show that, rather than being additive, the impact of both heavy cigarette smoking and alcohol drinking leads to more of a multiplicative risk increase for laryngeal cancer, when compared to those who don't smoke or drink (OR = 177.2, 95% CI: 64.99-483.28). The authors hypothesised that this may be due to some biological synergy. In a work published 7 years later, Hashibe et al. (2009) sought to examine the multiplicative joint effect of ever using tobacco and alcohol on head and neck cancer risk. The authors use a multiplicative interaction parameter ψ given by

$$\psi = \frac{\text{OR}_{11}}{\text{OR}_{01} \times \text{OR}_{10}},$$

where OR_{11} is the OR for ever tobacco/ever alcohol use, OR_{01} is the OR for never tobacco/ever alcohol use, and OR_{10} is the OR for ever tobacco/never alcohol use. $\psi > 1$ implies a joint effect that is greater than expected under the multiplicative model. The authors showed that there was a greater than multiplicative joint effect between ever using tobacco and alcohol for observed head and neck cancer ($\psi = 2.15$, 95% CI: 1.53-3.04). In addition to personal risk of diseases, the literature on injury epidemiology contains instances in which negative health behaviours are jointly used as risk factors for fatalities. Chihuri et al. (2017) show that for fatal crash involvement, in comparison to drivers testing negative for alcohol and marijuana, $\text{OR}_{01} = 16.33$ (95% CI: 14.23-18.75) for those testing positive for alcohol and negative for marijuana, $\text{OR}_{10} = 1.54$ (95% CI: 1.16-2.03) for those testing positive for marijuana and negative for alcohol, and $\text{OR}_{11} = 25.09$ (95% CI: 17.97-35.03) for those testing positive for both alcohol and marijuana. This suggested that, when alcohol and marijuana are used together, the observed fatal crash risk is not on the multiplicative

scale. In the literature on negative health behaviours, the need to consider outcomes in a multivariate manner is of increasing importance. This is due to a number of reasons, such as the fact that these outcomes are typically correlated. Lesaffre and Molenberghs (1991) take a bivariate outcome made up of cigarette smoking and drinking to illustrate the value of a multivariate modelling approach over modelling a pair of correlated responses separately. Azagba and Sharaf (2014) make use of a bivariate outcome constructed from two binary outcomes, binge drinking and marijuana use. Binge drinking was defined as having five or more drinks on one occasion at least once a month in the past year and marijuana use was defined by students who reported using marijuana in the last 12 months. The study was based on 4466 Canadian students in Grades 7-12. The authors found that menthol cigarette smokers are 6% (marginal effect (ME) = 0.06, 95% CI: 0.03–0.09) more likely to binge drink and 7% (ME = 0.07, 95% CI: 0.05–0.10) more likely to use marijuana. In the same model, they also found that class skippers are 14% (ME = 0.14, 95% CI: 0.08–0.21) more likely to binge drink and 10% (ME = 0.10, 95% CI: 0.04–0.17) more likely to use marijuana. As marijuana is classified as an illegal drug in many places, there are examples in the literature on multivariate negative health behaviours in which it is coupled with other illegal drugs, such as heroin and cocaine to form a trivariate outcome (see for example Ramful and Zhao (2009)).

The literature on single and multiple health behaviours is a vast and growing field of research. There are a number of approaches that have been used to model the uptake of single and multiple health behaviours. These approaches generally fall under one of two categories, network-based approaches and non-network-based approaches.

3.2 Networks

The literature on network analysis spans a number of related disciplines, such as sociology, psychology, and statistics. Some of the earliest and most prolific works in network literature

relate to the works carried out by Milgram (1967) and Travers and Milgram (1969) in the mid-20th century. A pioneering experimental study relating to the small worlds of social networks, the hypothesis that any two people in the world can be linked through a small number of intermediary acquaintances, was first conducted by Milgram (1967). Milgram set out a framework for testing this hypothesis, which was based on chains of letter forwarding. Travers and Milgram (1969) presented the first technical report on the method, in which 296 volunteers were recruited to start a chain of letter forwarding; 100 were solicited from a Boston newspaper advertisement and 196 were Nebraska residents solicited by mail, with 100 of the 196 specifically chosen to be blue-chip stock holders. The target for the chains of letters was unknown to the participants, a stock broker living in Massachusetts. The study found that, of the 64 completed chains, each chain passed through a mean of 5.2 intermediaries before reaching the designated target. Although the study had a number of limitations, such as recruiting blue-chip stock holders to start some of the chains of letters to a target stock broker, which may bias estimates relating to the chain length, the results were profound and illustrated how small the degrees of separation between two people in a network can be. Network literature of this type has since been expanded with much larger experiments that use the internet. Dodds et al. (2003) modernised Milgram's approach by conducting an email based study that focused on 18 targets from 13 different countries. 98,847 individuals were registered from across the world, with approximately 24,163 providing their personal information and initiating message chains. The study found that, of the 384 completed chains, each chain passed through a mean of 4.05 intermediaries before reaching the designated target. This study has limitations, such as the hypothesis that longer chains are less likely to be completed and included in the results, leading to the underestimation of quantities relating to chain length, but gives an idea of how small the degrees of separation in a network can be.

The network literature that illustrates and supports the small worlds notion of social networks is greatly complemented by the literature on how peers in networks can influence the

negative behaviour of an actor in the network. There are numerous theories in the network literature that seek to explain the role that social networks play in an individual consuming alcohol, tobacco, and marijuana. Social learning theory is based on the concept that individuals learn new behaviours by observing and experiencing the behaviours of others and mimicking them. The theory has been put forward as a theory to unify how influence can permeate through a network. This unifying theory has attempted to have its validity assessed (See Akers et al. (1979)). The social development model formulated by Hawkins and Weis (1985) is grounded in criminology and theorizes that behaviour is sequentially influenced by socialization, family, schools, peers, and community. This theoretical explanation for how negative behaviours can be transmitted through a network has been investigated in the network literature (See Catalano et al. (1996)). Oetting and Beauvais (1986) put forward a peer cluster theory for drug use, which states that small, identifiable peer clusters determine where, when, and how drugs are used and that these clusters specifically help shape attitudes and beliefs about drugs. The theory also considers the importance of psychosocial characteristics that form the basis of drug use and the environment for peer clusters to work. There are instances in the literature that seek to test the peer cluster theory (See Rose (1999); Kim et al. (2002)). The psychology and sociology network literature illustrates how the coupling of social network structures and influence can work in tandem to increase the risk of an adolescent engaging in negative health behaviours.

Statistical network literature has a broad scope, experiencing major developments alongside the developments in the psychology and sociology network literature during the 1970s and 1980s. The statistical network literature relating to the consumption of alcohol, tobacco, and marijuana varies in the approaches taken, ignoring the network entirely, utilizing descriptive network statistics and explicitly including the membership matrix of the network.

3.2.1 Non-network approaches

In a study of 171 boys and 274 girls aged between 14-18 years old that had lived in Palma De Mallorca for at least 2 years, Tur et al. (2003) model the alcohol consumption of adolescents in an approach which did not contain peer network information, but instead focused on parental information. The study found that, in comparison to boys with mothers that had an educational level of no studies, boys with mothers that had a university educational level had a lower observed odds of alcohol consumption (OR = 0.53, 95% CI: 0.40 - 0.69). In a larger study of 9920 middle and high school students from Ontario, with a mean age of 15.1 (SD = 1.8), Sampasa-Kanyinga et al. (2018) investigates the use of cannabis in Canadian schools. The study is inspired, in part, by the government of Canada's commitment to legalize the use, possession, purchase, and growth of recreational cannabis around the time of the research. The modelling done in the study did not contain any network information and instead placed a greater focus on the cigarette and alcohol use of adolescents to explain marijuana use over the past 12 months. The study found that, in comparison to those who did not use tobacco cigarettes over the past 12 months, those that did use tobacco cigarettes had a higher observed odds of having used marijuana over the past 12 months (AOR = 10.10, 95% CI: 8.68 - 13.92). Also, in comparison to those who did not use or just sipped alcohol over the past 12 months, those who reported drinking occasionally had a higher observed odds of having used marijuana over the past 12 months (AOR = 5.35, 95% CI: 4.01 - 7.13). A limitation of this study was the self-reported nature of cannabis use, which was considered an illegal substance during the time that the data was collected in 2015. In a smaller study of 543 males and 532 females with a mean age of 16.2 (SD = 0.47) from Kamianske,Ukraine, Hryhorczuk et al. (2019) model having ever used alcohol, having used alcohol in the past 12 months, and having used alcohol in the past 30 days. The models did not contain network information, but instead focused on covariates relating to leisure activities engaged in by the individual, such as watching TV, socializing with friends, reading

books, and using social media. In the sample of 1075 adolescents 886, 152, and 3 reported socializing with friends frequently, sometimes, and never respectively, 34 had missing values. The study found that, in comparison to those that did not use social media, those that did use social media had higher observed odds of having ever used alcohol (OR = 2.11, 95% CI: 1.40 - 3.19), having used alcohol in the past 12 months (OR = 2.60, 95% CI: 1.73 - 3.90), and having used alcohol in the past 30 days (OR = 2.35, 95% CI: 1.50 - 3.70). In comparison, a larger study of 1,400 males and 1,039 females from El Salvador and 1,439 males and 1,762 females from Peru aged between 13-18 years old, Prieto-Damm et al. (2019) model alcohol consumption using leisure activities, and not placing a great emphasis on social network information. The study, however, does include how often an individual hangs out with friends. Similarly to Hryhorczuk et al. (2019), covariates such as watching TV and reading books are included in the modelling process. The alcohol consumption response was generated by asking participants *"how frequently they consumed alcoholic beverages"*, with the options being *"never"*, *"almost never"*, *"sometimes"*, *"almost always"*, and *"always"*, those that have never consumed an alcoholic beverage were considered to be non-alcohol consuming and the rest were considered alcohol consuming. The study found that, in comparison to those who hung out with friends less than 1 day per month, those who hung out with friends more than 1 day per month had higher observed odds of alcohol consumption (AOR = 1.28, 95% CI: 1.09 - 1.49). In addition, in comparison to those who read books less than 1 day per month, those who read books more than 1 day per month had lower observed odds of alcohol consumption (AOR = 0.82, 95% CI: 0.70 - 0.95). A potential limitation of this study was the self-reported nature of the data.

3.2.2 Descriptive network statistics

In a manner that coincides with social learning theory, there are instances in the network literature which focus on the behaviours of close/best friends as social network information

to explain how an adolescent may engage in negative health behaviours. In this section, the focus is on how descriptive network statistics can be used to model health behaviours. In a study of 6,900 adolescents aged between 14-18 years old from across the United States of America, Wang et al. (1995) model smoking status of adolescents using summaries of peer network information. Each adolescent nominated four best male friends and four best female friends. The dependent variable was dichotomized to classify adolescents as current smokers and those that never smoked. The study found that for the observed males and females, the same-gender best friends' smoking status had the greatest effect on each group's odds of being a current smoker. In comparison to 18 year old males with 0 best male friends that smoke, 18 year old males with 1 - 2 best male friends that smoke had a higher observed odds of current regular smoking (OR = 7.3), which was significant at the 5% level. Similarly, in comparison to 18 year old females with 0 best male friends that smoke, 18 year old males with 1-2 best female friends that smoke had a higher observed odds of current regular smoking (OR = 5.7), which was significant at the 5% level. A limitation of the study is that the results presented did not include 95% confidence intervals, but instead just reported whether or not an effect was significant at the 5% level, which restricts the amount of inference that a reader can make about a parameter. In a study of 1411 Latina individuals aged between 14-24 years old who were clients at two federally funded family planning clinics in the United States of America, Kaplan et al. (2001) investigate the ways in which socioeconomic factors and network attributes can affect young Latinas propensity to smoke. The dependent variable in the model was whether a respondent had ever tried a cigarette (triers/regular smokers) or not (never smokers). Exposure of smoking behaviours of peers in the network was defined as the total number of friends, brothers, sister, and/or significant other who smoke. The study found that, in comparison to those with no peers who currently smoke, those with two or more peers who currently smoke had a higher observed odds of being a trier/regular smoker (OR = 2.49, 95% CI: 1.93 - 3.20). In comparison to those with no parents who currently smoke, those with at least one parent who currently smoke had a higher observed odds of

being a trier/regular smoker (OR = 1.65, 95% CI: 1.32 - 2.07). A major limitation of this study was the fact that the data used were gathered from a sample of clinics, which may lead to biased results and inferences which are in no way generalizable to the Latina population as a whole. In a larger sample of 3146 males and 3128 females in grades 8 to 10 with a mean age of 14.7 years old ($SD = 0.98$) from 14 rural communities in Minnesota, Eisenberg and Forster (2003) investigate how social norms, personal characteristics and family/peer smoking habits can affect their propensity to smoke. The study found that having friends that smoked was significantly related to having smoked in the past month (OR = 3.05, 95% CI: 2.80 - 3.14) and that the perceived adult disapproval of teen smoking was also significantly related to having smoked in the past month (OR = 0.85, 95% CI: 0.73 - 0.99). However, due to the fact that the data were gathered from small rural communities, these results may not easily be generalizable to populations of adolescents who live in urban areas. Kuntsche and Jordan (2006) study a cross-sectional sample of 3925 eighth and ninth grade students in Switzerland with a mean age of 15.2 years old ($SD = 0.9$). The study showed that having substance using peers increased the risk of both drunkenness and cannabis use. In a larger study, Murnaghan et al. (2008) investigate a sample of 4709 students from Prince Edward Island, Canada, investigate the school-based smoking policies and descriptive statistics of friendship networks to model smoking behaviours of adolescents. The dependent variable was whether an individual was a current non-smoker or occasional smoker. Current non-smokers were defined as having never smoked or only tried a cigarette once. Occasional smokers were defined as those who smoked less than weekly. The study found that, in comparison to those with no close friends, those with 1 close friend that smokes had a higher observed odds of being an occasional smoker (OR = 1.40, 95% CI: 1.14 - 1.70). Also, in comparison to those belonging to schools with no smoking programs or policies, those belonging to schools with a smoking program had a lower observed odds of being an occasional smoker (OR = 0.57, 95% CI: 0.44 - 0.75). However, a limitation of the study was the self-reported nature of the study. Students reported on their friends' smoking status, which may have generated

incorrect data. In a study of 1065 males and 908 females aged between 11-19 years old from 25 senior high schools in Ghana, Oppong Asante and Kugbey (2019) model alcohol use, lifetime drunkenness, and problem drinking. Alcohol use is defined as having drank in the past 30 days. Lifetime drunkenness was defined as having been drunk at least one time in the past. Problem drinking was defined as having at least being involved in one incident as a result of drinking. The study included crude binary network information, namely, the amount of close friends an individual has, 0 for zero friends and 1 for one or more close friends. The study found that, in comparison to those with no close friends, those with at least one close friend had a lower observed adjusted odds of current alcohol use only (AOR = 0.52, 95% CI: 0.28 - 0.96). At the 5% level, the binary network variable was not significant for lifetime drunkenness or problem drinking. A limitations of this study was the low prevalence of alcohol use (12.6%), lifetime drunkenness (11.1%), and problem drinking (6.8%) in the data. Including network information in such a manner may also be considered another limitation.

In comparison to network literature that focus on using the behaviours of close/best friends as social network information, some studies instead use measures of popularity, such as indegree, and other network statistics to explain how an adolescent may engage in negative health behaviours. In a national sample of 2525 adolescents from the United States of America with a mean age of 15.47 (SD = 1.5), Alexander et al. (2001) model cigarette smoking, placing a very strong emphasis on using descriptive statistics of the network as covariates. Those who reported smoking cigarettes on at least one of the past 30 days were considered current smokers and those that did not were considered non-current smokers. A proxy covariate entitled popularity was created by calculating the indegree of an individual in the friendship network. The study found that the popularity covariate was not a significant predictor of cigarette smoking (OR = 1.02, 95% CI: 0.90 - 1.16). However, the risk of cigarette smoking was increased with peer networks in which at least half smoked (OR = 1.91, 95% CI: 1.32 - 2.78). In a smaller sample of 1,486 sixth and seventh graders from 16

middle schools in southern California, Valente et al. (2005) seek to investigate whether or not popular adolescents in a network are more likely to smoke. The dependent variable, having ever smoked, was a self-reported answer to whether an individual had ever taken a puff, smoked a whole cigarette, or smoked any cigarette in the last 30 days at a 1-year follow-up, coded as no = 0 and yes = 1. The popularity covariate was generated by asking adolescents to name their 5 closest friends in the class and taking popularity to be the number of times a person was named as a friend divided by class size. The study found that in comparison to those who received no friend nominations, those that received friend nominations from everyone in the class would be 5.09 times more likely to have ever smoked. The popularity effect was significant at the 5% level. A potential limitation of the study was that the popularity covariate was generated by friendship nominations in a single classroom. An adolescent may belong to a class which they are not very popular in but in the school as a whole they might be very popular. Thus the way that the network metric was generated could potentially overestimate or underestimate a persons true popularity. In addition, it is possible that instead of popularity increasing the risk of someone smoking, adolescents may smoke to gain or maintain popularity. In a larger study of 2610 7th - 11th grade students aged between 12-19 years old from the United States of America, Mundt (2011) seek to investigate the impact of peer social networks on adolescent alcohol use. The study focused on four descriptive network statistics; indegree, centrality, 3-step reach, and density. Indegree was defined as the number of friendship nominations received by an adolescent. Centrality was the relative number of connections that an individual's friend has within the friendship network. 3-step reach is the degree to which a member of the peer social network can make contact with other members of the network through 3 steps of friendship connections. Density is the number of edges in the total school peer network divided by the number of possible network edges. The study found that for every additional 10 friends within 3-step reach of a nominated friend, risk of alcohol initiation increased by 3% (95% CI: 0.3% - 6%).

3.2.3 Network random effects

In contrast to network literature that uses descriptive statistics of networks to explain why an adolescent may engage in negative health behaviours, the matrix of network information may also be used to incorporate random effects within the modelling process. In a sample of 11,015 students with a mean age of 15.2 ($SD = 1.0$) from 50 schools across six European cities, Lorant and Tranmer (2019) seek to investigate how peer networks affect the risk of an individual using cigarettes, alcohol and marijuana. An adolescent was classified as a daily smoker if they reported having smoked at least one cigarette a day in the last 30 days. Adolescents that reported having consumed alcohol at least once a month in the past year were classified as alcohol users. An adolescent was classified as a monthly cannabis user if they reported having used cannabis at least once a month in the past year. The study combined the use of descriptive network statistics, such as indegree, closeness, and betweenness, with the incorporation of peers as random effect within the modelling process. The study found that, all else being held fixed, every unit increase in indegree increased the observed odds of daily smoking (OR = 1.06, 95% CI: 1.02 - 1.11) and monthly alcohol use (OR = 1.06, 95% CI: 1.03 - 1.09). The study also found that, after controlling for the explanatory variables, the unexplained variation for the network random effect for daily smoking, monthly alcohol use, and monthly cannabis use were $\hat{\sigma}^2 = 14.08$ (95% CI: 3.37-17.19), $\hat{\sigma}^2 = 9.16$ (95% CI: 0.06-11.60), and $\hat{\sigma}^2 = 12.31$ (95% CI: 4.94-16.18) respectively. A potential limitation of the study was that it treated daily cigarette use, monthly alcohol use, and monthly marijuana use in a univariate manner instead of a trivariate manner.

Chapter 4

Univariate and multivariate MMMC models

4.1 Introduction

The multivariate regression is a technique that takes into account potential dependence between more than one response variable. In contrast, performing univariate regressions on each of the response variables separately does not take into account any potential dependence. The literature on multiple health behaviours suggests that there may be a dependence between an individual partaking in alcohol, cigarette, and marijuana use. Thus, a multivariate regression approach may be more appropriate than the use of multiple univariate regressions on such data. The dependence between alcohol and tobacco use has been studied and observed in the population of adolescents (Koopmans et al. (1997); Jackson et al. (2002)) and college students (Hines et al. (1998); Saules et al. (2004); McKee et al. (2004)). Likewise, the dependence between alcohol and marijuana use has been studied and observed in the United States population (Wechsler et al. (1995); Clapp and Shillington (2001); Novak et al. (2016)) and parts of Europe (DiGrande et al. (2000)). The relationship between

tobacco and marijuana use has also been studied and observed in the United States (Rigotti et al. (2000); Ford et al. (2002)). There are a number of theoretical frameworks that have been put forward to explain why such observations may occur. Steele and Josephs (1990) explore the idea of *alcohol myopia*, which is based on the notion that alcohol produces a myopic effect that causes a user's behaviour to be overly influenced by salient environmental cues and limit their ability to consider future consequences. Thus salient environmental cues that promote smoking such as other people smoking and the smell of cigarette smoke may influence an individual to smoke. Kandel (2002) details the *gateway hypothesis* which builds on the work of Kandel (1975) and suggests that alcohol and tobacco use may serve as a gateway substance to cannabis use.

In this chapter I model multiple health behaviours of adolescents with the use of a univariate MMMC model, whose framework is described in Browne et al. (2001), and a multivariate MMMC model proposed here, comparing and contrasting each approach. The multivariate MMMC model extends the univariate MMMC model, which is a form of generalised linear mixed model, by allowing both the sets of random effects for the single membership and multiple membership classifications to be jointly modelled across responses. The covariance matrix assigned to the random effects can take on many structures that place various levels of restrictiveness on the model. In contrast to restrictive structures, such as compound symmetry and Toeplitz, an unstructured covariance matrix is a completely general covariance matrix and is the structure used in all applicable models in this chapter. The only known instance of an MMMC model being applied to health network data is provided by Lorant and Tranmer (2019), who model multiple health behaviours in a univariate way. These models have also been applied to other research areas such as academic performance Tranmer et al. (2014) and organizational performance Tranmer et al. (2016)

The remainder of the chapter is structured as follows. Section 4.1 sets out the motivation for this chapter and outlines the novel contributions. Section 4.2 provides a description and exploration of the network data used, and acts as a form of guided inspiration for the

multivariate MMMC model proposed in this chapter. Section 4.3 describes the multivariate MMMC model used in this chapter. Section 4.4 details the Markov chain Monte Carlo algorithm used for the aforementioned models in this chapter. Section 4.5 presents an application of the models discussed in this chapter on network data of multiple health behaviours of adolescents in Los Angeles, California. Section 4.6 concludes with a discussion.

4.1.1 Motivation

In the literature concerning network studies, one of the most common approaches is to employ a generalised linear model. There are a number of reasons why this is the case. Three of the most prominent reasons are the facts that generalised linear models are typically easy to use, understand, and implement. It is common for network studies that are concerned with binary outcomes to use a generalised linear model with a Bernoulli likelihood, and use the network information in the data by way of computing summary statistics of the network for an observation and feeding it into the model as a covariate. The general form of the model used in such studies is given as follows, where prior distributions are given for the parameters as inference will be performed in a Bayesian setting.

$$\begin{aligned} y_{ik} &\sim \text{Bernoulli}(\pi_{ik}) \quad i = 1, ..., N_k, \quad k = 1, ..., K, \\ \text{logit}(\pi_{ik}) &= \mathbf{x}_{ik}^{\top}\boldsymbol{\beta}, \\ \boldsymbol{\beta} &\sim \text{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}). \end{aligned} \tag{4.1}$$

The vector of a binary health behaviour for all adolescents in all K schools studied is denoted by $\mathbf{y} = (y_{11}, \ldots, y_{N_1 1}, \ldots, y_{1K}, \ldots, y_{N_K K})_{\sum_{k=1}^{K} N_k \times 1}$. The covariates for the ith adolescent in the kth school are included in a $p \times 1$ vector $\mathbf{x}_{ik}$, and a corresponding $p \times 1$ vector of fixed effect parameters are denoted by $\boldsymbol{\beta}$. Model (4.1) describes the Bayesian implemen-

tation of the Bernoulli generalised linear model. However, the prior on $\boldsymbol{\beta}$ can be removed, and a likelihood based approach to parameter estimation can be adopted, such as Fisher's scoring method. Fujimoto and Valente (2015) employed a likelihood based approach to network data of 1707 adolescents in Los Angeles, California with model (4.1). The researchers made use of a binary alcohol outcome and used two networks to construct covariates. One of the networks used was a friendship network for which individuals were asked to nominate friends resulting in a network. The other network used was a popularity network for which individuals were asked to nominate other individuals who they perceived to be popular resulting in a network. The two network covariates relate to the number of times an individual was nominated in each network. For example, in the popularity network, if an individual was nominated 5 times, they would have a value of 5 for the corresponding network covariate. The researchers fitted 3 different models to the data and found the friendship network covariate to be statistically significant at the 10% level and the popularity network covariate to be statistically significant at the 0.1% level across all three models. These results help to motivate the need to model the networks present in data.

MMMC models have been sparingly used in the literature on network studies, in part due to the large number of random effects it can create and the response being binary which can impact model fitting, with Lorant and Tranmer (2019) being the only known network study to apply it to several health outcomes individually. The work made use of the univariate MMMC model to estimate the relative share of variation in binary health behaviours of adolescents at the school and friendship network level after accounting for the fixed effects of covariates. The general form of the model used in the study is given as follows

$$
\begin{aligned}
y_{ik} &\sim \text{Bernoulli}(\pi_{ik}) \quad i = 1, ..., N_k, \quad k = 1, ..., K, \\
\text{logit}(\pi_{ik}) &= \mathbf{x}_{ik}^{\top}\boldsymbol{\beta} + v_k + \sum_{j \in \text{net}(ik)} w_{ikj} u_j, \\
\boldsymbol{\beta} &\sim \text{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \\
v_k &\sim \text{N}(0, \sigma_v^2) \quad k = 1, ..., K, \\
u_j &\sim \text{N}(0, \sigma_u^2) \quad j = 1, ..., J, \\
\sigma_v^2 &\sim \text{Inverse-Gamma}(\alpha_v, \xi_v), \\
\sigma_u^2 &\sim \text{Inverse-Gamma}(\alpha_u, \xi_u). \qquad (4.2)
\end{aligned}
$$

The vector of a binary health behaviour for all adolescents in all K schools studied is again denoted by $\mathbf{y} = (y_{11}, \ldots, y_{N_1 1}, \ldots, y_{1K}, \ldots, y_{N_K K})_{\sum_{k=1}^{K} N_k \times 1}$. The covariates for the ith adolescent in the kth school are included in a $p \times 1$ vector $\mathbf{x}_{ik}$, and a corresponding $p \times 1$ vector of fixed effect parameters are denoted by $\boldsymbol{\beta}$. The $K \times 1$ vector of random effects for the single membership school classifications are denoted by $\mathbf{v} = (v_1, \ldots, v_K)_{K \times 1}$. There are J alters, where an alter is a person who was nominated as a friend by an individual. The $J \times 1$ vector of alter random effects are denoted by $\mathbf{u} = (u_1, \ldots, u_J)_{J \times 1}$. net$(ik)$ is the set of alters that the ith adolescent in the kth school nominates as a friend such that net$(ik) \subset \{1, \ldots, J\}$. The weight of the jth multiple membership random effect for the ith adolescent in the kth school is given as follows

$$
w_{ikj} = \frac{1}{|\text{net}(ik)|}, \qquad (4.3)
$$

with the restriction that $|\text{net}(ik)| \neq 0$, thus $\sum_{j=1}^{J} w_{ikj} = 1$.

There are a number of reasons as to why this type of model is not widely used on network

data: firstly, the sum of alter random effects in Model (4.2) can be very costly to compute when implementing a Markov chain Monte Carlo algorithm, particularly when J and the amount of times each u_j appears in a likelihood is large. As the amount of times u_j appears in a likelihood gets large, the more likelihoods need to be evaluated to compute the acceptance probability for u_j. Specifically, the acceptance probability of a move from the current value $u_j^{(t)}$ to to the proposal value u_j^* is given as follows

$$\min\left\{1, \frac{\prod_{ik \ s.t. \ j\in \text{net}(ik)} \text{Bernoulli}(y_{ik}|\boldsymbol{\beta}^{(t)}, v_k^{(t)}, u_j^*, \mathbf{u}_{-j}^{(t)})\text{N}(u_j^*|\sigma_u^{2(t)})}{\prod_{ik \ s.t. \ j\in \text{net}(ik)} \text{Bernoulli}(y_{ik}|\boldsymbol{\beta}^{(t)}, v_k^{(t)}, u_j^{(t)}, \mathbf{u}_{-j}^{(t)})\text{N}(u_j^{(t)}|\sigma_u^{2(t)})}\right\},$$

thus every unit increase in the amount of times u_j appears in a likelihood results in an extra two likelihoods needing to be evaluated. Thus suppose that $J = 10,000$ and the user requires the sampler to run for 1,000,000 iterations, every unit increase in the amount of times all u_js appear in a likelihood would result in an additional $2 \times 10,000 \times 1,000,000 =$ 20 billion likelihoods needing to be evaluated. Secondly, in comparison to network data with a continuous or count responses, through experience, estimating the relative share of variation in multivariate binary data of the single/multiple membership classification can lead to convergence issues, as it is typically more challenging to fit both covariates and random effects when the response is binary and doesn't have as much variation as a continuous response. Thirdly, in the conception of the MMMC Model (see Browne et al. (2001)), the mapping of each of the $\sum_{k=1}^{K} N_k$ adolescents to the subset of J alters is assumed to result in a mapping such that the subset of J alters that the ith adolescent in the kth school maps to can not be empty. This means that something needs to be done about ith adolescents in the kth school who have $|\text{net}(ik)| = 0$. This is a possibility in network data when the ith adolescent in the kth school may not nominate anyone in the network and thus has no alters. There isn't much known formal guidance for what to do when this case arises. This situation occurred in Tranmer et al. (2014). As there were only a few cases of this,

Tranmer et al. (2014) included the adolescents in the analysis and set the weights of all their multiple membership random effects to 0. In the analysis conducted in this chapter, when this situation occurs, the individual will be assigned a special random effect which represents the fact that they haven't nominated a friend. This can be thought of as a random effect for loners.

4.1.2 Novel contributions

The quality and amount of network data being generated in various disciplines is ever increasing, with a strong emphasis on understanding the underlying process that generated the data becoming ever more important. This chapter provides two distinct novel contributions to the literature on network analysis and the statistical methodology of the MMMC model of Browne et al. (2001) and are given as follows

1. **A multivariate MMMC model**. A critique of the publication of Lorant and Tranmer (2019) was that it did not offer the reader a multivariate model approach to estimate the relative share of variation in multiple binary health behaviours, relating to exercise, alcohol, marijuana, and tobacco use, at the school and friendship network level. In examples, such as the one described, it is reasonable to imagine that alcohol and tobacco use are often correlated. Thus, as a result, it may be wholly appropriate to adopt a multivariate modelling approach. Section 4.3 describes the multivariate MMMC model to be implemented in the first known analysis of network data, which is presented in subsection 4.5.2.

2. **A univariate and multivariate MMMC comparison**. Given that it is not unusual for the different health behaviours of an individual to be correlated, there is great interest in applying a multivariate approach to the data and comparing how the results of a multivariate approach may differ from a univariate approach. To my knowledge,

subsection 4.5.2 is the only instance in which such a comparison has been conducted for network data.

4.2 Data: Social Networking Survey

This section describes and explores the Social Networking Survey (SNS) network data from Los Angeles, California (Valente (2010)). Subsection 4.2.1 provides a detailed description as to how the network data have been gathered and generated, and also what it contains. Subsection 4.2.2 goes on to provide an exploration of the network data.

4.2.1 Description of data

The network data used in this study relate to 555 male and 584 female 10th grade students from 5 schools in Los Angeles, California in 2010. The network data were generated by way of a survey provided to each of the students during regular school hours. The survey questions given to the students can be partitioned into several sections, with the sections of interest to us here being those that relate to the networks that individuals have formed in the 10th grade, along with their household information, personal information, and health behaviours. The binary health behaviours to be modelled are

1. Whether an individual has consumed at least one drink of alcohol in the past 30 days.
2. Whether an individual has smoked at least one cigarette in the past 30 days.
3. Whether an individual has ever tried marijuana.

In each of the 5 schools, a friendship network was derived for the adolescents. The friendship networks were obtained by asking each individual to nominate up to 19 individuals in the 10th grade. The friendship network for school 1 is shown in Figure 4.1 and shows how

individuals with a specific health behaviour response value tend to nominate others with the same health behaviour response value, i.e. individuals are typically joined to others of the same response value (identified by colour). Across the five schools, a total of 1069 individuals were nominated.

4.2.2 Data exploration

The exploration of the network data used in this study focuses largely on the network structures present, as these structures are what motivates the multivariate model described in Model (4.4), and are designed to make use of these structures to estimate the share of variation in the binary responses at the network level.

The summaries of the covariates in the data are shown in Table 4.1. The description of the covariates are as follows. *Gender* is the sex of the individual. *Rooms in house* is the number of rooms an individual has in their house or apartment, excluding the kitchen and bathroom. *Lunch eligible* is whether an individual is eligible for a free or reduced price lunch at school. *Exam grades* are the grades that an individual mainly achieved in school last year. *Father's education* is the highest level of education completed by an individual's father. *Mother's education* is the highest level of education completed by an individual's mother. *General health* is a self-reported measure of what an individual considers their overall health to be. *Facebook use* is a self-reported measure of how often an individual uses Facebook. Never, rarely, occasionally, frequently, and very frequently are considered to be about once a month or less, about once a week or less, about once every 2-3 days, and about once a day or more respectively. *Online gaming* is a self-reported measure of how often an individual uses online gaming services. *Home ownership* was derived by asking individuals "Does your family own its home or rent from a landlord?".

The joint distribution for alcohol use, cigarette use and having ever smoked marijuana in each of the five schools and as a whole provide evidence to support the notion that knowing

Covariates	School 1 ($n = 303$)	School 2 ($n = 204$)	School 3 ($n = 161$)	School 4 ($n = 272$)	School 5 ($n = 199$)	All schools ($n = 1139$)
Gender						
Female	166 (54.8%)	94 (46.1%)	86 (53.4%)	126 (46.3%)	112 (56.3%)	584 (51.3%)
Male	137 (45.2%)	110 (53.9%)	75 (46.6%)	146 (53.7%)	87 (43.7%)	555 (48.7%)
Lunch eligible						
Yes	246 (81.2%)	193 (94.6%)	148 (91.9%)	213 (78.3%)	169 (84.9%)	969 (85.1%)
No	57 (18.8%)	11 (5.4%)	13 (8.1%)	59 (21.7%)	30 (15.1%)	170 (14.9%)
Ages						
<= 14 years old	14 (4.6%)	9 (4.4%)	6 (3.7%)	16 (5.9%)	17 (8.6%)	62 (5.4%)
15 years old	236 (77.9%)	171 (83.8%)	149 (92.6%)	233 (85.7%)	174 (87.4%)	963 (84.6%)
>= 16 years old	53 (17.5%)	24 (11.8%)	6 (3.7%)	23 (8.5%)	8 (4%)	114 (10%)
Rooms in house						
<= 2 rooms	76 (25.1%)	88 (43.1%)	71 (44.1%)	57 (21%)	60 (30.2%)	352 (30.9%)
3 rooms	98 (32.3%)	62 (30.4%)	44 (27.3%)	91 (33.5%)	71 (35.7%)	366 (32.1%)
4 rooms	63 (20.8%)	38 (18.6%)	29 (18%)	67 (24.6%)	40 (20.1%)	237 (20.8%)
>= 5 rooms	66 (21.8%)	16 (7.8%)	17 (10.6%)	57 (21%)	28 (14.07%)	184 (16.2%)
Exam grades						
Mostly A's	50 (16.5%)	17 (8.3%)	10 (6.2%)	39 (14.3%)	17 (8.5%)	133 (11.7%)
Mostly A's and B's	71 (23.4%)	40 (19.6%)	42 (26.1%)	80 (29.4%)	57 (28.6%)	290 (25.5%)
Mostly B's	16 (5.3%)	15 (7.4%)	13 (8.1%)	19 (7%)	10 (5%)	73 (6.4%)
Mostly B's and C's	83 (27.4%)	59 (28.9%)	55 (34.1%)	67 (24.6%)	59 (29.7%)	323 (28.4%)
Mostly C's or lower	83 (27.4%)	73 (35.8%)	41 (25.5%)	67 (24.6%)	56 (28.1%)	320 (28.1%)
Father's education						
8th grade or less	33 (10.9%)	55 (27%)	36 (22.4%)	34 (12.5%)	49 (24.6%)	207 (18.2%)
Some high school	64 (21.1%)	43 (21.1%)	30 (18.6%)	46 (16.9%)	35 (17.6%)	218 (19.1%)
High school graduate	46 (15.2%)	25 (12.3%)	28 (17.4%)	51 (18.8%)	36 (18.1%)	186 (16.3%)
Some college	37 (12.2%)	12 (5.9%)	13 (8.1%)	40 (14.7%)	18 (9.1%)	120 (10.5%)
College graduate	28 (9.2%)	2 (1%)	2 (1.2%)	27 (9.9)	8 (4%)	67 (5.9%)
Advanced graduate	13 (4.3%)	3 (1.5%)	3 (1.9%)	8 (2.9%)	2 (1%)	29 (2.6%)
I don't know	82 (27.1%)	64 (31.4%)	49 (30.4%)	66 (24.3%)	51 (25.6%)	312 (27.4%)
Mother's education						
8th grade or less	40 (13.2%)	41 (20.1%)	38 (23.6%)	35 (12.9%)	42 (21.1%)	196 (17.2%)
Some high school	38 (12.5%)	47 (23%)	37 (23%)	45 (16.5%)	40 (20.1%)	207 (18.2%)
High school graduate	62 (20.5%)	33 (16.2%)	26 (16.2%)	63 (23.2%)	40 (20.1%)	224 (19.7%)
Some college	50 (16.5%)	16 (7.8%)	17 (10.6%)	35 (12.9%)	22 (11.1%)	140 (12.3%)
College graduate	35 (11.6%)	9 (4.4%)	5 (3.1%)	30 (11%)	12 (6%)	91 (8%)
Advanced graduate	11 (3.6%)	3 (1.5%)	3 (1.9%)	8 (2.9%)	2 (1%)	27 (2.4%)
I don't know	67 (22.1%)	55 (27%)	35 (21.8%)	56 (20.6%)	41 (20.6%)	254 (22.3%)
General health						
Excellent	42 (13.9%)	27 (13.2%)	38 (23.6%)	52 (19.1%)	37 (18.6%)	196 (17.2%)
Very good	106 (35%)	61 (30%)	39 (24.2%)	91 (33.5%)	62 (31.2%)	359 (31.6%)
Good	123 (40.6%)	81 (39.7%)	56 (34.8%)	93 (34.2%)	81 (40.7%)	434 (38.1%)
Fair or poor	32 (10.6%)	35 (17.2%)	28 (17.4%)	36 (13.2%)	19 (9.6%)	150 (13.2%)
Facebook use						
Never or rarely	91 (30%)	100 (49%)	95 (59%)	64 (23.5%)	108 (54.27%)	458 (40.2%)
Occasionally	89 (29.4%)	49 (24%)	34 (21.1%)	92 (33.8%)	51 (25.63%)	315 (27.7%)
Frequent or very frequently	123 (40.6%)	55 (27%)	32 (19.9%)	116 (42.7%)	40 (20.1%)	366 (32.1%)
Online gaming						
Never	160 (52.8%)	104 (51%)	96 (59.6%)	132 (48.5%)	118 (59.3%)	610 (53.6%)
Rarely	48 (15.8%)	22 (10.8%)	30 (18.6%)	38 (14%)	33 (16.6%)	171 (15%)
Occasionally	33 (10.9%)	26 (12.8%)	14 (8.7%)	48 (17.7%)	20 (10.1%)	141 (12.4%)
Frequent	23 (7.6%)	26 (12.8%)	8 (5%)	19 (7%)	12 (6%)	88 (7.7%)
Very frequently	39 (12.9%)	26 (12.8%)	13 (8.1%)	35 (12.9%)	16 (8%)	129 (11.3%)
Home ownership						
Own	134 (44.2%)	58 (28.4%)	38 (23.6%)	117 (43%)	77 (38.7%)	424 (37.2%)
Rent	131 (43.2%)	126 (61.8%)	103 (64%)	130 (47.8%)	99 (49.8%)	589 (51.7%)
I don't know	38 (12.5%)	20 (9.8%)	20 (12.4%)	25 (9.2%)	23 (12.6%)	126 (11.1%)

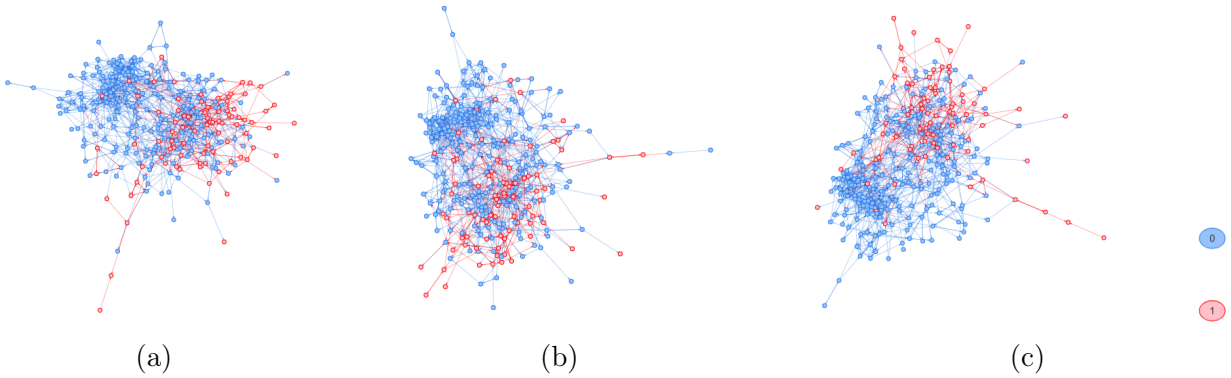
Table 4.1: Summaries of the number of individuals with given covariate values for each school. Decimals are rounded to 1 decimal place.

one or two health behaviours of an individual provides useful information that can be used to determine what the values of the other binary health behaviours may take, and that the distribution of these three behaviours are not independent.

- Let the random variable X be whether an individual has smoked at least one cigarette in the past 30 days.
- Let the random variable Y be whether an individual has ever tried marijuana.
- Let the random variable Z be whether an individual has consumed at least one drink of alcohol in the past 30 days.

Considering all schools together, in comparison to those who have not smoked at least one cigarette in the past 30 days ($X = 0$) and have never tried marijuana ($Y = 0$), the conditional probability of having used alcohol given that you have have smoked at least one cigarette in the past 30 days ($X = 1$) and have ever tried marijuana ($Y = 1$) is greater $\mathbb{P}(Z = 1|X = 1, Y = 1) = \mathbb{P}(Z = 1, X = 1, Y = 1)/\mathbb{P}(X = 1, Y = 1) = 0.14/0.2 = 0.7 > \mathbb{P}(Z = 1|X = 0, Y = 0) = 0.13$. Similarly, for all schools, the conditional probability of having used cigarettes in the past 30 days is greater for those who consumed alcohol in the past 30 days ($Z = 1$) and have ever used marijuana ($Y = 1$) when compared to those who reported never having done either ($Z = Y = 0$), $\mathbb{P}(X = 1|Y = 1, Z = 1) = 0.78 > \mathbb{P}(X = 1|Y = 0, Z = 0) = 0.1$. The same is also true for the conditional distribution of having ever used marijuana in all schools, $\mathbb{P}(Y = 1|X = 1, Z = 1) = 0.82 > \mathbb{P}(Y = 1|X = 0, Z = 0) = 0.07$.

The friendship network in school 1 is shown in Figure 4.1 and colour coded with respect to the binary response value for each of the three health behaviours under investigation. In comparison to individuals with a positive response (i.e. $X = 1$), those with a negative response (i.e. $X = 0$) are more likely to nominate at least one other person with a negative



(a) (b) (c)

Figure 4.1: The friendship network present in school 1. The individuals (vertices) are coloured by their alcohol (a), cigarette (b), and marijuana (c) response. 1, shown in red, is defined as yes and 0, shown in blue, is defined as no.

response. This helps to motivate the need to include the friendship network within each school to help model the uptake of each of the three health behaviours.

The justification for a multivariate modelling approach can be further made by evaluating the joint probability mass functions for having consumed alcohol in the past 30 days, having smoked cigarettes in the past 30 days, and having ever used marijuana and noticing that $\mathbb{P}(X = x, Y = y, Z = z) \neq \mathbb{P}(X = x)\mathbb{P}(Y = y)\mathbb{P}(Z = z)$ for all possible combinations of x, y, and z. For example, in school 1, $\mathbb{P}(X = 1, Y = 1, Z = 1) = 0.129 \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 1)\mathbb{P}(Z = 1) = 0.264 \times 0.257 \times 0.280 = 0.019$. Thus, the 3 outcomes are not independent and their correlation should be modelled.

4.3 The multivariate uniplex MMMC model

This section describes a multivariate MMMC model to be used on network data. The description of the model is compared and contrasted with the univariate model, to illustrate appropriate settings for its use. In addition, the possible benefits of the multivariate MMMC model over the univariate MMMC model are detailed in terms of the additional inference that it allows for when applied to particular types of network data.

The multivariate MMMC model described in this subsection is a multivariate extension of Model (4.2), which is given by

$$
\begin{aligned}
Y_{ikr} &\sim \text{Bernoulli}(\pi_{ikr}) \quad \text{for } i = 1, \ldots, N_k,\ k = 1, \ldots, K,\ r = 1, \ldots, R, \\
\ln\left(\frac{\pi_{ikr}}{1-\pi_{ikr}}\right) &= \mathbf{x}_{ik}^{\top}\boldsymbol{\beta}_r + v_{kr} + \sum_{j \in \text{net}(ik)} w_{ikj}u_{jr}, \\
\boldsymbol{\beta}_r &\sim \text{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \quad \text{for } r = 1, \ldots, R, \\
(v_{k1}, \ldots, v_{kR}) = \mathbf{v}_k &\sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}}) \quad \text{for } k = 1, \ldots, K, \\
(u_{j1}, \ldots, u_{jR}) = \mathbf{u}_j &\sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}}) \quad \text{for } j = 1, \ldots, J, \\
\boldsymbol{\Sigma}_{\mathbf{v}} &\sim \text{Inverse-Wishart}(\phi_{\mathbf{v}} + R - 1, 2\phi_{\mathbf{v}}\Delta_{\mathbf{v}}), \\
\boldsymbol{\Sigma}_{\mathbf{u}} &\sim \text{Inverse-Wishart}(\phi_{\mathbf{u}} + R - 1, 2\phi_{\mathbf{u}}\Delta_{\mathbf{u}}), \\
\Delta_{\mathbf{v}} &= \text{diag}(\lambda_{\mathbf{v}1}, \ldots, \lambda_{\mathbf{v}R}), \\
\lambda_{\mathbf{v}r} &\sim \text{Gamma}\left(\alpha_{\mathbf{v}r}, \frac{1}{\xi_{\mathbf{v}r}^2}\right) \quad \text{for } r = 1, \ldots, R, \\
\Delta_{\mathbf{u}} &= \text{diag}(\lambda_{\mathbf{u}1}, \ldots, \lambda_{\mathbf{u}R}), \\
\lambda_{\mathbf{u}r} &\sim \text{Gamma}\left(\alpha_{\mathbf{u}r}, \frac{1}{\xi_{\mathbf{u}r}^2}\right) \quad \text{for } r = 1, \ldots, R.
\end{aligned} \tag{4.4}
$$

Here y_{ikr} is the binary health behaviour of the rth response for the ith adolescent in the kth school. The vector of a multivariate binary health behaviour for all adolescents in all K schools studied is again denoted by $\mathbf{y} = (y_{111}, \ldots, y_{11R}, \ldots, y_{N_1 11}, \ldots y_{N_1 1R}, \ldots, y_{1K1}$ $, \ldots, y_{1KR}, \ldots, y_{N_K K1}, \ldots, y_{N_K KR})_{(R\times\sum_{k=1}^{K} N_k)\times 1}$. The covariates for the ith adolescent in the kth school are included in a $p \times 1$ vector $\mathbf{x}_{ik}$, and a corresponding $p \times 1$ vector of fixed effect parameters for the rth response are denoted by $\boldsymbol{\beta}_r$ which vary by the R outcomes. The $R \times 1$ vector of joint random effects for the kth school is denoted by $\mathbf{v}_k = (v_{k1}, \ldots, v_{kR})_{R\times 1}$ and has a joint Gaussian distribution. The unstructured covariance matrix $\boldsymbol{\Sigma}_{\mathbf{v}}$ captures the covariance between the R outcomes at the school level and is given by

$$\boldsymbol{\Sigma}_{\mathbf{v}} = \begin{bmatrix} \sigma_1^2 & \gamma_{12} & \cdots & \gamma_{1R} \\ \gamma_{21} & \sigma_2^2 & \cdots & \gamma_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{R1} & \gamma_{R2} & \cdots & \sigma_R^2 \end{bmatrix},$$

where σ_r^2 is the variance of the $K \times 1$ vector of school random effects relating to the rth response $\mathbf{v}_r = (v_{1r}, \ldots, v_{Kr})_{K\times 1}$. The covariance terms $\gamma_{rr'}$ $(r \neq r')$ are the covariances between two different $K \times 1$ vectors of school random effects $\mathbf{v}_r$ and $\mathbf{v}_{r'}$. The $R \times 1$ vector of joint random effects for the jth alter is denoted by $\mathbf{u}_j = (u_{j1}, \ldots, u_{jR})_{R\times 1}$ and has a joint Gaussian distribution. The unstructured covariance matrix $\boldsymbol{\Sigma}_{\mathbf{u}}$ captures the covariance between the R outcomes at the network level and is given by

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \begin{bmatrix} \tilde{\sigma}_1^2 & \tilde{\gamma_{12}} & \cdots & \tilde{\gamma}_{1R} \\ \tilde{\gamma}_{21} & \tilde{\sigma}_2^2 & \cdots & \tilde{\gamma}_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\gamma}_{R1} & \tilde{\gamma}_{R2} & \cdots & \tilde{\sigma}_R^2 \end{bmatrix},$$

where $\tilde{\sigma}_r^2$ is the variance of the $J \times 1$ vector of alter random effects relating to the rth response $\mathbf{u}_r = (u_{1r}, \ldots, u_{Jr})_{J\times 1}$. The covariance terms $\tilde{\gamma}_{rr'}$ $(r \neq r')$ are the covariances between two different $J \times 1$ vectors of alter random effects $\mathbf{u}_r$ and $\mathbf{u}_{r'}$. The weight of the jth alter random effect for the ith adolescent in the kth school is identical to the one in Model (4.2) and given in equation (4.3).

In Model (4.4) the log odds that the ith individual in the kth school has a positive value for the rth response $(y_{ikr} = 1)$ is denoted by $\ln\left(\frac{\pi_{ikr}}{1-\pi_{ikr}}\right)$. The systematic component of the model is comprised of three terms. $\mathbf{x}_{ik}^{\top}\boldsymbol{\beta}_r$ is a term that accounts for the covariate

values of an individual for the rth response. Whereas the two random effects terms v_{kr} and $\sum_{j\in \text{net}(ik)} w_{ikj}u_{jr}$ are used to provide an estimate of the influence of the school and network level after accounting for the covariate term in the model.

4.4 MCMC estimation algorithm

This section describes Markov chain Monte Carlo algorithms for Models (4.2) and (4.4). The joint distribution of $(\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \sigma_u^2, \sigma_v^2)$ for Model (4.2) is given by

$$\begin{aligned} f(\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \sigma_v^2, \sigma_u^2|\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}) f(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) f(\mathbf{v}|\sigma_v^2) f(\mathbf{u}|\sigma_u^2) f(\sigma_v^2) f(\sigma_u^2) \\ &= \prod_{k=1}^{K}\prod_{i=1}^{N_k} \text{Bernoulli}(y_{ik}|\boldsymbol{\beta}, v_k, \mathbf{u}) \text{N}(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\ &\times \prod_{k=1}^{K} \text{N}(v_k|\mu_0, \sigma_v^2) \prod_{j=1}^{J} \text{N}(u_j|\mu_0, \sigma_u^2) \\ &\times \text{Inverse-Gamma}(\sigma_v^2|a_v, b_v) \text{Inverse-Gamma}(\sigma_u^2|a_u, b_u) \end{aligned}$$

where $\mu_0 = 0$, and thus the prior distribution for all random effects have a mean of 0. In comparison to Model (4.2), The joint distribution of $(\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}})$ for Model (4.4) is given by

$$
\begin{aligned}
f(\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}} | \mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}) f(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) f(\mathbf{v}|\boldsymbol{\Sigma}_{\mathbf{v}}) f(\mathbf{u}|\boldsymbol{\Sigma}_{\mathbf{u}}) f(\boldsymbol{\Sigma}_{\mathbf{v}}|\boldsymbol{\lambda}_{\mathbf{v}}) \\
&\quad f(\boldsymbol{\Sigma}_{\mathbf{u}}|\boldsymbol{\lambda}_{\mathbf{u}}) f(\boldsymbol{\lambda}_{\mathbf{v}}) f(\boldsymbol{\lambda}_{\mathbf{u}}) \\
&= \prod_{k=1}^{K} \prod_{i=1}^{N_k} \prod_{r=1}^{R} \text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r, v_{kr}, \mathbf{u}_r) \prod_{r=1}^{R} \text{N}(\boldsymbol{\beta}_r|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\
&\times \prod_{k=1}^{K} \text{N}(\mathbf{v}_k|\boldsymbol{\mu}_{\mathbf{0}}, \boldsymbol{\Sigma}_{\mathbf{v}}) \prod_{j=1}^{J} \text{N}(\mathbf{u}_j|\boldsymbol{\mu}_{\mathbf{0}}, \boldsymbol{\Sigma}_{\mathbf{u}}) \\
&\times \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{v}}|\phi_{\mathbf{v}}, \Delta_{\mathbf{v}}) \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{u}}|\phi_{\mathbf{u}}, \Delta_{\mathbf{u}}) \\
&\times \prod_{r=1}^{R} \text{Gamma}(\lambda_{\mathbf{v}r}|\alpha_{\mathbf{v}r}, \xi^2_{\mathbf{v}r}) \prod_{r=1}^{R} \text{Gamma}(\lambda_{\mathbf{u}r}|\alpha_{\mathbf{u}r}, \xi^2_{\mathbf{u}r})
\end{aligned}
$$

where $\boldsymbol{\mu}_{\mathbf{0}} = \mathbf{0}$ and $\mathbf{u}_r = (u_{1r}, \ldots, u_{Jr})_{J\times 1}$ is the $J \times 1$ vector of alter random effects relating to the rth response. The following subsections in this section detail the full conditionals required for both Models (4.2) and (4.4).

There are many possible ways to apply Markov chain Monte Carlo estimation methods to MMMC models. Browne et al. (2001) proposed the first of such methods in the same publication that proposed the model. Browne et al. (2001) used single site Metropolis updates for the set of fixed effects ($\boldsymbol{\beta}$) and random effects ($\mathbf{v}$, $\mathbf{u}$). The proposal distributions for each of the fixed and random effects were made to be Gaussian with a fixed variance parameter. A Gibbs update was specified for the variance and covariance parameters, as the Inverse-Gamma distribution is a conjugate prior for the univariate Gaussian distribution and the Inverse-Wishart distribution is a conjugate prior for the multivariate Gaussian distribution. In the case of covariance matrices with an Inverse-Wishart prior, the scale matrix can be specified as an estimate of the covariance matrix obtained by using quasi-likelihood methods (see Browne et al. (2009)).

Multilevel models that use Markov chain Monte Carlo estimation can sometimes produce posterior chains that exhibit poor mixing (see Browne et al. (2009)). There are a number of

different transformation techniques that have been used to address this problem, and a brief review of such methods are given here. Hierarchical centering (see Gelfand et al. (1995)) is a type of reparameterization that is designed to substitute the original parameters in the model with a set of new parameters that are less correlated with each other in the joint posterior distribution. Browne et al. (2009) makes use of this method and show with an example how this reparameterization can increase the effective sample size (ESS) of fixed effect parameters by a factor of over 100. However, it is also shown that for models with a large number of random effects whose variance is small, the hierachical centering can have a negative effect on the ESS of parameters. An orthogonal transform of covariates (see Browne et al. (2009)) is a method designed to increase the ESS of fixed effects and has been shown to work well in models that contain a large number of random effects and a binomial likelihood. Parameter expansion (see Liu and Wu (1999)) has been used in Markov chain Monte Carlo estimation to introduce additional parameters to a multilevel model that allow for a set of random effects and its variance parameter to be updated jointly. Browne et al. (2009) show that coupling a parameter expansion with an orthogonal reparameterization of the covariates can lead to better mixing for the variance parameter of random effects. In this section hierarchical centering is used for all models as it is the simplest to implement.

4.4.1 Sampling from $f(\boldsymbol{\beta}_r|\boldsymbol{\beta}_{-r}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_\mathbf{u}, \boldsymbol{\Sigma}_\mathbf{v}, \boldsymbol{\lambda}_\mathbf{v}, \boldsymbol{\lambda}_\mathbf{u}, \mathbf{y})$

The full conditional of $\boldsymbol{\beta}_r$ for Model (4.4) is the product of $\sum_{k=1}^{K} N_k$ Bernoulli likelihoods and a Gaussian prior, which is shown below

$$f(\boldsymbol{\beta}_r|\boldsymbol{\beta}_{-r}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_\mathbf{v}, \boldsymbol{\Sigma}_\mathbf{u}, \boldsymbol{\lambda}_\mathbf{v}, \boldsymbol{\lambda}_\mathbf{u}, \mathbf{y}) \propto \prod_{k=1}^{K}\prod_{i=1}^{N_k} \text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r, v_{kr}, \mathbf{u}_r)\text{N}(\boldsymbol{\beta}_r|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}),$$

where $\boldsymbol{\beta}_{-r}$ is the set of parameters in $\boldsymbol{\beta}$ without the set of parameters in $\boldsymbol{\beta}_r$. Similarly to the full conditional of $\boldsymbol{\beta}$ for (4.2), the Gaussian prior is not conjugate to the Bernoulli likelihood, which results in a full conditional distribution which is again not standard. A Metropolis step is used to update $\boldsymbol{\beta}_r$, which is implemented in blocks. The proposal distribution for the block of parameters to be updated is given by $\mathrm{N}(\boldsymbol{\beta}_r^{(t)}, \sigma_{\boldsymbol{\beta}_r}^{2(t)}\mathbf{I})$, where $\boldsymbol{\beta}_r^{(t)}$ is the current value of the block of parameters and $\sigma_{\boldsymbol{\beta}_r}^{2(t)}$ is the current value of the adaptive tuning parameter for the block of parameters that is designed to keep the acceptance rate $\alpha_{\boldsymbol{\beta}_r}^{(t)}$ for the block of parameters to be between 0.3 - 0.5. The adaptive tuning parameter flattens the proposal distribution when the acceptance rate is less than 0.3 and compacts the proposal distribution when the acceptance rate is greater than 0.5. In the literature on adaptive tuning parameters, Gelman et al. (1997) found the asymptotically optimal acceptance rate for the random walk Metropolis algorithm to be 0.234 under quite general conditions. Browne and Draper (2000) recommend an acceptance rate of between 45%-60% for a wide variety of models and parameters. The acceptance probability of updating the block of current parameters $\boldsymbol{\beta}_r^{(t)}$ to proposed parameters $\boldsymbol{\beta}_r^*$ is given by

$$\min\left\{1, \frac{\prod_{k=1}^{K}\prod_{i=1}^{N_k}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^*, v_{kr}^{(t)}, \mathbf{u}_r^{(t)})\mathrm{N}(\boldsymbol{\beta}_r^*|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})}{\prod_{k=1}^{K}\prod_{i=1}^{N_k}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^{(t)}, v_{kr}^{(t)}, \mathbf{u}_r^{(t)})\mathrm{N}(\boldsymbol{\beta}_r^{(t)}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})}\right\}.$$

4.4.2 Sampling from $f(\mathbf{v}_k|\boldsymbol{\beta}, \mathbf{v}_{-k}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y})$

The full conditional of $\mathbf{v}_k$ for Model (4.4) is the product of K Bernoulli likelihoods and a Gaussian prior, which is shown below

$$f(\mathbf{v}_k|\boldsymbol{\beta}, \mathbf{v}_{-k}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y}) \;\propto\; \prod_{i=1}^{N_k}\prod_{r=1}^{R}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r, v_{kr}, \mathbf{u}_r)\mathrm{N}(\mathbf{v}_k|\boldsymbol{\mu}_{\mathbf{0}}, \boldsymbol{\Sigma}_{\mathbf{v}}).$$

Similarly to the full conditional of v_k for (4.2), the Gaussian prior is not conjugate to the Bernoulli likelihood, which results in a full conditional distribution which is not standard. A Metropolis step is used to update $\mathbf{v}_k$. The proposal distribution for $\mathbf{v}_k$ is given by $\mathrm{N}(\mathbf{v}_k^{(t)}, \sigma_{\mathbf{v}_k}^{2(t)}\mathbf{I})$, where $\mathbf{v}_k^{(t)}$ is the current value of $\mathbf{v}_k$ and $\sigma_{\mathbf{v}_k}^{2(t)}$ is the current value of the adaptive tuning parameter for the proposal distribution. The acceptance probability of updating the block of current parameters $\mathbf{v}_k^{(t)}$ to proposed parameters $\mathbf{v}_k^*$ is given by

$$\min\left\{1, \frac{\prod_{i=1}^{N_k}\prod_{r=1}^{R}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^{(t)}, v_{kr}^*, \mathbf{u}_r^{(t)})\mathrm{N}(\mathbf{v}_k^*|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{v}}^{(t)})}{\prod_{i=1}^{N_k}\prod_{r=1}^{R}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^{(t)}, v_{kr}^{(t)}, \mathbf{u}_r^{(t)})\mathrm{N}(\mathbf{v}_k^{(t)}|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{v}}^{(t)})}\right\}.$$

4.4.3 Sampling from $f(\mathbf{u}_j|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}_{-j}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y})$

The full conditional of $\mathbf{u}_j$ for Model (4.4) is the product of a Gaussian prior and Bernoulli likelihoods equal to the number of times $\mathbf{u}_j$ appears in a likelihood, which is shown below

$$\begin{aligned} f(\mathbf{u}_j|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}_{-j}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y}) &\propto \prod_{k=1}^{K}\prod_{i=1}^{N_k}\prod_{r=1}^{R}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r, v_{kr}, \mathbf{u}_r)\mathrm{N}(\mathbf{u}_j|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{u}}) \\ &\propto \prod_{ik\ s.t.\ j\in\text{net}(ik)}\prod_{r=1}^{R}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r, v_{kr}, \mathbf{u}_r)\mathrm{N}(\mathbf{u}_j|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{u}}). \end{aligned}$$

Similarly to the full conditional of $\mathbf{u}_j$ for (4.2), the Gaussian prior is not conjugate to the Bernoulli likelihood, which results in a full conditional distribution which is not standard. A Metropolis step is used to update $\mathbf{u}_j$. The proposal distribution for $\mathbf{u}_j$ is given by $\mathrm{N}(\mathbf{u}_j^{(t)}, \sigma_{\mathbf{u}_j}^{2(t)}\mathbf{I})$, where $\mathbf{u}_j^{(t)}$ is the current value of $\mathbf{u}_j$ and $\sigma_{\mathbf{u}_j}^{2(t)}$ is the current value of the adaptive tuning parameter for the proposal distribution. The acceptance probability of updating the block of current parameters $\mathbf{u}_j^{(t)}$ to proposed parameters $\mathbf{u}_j^*$ is given by

$$\min\left\{1, \frac{\prod_{ik \ s.t. \ j\in\text{net}(ik)} \prod_{r=1}^{R} \text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^{(t)}, v_{kr}^{(t)}, u_{jr}^{*}, \mathbf{u}_r^{(t)}) \text{N}(\mathbf{u}_j^{(t)}|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{u}}^{(t)})}{\prod_{ik \ s.t. \ j\in\text{net}(ik)} \prod_{r=1}^{R} \text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^{(t)}, v_{kr}^{(t)}, u_{jr}^{(t)}, \mathbf{u}_r^{(t)}) \text{N}(\mathbf{u}_j^{(t)}|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{u}}^{(t)})}\right\}.$$

4.4.4 Sampling from $f(\boldsymbol{\Sigma}_{\mathbf{v}}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y})$

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{v}}$ for Model (4.4) is the product of K Gaussian distributions and a conjugate Inverse-Wishart$(\boldsymbol{\Sigma}_{\mathbf{v}}|\phi_{\mathbf{v}}, \Delta_{\mathbf{v}})$ prior, which is shown below

$$f(\boldsymbol{\Sigma}_{\mathbf{v}}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y}) \ \propto \ \prod_{k=1}^{K} \text{N}(\mathbf{v}_k|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{v}}) \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{v}}|\alpha_{\mathbf{v}r}, \xi_{\mathbf{v}r}^2).$$

This results in an

$$\text{Inverse-Wishart}\left(\phi_{\mathbf{v}} + R + 1 + K, 2\phi_{\mathbf{v}}\Delta_{\mathbf{v}} + \sum_{k=1}^{K} \mathbf{v}_k \mathbf{v}_k^{\top}\right)$$

posterior distribution.

4.4.5 Sampling from $f(\boldsymbol{\Sigma}_{\mathbf{u}}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y})$

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{u}}$ for Model (4.4) is the product of J Gaussian distributions and a conjugate Inverse-Wishart$(\boldsymbol{\Sigma}_{\mathbf{u}}|\phi_{\mathbf{u}}, \Delta_{\mathbf{u}})$ prior, which is shown below

$$f(\boldsymbol{\Sigma}_{\mathbf{u}}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y}) \ \propto \ \prod_{j=1}^{J} \text{N}(\mathbf{u}_j|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_{\mathbf{u}}) \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{u}}|\phi_{\mathbf{u}}, \Delta_{\mathbf{u}}).$$

This results in an

$$\text{Inverse-Wishart}\left(\phi_{\mathbf{u}} + R + 1 + J, 2\phi_{\mathbf{u}}\Delta_{\mathbf{u}} + \sum_{j=1}^{J} \mathbf{u}_j \mathbf{u}_j^\top\right)$$

posterior distribution.

4.4.6 Sampling from $f(\lambda_{\mathbf{v}r}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}-r}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y})$

The full conditional of $\lambda_{\mathbf{v}r}$ for Model (4.4) is the product of an Inverse-Wishart$(\boldsymbol{\Sigma}_{\mathbf{v}}|\phi_{\mathbf{v}}, \Delta_{\mathbf{v}})$ and a Gamma$(\lambda_{\mathbf{v}r}|\alpha_{\mathbf{v}r}, \xi^2_{\mathbf{v}r})$ prior, which is shown below

$$f(\lambda_{\mathbf{v}r}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}-r}, \boldsymbol{\lambda}_{\mathbf{u}}, \mathbf{y}) \quad \propto \quad \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{v}}|\phi_{\mathbf{v}}, \Delta_{\mathbf{v}})\text{Gamma}(\lambda_{\mathbf{v}r}|\alpha_{\mathbf{v}r}, \xi^2_{\mathbf{v}r}).$$

This results in a

$$\text{Gamma}\left(\alpha_{\mathbf{v}r}, \phi_{\mathbf{v}}(\boldsymbol{\Sigma}_{\mathbf{v}}^{-1})_{rr} + \frac{1}{\xi^2_{\mathbf{v}r}}\right)$$

posterior distribution.

4.4.7 Sampling from $f(\lambda_{\mathbf{u}r}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}-r}, \mathbf{y})$

The full conditional of $\lambda_{\mathbf{u}r}$ for Model (4.4) is the product of an Inverse-Wishart$(\boldsymbol{\Sigma}_{\mathbf{u}}|\phi_{\mathbf{u}}, \Delta_{\mathbf{u}})$ and a Gamma$(\lambda_{\mathbf{u}r}|\alpha_{\mathbf{u}r}, \xi^2_{\mathbf{u}r})$ prior, which is shown below

$$f(\lambda_{\mathbf{u}r}|\boldsymbol{\beta}, \mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{v}}, \boldsymbol{\lambda}_{\mathbf{u}-r}, \mathbf{y}) \propto \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{u}}|\phi_{\mathbf{u}}, \Delta_{\mathbf{u}})\text{Gamma}(\lambda_{\mathbf{u}r}|\alpha_{\mathbf{u}r}, \xi^2_{\mathbf{u}r}).$$

This results in a

$$\text{Gamma}\left(\alpha_{\mathbf{u}r}, \phi_{\mathbf{u}}(\boldsymbol{\Sigma}^{-1}_{\mathbf{u}})_{rr} + \frac{1}{\xi^2_{\mathbf{u}r}}\right)$$

posterior distribution.

The univariate model described in (4.2) has a sampling procedure which is very similar to the ones described in this subsection. The software used to implement these algorithms are written in a mixture of `C++` and `R`. The sampling steps are performed in `C++` and the manipulation of the user-specified arguments passed to the function are done in `R`.

4.5 Results: Social Networking Survey

This section presents an analysis of network data from Los Angeles, California to compare and contrast the results of the different MMMC modelling approaches. Subsection 4.5.1 provides details on the methodology underpinning the selection of covariates to be used in the modelling process. Subsection 4.5.2 presents the results generated by the analysis. The aims of this section are given as follows

1. **A univariate and multivariate MMMC comparison**. The novel contribution relating to a univariate and multivariate MMMC comparison stated in Section 4.1.2 is carried out in this section.

2. **An evaluation of the network component of the model**. There are three main

research questions relating to the network level of the model.

Q1 What effect does not having anyone that you consider to be a friend have on the observed odds of engaging in alcohol, cigarette, and marijuana consumption?

Q2 How much variance is there at the network level of the model across the three responses and how does this compare to that of the school level?

Q3 How much correlation is there between the alter random effects across responses and which is greatest?

4.5.1 Variable selection

The variable selection process was conducted in this case study by applying a Bayesian model averaging approach for the univariate logistic regression model which uses all the covariates presented in Table 4.1. This is performed on all three univariate responses separately with the same set of covariates and the covariates present in the models with the greatest posterior probability are all used for the models fit in this section. The main advantage of this method is that it provides a quick and easy form of variable selection. The variable selection process in this study is seen as a means of enabling a comparison between the results generated by the Model (4.2) and (4.4) on the network data, which was the third novel contribution stated in subsection 4.1.2. The Bayesian model averaging procedure for the univariate logistic regression model in which whether an individual has consumed at least one drink of alcohol in the past 30 days was the response, resulted in the model containing gender and exam grade as covariates yielding the greatest posterior probability. The Bayesian model averaging procedure for the univariate logistic regression model in which whether an individual has smoked at least one cigarette in the past 30 days was the response, resulted in the model containing exam grade as a covariate yielding the greatest posterior probability. The Bayesian model averaging procedure for the univariate logistic regression model in which whether an individual has ever tried marijuana was the response, resulted in the model con-

taining exam grade as a covariate yielding the greatest posterior probability. Thus all models in this section were fit with gender and exam grade as covariates.

4.5.2 Results

This subsection presents the results of fitting Models (4.2) and (4.4), whose Markov chain Monte Carlo algorithm are described in Section 4.4. As there are three different responses of interest the set of covariates are used to model each of the responses using the univariate MMMC model, resulting in three univariate MMMC models. Similarly, the set of covariates are used to model the responses jointly using the multivariate MMMC model. The primary reason for this is to allow for a comparison between the results of the univariate and multivariate MMMC model using the same set of covariates.

Model convergence was monitored using the potential scale reduction factor (Gelman et al. (2003)). Each model in this section was run twice, simulating two sets of samples from the posterior distribution of each parameter to compute the potential scale reduction factor. For each run, 300,000 iterations were thinned by a factor of 20 after a burn-in period of 100,000 samples, resulting in 15,000 posterior samples for each run.

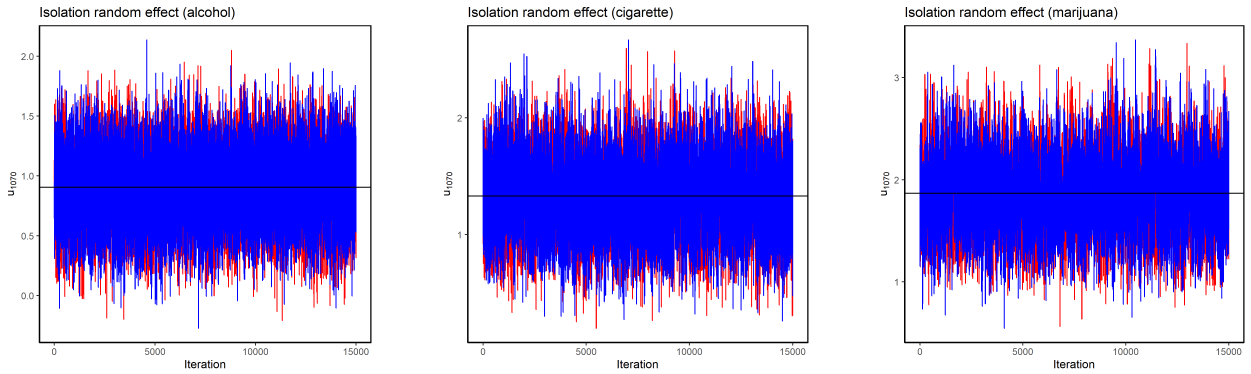
4.5.2.1 Multivariate model results

This section presents the results of fitting Model (4.4) with a hierarchical half-t prior on the Social Network Survey data.

	Alcohol ($r = 1$)				Cigarettes ($r = 2$)				Marijuana ($r = 3$)				Group size
	OR	95% C.I.	ESS	PSRF	OR	95% C.I.	ESS	PSRF	OR	95% C.I.	ESS	PSRF	
Intercept	0.09	(0.02, 0.46)	1668	1	0.05	(0.01, 0.27)	1850	1	0.03	(0, 0.21)	1487	1	-
Gender													
Female	-	-	-	-	-	-	-	-	-	-	-	-	584 (51.3%)
Male	0.27	(0.13, 0.57)	7286	1	1.48	(0.68, 3.29)	6136	1	0.7	(0.28, 1.7)	5606	1	555 (48.7%)
Exam grades													
Mostly A's	-	-	-	-	-	-	-	-	-	-	-	-	133 (11.7%)
Mostly A's and B's	1.82	(0.43, 7.24)	1959	1	0.94	(0.21, 4.26)	2358	1	0.8	(0.14, 4.31)	1843	1	290 (25.5%)
Mostly B's	1.78	(0.06, 46.99)	2059	1	0.02	(0, 0.93)	2171	1	0.19	(0, 9.78)	1934	1	73 (6.4%)
Mostly B's and C's	4.38	(1.17, 16.12)	1881	1	3.04	(0.76, 12.3)	2052	1	3.97	(0.84, 17.81)	1610	1	323 (28.4%)
Mostly C's or lower	16.5	(4.35, 62.18)	1918	1	11.29	(2.86, 45.15)	2183	1	25.78	(5.75, 114.43)	1768	1	320 (28.1%)
Covariance components	Post. Mean	95% C.I.	ESS	PSRF	Post. Mean	95% C.I.	ESS	PSRF	Post. Mean	95% C.I.	ESS	PSRF	
$\mathbf{\Sigma}_{\mathbf{v}rr}$	3.49	(0.06, 24.48)	30000	1.07	4.78	(0.03, 36.72)	30000	1.03	4.45	(0.04, 32.2)	29426	1.02	
$\mathbf{\Sigma}_{\mathbf{v}12}$	0.05	(-1.98, 2.12)	30000	1.01	-	-	-	-	-	-	-	-	
$\mathbf{\Sigma}_{\mathbf{v}13}$	0.12	(-1.93, 2.25)	30000	1.19	-	-	-	-	-	-	-	-	
$\mathbf{\Sigma}_{\mathbf{v}23}$	0.08	(-2.12, 2.31)	30000	1.04	-	-	-	-	-	-	-	-	
$\mathbf{\Sigma}_{\mathbf{u}rr}$	7.49	(4.25, 12.59)	658	1.01	11.85	(6.7, 20.14)	489	1	17.63	(9.82, 31.23)	469	1	
$\mathbf{\Sigma}_{\mathbf{u}12}$	8.48	(5.24, 13.17)	504	1	-	-	-	-	-	-	-	-	
$\mathbf{\Sigma}_{\mathbf{u}13}$	10.47	(6.51, 16.63)	541	1	-	-	-	-	-	-	-	-	
$\mathbf{\Sigma}_{\mathbf{u}23}$	13.58	(8.28, 22.11)	431	1	-	-	-	-	-	-	-	-	
Correlation components	Post. Mean	95% C.I.	ESS	PSRF	Post. Mean	95% C.I.	ESS	PSRF	Post. Mean	95% C.I.	ESS	PSRF	
ρ_{12}	0.91	(0.78, 0.97)	-	-	-	-	-	-	-	-	-	-	
ρ_{13}	0.92	(0.80, 0.97)	-	-	-	-	-	-	-	-	-	-	
ρ_{23}	0.95	(0.87, 0.98)	-	-	-	-	-	-	-	-	-	-	
DIC = 3112.071													

Table 4.2: Results of the multivariate model for alcohol, cigarettes, and marijuana consumption.

Table 4.2 presents the results of the multivariate model for whether an individual has consumed at least one drink of alcohol in the past 30 days, whether an individual has smoked at least one cigarette in the past 30 days, and whether an individual has ever tried marijuana. The results from this multivariate model show that, in comparison to females, males had a lower observed odds of consuming alcohol in the past 30 days (odds ratio (OR) = 0.27, 95% CI: (0.13, 0.57)). In comparison to those who got mostly A's in exams last year, those who got mostly B's and C's in exams last year had a greater observed odds of having consumed at least one drink of alcohol in the past 30 days (OR = 16.5, 95% CI: (4.35, 62.18)), smoking at least one cigarette in the past 30 days (OR = 11.29, 95% CI: (2.86, 45.15)), and having ever consumed marijuana (OR = 25.78, 95% CI: (5.75, 114.43)). The results also show that for the covariance matrix relating to the random effects of the network, the largest variance



(a) (b) (c)

Figure 4.2: Trace plots of the two chains (shown in red and blue) generated relating to the isolation random effect u_{1070} in the multivariate MMMC model for responses relating to alcohol (a), cigarettes (b), and marijuana (c). The posterior mean is denoted by the black horizontal line.

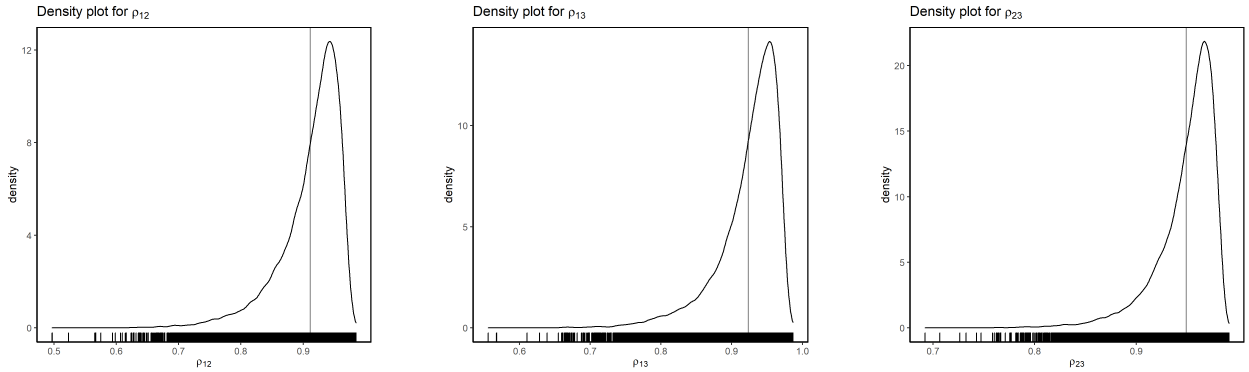
is for the response relating to marijuana ($\hat{\boldsymbol{\Sigma}}_{\mathbf{u}33} = 17.63$, 95% CI: (9.82, 31.23)).

Figures 4.2a, 4.2b, and 4.2c show the traces plots for the alter random effects $u_{1070,1}$, $u_{1070,2}$, and $u_{1070,3}$, which are the random effect of not nominating a friend, across each of the responses. The results show that the posterior mean of $u_{1070,1}$ is 0.90 (95% CI: (0.33, 1.48)), and thus not nominating a friend increases the observed probability of having consumed alcohol in the past 30 days, supporting the literature on loneliness increasing the propensity for an individual to abuse alcohol (see Akerlind and Hörnquist (1992)). The posterior mean of $u_{1070,2}$ is 1.32 (95% CI: (0.74, 1.93)), and thus not nominating a friend also increases the observed probability of having smoked at least one cigarette in the past 30 days. This result supports previous work that has found that loneliness has a significant effect on an individual's observed probability of smoking (see Lauder et al. (2006)) The posterior mean of $u_{1070,3}$ is 1.87 (95% CI: (1.21, 2.58)), and thus not nominating a friend also increases the observed probability of having ever tried marijuana. This answers **Q1** and suggests that the effect of not having a person you consider a friend in the network is to increase the observed probability of partaking in negative health behaviours.

In order to address how much variance there is at the network level of the model across the three responses and how this compares to that of the school level (**Q2**), the weighting of the alter random effects need to be considered (Fielding and Goldstein (2006)). Therefore, in Model (4.4) when considering the contribution to the overall variance of rth response's log odds, the contribution of the network level for the ith adolescent in the kth school is $\text{Var}(\sum_{j\in\text{net}(ik)} w_{ikj}u_{jr}) = \sum_{j\in\text{net}(ik)} w_{ikj}^2\text{Var}(u_{jr}) = \sum_{j\in\text{net}(ik)} w_{ikj}^2\sigma_r^2$ and not necessarily the same for each adolescent. In an attempt to make a direct comparison between the variance at the school and network levels, for each response, we average the network variance of all individuals to obtain an average variance component for the network level, as proposed in Tranmer et al. (2014). As a result, the average network variance contribution to the rth response's log odds is $\sum_{k=1}^{K}\sum_{i=1}^{N_k}(\sum_{j\in\text{net}(ik)} w_{ikj}^2\sigma_r^2)/\sum_{k=1}^{K} N_k$. The average network variance contribution to the log odds of whether an individual has consumed at least one drink of alcohol in the past 30 days, whether an individual has smoked at least one cigarette in the past 30 days, and whether an individual has ever tried marijuana are 3.07, 4.85, and 7.22 respectively. Thus, making a direct comparison with the school level variance contribution with these quantities, the average network variance contribution to the log odds of whether an individual has smoked at least one cigarette in the past 30 days and whether an individual has ever tried marijuana is greater than that of the school.

Figures 4.3a, 4.3b, and 4.3c show the density plots of samples for the correlation parameters relating to the sets of alter random effects. The 95% credible intervals for all three parameters do not contain the value of 0, providing evidence that there is significant correlation between the sets of alter random effects across the 3 responses. The posterior means for the correlation of the alter random effects ρ_{12}, ρ_{13}, and ρ_{23} are 0.91 (95% CI: (0.78, 0.97)), 0.92 (95% CI: (0.80, 0.97)), and 0.95 (95% CI: (0.87, 0.98)) respectively. Thus the posterior mean correlation for the sets alter random effects is greatest between the cigarette and marijuana related response and answers **Q3**. The 95% credible intervals for the correlation parameters relating to the sets of school random effects contain the value of 0, providing ev-

idence that there is no significant correlation between the sets of school level random effects across the 3 responses.



(a) (b) (c)

Figure 4.3: Density plots of samples for the correlation parameters relating to the alter random effects across responses ρ_{12} (a), ρ_{13} (b), and ρ_{23} (c). The posterior mean is denoted by the black vertical line.

4.5.2.2 Univariate model results

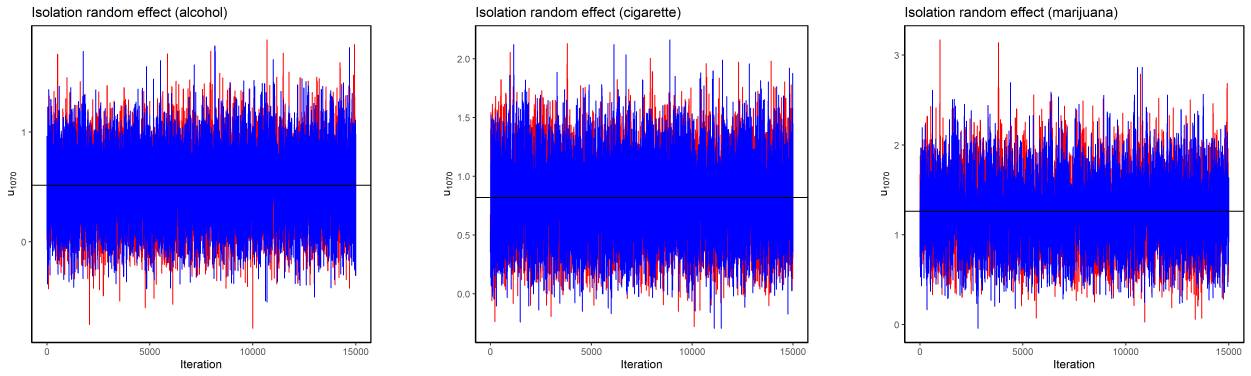
This section presents the results of fitting the univariate model to the Social Network Survey data. The univariate model in this section is fit by merging the 3 responses into 1 response so that the response vector for the univariate and multivariate models are identical, and so direct comparisons between the models can be made in terms of DIC, etc.

	Alcohol				Cigarettes				Marijuana			
	OR	95% C.I.	ESS	PSRF	OR	95% C.I.	ESS	PSRF	OR	95% C.I.	ESS	PSRF
Intercept	0.03	(0.01, 0.16)	1333	1	0.02	(0.01, 0.1)	1901	1	0.02	(0, 0.09)	1561	1.01
Gender												
Female	-	-	-	-	-	-	-	-	-	-	-	-
Male	0.31	(0.16, 0.59)	10905	1	1.21	(0.64, 2.32)	9864	1	0.61	(0.3, 1.25)	10006	1
Exam grades												
Mostly A's	-	-	-	-	-	-	-	-	-	-	-	-
Mostly A's and B's	2.91	(0.79, 11.36)	1573	1	1.72	(0.48, 6.23)	2337	1	1.73	(0.39, 7.46)	1902	1
Mostly B's	7.76	(0.31, 186.79)	1701	1	0.27	(0.01, 7.39)	2080	1	1.79	(0.04, 57.97)	1576	1
Mostly B's and C's	11.01	(3.19, 39.25)	1501	1	9.13	(2.8, 29.67)	2165	1	11.96	(3.03, 46.06)	1716	1.01
Mostly C's or lower	48.57	(14.01, 174.16)	1527	1	37.32	(11.59, 120.3)	2132	1	76.36	(20.29, 284.29)	1886	1.01
Variance components	Post. Mean	95% C.I.	ESS	PSRF	Post. Mean	95% C.I.	ESS	PSRF	Post. Mean	95% C.I.	ESS	PSRF
σ_v^2	2.93	(0.07, 19.6)	30000	1.01	5.64	(0.03, 39.36)	30000	1.05	4.16	(0.04, 30.51)	27129	1
σ_u^2	2	(0.41, 6.53)	907	1	2.07	(0.42, 6.82)	886	1	3.88	(0.58, 11.73)	837	1
DIC = 3670.767												

Table 4.3: Results of the univariate model for alcohol, cigarettes, and marijuana consumption.

Table 4.3 presents the results of the univariate model for whether an individual has consumed at least one drink of alcohol in the past 30 days, whether an individual has smoked at least one cigarette in the past 30 days, and whether an individual has ever tried marijuana. Similarly to the results of the multivariate model presented in Table 4.2, in comparison to females, males had a lower observed odds of consuming alcohol in the past 30 days (OR = 0.31, 95% CI: (0.16, 0.59)). In contrast to the results of the multivariate model presented in Table 4.2, the estimates of the ORs for those who got mostly C's or lower in exams last year, in comparison to those who got mostly A's, is much greater, i.e for the cigarette related response in the multivariate model the estimate is OR = 11.29 (95% CI: (2.86, 45.15)) whereas it is OR = 37.32 (95% CI: (11.59, 120.3)) in the univariate model. The results also show that for the variances of the alter random effects, the largest variance is for the response relating to marijuana ($\hat{\sigma}_3^2 = 3.88$, 95% CI: (0.58, 11.73)). This result is similar to that of the multivariate model.

Figures 4.4a, 4.4b, and 4.4c show the traces plots for the alter random effects $u_{1070,1}$, $u_{1070,2}$, and $u_{1070,3}$, which are the random effect of not nominating a friend, across each of



(a) (b) (c)

Figure 4.4: Trace plots of the two chains (shown in red and blue) generated relating to the isolation random effect u_{1070} in the univariate MMMC model for responses relating to alcohol (a), cigarettes (b), and marijuana (c). The posterior mean is denoted by the black horizontal line.

the responses. The results show that the posterior mean of $u_{1070,1}$ is 0.51 (95% CI: (-0.06, 1.11)). The posterior mean of $u_{1070,2}$ is 0.82 (95% CI: (0.25, 1.43)), and thus not nominating a friend increases the observed probability of having smoked at least one cigarette in the past 30 days. The posterior mean of $u_{1070,3}$ is 1.26 (95% CI: (0.62, 1.98)), and thus not nominating a friend also increases the observed probability of having ever tried marijuana.

4.5.2.3 Univariate and multivarate model comparison

This section directly compares results relating to the univariate and multivariate models, and seeks to address which model is preferred in this study. The two main reasons for the multivariate model being preferred are as follows

1. **The multivariate MMMC model results in a lower DIC.** The DIC for the multivariate model across the two runs is 3112.071, whereas the DIC for the univariate model across the two runs is 3670.767. As the multivariate model has a smaller estimated DIC, the multivariate model would best predict a replicate data set which has the same structure as that currently observed in the Social Networking Survey.

2. **The multivariate MMMC model allows for inference of the correlation between alter random effects across responses.** In addition to having a smaller DIC, unlike the univariate model, the multivariate model allows for inference of the correlation between alter random effects across responses. The multivariate model results show that the 95% credible interval for the correlation between the alter random effects for the cigarette and marijuana responses has a larger start and end point than for the alcohol and marijuana responses. This inference could be valuable to a policy maker who aims to prevent an adolescent from ever trying marijuana by making an intervention between an adolescent and their alters.

All computation in this section was performed on a desktop computer with Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz with 32GB of RAM. The time taken to implement the model presented in Table 4.3 took approximately 23 hours to generate 2 runs of 300,000 iterations thinned by a factor of 20 after a burn-in period of 100,000 samples. Whereas the time taken to implement the models presented in Table 4.2 took approximately 27.5 hours to generate 2 runs of 300,000 iterations thinned by a factor of 20 after a burn-in period of 100,000 samples.

4.6 Discussion and conclusions

This chapter has investigated the use of univariate and multivariate MMMC models on a set of network data, comparing and contrasting the results for both types of model. The results show that, across the univariate and multivariate models, the covariate with the largest impact on whether an individual has consumed at least one drink of alcohol in the past 30 days, whether an individual has smoked at least one cigarette in the past 30 days, and whether an individual has ever tried marijuana was the exam grades an individual achieved last year. In particular, holding all other covariates fixed, in comparison to those

that achieved mostly A's in exams last year, those that achieved mostly C's or lower had much higher odds of a positive value for all responses.

The results in this section also reveal how the social network component of a data set can impact the behaviour of an individual. It also supports the idea that the alter random effects may be correlated across different responses, and thus possessing positive alter influences on one behaviour may carry over to other behaviours. For example, the results in this section show that there is a strong positive linear relationship between the alter random effects relating to the cigarette and marijuana responses, $\rho_{23} = 0.95$ (95% CI: (0.87, 0.98)). Thus an individual having alters that positively influence whether they have smoked at least one cigarette in the past 30 days, may carry over to impacting their marijuana consumption.

There are a number of potential limitations and areas for further work in regards to certain parts of the case study. One of the first limitations is the use of 5 schools as random effects, which may not be wholly sensible. In the univariate model the estimate of the variance parameter for the school level random effects are based on 5 random effects, which may not be ideal. In the multivariate model, correlation estimates at the school level will be based on 2 vectors containing 5 elements, which may be viewed as not ideal. In addition to this, a further area of work would involve studying the effects of different types of reparamterizations (see Browne et al. (2009)) on the fixed effects, random effects, variance parameters, and covariance parameters in the MMMC models described in this chapter. There is also scope to extend the models described in this section by allowing for multiple networks.

Chapter 5

Prior specifications

5.1 Introduction

In this chapter, I investigate the use of different priors across a range of different models to see how a change in prior for the variance-components and covariance components in models previously presented in earlier chapters of this thesis may induce different properties for the parameters in the model. Bayesian models are known for being sensitive to the choice of prior distributions for variance and covariance parameters. The univariate and multivariate MMMC models presented in this thesis contain a number of variance and covariance parameters in their specification, and so it is of great importance to better understand how the change in priors for these parameters affect the simulated posterior distributions that the models yield. This chapter also provides an exploration of how changes in the grouping size and structure of random effects impact the results of these models.

In the Bayesian analysis of statistical models, there are two types of prior distributions that can be imposed on a parameter. The first is a diffuse prior and the second is an informative prior. In the case of the diffuse prior, little may be known about the parameter of interest or the researcher may want to minimise the effect of the prior distribution on the

resultant posterior distribution. In the case of the informative prior, prior information may be available for the parameter of interest, such as similar previous studies or comments from an expert within the field of research. In this section, the focus is on exploring different types of diffuse priors, which are compared and contrasted using bias and credible interval estimates produced by the models on simulated data for which we know the true value of parameters.

The literature pertaining to the specification of diffuse priors is vast and comprehensive for a variety of models (see Arnold and Villasenor (1997), Browne and Draper (2006), Gelman et al. (2013), McElreath (2016), Pateras et al. (2021)). In the literature on priors for variance parameters, one of the most common approaches is to impose a diffuse inverse-gamma distribution on variance parameters, as random effects are typically modelled using a normal distribution and the inverse-gamma distribution is conjugate to this, making it convenient to use. Similarly, in the literature on priors for covariance parameters, one of the most common approaches is to impose an inverse-Wishart distribution on covariance parameters, as the random effects in multivariate models are often modeled using a multivariate normal distribution and the inverse-Wishart distribution is conjugate to this, making it convenient to use.

The remainder of the chapter is structured as follows. Subsection 5.1.1 outlines the novel contributions for this chapter. Section 5.2 provides an exploration of different priors for variance components for a univariate MMMC model. The section also explores how changing the grouping size and structure of random effects impacts the results of these models. Section 5.3 presents a simulation study for the multivariate MMMC model. The section also explores the use of different prior specifications for covariance structures in the multivariate model. Section 5.4 concludes with a discussion.

5.1.1 Novel contributions

This section presents a simulation study to illustrate how the structure of the multiple classification random effects and choice of priors can impact the estimation of parameters in the model. It also illustrates the workings of the Bernoulli multivariate MMMC model with three different priors imposed on the covariance matrices, the Inverse-Wishart, scaled Inverse-Wishart, and hierarchical half-t priors. This simulation study has two aims and are given as follows

A1 **Parameter correctness**. The main purpose of this simulation study is to illustrate the correctness of the code written to implement each model using different priors.

A2 **An exploration of the properties of different priors**. A change in the prior for the variance parameter/covariance matrices in a model may induce different properties for the parameters in the model, such as different effective sample sizes (ESS), etc. Thus it is of interest to compare and contrast the different priors in this respect.

5.2 Variance-components prior models

There are a number of potential issues with using the Inverse-Gamma distribution as a prior for variance parameters in Bayesian hierarchical modelling. This subsection discusses a few of these issues and puts forward two alternative priors that can be used for variance parameters in Bayesian hierarchical models. The two alternatives are the uniform prior and the half-normal prior (see Pateras et al. (2021)). This subsection also addresses and investigates how the number of multiple membership random effects and how the distribution of observations assigned to those multiple membership groups can affect the simulated posterior distributions of parameters in the model.

5.2.1 Simulation study

Each simulation was performed using 100 data sets. The data sets are such that they differ for each scenario that is explored. There are 8 different scenarios under which the data has been generated, which are presented in Table 5.1. The scenarios are as follows: (i) There are 120 individuals assigned to possibly more than one of 10 multiple membership groups so that the number of individuals in each group is "unbalanced" (10 UB). (ii) There are 120 individuals assigned to possibly more than one of 10 multiple membership groups so that the number of individuals in each group is balanced (10 B). (iii) There are 300 individuals assigned to possibly more than one of 25 multiple membership groups so that the number of individuals in each group is "unbalanced" (25 UB). (iv) There are 300 individuals assigned to possibly more than one of 25 multiple membership groups so that the number of individuals in each group is balanced (25 B). (v) There are 480 individuals assigned to possibly more than one of 40 multiple membership groups so that the number of individuals in each group is "unbalanced" (40 UB). (vi) There are 480 individuals assigned to possibly more than one of 40 multiple membership groups so that the number of individuals in each group is balanced (40 B). (vii) There are 1100 individuals assigned to possibly more than one of 55 multiple membership groups so that the number of individuals in each group is "unbalanced" (55 UB). (viii) There are 1100 individuals assigned to possibly more than one of 55 multiple membership groups so that the number of individuals in each group is balanced (55 B).

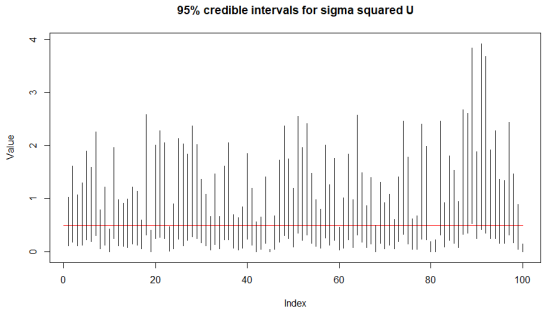
Each model was fit for a total of 20,000 iterations, with a burn-in period of 20,000 and thinning of 1. The model used in this subsection is that of Equation 4.2 in Chapter 4, with the exclusion of the random effects for the single membership classification $\mathbf{v}$ and its corresponding variance σ_v^2 and the use of a Poisson likelihood. 8 different priors are imposed on σ_u^2, namely, $\sigma_u^2 \sim \text{IG}(0.001, 0.001)$, $\sigma_u^2 \sim \text{IG}(0.01, 0.01)$, $\sigma_u^2 \sim \text{IG}(1, 1)$, $\sigma_u^2 \sim \text{IG}(1, 0.01)$, $\sigma_u^2 \sim \text{Uniform}(0, 1000)$ and $\sigma_u^2 \sim \text{HN}(10000)$. The true value of parameters is held constant across the study design and selected to be $\beta_0 = 0.444$, $\beta_1 = 0.477$, $\beta_2 = 0.656$ and $\sigma_u^2 = 0.5$.

Table 5.1: Summary of the study design.

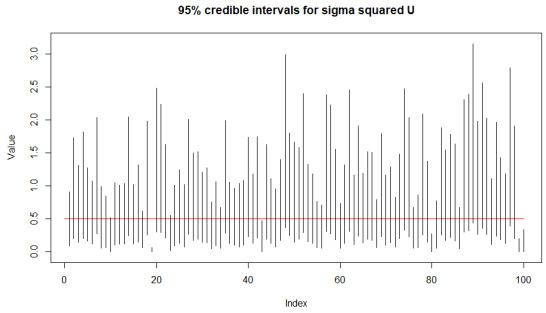
Design	# of individuals	# of individuals in each multiple classification group
10 UB	120	10 10 13 15 9 14 14 16 8 11
10 B	120	20 for all groups
25 UB	300	12 9 14 7 18 10 16 11 11 8 8 15 14 20 13 17 9 11 19 9 8 9 15 7 10
25 B	300	20 for all groups
40 UB	480	6 17 7 9 13 13 14 14 10 19 20 15 13 8 16 13 13 11 11 13 5 7 8 14 12 15 8 17 18 12 8 10 9 10 12 12 10 14 11 9 17
40 B	480	20 for all groups
55 UB	1100	6 8 11 13 10 10 12 18 12 14 13 9 12 10 9 8 15 9 13 16 13 13 17 10 6 12 10 19 10 8 16 15 13 11 11 10 20 9 16 14 15 14 9 14 6 13 17 11 13 8 8 13 17 11 10
55 B	1100	20 for all groups

Figure 5.1 focuses on comparing the 95% credible intervals produced by the Inverse-Gamma(0.001, 0.001) prior for the variance parameter σ_u^2, when there are 10, 25, 40 and 55 multiple membership groups that are balanced and unbalanced. The set of figures {5.1a, 5.1c, 5.1e, 5.1g} shows that, holding the prior and balancing fixed, the coverage for the parameter tends to improve, the more multiple membership groups there are. The coverage in Figures 5.1a, 5.1c, 5.1e and 5.1g are 0.9, 0.91, 0.95 and 0.96 respectively. In contrast, the set of Figure {5.1b, 5.1d, 5.1f, 5.1h} shows that holding the prior and balancing fixed, when there is balance, in groups the coverage is fairly good. The coverage in Figures 5.1b, 5.1d,

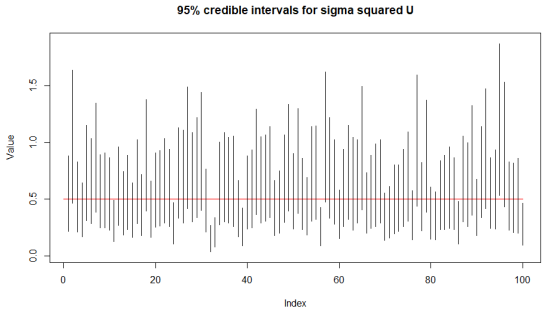
5.1f and 5.1h are 0.95, 0.92, 0.96 and 0.94 respectively. However, it is worth noting that the average width of the 95% credible interval does shrink, the more groups that there are. This pattern also appears for all of the priors explored.



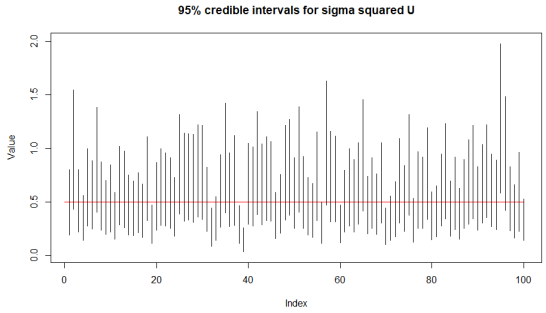
(a) IG(0.001, 0.001), 10 groups, unbalanced.



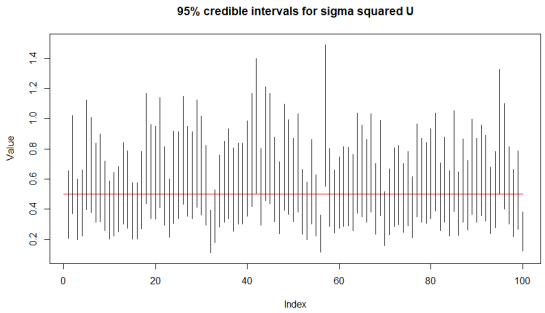
(b) IG(0.001, 0.001), 10 groups, balanced.



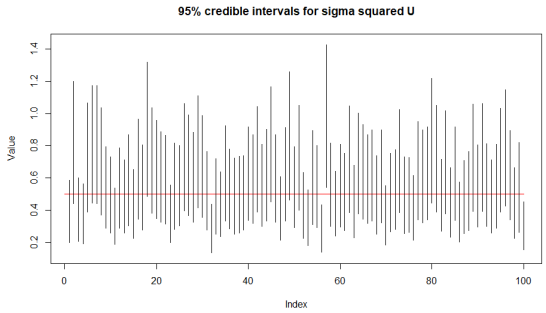
(c) IG(0.001, 0.001), 25 groups, unbalanced.



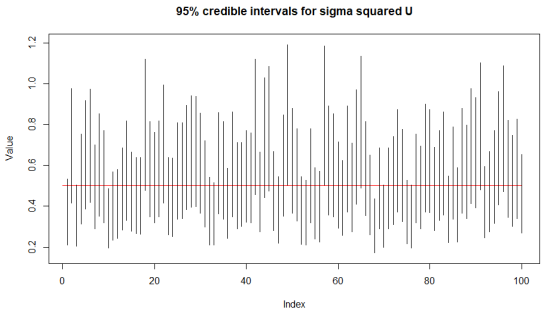
(d) IG(0.001, 0.001), 25 groups, balanced.



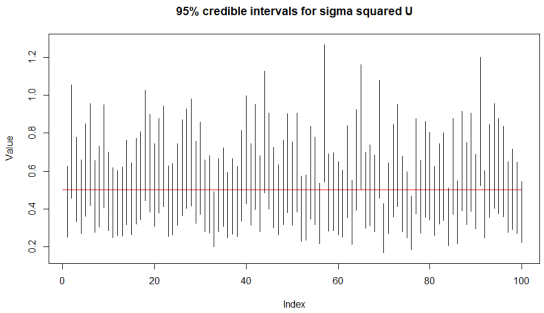
(e) IG(0.001, 0.001), 40 groups, unbalanced.



(f) IG(0.001, 0.001), 40 groups, balanced.



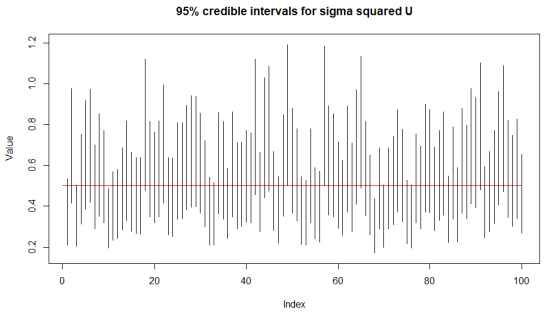
(g) IG(0.001, 0.001), 55 groups, unbalanced.



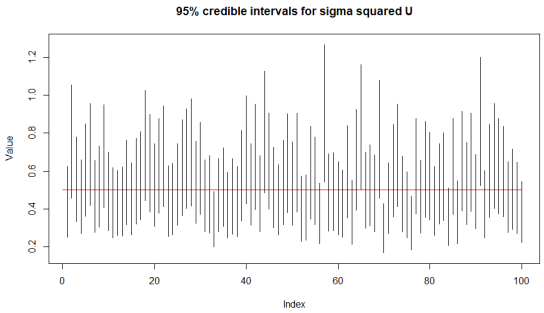
(h) IG(0.001, 0.001), 55 groups, balanced.

Figure 5.1: 95% credible intervals for the 100 simulations for the parameter σ_u^2 for the IG(0.001, 0.001) prior across for the balanced/unbalanced scenarios and all multiple membership group sizes.

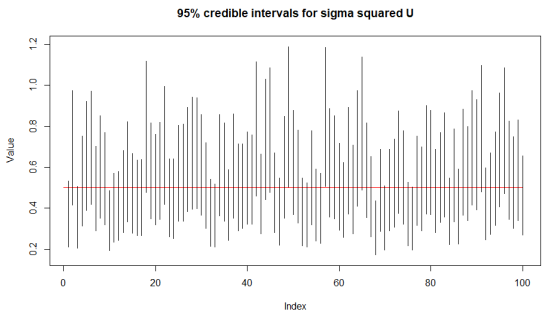
Figure 5.2 focuses on comparing the 95% credible intervals produced by four Inverse-Gamma priors for the variance parameter σ_u^2, when there are 55 multiple membership groups that are balanced and unbalanced. The coverage in Figures 5.2a, 5.2b, 5.2c, 5.2d, 5.2e, 5.2f, 5.2g and 5.2h are 0.96, 0.95, 0.96, 0.94, 0.96, 0.96, 0.92 and 0.94 respectively. In terms of coverage, the Inverse-Gamma(1, 0.01) prior with an unbalanced number of individuals across the 55 groups performed the worst with a coverage of 0.92.



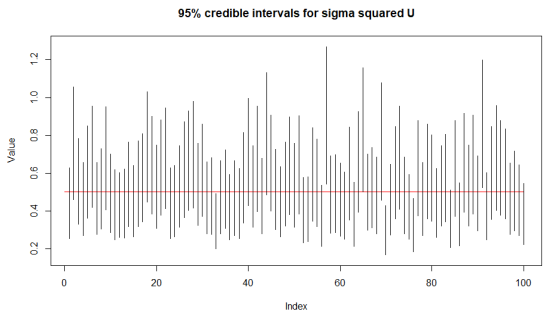
(a) IG(0.001, 0.001), 55 groups, unbalanced.



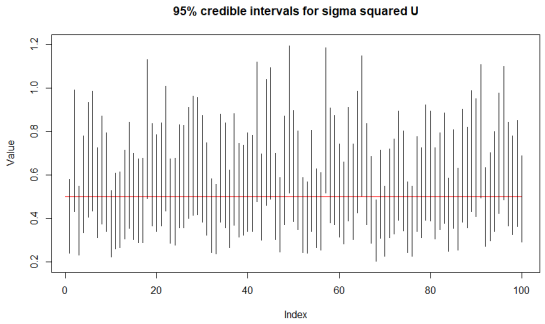
(b) IG(0.001, 0.001), 55 groups, balanced.



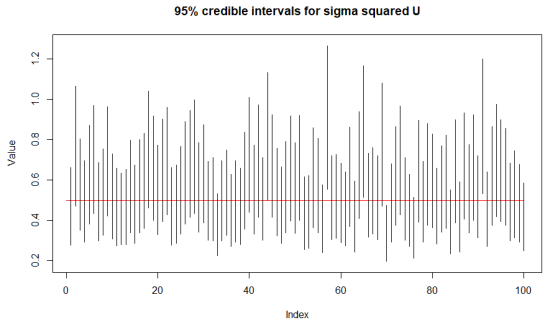
(c) IG(0.01, 0.01), 55 groups, unbalanced.



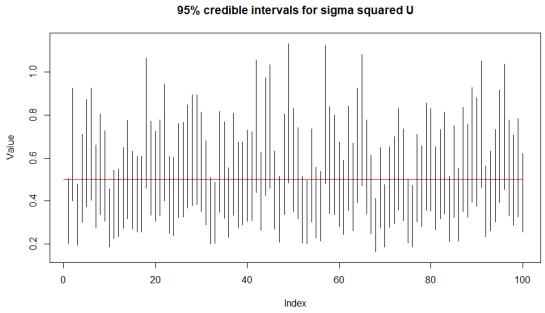
(d) IG(0.01, 0.01), 55 groups, balanced.



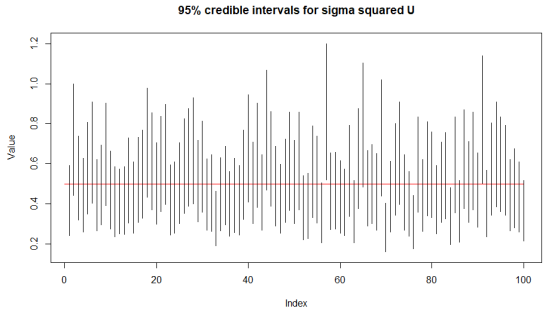
(e) IG(1, 1), 55 groups, unbalanced.



(f) IG(1, 1), 55 groups, balanced.



(g) IG(1, 0.01), 55 groups, unbalanced.



(h) IG(1, 0.01), 55 groups, balanced.

Figure 5.2: 95% credible intervals for the 100 simulations for the parameter σ_u^2 across the four Inverse-Gamma priors for the balanced/unbalanced scenario with 55 multiple membership groups.

Figures 5.3, 5.4, 5.5 and 5.6 show the bias of the model parameters β_0, β_1, β_2 and σ_u^2 as a function of the unbalanced study design across the 6 different priors for σ_u^2. Here the focus is on the unbalanced study design across the 6 different priors for σ_u^2, as this is the structure

that a researcher is more likely to come across in their work. As shown in Figures 5.4 and 5.5, each of the 6 priors for σ_u^2, have similar bias values across the unbalanced study design. As shown in Figures 5.3 and 5.6, the significant differences in the 6 different priors for σ_u^2 across the study design start to show. Figure 5.6 shows that the uniform and Inverse-Gamma(1, 1) prior imposed on σ_u^2 induce relatively large absolute biases for the corresponding parameters when the number of groups is small but settles down when the number of groups gets larger.

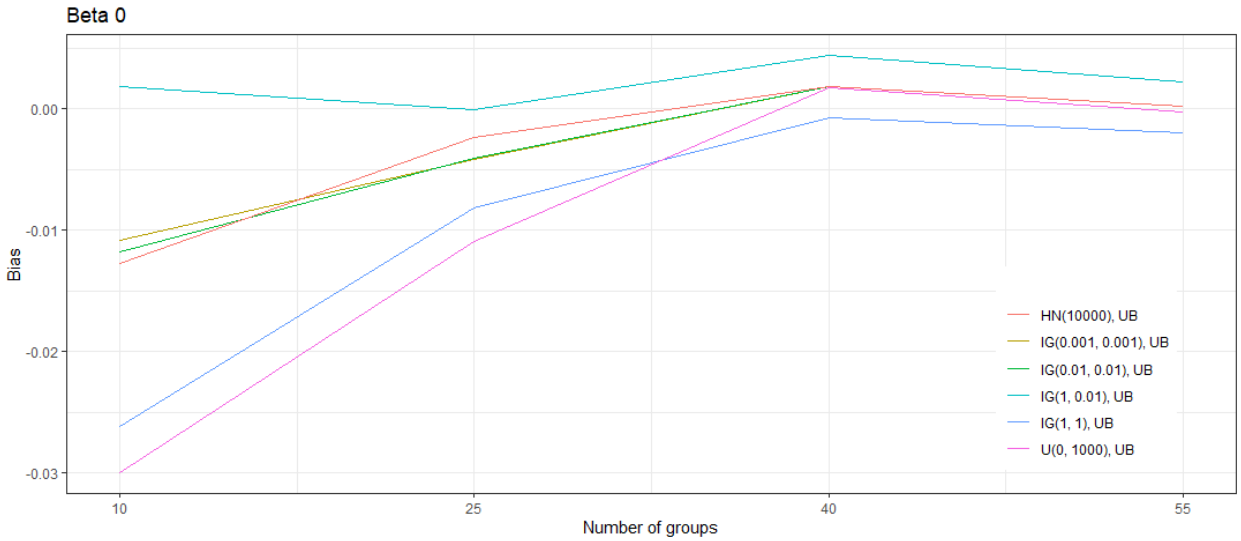


Figure 5.3: Bias of β_0 as a function of the unbalanced study design across the 6 priors.

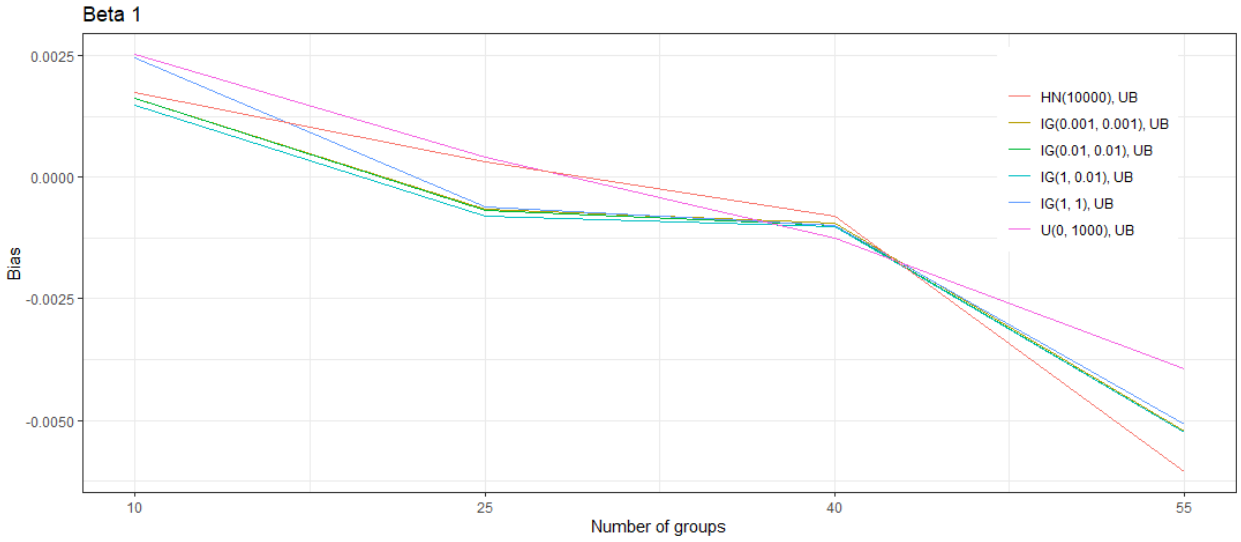


Figure 5.4: Bias of β_1 as a function of the unbalanced study design across the 6 priors.

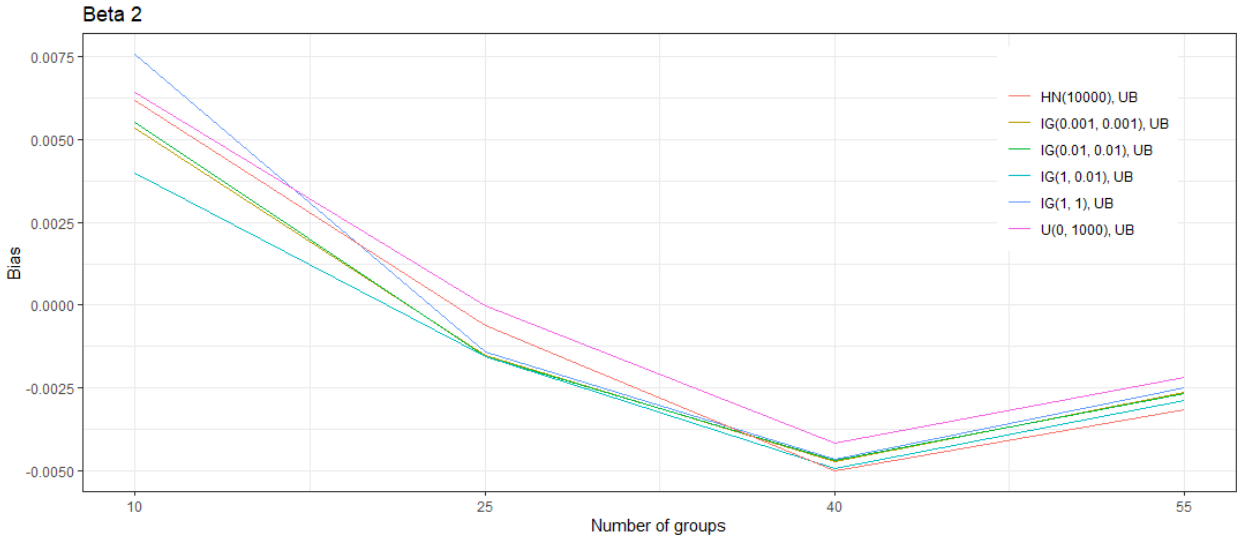


Figure 5.5: Bias of β_2 as a function of the unbalanced study design across the 6 priors.

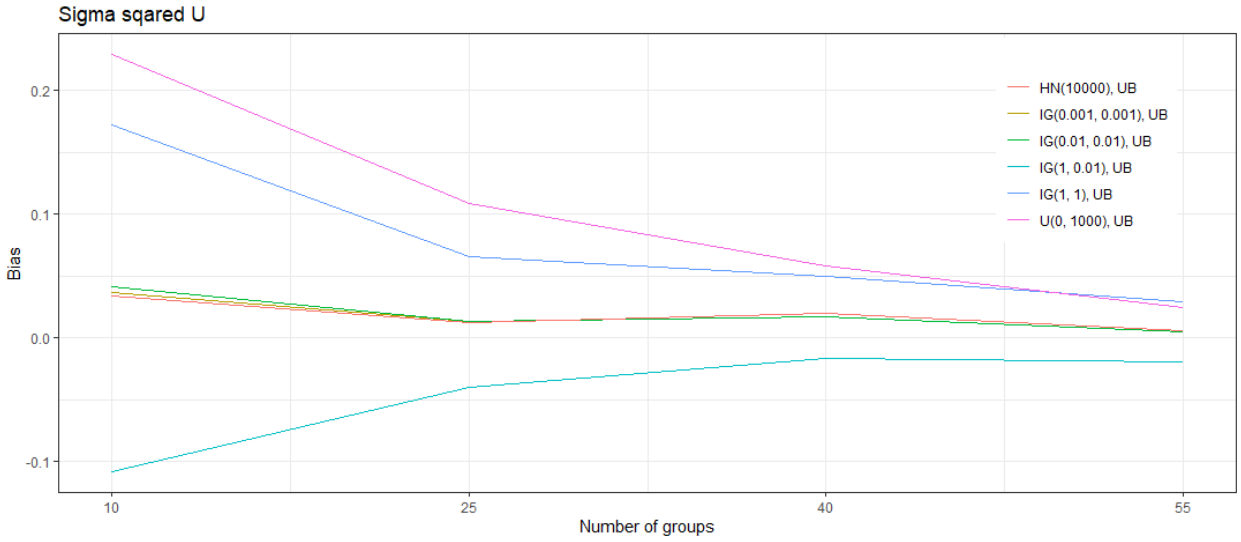


Figure 5.6: Bias of σ_u^2 as a function of the unbalanced study design across the 6 priors.

5.3 Covariance-components prior models

There are a number of potential issues with using the Inverse-Wishart distribution as a prior for covariance matrices in Bayesian hierarchical modelling. This subsection discusses a few of these issues and puts forward two alternative priors that can be used for covariance matrices

in Bayesian hierarchical models, such as the multivariate MMMC model described in model (4.4). The two alternatives are the scaled Inverse-Wishart prior and the hierarchical half-t prior and have been discussed in Alvarez et al. (2016). The scaled Inverse-Wishart prior for an $R \times R$ covariance matrix is given by

$$\begin{aligned} \mathbf{\Sigma} &= \text{diag}(\delta_1, \ldots, \delta_R)\ \mathbf{Q}\ \text{diag}(\delta_1, \ldots, \delta_R) \\ \delta_r &\sim \text{Lognormal}(\alpha_r, \omega_r^2) \quad \text{for } r = 1, \ldots, R \\ \mathbf{Q} &\sim \text{Inverse-Wishart}(\xi, \mathbf{\Omega}), \end{aligned} \tag{5.1}$$

where $\alpha_r = 0\ \forall r, \omega_r^2 = 100\ \forall r, \xi = R + 1, \mathbf{\Omega} = \mathbf{I}_{R\times R}$. These choices are based on guidance from Alvarez et al. (2016). This prior employs a separation strategy to construct the covariance matrix $\mathbf{\Sigma} = \text{diag}(\delta_1, \ldots, \delta_R)\ \mathbf{Q}\ \text{diag}(\delta_1, \ldots, \delta_R)$, where $\mathbf{Q}$ is a covariance matrix whose diagonal entries are independently scaled by a squared lognormal random variable δ_r^2 and the off-diagonals entries are scaled by a product of lognormal random variables $\delta_r \delta_{r'}$ $(r \neq r^{'})$. The purpose of these scaling procedures is to provide greater flexibility in the resultant elements of the covariance matrix which would otherwise be $\mathbf{Q}$. Alternatively, the hierarchical half-t prior for an $R \times R$ covariance matrix is given by

$$\begin{aligned} \mathbf{\Sigma} &\sim \text{Inverse-Wishart}(\phi + R - 1, 2\phi\Delta) \\ \Delta &= \text{diag}(\lambda_1, \ldots, \lambda_R) \\ \lambda_r &\sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{\xi_r^2}\right) \quad \text{for } r = 1, \ldots, R, \end{aligned} \tag{5.2}$$

where $\xi_r^2 = 1\ \forall r$ and $\phi = 2$. This prior implies a half-t distribution on the standard deviations, which is based on a result given in Wand et al. (2011) (Result 5) that states

that the half-t distribution can be written as a hierarchy of Inverse-Gamma distributions. Namely that

$$\begin{aligned} \text{if} \quad & \sigma^2|\lambda \sim \text{Inverse-Gamma}(\phi/2, \phi/\lambda) \text{ and } \lambda \sim \text{Inverse-Gamma}(1/2, 1/\xi^2), \\ \text{then} \quad & \sigma \sim \text{Half-}t(\xi, \phi). \end{aligned} \tag{5.3}$$

This results in $\sigma_r \sim$ Half-$t(1, 2)$ $\forall r$ for the prior in (5.2). This is clear to see by noting that in (5.2) $\sigma_r^2|\lambda_r \sim$ Inverse-Gamma$(\phi/2, 2\phi\lambda_r/2)$ and $\lambda_r \sim$ Gamma$(1/2, 1/\xi_r^2)$ $\forall r$, so $\sigma_r^2|\lambda_r \sim$ Inverse-Gamma$(\phi/2, \phi/\lambda_r)$ and $\lambda_r \sim$ Inverse-Gamma$(1/2, 1/\xi_r^2)$ $\forall r$. The use of half-t priors on standard deviation parameters is recommended in Gelman (2006) to achieve a weakly informative prior.

5.3.1 Simulation study

Each simulation was performed using 100 data sets. The data sets are such that $N = 1000$, $R = 3$, $K = 20$, and $J = 20$. The actual design matrix, allocation of individuals in schools matrix, and weight matrices are the same for all data sets and were specified to be similar to the real data. Each model was fit for a total of 20,000 iterations, with a burn-in period of 20,000 and thinning of 1. Summaries of the 95% coverage probability, effective sample size (ESS), bias for the posterior mean estimates, and RMSE for the posterior mean estimates are presented in Table 5.2. The summaries regarding $\mathbf{v}$ and $\mathbf{u}$ are averaged over all R, K and J elements for each response respectively. The variance components for each of the covariance matrices were chosen to take the value 1.5, with the purpose being to highlight the issue that the unscaled Inverse-Wishart prior could theoretically produce and was previously discussed.

Table 5.2 gives the summaries of the simulations. Table 5.2 shows that the biases of the posterior mean estimates for each parameter is relatively small over the 100 simulations and

	Inverse-Wishart Prior				Scaled Inverse-Wishart Prior				Hierarchical Half-t Prior			
	Posterior mean				Posterior mean				Posterior mean			
Parameter	Bias	RMSE	Coverage	$\overline{\text{ESS}}$	Bias	RMSE	Coverage	$\overline{\text{ESS}}$	Bias	RMSE	Coverage	$\overline{\text{ESS}}$
β_{11}	0.03	0.08	0.92	1378	0.03	0.08	0.92	1404	0.03	0.08	0.92	1418
β_{12}	0.01	0.07	0.95	1495	0.01	0.08	0.95	1510	0.01	0.08	0.94	1524
β_{13}	−0.05	0.09	0.92	1163	−0.04	0.09	0.94	1180	−0.04	0.09	0.93	1187
β_{21}	0.02	0.10	0.91	1451	0.02	0.10	0.91	1453	0.02	0.10	0.91	1453
β_{22}	0.00	0.08	0.93	1537	0.00	0.08	0.92	1523	0.00	0.08	0.92	1552
β_{23}	−0.05	0.10	0.91	1149	−0.04	0.10	0.94	1141	−0.04	0.10	0.95	1168
β_{31}	−0.02	0.09	0.91	1401	−0.01	0.09	0.91	1406	−0.01	0.09	0.91	1396
β_{32}	0.04	0.09	0.93	1261	0.03	0.09	0.94	1270	0.03	0.09	0.94	1272
β_{33}	0.00	0.07	0.97	1362	0.00	0.07	0.98	1374	0.00	0.07	0.98	1352
$\mathbf{\Sigma_{v}}_{11}$	−0.12	0.50	0.90	6147	0.14	0.61	0.94	3111	0.07	0.40	0.99	13410
$\mathbf{\Sigma_{v}}_{12}$	0.00	0.38	0.93	9217	0.00	0.39	0.94	10636	−0.05	0.35	0.94	10999
$\mathbf{\Sigma_{v}}_{13}$	0.05	0.35	0.94	9528	0.06	0.35	0.99	10942	0.00	0.31	0.99	11269
$\mathbf{\Sigma_{v}}_{22}$	−0.07	0.57	0.94	6289	0.20	0.70	0.93	3110	0.11	0.47	1.00	13320
$\mathbf{\Sigma_{v}}_{23}$	0.03	0.39	0.91	9606	0.04	0.40	0.94	11035	−0.02	0.36	0.94	11279
$\mathbf{\Sigma_{v}}_{33}$	0.00	0.50	0.93	6522	0.28	0.67	0.96	3283	0.16	0.43	1.00	12822
$\mathbf{\Sigma_{u}}_{11}$	−0.10	0.54	0.91	5648	0.17	0.65	0.96	3035	0.03	0.45	0.98	8712
$\mathbf{\Sigma_{u}}_{12}$	−0.03	0.41	0.88	8315	−0.03	0.41	0.94	9805	−0.08	0.38	0.92	10064
$\mathbf{\Sigma_{u}}_{13}$	0.00	0.39	0.91	8142	0.00	0.39	0.94	9534	−0.05	0.35	0.95	9953
$\mathbf{\Sigma_{u}}_{22}$	−0.12	0.49	0.90	5611	0.16	0.61	0.96	2965	0.06	0.39	0.99	12089
$\mathbf{\Sigma_{u}}_{23}$	−0.03	0.37	0.92	8206	−0.03	0.37	0.97	9663	−0.08	0.34	0.97	10115
$\mathbf{\Sigma_{u}}_{33}$	−0.12	0.55	0.88	5470	0.14	0.67	0.94	3025	0.07	0.43	0.99	12574
$\mathbf{v}$	0.00	0.35	0.94	3904	0.00	0.35	0.94	3954	0.00	0.35	0.94	4082
$\mathbf{u}$	0.00	0.38	0.93	3857	0.00	0.38	0.94	3914	0.00	0.38	0.94	4088

20,000 burn-in, thin = 1, and 20,000 post burn-in samples.

Table 5.2: Summary results from the simulation study.

each prior (**A1**). In addition to this, for each prior, the 95% coverage probability for each parameter is roughly 0.95 (**A1**). The boxplots of the ESS for the set of parameters in the covariance matrix for the school random effects under the three priors are shown in Figures 5.7a, 5.7b, 5.7c, 5.7d, 5.7e, and 5.7f. They show that, for variance components shown in Figures 5.7a, 5.7d, and 5.7f, the hierarchical half-t distribution resulted in larger ESSs when compared to the Inverse-Wishart and the scaled Inverse-Wishart priors (**A2**). It is also worth pointing out that the hierarchical half-t prior achieves this while also producing similar biases to the Inverse-Wishart and the scaled Inverse-Wishart priors for each parameter.

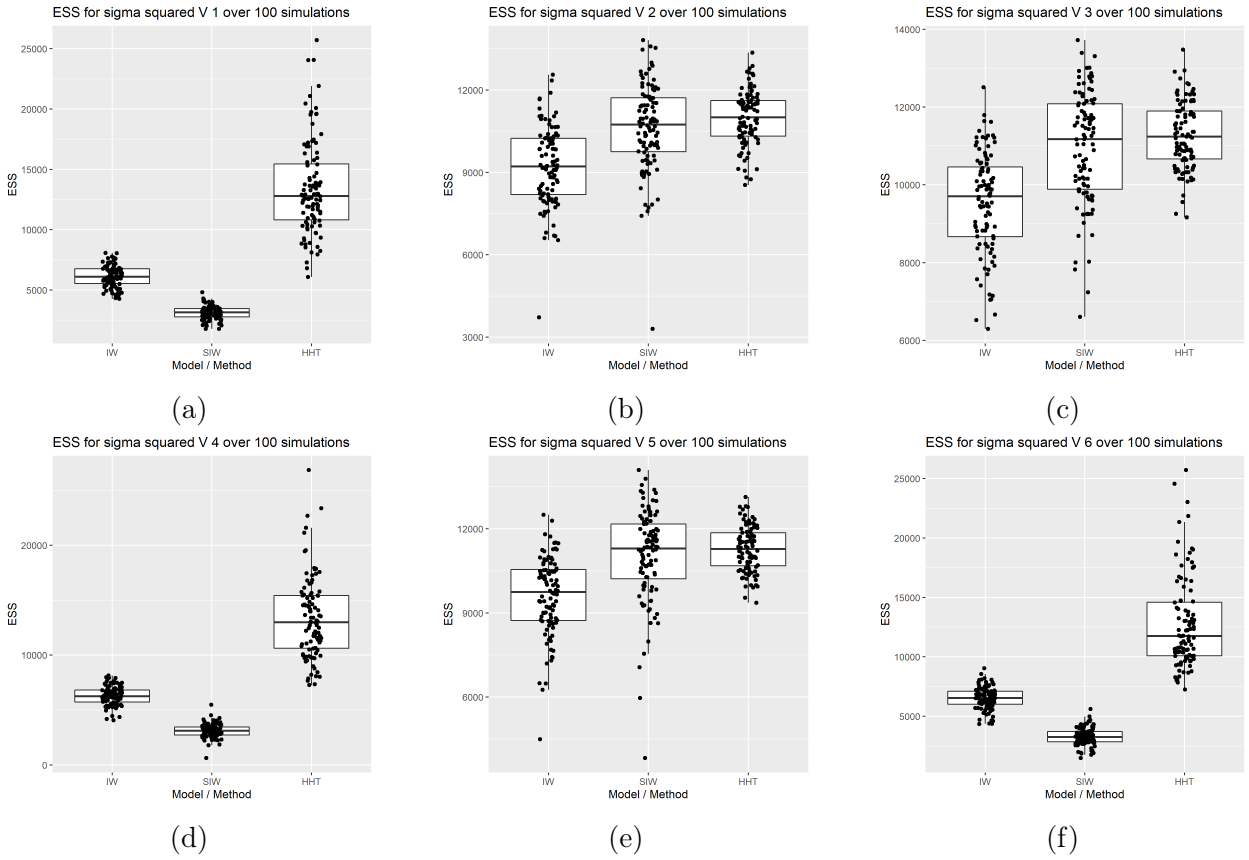
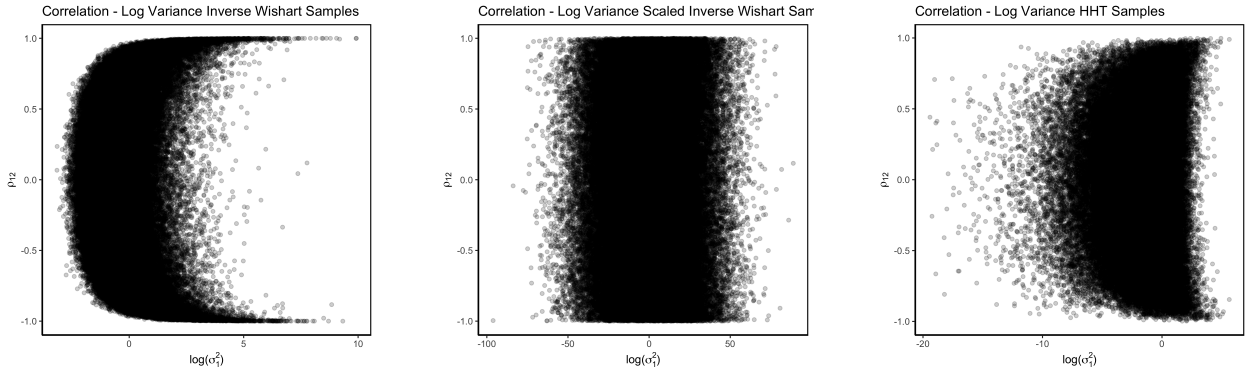


Figure 5.7: Boxplots of the ESSs for the 100 simulations for the parameters in $\mathbf{\Sigma_v}$ across the three different priors, $\mathbf{\Sigma_{v}}_{11}$ (a), $\mathbf{\Sigma_{v}}_{12}$ (b), $\mathbf{\Sigma_{v}}_{13}$ (c), $\mathbf{\Sigma_{v}}_{22}$ (d), $\mathbf{\Sigma_{v}}_{23}$ (e), and $\mathbf{\Sigma_{v}}_{33}$ (f).

The Inverse-Wishart distribution is a family of distributions that are parameterised by a scale matrix $\mathbf{\Omega}$ and a degrees of freedom ξ, which are selected to be $\xi = R + 1$ and $\mathbf{\Omega} = \mathbf{I}_{R\times R}$. A potential issue with using the Inverse-Wishart prior is that it results in all variance parameters being controlled by the single degree of freedom parameter. The Inverse-Wishart prior has the drawback that it implies a scaled inverse chi-squared prior for each variance $\sigma_r^2 \sim \text{inv-}\chi^2(\xi - R + 1, \frac{\mathbf{\Omega}_{rr}}{\xi - R + 1})$. Thus the implied scaled inverse chi-squared prior provides a prior on variance components that have very low mass for values close to 0. As a result, the posterior summaries, such as the posterior mean, of the variance parameters may be overestimated when their true values are small. In addition to this, using the Inverse-Wishart distribution as a prior on the covariance matrix in a hierarchical model induces a prior dependency on the variance and correlation parameters (**A2**). As shown

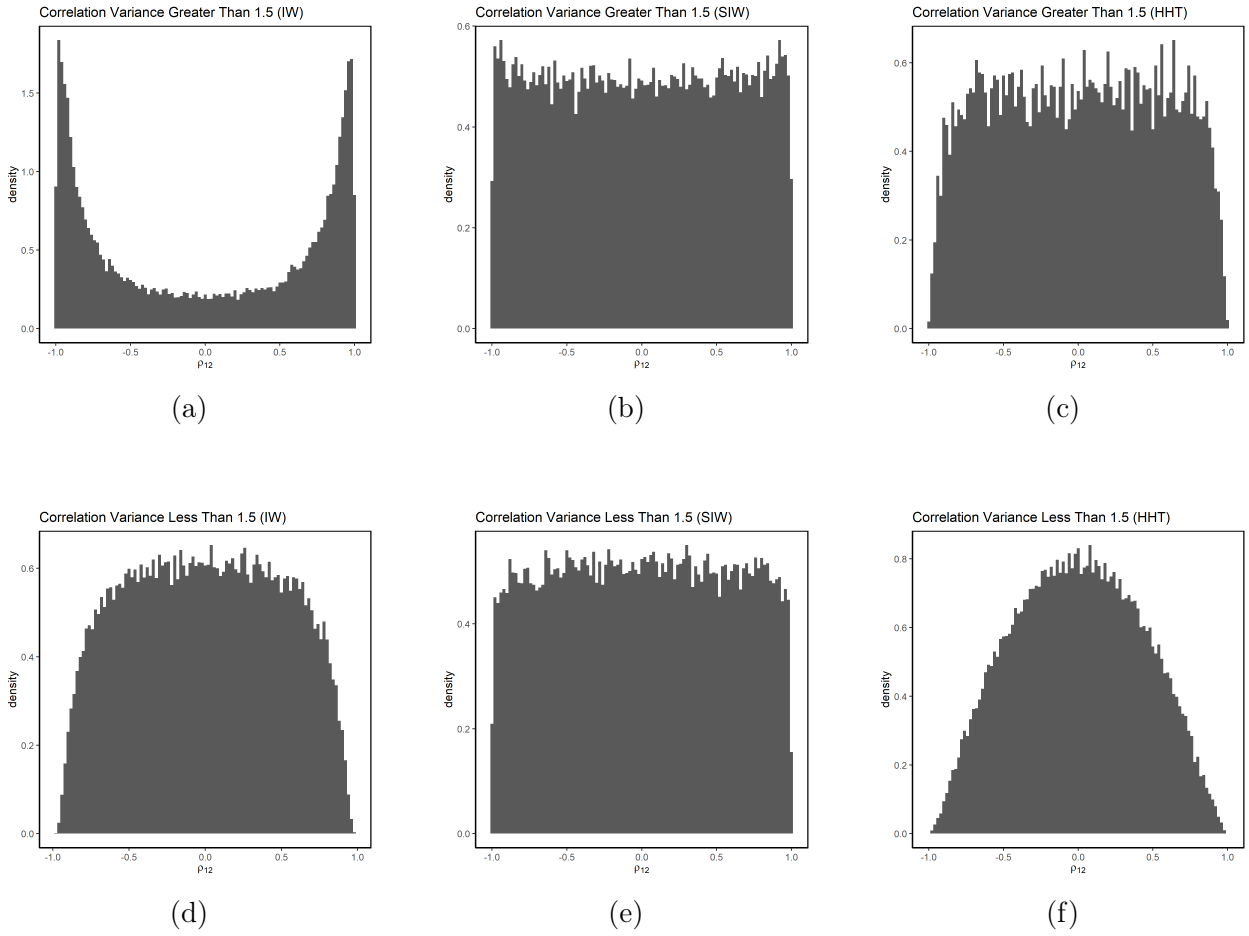
in Figure 5.8a, there is a distinct prior relationship between $\log(\sigma_1^2)$ and ρ_{12} induced by the Inverse-Wishart(4, $\mathbf{I}_{R\times R}$), with larger values of $\log(\sigma_1^2)$ appearing to induce correlations with large absolute values. This behaviour also persisted for the Inverse-Wishart(4, $3\mathbf{I}_{R\times R}$) prior distribution, where the scale matrix is not the identity matrix. On further inspection, defining a large variance as a variance that is greater than or equal to 1.5 and a small variance being less than 1.5, Figures 5.9a and 5.9d show that large variances are associated with correlations that have large absolute values and smaller variance are associated with correlations that have small absolute values.



(a) (b) (c)

Figure 5.8: Scatter plots of ρ_{12} against $\log(\sigma_1^2)$ from 100,000 samples of Inverse-Wishart(4, $\mathbf{I}_{R\times R}$) (a), the scaled Inverse-Wishart described in (5.1) (b), and the hierarchical half-t described in (5.2) (c).

In comparison, as shown in Figure 5.8b, the scaled Inverse-Wishart given above does not appear to associate a distinct relationship between $\log(\sigma_1^2)$ and ρ_{12}. In contrast, to the Inverse-Wishart, Figures 5.9b and 5.9d show that large and small variances are associated with correlations that evenly span the range [-1, 1]. Similarly, as shown in Figure 5.8c, the hierarchical half-t given above does not appear to associate a distinct relationship between $\log(\sigma_1^2)$ and ρ_{12}. Figures 5.9c and 5.9f show that large and small variances are associated with correlations that unimodal about 0, with neither greatly favouring correlations with a value of -1 or 1.



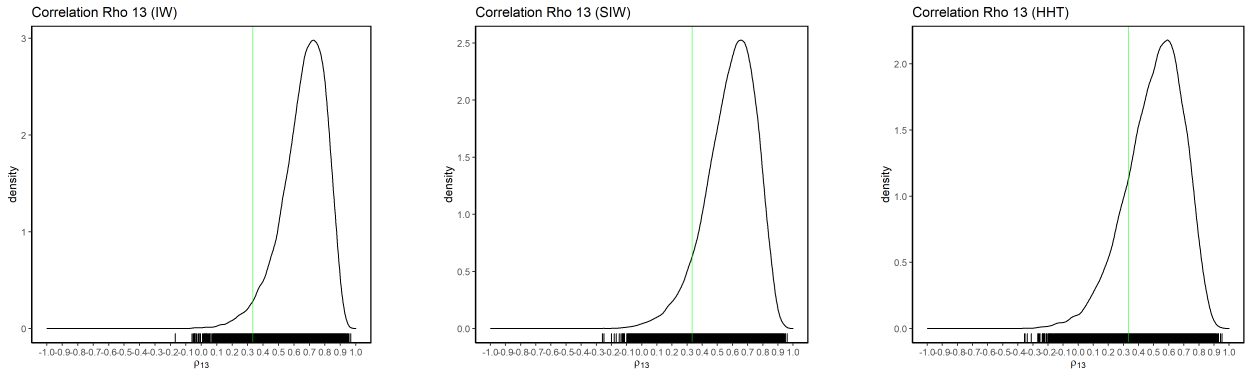
(a) (b) (c)

(d) (e) (f)

Figure 5.9: Histograms of 100,000 samples presented in Figures 5.8a, 5.8b, and 5.8c partitioned by whether $\sigma_1^2 \geq 1.5$ or $\sigma_1^2 < 1.5$. Inverse-Wishart$(4, \mathbf{I}_{R\times R})$ when $\sigma_1^2 \geq 1.5$ (a) and $\sigma_1^2 < 1.5$ (d). The scaled Inverse-Wishart described in (5.1) when $\sigma_1^2 \geq 1.5$ (b) and $\sigma_1^2 < 1.5$ (e). The hierarchical half-t described in (5.2) when $\sigma_1^2 \geq 1.5$ (c) and $\sigma_1^2 < 1.5$ (f).

Figure 5.10a, 5.10b, and 5.10c show the density plot of posterior samples for ρ_{13} under the Inverse-Wishart, scaled Inverse-Wishart, and hierarchical half-t prior for one of the 100 simulations. Unsurprisingly, Figure 5.10a shows that the Inverse-Wishart prior leads to an overestimation for the absolute value of ρ_{13} when the true value of $\mathbf{\Sigma_v}_{11} = 1.5$ and $\mathbf{\Sigma_v}_{33} = 1.5$, which may be classified as large. In comparison, as to be expected, the scaled Inverse-Wishart and hierarchical half-t priors do not overestimate the absolute value of ρ_{13}. As the correlation is a normalized form of covariance and not affected by scale, it is usually the parameter of

interest when it comes to measuring the linear relationship between variables (**A2**). Thus, the hierarchical half-t prior may be preferred to the Inverse-Wishart and scaled Inverse-Wishart priors when modelling covariance structures in a multivariate Bayesian hierarchical MMMC model due its desirable properties and the increased ESSs that it results in for variance parameters.



(a) (b) (c)

Figure 5.10: Density plot of samples of ρ_{13} under the Inverse-Wishart (a), scaled Inverse-Wishart (b), and hierarchical half-t (c) prior for one of the 100 simulations. The green vertical line represents the true value of $\rho_{13} = 1/3$.

5.4 Discussion and conclusions

This chapter has investigated the use of different priors across a range of different models to see how a change in prior for the variance-components and covariance components in models previously presented in earlier chapters of this thesis may induce different properties for the parameters in the model. The results show that under certain conditions, the choice of prior distributions for variance and covariance parameters can have a massive impact on the estimation of model parameters. Thus, it is of great importance for a user to carefully consider the choice of prior that they plan to use in their Bayesian models. In addition to this, this chapter has also shown how the structure of data pertaining to random effects in

the model can play a significant part in how the estimates of a parameter behave.

Chapter 6

Multivariate spatial MMMC models

6.1 Introduction

Bayesian spatial modelling is a technique that seeks to account for the underlying spatial trend created by an underlying process which may otherwise be ignored by non-spatial modelling approaches. The literature on multiple health behaviours suggests that there may be a dependence between an individual's behaviours and where they live. Thus, multivariate spatial models may be more appropriate to use than non-spatial models in such settings. The spatial effect that living in certain areas has on an individual's alcohol consumption has been observed across the regions of England (Shelton and Savell (2011); Castillo et al. (2017)). Shelton and Savell (2011) found that, in comparison to males living in London, males living in the North East had a larger observed odds of binge drinking in the past week (OR = 2.85, 95% CI: (1.72, 4.73)). Similarly, Castillo et al. (2017) found that, in comparison to males living in the North East, males living in London had a lower observed odds of binge drinking (OR = 0.418, 95% CI: (0.313, 0.557)). The spatial effect that living in certain areas has on an individual's smoking habits has also been observed across the regions of England. Beard et al. (2017) found that, in comparison to individuals living in the South West, individuals

living in the East Midlands had a lower observed relative risk of being a smoker (RR = 0.86, 95% CI: (0.79, 0.94)).

In this chapter, I build on Chapter 4 by modelling the multiple health behaviours of adolescents with the use of a proposed multivariate spatial MMMC model. The multivariate spatial MMMC model extends the multivariate MMMC model proposed in Chapter 4 by modelling spatial information through the use of conditional autoregressive random effects.

The remainder of the chapter is structured as follows. Subsection 6.1.1 outlines the novel contributions made. Section 6.2 provides an exploration of the spatial component of the network data used, and acts as a form of guided inspiration for the multivariate spatial MMMC models proposed in this chapter. Section 6.3 describes the multivariate spatial MMMC models proposed in this chapter. Section 6.4 details the Markov chain Monte Carlo algorithm used for the aforementioned models in this chapter. Section 6.5 presents an application of the models discussed in this chapter on network data of multiple health behaviours of adolescents in Los Angeles, California. Section 6.6 concludes with a discussion.

6.1.1 Novel Contributions

There is a great interest in understanding the underlying process of data with both a network and spatial component. This chapter provides a distinct novel contribution to the literature on spatial and network analysis which is given as follows

1. **Multivariate spatial MMMC models.** A potential limitation of the multivariate MMMC model proposed in Chapter 4 is that it does not model the potential spatial trend in data. In instances such as the one being dealt with in this chapter, the data may contain a spatial component that we wish to model. As a result, it may be reasonable to propose a multivariate MMMC models that takes account of the spatial and network component of the data. Section 6.3 describes the multivariate model that

is to be implemented in Section 6.5.

6.2 Data: Social Networking Survey

This section builds on the data description provided in the previous section and explores the spatial aspect of the Social Networking Survey (SNS) network data from Los Angeles, California.

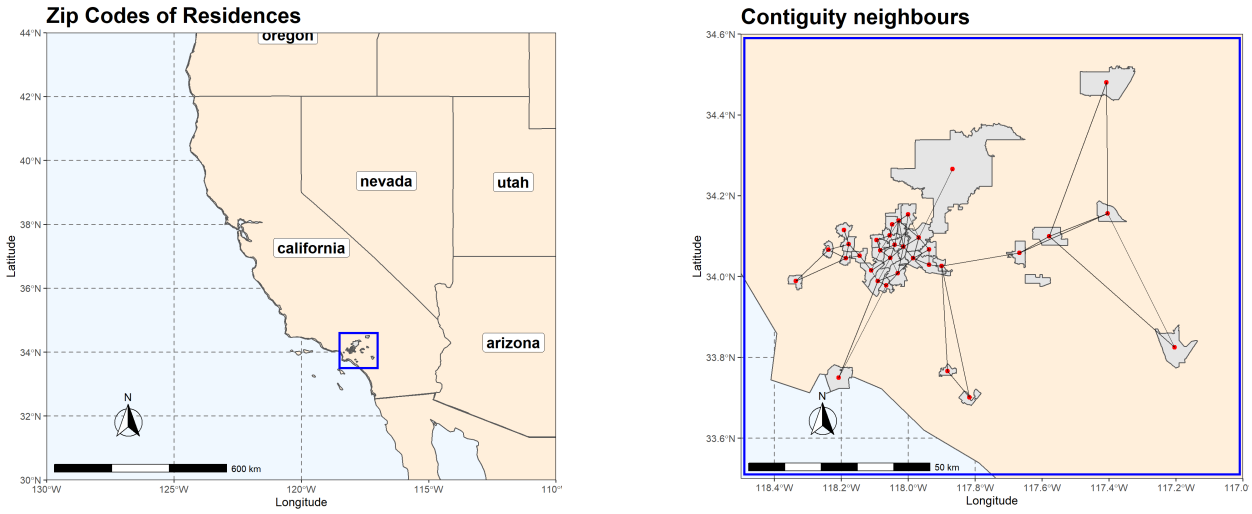
6.2.1 Spatial data exploration

The adolescents surveyed collectively reside in $S = 33$ non-overlapping administrative areas known as Zip Codes, which contain very unequal numbers of survey responders. For example, the two Zip Codes with the largest number of adolescents surveyed are El Monte (91732 - 378 individuals) and South El Monte (91733 - 271 individuals), while there are 19 instances in which only 1 adolescent surveyed resides in the Zip Code. The spatial configuration of the $S = 33$ Zip Codes is displayed in the right panel of Figure 6.1, which shows that while most of the Zip Codes are grouped together in the middle of the region, there are a small number of isolated Zip Codes that are not close to the remaining ones.

The spatial closeness between each pair of Zip Codes is encoded in the model described in the next section by a binary neighbourhood matrix denoted $\mathbf{A}_{33\times 33}$, where the ijth element $a_{ij} = 1$ if Zip Codes (i, j) share a common border and $a_{ij} = 0$ otherwise (and $a_{ii} = 0$ for all i). This *border sharing* specification is the most commonly used neighbourhood matrix in spatial areal unit modelling (see for example Bivand et al., 2013 and Jack et al., 2019), because of its sparsity and simplicity of construction (e.g. it does not have a tuning parameter as the k-nearest neighbours rule does). However, Zip Codes that are isolated share no neighbours under this definition, which means the conditional autoregressive prior outlined in the next section for capturing the spatial correlation has improper full conditional distributions for

these Zip Codes. Thus we make the commonly used adjustment to $\mathbf{A}$ for each isolated Zip Code i to rectify this problem, which is to make them a neighbour of the Zip Code j that is geographically closest (e.g. set $a_{ij} = a_{ji} = 1$). The final neighbourhood structure assumed when fitting the model is displayed by the connecting lines in the right panel of Figure 6.1, which shows that under this specification the Zip Codes comprise a single connected graph.

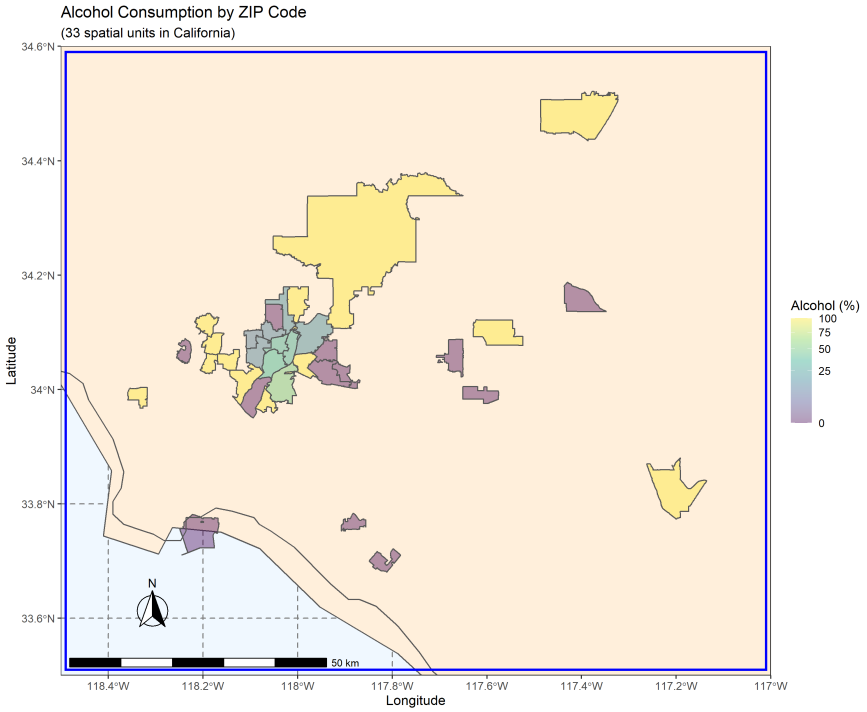
We assessed in an exploratory manner whether there are likely to be spatial effects in the data, i.e. whether adolescents in different Zip Codes have differing propensities for partaking in adverse health behaviours. We do this by computing the empirical probability of engaging in a specific health behaviour given that the individual is from a certain Zip Code. However, as previously discussed the distribution of individuals to the 33 Zip Codes is highly skewed, with a minimum, 1st quartile, median, 3rd quartile, and maximum of 1, 1, 1, 6, and 378 individuals respectively. Thus for a meaningful comparison we only consider the 5 Zip Codes containing the most individuals, which are Temple City (91780 - 31 individuals), El Monte (91731 - 154 individuals), Rosemead (91770 - 167 individuals), South El Monte (91733 - 271 individuals), and El Monte, (91732 - 378 individuals). For these Zip Codes the observed probabilities of consuming at least one drink of alcohol in the past 30 days are 0.16, 0.27, 0.15, 0.33, and 0.28 respectively. The corresponding probabilities of smoking at least one cigarette in the past 30 days are 0.13, 0.36, 0.17, 0.33, and 0.29 respectively, while for marijuana the probabilities are 0.16, 0.27, 0.15, 0.33, and 0.28. As these empirical probabilities show some variation by Zip Code, spatial effects are a plausible component to include in the model.



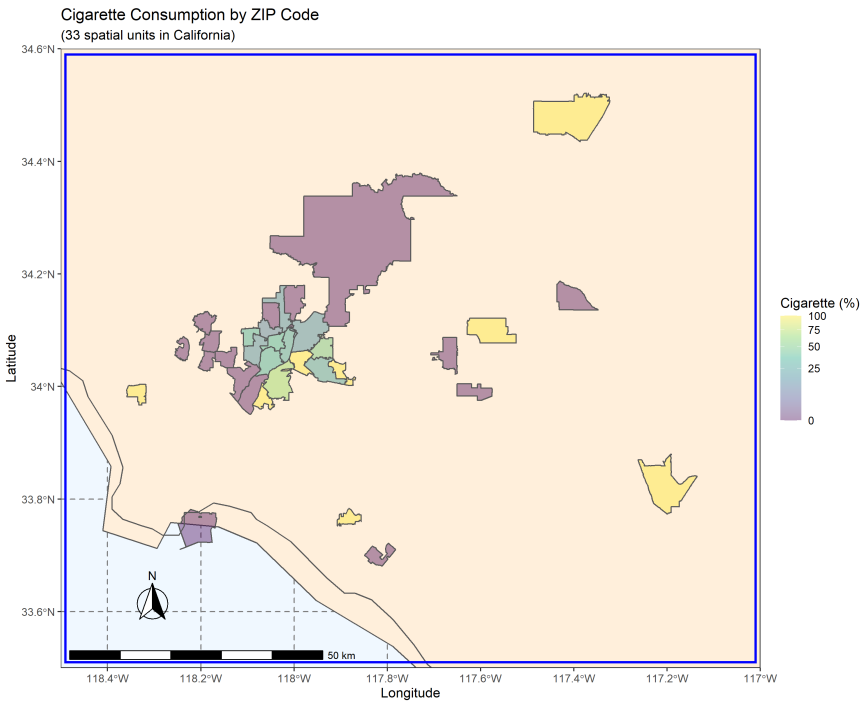
(a) (b)

Figure 6.1: Maps of California (a) and the spatial configuration of the Zip Codes (b). In the latter the lines denote the neighbour relationships between two Zip Codes assumed when fitting the model.

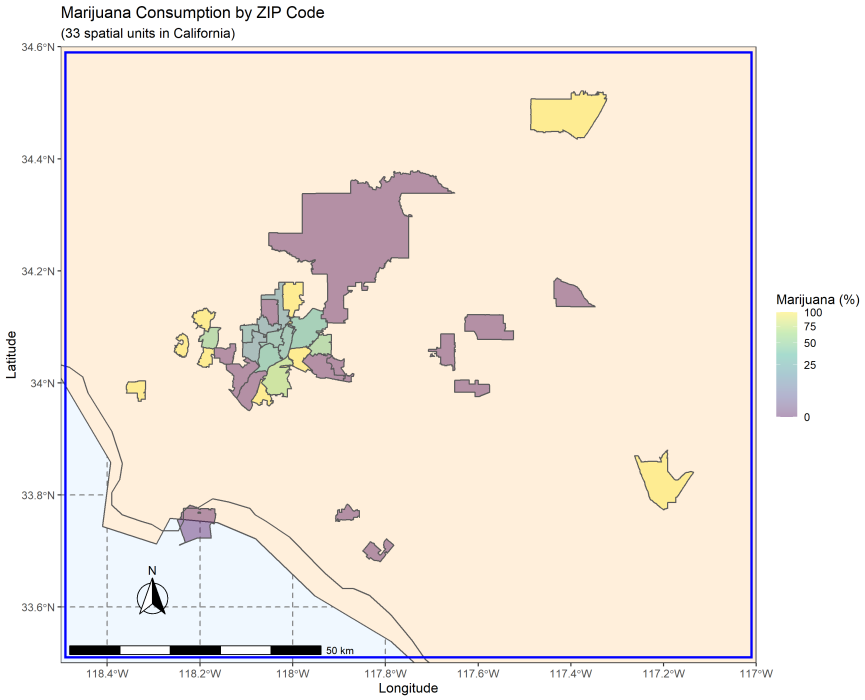
Figure 6.2 displays the percentages of adolescents who responded "yes" to whether they consumed at least one drink of alcohol in the past 30 days (Figure 6.2a), had smoked at least one cigarette in the past 30 days (Figure 6.2b), and had ever tried marijuana (Figure 6.2c) in each areal unit.


Alcohol Consumption by ZIP Code
(33 spatial units in California)
34.6°N
34.4°N
34.2°N
34°N
33.8°N
33.6°N
Latitude
N
50 km
118.4°W
118.2°W
118°W
117.8°W
117.6°W
117.4°W
117.2°W
117°W
Longitude
Alcohol (%)
100
75
50
25
0

(a)


Cigarette Consumption by ZIP Code
(33 spatial units in California)
34.6°N
34.4°N
34.2°N
34°N
33.8°N
33.6°N
Latitude
N
50 km
118.4°W
118.2°W
118°W
117.8°W
117.6°W
117.4°W
117.2°W
117°W
Longitude
Cigarette (%)
100
75
50
25
0

(b)



(c)

Figure 6.2: Map displaying the percentage of adolescents who responded "yes" to the alcohol (a), cigarette (b), and marijuana (c) related responses.

6.3 Multivariate spatial uniplex MMMC models

This section describes the multivariate spatial MMMC model to be used on network data with a spatial aspect.

6.3.1 The multivariate Leroux CAR uniplex MMMC model

As the adverse health behaviours are binary no/yes outcomes, the data likelihood model has a Bernoulli logistic regression form and is given by

$$\begin{aligned} Y_{i_s r} &\sim \text{Bernoulli}(\theta_{i_s r}) \quad \text{for } i = 1, \ldots, N_s,\ s = 1, \ldots, S,\ r = 1, \ldots, R, \\ \ln\left(\frac{\theta_{i_s r}}{1-\theta_{i_s r}}\right) &= \mathbf{x}_{i_s}^{\top}\boldsymbol{\beta}_r + \phi_{sr} + \sum_{j \in \text{net}(i_s)} w_{i_s j} u_{jr} + w_{i_s}^{*} u_r^{*}. \end{aligned} \tag{6.1}$$

Here $Y_{i_s r}$ denotes the binary ($Y_{i_s r} = 1$ - yes; $Y_{i_s r} = 0$ - no) adverse health behaviour for the rth response for the ith individual who lives in the sth spatial unit, where $r = 1, \ldots, R(= 3)$, $s = 1, \ldots, S(= 33)$ and $i = 1, \ldots, N_s$. The probability that individual i from spatial unit s partakes in adverse health behaviour r is denoted by $\theta_{i_s r}$, which is modelled on the logit scale by three separate components. The first is a $p \times 1$ vector of covariates $\mathbf{x}_{i_s}$, which is accompanied by a $p \times 1$ vector of fixed effect regression parameters $\boldsymbol{\beta}_r$ that vary by health behaviour r. The prior for these fixed effect parameters is given by $\boldsymbol{\beta}_r \sim \text{N}(\mathbf{0}, 100000\mathbf{I})$ independently for each outcome r, where $\mathbf{I}$ is the $p \times p$ identity matrix. This specification is chosen to be weakly informative, thus allowing the parameter estimates to be largely informed by the data. The other two components in the systematic part of the model are a friendship network effect and a spatial effect, and these two different levels in the model are described below. The friendship network effect captures correlations between the three health behaviours and between individuals (friends), while the spatial effect capture correlations between neighbouring Zip Codes.

6.3.2 Friendship network effects

Friendship network effects are accounted for in the model by the $\sum_{j\in\text{net}(i_s)} w_{i_sj}u_{jr} + w^*_{i_s}u^*_r$ component in Equation (6.1), where $\text{net}(i_s)$ denotes the set of individuals that the ith individual in the sth spatial unit nominates as friends. The latter part u^*_r is an *isolation effect* for health behaviour r, which is an effect for individuals who don't nominate any friends. This is achieved by setting $w^*_{i_s} = 1$ if individual i_s nominates no friends and $w^*_{i_s} = 0$ otherwise, and if $w^*_{i_s} = 1$ then clearly $\sum_{j\in\text{net}(i_s)} w_{i_sj}u_{jr} = 0$ as $\text{net}(i_s)$ is the empty set. The $\sum_{j\in\text{net}(i_s)} w_{i_sj}u_{jr}$ component is known as a multiple membership model, and was proposed by Browne et al. (2001). The friendship network structure for each individual is discussed in Section 4.2.2 and displayed visually in Figure 4.1, but for the purposes of the model is encoded into an $n \times n$ friendship network matrix $\mathbf{W} = (w_{i_sj})$, where $n = \sum_{s=1}^{S} N_s$ denotes the number of individuals in the survey. The values of the entries in this friendship network matrix are given by

$$w_{i_sj} = \begin{cases} 1/|\text{net}(i_s)| & \text{if individual } i \text{ in spatial unit } s \text{ nominates individual } j \text{ as a friend} \\ 0 & \text{Otherwise} \end{cases}, \tag{6.2}$$

where $|\text{net}(i_s)|$ is the cardinality of the set $\text{net}(i_s)$. Thus, the only non-zero entries in this matrix relate to friendships that one individual has with another individual. Note, this matrix is not necessarily symmetric because it represents a directed rather than an undirected graph. The values of the non-zero elements in this matrix are the reciprocal of the number of friends each individual nominates, which ensures that the matrix is row standardised (each row sums to 1).

Then for individual i in spatial unit s the friendship network component of the model contributes the term $\sum_{j\in\text{net}(i_s)} w_{i_sj}u_{jr}$ to the linear predictor. Here u_{jr} is a random effect

representing the peer effect that person j has on their friends in terms of partaking in health behaviour r. Thus $\sum_{j\in\text{net}(i_s)} w_{i_sj}u_{jr}$ simply represents the average (mean) effect that the friends of individual i in spatial unit s have on that individual in terms of partaking in health behaviour r. We allow these friendship random effects $\mathbf{u}_j = (u_{1j}, \ldots, u_{Rj})_{R\times 1}$ to be correlated across the different responses because we believe that if an individual encourages others to drink alcohol for example then they may also encourage others to smoke cigarettes or marijuana as well. Thus we specify the following multivariate normal prior distribution for each individual's friendship random effects $\mathbf{u}_j$:

$$\begin{aligned} \mathbf{u}_j &\sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\ \mathbf{u}^* &\sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\ \boldsymbol{\Sigma} &\sim \text{Inverse-Wishart}(4, \mathbf{I}). \end{aligned} \tag{6.3}$$

Between health behaviour correlation is allowed for each $\mathbf{u}_j$ and $\mathbf{u}^*$ via the $R \times R$ covariance matrix $\boldsymbol{\Sigma}$, which is assigned a weakly informative Inverse-Wishart prior distribution, which allows the estimation to mainly be informed by the data. We assume that $\mathbf{u}_j$ and $\mathbf{u}^*$ share the same covariance matrix $\boldsymbol{\Sigma}$ for convenience.

6.3.3 Spatial effects

The underlying process of the data suggests there may be spatial effects in the data, which we model using spatially correlated random effects. These random effects are assigned a conditional autoregressive (CAR) prior distribution, which is the most common approach to modelling spatial correlation in areal unit data (see for example Banerjee et al., 2004). As our adverse health behaviour response is multivariate, a multivariate CAR model could be adopted to model both spatial and between health behaviour correlations. Multivariate

CAR type models are an active research area, and numerous different approaches have been proposed including Gelfand and Vounatsou (2003), Jin et al. (2007), Martinez-Beneito (2013) and MacNab (2016). However, in this paper we model the correlations between the 3 adverse health behaviours via the friendship network component of the model as described above, and thus modelling these correlations a second time may cause parameter identifiability issues in the model. In fact, as we show in the next section, the friendship network effects are a much more important driver of adverse health behaviours than the spatial effect, so the between behaviour correlations are more prominently captured in that component of the model.

Therefore instead we specify independent CAR models for each adverse health behaviour r, which are specified as a prior distribution for a vector of spatial random effects $\boldsymbol{\phi}_r = (\phi_{1r}, \ldots, \phi_{Sr})$ for the $S = 33$ Zip Codes. Spatial correlation is induced into these random effects through the spatial neighbourhood matrix $\mathbf{A}$ described in Section 2.4, which is defined by the commonly used *border sharing* rule. We note here that as the spatial correlation structure is defined by $\mathbf{A}$, all inferences about this part of the model are conditional on the choice of $\mathbf{A}$. Following a comparative study of different CAR priors by Lee (2011), we use the CAR prior proposed by Leroux et al. (2000) due to its consistent superior performance. This prior has the joint distribution $\boldsymbol{\phi}_r \sim \mathrm{N}(\mathbf{0}, \tau_r^2[\gamma_r(\mathrm{diag}(\mathbf{A1}) - \mathbf{A}) + (1-\gamma_r)\mathbf{I}]^{-1})$ for each response r, where $\mathbf{1}$ is an $S \times 1$ vectors of ones, $\mathbf{I}$ is the $S \times S$ identity matrix and $\mathrm{diag}(\mathbf{A1})$ is a diagonal matrix with diagonal elements obtained by the matrix product $\mathbf{A1}$. Thus the joint distribution of the spatial random effects for all three health behaviours $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3)_{3S\times 1}$ is a zero-mean multivariate normal distribution, whose covariance matrix is block diagonal with three $S \times S$ blocks given by $\tau_r^2[\gamma_r(\mathrm{diag}(\mathbf{A1}) - \mathbf{A}) + (1-\gamma_r)\mathbf{I}]^{-1}$ for $r = 1, 2, 3$.

The precision matrix for the spatial random effects relating to health behaviour r is given by $\gamma_r(\mathrm{diag}(\mathbf{A1}) - \mathbf{A}) + (1-\gamma_r)\mathbf{I}$ and is a weighted average of correlated $(\mathrm{diag}(\mathbf{A1}) - \mathbf{A})$ and independent $(\mathbf{I})$ components, where the amount of spatial dependence is controlled by γ_r. However, the fact that this prior captures spatial correlation is much easier to see from its

univariate full conditional form, which is given, together with the hyperpriors, for response r by

$$\begin{aligned} \phi_{sr}|\boldsymbol{\phi}_{-sr} &\sim \text{N}\left(\frac{\gamma_r \sum_{j\neq s} a_{sj}\phi_{jr}}{\gamma_r \sum_{j\neq s} a_{sj} + 1 - \gamma_r}, \frac{\tau_r^2}{\gamma_r \sum_{j\neq s} a_{sj} + 1 - \gamma_r}\right), \\ \tau_r &\sim \text{Half-Normal}(10000), \\ \gamma_r &\sim \text{Uniform}(0,1). \end{aligned} \tag{6.4}$$

Here $\boldsymbol{\phi}_{-sr}$ denotes the vector of $S-1$ spatial random effects for outcome r excluding ϕ_{sr}. In this prior the spatial dependence parameter γ_r is assigned a non-informative prior on the unit interval, and if $\gamma_r = 1$ the model simplifies to the intrinsic CAR prior for strong spatial correlation proposed by Besag et al. (1991) because the conditional expectation is the mean of the random effects in neighbouring areas. In contrast, if $\gamma_r = 0$ it is trivial to see that the random effects are independent. Finally, a weakly informative (large variance) half normal prior centred on zero is specified for the spatial standard deviation τ_r as suggested by Gelman (2006), which again lets the data play the dominant role in the estimation of its value.

6.4 MCMC estimation algorithm

This section describes Markov chain Monte Carlo algorithms for Model (6.1). The joint distribution of ($\boldsymbol{\beta}, \mathbf{u}$, $\boldsymbol{\Sigma}_{\mathbf{u}}$, $\boldsymbol{\phi}$, $\boldsymbol{\tau}^2$, $\boldsymbol{\rho}$) for Model (6.1) is given by

$$
\begin{aligned}
f(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\Sigma}_{\mathbf{u}}, \boldsymbol{\lambda}_{\mathbf{u}}, \boldsymbol{\phi}, \boldsymbol{\tau}^2, \boldsymbol{\rho} | \mathbf{Y}) &\propto f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{u}) f(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) f(\boldsymbol{\phi}|\boldsymbol{\tau}^2, \boldsymbol{\rho}) f(\mathbf{u}|\boldsymbol{\Sigma}_{\mathbf{u}}) f(\boldsymbol{\tau}^2) f(\boldsymbol{\rho}) \\
&= \prod_{s=1}^{S} \prod_{i=1}^{N_s} \prod_{r=1}^{R} \text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r, \phi_{sr}, \mathbf{u}_r) \prod_{r=1}^{R} \text{N}(\boldsymbol{\beta}_r|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\
&\times \prod_{r=1}^{R} \text{N}(\boldsymbol{\phi}_r|\tau_r, \rho_r) \prod_{j=1}^{J} \text{N}(\mathbf{u}_j|\boldsymbol{\Sigma}_{\mathbf{u}}) \\
&\times \prod_{r=1}^{R} \text{Half-Normal}(\tau_r|a_\tau) \text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{u}}|\mathbf{u})
\end{aligned}
$$

where $\boldsymbol{\mu_0} = \boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = 10^5\mathbf{I}$, and $\mathbf{u}_r = (u_{1r}, \ldots, u_{Jr})_{J\times 1}$ is the $J \times 1$ vector of alter random effects relating to the rth response. The following subsections in this section detail the full conditionals required for Model (6.1).

6.4.1 Sampling from $f(\boldsymbol{\beta}_r|\boldsymbol{\phi}_r, \mathbf{u}_r, \mathbf{y})$

The full conditional of $\boldsymbol{\beta}_r$ for model (6.1) is the product of $\sum_{s=1}^{S} N_s$ Bernoulli likelihoods and a Gaussian prior, which is shown below

$$
f(\boldsymbol{\beta}_r|\boldsymbol{\phi}_r, \mathbf{u}_r, \mathbf{y}) \propto \prod_{s=1}^{S} \prod_{i=1}^{N_s} \text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r, \phi_{sr}, \mathbf{u}_r) \text{N}(\boldsymbol{\beta}_r|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}).
$$

The Gaussian prior is not conjugate to the Bernoulli likelihood, which results in a full conditional distribution which is not standard. A Metropolis step is used to update $\boldsymbol{\beta}_r$, which is implemented in blocks. The proposal distribution for the block of parameters to be updated is given by $\text{N}(\boldsymbol{\beta}_r^{(t)}, \sigma_{\boldsymbol{\beta}_r}^{2(t)}\mathbf{I})$, where $\boldsymbol{\beta}_r^{(t)}$ is the current value of the block of parameters and $\sigma_{\boldsymbol{\beta}_r}^{2(t)}$ is the current value of the adaptive tuning parameter for the block of parameters

that is designed to keep the acceptance rate $\alpha_{\boldsymbol{\beta}_r}^{(t)}$ for the block of parameters to be between 0.3 - 0.5. The adaptive tuning parameter flattens the proposal distribution when the acceptance rate is less than 0.3 and compacts the proposal distribution when the acceptance rate is greater than 0.5. The acceptance probability of updating the block of current parameters $\boldsymbol{\beta}_r^{(t)}$ to proposed parameters $\boldsymbol{\beta}_r^*$ is given by

$$\min\left\{1, \frac{\prod_{s=1}^{S}\prod_{i=1}^{N_s}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^*, \phi_{sr}^{(t)}, \mathbf{u}_r^{(t)})\text{N}(\boldsymbol{\beta}_r^*|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})}{\prod_{s=1}^{S}\prod_{i=1}^{N_s}\text{Bernoulli}(y_{ikr}|\boldsymbol{\beta}_r^{(t)}, \phi_{sr}^{(t)}, \mathbf{u}_r^{(t)})\text{N}(\boldsymbol{\beta}_r^{(t)}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})}\right\}.$$

6.4.2 Sampling from $f(\boldsymbol{\phi}_r|\boldsymbol{\beta}_r, \mathbf{u}_r, \tau_r, \rho_r, \mathbf{y})$

The full conditional of $\boldsymbol{\phi}_r$ for model (6.1) is the product of $\sum_{s=1}^{S} N_s$ Bernoulli likelihoods and a Gaussian prior, which is shown below

$$f(\boldsymbol{\phi}_r|\boldsymbol{\beta}_r, \mathbf{u}_r, \tau_r, \rho_r, \mathbf{y}) \quad \propto \quad \prod_{s=1}^{S}\prod_{i=1}^{N_s}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r, \phi_{sr}, \mathbf{u}_r)\text{N}(\boldsymbol{\phi}_r|\tau_r, \rho_r).$$

In this instance, the Gaussian prior is not conjugate to the Bernoulli likelihood, which results in a full conditional distribution which is not standard. A Metropolis step is used to update $\boldsymbol{\phi}_r$. The proposal distribution for $\boldsymbol{\phi}_r$ is given by $\text{N}(\boldsymbol{\phi}_r^{(t)}, \sigma_{\boldsymbol{\phi}_r}^{2(t)}\mathbf{I})$, where $\boldsymbol{\phi}_r^{(t)}$ is the current value of $\boldsymbol{\phi}_r$ and $\sigma_{\boldsymbol{\phi}_r}^{2(t)}$ is the current value of the adaptive tuning parameter for the proposal distribution. The acceptance probability of updating the block of current parameters $\boldsymbol{\phi}_r^{(t)}$ to proposed parameters $\boldsymbol{\phi}_r^*$ is given by

$$\min\left\{1, \frac{\prod_{s=1}^{S}\prod_{i=1}^{N_s}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r^{(t)}, \phi_{sr}^{*}, \mathbf{u}_r^{(t)})\text{N}(\boldsymbol{\phi}_r^{*}|, \tau_r^{(t)}, \rho_r^{(t)})}{\prod_{s=1}^{S}\prod_{i=1}^{N_s}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r^{(t)}, \phi_{sr}^{(t)}, \mathbf{u}_r^{(t)})\text{N}(\boldsymbol{\phi}_r^{(t)}|\tau_r^{(t)}, \rho_r^{(t)})}\right\}.$$

6.4.3 Sampling from $f(\mathbf{u}_j|\boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{u}_{-j}, \boldsymbol{\Sigma}_{\mathbf{u}}, \mathbf{y})$

The full conditional of $\mathbf{u}_j = (u_{1j}, ..., u_{Rj})$ for model (6.1) is the product of a Gaussian prior and Bernoulli likelihoods equal to the number of times $\mathbf{u}_j$ appears in a likelihood, which is shown below

$$\begin{aligned} f(\mathbf{u}_j|\boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{u}_{-j}, \boldsymbol{\Sigma}_{\mathbf{u}}, \mathbf{y}) &\propto \prod_{s=1}^{S}\prod_{i=1}^{N_s}\prod_{r=1}^{R}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r, \phi_{sr}, \mathbf{u}_r)\text{N}(\mathbf{u}_j|\boldsymbol{\Sigma}_{\mathbf{u}}) \\ &\propto \prod_{is\ s.t.\ j\in\text{net}(is)}\prod_{r=1}^{R}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r, \phi_{sr}, \mathbf{u}_r)\text{N}(\mathbf{u}_j|\boldsymbol{\Sigma}_{\mathbf{u}}). \end{aligned}$$

Similarly to the full conditional of $\mathbf{u}_j$ for (6.1), the Gaussian prior is not conjugate to the Bernoulli likelihood, which results in a full conditional distribution which is not standard. A Metropolis step is used to update $\mathbf{u}_j$. The proposal distribution for $\mathbf{u}_j$ is given by $\text{N}(\mathbf{u}_j^{(t)}, \sigma_{\mathbf{u}_j}^{2(t)}\mathbf{I})$, where $\mathbf{u}_j^{(t)}$ is the current value of $\mathbf{u}_j$ and $\sigma_{\mathbf{u}_j}^{2(t)}$ is the current value of the adaptive tuning parameter for the proposal distribution. The acceptance probability of updating the block of current parameters $\mathbf{u}_j^{(t)}$ to proposed parameters $\mathbf{u}_j^{*}$ is given by

$$\min\left\{1, \frac{\prod_{is\ s.t.\ j\in\text{net}(is)}\prod_{r=1}^{R}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r^{(t)}, \phi_{sr}^{(t)}, u_{jr}^{*}, \mathbf{u}_r^{(t)})\text{N}(\mathbf{u}_j^{*}|\boldsymbol{\Sigma}_{\mathbf{u}}^{(t)})}{\prod_{is\ s.t.\ j\in\text{net}(is)}\prod_{r=1}^{R}\text{Bernoulli}(y_{isr}|\boldsymbol{\beta}_r^{(t)}, \phi_{sr}^{(t)}, u_{jr}^{(t)}, \mathbf{u}_r^{(t)})\text{N}(\mathbf{u}_j^{(t)}|\boldsymbol{\Sigma}_{\mathbf{u}}^{(t)})}\right\}.$$

6.4.4 Sampling from $f(\tau_r|\boldsymbol{\phi}_r, \rho_r, \mathbf{y})$

The full conditional of τ_r for model (6.1) is the product of a Gaussian prior and a Half-Normal(a_τ). A Metropolis-Hastings step is used. As the proposal distribution to be used is a truncated normal distribution which is not symmetric and ensures that $\tau \in [0, \infty)$. The truncated normal distribution is given by $\text{TN}(\tau_r^{(t)}, \sigma_{\tau_r}^{2(t)}, 0, \infty)$, where $\tau_r^{(t)}$ is the current value of τ_r, $\sigma_{\tau_r}^{2(t)}$ is the current value of the adaptive tuning parameter for the proposal distribution, 0 is the minimum value of τ_r that can be drawn. The acceptance probability of updating $\tau_r^{(t)}$ to proposed parameters τ_r^* is given by

$$\min\left\{1, \frac{\text{N}(\mathbf{0}, \tau_r^{2*}(\rho_r^{(t)}(\text{diag}(\mathbf{A1}) - \mathbf{A}) + (1 - \rho_r^{(t)})\mathbf{I})^{-1})\ \text{Half-Normal}(\tau_r^*|a_\tau)\ \text{TN}(\tau_r^{(t)}|\tau_r^*, \sigma_{\tau_r^2}^{2(t)}, 0, \infty)}{\text{N}(\mathbf{0}, \tau_r^{2(t)}(\rho_r^{(t)}(\text{diag}(\mathbf{A1}) - \mathbf{A}) + (1 - \rho_r^{(t)})\mathbf{I})^{-1})\ \text{Half-Normal}(\tau_r^{(t)}|a_\tau)\ \text{TN}(\tau_r^*|\tau_r^{(t)}, \sigma_{\tau_r^2}^{2(t)}, 0, \infty)}\right\}.$$

6.4.5 Sampling from $f(\rho_r|\boldsymbol{\phi}_r, \tau_r, \mathbf{y})$

The full conditional of ρ_r for model (6.1) is $\text{N}(\mathbf{0}, \tau_r^2(\rho_r(\text{diag}(\mathbf{A1}) - \mathbf{A}) + (1 - \rho_r)\mathbf{I})^{-1})$, this results in a full conditional distribution which is not standard A Metropolis-Hastings step is used to update ρ_r, as the proposal distribution to be used is a truncated normal distribution which is not symmetric and ensures that $\rho \in [0, 1)$. The truncated normal distribution is given by $\text{TN}(\rho_r^{(t)}, \sigma_{\rho_r}^{2(t)}, 0, 1)$, where $\rho_r^{(t)}$ is the current value of ρ_r, $\sigma_{\rho_r}^{2(t)}$ is the current value of the adaptive tuning parameter for the proposal distribution, 0 is the minimum value of ρ_r that can be drawn, and 1 is the maximum value of ρ_r that can be drawn. The acceptance probability of updating $\rho_r^{(t)}$ to proposed parameters ρ_r^* is given by

$$\min\left\{1, \frac{\text{N}(\mathbf{0}, \tau_r^{2(t)}(\rho_r^*(\text{diag}(\mathbf{A1}) - \mathbf{A}) + (1 - \rho_r^*)\mathbf{I})^{-1})\ \text{TN}(\rho_r^{(t)}|\rho_r^*, \sigma_{\rho_r}^{2(t)}, 0, 1)}{\text{N}(\mathbf{0}, \tau_r^{2(t)}(\rho_r^{(t)}(\text{diag}(\mathbf{A1}) - \mathbf{A}) + (1 - \rho_r^{(t)})\mathbf{I})^{-1})\ \text{TN}(\rho_r^*|\rho_r^{(t)}, \sigma_{\rho_r}^{2(t)}, 0, 1)}\right\}.$$

6.4.6 Sampling from $f(\boldsymbol{\Sigma}_{\mathbf{u}}|\mathbf{u}, \mathbf{y})$

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{u}}$ for Model (6.1) is the product of J Gaussian distributions and a conjugate Inverse-Wishart$(R+1, \mathbf{I})$ prior, which is shown below

$$f(\boldsymbol{\Sigma}_{\mathbf{u}}|\mathbf{u}, \mathbf{y}) \propto \prod_{j=1}^{J} \text{N}(\mathbf{u}_j|\boldsymbol{\Sigma}_{\mathbf{u}})\text{Inverse-Wishart}(\boldsymbol{\Sigma}_{\mathbf{u}}|\mathbf{u}).$$

This results in an

$$\text{Inverse-Wishart}\left(R+1+J, \mathbf{I}+\sum_{j=1}^{J}\mathbf{u}_j\mathbf{u}_j^{\top}\right)$$

posterior distribution.

The sampling steps are performed in `C++` and the manipulation of the user-specified arguments passed to the function are done in `R`.

6.5 Results: Social Networking Survey

This section presents the results of the multivariate MMMC model with a spatial component on the Social Networking Survey data set. Subsection 6.5.1 presents the results generated by the analysis. The aims of this section are given as follows

1. **A multivariate MMMC model with a spatial component**. The novel contribution relating to the multivariate MMMC model with a spatial component stated in Subsection 6.1.1 is carried out in this section.

2. **An evaluation of the spatial and network component of the model**. There

are two main research questions relating to the network level of the model.

Q1 How much global spatial autocorrelation is there in the odds of engaging in alcohol, cigarette, and marijuana consumption when the network is taken into account?

Q2 How much variance is there in the network level and how does this compare to the amount of variation in the spatial random effects?

6.5.1 Results

This subsection presents the results of fitting Model (6.1), whose Markov chain Monte Carlo algorithm are described in Section 6.4. The model in this section was run twice, simulating two sets of samples from the posterior distribution of each parameter to compute the potential scale reduction factor. For each run, 200,000 iterations were thinned by a factor of 20 after a burn-in period of 200,000 samples, resulting in 10,000 posterior samples for each run.

6.5.1.1 Multivariate model results

We fit 8 different models to the survey data, which allows us to examine the relative importance of covariate effects, friendship network effects and spatial effects in explaining an adolescents' propensity to partake in adverse health behaviours. These 8 models are denoted $\mathcal{M}1$ to $\mathcal{M}8$ and contain all possible combinations of the three different model components, ranging from $\mathcal{M}1$ which only contains an intercept term through to $\mathcal{M}8$ which is the full model given by (6.1). A summary of the components included in each model is given in Table 6.1 for ease of reference. The covariates used in these models include the categorical variables gender, exam grades and school, which are summarised in Section 2.2.

Table 6.1: Summary and overall fit of the 8 models.

Model	**Covariates**	**Space**	**Network**	**DIC**	p_D
$\mathcal{M}1$	-	-	-	3,821	3.0
$\mathcal{M}2$	✓	-	-	3,463	30.1
$\mathcal{M}3$	-	✓	-	3,798	22.1
$\mathcal{M}4$	-	-	✓	2,904	427.5
$\mathcal{M}5$	✓	✓	-	3,463	29.9
$\mathcal{M}6$	✓	-	✓	2,891	408.6
$\mathcal{M}7$	-	✓	✓	2,907	417.7
$\mathcal{M}8$	✓	✓	✓	2,883	413.1

6.5.2 Model comparison

The overall fit of each model is summarised in Table 6.1, which displays the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) and the effective number of independent parameters (p_D). A comparison of the single component models $\mathcal{M}2$, $\mathcal{M}3$, and $\mathcal{M}4$ to the null (intercept only) model shows that the inclusion of the friendship network component leads to the greatest reduction in DIC compared to the intercept only model, as the DIC goes from 3,821 to 2,904, a reduction of 917. The sole inclusion of covariates has just under half this impact with a DIC reduction of 358, while the sole inclusion of a spatial component leads to a DIC reduction of just 23. Adding in the covariates ($\mathcal{M}6$) and then additionally the spatial component ($\mathcal{M}8$) to the friendship network model improves the fit to the data but only marginally, with the DIC reductions compared to the network only model ($\mathcal{M}4$) being only 13 and 21 respectively. Thus, while the full model with all three components has the lowest DIC value, the impact of adding in the covariates and the spatial components are small once the friendship network effects are included.

The results in Table 6.1 show that the effective number of independent parameters p_D went down for $\mathcal{M}8$ compared to $\mathcal{M}4$ and $\mathcal{M}7$, despite the former model being the most complex in terms of its parameterisation. The reason for this is that in the full model $\mathcal{M}8$ the variation in the data is jointly modelled by all three components, where as in $\mathcal{M}4$ for

example only the network component is included. Thus in $\mathcal{M}8$ the network component is having to model less of the variation in the data compared to in $\mathcal{M}4$, due to the covariates and to a much lesser degree the spatial effect modelling some of this variation. This results in a reduction in the effective number of independent parameters for the network component in model $\mathcal{M}8$ compared to $\mathcal{M}4$ due to a reduction in the variation in the random effects $\{\mathbf{u}_j\}$, which thus causes the reduced p_D. The remainder of this section present the results relating to the full model $\mathcal{M}8$, so that the effects of all three components can be observed.

6.5.3 Model fit

In order to confirm that the model fits the data adequately we simulate 1,000 trivariate samples $\{\tilde{\mathbf{y}}^{(1)}, \cdots, \tilde{\mathbf{y}}^{(1000)}\}$ from the posterior predictive distribution $f(\tilde{\mathbf{y}}|\mathbf{y})$, where $\mathbf{y}$ denotes the observed data. As both $(\tilde{\mathbf{y}}^{(j)}, \mathbf{y})$ are binary this posterior predictive check involves computing the probability that the observed data matches the simulated data generated from the posterior predictive distribution. Averaging over all individuals i, spatial units s, health behaviour r and posterior predictive samples j, the posterior predictive probability $\mathbb{P}(\tilde{y}_{i_s r} = y_{i_s r}) = 0.68$, suggesting that the model fits the data relatively well as it generates simulated data that are similar to the real data. The corresponding health behaviour specific values are $\mathbb{P}(\tilde{y}_{i_s 1} = y_{i_s 1}) = 0.67$, $\mathbb{P}(\tilde{y}_{i_s 2} = y_{i_s 2}) = 0.67$, and $\mathbb{P}(\tilde{y}_{i_s 3} = y_{i_s 3}) = 0.70$, suggesting that the model fits the marijuana response slightly better than the other two.

Additionally, Table 6.2 provides the posterior means and 95% credible intervals for the between health behaviour correlations from the friendship network component of the model, which allows us to examine the appropriateness of modelling all three health behaviours jointly. These correlations are captured in $\boldsymbol{\Sigma}$, and for example the correlation between alcohol and cigarettes is computed by $\rho_{12} = \boldsymbol{\Sigma}_{12}/\sqrt{\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22}}$. The table shows that the correlations are very high and close to one for each pair of adverse health behaviours, with posterior means ranging between 0.955 (alcohol and marijuana) and 0.975 (cigarettes and marijuana).

Table 6.2: Estimates and 95% credible intervals for the between health behaviour correlations.

Adverse health behaviours	Estimated correlation
$\rho_{12} = \boldsymbol{\Sigma}_{12}/\sqrt{\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22}}$ - alcohol and cigarettes	0.956 (0.808, 0.996)
$\rho_{13} = \boldsymbol{\Sigma}_{13}/\sqrt{\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{33}}$ - alcohol and marijuana	0.955 (0.792, 0.997)
$\rho_{23} = \boldsymbol{\Sigma}_{23}/\sqrt{\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{33}}$ - cigarettes and marijuana	0.975 (0.890, 0.998)

These strong correlations thus support the use of a joint modelling approach for our adverse health behaviours. Finally, the posterior samples of these network correlation parameters yield $\mathbb{P}(\rho_{23} > \rho_{12} \cap \rho_{23} > \rho_{13}) = 0.629$, suggesting that the correlation between the cigarette and marijuana responses is likely to be greater than both the correlations between the alcohol and cigarette responses and the alcohol and marijuana responses.

6.5.4 Covariate effects

Table 6.3 displays the estimated covariate effects (posterior means) and 95% credible intervals for each adverse health behaviour, and all results are presented as odds ratios relative to the baseline level of the factor (the first one in the table denoted by a "-"). The table shows that in comparison to females, the baseline level, males had a significantly reduced odds of consuming alcohol in the past 30 days, with an estimated odds ratio of 0.57. In contrast, the 95% credible intervals for the male covariate relating to the cigarette and marijuana responses show no statistically significant gender effect.

In contrast, the effect of exam performance is much more consistent than that of gender, with decreasing exam performance being significantly associated with higher odds of partaking in each adverse health behaviour. Here the baseline level is mostly A's, and decreasing the grade category almost always exhibits an increased and significant odds ratio. For example, adolescents who score mostly C's or below have significant odds ratios of 8.05 (alcohol), 5.46 (cigarettes) and 14.66 (marijuana), when compared to the baseline mostly

Table 6.3: Summary of the covariate effects as odds ratios and selected other parameters from model $\mathcal{M}8$.

	Alcohol	**Cigarettes**	**Marijuana**
Covariates			
Female	-	-	-
Male	0.57 (0.39, 0.83)	1.31 (0.86, 1.99)	0.91 (0.56, 1.46)
A's	-	-	-
A's and B's	2.49 (1.10, 6.12)	1.41 (0.61, 3.34)	2.48 (0.86, 7.82)
B's	2.66 (0.91, 7.85)	0.68 (0.19, 2.21)	2.35 (0.59, 9.55)
B's and C's	4.16 (1.86, 10.01)	2.60 (1.16, 6.07)	5.73 (2.09, 17.04)
C's or lower	8.05 (3.61, 19.74)	5.46 (2.47, 12.46)	14.66 (5.47, 45.05)
School 1	-	-	-
School 2	0.96 (0.35, 2.41)	1.12 (0.35, 3.38)	1.07 (0.28, 3.79)
School 3	1.42 (0.48, 3.85)	1.02 (0.27, 3.33)	1.09 (0.24, 4.40)
School 4	0.38 (0.12, 1.05)	0.78 (0.20, 2.76)	0.47 (0.10, 2.05)
School 5	1.07 (0.29, 3.56)	0.79 (0.15, 3.49)	0.80 (0.11, 4.79)
Space			
τ_r	0 (0, 6.6×10^{-53})	0 (0, 1.4×10^{-60})	0 (0, 2.3×10^{-62})
γ_r	0.419 (0.02, 0.92)	0.415 (0.02, 0.92)	0.417 (0.02, 0.92)
Network			
$\mathbf{\Sigma}_{rr}$	6.21 (3.29, 11.01)	10.32 (5.49, 19.59)	14.92 (7.12, 27.76)

A's category. Finally, the table shows that after accounting for all the other components in the model, there are no statistically significant school effects for schools 2, 3, 4 and 5 when compared to the reference level school 1, with all 95% credible intervals across the three responses containing the null odds ratio of 1.

6.5.5 Peer network effects

Table 6.3 provides the posterior means and 95% credible intervals for the variances relating to the network random effects in the model, which quantify the variation among the individual friendship network effects. The posterior means for alcohol, cigarettes and marijuana are respectively 6.21, 10.32 and 14.92, suggesting that the greatest level of variation is for marijuana. This finding is confirmed by the posterior probability that $\mathbb{P}(\mathbf{\Sigma}_{33} > \mathbf{\Sigma}_{11} \cap \mathbf{\Sigma}_{33} >$

$\boldsymbol{\Sigma}_{22}) = 0.892$, suggesting a very clear size ordering among these variances with the variance relating to the marijuana response being the largest.

The isolation random effects for not nominating a friend are denoted by u_1^*, u_2^*, and u_3^*, and on the odds ratio scale their estimates and 95% credible intervals are given by: Alcohol - 2.31 (1.34, 4.06); Cigarettes - 3.32 (1.88, 6.11); and Marijuana - 4.57 (2.39, 9.39). All these estimates and 95% credible intervals are greater than one, suggesting that being isolated from others (i.e. not nominating a friend) increases the likelihood of drinking alcohol, smoking cigarettes, and having used marijuana. The posterior mean marijuana isolation effect is the largest of the three but not significantly so, as all three 95% credible intervals overlap. However, that said there is relatively strong evidence of a clear size ordering in these isolation effects, because the model produced the following posterior probabilities: $\mathbb{P}(u_3^* > u_1^* \cap u_3^* > u_2^*) = 0.793$, $\mathbb{P}(u_3^* > u_1^*) = 0.962$, and $\mathbb{P}(u_3^* > u_2^*) = 0.806$. Thus it appears that isolation has the largest effect on the marijuana response.

Figure 6.3 displays the 95% credible intervals for the individual friendship random effects and the isolation effect relating to each response, namely $\{\mathbf{u}_1, u_1^*\}$ (alcohol), $\{\mathbf{u}_2, u_2^*\}$ (cigarettes), and $\{\mathbf{u}_3, u_3^*\}$ (marijuana). The effects are ordered by posterior mean on the horizontal axis, and those in black are not significantly different from zero at the 5% level. The instances in green are significantly different from zero at the 5% level and contain only negative values. There was only one case of this for each of the three responses, all attributable to the same individual. Thus, holding everything else equal, having nominated this individual as a friend was observed to have decreased the likelihood of drinking alcohol, smoking cigarettes, and having used marijuana. Those in red are significantly different from zero at the 5% level and contain only positive values. There were 38, 41, and 43 instances of this relating to the alcohol, cigarette, and marijuana response respectively, and nominating these individuals as friends increases ones propensity to smoke, drink, and use marijuana.

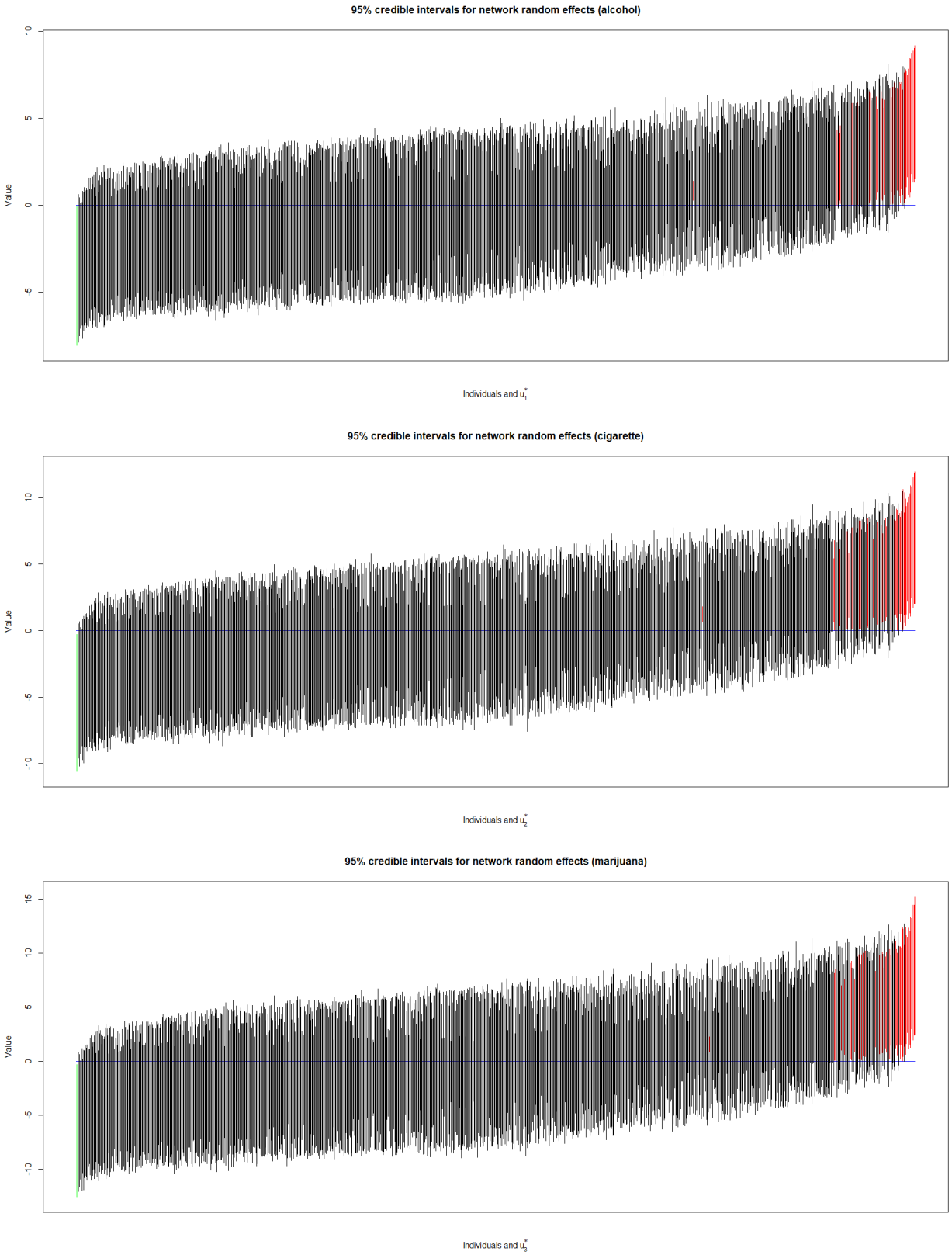


Figure 6.3: The 95% credible intervals for the network effects relating to each response, $\{\mathbf{u}_1, u_1^*\}$ (alcohol - top), $\{\mathbf{u}_2, u_2^*\}$ (cigarettes - middle), and $\{\mathbf{u}_3, u_3^*\}$ (marijuana - bottom). These effects are ordered by size, and those non-significant effects with 95% credible intervals that contain 0 are shown in black, those that contain values strictly less than 0 are shown in green, and those that contain values strictly greater than 0 are shown in red.

6.5.6 Spatial effects

The spatial standard deviation (τ_r) and dependence (γ_r) parameters are displayed in Table 6.3. The table shows that the spatial effect is essentially non-existent, as the posterior means for the conditional standard deviations τ_r are almost zero for all three adverse health behaviours, and the upper limit of the 95% credible interval is also very close to zero. This is further emphasised by the posterior means for the sets of spatial random effects ($\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3$), which range between -3.69×10^{-53} and 5.51×10^{-53}. This lack of a spatial effect after covariate and friendship network effects have been accounted for confirms the overall model fit results from Table 6.1, which shows that the DIC values for models $\mathcal{M}6$ and $\mathcal{M}8$ are almost identical.

6.6 Discussion and limitations

This section has proposed a novel spatio-network model for binary multiple health behaviour data, which jointly captures the potential effects of covariate factors, spatial location, friendship network effects and within-individual correlations between outcomes. The main advantage of our model over existing alternatives is its flexibility in being able to capture this wide range of drivers of adolescent health behaviours, whereas existing models only account for a subset of them. This has allowed us to examine which of these drivers are the most important for explaining an adolescent's propensity to drink alcohol or smoke cigarettes or marijuana, and in our California-based study, we have obtained a number of interesting findings.

Our main finding is that peer effects play a large role in determining whether adolescents partake in adverse health behaviours, as their addition to the null model leads to the greatest reduction in DIC when compared to just adding either the covariates or spatial component. Furthermore, once friendship effects are included in the model, there is only a small improvement in model fit when incorporating the covariates and/ r spatial component.

In future work, it would be interesting to see whether the relative importances of the three components are mirrored in terms of their out-of-sample predictive ability, for example using a cross-validation type approach. For the covariates, the only consistently significant effect on the participation in adverse health behaviours was school exam performance, with students having poorer exam results observed to be more likely to partake in these behaviours. In contrast, the spatial effect was essentially non-existent.

Our second main finding is that the effect that a friend has on an adolescent is strongly correlated across the three binary responses, with estimated pairwise correlations ranging between 0.956 and 0.975. These correlations support the notion of co-occurrence of risky behaviours in adolescents found by Hale and Viner (2016). Among the three pairs of jointly modelled friendship random effects, the results show that the pair relating to the cigarette and marijuana responses are the most correlated, although the correlations for the remaining pairs are only slightly smaller.

Chapter 7

Review: Multilevel statistical software

In this chapter, I provide an overview of some popular statistical software programs that allow users to fit multilevel models and or multiple membership multiple classification (MMMC) models. The remainder of this chapter is structured as follows. Section 7.1 reviews various statistical software programs that can be used to fit multilevel and MMMC models. Section 7.2 outlines the novel contributions that the software I propose will make to the landscape of multilevel and MMMC software.

7.1 Multilevel software

SPSS

SPSS is a statistical software program that uses a likelihood-based approach to allow for the fitting of hierarchical or multilevel models. The software was originally released in 1968 and has seen regular updates throughout its life, with Version 11.0 introducing the capability to fit hierarchical or multilevel models in 2001. The software is offered to users through paid subscriptions and perpetual/term licenses but also offers a free trial. SPSS is very user-friendly and easy for a novice to use due to its point-and-click GUI that allows the

user to perform complex tasks with minimal coding. The software package also offers SPSS syntax, which is a programming language unique to SPSS that can be used as an alternative to the drop-down menus for data manipulation and statistical analyses.

Mplus

Mplus (Muthén and Muthén (1998-2017)) is a statistical software program that can use Markov chain Monte Carlo methods to allow for the fitting of multilevel models. The first version of the software, Mplus Version 1, was released on 19th November 1998 and has undergone several major version updates with a few minor updates for each major version. Mplus Version 6, which was released in April 2010, introduced the capability to fit multilevel models in a Bayesian paradigm. The software is offered to users through paid subscriptions but offers a demo version that limits the number of variables that can be used in an analysis. Mplus is fairly user-friendly and uses an Mplus syntax to specify models, which the user is required to learn. Mplus is a very well documented piece of software that provides the user with an in-depth guide of the software. However, as Mplus is not open-source, it lacks the open-source community that is a feature of many other statistical software packages.

Bambi

Bambi (Capretto et al. (2022)) is a statistical software package that uses Markov chain Monte Carlo methods to allow for the fitting of hierarchical and multilevel models. The package is written in Python, working with the probabilistic programming framework PyMC, and is designed to make it extremely easy to fit Bayesian mixed-effects models common in biology, social sciences and other disciplines. The software is open-source, free to use and under an MIT license. Bambi uses a syntax very similar to lme4 formula syntax in R, which many packages in R emulate. This makes it very user-friendly for users who are familiar with specifying models in R. Bambi is a very well documented piece of software that provides the user with an in-depth guide of the software and online worked examples for the models in the package.

Stata

Stata is a statistical software program that can use Markov chain Monte Carlo methods to fit multilevel models through bayesmh. Stata introduced Bayesian analysis through bayesmh in its release of Stata 14 in April 2015. The software is offered to users through a variety of paid annual and perpetual licenses. Stata is very user-friendly due to its point-and-click GUI that allows the user to perform complex tasks with minimal coding. The software also offers scripting, which allows users to use Stata's programming language as an alternative to the drop-down menus for data manipulation and statistical modelling. Stata is a very well documented and provides the user with an in-depth guide of the software along with other online resources.

R

R is a programming language and free software environment designed for statistical computing and graphics. The language heavily supports the creation of packages by developers to expand the language and increase its functionality. These packages are hosted on the Comprehensive R Archive Network (CRAN), the main repository for R packages.

R package: MCMCglmm

MCMCglmm (Hadfield (2010)) is an R package that uses Markov chain Monte Carlo methods to fit generalized linear mixed-effects models in a Bayesian setting. The software is written in C, C++, and R. MCMCglmm was initially released in 2010 and created by Jarrod Hadfield. The software has been used in various research fields, such as behavioural research (Dean et al. (2017)), marketing (Chandrasekaran et al. (2019)), and global warming (Mostafa (2016)). MCMCglmm relies on various sampling procedures to produce simulations from the posterior distribution of a parameter. MCMCglmm is relatively simple to use for those with limited knowledge of Bayesian statistics. The software also has an accompanying course to provide the user with a deeper understanding of the software (see Hadfield (2019)). Furthermore, as MCMCglmm is a package in R, users have access to a number of other R

packages that can be used for data cleaning and visualisation, making the entire analysis a more seamless experience. However, in contrast to software such as WinBUGS (see below), the types of models that can be fit is more restrictive.

R package: bayesm

bayesm (Rossi (2022)) is an R package that uses Markov chain Monte Carlo methods to fit multilevel models in a Bayesian paradigm. The software is tailored to cover many important models used in marketing and micro-econometrics applications. The software was initially released in 2005 and is written in a mixture of C++ and R. The software is open source and under a GPL ≥ 2 license. bayesm is fairly simple to use, with access to a variety of supplementary online resources and the availability of an accompanying book (Rossi et al. (2012)) which accompany it that users can study.

R package: lme4

lme4 is an R package that is used to fit linear and generalized linear mixed-effects models. lme4 uses likelihood approaches to fit its models. The software is written in both C++ and R. lme4 was initially released in 2010 and created by Ben Bolker but has had many contributors. The software has been used in various research fields, such as language (Nixon et al. (2015)), agriculture (Wilson et al. (2010)), and medicine (Munoz-Zanzi et al. (2016)). lme4 is a very popular model fitting package and has very detailed documentation.

R package: brms

brms (Bürkner (2017)) is an R package that uses Markov chain Monte Carlo methods to fit generalized (non-)linear multivariate multilevel/MMMC models using Stan (Stan Development Team (2022)) for full Bayesian inference. brms was initially released in 2015 to address the hesitancy in researchers using Stan directly, as every model has to be written, debugged and possibly also optimized which may be time-consuming and error-prone process even for researchers familiar with Bayesian inference by allowing the user to benefit from the merits of Stan only by using simple, lme4-like formula syntax. The software is open

source and under a GPL ≥ 2 license. brms is fairly simple to use, with access to a variety of supplementary online resources that users can study.

WinBUGS

WinBUGS (Lunn et al. (2000)) is a statistical software program that uses Markov chain Monte Carlo methods to fit Bayesian models. The software was initially released in 1997 and is an upgraded and rewritten version of classic BUGS, a piece of statistical software written in Modular-2. WinBUGS was developed by a team of academics at the Medical research Council Biostatistics Research Unit in Cambridge and is written in Component Pascal, a successor of the Pascal programming language. The final release of WinBUGS is version 1.4.3 and was published on 6th August 2007. The software is under a Freeware license and has been used in a wide array of research areas, such as accident research (Islam and El-Basyouny (2015)), social science (Wong et al. (2009)), and infectious diseases (Yang et al. (2015)). WinBUGS does allow for the specification of arbitrary distributions through the use of syntactical tricks. Markov chain Monte Carlo estimation in WinBUGS is conducted using Gibbs sampling, the random walk Metropolis algorithm (Metropolis et al. (1953)), adaptive rejection sampling (Gilks et al. (1995)), and the slice sampling algorithm (Neal (1997)). WinBUGS is a very flexible piece of statistical software and has very detailed documentation. However, it requires the user to have a good grasp of Bayesian statistics in order to correctly specify an intended model. Additionally, for beginners, error messages in WinBUGS can be a source of frustration, as they can be difficult to understand. WinBUGS can also struggle to handle large data sets and can take a long time to perform a large number of Markov chain Monte Carlo iterations for complex models.

MLwiN

MLwiN (Charlton et al. (2019)) is a statistical software program that, similarly to WinBUGS, uses Markov chain Monte Carlo methods to fit multilevel and MMMC models in a Bayesian setting. MLwiN was initially released in 1998 and is a successor to MLn. MLwiN

was developed by the Centre for Multilevel Modelling at the University of Bristol. The software is under a proprietary license and has been used in various research fields, such as education (Durrant et al. (2018)), health (Barker et al. (2019)), and veterinary medicine (Aunsmo et al. (2009)). Markov chain Monte Carlo estimation for multilevel and MMMC models in MLwiN are carried out using Gibbs sampling and Metropolis-Hastings sampling. MLwiN is a very well documented piece of software that provides the user with an in-depth guide of the software. Another very valuable aspect of MLwiN is the equation window that it provides to users, which makes clear the model being fit by the user. However, MLwiN is not as flexible as a statistical program such as WinBUGS which allows users to specify a very wide array of models.

7.2 Novel contributions

As discussed above, there are a number of different statistical software programs that allow users to fit multilevel and or MMMC models. In this section I provide the contributions that my software, described in Chapter 8, makes for multilevel and MMMC statistical software.

In contrast to SPSS and lme4, which do not allow the user to perform Bayesian data analysis, the software I propose will use Markov chain Monte Carlo methods to fit multilevel and MMMC models in a Bayesian setting. Conversely to MCMCglmm, which does not allow users to fit MMMC models, the software proposed will allow MMMC models to be fit. In comparison to MLwiN, which does allow users to fit MMMC models, the software proposed will be open-source. Although WinBUGS allows users to fit almost all types of statistical models, WinBUGS is not dedicated to fitting MMMC models which results can result in the software forfeiting some efficiency and being difficult to use. Conversely to MLwiN, which requires most user to purchase a license, the software proposed will be free for all.

Chapter 8

`netcmc`: An `R` package for social network modeling

8.1 Introduction

Social network structures are a prominent feature in many different fields of research, such as sexual disease transmission networks (Neaigus et al., 1996) and the impact of friendship networks on adolescent health behaviours (Alexander et al., 2001; Lorant and Tranmer, 2019; Tranmer et al., 2014). The social network structures that are present in a data set can be gathered in many different ways, with one of the most popular ways in the adolescent health behaviours literature being through surveys. In such surveys, adolescents are asked to nominate a given amount of alters (individuals) as peers, using a particular criteria (i.e. friend, romantic relationship, etc.), resulting in a social network structure. Data sets with social network structures typically exhibit social network autocorrelation, with adolescents tending to have similar behaviours to their peers. However, there aren't many resources to help a user to model such data.

In this chapter, we describe work completed where we seek to make modeling social net-

work structures in data more accessible by creating an R package titled netcmc, which will allow a user to fit a wide range of social network models in a Bayesian paradigm. This chapter pulls together many of the concepts and developments of methods provided in earlier chapters, such as the Bayesian methods presented in Chapter 2 and the univariate and multivariate models presented across Chapters 4 and 6. netcmc can be downloaded from The Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/web/packages/netcmc/index.html and GitHub at https://github.com/GNG3/netcmc for macOS, Linux, and Windows platforms.

The remainder of the chapter is structured as follows. Subsection 8.1.1 outlines the main motivations behind the creation of the netcmc package. Subsection 8.1.2 outlines the novel contributions made. Section 8.2 describes the variety of models that can be implemented in the netcmc package. Section 8.3 describes how the software can be installed and used. Section 8.4 presents a simulation study for the package. Section 8.5 presents an application of a model implemented in the netcmc package on network data of delinquent behaviours of adolescents in the Netherlands. Section 8.6 concludes with a discussion.

8.1.1 Motivation

The models outlined above are usually implemented in a Bayesian paradigm, with inference based on Markov chain Monte Carlo (MCMC) simulation. The most commonly used software to fit the models previously described are WinBUGS and MLwiN. WinBUGS (Lunn et al. (2000)) is a statistical software program that uses Markov chain Monte Carlo methods to fit Bayesian models. The software is under a Freeware license. Markov chain Monte Carlo estimation in WinBUGS is conducted using Gibbs sampling, the random walk Metropolis algorithm (Metropolis et al. (1953)), adaptive rejection sampling (Gilks et al. (1995)), and the slice sampling algorithm (Neal (1997)). MLwiN (Charlton et al. (2019)) is a statistical software program that, similarly to WinBUGS, uses Markov chain Monte Carlo methods to fit

multilevel and multiple membership multiple classification (MMMC) models in a Bayesian setting. MLwiN was initially released in 1998 and is a successor to MLn. MLwiN was developed by the Centre for Multilevel Modelling at the University of Bristol. The software is under a proprietary license.

However, each of the software packages mentioned above are limited in a number of ways. MLwiN has 2 main limitations: (1) MLwiN requires the purchase of a license for those who are not UK-academics; (2) it is not open-source, and so users aren't able to view the source code. WinBUGS' main limitation is it's non-ease of use when it comes to specifying a desired model.

8.1.2 Novel contributions

This chapter provides three distinct novel contributions of netcmc to the literature on social network modeling in a Bayesian paradigm, and are given as follows

1. **An R package which is free to use for all.** As MLwiN requires the purchase of a license for those who are not UK-academics, the software may limit the adoption of social network modeling for those who are not UK-academics. Thus, as a result, the netcmc package in R is made free to use for all.

2. **An open-source R package for social network modeling.** netcmc it is open-source package for social network modeling, and so allows users to view the source code, and even contribute to the project. This point should also contribute to the reproducibility of research in the field of social network modeling.

3. **An easy to use, single call software package.** The main advantage netcmc has over WinBUGS is ease of use, because (1) the network adjacency matrix is easy to specify as a binary neighbourhood matrix; and (2) given the required argument inputs, the models can be implemented by a single call in R.

8.2 Available models

This section describes the set of Bayesian hierarchical models that can be implemented in the `netcmc` software, which broaden the univariate and multivariate models presented across Chapters 4 and 6, making the software useable for a wider array of problems.

8.2.1 Data likelihoods

The general univariate model that can be implemented in `netcmc` is a variant of a generalized linear model which incorporates covariates, spatial, and network structure and is given by

$$\begin{aligned} Y_{is}|\mu_{is} &\sim f(y_{is}|\mu_{is},\nu^2) \quad i=1,\ldots,N_s,\ s=1,\ldots,S, \\ g(\mu_{is}) &= \mathbf{x}_{is}^\top\boldsymbol{\beta}+\psi_{is}, \\ \boldsymbol{\beta} &\sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}). \end{aligned} \tag{8.1}$$

The covariates for the ith individual in the sth spatial unit or other grouping are included in a $p\times 1$ vector $\mathbf{x}_{is}$. The corresponding $p\times 1$ vector of fixed effect parameters are denoted by $\boldsymbol{\beta}$, which has an assumed multivariate Gaussian prior with mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ that can be chosen by the user. Additional random effects relating to the spatial unit or other grouping and the network structure in the data are denoted by ψ_{is}, and can take many different forms. The response for the ith individual in the sth spatial unit Y_{is} come from an exponential family of distributions $f(y_{is}|\mu_{is},\nu^2)$. The `netcmc` package can implement Equation (8.1) for Binomial, Gaussian, and Poisson data and the exact specification of each are given below:

- Binomial: $Y_{is}\sim$ Binomial(n_{is},θ_{is}) and $\ln(\theta_{is}/(1-\theta_{is}))=\mathbf{x}_{is}^\top\boldsymbol{\beta}+\psi_{is}$.

- Gaussian: $Y_{is} \sim \mathrm{N}(\mu_{is}, \nu^2)$ and $\mu_{is} = \mathbf{x}_{is}^\top \boldsymbol{\beta} + \psi_{is}$.
- Poisson: $Y_{is} \sim \mathrm{Poisson}(\mu_{is})$ and $\ln(\mu_{is}) = \mathbf{x}_{is}^\top \boldsymbol{\beta} + \psi_{is}$.

Additionally, `netcmc` can also model an extension of (8.1) for multivariate data (for outcomes $r = 1, \cdots, R$) which is given by

$$\begin{aligned} Y_{isr}|\mu_{isr} &\sim f(y_{isr}|\mu_{isr}, \nu^2) \quad i = 1, \ldots, N_s,\ s = 1, \ldots, S,\ r = 1, \ldots, R, \\ g(\mu_{isr}) &= \mathbf{x}_{is}^\top \boldsymbol{\beta}_r + \psi_{isr}, \\ \boldsymbol{\beta}_r &\sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}). \end{aligned} \tag{8.2}$$

In this model the $p \times 1$ vector of fixed effect parameters relating to the rth response are denoted by $\boldsymbol{\beta}_r$ which has an assumed multivariate Gaussian prior with mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ that can be chosen by the user. Additional random effects allow for correlation between outcome r, the spatial unit or other grouping and the network are denoted by ψ_{isr}, and can take many different forms. As with the univariate model, the response Y_{isr} come from an exponential family of distributions $f(y_{isr}|\mu_{isr}, \nu^2)$. The `netcmc` package can implement Equation 8.2 for Binomial, Gaussian, and Poisson data and are simple extensions to those outlined above. Table 8.1 provides a summary of models that can be implemented in `netcmc`.

8.2.2 Univariate models

This subsection describes the set of univariate Bayesian hierarchical models that can be implemented in the `netcmc` software.

`uni()` is similar to that proposed by Nelder and Wedderburn (1972), and is a standard univariate generalized linear model which is called by the `uni()` function. The model takes

Model	Eq.	Data	Description
`uni()`	(8.3)	Covariate.	This model is similar to that proposed by Nelder and Wedderburn (1972), and is a standard univariate generalized linear model.
`uniNet()`	(8.4)	Covariate, Network.	This model is similar to that proposed by Browne et al. (2001), and represents the social network pattern in the mean response with a single set of weighted random effects.
`uniNetRand()`	(-)	Covariate, Network.	This model is similar to that proposed by Browne et al. (2001), and has the same network random effects structure as the `uniNet()` model. Additionally, this model allows for a set of independent random effects to be incorporated into the modeled.
`uniNetLeroux()`	(8.5)	Covariate, Network, Spatial.	This model represents the spatio-network pattern in the mean response with a set of spatial random effects and weighted random effects. The spatial effect is modeled by the conditional autoregressive prior proposed by Leroux et al. (2000).
`multiNet()`	(8.6)	Covariate, Network.	This model is an extension to that proposed by Browne et al. (2001), and is a multivariate extension of the `uniNet()` model. In this extension, the prior structure of the network random effects is jointly modeled across the multiple responses.
`multiNetRand()`	(8.7)	Covariate, Network.	This model is an extension to that proposed by Browne et al. (2001), and is a multivariate extension of the `uniNetRand()` model. In this extension, the prior structure of these random effects is jointly modeled across the multiple responses.
`multiNetLeroux()`	(-)	Covariate, Network Spatial.	This model is a multivariate extension of the `uniNetLeroux()` model. In this extension, the prior structure of the network random effects is jointly modeled across the multiple responses. However, the spatial random effects are modeled independently.

Table 8.1: Overview of various models available in the `netcmc` package together with the equation numbers (Eq.) defining them mathematically and what data structures they are designed for.

the form given in Equation (8.1) and is given by

$$\psi_{is} = 0. \tag{8.3}$$

Thus there are no spatial or network random effects included in the model.

`uniNet()` is similar to that proposed by Browne et al. (2001), and represents the social network pattern in the mean response with a single set of weighted random effects. This model has the same prior structure as that presented in Equation (4.2) of Chapter 4 with the exclusion of single membership random effects. This model is a univariate multiple classification model which takes the form given in Equation (8.1) and is given by

$$\begin{aligned} \psi_{is} &= \sum_{j \in \text{net}(is)} w_{isj} u_j, \\ u_j &\sim \text{N}(0, \sigma_u^2), \\ \sigma_u^2 &\sim \text{Inverse-Gamma}(\alpha_u, \xi_u), \\ w_{isj} &= \frac{1}{|\text{net}(is)|}. \end{aligned} \tag{8.4}$$

There are J alters, where an alter is a person who was nominated as a friend by an individual. The $J \times 1$ vector of alter random effects are denoted by $\mathbf{u} = (u_1, \ldots, u_J)_{J \times 1}$. net$(is)$ is the set of alters that the ith individual in the sth spatial unit nominates as a friend such that net$(is) \subset \{1, \ldots, J\}$. The weight of the jth multiple membership random effect for the ith individual in the sth spatial unit is given as follows $w_{isj} = \frac{1}{|\text{net}(is)|}$ with the restriction that $|\text{net}(is)| \neq 0$, thus $\sum_{j=1}^{J} w_{isj} = 1$. Conjugate Inverse-Gamma priors are specified for the random effects variance σ_u^2. The corresponding hyperparamaterers (α_u, ξ_u) can be chosen by the user, and the default values in the software are $(\alpha_u = 1, \xi_u = 0.001)$.

`uniNetRand()` has the same prior structure to that presented in Equation (4.2) of Chapter 4.

`uniNetLeroux()` represents the spatio-network pattern in the mean response with a set of spatial random effects and weighted random effects. The spatial effect that an individual experiences is modeled by the conditional autoregressive prior proposed by Leroux et al. (2000). This model builds on the prior structure of the model presented in Equation (4.2) of Chapter 4 with the inclusion of a conditional autoregressive prior. This is a univariate model which takes the form given in Equation (8.1) and is given by

$$
\begin{aligned}
\psi_{is} &= \phi_s + \sum_{j \in \text{net}(is)} w_{isj} u_j, \qquad (8.5)\\
\phi_s|\boldsymbol{\phi}_{-s} &\sim \text{N}\left(\frac{\rho \sum_{l=1}^{S} a_{sl}\phi_l}{\rho \sum_{l=1}^{S} a_{sl} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{l=1}^{S} a_{sl} + 1 - \rho}\right),\\
u_j &\sim \text{N}(0, \sigma_u^2),\\
\tau^2 &\sim \text{Inverse-Gamma}(a_\tau, b_\tau),\\
\rho &\sim \text{Uniform}(0, 1),\\
\sigma_u^2 &\sim \text{Inverse-Gamma}(\alpha_u, \xi_u),\\
w_{isj} &= \frac{1}{|\text{net}(is)|}.
\end{aligned}
$$

The random effect for the sth spatial unit is denoted by ϕ_s and has a conditional autoregressive prior. In Equation (8.5), the study region where the individuals live is partitioned into S non-overlapping areal units $\mathcal{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_S\}$. $\mathbf{A} = (a_{sl})_{S\times S}$ is a non-negative spatial adjacency matrix in Equation (8.5) which defines how spatially close the S areal units are to each other. The elements of $\mathbf{A}_{S\times S}$ can be binary or non-binary. In the more common binary case, $a_{sl} = 1$ if a pair of areal units $(\mathcal{G}_s, \mathcal{G}_l)$ share a common border or are considered neighbours by some other measure, and $a_{sl} = 0$ otherwise. τ^2 is a measure of

the variance relating to the spatial random effects $\boldsymbol{\phi}$. The parameter ρ controls the level of spatial autocorrelation, with values close to one and zero representing strong autocorrelation and independence respectively. A non-conjugate uniform prior on the unit interval is specified for the single level of spatial autocorrelation ρ. Conjugate Inverse-Gamma priors are specified for the random effects variances (τ^2, σ_u^2). The corresponding hyperparamaterers ($\alpha_\tau, \xi_\tau,\ \alpha_u, \xi_u$) can be chosen by the user, and the default values in the software are ($\alpha_\tau = 1, \xi_\tau = 0.001,\ \alpha_u = 1, \xi_u = 0.001$).

8.2.3 Multivariate models

This subsection describes the set of multivariate Bayesian hierarchical models that can be implemented in the `netcmc` software.

`multiNet()` is an extension to that proposed by Browne et al. (2001), and is a multivariate extension of the `uniNet()` model. In this extension, the prior structure of these network random effects is jointly modeled across the multiple responses. This model has a similar prior structure as the multivariate model presented in Chapter 6 with the exclusion of the spatial random effects term, making it suitable for data without this component. This is a multivariate multiple classification model which takes the form given in Equation (8.2) and is given by

$$\begin{aligned} \psi_{isr} &= \sum_{j \in \text{net}_{(is)}} w_{isj} u_{jr}, \qquad (8.6) \\ \mathbf{u}_j = (u_{j1}, \ldots, u_{jR}) &\sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{u}), \\ \boldsymbol{\Sigma}_\mathbf{u} &\sim \text{Inverse-Wishart}(\xi_\mathbf{u}, \boldsymbol{\Omega}_\mathbf{u}), \\ w_{isj} &= \frac{1}{|\text{net}(is)|}. \end{aligned}$$

The $R \times 1$ vector of joint random effects for the jth alter is denoted by $\mathbf{u}_j = (u_{j1}, \ldots, u_{jR})_{R\times 1}$ and has a joint Gaussian distribution. The unstructured covariance matrix $\mathbf{\Sigma_u}$ captures the covariance between the R outcomes at the network level. A conjugate Inverse-Wishart prior is specified for the random effects covariance matrix $\mathbf{\Sigma_u}$. The corresponding hyperparamaterers $(\xi_\mathbf{u}, \mathbf{\Omega_u})$ can be chosen by the user, and the default values in the software are $(\xi_\mathbf{u} = R + 1, \mathbf{\Omega_u} = \mathbf{I})$.

`multiNetRand()` is an extension to that proposed by Browne et al. (2001), and is a multivariate extension of the `uniNetRand()` model. In this extension, the prior structure of these random effects is jointly modelled across the multiple responses. This model has a similar prior structure as the multivariate model presented in Chapter 6 with the exclusion of the spatial random effects term for the addition of a multivariate single classification term. This is a multivariate multiple membership multiple classification model which takes the form given in Equation (8.2) and is given by

$$
\begin{aligned}
\psi_{isr} &= v_{sr} + \sum_{j \in \text{net}(is)} w_{isj} u_{jr}, \qquad (8.7)\\
\mathbf{v}_s = (v_{s1}, \ldots, v_{sR}) &\sim \text{N}(\mathbf{0}, \mathbf{\Sigma_v}),\\
\mathbf{u}_j = (u_{j1}, \ldots, u_{jR}) &\sim \text{N}(\mathbf{0}, \mathbf{\Sigma_u}),\\
\mathbf{\Sigma_v} &\sim \text{Inverse-Wishart}(\xi_\mathbf{v}, \mathbf{\Omega_v}),\\
\mathbf{\Sigma_u} &\sim \text{Inverse-Wishart}(\xi_\mathbf{u}, \mathbf{\Omega_u}),\\
w_{isj} &= \frac{1}{|\text{net}(is)|},
\end{aligned}
$$

where the group random effects are independent across the r responses. The $R \times 1$ vector of joint random effects for the sth single membership classification is denoted by $\mathbf{v}_s = (v_{s1}, \ldots, v_{sR})_{R\times 1}$ and has a joint Gaussian distribution. The unstructured covariance matrix $\mathbf{\Sigma_v}$ captures the covariance between the R outcomes at the single membership level.

Conjugate Inverse-Wishart priors are specified for the random effects covariance matrices ($\boldsymbol{\Sigma}_{\mathbf{v}}$, $\boldsymbol{\Sigma}_{\mathbf{u}}$). The corresponding hyperparameters ($\xi_{\mathbf{v}}, \boldsymbol{\Omega}_{\mathbf{v}}$, $\xi_{\mathbf{u}}, \boldsymbol{\Omega}_{\mathbf{u}}$) can be chosen by the user, and the default values in the software are ($\xi_{\mathbf{v}} = R + 1, \boldsymbol{\Omega}_{\mathbf{v}} = \mathbf{I}$, $\xi_{\mathbf{u}} = R + 1, \boldsymbol{\Omega}_{\mathbf{u}} = \mathbf{I}$).

`multiNetLeroux()` has a similar prior structure as the multivariate model presented in Chapter 6 but differs in that instead of τ_r following a half-normal distribution, τ_r^2 follows an Inverse-Gamma distribution.

8.2.4 Inference

The models in this package are fitted using the Bayesian paradigm, with inference being based on Markov chain Monte Carlo simulation. All parameters whose full conditional distributions are of a closed form are sampled using a Gibbs step, which includes the fixed effects regression parameters $\boldsymbol{\beta}$ and the random effects ($\boldsymbol{\phi}$, $\mathbf{u}$, etc.) when the data are Gaussian, and the variance (σ_v^2, σ_u^2, etc.) and covariance parameters ($\boldsymbol{\Sigma}_{\boldsymbol{\phi}}$, $\boldsymbol{\Sigma}_{\mathbf{u}}$, etc.) for all data likelihoods. The remaining parameters are updated using Metropolis and Metropolis-Hastings steps, which are automatically tuned to have acceptance rates of between 30%-50%. The functions used to implement the Markov chain Monte Carlo algorithms are written in `R`, with the computationally intense updating steps written in more computationally efficient `C++` using the `R` package `Rcpp`. In addition, the sparsity of adjacency matrices $\mathbf{A}$ and $\mathbf{W}$ relating to the spatial and social network structures present in the data are converted into triplet form when updating of random effects, which increases the efficiency of the software.

The validity of inference based on Markov chain Monte Carlo simulations are subject to samples being accurately drawn from the target posterior distribution. Thus determining if a Markov chain has converged to its target distribution is of major importance. `netcmc` provides a number of visual diagnostic checks to assess convergence of the simulated posterior distributions of parameters using trace, density, and ACF plots through the `plot()` function.

8.3 Installing and using the software

8.3.1 Installing the software

`netcmc` is a package for the statistical computing environment `R` (R Core Team, 2013) and can be downloaded from CRAN (Gerogiannis, 2022) and GitHub (`https://github.com/GNG3/netcmc`) for macOS, Linux, and Windows platforms. The package requires `R` ($\geq$ 3.6.1) and depends on packages `Rcpp` (Eddelbuettel and François, 2011), `RcppArmadillo` (Eddelbuettel and Sanderson, 2014), `RcppProgress`. Additionally, the package imports functionality from `coda` (Plummer et al., 2006), `ggplot2` (Wickham, 2016) and `mvtnorm` (Genz et al., 2020). Once the package has been installed, it can be loaded using the command `library("netcmc")`.

8.3.2 Using the software

The software can be used to fit 4 different types of univariate models: `uni()`, `uniNet()`, `uniNetRand()` and `uniLeroux()`, and 3 different types of multivariate models: `multiNet()`, `multiNetRand()` and `multiNetLeroux()`, which are variations of the models for in Chapters 4 and 6. The availability of these easy-to-use models will help to strengthen the amount of research that can be conducted in this area, allowing researchers to conduct studies similar to the studies presented in Chapters 4 and 6 of this thesis for a range of likelihoods. Full details of the arguments available for each function are provided in the help files, but the main arguments required for an analysis using default priors are as follows.

- `formula:` A formula for the covariate part of the model using a similar syntax to that used in the `lm()` function. For the multivariate model the response is read in as a matrix with R columns.
- `family:` The data likelihood model that must be `"binomial"`, `"gaussian"` or `"pois-`

son".

- `trials`: This is only required if `family = "binomial"`, and is a vector of the same length and in the same order as the response containing the total number of trials for each data point.

- `A`: This is only required for the spatial models, and is an $S \times S$ symmetric and non-negative neighbourhood matrix whose row sums must all be positive. Typically a binary specification is used, where the slth element a_{sl} equals one if areas ($\mathcal{G}_s$, $\mathcal{G}_s$) are spatially close (e.g. share a common border) and is zero otherwise.

- `spatialAssignment`: The $\sum_{s=1}^{S} N_s \times S$ binary matrix of individual's assignment to spatial areas used in the model fitting process. This output is only applicable for models that have a spatial/grouping component, namely `uniNetRand()`, `uniNetLeroux()`, `multiNetRand()` and `multiNetLeroux()`.

- `W`: A $\sum_{s=1}^{S} N_s \times J$ matrix that encodes the social network structure and whose rows sum to 1. Note, the order of the rows in this matrix must correspond to the order of the data points in the formula argument. Also, this matrix is not required in the `uni()` model.

- `numberOfSamples`: The number of MCMC samples to generate in total before thinning of the Markov chain or removing of the burn-in period.

- `burn-in`: The number of MCMC samples to discard as the burn-in period.

- `thin`: The value by which to thin the MCMC samples.

When a model is run within the package the user is updated on its progress via a progress bar in the `R` console. Then when the model has finished the `netcmc` functions `summary()` and `plot()` can be applied to elements of the model object to summarize the results.

- `summary()`: Returns summaries and diagnostics for the posterior distribution of each parameter in the model excluding the random effects.

- `plot()`: Returns trace, kernel density, and autocorrelation function (ACF) plots for the samples of a parameter(s) from its posterior distribution.

Once the model has finished running the model object returned is a list with the following components.

- `call`: A text string containing the function call including the arguments input by the user.

- `y`: The $(R\sum_{s=1}^{S} N_s) \times 1$ response vector used in the model fitting process.

- `X`: The $\sum_{s=1}^{S} N_s \times p$ design matrix of covariates used in the model fitting process.

- `squareSpatialNeighbourhoodMatrix`: The $S \times S$ binary spatial neighbourhood matrix $\mathbf{A}$ used in the model fitting process. This output is only applicable for models that have a spatial component, namely `uniNetLeroux()` and `multiNetLeroux()`.

- `spatialAssignment`: The $\sum_{s=1}^{S} N_s \times S$ binary matrix of individual's assignment to spatial areas used in the model fitting process. This output is only applicable for models that have a spatial/grouping component, namely `uniNetRand()`, `uniNetLeroux()`, `multiNetRand()` and `multiNetLeroux()`.

- `W`: The $\sum_{s=1}^{S} N_s \times J$ social network matrix used in the model fitting process. This output is only applicable for models that have a network component, namely `uniNet()`, `uniNetRand()`, `uniNetLeroux()`, `multiNet()`, `multiNetRand()` and `multiNetLeroux()`.

- `samples`: The matrix of MCMC samples generated from the posterior distribution of each parameter in the model excluding the random effects.

- `spatialRandomEffectsSamples`: The matrix of MCMC samples generated from the posterior distribution of the spatial/grouping random effects in the model. This output is only applicable for models that have a spatial/grouping component, namely `uniNetRand()`, `uniNetLeroux()`, `multiNetRand()` and `multiNetLeroux()`.

- `uRandomEffectsSamples`: The matrix of MCMC samples generated from the posterior distribution of the network random effects in the model. This output is only applicable for models that have a network component, namely `uniNet()`, `uniNetRand()`, `uniNetLeroux()`, `multiNet()`, `multiNetRand()` and `multiNetLeroux()`.

- `acceptanceRates`: The acceptance rates of the model parameters from the MCMC run excluding the random effects.

- `spatialRandomEffectsAcceptanceRate`: The acceptance rates of the spatial/grouping random effects from the MCMC run. This output is only applicable for models that have a spatial/grouping component, namely `uniNetRand()`, `uniNetLeroux()`, `multiNetRand()` and `multiNetLeroux()`.

- `uRandomEffectsAcceptanceRate`: The acceptance rates of the network random effects from the MCMC run. This output is only applicable for models that have a network component, namely `uniNet()`, `uniNetRand()`, `uniNetLeroux()`, `multiNet()`, `multiNetRand()` and `multiNetLeroux()`.

- `timeTaken`: The time taken to fit the model.

The remainder of this chapter illustrates the `netcmc` package via a small simulation study to illustrate the correctness of the MCMC algorithms, as well as a worked example.

8.4 Simulation study

This section illustrates the correctness of the `netcmc` implementation of the `uniNet()` model with a Poisson data likelihood, through generating 100 simulated data sets with known parameters and summarizing the bias, 95% coverage probabilities, and average effective sample size (ESS, approximate number of independent samples) of the estimated model parameters. In this study `netcmc` is compared to the equivalent model available in `R2MLwiN`, which is the most commonly used software for fitting this type of model.

8.4.1 Data generation

In what follows all 100 simulated data sets have 1 response (`R` = 1) for `N` = 500 individuals, 3 covariates including an intercept term, and 50 alters (`J` = 50). This information is specified by

```
R> N = 500
R> J = 50
R> p = 3
```

The true value of the `p` × 1 vector of fixed effects $\boldsymbol{\beta}$ has each of its elements drawn independently from a Gaussian distribution with mean 0.5 and standard deviation 0.1, while the covariates (excluding the intercept term) are drawn independently from a Gaussian distribution with mean 0 and standard deviation 1. This is all specified by

```
R> beta = rnorm(p, mean = 0.5, sd = 0.1)
R> Covariates = matrix(rnorm(2*N, 0, 1), ncol = (p - 1), nrow = N)
R> X = cbind(1, Covariates)
```

The social network random effects variance is fixed at $\sigma^2_{\mathbf{u}} = 0.5$, and the corresponding random effects $\mathbf{u}$ can be created and mean centred using the following R code.

```
R> sigmaSquaredU = 0.5
R> u = rnorm(J, 0, sqrt(sigmaSquaredU))
R> u = u - mean(u)
```

The network structure encoded in the matrix $\mathbf{W}_{\mathtt{N}\times\mathtt{J}}$ can be created using the R code below, which consists of 4 lines of code followed by a loop and a final line of code. The 1^{st} line of code creates an $\mathtt{N} \times \mathtt{J}$ matrix with entries of ones and zeros with probability 0.05 and 0.95 respectively, which assigns the N individuals to a subset of the J alters. As the ones and zeros were randomly sampled to create $\mathbf{W}$, there is a possibility that some individuals may not have an alter. Thus the 2^{nd} and 3^{rd} lines identify which (and how many) individuals do not have alters. Then the 4^{th} line randomly samples alters for these individuals to have, and the $\mathbf{W}$ matrix is updated in the for loop. In the final line of code, the matrix is row normalized to produce the final $\mathbf{W}_{\mathtt{N}\times\mathtt{J}}$. Note, in this example every individual has at least one alter (peer in the network), and thus isolation effects are not required.

```
R> W = matrix(rbinom(N * J, 1, 0.05), ncol = J)
R> numberOfIndividualsWithNoAlters = sum(apply(W, 1, function(x) { sum(x) == 0 }))
R> individualsWithNoAlters = which(apply(W, 1, function(x) { sum(x) == 0 }))
R> peers = sample(1:J, numberOfIndividualsWithNoAlters, TRUE)
R>      for(i in 1:numberOfIndividualsWithNoAlters)
+       {
+       W[individualsWithNoAlters[i], peers[i]] = 1
+       }
R> W = t(apply(W, 1, function(x) { x / sum(x) }))
```

Finally, the response vector $\mathbf{Y}$ is generated by first calculating the corresponding vector

of Poisson means $\boldsymbol{\mu} = \exp(\mathbf{X}_{\mathtt{N}\times\mathtt{p}}\boldsymbol{\beta}_{\mathtt{p}\times 1} + \mathbf{W}_{\mathtt{N}\times\mathtt{J}}\mathbf{u}_{\mathtt{J}\times 1})$, and then drawing samples from a Poisson distribution with these means. This can be specified using the following R code.

```
R> logTheta  =  X %*% beta + W %*% u
R> y = rpois(n = N, lambda = exp(logTheta))
```

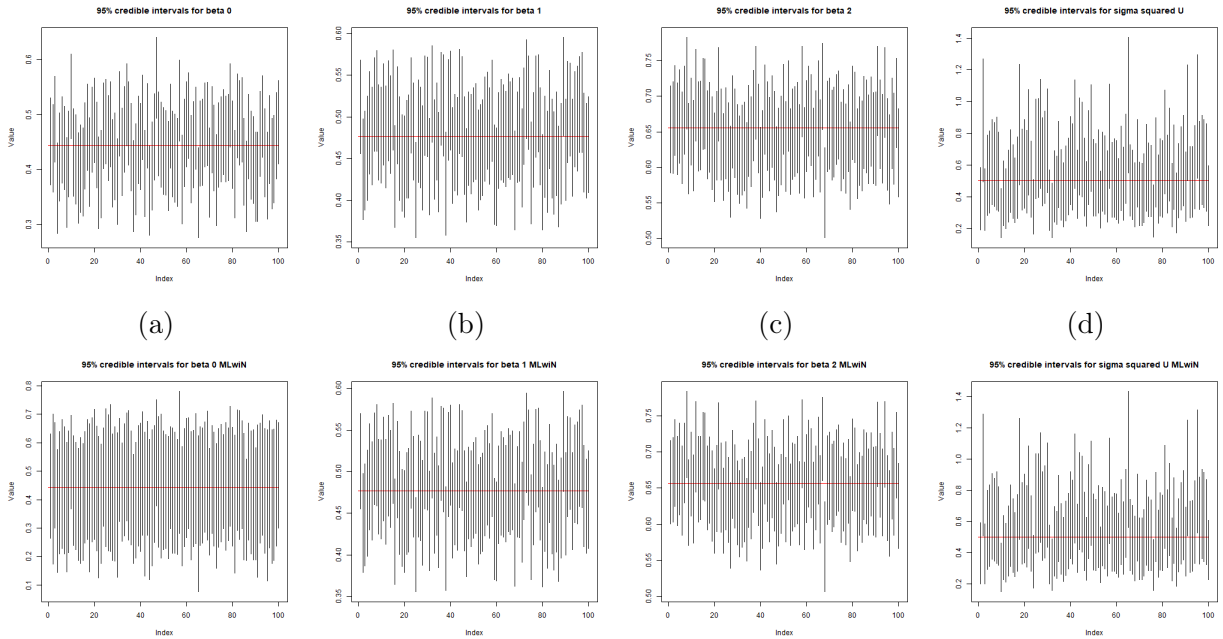
8.4.2 Results

netcmc and R2MLwiN were used to fit model (8.1) and (8.4) to each of the 100 simulated data sets generated as outlined above. In each case the models were run for 400,000 iterations with a burn-in period of 200,000 and a thinning value of 1, resulting in 200,000 samples for inference. The results from the study are presented in Table 8.2, which displays the bias, 95% coverage probabilities, and average ESS for β_0, β_1, β_2, $\sigma^2_{\mathbf{u}}$ and $\mathbf{u}$ averaged over all 100 simulated data sets.

Table 8.2 shows that both netcmc and R2MLwiN produce largely unbiased parameter estimates in all cases, while the corresponding coverage probabilities are all close to their nominal 0.95 level, suggesting that the 95% credible intervals have the correct width. Figure 8.1 displays the 95% credible intervals over the 100 simulated data sets for β_0, β_1, β_2, and $\sigma^2_{\mathbf{u}}$ produced by netcmc and R2MLwiN. The plots show that over the 100 simulated data sets the 95% credible intervals produced for the fixed effects and network variance have similar widths for both netcmc and R2MLwiN. Thus we have illustrated that for the uniNet() model that can be fitted by both software packages, the results are similar and show negligible bias and appropriate uncertainty quantification, suggesting that it has been appropriately implemented in netcmc. The average time taken for netcmc and R2MLwiN to fit the model to the data were 359.57 and 19.40 seconds respectively.

		Bias		Coverage probability		$\overline{\text{ESS}}$	
Parameter	True value	`netcmc`	`R2MLwiN`	`netcmc`	`R2MLwiN`	`netcmc`	`R2MLwiN`
β_0	0.444	0.001	-0.001	0.95	1	8479	2125
β_1	0.477	0.001	0.001	0.99	0.99	13052	24858
β_2	0.656	-0.005	-0.002	0.99	0.97	10500	21576
$\sigma^2_{\mathbf{u}}$	0.5	0.003	0.014	0.94	0.93	30827	36480
$\mathbf{u}$	-	1.44×10^{-18}	1.81×10^{-4}	0.956	0.9644	46293	16690

Table 8.2: Summary of the simulation study to assess the bias, ESS, and 95% coverage probabilities of the parameter estimates from the `uniNet()` model with a Poisson likelihood for `netcmc` and `R2MLwiN`. All results are based on 100 simulated data sets generated as outlined above.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 8.1: 95% credible intervals over the 100 simulated data sets for β_0 (8.1a, 8.1e), β_1 (8.1b, 8.1f), β_2 (8.1c, 8.1g), and $\sigma^2_{\mathbf{u}}$ (8.1d, 8.1h) produced by `netcmc` and `R2MLwiN` respectively.

8.5 Illustrative example: Adolescent delinquent behaviour in the Netherlands

We illustrate the `netcmc` software by modeling the effects of a social network on the prevalence of delinquent behaviour amongst 151 adolescents belonging to one of 4 schools in the Netherlands from 1994. Our analysis has three main aims:

A1 Which factors significantly affect delinquent behaviour in adolescents?

A2 How much variation in the delinquent behaviour can be explained by the social network component of the model?

A3 Does a network based modelling approach perform better than a non-network based approach for modelling the delinquent behaviour in these adolescents?

8.5.1 Data and exploratory analysis

The data come from the social behavior study conducted by Houtzager and Baerveldt (1999), and are available from https://www.stats.ox.ac.uk/ snijders/siena/BaerveldtData.html. The data come in 8 separate parts (`N34_1.DAT`, `CBE1.DAT`, `N34_3.DAT`, `CBE3.DAT`, `N34_4.DAT`, `CBE4.DAT`, `N34_6.DAT`, `CBE6.DAT`), which are described in Table 8.3. Note, each of the 4 schools contributes 2 of these files.

<table>
<tr><th>Data file</th><th>Description</th></tr>
<tr><td>N34_h.DAT</td><td>This file contains data on the individual adolescents from the hth school, and contains the following variables.
<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td>gender</td><td>A binary variable indicating the gender of the adolescent, where 1 = girls and 2 = boys.</td></tr>
<tr><td>importance</td><td>A factor variable denoting the importance of school friends, which varies from 1 = very important to 4 = unimportant. Note, no adolescent had a value of 4 for this variable, so in practice this variable has 3 levels $\{1, 2, 3\}$.</td></tr>
<tr><td>delinquency</td><td>A continuous measure of delinquent behaviour, specifically the number of minor offences that the respondent states to have committed transformed by the formula ln(1-x). This transformation is to be undone to turn the continuous measure back to a count.</td></tr>
</table></td></tr>
<tr><td>CBEh.DAT</td><td>This file is a social network matrix for the adolescents in the hth school. Specifically, peers are defined as giving and receiving emotional support, i.e. there is a connections from adolescent i to adolescent j if i says that they receive and/or give support to j.</td></tr>
</table>

Table 8.3: Description of data files.

After having downloaded the data from the online repository, the 8 files can be read into R using the following lines of code.

```
R> CB01.w1=as.matrix(read.table("N34_1.DAT"))
```

```
R> CB01.m=as.matrix(read.table("CBE1.DAT"))
R> CB03.w1=as.matrix(read.table("N34_3.DAT"))
R> CB03.m=as.matrix(read.table("CBE3.DAT"))
R> CB04.w1=as.matrix(read.table("N34_4.DAT"))
R> CB04.m=as.matrix(read.table("CBE4.DAT"))
R> CB06.w1=as.matrix(read.table("N34_6.DAT"))
R> CB06.m=as.matrix(read.table("CBE6.DAT"))
```

Once the 8 files have been loaded into the R workspace, the response and covariate data for the 4 schools can be combined and formatted using the following R code

```
R> school = c(rep(1, nrow(CB01.m)), rep(2, nrow(CB03.m)),
              rep(3, nrow(CB04.m)), rep(4, nrow(CB06.m)))
R> data = as.data.frame(rbind(CB01.m, CB03.m, CB04.m, CB06.m))
R> data = as.data.frame(cbind(data, school))
R> colnames(data) = c("gender", "delinquency", "importance", "school")
R> data = ifelse(data == 1, "girls", "boys")
```

where the final line recodes the gender covariate. Then the continuous `delinquency` variable can be transformed back into a count of the number of offences which the respondent states to have committed using the following R code.

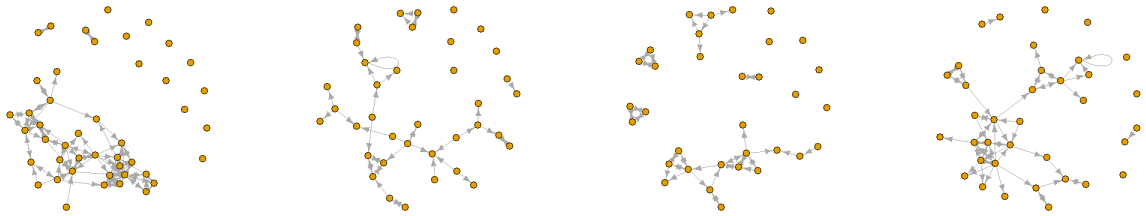
```
R> data$delinquency = round(exp(data$delinquency) - 1)
```

In order to illustrate the social network connections within the 4 schools we employ the `igraph` package. For example, the social network structure within school 1 is shown in Figure 8.2a and can be created using the following R code. Analogous plots for the other three schools are also shown in the same figure and the R code is not shown for brevity.

```
R> library(igraph)
R> graph = igraph::graph.adjacency(CB01.w1, mode = "directed")
R> layout = igraph::layout.fruchterman.reingold(graph)
R> plot(graph, layout = layout, vertex.size = 7, vertex.label=NA)
```

Then the adjacency matrices corresponding to these social network graphs for the 4 schools can be combined using the following R code.

```
R> networkData = as.data.frame(adiag(CB01.w1, CB03.w1, CB04.w1, CB06.w1))
```



(a) (b) (c) (d)

Figure 8.2: The network within school 1 (8.2a), school 2 (8.2b), school 3 (8.2c), and school 4 (8.2d).

As shown in Figure 8.2 there are instances of isolates in each of the 4 schools, which are individuals who have not nominated any alters (friends) as providing emotional support in the survey. To handle these isolate adolescents we assign them an `isolate` column in the social network matrix contained in `networkData`, which can be done using the following R code.

```
R> networkData$isolate = ifelse(rowSums(networkData) == 0, 1, 0)
```

Finally, this social network structure encoded in `networkData` is row normalized to produce the final matrix `W` using the following R code.

```
R> W = networkData/rowSums(networkData)
```

8.5.2 Non-network modeling

In order to assess whether a social network model fits the data better than a simpler non-network model (**A3**), we first fit the following simple Poisson generalized linear model to the data.

$$\begin{aligned}
\texttt{delinquency}_{i_s} &\sim \text{Poisson}(\mu_{i_s}) \quad i = 1, \ldots, N_s,\ s = 1, \ldots, S, \\
\log(\mu_{i_s}) &= \beta_0 + \texttt{gender}_{i_s}\beta_1 + \texttt{importance2}_{i_s}\beta_2 + \texttt{importance3}_{i_s}\beta_3 + \texttt{school2}_{i_s}\beta_4 \\
&+ \texttt{school3}_{i_s}\beta_5 + \texttt{school4}_{i_s}\beta_6, \\
\boldsymbol{\beta} &\sim \text{N}(\mathbf{0}, 10^5\mathbf{I}).
\end{aligned} \tag{8.8}$$

Here $\texttt{delinquency}_{i_s}$ is the count version of the delinquency measure of the ith adolescent in the sth school, while β_1 denotes the effect of being a girl (boy is the baseline). Additionally, (β_2, β_3) denote the effects of being in levels 2 and 3 of the importance variable, while level 1 is the baseline. $(\beta_4, \beta_5, \beta_6)$ denote the effects of being in levels 2, 3 and 4 of the school variable, while school 1 is the baseline. Model (8.8) is an example of (8.3) with a Poisson data likelihood, and thus can be implemented in `netcmc` using the following `R` code.

```
R> model = uni(formula = delinquency ~ factor(gender) + factor(importance) +
+              factor(school),
+              data = data,
+              family = "poisson",
+              numberOfSamples = 200000,
+              burnin = 200000,
+              thin = 10,
+              seed = 1)
```

Inference for this model is based on 20,000 MCMC samples, which were obtained by running the chain for 400,000 iterations, with 200,000 being discarded as the burn-in period and the remaining 200,000 being thinned by 10 to reduce the autocorrelation and size of output that is stored in the computer's memory. When the results from the model are printed using the `summary()` function, the output displayed below is produced.

```
R> summary(model)


Call:
uni(formula = delinquency ~ factor(gender) + factor(importance) +
factor(school), data = data, family = "poisson", numberOfSamples = 2e+05,
burnin = 2e+05, thin = 10, seed = 1)


MCMC Coefficients:
                      Mean Variance   2.5%  Median  97.5%  ESS Accept. %
(Intercept)          3.134    0.003  3.024   3.135  3.240 8315     39.99
factor(gender)girls -1.164    0.005 -1.296  -1.164 -1.034 4549     39.99
factor(importance)2 -0.236    0.004 -0.354  -0.235 -0.116 6149     39.99
factor(importance)3 -0.619    0.011 -0.833  -0.617 -0.415 4545     39.99
factor(school)2     -0.591    0.006 -0.745  -0.592 -0.438 5118     39.99
factor(school)3     -0.290    0.005 -0.431  -0.289 -0.149 5957     39.99
factor(school)4     -0.727    0.006 -0.883  -0.727 -0.571 4624     39.99


MCMC Diagnostics:


                    Geweke Z
(Intercept)           -0.790
factor(gender)girls    0.691
```

```
factor(importance)2    1.035
factor(importance)3    0.151
factor(school)2        0.243
factor(school)3        0.329
factor(school)4       -1.401

MCMC Model Information:

Number of observations: 151
DIC: 1252.931, D.bar: 1245.928
pd: 7.003, log likelihood: -619.463

MCMC Information:
Number of samples: 20000, Burn-in: 2e+05
Thin: 10, Time Taken: 15.58 Secs
```

This model summary is partitioned into 5 parts. The first part prints the user specified function that was used to run the model and falls under the section `Call:`. The second part falls under the `MCMC Coefficients:`, and includes the posterior mean (`Mean`), posterior variance (`Variance`), posterior median (`Median`), and 95% credible intervals (`2.5%`, `97.5%`) for all parameters excluding the random effects. In addition to this, the effective sample sizes (`ESS`) and acceptance rate (`Accept.%`) for each parameter is provided. The third part falls under the `MCMC Diagnostics:`, and includes output of the convergence diagnostic proposed by Geweke (1991) (`Geweke Z`). The fourth part falls under the `MCMC Model Information:`, and includes output relating to the number of observations and overall model fit including the deviance information criteria (DIC, Spiegelhalter et al., 2002) and the effective number of independent parameters in the model (`pd`). The fifth and final part falls under the `MCMC Information:`, and includes information regarding the number of samples used for inference

(`Number of samples`), the length of burn-in (`Burn-in`), the amount of thinning (`Thin`), and the time taken to run the model (`Time Taken`).

Convergence of the MCMC samples from this model was assessed by examining trace plots for the posterior samples for each parameter. Traceplots are produced using the `plot()` function and are not shown for brevity, but this function is illustrated for the network model fitted in the next section.

The output from Model (8.8) shows that the effect of `gender` is statistically significant, as the 95% credible interval for its effect does not contain 0. Holding everything else constant, in comparison to `boys` the relative rate of delinquency for `girls` is around a third ($\exp(\beta_1) = 0.31$, 95% C.I.: (0.27, 0.36)). The effects of both `importance2` and `importance3` are also statistically significantly different from that of `importance1`, as the 95% credible intervals for each of these parameters do not contain 0. Holding everything else constant, in comparison to adolescents with a value of `importance1`, adolescents with a value of `importance2` had a lower observed rate of delinquency ($\exp(\beta_2) = 0.79$, 95% C.I.: (0.70, 0.89)), around a 21% reduction.

The validity of the parameter estimates from `netcmc` were assessed by fitting the same model in the `R2MLwiN` software. The results of this comparison are displayed in Table 8.4, which shows the posterior means from the two software packages as well as the percentage absolute difference relative to the larger of the two estimates. Overall, the table shows good agreement between the two sets of point estimates, with percentage differences equal to or less than 0.69% for each parameter.

Parameter	netcmc	95% C.I.	R2MLwiN	95% C.I.	% difference
(Intercept)	3.134	(3.024, 3.240)	3.133	(3.023, 3.240)	0.03
gender	-1.164	(-1.296, -1.034)	-1.162	(-1.294, -1.032)	0.17
importance2	-0.236	(-0.354, -0.116)	-0.236	(-0.355, -0.117)	0
importance3	-0.619	(-0.833, -0.415)	-0.621	(-0.834, -0.414)	0.32
school2	-0.591	(-0.745, -0.438)	-0.589	(-0.746, -0.438)	0.34
school3	-0.290	(-0.431, -0.149)	-0.288	(-0.428, -0.147)	0.69
school4	-0.727	(-0.883, -0.571)	-0.725	(-0.878, -0.572)	0.28

Table 8.4: Comparison of the parameter estimates (posterior means) and 95% credible intervals from the netcmc and the R2MLwiN software packages. The final column displays the percentage difference in the estimates relative to the larger of the two estimates.

8.5.3 Network modeling

The residual social network effects that are likely to be present in the data can be estimated by Model (8.4) via the uniNet() function within netcmc. The specific model fitted is given by

$$
\begin{aligned}
\texttt{delinquency}_{i_s} &\sim \text{Poisson}(\mu_{i_s}) \quad i = 1, \ldots, N_s,\ s = 1, \ldots, S, \qquad (8.9)\\
\log(\mu_{i_s}) &= \beta_0 + \texttt{gender}_{i_s}\beta_1 + \texttt{importance2}_{i_s}\beta_2 + \texttt{importance3}_{i_s}\beta_3 + \texttt{school2}_{i_s}\beta_4\\
&+ \texttt{school3}_{i_s}\beta_5 + \texttt{school4}_{i_s}\beta_6 + \sum_{j \in \text{net}(i_s)} w_{i_s j} u_j + w^*_{i_s} u^*,\\
\boldsymbol{\beta} &\sim \text{N}(\mathbf{0}, 10^5\mathbf{I}),\\
u_j &\sim \text{N}(0, \sigma_u^2) \quad j = 1, ..., J,\\
u^* &\sim \text{N}(0, \sigma_u^2),\\
\sigma_u^2 &\sim \text{Inverse-Gamma}(0.001, 0.001),
\end{aligned}
$$

and is an example of Model (8.4) with a Poisson data likelihood. It can be implemented in netcmc using the following R code.

```
R> modelNetwork = uniNet(formula = delinquency ~ factor(gender) + factor(importance) +
+                        factor(school),
+                        data = data,
+                        W = W,
+                        family = "poisson",
+                        numberOfSamples = 200000,
+                        burnin = 200000,
+                        thin = 10,
+                        a2 = 0.001,
+                        b2 = 0.001,
+                        seed = 1)
```

The results of this model can again be printed using the summary() function as follows:

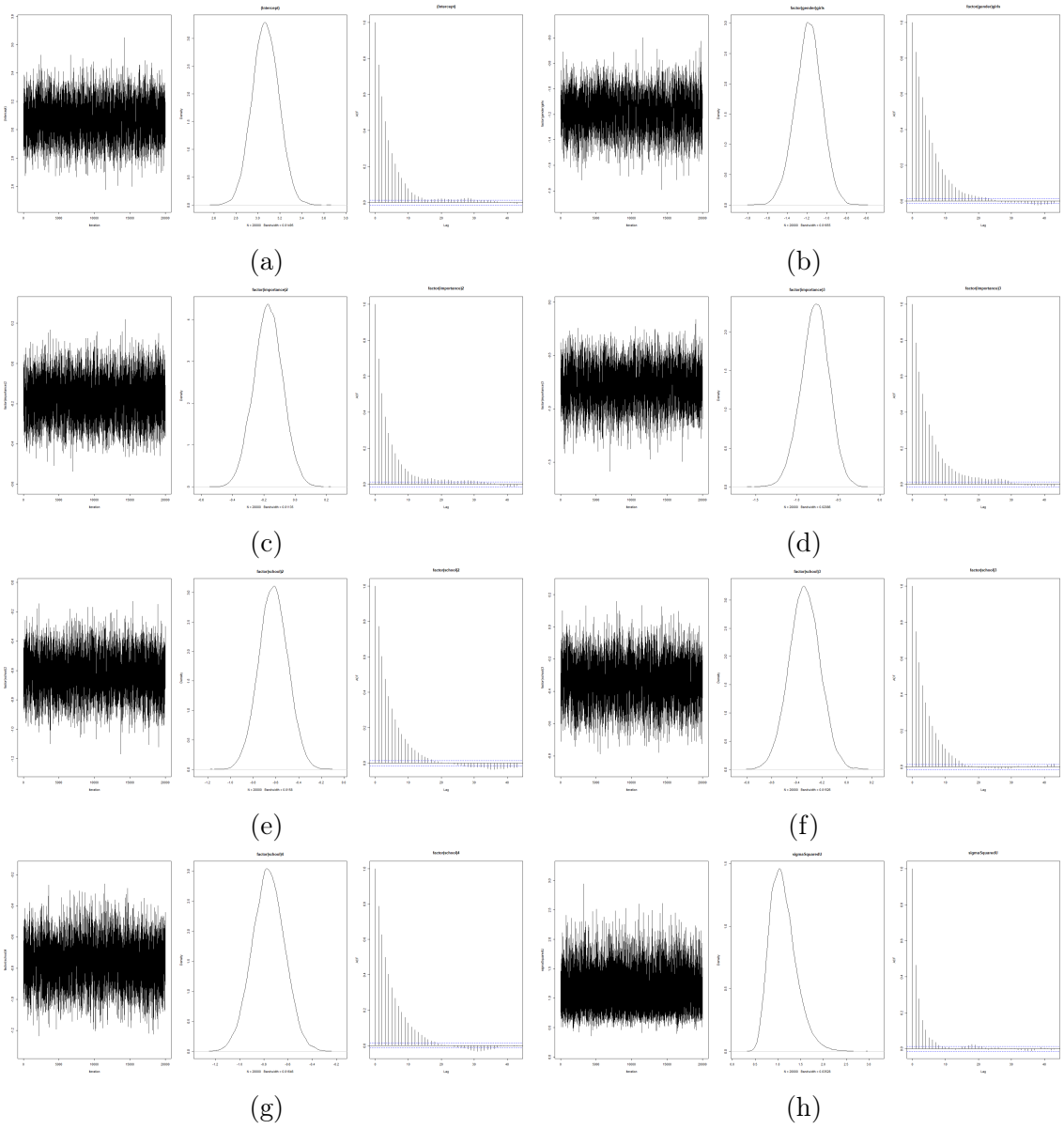
```
R> summary(modelNetwork)

Call:

uniNet(formula = delinquency ~ factor(gender) + factor(importance) +
factor(school), data = data, family = "poisson", W = as.matrix(W),
numberOfSamples = 2e+05, burnin = 2e+05, thin = 10, seed = 1,
a2 = 0.001, b2 = 0.001)

MCMC Coefficients:
                      Mean Variance   2.5%  Median   97.5%  ESS Accept. %
```

```
(Intercept)           3.070    0.014  2.834   3.069   3.305 2672     39.51
factor(gender)girls  -1.185    0.018 -1.462  -1.183  -0.925 1810     39.51
factor(importance)2  -0.173    0.008 -0.353  -0.173   0.009 3041     39.51
factor(importance)3  -0.784    0.029 -1.125  -0.780  -0.460 2244     39.51
factor(school)2      -0.622    0.016 -0.872  -0.623  -0.372 2309     39.51
factor(school)3      -0.339    0.015 -0.581  -0.339  -0.098 2535     39.51
factor(school)4      -0.763    0.018 -1.023  -0.763  -0.505 2213     39.51
sigmaSquaredU         1.118    0.085  0.653   1.082   1.790 5832    100.00

MCMC Diagnostics:
                    Geweke Z
(Intercept)           -1.475
factor(gender)girls    1.066
factor(importance)2   -0.979
factor(importance)3   -0.245
factor(school)2        1.407
factor(school)3        2.136
factor(school)4        0.846
sigmaSquaredU         -2.241

MCMC Model Information:

Number of observations: 151
Number of network random effects: 152

DIC: 1053.537, D.bar: 978.783
pd: 74.754, log likelihood: -452.014
```

```
MCMC Information:


Number of samples: 20000, Burn-in: 2e+05
Thin: 10, Time Taken: 350.83 Sec
```

This time we illustrate MCMC convergence checking via traceplots, which is achieved using the `plot()` function to produce Figures 8.3a, 8.3b, 8.3c, 8.3d, 8.3e, 8.3f, 8.3g, and 8.3h via the following code.

```
R> plot(modelOnePoissonNetwork)
```



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 8.3: Trace, density, and ACF plots from the `plot()` function for β_0 (8.3a), β_1 (8.3b), β_2 (8.3c), β_3 (8.3d), β_4 (8.3e), β_5 (8.3f), β_6 (8.3g), σ_u^2 (8.3h).

In all cases the trace plots suggest the MCMC samples have converged, which allows us to now interpret our findings. The output from Model (8.9) shows that the effect of `gender` is still statistically significant, as the 95% credible interval for the effect does not contain 0. Holding everything else constant, in comparison to `boys`, `girls` have a lower relative rate of delinquency by $\exp(\beta_1) = 0.31$, 95% C.I.: (0.23, 0.40), which is a slightly larger effect size compared to that obtained from the non-network model in the previous

section. The effect of `importance3`, relative to the baseline level `importance1`, also remains statistically significant, as the 95% credible intervals for the regression parameter does not include 0. Holding everything else constant, in comparison to adolescents with a value of `importance1`, adolescents with a value of `importance3` have a lower relative rate of delinquency of $\exp(\beta_3) = 0.46$, 95% C.I.: (0.32, 0.63).

The output from Model (8.9) also shows that the posterior mean of the network random effects variance is $\sigma_u^2 = 1.118$ with a 95% credible interval of $(0.65, 1.79)$. As this variance is not close to zero it suggests that there are substantial network effects present in the data that this model is capturing. This is confirmed by comparing the DIC values from models (8.8) and (8.9), which quantify the overall fit to the data of each model. The model including the social network component has a value that is lower by 15.9% compared to the covariate only model, with DIC values of 1,252.931 and 1,053.537 respectively. The model outputs also show that the effective number of independent parameters rises from 7.00 to 74.75 when incorporating the social network component into the model, again suggesting that it plays a sizeable role in explaining adolescents propensity to delinquency.

Figure 8.4 shows the 95% credible intervals for the network alter random effects $\mathbf{u}$ from the model, ordered by posterior medians. There are 8 adolescents whose 95% credible intervals are strictly positive and thus have the effect of increasing the rate of minor offences that adolescents that they are connected with commit. These 8 adolescents have themselves have committed 1, 40, 40, 13, 9, 2, 3, and 12 minor offenses, while only two other adolescents in the data have committed more than 40 minor offences, with values of 43 and 44. In contrast, there were 7 adolescents whose 95% credible intervals are strictly negative and thus have the effect of decreasing the rate of minor offences that adolescents that they are connected with commit. These 7 adolescents have themselves committed 1, 2, 2, 3, 13, 2, and 0 minor offenses.

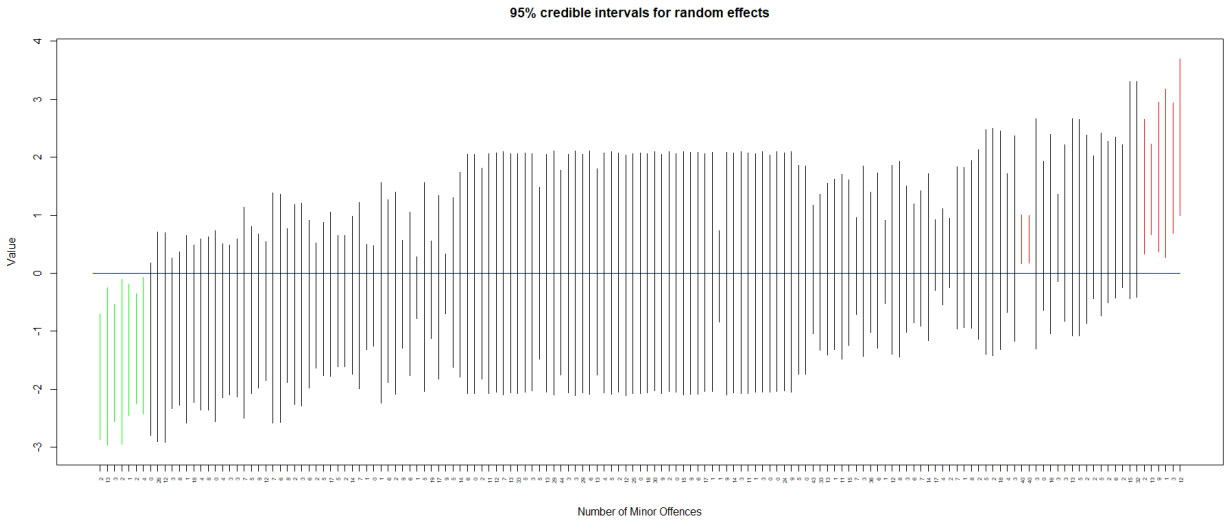


Figure 8.4: 95% credible intervals for the 151 random effects (excluding the isolation random effect), ordered by posterior medians. Those whose interval is entire positive are shown in red and those whose interval is entirely negative are shown in green.

Finally, a comparison of the parameter estimates and 95% credible intervals for this model when fitted in `netcmc` and `R2MLwiN`, with the same choice of priors, is given in Table 8.5, which again shows good agreement between the estimates from the two software packages. The times taken to run the model were 350.83 seconds and 12.63 seconds for `netcmc` and `R2MLwiN` respectively.

Parameter	netcmc	95% C.I.	R2MLwiN	95% C.I.	% difference
(Intercept)	3.070	(2.834, 3.305)	3.071	(2.775, 3.351)	0.03
gender	-1.185	(-1.462, -0.925)	-1.184	(-1.399, -0.973)	0.08
importance2	-0.173	(-0.353, 0.009)	-0.172	(-0.327, -0.018)	0.58
importance3	-0.784	(-1.125, -0.460)	-0.781	(-1.059, -0.507)	0.38
school2	-0.622	(-0.872, -0.372)	-0.626	(-0.838, -0.416)	0.64
school3	-0.339	(-0.581, -0.098)	-0.341	(-0.549, -0.140)	0.59
school4	-0.763	(-1.023, -0.505)	-0.769	(-0.993, -0.551)	0.79
sigmaSquaredU	1.118	(0.653, 1.790)	1.139	(0.670, 1.813)	1.88

Table 8.5: Comparison of the parameter estimates (posterior means) and 95% credible intervals from the netcmc and the R2MLwiN software packages. The final column displays the percentage difference in the estimates relative to the larger of the two estimates.

8.6 Discussion

This chapter has presented the netcmc package, which is the first R package dedicated to fitting network models, and the first software package that specialises in fitting spatio-network models. This chapter has outlined the library of models that can be implemented in the software along with the with the types of sampling schemes used to fit these models. It then illustrated the correctness of the MCMC estimation algorithms with a small simulation study, before applying the models to a worked exemplar. In comparison to R2MLwiN, the main advantage of this software is the fact that it is free for all, open-source, and has the ability to extend network models to include a spatial component.

Plans to further develop the software will be along two major lines. The first is to increase the number of data likelihood models that can be implemented in the software, which will allow users to perform a wider array of analysis in this single software environment and allow for a wider range of responses to be modeled. Secondly, as the user may have multiple

social network structures available in their data, such as a friendship network and a separate romantic relationship network, we aim to extend the software by developing models that can account for multiple network structures within the same model.

Chapter 9

General discussion and conclusion

The main focus of this thesis was to study the relationship between an individual's social network and their uptake of specific, correlated, multiple health behaviours, which have become an increasing public health concern. Most of the literature on the association of social networks and health behaviours takes the following approach. Descriptive network statistics are obtained from the raw network to estimate the observed effect on a health outcome of an arbitrary concept, such as an individual's popularity in the network. In contrast, in this thesis, methods and models have been developed and applied to estimate the observed effect on health behaviours that each individual has on others in their network, whilst also assessing variations in health behaviours by place of residence, and taking into account individual-level factors. Data with these type of structures are increasingly becoming available for researchers to study, having taken the appropriate confidentiality measures, and pose various statistical challenges when being modeled. Chapter 2 provides a review of statistical methods that form the basis of the novel statistical models that are formulated in Chapters 4 and 6. Chapter 3 provides a review that outlines some of the many challenges that researchers face when seeking to model the role that exists between social networks and health behaviours in both formulating a statistical model and implementing them in software. This review compares and contrasts various approaches and highlights their shortcomings in modelling

the role that exists between social networks and health behaviours. Chapters 4, 6 and 8 provide novel contributions to the literature on social network modelling, with Chapters 4 and 6 focusing on the formulation of novel statistical models and Chapter 8 focusing on their implementation in an `R` software package, `netcmc`. The motivation behind the construction of these three novel chapters is outlined below.

Chapter 4 models the role that exists between social networks and health behaviours with various multivariate extensions of the multiple membership multiple classification model. The multivariate multiple membership multiple classification model extends the univariate model, which is a form of generalised linear mixed model, by allowing both the sets of random effects for the single membership and multiple membership classifications to be jointly modeled across responses. In real world examples, in which a researcher seeks to model the effect that a social network has on multiple health behaviours, it is reasonable to imagine that the effects of the social network across the multiple health behaviours are often correlated. Thus, as a result, it may be wholly appropriate to adopt a multivariate modelling approach. In furtherance of studying the multivariate extension of the model, a comparison of priors for covariance structures in the multivariate multiple membership multiple classification model is conducted in Chapter 5 through a simulation study to compare and contrast the effects of using an Inverse-Wishart, scaled Inverse-Wishart, or hierarchical half-t prior for covariance matrices. Chapter 6 proposes a multivariate spatial multiple membership multiple classification model, which examines the relationship that a spatial structure may play in modelling multiple health behaviours. This model extends the model presented in Chapter 4 by modelling the spatial configuration of areal units through multivariate random effects that are assigned a conditional autoregressive prior distribution. There are a number of avenues of future work that naturally result from this chapter, such as extending the spatial model. Such extensions could be to examine the effects of changing the specification of the neighbourhood matrix to see what impact this has on the results, as well as allowing for between outcome correlations via a multivariate CAR type model. Also, a spatio-network interaction

involving the sets of spatial and friendship network random effects could be explored, as it may be of interest to study whether friendship effects differ depending on the Zip Code in which an individual lives. Chapter 8 seeks to make modelling social network and spatial structures in data more accessible by creating an R package titled `netcmc`, which will allow a user to fit a wide range of social network models in a Bayesian paradigm. This chapter and the creation of `netcmc` provides three novel contributions to the software landscape of social network modelling. Firstly, `netcmc` is an R package which is free to use for all. In contrast to other software that requires the purchase of a license for those who are not UK-academics, `netcmc` seeks to increase the adoption of social network modelling for those who are not UK-academics and so is made free to use for all. Secondly, in contrast to much of the software in this space, `netcmc` is an open-source package that allows users to view the source code and even contribute to the project. This point should also contribute to the reproducibility of research in the field of social network modelling. Thirdly, `netcmc` is an easy to use, single-call software package. In contrast to competing software packages, `netcmc` allows the network adjacency matrix to be specified as a binary neighbourhood matrix and given the required argument inputs, the models can be implemented by a single call in `R`.

The novel contributions in this thesis are centered around three main areas, which seek to enrich the literature on social network modelling. The first of these areas focuses on how to model multiple health behaviours, which are often present in health data, and how to go about doing so with the use of a social network structure also present in the data. The second area focuses on how to model the presence of spatial effects on an individual's propensity to engage in a particular health behaviours, and more specifically multiple health behaviours that may be correlated. The third area focuses on the dissemination of the novel ideas proposed in this thesis. Each of these three areas is discussed below in greater detail.

9.1 Multivariate random effects

The multiple health behaviour survey data used through out this thesis contains covariate information on the individual as well as a social network structure, which had been induced by the survey through asking each individual to nominate up to 19 individuals in the 10th grade that they considered to be their friend. Many of the covariates in the data are proxy measures for the financial background from which the adolescent comes from, such as their *father's education*, *mother's education*, *rooms in house*, free *lunch eligibility*, *general health* and family's *home ownership*. Other covariates seek to understand how the adolescent chooses to use their spare time, through covariates such as Facebook use, online gaming habits and last year's *exam grades*. The remaining covariates focus on the adolescent's gender and age. Many of the studies in the literature have selected a subset of covariates from a similar pool of covariates and included descriptive network statistics as covariates (Valente et al., 2005) or univariate random effects (Lorant and Tranmer, 2019) to account for the network structure in the data. In Chapter 4 multiple health behaviour data are studied in a an approach which allows for multivariate random effects to be used in the modelling of multiple health behaviours. This approach offers several advantages over the two approaches previously mentioned. Firstly, modelling random effects in a multivariate fashion allows for the researcher to compute the correlation between sets of random effect relating to pairs of health behaviours, allowing the research to make statements about how the social network random effects are or are not correlated for a given pair of health behaviours. Secondly, in contrast to the approach of using descriptive network statistics, the multivariate random effects approach allows the researcher to identify the effects that each individual in the network has on their peers, which is a very powerful tool that allows for positive and negative influencers to be identified in a social network.

The application of the multivariate multiple membership multiple classification model described in Chapter 4 to the survey network data revealed how the social network structure

in data can impact the health behaviour of an individual. It provided evidence of the way in which social network random effects may be correlated across different health behaviour responses. Thus, possessing positive/negative alter influences on one's behaviour may carry over to other health behaviours. For example, the results in Chapter 4 showed that there was a strong positive linear relationship observed between the alter random effects relating to the cigarette and marijuana responses. Thus an individual having alters that positively influence whether they have smoked at least one cigarette in the past 30 days, may carry over to impacting their marijuana consumption. Chapter 4 also compares the results of fitting a multivariate multiple membership multiple classification model to that of its univariate counter-part. The results showed that the estimated DIC of the multivariate model was lower than that of the multivariate model that induced independence across the three responses, suggesting that the multivariate extension of the model would best predict a replicate data set which has the same structure as that being observed. This result further highlights the advantages that a multivariate extension of the univariate multiple membership multiple classification model may have in being applied to health behaviour data that has multiple responses and a social network structure.

Focusing on covariate effects, the results of Chapter 4 show that, across the univariate and multivariate models, the covariate with the largest impact on whether an individual has consumed at least one drink of alcohol in the past 30 days, whether an individual has smoked at least one cigarette in the past 30 days, and whether an individual has ever tried marijuana was the exam grades an individual achieved last year. In particular, holding all other covariates fixed, in comparison to those that achieved mostly A's in exams last year, those that achieved mostly C's or lower had much higher odds of a positive value for all responses.

9.2 Spatio-network models

Chapter 6 investigates whether a multivariate relationship exists between the non-overlapping areal units (Zip Codes) that each individual belongs to and one of the multiple health behaviours that they self-report. The spatial closeness between each pair of Zip Codes is encoded in the model described in the next section by a binary neighbourhood matrix denoted $\mathbf{A}_{33\times33}$, where the ijth element $a_{ij} = 1$ if Zip Codes (i, j) share a common border and $a_{ij} = 0$ otherwise (and $a_{ii} = 0$ for all i). The *border sharing* specification is the neighbourhood matrix used in the spatial modelling because of its sparsity and simplicity of construction. However, Zip Codes that are isolated share no neighbours under this definition, which means the conditional autoregressive prior outlined in the model for capturing the spatial correlation has improper full conditional distributions for these Zip Codes. Thus a commonly used adjustment to $\mathbf{A}$ for each isolated Zip Code i was used to rectify this problem, which is to make them a neighbour of the Zip Code j that is geographically closest (e.g. set $a_{ij} = a_{ji} = 1$). In order to provide a comprehensive examination of the effects that covariates, spatial and social network structures have on the set of multiple health behaviours, each permutation of possible model was fit to the data. This led to 8 models being fit to the data, ranging from the null model to the full, saturated model, which contained covariates, a multivariate random effects structure for spatial units and a multivariate random effects structure for the social network structure.

The results showed that peer effects play a large role in determining whether adolescents partake in a given health behaviour, as their addition to the null model leads to the greatest reduction in DIC when compared to just adding either the covariates or spatial component. Also, once network random effects are included in the model, there is only a small improvement in model fit when incorporating the covariates and or a spatial component. For the covariates, the only consistently significant effect on the participation in adverse health behaviours was school exam performance, with students having poorer exam results observed

to be more likely to partake in these behaviours. In contrast, the spatial effect was essentially non-existent.

The results also showed that the effect that a friend has on an adolescent is strongly correlated across the three binary responses, with estimated pairwise correlations ranging between 0.956 and 0.975. These correlations support the notion of co-occurrence of risky behaviours in adolescents. Among the three pairs of jointly modeled friendship random effects, the results show that the pair relating to the cigarette and marijuana responses are the most correlated, although the correlations for the remaining pairs are only slightly smaller.

9.3 Dissemination - `netcmc`

Chapter 8 focuses on the dissemination of the novel concepts presented in this thesis through the creation of the `netcmc` package in `R`. This statistical software package is a free to use, open-source, single call software package. `netcmc` extends the novel work presented in Chapters 4 and 6 in two major ways, which are described below.

Firstly, whereas Chapters 4 and 5 use a Bernoulli likelihood to model the multiple binary health behaviours in the data that they are applied to, `netcmc` allows the user to fit models to Binomial, Gaussian and Poisson data. This is done to allow a user to fit the models described in Chapters 4 and 5 to continuous, count and binomial data, which is commonly found in the work of researchers.

Secondly, `netcmc` allows the user to fit a class of 4 univariate models and 3 multivariate models that can be used when only covariates or covariates and network structure or covariates, network structure and spatial structure is present in the data. The call `uni()` implements a covariate only univariate model that is similar to that proposed by Nelder and Wedderburn (1972), which is a standard univariate generalized linear model. The call

`uniNet()` implements a covariate and network univariate model that is similar to that proposed by Browne et al. (2001), and represents the social network pattern in the mean response with a single set of weighted random effects. The call `uniNetRand()` implements a covariate, network and random effect univariate model that is similar to that proposed by Browne et al. (2001), and has the same network random effects structure as the `uniNet()` model. Additionally, this model allows for a set of independent random effects to be incorporated into the modeled. The call `uniNetLeroux()` implements a covariate, network and spatial univariate model that is similar to that represents the spatio-network pattern in the mean response with a set of spatial random effects and weighted random effects. The spatial effect is modeled by the conditional autoregressive prior proposed by Leroux et al. (2000). The call `multiNet()` implements a covariate and network multivariate model that is an extension to that proposed by Browne et al. (2001), and is a multivariate extension of the `uniNet()` model. In this extension, the prior structure of the network random effects is jointly modeled across the multiple responses. The call `multiNetRand()` implements a covariate, network and random effect multivariate model that is similar to that is an extension to that proposed by Browne et al. (2001), and is a multivariate extension of the `uniNetRand()` model. In this extension, the prior structure of these random effects is jointly modeled across the multiple responses. The call `multiNetLeroux()` implements a covariate, network and spatial multivariate model that is a multivariate extension of the `uniNetLeroux()` model. In this extension, the prior structure of the network random effects is jointly modeled across the multiple responses. However, the spatial random effects are modeled independently.

9.4 Summary

In summary, the novel models developed and the results derived from the models should expand the literature on social network modelling and allow for alternative approaches to modelling the effects that a social network is observed to have on an individual's propensity

to partake in a given health behaviour. In addition to this, in the multivariate modelling approach, it should also provide researchers with a greater understanding of the underlying process between the effects that individuals in a social network have on each other across multiple health behaviours. A major limitation of this thesis and the literature on network modelling is the availability of data to study. Network data often contains very sensitive information on individual people and so is often subject to strict regulation when it comes to data sharing. This can make it difficult for researchers to obtain and subsequently develop novel approaches for, which limits the way in which the literature on network modelling develops. There are a number of avenues of future work that naturally result from this thesis. The first of which is the development of a multivariate model that considers a spatio-network interaction involving the sets of spatial and network random effects, as it may be of interest to study whether friendship effects differ depending on the areal unit in which an individual lives. The second is to extend the methods to allow for longitudinal data to modeled with these techniques, as it may be of interest to study whether friendship effects differ through time. The third concerns the further development of the `netcmc` software package. Plans to further develop the software will be along two major lines. The first is to increase the number of data likelihood models that can be implemented in the software, which will allow users to perform a wider array of analysis in this single software environment and allow for a wider range of responses to be modeled. Secondly, as the user may have multiple social network structures available in their data, such as a friendship network and a separate romantic relationship network, we aim to extend the software by developing models that can account for multiple network structures within the same model.

Bibliography

Akerlind, I., Hörnquist, J.O., 1992. Loneliness and alcohol abuse: A review of evidences of an interplay. Social Science & Medicine (1982) 34, 405–414. doi:10.1016/0277-9536(92)90300-f.

Akers, R.L., Krohn, M.D., Lanza-Kaduce, L., Radosevich, M., 1979. Social Learning and Deviant Behavior: A Specific Test of a General Theory. American Sociological Review 44, 636–655. doi:10.2307/2094592.

Alexander, C., Piazza, M., Mekos, D., Valente, T., 2001. Peers, Schools, and Adolescent Cigarette Smoking. The Journal of adolescent health : official publication of the Society for Adolescent Medicine 29, 22–30. doi:10.1016/S1054-139X(01)00210-5.

Alvarez, I., Niemi, J., Simpson, M., 2016. Bayesian inference for a covariance matrix. arXiv arXiv:1408.4050.

Arnold, B.C., Villasenor, J.A., 1997. Variance estimation with diffuse prior information. Statistics & Probability Letters 33, 35–39. doi:10.1016/S0167-7152(96)00107-1.

Aunsmo, A., Øvretveit, S., Breck, O., Valle, P.S., Larssen, R.B., Sandberg, M., 2009. Modelling sources of variation and risk factors for spinal deformity in farmed Atlantic salmon using hierarchical- and cross-classified multilevel models. Preventive Veterinary Medicine 90, 137–145. doi:10.1016/j.prevetmed.2009.03.015.

Azagba, S., Sharaf, M.F., 2014. Binge drinking and marijuana use among menthol and

non-menthol adolescent smokers: Findings from the Youth Smoking Survey. Addictive Behaviors 39, 740–743. doi:10.1016/j.addbeh.2013.12.005.

Bagnardi, V., Blangiardo, M., Vecchia, C.L., Corrao, G., 2001. A meta-analysis of alcohol drinking and cancer risk. British Journal of Cancer 85, 1700–1705. doi:10.1054/bjoc.2001.2140.

Balsa, A.I., Giuliano, L.M., French, M.T., 2011. The effects of alcohol use on academic achievement in high school. Economics of education review 30, 1–15. doi:10.1016/j.econedurev.2010.06.015.

Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall, Boca Raton.

Barker, K.M., Subramanian, S.V., Berkman, L., Austin, S.B., Evans, C.R., 2019. Adolescent Sexual Initiation: A Cross-Classified Multilevel Analysis of Peer Group-, School-, and Neighborhood-Level Influences. Journal of Adolescent Health 65, 390–396. doi:10.1016/j.jadohealth.2019.03.002.

Beard, E., Brown, J., West, R., Angus, C., Kaner, E., Michie, S., 2017. Healthier central England or North–South divide? Analysis of national survey data on smoking and high-risk drinking. BMJ Open 7, e014210. doi:10.1136/bmjopen-2016-014210.

Berndt, T.J., 1979. Developmental changes in conformity to peers and parents. Developmental Psychology 15, 608–616.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistics and Mathematics 43, 1–59.

Bivand, R., Pebesma, E., Gomez-Rubio, V., 2013. Applied Spatial Data Analysis with R. 2 ed., Springer, New York.

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., Fowler, J.H., 2012. A 61-million-person experiment in social influence and political mobilization 489, 10.1038/nature11421. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3834737/, doi:10.1038/nature11421.

Brook, J.S., Brook, D.W., Arencibia-mireles, O., Richter, L., Whiteman, M., 2001. Risk Factors for Adolescent Marijuana Use Across Cultures and Across Time. The Journal of Genetic Psychology 162, 357–374. doi:10.1080/00221320109597489.

Browne, W.J., Draper, D., 2000. Implementation and performance issues in the bayesian and likelihood fitting of multilevel models 15, 391–420. URL: https://doi.org/10.1007/s001800000041, doi:10.1007/s001800000041.

Browne, W.J., Draper, D., 2006. A comparison of bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis 1, 473–514. doi:10.1214/06-BA117.

Browne, W.J., Goldstein, H., Rasbash, J., 2001. Multiple membership multiple classification (MMMC) models. Statistical Modelling 1, 103–124. doi:10.1177/1471082X0100100202.

Browne, W.J., Steele, F., Golalizadeh, M., Green, M.J., 2009. The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. Journal of the Royal Statistical Society. Series A, (Statistics in Society) 172, 579–598. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2718325/, doi:10.1111/j.1467-985X.2009.00586.x.

Bürkner, P.C., 2017. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software 80, 1–28. doi:10.18637/jss.v080.i01.

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., Martin, O.A., 2022. Bambi: A simple interface for fitting bayesian linear models in python. Journal of Statistical Software 103, 1–29. URL: https://www.jstatsoft.org/index.php/jss/article/view/v103i15, doi:10.18637/jss.v103.i15.

Castillo, J.M., Jivraj, S., Ng Fat, L., 2017. The regional geography of alcohol consumption in England: Comparing drinking frequency and binge drinking. Health & Place 43, 33–40. doi:10.1016/j.healthplace.2016.11.007.

Catalano, R.F., Kosterman, R., Hawkins, J.D., Newcomb, M.D., Abbott, R.D., 1996. Modeling the Etiology of Adolescent Substance Use: A Test of the Social Development Model. Journal of drug issues 26, 429–455.

Chandrasekaran, S., Annamalai, B., De, S.K., 2019. Evaluating marketer generated content popularity on brand fan pages – A multilevel modelling approach. Telematics and Informatics 44, 101266. doi:10.1016/j.tele.2019.101266.

Charlton, C., Rasbach, J., Browne, W., Cameron, B., 2019. MLwiN version 3.04. Centre for Multilevel Modelling, University of Bristol .

Chassin, L., Presson, C.C., Sherman, S.J., Corty, E., Olshavsky, R.W., 1984. Predicting the Onset of Cigarette Smoking in Adolescents: A Longitudinal Study1. Journal of Applied Social Psychology 14, 224–243. doi:10.1111/j.1559-1816.1984.tb02233.x.

Chihuri, S., Li, G., Chen, Q., 2017. Interaction of marijuana and alcohol on fatal motor vehicle crash risk: A case–control study. Injury Epidemiology 4. doi:10.1186/s40621-017-0105-z.

Choo, E.K., Benz, M., Zaller, N., Warren, O., Rising, K.L., McConnell, K.J., 2014. The Impact of State Medical Marijuana Legislation on Adolescent Marijuana Use. Journal of Adolescent Health 55, 160–166. doi:10.1016/j.jadohealth.2014.02.018.

Christakis, N.A., Fowler, J.H., 2007. The spread of obesity in a large social network over 32 years 357, 370–379. URL: https://www.nejm.org/doi/10.1056/NEJMsa066082, doi:10.1056/NEJMsa066082. publisher: Massachusetts Medical Society.

Clapp, J.D., Shillington, A.M., 2001. ENVIRONMENTAL PREDICTORS OF HEAVY EPISODIC DRINKING. The American Journal of Drug and Alcohol Abuse 27, 301–313. doi:10.1081/ADA-100103711.

Cramb, S., Duncan, E., Baade, P., Mengersen, K., 2017. Investigation of bayesian spatial models. .

Dean, D.O., Bauer, D.J., Prinstein, M.J., 2017. Friendship Dissolution Within Social Networks Modeled Through Multilevel Event History Analysis. Multivariate Behavioral Research 52, 271–289. doi:10.1080/00273171.2016.1267605.

DiGrande, L., Perrier, M.P., Lauro, M.G., Contu, P., 2000. Alcohol Use and Correlates of Binge Drinking among University Students on the Island of Sardinia, Italy. Substance Use & Misuse 35, 1471–1483. doi:10.3109/10826080009148226.

Dishion, T., Loeber, R., 1985. Adolescent Marijuana and Alcohol Use: The Role of Parents and Peers Revisited. The American journal of drug and alcohol abuse 11, 11–25. doi:10.3109/00952998509016846.

Dodds, P.S., Muhamad, R., Watts, D.J., 2003. An Experimental Study of Search in Global Social Networks. Science 301, 827–829. doi:10.1126/science.1081058.

Doll, R., Hill, A.B., 1950. Smoking and Carcinoma of the Lung. Br Med J 2, 739–748. doi:10.1136/bmj.2.4682.739.

Doll, R., Hill, A.B., 1954. The Mortality of Doctors in Relation to Their Smoking Habits. Br Med J 1, 1451–1455. doi:10.1136/bmj.1.4877.1451.

Donaldson, L., 2009. Guidance on the consumption of alcohol by children and young people. A report by the Chief Medical Officer. URL: https://webarchive.nationalarchives.gov.uk/20130103185806/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_110258.

Durrant, G.B., Vassallo, R., Smith, P.W.F., 2018. Assessment of Multiple Membership Multilevel Models: An Application to Interviewer Effects on Nonresponse. Multivariate Behavioral Research 53, 595–611. doi:10.1080/00273171.2018.1465809.

Eddelbuettel, D., François, R., 2011. Rcpp: Seamless R and C++ integration. Journal of Statistical Software 40, 1–18. URL: http://www.jstatsoft.org/v40/i08/, doi:10.18637/jss.v040.i08.

Eddelbuettel, D., Sanderson, C., 2014. Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. Computational Statistics and Data Analysis 71, 1054–1063. URL: http://dx.doi.org/10.1016/j.csda.2013.02.005.

Eisenberg, M.E., Forster, J.L., 2003. Adolescent smoking behavior: Measures of social norms. American Journal of Preventive Medicine 25, 122–128. doi:10.1016/S0749-3797(03)00116-8.

Elder, J.P., Campbell, N.R., Litrownik, A.J., Ayala, G.X., Slymen, D.J., Parra-Medina, D., Lovato, C.Y., 2000. Predictors of Cigarette and Alcohol Susceptibility and Use among Hispanic Migrant Adolescents. Preventive Medicine 31, 115–123. doi:10.1006/pmed.2000.0693.

Fergusson, D.M., Horwood, L.J., Swain-Campbell, N.R., 2003. Cannabis dependence and psychotic symptoms in young people. Psychological Medicine 33, 15–21. doi:10.1017/S0033291702006402.

Fielding, A., Goldstein, H., 2006. Cross-classified and multiple membership structures in multilevel models: An introduction and review. URL: https://dera.ioe.ac.uk/6469/.

Ford, D.E., Vu, H.T., Anthony, J.C., 2002. Marijuana use and cessation of tobacco smoking in adults from a community sample. Drug and Alcohol Dependence 67, 243–248. doi:10.1016/S0376-8716(02)00066-2.

Fujimoto, K., Valente, T.W., 2015. Multiplex congruity: Friendship networks and perceived popularity as correlates of adolescent alcohol use. Social Science & Medicine 125, 173–181. doi:10.1016/j.socscimed.2014.05.023.

Fujimoto, K., Wang, P., Valente, T.W., 2013. The decomposed affiliation exposure model: A network approach to segregating peer influences from crowds and organized sports. Network Science (Cambridge University Press) 1, 154–169. doi:10.1017/nws.2013.7.

Gelfand, A., Vounatsou, P., 2003. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics 4, 11–15. doi:10.1093/biostatistics/4.1.11.

Gelfand, A.E., Sahu, S.K., Carlin, B.P., 1995. Efficient parametrisations for normal linear mixed models. Biometrika 82, 479–488. URL: http://www.jstor.org/stable/2337527.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. 3rd ed. CRC Press LLC.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis. 2nd ed. ed., Chapman and Hall/CRC.

Gelman, A., Gilks, W.R., Roberts, G.O., 1997. Weak convergence and optimal scaling of random walk metropolis algorithms 7, 110–120. doi:10.1214/aoap/1034625254.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6, 721–741. doi:10.1109/TPAMI.1984.4767596.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T., 2020. mvtnorm: Multivariate Normal and t Distributions. URL: https://CRAN.R-project.org/package=mvtnorm. r package version 1.1-1.

Gerogiannis, G., 2022. netcmc: Spatio-Network Generalised Linear Mixed Models for Areal Unit and Network Data. URL: https://CRAN.R-project.org/package=netcmc. r package version 1.0.

Geweke, J., 1991. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Federal Reserve Bank of Minneapolis.

Gilks, W.R., Best, N.G., Tan, K.K.C., 1995. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. Journal of the Royal Statistical Society. Series C (Applied Statistics) 44, 455–472. doi:10.2307/2986138.

Gold, D.R., Wang, X., Wypij, D., Speizer, F.E., Ware, J.H., Dockery, D.W., 1996. Effects of Cigarette Smoking on Lung Function in Adolescent Boys and Girls. New England Journal of Medicine 335, 931–937. doi:10.1056/NEJM199609263351304.

Goldstein, H., Browne, W., 2016. Multilevel models. doi:10.1002/9781118445112.stat05764.pub2.

Hadfield, J., 2019. Mcmcglmm course notes. URL: https://mran.microsoft.com/snapshot/2020-12-04/web/packages/MCMCglmm/vignettes/CourseNotes.pdf.

Hadfield, J.D., 2010. Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. Journal of Statistical Software 33, 1–22. URL: http://www.jstatsoft.org/v33/i02/.

Hale, D.R., Viner, R.M., 2016. The correlates and course of multiple health risk behaviour in adolescence. BMC Public Health 16, 458. doi:10.1186/s12889-016-3120-z.

Hammerton, G., Mahedy, L., Murray, J., Maughan, B., Edwards, A.C., Kendler, K.S., Hickman, M., Heron, J., 2017. Effects of Excessive Alcohol Use on Antisocial Behavior Across Adolescence and Early Adulthood. Journal of the American Academy of Child & Adolescent Psychiatry 56, 857–865. doi:10.1016/j.jaac.2017.07.781.

Hashibe, M., Brennan, P., Chuang, S.c., Boccia, S., Castellsague, X., Chen, C., Curado, M.P., Maso, L.D., Daudt, A.W., Fabianova, E., Fernandez, L., Wünsch-Filho, V., Franceschi, S., Hayes, R.B., Herrero, R., Kelsey, K., Koifman, S., La Vecchia, C., Lazarus, P., Levi, F., Lence, J.J., Mates, D., Matos, E., Menezes, A., McClean, M.D., Muscat, J., Eluf-Neto, J., Olshan, A.F., Purdue, M., Rudnai, P., Schwartz, S.M., Smith, E., Sturgis, E.M., Szeszenia-Dabrowska, N., Talamini, R., Wei, Q., Winn, D.M., Shangina, O., Pilarska, A., Zhang, Z.F., Ferro, G., Berthiller, J., Boffetta, P., 2009. Interaction between tobacco and alcohol use and the risk of head and neck cancer: Pooled analysis in the INHANCE consortium. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 18, 541–550. doi:10.1158/1055-9965.EPI-08-0347.

Hastings, W.K., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57, 97–109. URL: http://www.jstor.org/stable/2334940.

Hawkins, J.D., Weis, J.G., 1985. The social development model: An integrated approach to delinquency prevention. Journal of Primary Prevention 6, 73–97. doi:10.1007/BF01325432.

Hines, D., Fretz, A.C., Nollen, N.L., 1998. Regular and Occasional Smoking by College Students: Personality Attributions of Smokers and Nonsmokers. Psychological Reports 83, 1299–1306. doi:10.2466/pr0.1998.83.3f.1299.

Houtzager, B., Baerveldt, C., 1999. Just like normal: A social network study of the relation between petty crime and the intimacy of adolescent friendships. Social Behavior and Personality 27, 177–192.

Hryhorczuk, N., Zvinchuk, A., Shkiriak-Nyzhnyk, Z., Gonzales, N., Hryhorczuk, D., 2019. Leisure activity and alcohol use among Ukrainian adolescents. Addictive Behaviors Reports 10, 100201. doi:10.1016/j.abrep.2019.100201.

Huang, G.C., Unger, J.B., Soto, D., Fujimoto, K., Pentz, M.A., Jordan-Marsh, M., Valente, T.W., 2014. Peer Influences: The Impact of Online and Offline Friendship Networks on Adolescent Smoking and Alcohol Use. Journal of Adolescent Health 54, 508–514. doi:10.1016/j.jadohealth.2013.07.001.

Islam, M.T., El-Basyouny, K., 2015. Multilevel models to analyze before and after speed data. Analytic Methods in Accident Research 8, 33–44. doi:10.1016/j.amar.2015.10.001.

Jack, E., Lee, D., Dean, N., 2019. Estimating the changing nature of Scotland's health inequalities by using a multivariate spatiotemporal model. Journal of the Royal Statistical Society Series A (Statistics in Society) 182, 1061–1080. doi:10.1111/rssa.12447.

Jackson, K.M., Sher, K.J., Cooper, M.L., Wood, P.K., 2002. Adolescent alcohol and tobacco use: Onset, persistence and trajectories of use across two samples. Addiction (Abingdon, England) 97, 517–531.

Jacobs, W., Barry, A.E., Xu, L., Valente, T.W., 2016a. Hispanic/Latino Adolescents' Alcohol Use: Influence of Family Structure, Perceived Peer Norms, and Family Members' Alcohol Use. American journal of health education 47, 253–261. doi:10.1080/19325037.2016.1179141.

Jacobs, W., Goodson, P., Barry, A.E., McLeroy, K.R., McKyer, E.L.J., Valente, T.W., 2017. Adolescent Social Networks and Alcohol Use: Variability by Gender and Type. Substance use & misuse 52, 477–487. doi:10.1080/10826084.2016.1245333.

Jacobs, W., Jeon, K.C., Goodson, P., Valente, T.W., 2016b. What's love got to do with it? Adolescent romantic networks and substance use. International Journal of Adolescence and Youth 21, 513–522. doi:10.1080/02673843.2015.1122643.

Jin, X., Banerjee, S., Carlin, B., 2007. Order-free co-regionalized areal data models with

application to multiple-disease mapping. Journal of the Royal Statistical Society Series B (Statistical Methodology) 69, 817–838.

Kandel, D., 1975. Stages in adolescent involvement in drug use. Science 190, 912–914. doi:10.1126/science.1188374.

Kandel, D., 2002. Stages and Pathways of Drug Involvement: Examining the Gateway Hypothesis. Cambridge University Press. doi:10.1017/CBO9780511499777.

Kaplan, C.P., Nápoles-Springer, A., Stewart, S.L., Pérez-Stable, E.J., 2001. Smoking acquisition among adolescents and young Latinas: The role of socioenvironmental and personal factors. Addictive Behaviors 26, 531–550. doi:10.1016/S0306-4603(00)00143-X.

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R.M., 1998. Markov chain monte carlo in practice: A roundtable discussion. The American Statistician 52, 93–100. URL: http://www.jstor.org/stable/2685466.

Kim, I.J., Zane, N.W.S., Hong, S., 2002. Protective Factors Against Substance Use Among Asian American Youth: A Test of the Peer Cluster Theory. Journal of Community Psychology .

Komro, K.A., Williams, C.L., Forster, J.L., Perry, C.L., Farbakhsh, K., Stigler, M.H., 2000. The Relationship Between Adolescent Alcohol Use and Delinquent and Violent Behaviors. Journal of Child & Adolescent Substance Abuse 9, 13–28. doi:10.1300/J029v09n02_02.

Koopmans, J.R., van Doornen, L.J.P., Boomsma, D.I., 1997. Association between Alcohol Use and Smoking in Adolescent and Young Adult Twins: A Bivariate Genetic Analysis. Alcoholism: Clinical and Experimental Research 21, 537–546. doi:10.1111/j.1530-0277.1997.tb03800.x.

Krosnick, J., Judd, C., 1982. Transitions in social influence at adolescence: Who induces cigarette smoking? 18, 359–368. doi:10.1037/0012-1649.18.3.359.

Kuntsche, E., Jordan, M.D., 2006. Adolescent alcohol and cannabis use in relation to peer and school factors: Results of multilevel analyses. Drug and Alcohol Dependence 84, 167–174. doi:10.1016/j.drugalcdep.2006.01.014.

Lauder, W., Mummery, K., Jones, M., Caperchione, C., 2006. A comparison of health behaviours in lonely and non-lonely populations. Psychology, Health & Medicine 11, 233–245. doi:10.1080/13548500500266607.

Lee, D., 2011. A comparison of conditional autoregressive models used in Bayesian disease mapping. Spatial and Spatio-temporal Epidemiology 2, 79–89.

Lee, D., Mitchell, R., 2012. Boundary detection in disease mapping studies. Biostatistics 13, 415–426. doi:10.1093/biostatistics/kxr036.

Lee, D., Mitchell, R., 2013. Locally adaptive spatial smoothing using conditional auto-regressive models. Journal of the Royal Statistical Society Series C: Applied Statistics 62, 593–608. doi:10.1111/rssc.12009.

Lee, D., Sarran, C., 2015. Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. Environmetrics 26, 477–487. doi:https://doi.org/10.1002/env.2348.

Lees, B., Meredith, L.R., Kirkland, A.E., Bryant, B.E., Squeglia, L.M., 2020. Effect of alcohol use on the adolescent brain and behavior. Pharmacology Biochemistry and Behavior 192, 172906. doi:10.1016/j.pbb.2020.172906.

Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence, in: Halloran, M.E., Berry, D. (Eds.), Statistical Models in Epidemiology, the Environment, and Clinical Trials, Springer, New York, NY. pp. 179–191. doi:10.1007/978-1-4612-1284-3_4.

Lesaffre, E., Molenberghs, G., 1991. Multivariate probit analysis: A neglected procedure in medical statistics. Statistics in Medicine 10, 1391–1403. doi:10.1002/sim.4780100907.

Ling, M.Y.J., Rodzlan Hasani, W.S., Mohd Yusoff, M.F., Abd Hamid, H.A., Lim, K.H., Tee, G.H., Baharom, N., Ab Majid, N.L., Robert Lourdes, T.G., Mat Rifin, H., Saminathan, T.A., 2019. Cigarette Smoking Among Secondary School-Going Male Adolescents in Malaysia: Findings From the National Health and Morbidity Survey: Adolescent Health Survey 2017. Asia Pacific Journal of Public Health , 1010539519874948doi:10.1177/1010539519874948.

Liu, J.S., Wu, Y.N., 1999. Parameter Expansion for Data Augmentation. Journal of the American Statistical Association 94, 1264–1274. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10473879, doi:10.1080/01621459.1999.10473879.

Lorant, V., Tranmer, M., 2019. Peer, school, and country variations in adolescents' health behaviour: A multilevel analysis of binary response variables in six European cities. Social Networks 59, 31–40. doi:10.1016/j.socnet.2019.05.004.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. Winbugs a bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 10, 325–337. URL: http://portal.acm.org/citation.cfm?id=599385, doi:http://dx.doi.org/10.1023/A:1008929526011.

MacNab, Y., 2016. Linear models of coregionalization for multivariate lattice data: a general framework for coregionalized multivariate car models. Statistics in Medicine 35, 3827–3850.

Martinez-Beneito, M., 2013. A general modelling framework for multivariate disease mapping. Biometrika 100, 539–553.

McElreath, R., 2016. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. Chapman and Hall/CRC. doi:10.1201/9781315372495.

McKee, S., Hinson, R., Rounsaville, D., Petrelli, P., 2004. Survey of subjective effects of smoking while drinking among college students. Nicotine & Tobacco Research 6, 111–117. doi:10.1080/14622200310001656939.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics 21, 1087–1092. doi:10.1063/1.1699114.

Metropolis, N., Ulam, S., 1949. The monte carlo method. Journal of the American Statistical Association 44, 335–341. doi:10.1080/01621459.1949.10483310. pMID: 18139350.

Milgram, S., 1967. The Small World Problem. Psychology Today 1, 61–67.

Mostafa, M.M., 2016. Post-materialism, Religiosity, Political Orientation, Locus of Control and Concern for Global Warming: A Multilevel Analysis Across 40 Nations. Social Indicators Research 128, 1273–1298. doi:10.1007/s11205-015-1079-2.

Müller, F.H., 1939. Tabakmißbrauch und Lungencarcinom. Zeitschrift für Krebsforschung 49, 57–85. doi:10.1007/BF01633114.

Mundt, M.P., 2011. The Impact of Peer Social Networks on Adolescent Alcohol Use Initiation. Academic Pediatrics 11, 414–421. doi:10.1016/j.acap.2011.05.005.

Munoz-Zanzi, C., Campbell, C., Berg, S., 2016. Seroepidemiology of toxoplasmosis in rural and urban communities from Los Rios Region, Chile. Infection Ecology & Epidemiology 6. doi:10.3402/iee.v6.30597.

Murnaghan, D.A., Leatherdale, S.T., Sihvonen, M., Kekki, P., 2008. A multilevel analysis examining the association between school-based smoking policies, prevention programs and youth smoking behavior: Evaluating a provincial tobacco control strategy. Health Education Research 23, 1016–1028. doi:10.1093/her/cyn034.

Muthén, L.K., Muthén, B.O., 1998-2017. Mplus User's Guide (8th ed.). Los Angeles, CA: Muthén & Muthén. URL: http://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf.

Nam, S.J., Kim, H.M., Kang, T., Park, C.Y., 2019. A Study of Machine Learning Models in Predicting the Intention of Adolescents to Smoke Cigarettes. arXiv:1910.12748 [cs, stat] arXiv:1910.12748.

Neaigus, A., Friedman, S.R., Jose, B., Goldstein, M.F., Curtis, R., Ildefonso, G., Des Jarlais, D.C., 1996. High-Risk Personal Networks and Syringe Sharing as Risk Factors for HIV Infection Among New Drug Injectors:. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 11, 499–509. doi:10.1097/00042560-199604150-00011.

Neal, R.M., 1997. Markov Chain Monte Carlo Methods Based on 'Slicing' the Density Function. Technical Report.

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General) 135, 370–384. doi:10.2307/2344614.

NHS, 2019. Smoking, Drinking and Drug Use among Young People in England 2018. URL: https://digital.nhs.uk/data-and-information/publications/statistical/smoking-drinking-and-drug-use-among-young-people-in-england/2018.

Nixon, J.S., Chen, Y., Schiller, N.O., 2015. Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. Language, Cognition and Neuroscience 30, 491–505. doi:10.1080/23273798.2014.942326.

Novak, S.P., Peiper, N.C., Zarkin, G.A., 2016. Nonmedical prescription pain reliever and alcohol consumption among cannabis users. Drug and Alcohol Dependence 159, 101–108. doi:10.1016/j.drugalcdep.2015.11.039.

NTP, 2016. National Toxicology Program: 14th Report on Carcinogens. https://ntp.niehs.nih.gov/go/roc14.

Oetting, E.R., Beauvais, F., 1986. Peer Cluster Theory: Drugs and the Adolescent. Journal of Counseling & Development 65, 17–22. doi:10.1002/j.1556-6676.1986.tb01219.x.

Oldham, M., Holmes, J., Whitaker, V., Fairbrother, H., Curtis, P., 2018. Youth Drinking in Decline .

Oppong Asante, K., Kugbey, N., 2019. Alcohol use by school-going adolescents in Ghana: Prevalence and correlates. Mental Health & Prevention 13, 75–81. doi:10.1016/j.mhp.2019.01.009.

Pateras, K., Nikolakopoulos, S., Roes, K.C.B., 2021. Prior distributions for variance parameters in a sparse-event meta-analysis of a few small trials. Pharmaceutical Statistics 20, 39–54. doi:10.1002/pst.2053.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. Coda: Convergence diagnosis and output analysis for mcmc. R News 6, 7–11. URL: https://journal.r-project.org/archive/.

Prieto-Damm, B., de la Rosa, P.A., Lopez-del Burgo, C., Calatrava, M., Osorio, A., Albertos, A., de Irala, J., 2019. Leisure activities and alcohol consumption among adolescents from Peru and El Salvador. Drug and Alcohol Dependence 199, 27–34. doi:10.1016/j.drugalcdep.2019.03.002.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Ramful, P., Zhao, X., 2009. Participation in marijuana, cocaine and heroin consumption in Australia: A multivariate probit approach. Applied Economics 41, 481–496. doi:10.1080/00036840701522853.

Rigotti, N.A., Lee, J.E., Wechsler, H., 2000. US College Students' Use of Tobacco Products: Results of a National Survey. JAMA 284, 699–705. doi:10.1001/jama.284.6.699.

Ripley, B.D., 1987. Stochastic Simulation. John Wiley Sons, Ltd. doi:https://doi.org/10.1002/9780470316726.

Rose, C.D., 1999. Peer Cluster Theory and Adolescent Alcohol Use: An Explanation of Alcohol Use and a Comparative Analysis between Two Causal Models. Journal of Drug Education 29, 205–215. doi:10.2190/GEJL-0XM8-2C0K-V5AQ.

Rossi, P., 2022. bayesm: Bayesian Inference for Marketing/Micro-Econometrics. URL: https://CRAN.R-project.org/package=bayesm. r package version 3.1-5.

Rossi, P., Allenby, G., McCulloch, R., 2012. Bayesian Statistics and Marketing. Wiley Series in Probability and Statistics, Wiley. doi:10.1002/0470863692.

Sampasa-Kanyinga, H., Hamilton, H.A., LeBlanc, A.G., Chaput, J.P., 2018. Cannabis use among middle and high school students in Ontario: A school-based cross-sectional study. CMAJ Open 6, E50–E56. doi:10.9778/cmajo.20170159.

Saules, K.K., Pomerleau, C.S., Snedecor, S.M., Mehringer, A.M., Shadle, M.B., Kurth, C., Krahn, D.D., 2004. Relationship of onset of cigarette smoking during college to alcohol use, dieting concerns, and depressed mood: Results from the Young Women's Health Survey. Addictive Behaviors 29, 893–899. doi:10.1016/j.addbeh.2004.02.015.

Schairer, E., Schöniger, E., 1943. Lungenkrebs und Tabakverbrauch. Zeitschrift für Krebsforschung 54, 261–269. doi:10.1007/BF01628727.

Schepis, T.S., Desai, R.A., Cavallo, D.A., Smith, A.E., McFetridge, A., Liss, T.B., Potenza, M.N., Krishnan-Sarin, S., 2011. Gender differences in adolescent marijuana use and associated psychosocial characteristics. Journal of Addiction Medicine 5, 65–73. doi:10.1097/ADM.0b013e3181d8dc62.

Secretan, B., Straif, K., Baan, R., Grosse, Y., Ghissassi, F.E., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., Cogliano, V., 2009. A review of human carcinogens—Part E: Tobacco, areca nut, alcohol, coal smoke, and salted fish. The Lancet Oncology 10, 1033–1034. doi:10.1016/S1470-2045(09)70326-2.

Shelton, N., Savell, E., 2011. The geography of binge drinking: The role of alcohol-related knowledge, behaviours and attitudes. Results from the Health Survey for England 2007. Health & Place 17, 784–792. doi:10.1016/j.healthplace.2011.02.004.

Spiegelhalter, D., Best, N., Carlin, B., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B (Statistical Methodology) 64, 583–639.

Stan Development Team, 2022. Stan Modeling Language User's Guide and Reference Manual, Version 2.31. URL: http://mc-stan.org/.

Steele, C., Josephs, R., 1990. Alcohol Myopia. Its Prized and Dangerous Effects. The American psychologist 45, 921–33. doi:10.1037//0003-066X.45.8.921.

Steinberg, L., Monahan, K.C., 2007. Age differences in resistance to peer influence 43, 1531–1543. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2779518/, doi:10.1037/0012-1649.43.6.1531.

Talamini, R., Bosetti, C., La Vecchia, C., Dal Maso, L., Levi, F., Bidoli, E., Negri, E., Pasche, C., Vaccarella, S., Barzan, L., Franceschi, S., 2002. Combined effect of tobacco and alcohol on laryngeal cancer risk: A case–control study. Cancer Causes & Control 13, 957–964. doi:10.1023/A:1021944123914.

Tranmer, M., Pallotti, F., Lomi, A., 2016. The embeddedness of organizational performance: Multiple Membership Multiple Classification Models for the analysis of multilevel networks. Social Networks 44, 269–280. doi:10.1016/j.socnet.2015.06.005.

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, 439–455. doi:10.1111/rssa.12021.

Travers, J., Milgram, S., 1969. An Experimental Study of the Small World Problem. Sociometry 32, 425–443. URL: www.jstor.org/stable/2786545, doi:10.2307/2786545.

Tur, J.A., Puig, M., Pons, A., Benito, E., 2003. Alcohol consumption among school adolescents in Palma de Mallorca. Alcohol and alcoholism (Oxford, Oxfordshire) 38, 243–8.

Valente, T., 2010. Social networking survey 2010 [data file]. Available from author.

Valente, T.W., Fujimoto, K., Unger, J.B., Soto, D.W., Meeker, D., 2013. Variations in network boundary and type: A study of adolescent peer influences. Social Networks 35, 309–316. doi:10.1016/j.socnet.2013.02.008.

Valente, T.W., Unger, J.B., Johnson, C.A., 2005. Do popular students smoke? The association between popularity and smoking among middle school students. Journal of Adolescent Health 37, 323–329. doi:10.1016/j.jadohealth.2004.10.016.

Wand, M.P., Ormerod, J.T., Padoan, S.A., Frühwirth, R., 2011. Mean Field Variational Bayes for Elaborate Distributions. Bayesian Analysis 6, 847–900. doi:10.1214/11-BA631.

Wang, M.Q., Fitzhugh, E.C., Westerfield, R.C., Eddy, J.M., 1995. Family and peer influences on smoking behavior among American adolescents: An age trend. Journal of Adolescent Health 16, 200–203. doi:10.1016/1054-139X(94)00097-X.

Watanabe, S., 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory 11, 3571–3594. URL: http://jmlr.org/papers/v11/watanabe10a.html.

Wechsler, H., Dowdall, G.W., Davenport, A., Castillo, S., 1995. Correlates of college student binge drinking. American Journal of Public Health 85, 921–926.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. URL: https://ggplot2.tidyverse.org.

Wilson, M.W., Gittings, T., Kelly, T.C., O'Halloran, J., 2010. The importance of non-crop vegetation for bird diversity in Sitka spruce plantations in Ireland. Bird Study 57, 116–120. doi:10.1080/00063650903150676.

Wilson, N., Syme, S.L., Boyce, W.T., Battistich, V.A., Selvin, S., 2005. Adolescent Alcohol, Tobacco, and Marijuana Use: The Influence of Neighborhood Disorder and Hope. American Journal of Health Promotion 20, 11–19. doi:10.4278/0890-1171-20.1.11.

Wong, I.O.L., Cowling, B.J., Lo, S.V., Leung, G.M., 2009. A multilevel analysis of the effects of neighbourhood income inequality on individual self-rated health in Hong Kong. Social Science & Medicine 68, 124–132. doi:10.1016/j.socscimed.2008.09.064.

Wynder, E.L., Graham, E.A., Croninger, A.B., 1953. Experimental Production of Carcinoma with Cigarette Tar. Cancer Research 13, 855–864.

Yang, K., Zhou, X.N., Jia, T.W., Yang, G.J., Wu, X.H., Shi, X.W., Li, H.J., Steinmann, P., Utzinger, J., Bergquist, R., 2015. Eco-social determinants of Schistosoma japonicum infection supported by multi-level modelling in Eryuan county, People's Republic of China. Acta Tropica 141, 391–398. doi:10.1016/j.actatropica.2014.04.013.

Zammit, S., Allebeck, P., Andreasson, S., Lundberg, I., Lewis, G., 2002. Self reported cannabis use as a risk factor for schizophrenia in Swedish conscripts of 1969: Historical cohort study. BMJ 325, 1199. doi:10.1136/bmj.325.7374.1199.

Znaor, A., Brennan, P., Gajalakshmi, V., Mathew, A., Shanta, V., Varghese, C., Boffetta, P., 2003. Independent and combined effects of tobacco smoking, chewing and alcohol drinking on the risk of oral, pharyngeal and esophageal cancers in Indian men. International Journal of Cancer 105, 681–686. doi:10.1002/ijc.11114.