Bayesian estimation of a phenotype network structure using reversible jump Markov chain Monte Carlo

by

Lisa Rae Woods

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy

Victoria University of Wellington

2015

Abstract

In this thesis we aim to estimate the unknown phenotype network structure existing among multiple interacting quantitative traits, assuming the genetic architecture is known.

We begin by taking a frequentist approach and implement a score-based greedy hill-climbing search strategy using AICc to estimate an unknown phenotype network structure. This approach was inconsistent and overfitting was common, so we then propose a Bayesian approach that extends on the reversible jump Markov chain Monte Carlo algorithm. Our approach makes use of maximum likelihood estimates in the chain, so we have an efficient sampler using well-tuned proposal distributions. The common approach is to assume uniform priors over all network structures; however, we introduce a prior on the number of edges in the phenotype network structure, which prefers simple models with fewer directed edges. We determine that the relationship between the prior penalty and the joint posterior probability of the true model is not monotonic, there is some interplay between the two.

Simulation studies were carried out and our approach is also applied to a published data set. It is determined that larger trait-to-trait effects are required to recover the phenotype network structure; however, mixing is generally slow, a common occurrence with reversible jump Markov chain Monte Carlo methods. We propose the use of a double step to combine two steps that alter the phenotype network structure. This proposes larger steps than the traditional birth and death

move types, possibly changing the dimension of the model by more than one. This double step helped the sampler move between different phenotype network structures in simulated data sets.

Acknowledgments

Thank you to Kent Eskridge, P. Stephen Baenziger and B. Todd Campbell for allowing me to use your winter wheat data set.

This work would not have been possible without the support and guidance of many people. Most importantly I would like to express the most sincere appreciation and gratitude to my supervisor, Dr. Nokuthaba Sibanda, for her encouragement, guidance and patience during the last few years. I would also like to thank my secondary supervisor, Dr. Richard Arnold, who always had wise advice when I needed it most.

Thank you to my fellow postgraduate students, and the both the admin and teaching staff within the School of Mathematics, Statistics and Operations Research; you have all played a part in my time at Victoria and make it very hard to leave. A special thanks to Yuichi for all the cups of hot chocolate and the extra advice, and to Ray for all the computing help.

A huge thank you to my friends and family, seeing you all everyday has helped me keep my sanity. Lastly, to my parents, thank you for your unconditional support — I wouldn't be where I am without you.

iv

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A quantitative trait (or phenotype) is a trait that can be measured on a continuous scale, e.g. grain yield (measured in kg). Quantitative traits are controlled by genes, inherited from parents, therefore the section of the genome containing (or linked to) the genes controlling the trait(s) are called quantitative trait loci (QTL). QTL mapping is the identification of QTL affecting single (or multiple) quantitative trait(s); involving, among other things, the estimation of the QTL location within the genome, the QTL genotype and the effect the QTL has on the quantitative trait(s) of interest. These properties of a QTL are explained in Section 1.1.

As QTL mapping approaches were developed to incorporate multiple traits it became necessary to account for the causal relationships that exist within a given set of traits. The estimation of these causal network structures are the focus of more recent QTL mapping approaches and are introduced in Section 1.2. The objectives of QTL mapping over the years are discussed in Section 1.3, where we outline the contribution of this thesis to the field of causal network structure estimation.

Accompanying this introductory chapter is Appendix A, which details Mendelian

1

inheritance and the basic biological principles underlying QTL mapping, as well as a brief description of the breeding designs and the resulting populations commonly used in the QTL mapping literature.

1.1 Properties of a QTL

A QTL is defined by its location within the genome, the genotype, and the effect on the quantitative trait(s) of interest.

1.1.1 QTL location

The genome is the genetic material of an organism, encoded in DNA (deoxyribonucleic acid) and formed into chromosomes. A genetic marker is a gene (or sequence of DNA) with a known location within the genome (i.e. a specific location on a given chromosome), for all individuals of a given species. A set of genetic markers, spread over the chromosome(s) of interest, form the base of the genetic map on which we may estimate the location of putative QTL.

A locus (plural loci) is a particular location within the genome, here this could refer to a genetic marker or putative QTL. Appendix A.1 defines how the distance between two loci is recorded within the genome using Haldane's mapping function. The locations of the genetic markers may be defined using recombination fractions (or recombination frequencies), with the distance between marker 1 and marker 2 on the chromosome labelled k (markers labelled $k1$ and $k2$, respectively) being denoted by r_{k1k2}. The matrix $\mathbf{r}$ denotes the pairwise distances and thus the locations of all genetic markers in the study. Using the known location of genetic markers within the genome we may specify the location of a QTL. A QTL will be located within a marker interval, i.e. between two markers which are referred

to as flanking markers. The ℓth QTL has its location denoted by $\lambda_\ell = (\xi_\ell, \gamma_\ell, \theta_\ell)$; defined by

- the chromosome label (ξ_ℓ),

- the left flanking marker (γ_ℓ), i.e. the genetic marker located to the left of the QTL on the genetic map

- the recombination fraction between the ℓth QTL and the left flanking marker (θ_ℓ).

Figure 1.1 illustrates how the QTL location is defined.
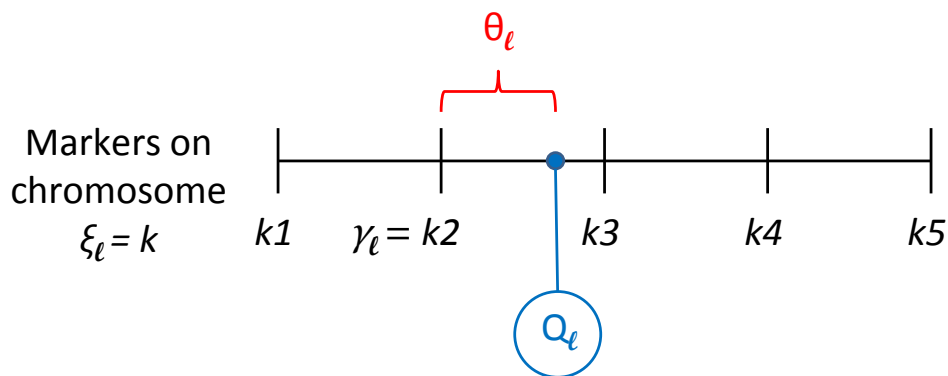
Figure 1.1: *The location of the ℓth QTL is defined by the chromosome it is located on ($\xi_\ell = k$), the left flanking marker ($\gamma_\ell = k2$) and the recombination fraction (θ_ℓ).*

1.1.2 QTL genotype

An allele is the form of a gene existing at a locus within the genome. Diploid species, such as humans, are individuals with two copies of each chromosome, and therefore two alleles at each locus.

Alleles can be either dominant or recessive and are denoted, for example, *A* and *a* respectively. If the two alleles are the same, e.g. *AA* or *aa*, the QTL genotype is said to be homozygous; homozygous dominant for *AA*, and homozygous recessive for *aa*. A heterozygous QTL genotype refers to the presence of one dominant and one recessive allele, e.g. *Aa*. A recessive allele is only fully expressed in the phenotype when the genotype is homozygous recessive. However, the dominant allele is expressed in the phenotype with a homozygous dominant genotype, or with a heterozygous genotype where it may mask the effect of the recessive allele, according to Mendel's law of dominance (see Appendix A.1).

The QTL genotypes present in a population are dependent upon the type of breeding design (or "cross") and the parent genotypes. For example, a doubled haploid population can only yield individuals with a homozygous genotype. Populations commonly referred to in QTL mapping literature are outlined in Appendix A.2.

1.1.3   QTL effect

For the analysis of a single trait, each identified QTL is assumed to have an effect (possibly zero) on the trait of interest and there are multiple ways to model this effect, depending on the population used and the resulting genotypes present. The ℓth QTL is said to have effect $\phi_{\ell t}$ on the tth trait, and we refer to this as the QTL-to-trait effect or simply the QTL effect as QTL can affect traits, but traits cannot affect QTL. For populations with both homozygous and heterozygous genotypes (e.g. an F_2 population) we must be able to account for the dominance effect which occurs with only one copy of the dominant allele (a heterozygous genotype). The widely used Cockerham genetic model is used for this (see Cockerham 1954);

however, the simulated data in this thesis uses a more simple breeding design so
that we only model additive effects.

The analysis of multiple traits allows for more complicated genetic models,
whereby each trait is able to be affected by its own unique set of QTL, i.e. not all
QTL affect all traits.

1.2 Causal network structures

For QTL mapping with multiple quantitative traits, it is necessary to incorporate
the entire causal network structure. We separate the causal network structure into
two components: the genetic component and the phenotype network structure.
The genetic component includes the QTL effects, and what we refer to as the
genetic architecture. We separate the QTL effects from the genetic architecture
as in our approach, in Chapters 4 and 5, we assume that the genetic architecture
is known, but we still wish to estimate the QTL effects. The genetic architecture
includes:

- the number of QTL

- the QTL locations

- the QTL genotypes

- knowledge of which subset of QTL affect each trait.

Knowing which QTL have an effect on each trait means that we know which QTL
effect parameters we need to estimate, which is important as in a multiple trait
analysis not all QTL affect all traits. QTL mapping focuses on the estimation
of the genetic architecture and the QTL effects, see the QTL mapping objectives

listed in Section 1.3. The analysis of multiple traits, however, also requires esti-
mation of the phenotype network structure, i.e. the estimation of the causal rela-
tionships between traits.

Casual network structures (or causal networks) may be represented by a di-
rected acyclic graph (DAG) which is a directed graph without any directed cycles,
or by a system of linear equations as a structural equation model (SEM) (see for
example Wright 1921, Haavelmo 1943, Koopmans et al. 1950, Koopmans 1953,
Pearl 2000, Mi et al. 2010). Each DAG is comprised of a set of nodes to represent
both the QTL and quantitative traits, connected by a set of directed edges which
correspond to the causal relationships existing between QTL and traits (the QTL
effects) and the traits themselves (trait-to-trait effects). An example causal net-
work structure is given in Figure 1.2, with three quantitative traits, denoted Y_1, Y_2
and Y_3, and three QTL, denoted Q_1, Q_2 and Q_3.
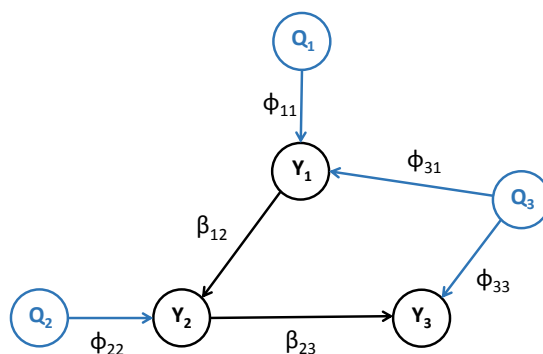
Figure 1.2: *An example causal network structure with three quantitative traits*
(Y_1, Y_2, Y_3), three QTL (Q_1, Q_2, Q_3), two trait-to trait effects (β_{12}, β_{23}) and four
QTL effects (ϕ_{11}, ϕ_{22}, ϕ_{31}, ϕ_{33}). The genetic architecture is assumed known and
is highlighted in blue, the phenotype network structure and QTL effects are black
and require estimation.

With the analysis of multiple traits and the incorporation of trait-to-trait effects into the model, a causal relationship can occur either directly or indirectly. If there is a directed edge between two nodes, say Q_1 and Y_1, such that

$$Q_1 \rightarrow Y_1$$

then Q_1 is said to have a direct effect on Y_1. An indirect effect is mediated by one (or more) other variables, e.g. Q_1 is said to have an indirect effect on Y_2 when

$$Q_1 \rightarrow Y_1 \rightarrow Y_2.$$

It is possible that, when added together to obtain the total effect, the direct and indirect effects can cancel each other out (i.e. one positive effect, one negative effect), resulting in the total effect being close to zero. The motivation to analyse both the direct and indirect effects is that by only considering the total effect of a QTL, methods can fail to detect QTL with significant direct and indirect effects (Mi et al. 2010).

There are two causal relationships existing between traits in the example causal network structure given in Figure 1.2:

$$Y_1 \rightarrow Y_2 \quad \text{and} \quad Y_2 \rightarrow Y_3$$

with the parameters β_{12} and β_{23}, respectively. The following relationships between QTL and traits also exist in Figure 1.2:

$$Q_1 \rightarrow Y_1, \quad Q_2 \rightarrow Y_2, \quad Q_3 \rightarrow Y_1 \quad \text{and} \quad Q_3 \rightarrow Y_3$$

corresponding to the QTL effect parameters ϕ_{11}, ϕ_{22}, ϕ_{31} and ϕ_{33}, respectively.

The DAG in Figure 1.2 has much in common with a structural equation model (SEM), as a SEM is a generalization of simultaneous equation procedures used in path analysis, developed by Wright (1921), Haavelmo (1943), Koopmans et al. (1950), and Koopmans (1953). Here, we use SEM within a specific genetic setting, and incorporate further information regarding the genetic markers and genetic map, see for example Mi et al. (2010).

It follows that the graphical model in Figure 1.2 (p. 6) can also be written as a structural equation model:

$$
\begin{aligned}
y_{i1} &= \beta_{01} + \phi_{11}q_{i1} + \phi_{31}q_{i3} + \varepsilon_{i1} \\
y_{i2} &= \beta_{02} + \phi_{22}q_{i2} + \beta_{12}y_{i1} + \varepsilon_{i2} \\
y_{i3} &= \beta_{03} + \phi_{33}q_{i3} + \beta_{23}y_{i2} + \varepsilon_{i3}
\end{aligned}
$$

where for individual i, the QTL genotypes are denoted $q_{i\ell}$ for the ℓth QTL ($\ell = 1, 2, 3$) and for trait t ($t = 1, 2, 3$); the quantitative trait values are denoted y_{it}. The intercept is denoted β_{0t} and the residuals are denoted ε_{it}:

$$
\varepsilon_{it} \overset{i.i.d}{\sim} N(0, \sigma_t^2).
$$

1.3   QTL mapping objectives

The goals of QTL mapping have evolved over the years and with this the mapping methods have as well. Over time, QTL mapping has extended from mapping QTL for only one quantitative trait, to the inclusion of multiple traits, and is currently focused on the estimation of the phenotype network structure existing between

multiple interacting traits. Although the focus is to map the causal relationships between multiple traits, this is still a QTL mapping problem as the estimated trait-to-trait effects are used to determine the indirect effect of a QTL on a given trait. QTL mapping approaches have focused on any or all of the following six objectives:

1. identify the locations of any QTL affecting each trait of interest

2. identify the QTL genotype for each individual

3. identify the size of the QTL effects on the trait(s) of interest

4. identify the number of QTL affecting the trait(s) of interest

5. identify the size of the trait-to-trait effects within a given causal network structure

6. identify the phenotype network structure between correlated traits, i.e. identify which traits have an effect on other traits, corresponding to indirect QTL effects.

This thesis focuses on the estimation of the (unknown) phenotype network structure. The estimation of an unknown causal network structure is important as they can be used to model the biological pathways influencing complex diseases such as diabetes (Neto et al. 2010), and knowledge of a causal network structure can allow for predictions to be made about the effect of certain interventions (Heckerman et al. 1995). For example, a causal network structure could signal that certain genes imply a susceptibility to a disease, and interventions could focus on selective breeding programs to diminish this. There are challenges with estimating an unknown causal network or phenotype network structure, however. Most

importantly, an efficient search algorithm is required to move between candidate models, as the number of possible causal network structures increases rapidly as the number of traits increases.

Chapter 2 reviews existing QTL mapping methods which meet the sixth QTL mapping objective, which is to say that they are able to identify the phenotype network structure existing between a set of quantitative traits. The estimation of an unknown causal or phenotype network structure makes this a model determination problem (Neto et al. 2008, 2010). The necessary graphical model terminology are also outlined to help describe the comparison of causal network structures.

In Chapter 3 we use a score-based greedy hill-climbing search strategy to carry out a simulation study to determine how the model space can be searched, assuming that the genetic architecture has already been determined using established methods. We fit a system of linear equations and assess the goodness-of-fit of a model by obtaining the maximum likelihood estimates and calculating the corrected Akaike Information Criterion (the AICc) which is used for model comparison.

Building on the existing QTL mapping and causal network structure estimation methods we form a trans-dimensional Bayesian approach to estimate an unknown phenotype network structure using reversible jump Markov chain Monte Carlo in Chapters 4 and 5. We address the third, fifth and sixth QTL mapping objectives with our approach, recommending the use of existing methods to estimate the genetic architecture. We incorporate a prior for the phenotype network structure which discourages the chain from moving to overly complicated models with too many causal relationships, and make use of maximum likelihood estimates in the chain to create an efficient sampler using well-tuned proposal distributions. In Chapter 5 we also propose a 'double' step which combines two of the phenotype

structure altering steps used here, creating six possible types of double step: add two edges, add an edge and remove an existing edge, add an edge and reverse an existing edge, remove two existing edges, remove one edge and reverse another, and reverse two existing edges. This new move type allows the sampler to propose larger steps to move between models, increasing mixing.

The final chapter, Chapter 6, summarizes our findings and discusses the contribution of this thesis to the field of causal network structure estimation, which is to extend the traditional trans-dimensional reversible jump algorithm to develop a Bayesian approach to estimate an unknown phenotype network structure between multiple interacting traits.

Chapter 2

Review of methods estimating a causal network

This thesis focuses on the estimation of a phenotype network structure, therefore we will not provide details for earlier QTL mapping approaches which do not incorporate the phenotype network structure into the analysis. Zeng (1994) provides a good summary of established QTL mapping methods for a single quantitative trait, including the use of simple linear regression (see also Soller et al. 1976), interval mapping (see also Lander & Botstein 1989), and composite interval mapping. Many of these methods rely on strong simplifying assumptions, in particular often assuming that the number of QTL is known. The reversible jump Markov chain Monte Carlo algorithm proposed by Green (1995) is the exception, treating the number of QTL as an unknown parameter of interest.

In QTL mapping experiments one may obtain observations on multiple quantitative traits. When jointly analysing multiple quantitative traits, the correlation structure existing between them should be incorporated in order to improve the power and precision of mapping QTL, compared to methods which analyse traits

separately (Jiang & Zeng 1995, Banerjee et al. 2008). Notable methods for multiple trait QTL mapping include composite interval mapping (Jiang & Zeng 1995), multivariate regression (Hackett et al. 2001), the seemingly unrelated regression model (Banerjee et al. 2008), and the structural equation model (Mi et al. 2010).

Many biological studies focus on correlated phenotypes, but traditional QTL mapping methods fail to incorporate the causal network structure of the phenotypes into the analysis, also referred to as the phenotype network. As a result, the genetic architecture inferred from the analysis is often misleading because such methods can identify QTL as having a direct effect on the phenotype of interest, when in fact, they may have an indirect effect. A QTL with an indirect effect is one which is conditionally independent of the phenotype of interest, given an intervening trait; it affects other phenotypes which in turn affect the phenotype of interest. These causal networks, consisting of QTL, phenotypes, and the causal relationships between them, can become very intertwined. Therefore, in order to correctly infer the genetic architecture, the phenotype network structure must be included in the analysis.

This chapter briefly summarises existing methods for estimating the direction of causality and the causal effect parameters when the causal direction is unknown. Existing methods include the use of structural equation modeling combined with model assessments such as the goodness-of-fit statistic or the AIC to refine the causal network structure, as well as the use of the Markov chain Monte Carlo algorithm to jointly infer the genetic architecture and the phenotype network; see Sections 2.2 and 2.3, respectively. We incorporate some of the key components from these analyses into our extension of the traditional trans-dimensional reversible jump algorithm, which estimates an unknown phenotype network structure between multiple interacting traits.

As the causal network structure is unknown, QTL mapping approaches require ways to move between candidate causal network structures which may differ in dimension. There must be a clearly defined way in which they distinguish between different causal relationships between traits. Causal networks can be represented as a directed graph without any cycles; as a directed acyclic graph, or DAG. These networks are comprised of nodes connected by directed edges which imply causality. The directed edges help to distinguish between networks which, if undirected, would be considered likelihood equivalent (having the same joint density). Causal networks incorporate the causal relationships existing between both the QTL and phenotypes, and between the phenotypes themselves. In order to infer an unknown phenotype network structure we must define the terminology we use to describe graphical models; see Section 2.1.

2.1 Graphical model terminology

A graphical model, G, is defined by a set of nodes (V) and a set of directed edges (E) such that $G = (V, E)$. There are a number of terms which may be used to describe a graphical model. This section aims to give a brief introduction to some of the graphical model terminology used when estimating the phenotype network structure.

> **Definition 1:** *direct effect*. A direct effect is represented by a directed edge. E.g. $Y_1 \rightarrow Y_2$ denotes the direct effect that the trait Y_1 has on Y_2.

> **Definition 2:** *directed path*. A directed path is a sequence of edges following the direction of causality between two specified nodes. Figure 2.1 gives an example of a directed path from node Y_1 to node Y_4.
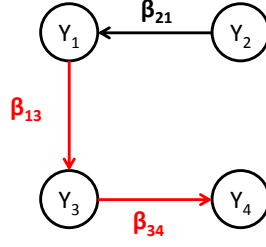
Figure 2.1: *The directed path from Y_1 to Y_4 is highlighted in red ($Y_1 \to Y_3 \to Y_4$), and is comprised of the edges denoted by β_{13} and β_{34}.*

Definition 3: *indirect effect*. When there is a directed path from node *A* to node *B*, mediated by at least one other variable, node *A* is said to have an indirect effect on node *B*. Note that node *A* can be either a QTL or a trait; however, node *B* must be a trait as QTL can affect traits, but traits cannot affect QTL. The size of an indirect effect is calculated by evaluating the product of path coefficients for each path from the associated variable to the dependent variable. For example:

$$
\begin{aligned}
y_{i1} &= \beta_{01} + \phi_{11}q_{i1} + \varepsilon_{i1} \\
y_{i2} &= \beta_{02} + \beta_{12}y_{i1} + \varepsilon_{i2}
\end{aligned}
$$

where β_{0t} is the intercept for the *t*th trait, and for individual *i*, the observed trait value for the *t*th trait is denoted y_{it}, and the ℓth QTL genotype is denoted $q_{i\ell}$ with the QTL effect $\phi_{\ell t}$ on the *t*th trait. The residuals are denoted ε_{it}. QTL Q_1 has a direct effect on trait Y_1 (ϕ_{11}). As Y_1 has a direct effect on trait Y_2 (β_{12}), QTL Q_1 is said to have a indirect effect on Y_2 as for any *i*th individual:

$$
\begin{aligned}
y_{i2} &= \beta_{02} + \beta_{12}\left(\beta_{01} + \phi_{11}q_{i1} + \varepsilon_{i1}\right) + \varepsilon_{i2} \\
&= \left(\beta_{02} + \beta_{01}\beta_{12}\right) + \beta_{12}\phi_{11}q_{i1} + \left(\beta_{12}\varepsilon_{i1} + \varepsilon_{i2}\right).
\end{aligned}
$$

The indirect effect of of QTL Q_1 on trait Y_2 is therefore equal to the product of the path coefficients, $\beta_{12}\phi_{11}$.

Definition 4: *total effect*. The total effect a QTL or trait has on a given trait is the sum of both the direct and indirect effects (Mi et al. 2010).

Definition 5: *fork*. A fork is formed by two directed edges originating from the same node, to two different nodes. An example is given in Figure 2.2.
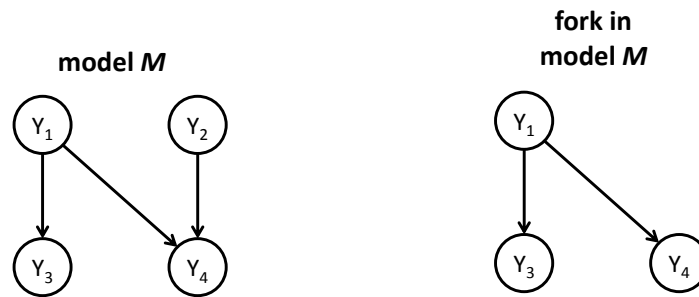
Figure 2.2: *A fork exists within the phenotype network structure created by the nodes Y_1, Y_2 and Y_3.*

Definition 6: *v-structure*. A v-structure is formed by two converging directed edges, where the tails are not connected by an edge. An example is given in Figure 2.3.
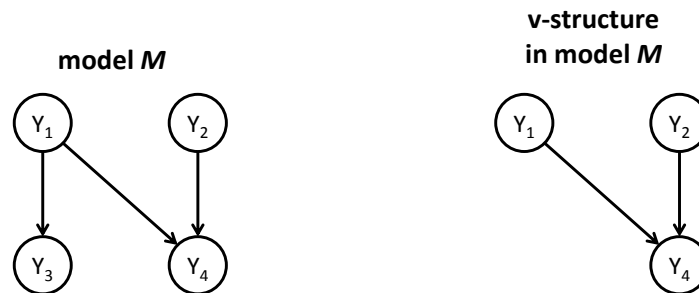
Figure 2.3: *A v-structure exists within the phenotype network structure created by the nodes Y_1, Y_2 and Y_3.*

Definition 7: *d-separated.* A set of nodes, $\mathbf{Z}$, is said to d-separate set $\mathbf{X}$ from $\mathbf{Y}$ if and only if the nodes in $\mathbf{Z}$ block all paths from a node in $\mathbf{X}$ to a node in $\mathbf{Y}$. If sets $\mathbf{X}$ and $\mathbf{Y}$ are not d-separated, then they are said to be d-connected (Pearl 1988, 2000). Example phenotype networks M_s and M_c are given in Figure 2.4 to display a d-separated and a d-connected structure, respectively. For both M_s and M_c, let $\mathbf{Z} = \{Y_1, Y_4\}$, $\mathbf{X} = \{Y_2\}$ and $\mathbf{Y} = \{Y_6\}$.
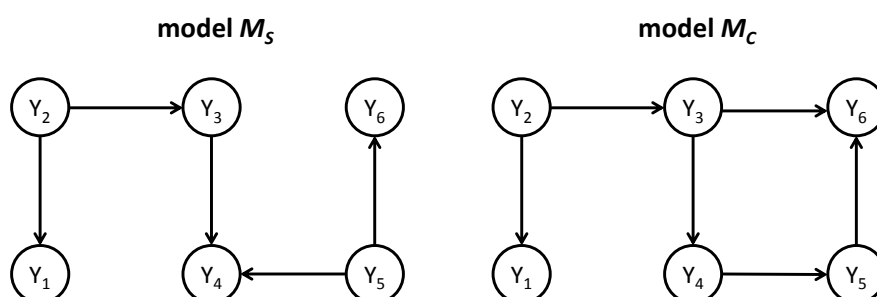
Figure 2.4: *The phenotype network M_s is d-separated as the node Y_4 (in $\mathbf{Z}$) blocks all paths from Y_2 ($\mathbf{X}$) to Y_6 ($\mathbf{Y}$). The phenotype network M_c is d-connected as the node Y_4 (in $\mathbf{Z}$) does not block the path from Y_2 ($\mathbf{X}$) to Y_6 ($\mathbf{Y}$).*

Definition 8: *skeleton.* The skeleton of a graph is obtained by removing the direction of the edges — creating an undirected graph, or UDG. An example is given in Figure 2.5.
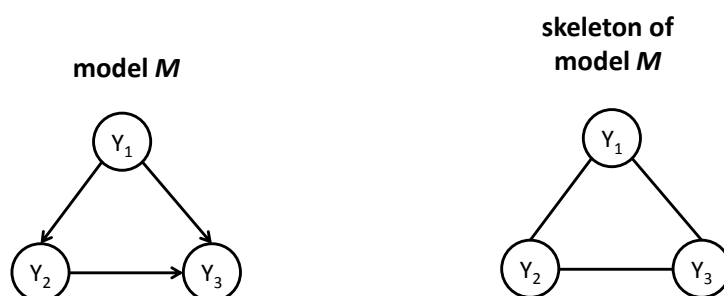
Figure 2.5: *The causal network structure M is a directed acyclic graph, the skeleton of M is an undirected graph.*

Definition 9: *directed cycle.* A directed cycle is a directed path that begins and ends at the same node (Lauritzen 1996). In this thesis we work with directed edges only, and so a directed cycle is simply referred to as a cycle and any causal network structure containing a cycle is said to be cyclic. An example is given in Figure 2.6.
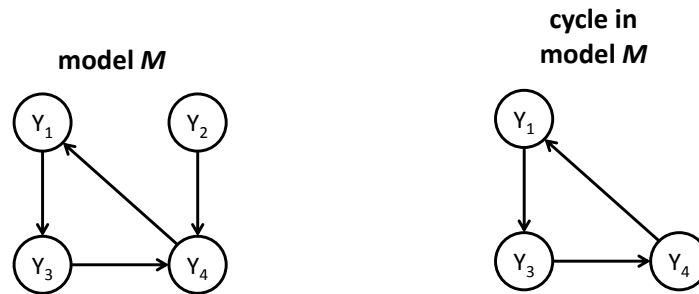
Figure 2.6: *The phenotype network structure M is cyclic as it contains a cycle.*

Definition 10: *likelihood equivalent.* Two causal networks with the same joint densities are said to be likelihood equivalent. An example of likelihood equivalence is given below.

Causal networks can often be likelihood equivalent, and in such instances model selection methods are unable to distinguish between the joint densities and the direction of a causal relationship cannot be determined. Therefore, causal relationships between phenotypes require the use of conditional probability in order to distinguish between different causal networks (Neto et al. 2010). This is demonstrated in the following example, adapted from Neto et al. (2008). For two quantitative traits, there exist two possible causal network structures containing a single directed edge, given in Figure 2.7. Models M_1 and M_2 are likelihood equivalent as the joint densities, denoted $f(y_1, y_2)$, are the same:

Model M_1 Model M_2

Figure 2.7: *Two possible phenotype network structures for traits Y_1 and Y_2.*

$$f(y_1)f(y_2|y_1) = f(y_1, y_2) = f(y_2)f(y_1|y_2)$$

(Neto et al. 2008). Now suppose that there exist three quantitative traits; Y_1, Y_2 and Y_3. Two phenotype network structures, models M_A and M_B, are given in Figure 2.8. Phenotype networks M_A (consisting of a v-structure) and M_B (consisting of a fork) are not likelihood equivalent as the joint densities are not equal. For M_A we have:

$$f(Y_1, Y_2, Y_3) = f(Y_2|Y_1, Y_3)f(Y_1)f(Y_3)$$

and for M_B:

$$f(Y_1, Y_2, Y_3) = f(Y_1|Y_2)f(Y_3|Y_2)f(Y_2)$$

M_A   **M_B**

Figure 2.8: Example phenotype networks M_A and M_B .

such that

$$
\begin{aligned}
f(Y_1, Y_2, Y_3) &= f(Y_1|Y_2)f(Y_3|Y_2)f(Y_2) \\
&= f(Y_1|Y_2)\frac{f(Y_2|Y_3)f(Y_3)}{f(Y_2)}f(Y_2) \\
&= f(Y_1|Y_2)f(Y_2|Y_3)f(Y_3) \\
&= \frac{f(Y_2|Y_1)f(Y_1)}{f(Y_2)}f(Y_2|Y_3)f(Y_3) \\
&= \frac{f(Y_2|Y_1)f(Y_2|Y_3)}{f(Y_2)}f(Y_1)f(Y_3) \\
&\neq f(Y_2|Y_1, Y_3)f(Y_1)f(Y_3).
\end{aligned}
$$

Model selection methods which incorporate the joint density function are able to distinguish between models.

2.2 SEM estimation of an unknown causal network structure

A generalization of simultaneous equation procedures used in path analysis, developed by Wright (1921), Haavelmo (1943), Koopmans et al. (1950), Koopmans (1953), the structural equation model (SEM) incorporates the causal relationships existing between traits and estimates the direct and indirect effect each QTL has on each trait (Mi et al. 2010). The drawback of the structural equation model is that the causal network structure has to be known *a priori*.

For the *i*th individual, there are observations on N_T causally related quantitative traits, denoted by the matrix $\mathbf{Y}$, where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_i, \cdots, \mathbf{y}_n]$ and $\mathbf{y}_i = [y_{1i}, \cdots, y_{iN_T}]^\top$. These traits are affected (additively) by N_{qtl} QTL, located at positions $\lambda_1, \lambda_2, \cdots, \lambda_{N_{qtl}}$ in N_{qtl} different marker intervals. For a RIL cross, genotypes

are homozygous ('*AA*' and '*aa*', for example) (see Appendix A.2.4). The structural equation model is as follows:

$$\mathbf{y}_i = \mathbf{B}\mathbf{y}_i + \alpha_\ell \mathbf{q}_i + \zeta_i \tag{2.1}$$

$$
\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iN_T} \end{bmatrix} =
\begin{bmatrix} 0 & \beta_{12} & \cdots & \beta_{1N_T} \\ 0 & 0 & \cdots & \beta_{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}
\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iN_T} \end{bmatrix} +
\begin{bmatrix} \alpha_{11} & \cdots & \alpha_{N_{qtl}1} \\ \alpha_{12} & \cdots & \alpha_{N_{qtl}2} \\ \vdots & \ddots & \vdots \\ \alpha_{1N_T} & \cdots & \alpha_{N_{qtl}N_T} \end{bmatrix}
\begin{bmatrix} q_{i1} \\ q_{i2} \\ \vdots \\ q_{iN_{qtl}} \end{bmatrix} +
\begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iN_T} \end{bmatrix}
\tag{2.2}
$$

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{a}\mathbf{Q} + \mathbf{E} \tag{2.3}$$

where β_{hk} is the effect of trait h on trait k ($\mathbf{B}$ is an $N_T \times N_T$ matrix comprised of β_{hk} for $h = 1, \cdots, N_T$ and $k = 1, \cdots, N_T$), and $\beta_{hk} = 0$ for $k \le h$. Note that the constraint $\beta_{hk} = 0$ for $k \le h$ means that a DAG structure is required for identifiability. The direct effect of the ℓth QTL on trait k is denoted $\alpha_{\ell k}$ ($\mathbf{a} = \{\alpha_{\ell k}\}$ for $\ell = 1, \cdots, N_{qtl}$ and $k = 1, \cdots, N_T$), and $q_{i\ell}$ is the genotype of the ℓth QTL for the ith individual such that $\mathbf{Q}$ is a matrix, $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_i, \cdots, \mathbf{q}_n]$ and $\mathbf{q}_i = [q_{i1}, \cdots, q_{iN_{qtl}}]^\top$. The residual effect on trait k for individual i, is denoted e_{ik} ($\mathbf{E} = \{e_{ik}\}$ for $i = 1, \cdots, n$ and $k = 1, \cdots, N_T$) and is assumed to have a multivariate normal distribution with mean zero and covariance matrix Ψ:

$$
\Psi = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N_T}^2 \end{pmatrix}.
\tag{2.4}
$$

The structural equation model given in Equations 2.1 – 2.3 incorporates the

QTL effects; however, the structural equation model can also model genotype-by-environment interactions (see Dhungana et al. 2007). The parameters can be estimated via a Markov chain Monte Carlo (MCMC) approach (see Mi et al. 2010, for example) or by maximum likelihood estimation (see Dhungana et al. 2007, for example).

In order to estimate an unknown causal network structure, Li et al. (2006) use SEM to assess the goodness-of-fit of a given model, and propose small adjustments to the causal structure by adding, removing or reversing a single causal relationship in order to explore the model space and identify the causal network structure which best fits the data. The approach is summarized into five steps, given in Section 2.2.1 as published in Li et al. (2006). The approach begins with the detection of QTL with effects on the given quantitative traits and the identification of any pleiotropic QTL (affecting multiple traits) using existing methods. An initial causal network structure is proposed using the QTL information obtained, and is then iteratively refined using model assessments such as the goodness-of-fit statistic or the AIC.

2.2.1 Development of an unknown causal network structure

1. *Identify QTL for individual phenotypes:* use existing QTL mapping methods to carry out a single locus genome scan in search of a QTL affecting each of the quantitative traits, see e.g. Sen & Churchill (2001). The presence of any QTL can be determined via likelihood ratios using the following linear model:

$$Y = \beta_0 + \beta_1 Q + \varepsilon \tag{2.5}$$

which incorporates only additive QTL effects (Li et al. 2006). The quantitative trait value is denoted by the $n \times 1$ vector Y, and the QTL genotypes are denoted by the $n \times 1$ vector Q. The intercept is denoted β_0; β_1 is the QTL effect, and the $n \times 1$ residual vector is denoted ε.

2. *Identify pleiotropic QTL:* conditional genome scans are carried out using the following model:

$$Y = \beta_0 + \beta_1 Q + \beta_2 X + \varepsilon \qquad (2.6)$$

which is the same as the linear model used in step 1, but now another trait, denoted X, is used as a covariate such that β_2 is the effect of trait X on trait Y, adjusted for the QTL effect. Note that many (even all possible) covariates could be (individually) fitted. A large change in the likelihood ratios from steps 1 and 2 indicates that trait X is causally connected to trait Y and QTL Q (Li et al. 2006).

3. *Define an initial path model:* each QTL and trait of interest is represented by a node in the causal network, and edges should be added from each QTL to the appropriate traits (according to the results from step 1). A significant change in the likelihood ratios in step 2 indicates that directed edges should be placed from QTL Q to both traits X and Y, with an additional edge from the conditioning trait, X, to the response, Y. Step 5, the refinement step, will assess both the significance and the causal direction of this edge.

4. *Assessment of the model:* construct t-tests to test the significance of each of the individual path coefficients, and use the goodness-of-fit test statistic to determine whether or not the model provides a good fit to the data (Li

et al. 2006, Bentler & Bonett 1980, Joreskog 1970). If the model does not provide a good fit to the data (the goodness-of-fit test statistic will be significant, $p < 0.05$), the model should be refined (Li et al. 2006).

5. *Refine the model:* in order to refine the model, new candidate causal network structures are proposed by either adding a new edge, removing an existing edge or reversing the causal direction of an edge (between the phenotypes), providing the graph remains acyclic. These models form the neighbourhood, and the goodness-of-fit of each causal network structure within the neighbourhood is assessed and the best model selected (Li et al. 2006). Different model selection methods may be used: the goodness-of-fit test statistic (for nested models) or the AIC (for non-nested models), for example.

Steps 4 and 5 are repeated until a final model has been selected and can't be improved. Note that the proposed alterations to the causal network structure are commonly used, and not unique to Li et al. (2006); see for example Heckerman et al. (1995). The final model should have the following properties specified by Li et al. (2006):

1. The final model must have at least one degree of freedom, i.e. there exist more data points than parameters.

2. The p-value for the goodness-of-fit test should be greater than 0.05, indicating that the final model provides a good fit to the data.

3. The absolute value of each standardized residual should be less than 2.

4. The individual t-tests on each path coefficient should indicate significant (non-zero) effects.

5. The standardized path coefficients should have absolute values exceeding 0.05 (i.e. effects should not be trivial).

6. A large proportion of the phenotypic variance of the dependent variables should be explained by the model.

Li et al. (2006) use the structural equation model to assess the goodness-of-fit of the current model, and then propose small changes to the causal network structure in order to iteratively search the model space. This approach is able to meet all six of the QTL mapping objectives given in Section 1.3, as it begins by scanning the genome for QTL, and also provides the framework to alter the phenotype network structure.

As with all approaches to estimating an unknown causal network structure, missing or misspecified model components can result in incorrect causal relationships being inferred. Li et al. (2006) use likelihood ratios to create a likely initial model, which starts the iterative search in a better position than a randomly selected initial model would. However, with greedy approaches it is normally advantageous to repeat the search from a variety of different positions, with different initial models in order to avoid local maxima. Li et al.'s initial model starts the search in a good position from where it is likely to find the global maxima; however, this is not guaranteed. It is recommended to repeat the search from a variety of different positions.

2.3   The QTLnet algorithm

Neto et al. (2010) created the QTLnet algorithm as a method to jointly infer the

phenotype network and the genetic architecture. The QTLnet algorithm infers the genetic architecture for each phenotype, conditional on the phenotype network, and is therefore ideal when trying to estimate an unknown causal network structure. The model used is similar to the seemingly unrelated regression model given in Banerjee et al. (2008) as both QTLs and covariates enter the genetic model through the mean:

$$y_{it} = \mu_{it}^* + \sum_{v \in \upsilon_Y(t)} \beta_{vt} y_{iv} + \varepsilon_{it} \tag{2.7}$$

for individual i ($i = 1, \cdots, n$) and phenotype t ($t = 1, \cdots, N_T$) (Neto et al. 2010). The phenotypic trait value is denoted by y_{it}, with $\mathbf{y} = (\mathbf{y}_1, \cdots, \mathbf{y}_i, \cdots, \mathbf{y}_n)$ and $\mathbf{y}_i = (y_{1i}, \cdots, y_{ti}, \cdots, y_{N_Ti})^\top$. The 'parents' of the tth trait, $\upsilon_Y(t)$, refer to all traits with a direct effect on trait t. The mean of the tth trait for the ith individual is given by μ_{it}^*, $\mu_{it}^* = \mu_t + \mathbf{X}_{it}\theta_t$, such that μ_t is the overall mean for a given trait t, the column vector θ_t represents the genetic effect of QTL on trait t, and the row vector $\mathbf{X}_{it}$ denotes the genetic predictors derived from QTL genotypes. β_{vt} is the trait-to-trait effect of trait v on trait t, and the errors are assumed to follow a Normal distribution, such that $\varepsilon_{it} \sim N(0, \sigma_t^2)$.

Assuming that the causal network structure, denoted M, with N_T quantitative traits is a DAG, the likelihood function for individual i factors as follows:

$$p(\mathbf{y}_i | \mathbf{q}_i, \Gamma, M) = \prod_{t=1}^{N_T} p(y_{it} | \mathbf{q}_{it}, \upsilon_Y(t)) \tag{2.8}$$

where the unobserved QTL genotypes for individual i are denoted by $\mathbf{q}_i$, and for the tth trait of the ith individual, $\mathbf{q}_{it}$ (Neto et al. 2010). All other model parameters

are denoted by Γ. Following from the model given in Equation 2.7:

$$p(y_{it}|\mathbf{q}_{it}, \upsilon_Y(t)) = N\left(\mu_{it}^* + \sum_{v \in \upsilon_Y(t)} \beta_{vt} y_{iv}, \sigma_t^2 \right). \tag{2.9}$$

There are many advantages to using this model, most notably, it allows each phenotype to have its own unique genetic architecture.

Neto et al. (2010) utilise a Metropolis-Hastings algorithm and propose modifications to the DAG representing the current model. Such modifications include the addition, deletion or reversal of a single causal relationship, existing within the current model, provided this does not result in a cyclic network. These are standard moves used to alter a causal network structure; see for example Heckerman et al. (1995) and Li et al. (2006). The DAGs resulting from these modifications make up the neighbourhood of model M, denoted $Ne(M)$ (Neto et al. 2010). The number of candidate causal network structures in the neighbourhood of model M is denoted $|Ne(M)|$. The algorithm is as follows:

1. Given a current model or phenotype network structure, $M_{current}$, the neighbourhood is determined and a candidate model selected from this with proposal probability

$$q(M_{candidate}|M_{current}) = \frac{1}{|Ne(M_{current})|}.$$

   The candidate model, $M_{candidate}$, is the result of a single modification (the addition, deletion or reversal of a causal relationship), and because the neighbourhood excludes cyclic networks it can be assumed that the candidate model is a DAG.

2. Factor the likelihood according to the candidate model and for each of the $t = 1, \cdots, N_T$ phenotypes, perform a QTL mapping analysis conditional on its parents, $\upsilon_Y(t)$.

3. Obtain $\widehat{p}(\mathbf{y}|\mathbf{q}, M_{candidate})$ by individually estimating the components of the marginal likelihood $p(y_{it}|\mathbf{q}_{it}, \upsilon_Y(t))$, see Equation 2.9.

4. The candidate model is accepted with probability

$$\alpha = min\left\{ 1, \frac{p(\mathbf{y}|\mathbf{q}, M_{candidate})\, p(M_{candidate})}{p(\mathbf{y}|\mathbf{q}, M_{current})\, p(M_{current})} \frac{q(M_{current}|M_{candidate})}{q(M_{candidate}|M_{current})} \right\}$$

(Neto et al. 2010).

5. Repeat steps 1 – 4 until the chain has converged to a stationary state and a large number of samples from the posterior have been drawn.

The Metropolis–Hastings algorithm proposed by Neto et al. (2010) only makes small modifications at each step and is therefore slow. In order for the Markov chain to mix well, long chains and large thinning windows are necessary. Neto et al. (2010) suggest a few ways to improve computation efficiency when using the QTLnet algorithm. Firstly, step 2 is identified as the most computationally demanding step. The current and candidate models will only differ slightly, so it is best to carry out mapping analysis only on the two nodes affected by an edge reversal, or on the single node affected by the addition or deletion of an edge. For an edge reversal, it is necessary to redo the mapping analysis for both nodes, say reversing the edge from $y_1 \rightarrow y_2$ to $y_1 \leftarrow y_2$, because the conditional relationships change: y_1 is now conditional on y_2.

The second recommendation given by Neto et al. (2010) is to conduct a multiple-trait multiple-QTL analysis prior to implementing the QTLnet algorithm. The seemingly unrelated regression model (Banerjee et al. 2008) is recommended. This prior analysis will identify all QTL linked to each of the phenotypes, individually. Incorporating the resulting relationships into the QTLnet algorithm then limits the possible QTL locations to those already identified. As the QTL locations are already known, only estimates for their effects on each phenotype are required; reducing the number of computations required by the QTLnet algorithm.

Figure 2.9, adapted from Husmeier (2003) and Neto et al. (2010), gives an example of a single step in the Metropolis-Hastings algorithm outlined above. Suppose there is a current model with three quantitative traits (Y_1, Y_2, Y_3) and causal relationships between Y_1 and Y_3 ($Y_1 \rightarrow Y_3$) and Y_3 and Y_2 ($Y_3 \rightarrow Y_2$). The neighbourhood for the current phenotype network structure is given on the left of Figure 2.9, displaying the set of candidate models to choose from. There are just five DAGs comprising the neighbourhood for the current model as the sixth phenotype network is actually cyclic so it must be excluded, see Figure 2.10. A candidate model is therefore selected with probability $q(M_{candidate}|M_{current}) = \frac{1}{5}$; this is the proposal probability of moving from the current to the candidate model.

Suppose the deletion of the causal relationship between items Y_3 and Y_2 is selected. The neighbourhood of the candidate model includes six DAGs, making the proposal probability to move from the candidate model back to the current model $q(M_{current}|M_{candidate}) = \frac{1}{6}$. In this instance, the Metropolis-Hastings proposal ratio is

$$\frac{q(M_{current}|M_{candidate})}{q(M_{candidate}|M_{current})} = \frac{5}{6}.$$

The QTLnet algorithm proposed by Neto et al. (2010) is able to meet all six of
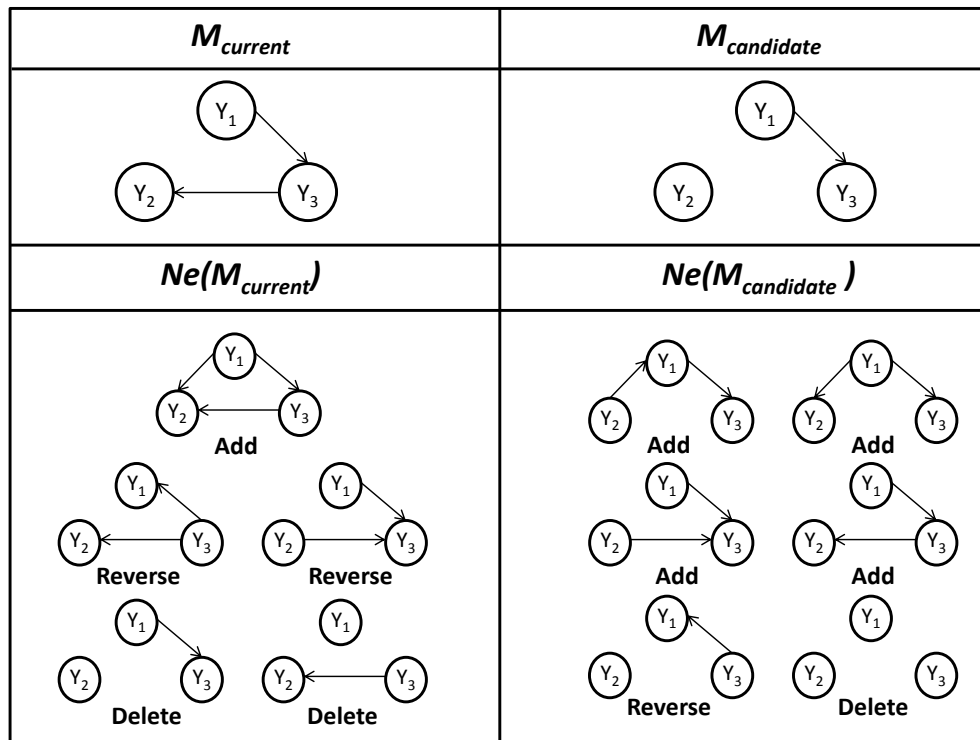
Figure 2.9: *The neighbourhoods for the current and candidate phenotype networks; here* $|Ne(M_{current})| = 5$ *and* $|Ne(M_{candidate})| = 6$.
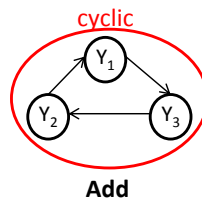
Figure 2.10: *This phenotype network was removed from the neighbourhood of the current phenotype network as it is cyclic.*

the QTL mapping objectives given in Section 1.3, as it alters the phenotype network structure and then infers the genetic architecture for each phenotype, conditional on the current phenotype network.

Both the approach proposed by Li et al. (2006) and the QTLnet algorithm proposed by Neto et al. (2010) move between candidate models in the same way (via the widely used steps which add an edge, remove an edge, or reverse an edge). Li et al. (2006) take a score-based approach, and Neto et al. (2010) take an MCMC approach. A score-based approach is generally faster than an MCMC approach; however, the number of candidate models in the neighbourhood increases rapidly as the number of quantitative traits increases (Robinson 1977). So as the causal network structures become more complex, the search strategy becomes slower, due to the increase in the number of models it must fit. In Chapter 3 we investigate the use of a score-based approach to determine the unknown phenotype network structure. We consider two approaches: the first compares all possible causal network structures and selects the model with the best fit as determined by the AICc, and the second approach sequentially builds up the model using the greedy hill-climbing search strategy. Following Li et al. (2006) and Neto et al. (2010) we consider steps which add an edge, remove an edge, or reverse an edge within the current causal network structure. In Chapters 4 and 5 we propose a reversible jump Markov chain Monte Carlo algorithm to estimate an unknown phenotype network structure.

Chapter 3

Estimating an unknown phenotype network structure using AICc

Chapter 2 introduced the QTLnet algorithm proposed by Neto et al. (2010), capable of jointly inferring the phenotype network structure and the genetic architecture. As numerous efficient methods for estimating the genetic architecture of a trait have already been established (see for example, Stephens & Fisch 1998; Sibanda 2002; Satagopan et al. 1996; Heath 1997), in this thesis we focus on the proposal of an approach to estimate the unknown phenotype network structure of a given set of quantitative traits, assuming that the genetic architecture underlying these traits is known. In other words, both the location of the N_{qtl} QTL affecting N_T traits and the genotype of each QTL for all n individuals are known *a priori*, and we also know which QTL affect which traits (i.e. the parent QTL nodes, $\upsilon_Q(t)$, are known for all t traits), but the QTL effects are unknown.

We use a score-based approach, see for example Heckerman et al. (1995) and Friedman et al. (1999), over a constraint based approach like that of Tsamardinos et al. (2006) and Spirtes et al. (2000), as constraint based approaches are unable to

33

reliably identify the conditional independence properties (Margaritis 2003). Furthermore, constraint-based approaches require two separate steps in order to obtain an estimate of a causal network structure with directed edges. The first step uses conditional independence tests to identify a set of conditional independence properties, forming the skeleton of a graph (a UDG), see Section 2.1 (Zhou 2011). Given the UDG, the second step of a constraint based approach determines the DAG.

Score-based approaches assign a score to each model (with a specific causal network structure) given the data. Here we use the Akaike Information Criterion with a correction for the sample size (the AICc). The Akaike Information Criterion (AIC) is based upon Fisher's maximized log-likelihood and is an estimator of divergence between the true model of the data and the approximating model as given by the Kullback-Leibler distance (Akaike 1973, Burnham & Anderson 2002). Both the AIC and AICc penalize models with many parameters (see, for example, Akaike 1973 and Hurvich & Tsai 1989). The AICc is the AIC with a correction for the sample size, such that as the sample size increases the AICc converges to the AIC. Therefore, regardless of the sample size, the use of the AICc is recommended (Burnham & Anderson 2002). Here we use an approximation to the multivariate AICc as defined by Fujikoshi & Satoh (1997), which includes a recommended correction for over- and under-defined models:

$$
\begin{aligned}
\text{AICc} &= \text{AIC} + 2\frac{p(p+v)}{(nN_T - p - v)} \\
&= -2\ell + 2p + 2\frac{p(p+v)}{(nN_T - p - v)}.
\end{aligned}
\tag{3.1}
$$

The maximized log-likelihood is given by ℓ, p is the number of parameters estimated (including the intercepts and residual trait variances), and v is the num-

ber of distinct parameters estimated in the covariance matrix, where $1 \leq v \leq N_T(N_T+1)/2$ (Fujikoshi & Satoh 1997, Burnham & Anderson 2002). We have n independent individuals, with N_T non-independent observations per individual.

Within a set of candidate models, the model which best fits the data is associated with the best score, here this is the lowest AICc value (Zhou 2011). However, when using the AIC there is a rule of thumb whereby two models with AIC values with a difference less than 2 can be considered indistinguishable. The AICc converges to the AIC with large n, and so we apply this rule here — considering models of a similar fit instead of simply selecting the model with the lowest AICc value.

This chapter considers two score-based approaches to estimate the unknown phenotype network structure. The first approach fits and compares all possible causal network structures and selects the model with the best fit as determined by the AICc. The second approach sequentially builds up the model using a standard heuristic search strategy — the greedy hill-climbing search strategy (see, for example, Heckerman et al. 1995 and Tsamardinos et al. 2002). With this search strategy, at each step we have a current model structure which we compare to a set of candidate models, each one formed by either the addition of a directed edge (i.e. one trait-to-trait effect) not included in the current model, or the removal or reversal of a directed edge within the current model. The model with the greatest improvement in the score is then selected to become the current model (Friedman et al. 1999). The search continues until the AICc of the current model can no longer be improved.

This second approach is necessary for larger datasets comprised of many quantitative traits. Robinson (1977) introduced the following recursive function to de-

termine the number of existing (acyclic) models for N_T traits:

$$f(N_T) = \sum_{i=1}^{N_T} (-1)^{i+1} \binom{N_T}{i} 2^{i(N_T - i)} f(N_T - i) \tag{3.2}$$

where $f(0) = 1$. As the number of traits increases the number of possible models increases rapidly and so the first approach, where we assign a score to each possible model, becomes computationally infeasible. For example, this function states that there exist 25 possible acyclic models for three quantitative traits, and $29,281$ for five quantitative traits.

The two score-based approaches to estimating an unknown phenotype network structure are described in Sections 3.2 and 3.3, along with simulated examples to illustrate each approach. Both use the approximation to the multivariate AICc as given in Equation 3.1 (p. 34), where the least squares estimates maximise the log likelihood as detailed in Section 3.1.

3.1 The maximised log likelihood

The notation used in this chapter is continued throughout the remainder of this thesis, and is summarized in Table 3.1 for convenience. Let the causal network structure be denoted $\mathbf{S}$, comprised of a set of nodes connected by a set of directed edges. Let the quantitative trait measurements be denoted by the $n \times N_T$ matrix $\mathbf{Y}$. The QTL genotypes ($\mathbf{Q}$), QTL locations (Λ), marker genotypes ($\mathbf{G}$) and recombination fractions between markers ($\mathbf{r}$) are assumed to be known, and are as described in Table 3.1. Let Ψ denote the parameters assumed to be known, such that $\Psi = \{\mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}\}$. The unknown parameters are denoted by Ω:

$$\Omega = \{\beta_{01}, \cdots, \beta_{0N_T}, \beta_{\upsilon_Y(1)1}, \cdots, \beta_{\upsilon_Y(N_T)N_T}, \phi_{\upsilon_Q(1)1}, \cdots, \phi_{\upsilon_Q(N_T)N_T}, \sigma_1^2, \cdots, \sigma_{N_T}^2\}.$$

Table 3.1: *A summary of the notation used, as well as an indication of whether or not each is included in Ψ or Ω.*

Notation Used		Description	
n		number of individuals	
N_T		number of quantitative traits	
N_{qtl}		number of QTL	
N_ϕ		number of QTL effects	
$\mathbf{Y}$	$\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_t, \cdots, \mathbf{y}_{N_T}]$	quantitative trait measurements	
	where $\mathbf{y}_t = [y_{i1}, \cdots, y_{it}, \cdots, y_{it}]^\top$		
	$\upsilon_Y(t)$	the set of quantitative traits with a direct effect on trait t	
	$|\upsilon_Y(t)|$	the number of quantitative traits with a direct effect on trait t	
	$\mathbf{y}_{i\upsilon_Y(t)}$	the set of trait values for the ith individual, for those traits with a direct effect on trait t	
$\mathbf{G}$	$\mathbf{G} = \{\mathbf{g}_i\}_{i=1}^n$	marker genotypes	(Ψ)
	where $\mathbf{g}_i = \{g_{km}^{(i)} : (k=1,\cdots,K), (m=1,\cdots,N_M^{(k)})\}$	$g_{km} \in \{0,1\}$ for a doubled haploid population, for example	
		for K chromosomes and $N_M^{(k)}$ markers on chromosome k	
$\mathbf{r}$	$\mathbf{r} = \{r_{km} : (k=1,\cdots,K), (m=1,\cdots,N_M-1)\}$	recombination fractions between markers	(Ψ)
		r_{km} is the recombination fraction between markers	
		m and $m+1$ on chromosome k	
$\mathbf{Q}$	$\mathbf{Q} = \{\mathbf{q}_1, \cdots, \mathbf{q}_\ell, \cdots, \mathbf{q}_{N_{qtl}}\}$	QTL genotypes	(Ψ)
	where $\mathbf{q}_\ell = \{q_{i\ell}\}_{i=1}^n$	$q_{i\ell} \in \{0,1\}$ for a doubled haploid population, for example	
	for $\ell = 1, \cdots, N_{qtl}$	for the N_{qtl} QTL	
	$\upsilon_Q(t)$	the set of QTL with a direct effect on trait t	
	$|\upsilon_Q(t)|$	the number of QTL with direct effects on trait t	
	$\mathbf{q}_{i\upsilon_Y(t)}$	the set of QTL genotypes for the ith individual, for those QTL with a direct effect on trait t	
Λ	$\Lambda = \{\lambda_1, \cdots, \lambda_\ell, \cdots, \lambda_{N_{qtl}}\}$ where $\lambda_\ell = (\xi_\ell, \gamma_\ell, \theta_\ell)$	genomic location of the ℓth QTL	(Ψ)
	ξ_ℓ	chromosome on which the ℓth QTL is located	
	γ_ℓ	left flanking marker of the ℓth QTL	
	θ_ℓ	recombination fraction between γ_ℓ and the ℓth QTL	
Φ	$\Phi = \{\Phi_{\upsilon_Q(1)1}, \cdots, \Phi_{\upsilon_Q(t)t}, \cdots, \Phi_{\upsilon_Q(N_T)N_T}\}$	QTL effects	(Ω)
	$\phi_{\upsilon_Q(t)t}$	the set of $|\upsilon_Q(t)|$ QTL effects the parent QTL have on trait t	
β_{0t}	$\{\beta_{01}, \cdots, \beta_{0N_T}\}$	trait intercepts for trait $t = 1, \cdots, N_T$	(Ω)
$\beta_{\upsilon_Y(t)t}$	for $t = 1, \cdots, N_T$	the set of $|\upsilon_Y(t)|$ trait-to-trait effects the parent traits have on trait t	(Ω)
	$\beta_{t(t+1)}$	the trait-to-trait effect of trait t on trait $(t+1)$	
σ_t^2	for $t = 1, \cdots, N_T$	residual variance for trait t	(Ω)

For traits $t = 1, \cdots, N_T$, β_{0t} is the intercept for trait t, and $\upsilon_Y(t)$ and $\upsilon_Q(t)$ denote those traits and QTL directly upstream of trait $\mathbf{y}_t$, respectively. The terms $\beta_{\upsilon_Y(t)t}$ and $\phi_{\upsilon_Q(t)t}$ then denote the direct effect of the traits $\upsilon_Y(t)$ and QTL $\upsilon_Q(t)$ on trait t, respectively. Together the graph structure and the parameters define the model M; $M = \{\Omega, \mathbf{S}\}$. Given data $\mathbf{Y}$ and Ψ, the posterior distribution for model M is:

$$
\begin{aligned}
p(M|\mathbf{Y}, \Psi) &= \frac{p(\mathbf{Y}|M, \Psi)p(M|\Psi)}{p(\mathbf{Y}|\Psi)} \\
&\propto p(\mathbf{Y}|\Omega, \mathbf{S}, \Psi)p(\Omega, \mathbf{S}|\Psi) \\
&\propto p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S}). \quad (3.3)
\end{aligned}
$$

QTL can have both direct and indirect effects on quantitative traits. The genetic model is:

$$
y_{it} = \beta_{0t} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} + \varepsilon_{it} \quad (3.4)
$$

For the tth quantitative trait of the ith individual, denoted y_{it}:

$$
y_{it} = \mu_{it} + \varepsilon_{it}
$$

where
$$
\varepsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_t^2)
$$

$$
\mu_{it} = \beta_{0t} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell}.
$$

The log likelihood is:

$$
\begin{aligned}
&\ell(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) \\
&= \sum_{i=1}^{n} \log \left(\prod_{\ell=1}^{N_{qtl}} \left[p(q_{i\ell}|\lambda_\ell, \mathbf{G}, \mathbf{r})p(\lambda_\ell|\mathbf{G}, \mathbf{r}) \right] \prod_{t=1}^{N_T} p(y_{it}|\mathbf{y}_{i\upsilon_Y(t)}, \mathbf{q}_{i\upsilon_Q(t)}, \Omega) \right) \\
&= c + \sum_{i=1}^{n} \log \left(\prod_{t=1}^{N_T} p(y_{it}|\mathbf{y}_{i\upsilon_Y(t)}, \mathbf{q}_{i\upsilon_Q(t)}, \Omega) \right) \quad (3.5)
\end{aligned}
$$

where $\mathbf{y}_{i\upsilon_Y(t)}$ denotes the set of observed trait values for the parent traits of the tth trait, and $\mathbf{q}_{i\upsilon_Q(t)}$ denotes the set of observed QTL genotypes for the parent QTL of the tth trait.

The probability of the ℓth QTL genotype for individual i having the value $q_{i\ell}$ is denoted by $p(q_{i\ell}|\lambda_\ell, \mathbf{G}, \mathbf{r})$ and is conditional on the location of the QTL (defined by λ_ℓ; the chromosome it is located on, the left flanking marker and the recombination fraction), the marker genotypes ($\mathbf{G}$) and marker locations ($\mathbf{r}$). Similarly, $p(\lambda_\ell|\mathbf{G}, \mathbf{r})$ is the probability of the ℓth QTL being located at the genomic position indicated by λ_ℓ. Therefore, $\prod_{\ell=1}^{N_{qtl}}[p(q_{i\ell}|\lambda_\ell, \mathbf{G}, \mathbf{r})p(\lambda_\ell|\mathbf{G}, \mathbf{r})]$ denotes the probability of observing the given set of N_{qtl} QTL for individual i. Here, we assume the QTL locations and genotypes are known, giving $\prod_{\ell=1}^{N_{qtl}}[p(q_{i\ell}|\lambda_\ell, \mathbf{G}, \mathbf{r})p(\lambda_\ell|\mathbf{G}, \mathbf{r})]$ a single fixed value, denoted by the constant c.

Section 3.1.1 outlines the calculation of the least squares estimates (LSE), and demonstrates that in this scenario they are equal to the maximum likelihood estimates (MLE).

3.1.1 Least squares estimates

All traits are modeled using additive linear models, and we have N_T separate linear regressions, one for each trait. The quantitative trait measurements for the tth trait are recorded in the $n \times 1$ column vector, $\mathbf{y}_t = [y_{1t}, \cdots, y_{nt}]^\top$. Let the $n \times (N_{\upsilon_Y(t)} + N_{\upsilon_Q(t)} + 1)$ design matrix be denoted $\mathbf{X}^{(t)}$ where there are $N_{\upsilon_Y(t)}$ traits and $N_{\upsilon_Q(t)}$ QTL affecting trait t. It follows that the $(N_{\upsilon_Y(t)} + N_{\upsilon_Q(t)} + 1) \times 1$ parameter vector $\underline{\beta}^{(t)}$ is comprised of the intercept, β_{0t}, the $N_{\upsilon_Y(t)}$ trait-to-trait effects ($\{\beta_{kt}\}_{k \in \upsilon_Y(t)}$) and the $N_{\upsilon_Q(t)}$ QTL effects ($\{\phi_{\ell t}\}_{\ell \in \upsilon_Q(t)}$). We assume that the genetic architecture is known *a priori* and so the number of QTL with a direct effect on trait t, $N_{\upsilon_Q(t)}$,

is known. It follows that:

$$\mathbf{y}_t = \mathbf{X}^{(t)}\underline{\beta}^{(t)} + \underline{\varepsilon}_t, \qquad \underline{\varepsilon}_t \sim N(0, \sigma_t^2)I_n. \tag{3.6}$$

Given that the quantitative trait values are normally distributed with variance σ_t^2, the probability density function is:

$$p(\mathbf{y}_t | \mathbf{X}\underline{\beta}^{(t)}, \sigma_t^2) = \left(\frac{1}{(2\pi\sigma_t^2)} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_t^2}(\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\beta}^{(t)})^\top (\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\beta}^{(t)}) \right) \tag{3.7}$$

and the log-likelihood:

$$\ell(\mathbf{y}_t | \mathbf{X}\underline{\beta}^{(t)}, \sigma_t^2) = -\frac{n}{2}\log(2\pi\sigma_t^2) - \frac{1}{2\sigma_t^2}(\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\beta}^{(t)})^\top (\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\beta}^{(t)}). \tag{3.8}$$

In order to find the least squares parameter estimates we need to minimise the SSE (sum of squared errors):

$$(\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\beta}^{(t)})^\top (\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\beta}^{(t)}).$$

It follows that by minimising the SSE we maximise the log likelihood, and we conclude that the least squares estimates are equal to the maximum likelihood estimates. These terms are therefore used interchangeably in this thesis.

The least squares estimates (LSE) are obtained by partially differentiating the SSE with respect to the parameter vector $\underline{\beta}^{(t)}$, and setting equal to zero to get:

$$\widehat{\underline{\beta}}^{(t)} = \left(\mathbf{X}^{(t)\top}\mathbf{X}^{(t)} \right)^{-1} \mathbf{X}^{(t)\top}\mathbf{y}_t. \tag{3.9}$$

We have not assumed equal variances across traits, so the variance of trait t is:

$$\hat{\sigma}_t^2 = \frac{1}{n - p^{(t)}} \left(\mathbf{y}_t - \mathbf{X}^{(t)}\widehat{\underline{\beta}}^{(t)} \right)^\top \left(\mathbf{y}_t - \mathbf{X}^{(t)}\underline{\widehat{\beta}}^{(t)} \right), \tag{3.10}$$

where $p^{(t)}$ is the number of parameters estimated including the intercept, $p^{(t)} = N_{v_Y(t)} + N_{v_Q(t)} + 1$. Given these individual trait variances, the variances of the parameter estimates relating to trait t, $\widehat{\underline{\beta}}^{(t)}$, is:

$$\widehat{Var}(\widehat{\underline{\beta}}^{(t)}) = \hat{\sigma}_t^2 \left(\mathbf{X}^{(t)\top} \mathbf{X}^{(t)} \right)^{-1}. \tag{3.11}$$

3.2 Approach 1: consider all possible causal models

Assuming an unknown phenotype network structure, one approach would be to compare all possible models (each representing a different phenotype network structure) and select the best model given the data, i.e. the model with the lowest AICc value. Furthermore, we may obtain model weights so as to compare the fit of each model relative to the model of best fit — normalizing the model likelihoods so that we can consider them as probabilities (Burnham & Anderson 2002, Anderson 2008). This approach is straightforward when only a small number of quantitative traits are included in the study; however, as this number increases the number of possible models increases rapidly (see Equation 3.2 from Robinson 1977 on p. 36) and it is no longer computationally feasible to fit all possible phenotype networks existing between N_T traits.

3.2.1 Summary of Approach 1

1. Determine the set of M candidate models for a given number of traits, N_T.

2. Calculate the LSE for each candidate model using Equations 3.9 and 3.10.

3. Calculate the AICc as given in Equation 3.1 (p. 34) for each model in the candidate set; denoted $AICc_m$ for the mth candidate model ($m = 1, \cdots, M$).

4. Identify the model with the lowest AICc, AICc$_{min}$, and calculate the difference between the AICc of each mth model and AICc$_{min}$;

$$\triangle_m = \text{AICc}_m - \text{AICc}_{min}. \tag{3.12}$$

5. Calculate the Akaike weight for all candidate models:

$$p(m) = \frac{\exp(-\frac{1}{2}\triangle_m)}{\sum\limits_{i=1}^{M} \exp(-\frac{1}{2}\triangle_i)} \tag{3.13}$$

for $m = 1, \cdots, M$ (Burnham & Anderson 2002).

The model with the lowest AICc provides the best fit to the data, but by using the Akaike weights we also have a measure of how different the fit of each model is compared to the best model. In other words, the Akaike weights are used to determine the relative standing of the candidate models regardless of the AICc scale (Burnham & Anderson 2002).

3.2.2   Example with three quantitative traits

Consider an example where there exist three quantitative traits and three QTL, simulated according to Appendix B and the causal network structure given in Figure 3.1. The true causal network structure corresponds to Model 6 in Figure 3.2, which displays all possible phenotype networks for $N_T = 3$ quantitative traits. For each trait t ($t = 1, \cdots, N_T$), $\upsilon_Q(t)$ is assumed known, but the size of the QTL effects are to be estimated, along with the phenotype network structure.

Each of these candidate models were individually fitted to the simulated data to obtain the LSE of each parameter, and the respective multivariate AICc values

(determined using Equation 3.1, p. 34) are given in Table 3.2 and plotted in Figure 3.3 (p. 45).
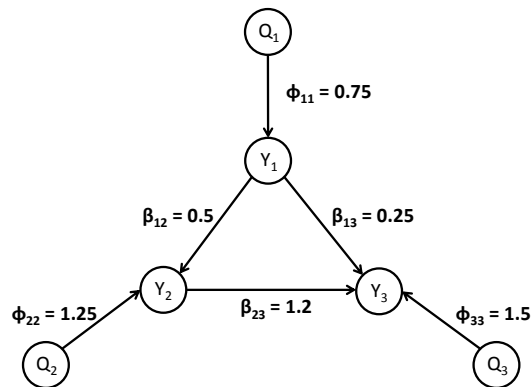
Figure 3.1: *The true causal network structure and genetic architecture.*
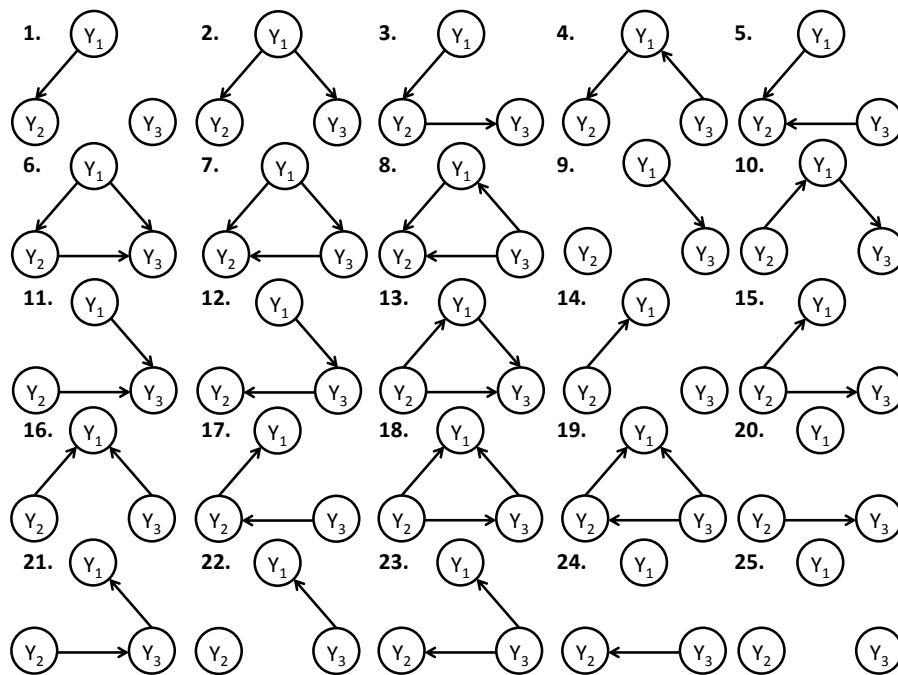
Figure 3.2: *There exist 25 candidate phenotype network structures between $N_T = 3$ quantitative traits. Model 6 is the phenotype network structure of the true model.*

Table 3.2: *The multivariate AICc values, change in AICc ($\triangle_m = AICc_m - AICc_{min}$), and the Akaike weights for all 25 models in the candidate set, indexed as in Figure 3.2.*

model number m	AICc	$\triangle_m$	Akaike weight $p(m)$
1	2280.77	347.00	0.00
2	2212.86	279.09	0.00
3	1939.47	5.71	0.05
4	2247.00	313.23	0.00
5	2134.78	201.01	0.00
6	1933.77	0.00	0.95
7	2066.87	133.10	0.00
8	2101.02	167.25	0.00
9	2277.50	343.73	0.00
10	2245.79	312.03	0.00
11	1998.41	64.64	0.00
12	2073.76	139.99	0.00
13	1966.71	32.94	0.00
14	2313.71	379.94	0.00
15	1972.41	38.64	0.00
16	2311.12	377.35	0.00
17	2109.97	176.20	0.00
18	1969.83	36.06	0.00
19	2107.38	173.61	0.00
20	2004.12	70.35	0.00
21	1970.35	36.58	0.00
22	2311.65	377.88	0.00
23	2107.91	174.14	0.00
24	2141.68	207.91	0.00
25	2367.60	433.83	0.00
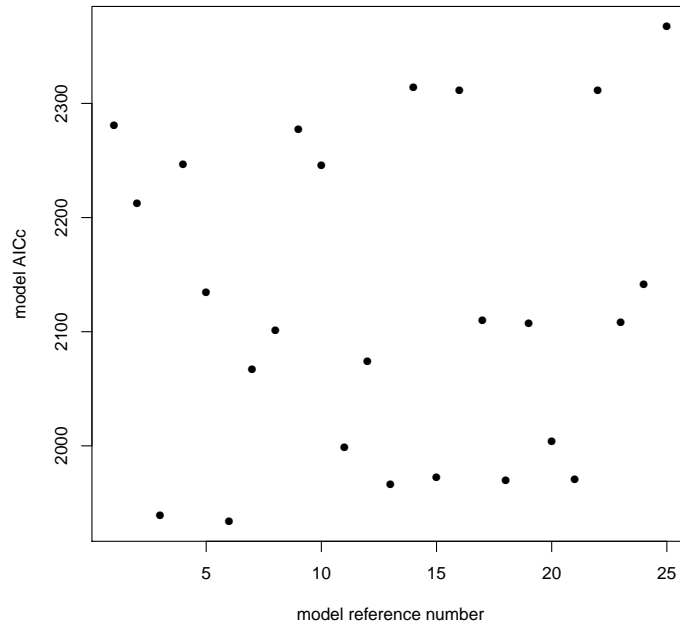
Figure 3.3: *The multivariate AICc values for all 25 candidate models incorporating three quantitative traits. Each model number corresponds to the models illustrated in Figure 3.2.*
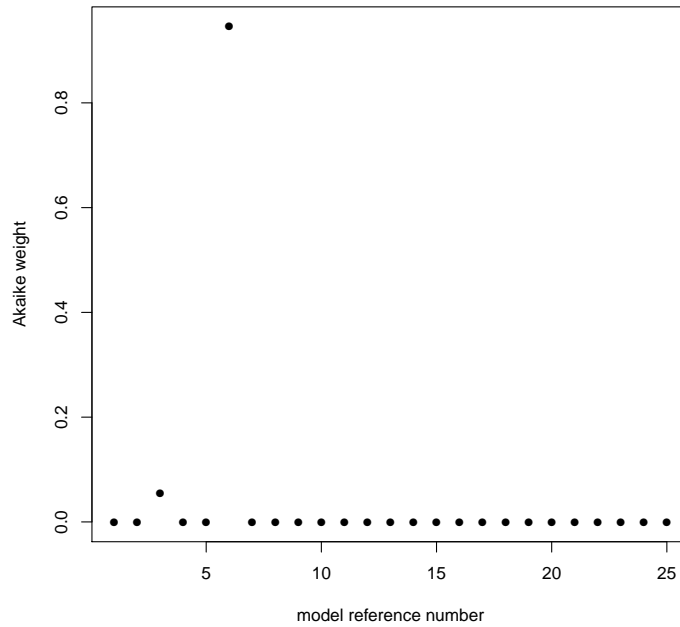
Figure 3.4: *The Akaike weights for all 25 candidate models incorporating three quantitative traits. Each model number corresponds to the models illustrated in Figure 3.2.*

Both Table 3.2 and Figure 3.3 show that Model 6 (the true model) has the lowest AICc value, indicating that this is the model of best fit. The Akaike weights/model probabilities are given in Table 3.2 and displayed in Figure 3.4. Model 6 has an Akaike weight of 0.95, significantly higher than the other possible models. Model 3 has a phenotype network structure similar to Model 6, although it is missing the (weakest) trait-to-trait effect β_{13}. Model 3 is the model with the only other Akaike weight where $p(m) > 0.005$. The addition of the trait-to-trait effect β_{13} reduced the AICc by more than 2, so we conclude that Model 6 is the correct phenotype network structure, given the data. Parameter estimates for Model 6 are given in Table 3.3. The least squares parameter estimates are close to the true values, with narrow confidence intervals each containing the true parameter value. This approach of fitting and comparing all possible models is able to both identify the correct phenotype network structure and correctly estimate the

Table 3.3: *Parameter estimates for the true causal network structure of Model 6, as given in Figure 3.2 (p. 43). Intervals including the true value are indicated by an asterisk (*).*

parameter	true value	estimate	95% confidence interval
β_{01}	0.50	0.41	$(0.21, 0.61)^*$
β_{02}	0.50	0.34	$(0.08, 0.60)^*$
β_{03}	0.50	0.55	$(0.28, 0.83)^*$
β_{12}	0.50	0.61	$(0.47, 0.74)^*$
β_{13}	0.25	0.21	$(0.06, 0.36)^*$
β_{23}	1.20	1.24	$(1.13, 1.36)^*$
ϕ_{11}	0.75	0.87	$(0.60, 1.14)^*$
ϕ_{22}	1.25	1.30	$(1.00, 1.59)^*$
ϕ_{33}	1.50	1.27	$(0.99, 1.56)^*$
σ_1	1.00	0.97	$(0.80, 1.19)^*$
σ_2	1.00	1.06	$(0.88, 1.30)^*$
σ_3	1.00	1.03	$(0.85, 1.26)^*$

true parameter values. However, once the number of quantitative traits included in the study increases, the number of models increases rapidly, creating the need for a search algorithm to efficiently search the model space.

3.3 Approach 2: greedy hill-climbing search strategy

As the number of quantitative traits increases, the number of possible causal network structures increases rapidly (Robinson 1977) and so it is not always feasible to fit all possible models. One way to approach this problem is to build the candidate phenotype network structure one causal relationship at a time via a greedy hill-climbing search strategy (see, for example, Heckerman et al. 1995 and Tsamardinos et al. 2002). In order to create a flexible search strategy, we assume we can add any directed edge to the current model, provided it does not create a cyclic model. This allows for the estimation of causal network structures that are disconnected — that is, models where if the UDG is considered, there exist two nodes such that there is no path connecting the two, see the example in Figure 3.5. If many quantitative traits are included in a study, it is possible that a small subset are not causally related to the other traits. If we conditioned the search method to only add directed edges (or causal relationships) connecting to the current model, we would not be able to identify models like that given in Figure 3.5. Furthermore, as this is a hill-climbing algorithm, we randomly select multiple initial models to repeat the search from different areas of the model space in an effort to avoid local maxima. The approach is summarized in Section 3.3.1 and we provide simulated examples for two different causal network structures in Sections 3.3.2 and 3.3.3.
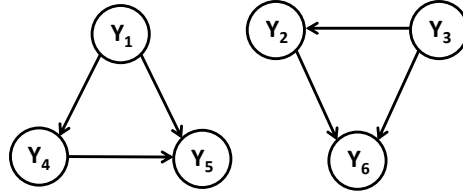
Figure 3.5: *An example of a disconnected phenotype network structure. There is no undirected path from Y_1 to Y_2, for example.*

3.3.1   Summary of Approach 2

1. Set five initial models, each denoted M_0; one with no trait-to-trait effects, one with the maximum number of trait-to-trait effects, and 3 randomly selected initial models. For each initial model complete steps $2 - 6$.

2. Fit the initial model, M_0, by estimating the LSE as outlined in Section 3.1.1. Calculate the multivariate AICc given in Equation 3.1 (p. 34), this is now the current model.

3. Determine the set of *M* acyclic candidate models resulting from:

[a.] the addition of a single directed edge not in the current model,

[b.] the removal of a single directed edge in the current model,

[c.] the reversal of a single directed edge in the current model.

4. Obtain the LSE for all *m* candidate models ($m = 1, \cdots, M$) following Section 3.1.1 and calculate the multivariate AICc for each, $AICc_m$, given in Equation 3.1 (p. 34).

5. Select the model with the lowest AICc, *AICc$_{min}$*. This model is now the current model.

6. Continue steps 3 through 5 until the current model is the model with the lowest AICc value.

7. Compare the models estimated from each of the five searches. If the models differ, the one with the lowest AICc is the final model, unless the AICc values differ by less than 2 and then the models are comparable.

3.3.2   Approach 2, example 1: five connected nodes

Following from Equation 3.2 (p. 36), for five traits there exist 29,281 possible models, making the first approach of fitting all possible models computationally infeasible. Data for $n = 500$ individuals was simulated for five traits and five QTL as described in Appendix B, according to the true causal network structure given in Figure 3.6.
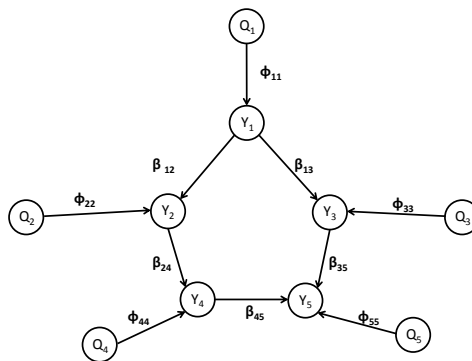
Figure 3.6: *The general form of the causal network structure; the true values of β_{12}, β_{13}, β_{24}, β_{35}, β_{45}, ϕ_{11}, ϕ_{22}, ϕ_{33} ϕ_{44} and ϕ_{55} are all varied according to the parameter sets given in Table 3.4.*

The true parameter values were varied in order to determine how the greedy hill-climbing search approach is influenced by different combinations of QTL and trait-to-trait effects. Both small and large QTL effects and trait-to-trait effects are selected with values between 0.1 and 1.2. This range was based upon those values in published data sets; for example, Dhungana et al. (2007) and Mi et al. (2010). We note that it is possible for the effects to be positive or negative. The true parameter values for sets 3.1 – 3.4 are listed in Table 3.4 (p. 51).

Parameter set 3.1: small QTL effects and small trait-to-trait effects,

Parameter set 3.2: small QTL effects and large trait-to-trait effects,

Parameter set 3.3: large QTL effects and small trait-to-trait effects,

Parameter set 3.3: large QTL effects and large trait-to-trait effects.

For each parameter set we simulate one initial data set and present the marginal posterior probabilities for each directed edge, as well as the model joint posterior probabilities and model averaged parameter estimates. We then simulate a further 29 data sets for each parameter set in order to determine the consistency of this approach. The parameter sets are labelled 3.1 – 3.4 to indicate that they relate to this chapter, Chapter 3, and the jth data set for parameter set i is referred to as 3.i.j. The simulation results from these four different parameter sets will give an indication of how this approach would perform in real QTL mapping scenarios.

The resulting LSE for the simulated data sets 3.1.1 – 3.4.1 are given in Tables 3.6 – 3.9 at the end of this section (pp. 54 – 57). We note that the true causal network structure has been identified for all combinations of effect sizes. All least squares estimates are close to the true values with narrow confidence intervals, so it appears as though this score-based hill-climbing approach works well for the

Table 3.4: *True parameter values for the parameter sets, 3.1 – 3.4, each with a causal network structure corresponding to that given in Figure 3.6 (p. 49).*

parameter	parameter set			
	3.1	3.2	3.3	3.4
β_{01}	0.5	0.5	0.5	0.5
β_{02}	0.5	0.5	0.5	0.5
β_{03}	0.5	0.5	0.5	0.5
β_{04}	0.5	0.5	0.5	0.5
β_{05}	0.5	0.5	0.5	0.5
β_{12}	0.2	1.1	0.2	1.1
β_{13}	0.2	1.1	0.2	1.1
β_{24}	0.2	1.1	0.2	1.1
β_{35}	0.2	1.1	0.2	1.1
β_{45}	0.2	1.1	0.2	1.1
ϕ_{11}	0.1	0.1	1.2	1.2
ϕ_{22}	0.1	0.1	1.2	1.2
ϕ_{33}	0.1	0.1	1.2	1.2
ϕ_{44}	0.1	0.1	1.2	1.2
ϕ_{55}	0.1	0.1	1.2	1.2
σ_1^2	1.0	1.0	1.0	1.0
σ_2^2	1.0	1.0	1.0	1.0
σ_3^2	1.0	1.0	1.0	1.0
σ_4^2	1.0	1.0	1.0	1.0
σ_5^2	1.0	1.0	1.0	1.0

range of values simulated. Some of the confidence intervals for the QTL effects in data sets 3.1.1 and 3.2.1 do contain 0, although this is due to the true value of the QTL effects being so close to zero.

To comment on the consistency of this approach, for each parameter set in Table 3.4 we simulate a further 29 data sets according to Appendix B and note the number of times the approach correctly identifies the true causal network structure. Table 3.5 gives the number of different models found across the 30 different data sets for each parameter set. We see that all combinations of QTL and trait-to-trait effects identify numerous phenotype network structures, with those parameter sets comprised of small trait-to-trait effects (parameter sets 3.1 and 3.3 in Table 3.4) estimating a wider range of causal network structures. With small QTL effects and small trait-to-trait effects (parameter set 3.1) the approach had trouble identifying the direction of causality, especially between traits Y_1 and Y_2, and Y_1 and Y_3, where the reversed edges β_{21} and β_{31} were estimated about in about 40% of the estimated phenotype networks. On the other hand, parameter sets 3.3 and 3.4, with large QTL effects, resulted in the identification of a smaller range of models. Parameter set 3.4, with large QTL effects and large trait-to-trait effects, produced the most consistent results, with 53% of the estimated phenotype network structures corresponding to the true model.

The edges comprising the true model (β_{12}, β_{13}, β_{24}, β_{35}, β_{45}) are found in at least 60% of the estimated models (regardless of the true parameter values), although additional edges were also present. If we look at the extra edges that are included we see that many of these are simply indirect effects, e.g. β_{14}, β_{15} and β_{25}. This overfitting suggests that a penalty is required in order to estimate the correct causal network structure, and the penalty applied by the AICc is not always large enough.

Table 3.5: *For each of the parameter sets 3.1 – 3.4 (given in Table 3.4), we simulated 30 data sets, recording the number of different models estimated, the number of times the true model was estimated, and the number (n) and proportion (p) of times each directed edge was included in the estimated causal network structure.*

parameter set	3.1		3.2		3.3		3.4	
number of different models	23		13		20		9	
number of true models	3		12		6		16	
	(0.10)		(0.40)		(0.20)		(0.53)	
trait-to-trait effect	n	(p)	n	(p)	n	(p)	n	(p)
β_{12}	18	(0.60)	30	(1.00)	27	(0.90)	30	(1.00)
β_{13}	18	(0.60)	30	(1.00)	29	(0.97)	30	(1.00)
β_{14}	0	(0.00)	6	(0.20)	10	(0.33)	2	(0.07)
β_{15}	0	(0.00)	4	(0.13)	6	(0.20)	5	(0.17)
β_{21}	12	(0.40)	0	(0.00)	3	(0.10)	0	(0.00)
β_{23}	0	(0.00)	3	(0.10)	3	(0.10)	2	(0.07)
β_{24}	19	(0.63)	30	(1.00)	29	(0.97)	30	(1.00)
β_{25}	3	(0.10)	4	(0.13)	7	(0.23)	3	(0.10)
β_{31}	12	(0.40)	0	(0.00)	1	(0.03)	0	(0.00)
β_{32}	3	(0.10)	2	(0.07)	2	(0.07)	3	(0.10)
β_{34}	2	(0.07)	2	(0.07)	5	(0.17)	0	(0.00)
β_{35}	22	(0.73)	30	(1.00)	28	(0.93)	30	(1.00)
β_{41}	3	(0.10)	0	(0.00)	1	(0.03)	0	(0.00)
β_{42}	11	(0.37)	0	(0.00)	1	(0.03)	0	(0.00)
β_{43}	2	(0.07)	6	(0.20)	5	(0.17)	3	(0.10)
β_{45}	19	(0.63)	30	(1.00)	29	(0.97)	30	(1.00)
β_{51}	3	(0.10)	0	(0.00)	0	(0.00)	0	(0.00)
β_{52}	3	(0.10)	0	(0.00)	0	(0.00)	0	(0.00)
β_{53}	8	(0.27)	0	(0.00)	2	(0.07)	0	(0.00)
β_{54}	11	(0.37)	0	(0.00)	1	(0.03)	0	(0.00)

Table 3.6: *The least squares estimates for data set 3.1.1. Intervals including the true parameter value are indicated by an asterisk (*).*

parameter	true value	estimate	95% confidence interval
β_{01}	0.5	0.53	$(0.41, 0.66)^*$
β_{02}	0.5	0.53	$(0.40, 0.66)^*$
β_{03}	0.5	0.36	$(0.21, 0.51)^*$
β_{04}	0.5	0.55	$(0.41, 0.69)^*$
β_{05}	0.5	0.41	$(0.26, 0.56)^*$
β_{12}	0.2	0.15	$(0.06, 0.23)^*$
β_{13}	0.2	0.21	$(0.12, 0.31)^*$
β_{24}	0.2	0.16	$(0.07, 0.26)^*$
β_{35}	0.2	0.19	$(0.11, 0.27)^*$
β_{45}	0.2	0.20	$(0.12, 0.29)^*$
ϕ_{11}	0.1	0.23	$(0.06, 0.41)^*$
ϕ_{22}	0.1	0.21	$(0.05, 0.38)^*$
ϕ_{33}	0.1	0.37	$(0.18, 0.56)$
ϕ_{44}	0.1	0.00	$(-0.17, 0.18)^*$
ϕ_{55}	0.1	0.30	$(0.12, 0.47)$
σ_1^2	1.0	0.99	$(0.88, 1.12)^*$
σ_2^2	1.0	0.94	$(0.83, 1.07)^*$
σ_3^2	1.0	1.06	$(0.94, 1.21)^*$
σ_4^2	1.0	1.01	$(0.90, 1.15)^*$
σ_5^2	1.0	0.98	$(0.87, 1.12)^*$

Table 3.7: *The least squares estimates for data set 3.2.1. Intervals including the true parameter value are indicated by an asterisk (*).*

parameter	true value	estimate	95% confidence interval
β_{01}	0.5	0.61	$(0.49, 0.74)^*$
β_{02}	0.5	0.50	$(0.36, 0.65)^*$
β_{03}	0.5	0.49	$(0.35, 0.64)^*$
β_{04}	0.5	0.53	$(0.39, 0.67)^*$
β_{05}	0.5	0.59	$(0.43, 0.74)^*$
β_{12}	1.1	1.10	$(1.01, 1.19)^*$
β_{13}	1.1	1.08	$(0.99, 1.17)^*$
β_{24}	1.1	1.12	$(1.06, 1.17)^*$
β_{35}	1.1	1.11	$(1.05, 1.18)^*$
β_{45}	1.1	1.08	$(1.03, 1.13)^*$
ϕ_{11}	0.1	0.15	$(-0.03, 0.32)^*$
ϕ_{22}	0.1	0.20	$(0.02, 0.39)^*$
ϕ_{33}	0.1	0.21	$(0.03, 0.40)^*$
ϕ_{44}	0.1	0.12	$(-0.06, 0.29)^*$
ϕ_{55}	0.1	0.08	$(-0.10, 0.25)^*$
σ_1^2	1.0	0.99	$(0.88, 1.13)^*$
σ_2^2	1.0	1.03	$(0.91, 1.17)^*$
σ_3^2	1.0	1.03	$(0.92, 1.18)^*$
σ_4^2	1.0	1.00	$(0.88, 1.13)^*$
σ_5^2	1.0	0.98	$(0.87, 1.12)^*$

Table 3.8: *The least squares estimates for data set 3.3.1. Intervals including the true parameter value are indicated by an asterisk (*).*

parameter	true value	estimate	95% confidence interval
β_{01}	0.5	0.43	$(0.31, 0.55)^*$
β_{02}	0.5	0.45	$(0.30, 0.60)^*$
β_{03}	0.5	0.43	$(0.28, 0.58)^*$
β_{04}	0.5	0.68	$(0.52, 0.84)^*$
β_{05}	0.5	0.54	$(0.36, 0.73)^*$
β_{12}	0.2	0.20	$(0.12, 0.27)^*$
β_{13}	0.2	0.26	$(0.18, 0.33)^*$
β_{24}	0.2	0.14	$(0.07, 0.21)^*$
β_{35}	0.2	0.18	$(0.11, 0.25)^*$
β_{45}	0.2	0.15	$(0.07, 0.22)^*$
ϕ_{11}	1.2	1.31	$(1.14, 1.48)^*$
ϕ_{22}	1.2	1.20	$(1.02, 1.38)^*$
ϕ_{33}	1.2	1.21	$(1.03, 1.39)^*$
ϕ_{44}	1.2	1.02	$(0.84, 1.19)$
ϕ_{55}	1.2	1.36	$(1.19, 1.54)^*$
σ_1^2	1.0	0.97	$(0.86, 1.10)^*$
σ_2^2	1.0	1.00	$(0.89, 1.14)^*$
σ_3^2	1.0	1.00	$(0.89, 1.15)^*$
σ_4^2	1.0	0.98	$(0.87, 1.11)^*$
σ_5^2	1.0	0.98	$(0.87, 1.11)^*$

Table 3.9: *The least squares estimates for data set 3.4.1. Intervals including the true parameter value are indicated by an asterisk (*).*

parameter	true value	estimate	95% confidence interval
β_{01}	0.5	0.46	$(0.34, 0.59)^*$
β_{02}	0.5	0.51	$(0.36, 0.66)^*$
β_{03}	0.5	0.56	$(0.41, 0.70)^*$
β_{04}	0.5	0.57	$(0.41, 0.73)^*$
β_{05}	0.5	0.56	$(0.36, 0.76)^*$
β_{12}	1.1	1.05	$(0.98, 1.13)^*$
β_{13}	1.1	1.08	$(1.01, 1.15)^*$
β_{24}	1.1	1.04	$(0.99, 1.09)$
β_{35}	1.1	1.08	$(1.02, 1.14)^*$
β_{45}	1.1	1.10	$(1.05, 1.14)^*$
ϕ_{11}	1.2	1.24	$(1.06, 1.42)^*$
ϕ_{22}	1.2	1.26	$(1.08, 1.44)^*$
ϕ_{33}	1.2	1.19	$(1.02, 1.36)^*$
ϕ_{44}	1.2	1.27	$(1.11, 1.44)^*$
ϕ_{55}	1.2	1.18	$(1.01, 1.35)^*$
σ_1^2	1.0	1.03	$(0.92, 1.17)^*$
σ_2^2	1.0	1.03	$(0.91, 1.17)^*$
σ_3^2	1.0	0.96	$(0.85, 1.09)^*$
σ_4^2	1.0	0.95	$(0.84, 1.08)^*$
σ_5^2	1.0	0.99	$(0.88, 1.13)^*$

3.3.3   Approach 2, example 2: a disconnected graph structure

This section includes a small simulated example to illustrate how the greedy hill-climbing search strategy (summarized in Section 3.3.1) is able to identify a disconnected phenotype network structure. Data was simulated for five traits and five QTL as described in Appendix B, according to the true causal network structure given in Figure 3.7 (with a disconnected UDG).

   The least squares estimates from the first simulated data set are given in Table 3.10. All least squares estimates are close to the true values with narrow confidence intervals, so this score-based hill-climbing approach works well for our simulated disconnected phenotype network structure.
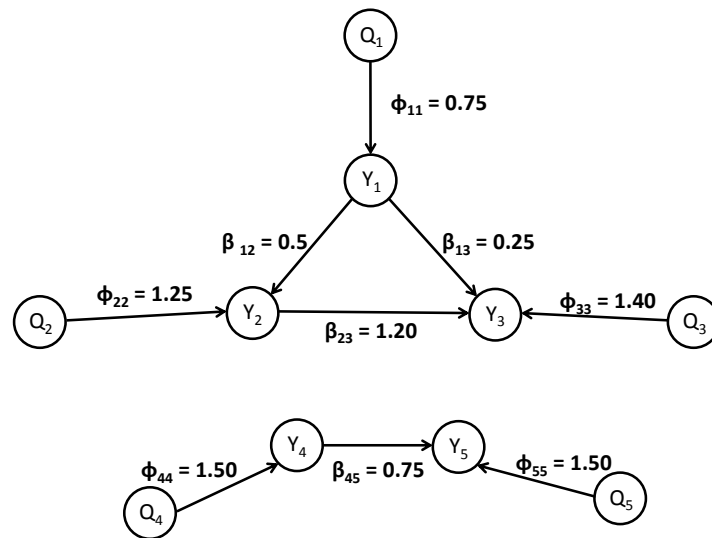
Figure 3.7: *The true causal network structure for five quantitative traits and five QTL; a disconnected graph structure used for an example of Approach 2.*

Table 3.10: *The least squares estimates for the first data set simulated according to Figure 3.7. Intervals including the true parameter value are indicated by an asterisk (*).*

parameter	true value	estimate	95% confidence interval
β_{01}	0.50	0.49	$(0.46, 0.71)$*
β_{02}	0.50	0.55	$(0.36, 0.63)$*
β_{03}	0.50	0.48	$(0.37, 0.73)$*
β_{04}	0.50	0.55	$(0.35, 0.61)$*
β_{05}	0.50	0.62	$(0.41, 0.69)$*
β_{12}	0.50	0.52	$(0.44, 0.60)$*
β_{13}	0.25	0.22	$(0.13, 0.31)$*
β_{23}	1.20	1.18	$(1.10, 1.25)$*
β_{45}	0.75	0.72	$(0.65, 0.78)$*
ϕ_{11}	0.75	0.62	$(0.44, 0.80)$*
ϕ_{22}	1.25	1.26	$(1.10, 1.42)$*
ϕ_{33}	1.40	1.46	$(1.29, 1.64)$*
ϕ_{44}	1.50	1.52	$(1.34, 1.70)$*
ϕ_{55}	1.50	1.52	$(1.35, 1.69)$*
σ_1^2	1.00	1.02	$(0.90, 1.16)$*
σ_2^2	1.00	0.92	$(0.82, 1.05)$*
σ_3^2	1.00	1.01	$(0.89, 1.14)$*
σ_4^2	1.00	1.02	$(0.90, 1.16)$*
σ_5^2	1.00	0.96	$(0.85, 1.09)$*

A further 29 data sets were simulated from the true causal network structure given in Figure 3.7 (30 data sets in total). Table 3.11 shows the consistency of this approach and we note that the true phenotype network structure was estimated fewer times than with a connected causal network structure. Here only 33% of the estimated models identified corresponded to the true model. Overfitting appears common in a disconnected structure — in fact, all of the additional edges, if individually added to the true causal network structure, would create a connected graph structure.

Table 3.11: *A total of 30 data sets were simulated according to the true parameter values given in Figure 3.7. We record the number of different models estimated, the number of times the true model was estimated, and the number (n) and proportion (p) of times each directed edge was included in the estimated causal network structure.*

number of different models	18	
number of true models	10	
	(0.33)	
trait-to-trait effect	n	(p)
β_{12}	30	(1.00)
β_{13}	30	(1.00)
β_{14}	2	(0.07)
β_{15}	3	(0.10)
β_{21}	0	(0.00)
β_{23}	30	(1.00)
β_{24}	1	(0.03)
β_{25}	4	(0.13)
β_{31}	0	(0.00)
β_{32}	0	(0.00)
β_{34}	3	(0.10)
β_{35}	3	(0.10)
β_{41}	3	(0.10)
β_{42}	1	(0.03)
β_{43}	5	(0.17)
β_{45}	30	(1.00)
β_{51}	2	(0.07)
β_{52}	4	(0.13)
β_{53}	4	(0.13)
β_{54}	0	(0.00)

In summary, given the genetic architecture, the (AICc) score-based greedy hill-climbing search algorithm was able to estimate the true causal network structure up to 53% of the time in a simulated example, provided the true causal network structure was connected. The presence of both large QTL effects and large trait-to-trait effects resulted in a higher number of overfitted models. With a disconnected causal network structure the method was not very consistent, estimating the true causal network structure in just 33% of the simulated data sets, often resulting in the estimation of many overfitted models. Overfitting would imply that the model has been influenced by the random error, rather than the effects (both trait-to-trait and QTL) comprising the causal network structure. This has resulted due to the complexity of the model fitted, with a large number of parameters (compared to n) and indicates that a Bayesian approach is preferred in which we will be able to construct priors which penalize models with complicated graph structures.

Chapter 4

Bayesian estimation of a phenotype network structure I

In Chapter 2 we described how the causal network structure can be represented by a graphical model, and in Chapter 3 we determined that an optimisation search strategy could be used to find an approximation to the optimum, because the optimum cannot be computed in an acceptable amount of time. Following Neto et al. (2010), we propose a Bayesian approach to identifying an unknown phenotype network structure using reversible jump Markov chain Monte Carlo. Here, in Chapter 4, we estimate the phenotype network structure with a fixed number of directed edges. This is not advisable in practise as it will result in poor mixing, as shown by a simulation study later in this chapter; however, restricting the number of edges does allow us to gradually introduce the move types used in our approach. We detail the update, reverse and relocate move types which keep the number of edges fixed, and in Chapter 5 we use these move types in addition to the add, remove and double steps which allow the number of edges in the model to vary. Together Chapters 4 and 5 form a Bayesian approach capable of estimating

an unknown phenotype network structure. It is important to note that this chapter still uses a reversible jump Markov chain Monte Carlo algorithm as even though the number of edges is the same in each model, the reverse and relocate move types are adding and removing parameters.

The differences between this approach and the QTLnet algorithm proposed by Neto et al. (2010) are summarized below.

- Neto et al. (2010) propose the joint inference of the phenotype network structure and the genetic components, whereas we assume that the genetic architecture is known and focus on the subproblem of estimating the unknown phenotype network structure and the QTL effects.

- Our approach makes use of LSE (equivalent to the maximum likelihood estimates) in the chain, obtained from separate regressions of each trait on the traits and QTL directly affecting it. This makes for a fast and efficient chain so we have an efficient sampler using well-tuned proposal distributions.

- Neto et al. (2010) experienced slow mixing of the Markov chain. This approach incorporates additional move types; here in Chapter 4 we detail the relocate step which relocates a directed edge, then in Chapter 5 we propose a double step to combine two steps that alter the phenotype network structure.

- We define the neighbourhood of a phenotype network structure as being conditional on a selected move type x; instead of including all possible candidate phenotype network structures given all move types, as in Neto et al. (2010). This makes it faster to obtain the neighbourhood for a given model when the number of traits is large.

- Neto et al. (2010) assume uniform priors over all network structures, whereas in Chapter 5 we introduce a prior on the number of directed edges in a model which causes simpler models, i.e. models with fewer directed edges, to be preferred.

The genetic components estimated by Neto et al. (2010) include the QTL locations, QTL genotypes and QTL-to-trait effects (also referred to as QTL effects). We, however, assume that the number of QTL, QTL locations and QTL genotypes are known, and that we have knowledge of which subset of QTL affect each trait (so we know which QTL effect parameters require estimation). We refer to this as the genetic architecture (see Section 1.2), and focus on estimating the unknown phenotype network structure and the QTL effects. We may assume that the genetic architecture is known as many methods have been established to map the genetic architecture of a trait (or multiple traits). Neto et al. (2010), for example, recommend the use of the seemingly unrelated regression model to estimate the genetic components; see Banerjee et al. (2008). It follows that our approach is a subproblem of Neto et al. (2010).

As described earlier, a causal network structure is defined by a set of nodes (the quantitative traits and QTL) connected by a set of directed edges, forming a directed acyclic graph (or DAG). We assume the genetic architecture is known, so that the objective of mapping the causal network structure is to estimate the phenotype network by selecting the set of directed edges which best describe the true causal network structure. So while we may refer to the estimation of the causal network structure, we are only estimating the phenotype network structure. Figure 1.2, included here for convenience, illustrates an example causal network structure with the genetic architecture highlighted in blue, and the phenotype network
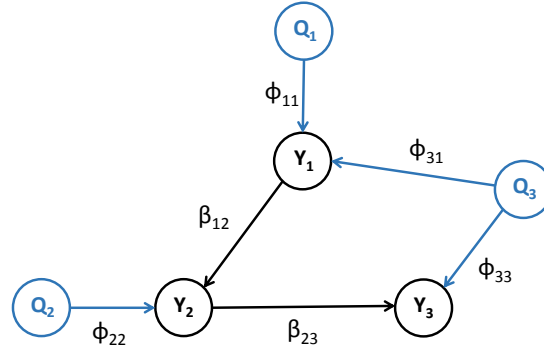
An example causal network structure with three quantitative traits (Y_1, Y_2, Y_3), three QTL (Q_1, Q_2, Q_3), two trait-to trait effects (β_{12}, β_{23}) and four QTL effects (ϕ_{11}, ϕ_{22}, ϕ_{31}, ϕ_{33}). The genetic architecture is assumed known and is highlighted in blue, the phenotype network structure and QTL effects are black and require estimation.

structure and QTL effects given in black. As we are keeping the number of edges fixed, at each step we propose one of the following three move types (or steps):

- **update** (x^U): update all parameters in the current model

- **reverse** (x^S): reverse the direction of one of the directed edges in the current phenotype network

- **relocate** (x^C): relocate one of the directed edges in the current phenotype network, i.e. delete one edge and propose a new, distinct directed edge.

A reverse step is a special case of a relocate step; however, we define the relocate step to exclude the reversal of a directed edge.

Each time a new model is proposed, new parameter estimates are generated from corresponding proposal distributions as necessary, such that those parameters unaffected by the reversal or relocation of an edge remain unchanged. This ensures that the current and candidate models differ as little as possible, to keep

steps small and increase the likelihood that proposed move will be accepted. This is possible because of conditional independence, whereby a trait t is independent of all other traits, given those traits and QTL directly affecting it: denoted $\upsilon_Y(t)$ and $\upsilon_Q(t)$, respectively. The model used is that given in Equation 3.4 (p. 38), restated here for convenience:

$$y_{it} = \beta_{0t} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik} + \varepsilon_{it}$$

where the model terms are as defined in Table 3.1 (p. 37). For the tth quantitative trait of the ith individual,

$$y_{it} = \mu_{it}^* + \varepsilon_{it},$$

where
$$\mu_{it}^* = \beta_{0t} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik}$$
$$\varepsilon_{it} \stackrel{i.i.d.}{\sim} N(0, \sigma_t^2).$$

As the genetic architecture is known, the sets of QTL parents for each trait are fixed for any causal network structure with N_T quantitative traits:

$$\upsilon_Q = (\upsilon_Q(1), \cdots, \upsilon_Q(N_T)).$$

The phenotype network is estimated using a model selection procedure, and proposing changes to the components of the phenotype network structure means that we are proposing changes to the sets of trait parents of each trait

$$\upsilon_Y = (\upsilon_Y(1), \cdots, \upsilon_Y(N_T)).$$

In this chapter, the proposed change to the phenotype network structure is either the reversal or relocation of a directed edge. A reversible jump MCMC algorithm is used to propose candidate models, and this is introduced for the general case in Section 4.1. Each of the update, reverse and relocate move types are detailed in Sections 4.2, 4.3 and 4.4, respectively. A brief summary of the approach is given in Section 4.5, and a simulation study to show the limitations of this approach while keeping the number of directed edges fixed is included in Section 4.6.

4.1   Reversible jump algorithm for model updates, keeping the number of edges fixed

We have a model space $\mathbb{M}$ with countable elements $M \in \mathbb{M}$. Model M is defined by its causal network structure $\mathbf{S}$ (comprised of a set of nodes connected by a set of directed edges), and the parameters Ω. We partition the model space into the structure space $\mathbb{S}$, and the parameter space $\mathbb{O}$, such that $\mathbf{S} \in \mathbb{S}$ and $\Omega \in \mathbb{O}$. Let us propose a move from the current model, M, to a candidate model, M', by selecting a move type x from the move space $\mathbb{X}$, such that a unique move type x changes model M into M'. Here, x is one of the update (x^U), reverse (x^S) or relocate (x^C) steps mentioned earlier. The phenotype network structures of models M and M' usually differ by a single directed edge, unless an update step is selected and then the phenotype network structure remains unchanged but all parameters are updated.

Let the quantitative trait measurements be denoted by the $n \times N_T$ matrix $\mathbf{Y}$, where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_t, \cdots, \mathbf{y}_{N_T}]$ and $\mathbf{y}_t = [y_{1t}, \cdots, y_{nt}]^\top$, for n individuals. The QTL genotypes ($\mathbf{Q}$), QTL locations (Λ), marker genotypes ($\mathbf{G}$) and recombination

fractions (**r**) are as described in Table 3.1 (p. 37), and are assumed to be known. Let Ψ denote the known parameters, such that $\Psi = \{\mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}\}$. Given the data **Y** and Ψ, the posterior distribution for model *M* is:

$$p(M|\mathbf{Y}, \Psi) \propto p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) p(\Omega|\mathbf{S}) p(\mathbf{S})$$

as detailed in Section 4.1.2 and given in Equation 4.4 (p. 72).

Whenever a Markov chain Monte Carlo step is carried out, conditional on the current state of the chain (i.e. model $M = \{\Omega, \mathbf{S}\}$), the following occur:

- select a move type *x* with probability $q(x|\Omega, \mathbf{S})$ where

$$\sum_{x \in \mathbb{X}} q(x|\Omega, \mathbf{S}) = 1 \quad \text{for all } (\Omega, \mathbf{S})$$

- given move type *x*, propose a candidate phenotype network structure $\mathbf{S}'$ with probability $q(\mathbf{S}'|x, \Omega, \mathbf{S})$ where

$$\sum_{\mathbf{S}' \in \mathbb{S}} q(\mathbf{S}'|x, \Omega, \mathbf{S}) = 1 \quad \text{for all } (x, \Omega, \mathbf{S})$$

- generate a set of random variables $\mathbf{u}_{M'}$ within their support space $\mathbb{U}_{M'}$, with probability density

$$q(\mathbf{u}_{M'}|x, \mathbf{S}', \Omega, \mathbf{S})$$

- the parameters of the new state of the chain, Ω', are generated by the deterministic function $g_{MM'}$, where

$$\Omega' = g_{MM'}(\Omega, \mathbf{u}_{M'}).$$

The general form of the acceptance probability for the update, reverse and relocate steps, is defined as $\alpha = min(1, r)$, where

$$r = \frac{p(M'|\mathbf{Y}, \Psi)}{p(M|\mathbf{Y}, \Psi)} \times \frac{q(\mathbf{u}'_M, \mathbf{S}, x'|\Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x|\Omega, \mathbf{S})} \times \left| \frac{\partial(\Omega', \mathbf{u}'_M)}{\partial(\Omega, \mathbf{u}_{M'})} \right| \tag{4.1}$$

(Green 1995). The move type transforming model M into M' is denoted x, and x' is the reverse of move type x — transforming model M' into M. The posterior distributions are denoted by $p()$ for models M and M', and $q()$ denotes the proposal distributions. It follows that:

$$\begin{aligned}
\alpha &= min(1, r) \quad \text{where} \\
r &= \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) p(\Omega'|\mathbf{S}') p(\mathbf{S}')}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) p(\Omega|\mathbf{S}) p(\mathbf{S})} \frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \times \\
&\quad \frac{q(\mathbf{S}|x', \Omega', \mathbf{S}')}{q(\mathbf{S}'|x, \Omega, \mathbf{S})} \frac{q(x'|\Omega', \mathbf{S}')}{q(x|\Omega, \mathbf{S})} \left| \frac{\partial(\Omega', \mathbf{u}'_M)}{\partial(\Omega, \mathbf{u}_{M'})} \right|.
\end{aligned} \tag{4.2}$$

4.1.1   Selecting the move type

At each step in the chain a move type is chosen at random from the update, reverse and relocate steps. The probabilities associated with the selection of each move type are chosen to allow equal probability of model parameter updates and exploration of the various phenotype network structures. The probability of selecting each move type is given in Table 4.1.

We define the operators U, S_g and $C_{g,f}$ such that:

- $U[M]$ is the existing model M with updated parameters Ω, but $\mathbf{S}$ unchanged

- $S_g[M] \in \mathbb{M}$ is a new model which reverses edge g (in $\mathbf{S}$) with an associated generation of new parameter values

Table 4.1: *The probability of each move type, as well as the number of candidate models given model M and move type x, denoted $|Ne(M,x)|$. There exist N_T traits, such that there are $N_T(N_T-1)$ unique directed edges possible; subsets of these edges form the current and candidate phenotype network structures. Model M has E_M directed edges.*

move type	(x)	probability	size of the neighbourhood of model M $\|Ne(M,x)\|$
Update	(x^U)	0.50	1
Reverse	(x^S)	0.25	$\sum_{g=1}^{E_M} I_a(S_g[M])$
Relocate	(x^C)	0.25	$\sum_{g=1}^{E_M} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M]) I_R(g,f\|\mathbf{S}) I_\in(f\|\mathbf{S})$

- $C_{g,f}[M] \in \mathbb{M}$ is a new model which relocates edge g to new location f (in **S**) with an associated generation of new parameter values.

Those parameters associated with the reversal or relocation of an edge include the new trait-to-trait effect, as well as select trait intercepts and residual variances. The exact parameters updated are stated in Sections 4.3 and 4.4 for the reverse and relocate steps, respectively.

Suppose we select move type x. Let the neighbourhood of the current model, $Ne(M,x)$, denote the set of all acyclic phenotype network structures given move type x. The size of $Ne(M,x)$, i.e. the number of distinct acyclic phenotype networks given model M and move type x, is denoted $|Ne(M,x)|$ and is given in Table 4.1 for all move types. The following indicator functions are used to identify and count the number of acyclic models:

$$I_a(M) = \begin{cases} 1, & \text{if model } M \text{ is acyclic} \\ 0, & \text{otherwise} \end{cases}$$

$$
I_R(g, f | \mathbf{S}) \;=\; I(g \neq rev(f) | \mathbf{S}) = \begin{cases} 1, & \text{if edge } g \text{ is not the reverse of edge } f \text{ in } \mathbf{S} \\ 0, & \text{otherwise} \end{cases}
$$

$$
I_{\in}(f | \mathbf{S}) \;=\; I(f \notin \mathbf{S}) = \begin{cases} 1, & \text{if edge } f \text{ is not an edge in } \mathbf{S} \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}
$$

Therefore, according to the size of the neighbourhood given in Table 4.1, the neighbourhood for a relocate step includes only models where edge g is relocated to a new, distinct location f, forming an acyclic model where edge f is not the reverse of edge g.

4.1.2 Posterior distribution

The causal network structure for model M is denoted $\mathbf{S}$, comprised of a set of nodes connected by a set of directed edges. Again, let Ψ denote the known parameters, such that $\Psi = \{\mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}\}$. Given the data $\mathbf{Y}$ and Ψ, the posterior distribution for model M, $M = \{\Omega, \mathbf{S}\}$, is:

$$
\begin{aligned}
p(M | \mathbf{Y}, \Psi) &= \frac{p(\mathbf{Y} | M, \Psi) p(M | \Psi)}{p(\mathbf{Y} | \Psi)} \\
&\propto p(\mathbf{Y} | \Omega, \mathbf{S}, \Psi) p(\Omega, \mathbf{S} | \Psi) \\
&\propto p(\mathbf{Y} | \Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) p(\Omega | \mathbf{S}) p(\mathbf{S}).
\end{aligned} \tag{4.4}
$$

The general expression of the likelihood is:

$$
p(\mathbf{Y} | \Omega, \mathbf{S}, \Psi) \propto \left(\prod_{t=1}^{N_T} \sigma_t \right)^{-n} \prod_{i=1}^{n} \exp\left[-\frac{1}{2} \sum_{t=1}^{N_T} \frac{\left(y_{it} - \beta_{0t} - \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} - \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik} \right)^2}{\sigma_t^2} \right].
$$

$$\tag{4.5}$$

where n is the number of individuals, and there are N_T quantitative traits and N_{qtl} QTL. The observed value of trait t for individual i is denoted by y_{it}, and the variance of trait t is denoted σ_t^2 . The observed QTL genotype for the ith individual is denoted $q_{i\ell}$. The effect of the ℓth QTL, $\ell \in \upsilon_Q(t)$, on trait t is denoted $\phi_{\ell t}$. The intercept for trait t is β_{0t}, and the effect of trait k, $k \in \upsilon_Y(t)$, on trait t is denoted β_{kt}.

The prior for model M is:

$$
\begin{aligned}
p(M) &= p(\Omega, \mathbf{S}) \\
&= p(\Omega|\mathbf{S})p(\mathbf{S})
\end{aligned}
$$

where the joint parameter prior is:

$$
\begin{aligned}
p(\Omega|\mathbf{S}) &= p(\boldsymbol{\beta_0}, \beta, \Phi, \sigma^2|\mathbf{S}) \\
&= p(\boldsymbol{\beta_0}|\mathbf{S})p(\beta|\mathbf{S})p(\Phi|\mathbf{S})p(\sigma^2|\mathbf{S}) \quad (4.6)
\end{aligned}
$$

for $\beta_0 = (\beta_{01}, \cdots, \beta_{0N_T})$, $\beta = (\beta_{\upsilon_Y(1)1}, \cdots, \beta_{\upsilon_Y(N_T)N_T})$, $\Phi = (\phi_{\upsilon_Q(1)1}, \cdots \phi_{\upsilon_Q(N_T)N_T})$, $\sigma^2 = (\sigma_1^2, \cdots, \sigma_{N_T}^2)$. For convenience, we assume independence of parameters in the prior distribution.

$$
\begin{aligned}
\beta_{0t}|\mathbf{S} &\sim N(0, \sigma_{\beta_{0t}}^2) \text{ for } t = 1, \cdots, N_T \\
\beta_{st}|\mathbf{S} &\sim N(0, \sigma_{\beta_t}^2) \text{ for } s = 1, \cdots, N_T \text{ and } t = 1, \cdots, N_T, \text{ for } s \neq t \\
\phi_{\ell t}|\mathbf{S} &\sim N(0, \sigma_{\phi_\ell}^2) \text{ for } \ell = 1, \cdots, N_{qtl} \text{ and } t = 1, \cdots, N_T \\
\sigma_t^{-2}|\mathbf{S} &\sim \Gamma(a_0, b_0) \text{ for } t = 1, \cdots, N_T. \quad (4.7)
\end{aligned}
$$

We use $\sigma_{\beta_{0t}}^2 = \sigma_{\beta_0}^2$ and $\sigma_{\beta_t}^2 = \sigma_{\beta}^2$ for all $t = 1, \cdots, N_T$.

To obtain a prior for the phenotype network of model M, a prior probability

Possible states of a connection between traits Y_1 and Y_2

Reversed   Absent   Forwards
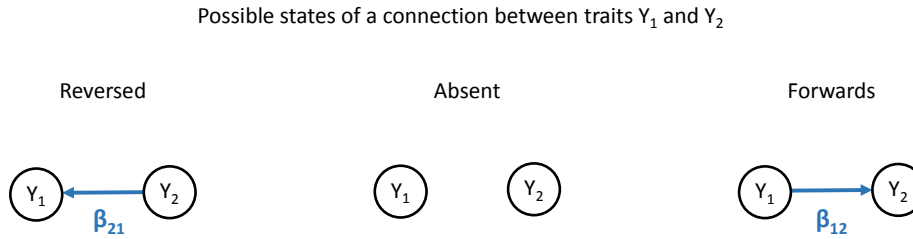
Figure 4.1: *A connection may have one of three possible states between two traits, where the directed edge is said to be reversed, absent, or forwards.*

distribution is set separately for the state of a connection between each pair of trait nodes; these are then normalised over all combinations of possible connections (i.e. all possible models). There are three possible states for the connection between any two trait nodes: absent, plus two possible directed edges. Each of the N_T traits are arbitrarily assigned a numeric label $(1, \cdots, t, \cdots, N_T)$ and by convention a directed edge is said to be forwards when the trait with the lower numeric label is affecting the trait with the higher numeric label, and reversed when the trait with the higher numeric label has an effect on the trait with the lower numeric label. Possible states for any connection between two nodes are therefore denoted reversed, absent, and forwards, as shown in Figure 4.1. Note that there is no particular meaning associated with the nomenclature forwards versus reversed. This type of distinction between the states of a connection between two nodes is also taken in Scutari (2013), where a multivariate trinomial random variable is used to model the directed graph.

The prior on the state of the gth connection between two trait nodes in the phenotype network structure $\mathbf{S}$ (for model M), denoted p_{gj}, is dependent on the directed edge present between these two nodes, as specified in Equation 4.8.

$$p_{gj} = \begin{cases} p_{g1}, & \text{if edge } g \text{ is reversed (e.g. } Y_1 \leftarrow Y_2) \\ p_{g2}, & \text{if edge } g \text{ is absent} \\ p_{g3}, & \text{if edge } g \text{ is forwards (e.g. } Y_1 \rightarrow Y_2). \end{cases} \tag{4.8}$$

where $\sum_{j=1}^{3} p_{gj} = 1 \ \forall g$. We may assign any values to p_{gj}, perhaps using prior knowledge of the phenotype network structure. For example, for edge g we could set the prior probability of the edge being absent as $p_{absent} = p_{g2}$, then the prior probability of the edge being present is $p_{present} = 1 - p_{absent}$ and the prior probability of the edge being forwards or reversed is then $p_{g1} = p_{g3} = p_{present}/2$.

Assuming *a priori* independence among connections, the joint probability for a given phenotype network **S**, is given by;

$$p^*(\mathbf{S}) = \prod_{g=1}^{max(E)} p_{gj}, \text{ for } j \in \{1,2,3\} \tag{4.9}$$

where there exist $max(E) = \frac{1}{2} N_T (N_T - 1)$ possible connections (or directed edges) in any model $M \in \mathbb{M}$ with N_T traits, each with state j in model M, for $j \in \{1,2,3\}$ (see Equation 4.8). The prior probability distribution of all possible acyclic phenotype network structures is then obtained by normalising these joint probabilities over all K acyclic phenotype network structures:

$$p(\mathbf{S}) = \frac{p^*(\mathbf{S}) I_a(\mathbf{S})}{\sum_{k=1}^{K} p^*(\mathbf{S}_k) I_a(\mathbf{S}_k)} \tag{4.10}$$

where

$$I_a(\mathbf{S}) = \begin{cases} 1, & \text{if graph structure } \mathbf{S} \text{ is acyclic} \\ 0, & \text{otherwise.} \end{cases}$$

We note that a uniform prior on the state of a connection between any two trait nodes, such that $p_{g1} = p_{g2} = p_{g3} = \frac{1}{3}$, will result in the prior on the phenotype network structure cancelling out of the acceptance probability (in this chapter). However, a uniform prior does imply that an edge is more likely to be included in the phenotype network structure (in one of two directions) than not.

4.1.3   Proposal distributions

The proposal of model M' given model M can be split into three parts:

- the selection of the move type,

- the selection of the candidate phenotype network structure (given the move type),

- the generation of new candidate parameters (given the move type, and the candidate phenotype network structure).

It follows that the proposal ratio for a move from model M to M', is:

$$\frac{q(\mathbf{u}'_M, \mathbf{S}, x' | \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x | \Omega, \mathbf{S})} = \frac{q(\mathbf{u}'_M | \mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'} | \mathbf{S}', x, \Omega, \mathbf{S})} \frac{q(\mathbf{S} | x', \Omega', \mathbf{S}')}{q(\mathbf{S}' | x, \Omega, \mathbf{S})} \frac{q(x' | \Omega', \mathbf{S}')}{q(x | \Omega, \mathbf{S})}.$$

The probability of selecting move type x given M, and x' given M', are denoted $q(x | \Omega, \mathbf{S})$ and $q(x' | \Omega', \mathbf{S}')$, respectively, and are given in Table 4.1 (p. 71). Given move type x to move from M to M', the phenotype network structure $\mathbf{S}'$ is randomly selected with probability

$$q(\mathbf{S}' | x, \Omega, \mathbf{S}) = \frac{1}{|Ne(M, x)|}$$

and conversely, given move type x' to move from M' to M, the phenotype network structure $\mathbf{S}$ is randomly selected with probability

$$q(\mathbf{S}|x',\Omega',\mathbf{S}') = \frac{1}{|Ne(M',x')|}.$$

The proposal distributions for the parameters of the current and candidate models are denoted $q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')$ and $q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})$, respectively. The proposal distributions used depend on whether or not the move types x and x' alter the phenotype network structure. An update step updates parameters using a random walk sampler, conditional on the current parameter estimates. These proposal distributions are given in Section 4.2. For the reverse and relocate steps, however, the phenotype network structure is altered so we use an independence sampler to propose candidate parameter values. The proposal distributions are based on the least squares estimates (LSE) of the parameters conditional on the proposed phenotype network structure. The LSE are effectively obtained from separate regressions of each trait on the traits and QTL directly affecting it ($\upsilon_Y(t)$ and $\upsilon_Q(t)$, respectively). We note that the use of the multivariate normal distribution would be more efficient; however, we still have an efficient sampler using well-tuned proposal distributions.

Let us consider a move from model M to M'. As minimising the SSE maximises the log likelihood (see Section 3.1.1), we calculate the LSE and therefore require the design matrix in order to form the proposal distribution for each parameter. In model M', let trait t be affected by $|\upsilon_Q(t)|$ QTL and $|\upsilon_Y(t)'|$ quantitative traits. Each of the t quantitative traits within model M' has a design matrix, denoted $\mathbf{X}_t{}'$, with n rows (for n individuals) and $n_c = |\upsilon_Q(t)| + |\upsilon_Y(t)'| + 1$ columns.

Note that here we use a prime to refer to the candidate model, and a '$\top$' denotes the transpose of a matrix. Let

$$\underline{\mathbf{o}}'_t = \left(\mathbf{X}'_t{}^\top \mathbf{X}'_t \right)^{-1},$$

and $\underline{\mathbf{d}}^p$ be an indicator vector of length n_c such that:

$$\underline{\mathbf{d}}^p = (d_1{}^p, \cdots, d_{n_c}{}^p)^\top$$

where

$$d_e{}^p = \begin{cases} 1, & \text{if parameter } p \text{ being estimated corresponds to column } e \text{ in } \mathbf{X}'_t \\ 0, & \text{otherwise} \end{cases}$$

for $e = 1, \cdots, n_c$. The term

$$V_p = \underline{\mathbf{d}}^p{}^\top \underline{\mathbf{o}}'_t \underline{\mathbf{d}}^p \tag{4.11}$$

is therefore used to define the element in the matrix $\left(\mathbf{X}'_t{}^\top \mathbf{X}'_t \right)^{-1}$ that corresponds to the variance of parameter p. This notation also applies to a move from model M' to M, using the design matrix for the current model ($\mathbf{X}_t$) instead.

Consider the following example to illustrate the use of the design matrix. Suppose we propose a move from model M' to M, where trait t is affected by one QTL (Q_1) and one trait (Y_1). The design matrix for the tth trait in model M with n individuals is

$$\mathbf{X}_t = \begin{bmatrix} 1 & y_{11} & q_{11} \\ \vdots & \vdots & \vdots \\ 1 & y_{i1} & q_{i1} \\ \vdots & \vdots & \vdots \\ 1 & y_{n1} & q_{n1} \end{bmatrix}_{n \times 3}$$

with columns corresponding to the subset of parameters $\{\beta_{0t}, \beta_{1t}, \phi_{1t}\}$. We have

$$\underline{\mathbf{o}}_t = \left(\mathbf{X}_t^\top \mathbf{X}_t \right)^{-1}$$

and the following indicator vectors:

$$
\begin{aligned}
\underline{\mathbf{d}}^{\beta_{0t}} &= [1,0,0]^\top \\
\underline{\mathbf{d}}^{\beta_{1t}} &= [0,1,0]^\top \\
\underline{\mathbf{d}}^{\phi_{1t}} &= [0,0,1]^\top
\end{aligned}
$$

such that:

$$
\begin{aligned}
V_{\beta_{0t}} &= [1,0,0]\underline{\mathbf{o}}_t[1,0,0]^\top \\
V_{\beta_{1t}} &= [0,1,0]\underline{\mathbf{o}}_t[0,1,0]^\top \\
V_{\phi_{1t}} &= [0,0,1]\underline{\mathbf{o}}_t[0,0,1]^\top.
\end{aligned}
$$

Given the above description of the design matrix for a move altering the phenotype network structure (a reverse or relocate step), let the intercept for the tth trait in model M' be denoted β'_{0t}, the effect of trait s on trait t be denoted β'_{st}, and the residual variance for trait t be denoted $\sigma_t'^2$, with the LSE $\widehat{\beta}'_{0t}$, $\widehat{\beta}'_{st}$ and $\widehat{\sigma}_t'^2$, respectively. Let τ be a constant used to scale the variance of the least squares estimates in all proposal distributions for the trait intercept and trait-to-trait effects (for all move types). As τ is constant, we will omit it from the conditional statements for simplicity of notation. We have the following proposal distributions for a move from model M to M':

- for the intercept for the tth trait:

$$\beta'_{0t} | \widehat{\beta}'_{0t}, \widehat{\sigma}_t'^2, V_{\beta'_{0t}} \sim N\left(\widehat{\beta}'_{0t}, \frac{1}{\tau}\widehat{\sigma}_t'^2 V_{\beta'_{0t}} \right) \tag{4.12}$$

- for the trait-to-trait effect of trait s on trait t, resulting from either a reversal or relocation of an edge:

$$\beta'_{st}|\widehat{\beta}'_{st},\widehat{\sigma}'^2_t,V_{\beta'_{st}} \sim N\left(\widehat{\beta}'_{st},\frac{1}{\tau}\widehat{\sigma}'^2_t V_{\beta'_{st}}\right) \tag{4.13}$$

- for the residual variance for trait t:

$$\sigma'^2_t|\widehat{\sigma}'^2_t \sim \text{Unif}\left(0.5\widehat{\sigma}'^2_t,1.5\widehat{\sigma}'^2_t\right), \tag{4.14}$$

such that the expected values of the proposal distributions are conditional means:
$\text{E}(\beta'_{0t}|\widehat{\beta}'_{0t},\widehat{\sigma}'^2_t,V_{\beta'_{0t}}) = \widehat{\beta}'_{0t}$, $\text{E}(\beta'_{st}|\widehat{\beta}'_{st},\widehat{\sigma}'^2_t,V_{\beta'_{st}}) = \widehat{\beta}'_{st}$, and $\text{E}(\sigma'^2_t|\widehat{\sigma}'^2_t) = \widehat{\sigma}'^2_t$.

The proposal distributions to move from model M' to M are the same but with the following conditional means: $\text{E}(\beta_{0t}|\widehat{\beta}_{0t},\widehat{\sigma}^2_t,V_{\beta_{0t}}) = \widehat{\beta}_{0t}$, $\text{E}(\beta_{st}|\widehat{\beta}_{st},\widehat{\sigma}^2_t,V_{\beta_{st}}) = \widehat{\beta}_{st}$, and $\text{E}(\sigma^2_t|\widehat{\sigma}^2_t) = \widehat{\sigma}^2_t$:

$$\beta_{0t}|\widehat{\beta}_{0t},\widehat{\sigma}^2_t,V_{\beta_{0t}} \quad \sim \quad N\left(\widehat{\beta}_{0t},\frac{1}{\tau}\widehat{\sigma}'^2_t V_{\beta_{0t}}\right) \tag{4.15}$$

$$\beta_{st}|\widehat{\beta}_{st},\widehat{\sigma}^2_t,V_{\beta_{st}} \quad \sim \quad N\left(\widehat{\beta}_{st},\frac{1}{\tau}\widehat{\sigma}'^2_t V_{\beta_{st}}\right) \tag{4.16}$$

$$\sigma^2_t|\widehat{\sigma}^2_t \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}^2_t,1.5\widehat{\sigma}^2_t\right). \tag{4.17}$$

Note that the QTL effect parameters are only updated in an update step (using a random walk sampler), as we have assumed that the genetic architecture is known, i.e. υ_Q is fixed, and therefore only the size of the QTL effect changes.

To summarize, the proposal ratio for the update, reverse and relocate steps can now be written as:

$$\begin{aligned}
\frac{q(\mathbf{u}'_M,\mathbf{S},x'|\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'},\mathbf{S}',x|\Omega,\mathbf{S})} &= \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})}\frac{q(\mathbf{S}|x',\Omega',\mathbf{S}')}{q(\mathbf{S}'|x,\Omega,\mathbf{S})}\frac{q(x'|\Omega',\mathbf{S}')}{q(x|\Omega,\mathbf{S})} \\
&= \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})}\frac{|Ne(M,x)|}{|Ne(M',x')|}
\end{aligned} \tag{4.18}$$

since $q(x'|\Omega',\mathbf{S}') = q(x|\Omega,\mathbf{S})$ as the same move type is required (and selected with the same probability) for $M \to M'$ and $M' \to M$, i.e. $x = x'$.

For an update step the proposal ratio further simplifies to

$$\frac{q(\mathbf{u}'_M,\mathbf{S},x'|\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'},\mathbf{S}',x|\Omega,\mathbf{S})} = \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \tag{4.19}$$

as $\mathbf{S} = \mathbf{S}'$ and therefore $|Ne(M,x=x^U)| = |Ne(M',x'=x^U)| = 1$.

4.1.4 The Jacobian

The determinant of the Jacobian matrix, referred to as simply the Jacobian, arises as a result of the transition from the current model to the candidate model. The current model parameters, Ω, and the candidate model parameters, Ω', may share a set of parameters, such that

$$\begin{aligned}
\Omega &= (\Omega_{MM'}, \Omega_{M\backslash M'}) \\
\Omega' &= (\Omega'_{M'M}, \Omega'_{M'\backslash M})
\end{aligned}$$

where $\Omega_{MM'}$ and $\Omega_{M'M}$ denote those parameters in both models M and M', $\Omega_{M\backslash M'}$ denotes those parameters in model M but not in model M', and $\Omega'_{M'\backslash M}$ denotes those parameters in model M' but not in model M.

The move from $M \to M'$ via move type x has the reverse move x' to move from $M' \to M$. When making the move x between models M and M' we generate the random variable $\mathbf{u}_{M'}$ from the distribution $q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})$. Similarly, when making the move x' between models M' and M we generate the random variable $\mathbf{u}'_M$ from the distribution $q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')$.

The dimension of each is then:

$$|\Omega_{MM'}| = d_{MM'} \qquad\qquad |\Omega'_{M'M}| = d_{M'M}$$

$$|\Omega_{M\setminus M'}| = d_{M\setminus M'} \qquad\qquad |\Omega'_{M'\setminus M}| = d_{M'\setminus M}$$

$$|\mathbf{u}_{M'}| = c_{M'} \qquad\qquad |\mathbf{u}'_M| = c_M$$

with the dimension matching constraint met, as specified by Green (1995):

$$|\Omega| + |\mathbf{u}_{M'}| = |\Omega'| + |\mathbf{u}'_M|$$

$$|\Omega_{MM'}| + |\Omega_{M\setminus M'}| + |\mathbf{u}_{M'}| = |\Omega'_{M'M}| + |\Omega'_{M'\setminus M}| + |\mathbf{u}'_M|$$

$$|\Omega_{M\setminus M'}| + |\mathbf{u}_{M'}| = |\Omega'_{M'\setminus M}| + |\mathbf{u}'_M|$$

$$d_{M\setminus M'} + c_{M'} = d_{M'\setminus M} + c_M$$

where $|\Omega_{MM'}| = |\Omega'_{M'M}|$.

Parameters for the new state of the chain (the candidate model) are generated using the deterministic function $g_{MM'}$, such that

$$\Omega' = g_{MM'}(\Omega, \mathbf{u}_{M'})$$

$$= g_{MM'}(\Omega_{MM'}, \Omega_{M\setminus M'}, \mathbf{u}_{M'}).$$

Similarly, we have the function $h_{M'M}$ to reverse this:

$$\mathbf{u}'_M = h_{MM'}(\Omega, \mathbf{u}_{M'})$$

$$= h_{MM'}(\Omega_{MM'}, \Omega_{M\setminus M'}, \mathbf{u}_{M'}),$$

and the calculation of the Jacobian is given in Equation 4.20.

$$\left| \frac{\partial (\Omega', \mathbf{u}'_M)}{\partial (\Omega, \mathbf{u}_{M'})} \right| = \left| \frac{\partial (\Omega'_{M'M}, \Omega'_{M'\setminus M}, \mathbf{u}'_M)}{\partial (\Omega_{MM'}, \Omega_{M\setminus M'}, \mathbf{u}_{M'})} \right|$$

$$= \begin{vmatrix} \frac{\partial \Omega'_{M'M}}{\partial \Omega_{MM'}} & \frac{\partial \Omega'_{M'M}}{\partial \Omega_{M\setminus M'}} & \frac{\partial \Omega'_{M'M}}{\partial \mathbf{u}_{M'}} \\ \frac{\partial \Omega'_{M'\setminus M}}{\partial \Omega_{MM'}} & \frac{\partial \Omega'_{M'\setminus M}}{\partial \Omega_{M\setminus M'}} & \frac{\partial \Omega'_{M'\setminus M}}{\partial \mathbf{u}_{M'}} \\ \frac{\partial \mathbf{u}'_M}{\partial \Omega_{MM'}} & \frac{\partial \mathbf{u}'_M}{\partial \Omega_{M\setminus M'}} & \frac{\partial \mathbf{u}'_M}{\partial \mathbf{u}_{M'}} \end{vmatrix}$$

$$= \begin{vmatrix} I & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{vmatrix}$$

$$= 1. \tag{4.20}$$

The update step, outlined in Section 4.2, only updates parameters, such that Ω and Ω' are in the same space. Therefore, $\Omega' = \mathbf{u}_{M'}$ and $\mathbf{u}'_M = \Omega$ and the Jacobian is 1. The reverse and relocate move types are outlined in Sections 4.3 and 4.4, and as these move types delete one element of Ω and create a new element, such that Ω and Ω' are not in the same space, the Jacobian is a little more complex. However, for these two move types, the Jacobian is also 1, see Appendix C.

4.2 The update step

For model M, $M = \{\Omega, \mathbf{S}\}$, an update step updates all of the parameters within the current model, without altering the causal network structure, $\mathbf{S}$. A random walk sampler is used as the model structure is not changed, with the Metropolis–Hastings acceptance probability given in Equation 4.31 (p. 86). An example update step and the calculation of the corresponding acceptance probability is given

in Section 4.2.1.

Suppose that there exist N_T traits, and we are proposing a move from model M to M'. Given in Equation 4.19 (p. 81), the proposal ratio for an update step is:

$$\frac{q(\mathbf{u}'_M, \mathbf{S}, x'|\Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x|\Omega, \mathbf{S})} = \frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})}$$

$$= \frac{q(\Omega|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\Omega'|\mathbf{S}', x, \Omega, \mathbf{S})}$$

as all parameters are updated $\mathbf{u}'_M = \Omega$ and $\mathbf{u}_{M'} = \Omega'$, and it follows that:

$$\frac{q(\Omega|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\Omega'|\mathbf{S}', x, \Omega, \mathbf{S})} = \frac{\prod_{t=1}^{N_T}\left\{q(\beta_{0t}|\beta'_{0t}, \sigma'^2_t, V_{\beta'_{0t}})q(\sigma^2_t|\sigma'^2_t)\right\}}{\prod_{t=1}^{N_T}\left\{q(\beta'_{0t}|\beta_{0t}, \sigma^2_t, V_{\beta_{0t}})q(\sigma'^2_t|\sigma^2_t)\right\}} \times$$

$$\left[\frac{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}q(\beta_{st}|\beta'_{st}, \sigma'^2_t, V_{\beta'_{st}})}{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}q(\beta'_{st}|\beta_{st}, \sigma^2_t, V_{\beta_{st}})}\right] \times$$

$$\left[\frac{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}q(\phi_{\ell t}|\phi'_{\ell t}, \sigma'^2_t, V_{\phi'_{\ell t}})}{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}q(\phi'_{\ell t}|\phi_{\ell t}, \sigma^2_t, V_{\phi_{\ell t}})}\right]. \qquad (4.21)$$

The trait intercepts are denoted β_{0t} for $t = 1, \cdots, N_T$, β_{st} is the effect of trait s on trait t, the effect of the ℓth QTL on the tth trait is denoted by $\phi_{\ell t}$, and σ^2_t is the residual variance of trait t; where parameters relating to the candidate model are indicated by a prime ($'$). Each parameter is updated via its corresponding proposal distribution, given below for $q(\Omega'|\mathbf{S}', x, \Omega, \mathbf{S})$:

$$\beta'_{0t}|\beta_{0t}, \sigma^2_t, V_{\beta_{0t}} \sim N\left(\beta_{0t}, \frac{1}{\tau}\sigma_t^2 V_{\beta_{0t}}\right) \quad \text{for } t = 1, \cdots, N_T \qquad (4.22)$$

$$\beta'_{st}|\beta_{st}, \sigma^2_t, V_{\beta_{st}} \sim N\left(\beta_{st}, \frac{1}{\tau}\sigma_t^2 V_{\beta_{st}}\right) \quad \text{for } s = 1, \cdots, N_T \text{ and } t = 1, \cdots, N_T \text{ for } s \neq t$$

$$(4.23)$$

$$\phi'_{\ell t}|\phi_{\ell t}, \sigma_t^2, V_{\phi_{\ell t}} \quad \sim \quad N\left(\phi_{\ell t}, \frac{1}{\tau}\sigma_t^2 V_{\phi_{\ell t}}\right) \quad \text{for } \ell = 1, \cdots, N_{qtl} \text{ and } \text{ for } t = 1, \cdots, N_T$$
(4.24)

$$\sigma_t'^2|\sigma_t^2 \quad \sim \quad \text{Unif}\left(0.5\sigma_t^2, 1.5\sigma_t^2\right) \quad \text{for } t = 1, \cdots, N_T$$
(4.25)

such that we have the conditional means: $\text{E}(\beta'_{0t}|\beta_{0t}, \sigma_t^2, V_{\beta_{0t}}) = \beta_{0t}$, $\text{E}(\beta'_{st}|\beta_{st}, \sigma_t^2, V_{\beta_{st}}) = \beta_{st}$, $\text{E}(\phi'_{\ell t}|\phi_{\ell t}, \sigma_t^2, V_{\phi_{\ell t}}) = \phi_{\ell t}$ and $\text{E}(\sigma_t'^2|\sigma_t^2) = \sigma_t^2$. Let parameter p be any of $\beta_{0t}, \beta_{st}, \phi_{\ell t}$ for $(\ell = 1, \cdots, N_{qtl})$, $(t = 1, \cdots, N_T)$ and $(s = 1, \cdots, N_T)$ where $s \neq t$; for parameter p the term V_p is defined in Equation 4.11 (p. 78).

Conversely, the proposal distributions for the parameters in model M given model M', corresponding to $q(\mathbf{u}'_M, \mathbf{S}, x'|\Omega', \mathbf{S}')$, are:

$$\beta_{0t}|\beta'_{0t}, \sigma_t'^2, V_{\beta'_{0t}} \quad \sim \quad N\left(\beta'_{0t}, \frac{1}{\tau}\sigma_t'^2 V_{\beta'_{0t}}\right) \quad \text{for } t = 1, \cdots, N_T$$
(4.26)

$$\beta_{st}|\beta'_{st}, \sigma_t'^2, V_{\beta'_{st}} \quad \sim \quad N\left(\beta'_{st}, \frac{1}{\tau}\sigma_t'^2 V_{\beta'_{st}}\right) \quad \text{for } s = 1, \cdots, N_T \text{ and } t = 1, \cdots, N_T \text{ for } s \neq t$$
(4.27)

$$\phi_{\ell t}|\phi'_{\ell t}, \sigma_t'^2, V_{\phi'_{\ell t}} \quad \sim \quad N\left(\phi'_{\ell t}, \frac{1}{\tau}\sigma_t'^2 V_{\phi'_{\ell t}}\right) \quad \text{for } \ell = 1, \cdots, N_{qtl} \text{ and } \text{ for } t = 1, \cdots, N_T$$
(4.28)

$$\sigma_t^2|\sigma_t'^2 \quad \sim \quad \text{Unif}(0.5\sigma_t'^2, 1.5\sigma_t'^2) \quad \text{for } t = 1, \cdots, N_T$$
(4.29)

such that we have the conditional means $\text{E}(\beta_{0t}|\beta'_{0t}, \sigma_t'^2, V_{\beta'_{0t}}) = \beta'_{0t}$, $\text{E}(\beta_{st}|\beta'_{0t}, \sigma_t'^2, V_{\beta'_{0t}}) = \beta'_{st}$, $\text{E}(\phi_{\ell t}|\phi'_{\ell t}, \sigma_t'^2, V_{\phi'_{\ell t}}) = \phi'_{\ell t}$ and $\text{E}(\sigma_t^2|\sigma_t'^2) = \sigma_t'^2$.

The prior ratio for an update step is:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \left[\frac{\prod_{t=1}^{N_T}\left\{p(\beta'_{0t}|\mathbf{S}')p(\sigma_t'^2|\mathbf{S}')\right\}}{\prod_{t=1}^{N_T}\left\{p(\beta_{0t}|\mathbf{S})p(\sigma_t^2|\mathbf{S})\right\}}\right]\left[\frac{\prod_{t=1}^{N_T}\prod_{s\in v_Y(t)}p(\beta'_{st}|\mathbf{S}')}{\prod_{t=1}^{N_T}\prod_{s\in v_Y(t)}p(\beta_{st}|\mathbf{S})}\right] \times$$
$$\left[\frac{\prod_{t=1}^{N_T}\prod_{\ell\in v_Q(t)}p(\phi'_{\ell t}|\mathbf{S}')}{\prod_{t=1}^{N_T}\prod_{\ell\in v_Q(t)}p(\phi_{\ell t}|\mathbf{S})}\right]$$
(4.30)

where the individual prior distributions are given in Equation 4.7 (p. 73) and the prior on the graph structure cancels out as $\mathbf{S} = \mathbf{S}'$.

Following the general form of the acceptance probability given in Equation 4.2 (p. 70), and substituting in the proposal and prior ratios in Equations 4.21 and 4.30 (p. 84), the acceptance probability for an update step is:

$$
\begin{aligned}
\alpha_{update} &= min(1, r_{update}), \quad \text{where} \\
r_{update} &= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \times \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \times \left|\frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})}\right| \\
&= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})} \left[\frac{\prod_{t=1}^{N_T}\left\{p(\beta'_{0t}|\mathbf{S}')p(\sigma_t'^2|\mathbf{S}')\right\}}{\prod_{t=1}^{N_T}\left\{p(\beta_{0t}|\mathbf{S})p(\sigma_t^2|\mathbf{S})\right\}}\right] \left[\frac{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}p(\beta'_{st}|\mathbf{S}')}{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}p(\beta_{st}|\mathbf{S})}\right] \times \\
&\quad \left[\frac{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}p(\phi'_{\ell t}|\mathbf{S}')}{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}p(\phi_{\ell t}|\mathbf{S})}\right] \left[\frac{\prod_{t=1}^{N_T}\left\{q(\beta_{0t}|\sigma_t'^2,\beta'_{0t},V_{\beta'_{0t}})q(\sigma_t^2|\sigma_t'^2)\right\}}{\prod_{t=1}^{N_T}\left\{q(\beta'_{0t}|\sigma_t^2,\beta_{0t},V_{\beta_{0t}})q(\sigma_t'^2|\sigma_t^2)\right\}}\right] \times \\
&\quad \left[\frac{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}q(\beta_{st}|\sigma_t'^2,\beta'_{st},V_{\beta'_{st}})}{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}q(\beta'_{st}|\sigma_t^2,\beta_{st},V_{\beta_{st}})}\right] \left[\frac{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}q(\phi_{\ell t}|\phi'_{\ell t},\sigma_t'^2,V_{\phi'_{\ell t}})}{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}q(\phi'_{\ell t}|\phi_{\ell t},\sigma_t^2,V_{\phi_{\ell t}})}\right] \quad (4.31)
\end{aligned}
$$

The general expression of the likelihood is given in Equation 4.5 (p. 72). The parameter priors are defined in Equation 4.7 (p. 73), and as the phenotype network structure is not altered, the prior on the phenotype network for both models M and M' are equal, so these terms cancel out of the acceptance probability. As Ω and Ω' are in the same space, the Jacobian is 1.

4.2.1   Example of an update step

Suppose that we want to carry out an update step to move from model M to M'; the example causal network structures are given in Figure 4.2. In this example we have a set of current parameters, Ω, and a set of candidate parameters, Ω':

$$
\begin{aligned}
\Omega &= \{\beta_{01},\beta_{02},\beta_{03},\beta_{12},\beta_{13},\beta_{23},\phi_{11},\phi_{22},\phi_{33},\sigma_1^2,\sigma_2^2,\sigma_3^2\} \\
\Omega' &= \{\beta'_{01},\beta'_{02},\beta'_{03},\beta'_{12},\beta'_{13},\beta'_{23},\phi'_{11},\phi'_{22},\phi'_{33},\sigma_1'^2,\sigma_2'^2,\sigma_3'^2\}.
\end{aligned}
$$
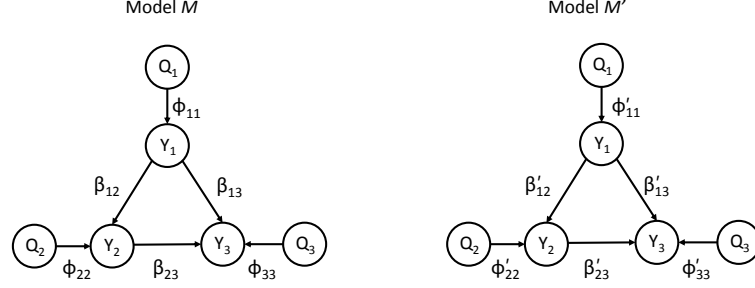
Figure 4.2: *Example causal network structures for both the current (M) and candidate (M′) models, given an update step. Here, the causal structure remains the same and we propose new parameter estimates.*

To move from model M to M', new parameters are generated from the following proposal distributions:

$$\beta'_{0t}|\beta_{0t},\sigma_t^2,V_{\beta_{0t}} \sim N\left(\beta_{0t},\frac{1}{\tau}\sigma_t^2 V_{\beta_{0t}}\right) \quad \text{for } t = 1,2,3 \quad \text{(see Eqn. 4.22)}$$

$$\beta'_{st}|\beta_{st},\sigma_t^2,V_{\beta_{st}} \sim N\left(\beta_{st},\frac{1}{\tau}\sigma_t^2 V_{\beta_{st}}\right) \quad \text{for } \{st\} = \{12,13,23\} \quad \text{(see Eqn. 4.23)}$$

$$\phi'_{\ell t}|\phi_{\ell t},\sigma_t^2,V_{\phi_{\ell t}} \sim N\left(\phi_{\ell t},\frac{1}{\tau}\sigma_t^2 V_{\phi_{\ell t}}\right) \quad \text{for } \{\ell t\} = \{11,22,33\} \quad \text{(see Eqn. 4.24)}$$

$$\sigma_t'^2|\sigma_t^2 \sim \text{Unif}\left(0.5\sigma_t^2,1.5\sigma_t^2\right) \quad \text{for } t = 1,2,3 \quad \text{(see Eqn. 4.25)}.$$

To move from model M' to M, the parameter proposal distributions are:

$$\beta_{0t}|\beta'_{0t},\sigma_t'^2,V_{\beta'_{0t}} \sim N\left(\beta'_{0t},\frac{1}{\tau}\sigma_t'^2 V_{\beta'_{0t}}\right) \quad \text{for } t = 1,2,3 \quad \text{(see Eqn. 4.26)}$$

$$\beta_{st}|\beta'_{st},\sigma_t'^2,V_{\beta'_{st}} \sim N\left(\beta'_{st},\frac{1}{\tau}\sigma_t'^2 V_{\beta'_{st}}\right) \quad \text{for } \{st\} = \{12,13,23\} \quad \text{(see Eqn. 4.27)}$$

$$\phi_{\ell t}|\phi'_{\ell t},\sigma_t'^2,V_{\phi'_{\ell t}} \sim N\left(\phi'_{\ell t},\frac{1}{\tau}\sigma_t'^2 V_{\phi'_{\ell t}}\right) \quad \text{for } \{\ell t\} = \{11,22,33\} \quad \text{(see Eqn. 4.28)}$$

$$\sigma_t^2|\sigma_t'^2 \sim \text{Unif}\left(0.5\sigma_t'^2,1.5\sigma_t'^2\right) \quad \text{for } t = 1,2,3 \quad \text{(see Eqn. 4.29)}.$$

For this example the proposal ratio is therefore:

$$
\frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \;=\; \left[\frac{\prod_{t=1}^{3}\left\{q(\beta_{0t}|\beta'_{0t},\sigma'^2_t,V_{\beta'_{0t}})q(\sigma^2_t|\sigma'^2_t)\right\}}{\prod_{t=1}^{3}\left\{q(\beta'_{0t}|\beta_{0t},\sigma^2_t,V_{\beta_{0t}})q(\sigma'^2_t|\sigma^2_t)\right\}}\right]\frac{q(\beta_{12}|\beta'_{12},\sigma'^2_2,V_{\beta'_{12}})}{q(\beta'_{12}|\beta_{12},\sigma^2_2,V_{\beta_{12}})} \times
$$

$$
\frac{q(\beta_{13}|\beta'_{13},\sigma'^2_3,V_{\beta'_{13}})q(\beta_{23}|\beta'_{23},\sigma'^2_3,V_{\beta'_{23}})}{q(\beta'_{13}|\beta_{13},\sigma^2_3,V_{\beta_{13}})q(\beta'_{23}|\beta_{23},\sigma^2_3,V_{\beta_{23}})} \times
$$

$$
\frac{q(\phi_{11}|\phi'_{11},\sigma'^2_1,V_{\phi'_{11}})q(\phi_{22}|\phi'_{22},\sigma'^2_2,V_{\phi'_{22}})q(\phi_{33}|\phi'_{33},\sigma'^2_3,V_{\phi'_{33}})}{q(\phi'_{11}|\phi_{11},\sigma^2_1,V_{\phi_{11}})q(\phi'_{22}|\phi_{22},\sigma^2_2,V_{\phi_{22}})q(\phi'_{33}|\phi_{33},\sigma^2_3,V_{\phi_{33}})}
$$

and the prior ratio:

$$
\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} \;=\; \frac{p(\beta'_{01}|\mathbf{S}')p(\beta'_{02}|\mathbf{S}')p(\beta'_{03}|\mathbf{S}')}{p(\beta_{01}|\mathbf{S})p(\beta_{02}|\mathbf{S})p(\beta_{03}|\mathbf{S})}\frac{p(\beta'_{12}|\mathbf{S}')p(\beta'_{13}|\mathbf{S}')p(\beta'_{23}|\mathbf{S}')}{p(\beta_{12}|\mathbf{S})p(\beta_{13}|\mathbf{S})p(\beta_{23}|\mathbf{S})} \times
$$

$$
\frac{p(\phi'_{11}|\mathbf{S}')p(\phi'_{22}|\mathbf{S}')p(\phi'_{33}|\mathbf{S}')}{p(\phi_{11}|\mathbf{S})p(\phi_{22}|\mathbf{S})p(\phi_{33}|\mathbf{S})}\frac{p(\sigma'^2_1|\mathbf{S}')p(\sigma'^2_2|\mathbf{S}')p(\sigma'^2_3|\mathbf{S}')}{p(\sigma^2_1|\mathbf{S})p(\sigma^2_2|\mathbf{S})p(\sigma^2_3|\mathbf{S})},
$$

where the prior distributions for each parameter are given in Equation 4.7 (p. 73) and the prior on the phenotype network structure cancels out as $\mathbf{S}=\mathbf{S}'$.
The acceptance probability for this example update step is therefore:

$$
\alpha_{update} \;=\; min(1, r_{update}), \quad \text{where}
$$

$$
r_{update} \;=\; \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}\left[\frac{\prod_{t=1}^{3}p(\beta'_{0t}|\mathbf{S}')p(\sigma'^2_t|\mathbf{S}')}{\prod_{t=1}^{3}p(\beta_{0t}|\mathbf{S})p(\sigma^2_t|\mathbf{S})}\right] \times
$$

$$
\frac{p(\beta'_{12}|\mathbf{S}')p(\beta'_{13}|\mathbf{S}')p(\beta'_{23}|\mathbf{S}')}{p(\beta_{12}|\mathbf{S})p(\beta_{13}|\mathbf{S})p(\beta_{23}|\mathbf{S})}\frac{p(\phi'_{11}|\mathbf{S}')p(\phi'_{22}|\mathbf{S}')p(\phi'_{33}|\mathbf{S}')}{p(\phi_{11}|\mathbf{S})p(\phi_{22}|\mathbf{S})p(\phi_{33}|\mathbf{S})} \times
$$

$$
\left[\frac{\prod_{t=1}^{3}q(\beta_{0t}|\beta'_{0t},\sigma'^2_t,V_{\beta'_{0t}})q(\sigma^2_t|\sigma'^2_t)}{\prod_{t=1}^{3}q(\beta'_{0t}|\beta_{0t},\sigma^t_1,V_{\beta_{0t}})q(\sigma'^2_t|\sigma^2_t)}\right] \times
$$

$$
\frac{q(\beta_{12}|\beta'_{12},\sigma'^2_2,V_{\beta'_{12}})q(\beta_{13}|\beta'_{13},\sigma'^2_3,V_{\beta'_{13}})q(\beta_{23}|\beta'_{23},\sigma'^2_3,V_{\beta'_{23}})}{q(\beta'_{12}|\beta_{12},\sigma^2_2,V_{\beta_{12}})q(\beta'_{13}|\beta_{13},\sigma^2_3,V_{\beta_{13}})q(\beta'_{23}|\beta_{23},\sigma^2_3,V_{\beta_{23}})} \times
$$

$$
\frac{q(\phi_{11}|\phi'_{11},\sigma'^2_1,V_{\phi'_{11}})q(\phi_{22}|\phi'_{22},\sigma'^2_2,V_{\phi'_{22}})q(\phi_{33}|\phi'_{33},\sigma'^2_3,V_{\phi'_{33}})}{q(\phi'_{11}|\phi_{11},\sigma^2_1,V_{\phi_{11}})q(\phi'_{22}|\phi_{22},\sigma^2_2,V_{\phi_{22}})q(\phi'_{33}|\phi_{33},\sigma^2_3,V_{\phi_{33}})}.
$$

The general expression of the likelihood is given in Equation 4.5 (p. 72).

4.3 The reverse step

A reverse step proposes the reversal of an existing directed edge, or trait-to-trait effect, within model M to form model M'. The reversal of an existing directed edge in model M means that both M and M' have phenotype network structures with the same number of directed edges; model M has E_M edges and model M' has $E_{M'}$ edges, and $E_M = E_{M'}$.

Given in Equation 4.18 (p. 80), the proposal ratio for the reversal of any trait-to-trait effect β_{st} to β'_{ts} is:

$$
\begin{aligned}
\frac{q(\mathbf{u}'_M, \mathbf{S}, x' | \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x | \Omega, \mathbf{S})} &= \frac{q(\mathbf{u}'_M | \mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'} | \mathbf{S}', x, \Omega, \mathbf{S})} \times \frac{|Ne(M, x = x^S)|}{|Ne(M', x' = x^S)|} \\
&= \frac{q(\beta_{0t} | \widehat{\beta}_{0t}, \widehat{\sigma}_t^2, V_{\beta_{0t}})}{q(\beta'_{0t} | \widehat{\beta}'_{0t}, \widehat{\sigma}_t'^2, V_{\beta'_{0t}})} \frac{q(\beta_{0s} | \widehat{\beta}_{0s}, \widehat{\sigma}_s^2, V_{\beta_{0s}})}{q(\beta'_{0s} | \widehat{\beta}'_{0s}, \widehat{\sigma}_s'^2, V_{\beta'_{0s}})} \frac{q(\beta_{st} | \widehat{\beta}_{st}, \widehat{\sigma}_t^2, V_{\beta_{st}})}{q(\beta'_{ts} | \widehat{\beta}'_{ts}, \widehat{\sigma}_s'^2, V_{\beta'_{ts}})} \times \\
&\qquad \frac{q(\sigma_s^2 | \widehat{\sigma}_s^2)}{q(\sigma_s'^2 | \widehat{\sigma}_s'^2)} \frac{\sum_{g=1}^{E_M} I_a(S_g[M])}{\sum_{g=1}^{E_{M'}} I_a(S_g[M'])}.
\end{aligned}
\tag{4.32}
$$

Using the indicator function $I_a(M)$ as defined in Equation 4.3 (p. 72), with $S_g[M]$ denoting the model with edge g reversed (in $\mathbf{S}$) as stated in Section 4.1.1, we define the size of the neighbourhoods as:

$$
|Ne(M, x = x^S)| = \sum_{g=1}^{E_M} I_a(S_g[M])
$$

and

$$
|Ne(M', x' = x^S)| = \sum_{g=1}^{E_{M'}} I_a(S_g[M']).
$$

The parameters which get updated differ depending on the proposed edge to reverse. This is due to conditional independence, whereby a trait t is independent of all other traits given the traits and QTL with direct effects on trait t ($\upsilon_Y(t)$ and $\upsilon_Q(t)$, respectively). Therefore, if we propose the reversal of the edge corre-

sponding to the trait-to-trait effect β_{st}, we generate β'_{ts}, in addition to β'_{0t}, β'_{0s}, σ'^2_t, and σ'^2_s for model M'. These proposal distributions are given in Equations 4.12 – 4.14 (p. 79), and the proposal distributions for $M' \to M$ are given in Equations 4.15 – 4.17 (p. 80). For the reversal of the edge corresponding to β_{st}, we have the following ratio of priors:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{0t}|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})} \frac{p(\beta'_{0s}|\mathbf{S}')}{p(\beta_{0s}|\mathbf{S})} \frac{p(\beta'_{ts}|\mathbf{S}')}{p(\beta_{st}|\mathbf{S})} \frac{p(\sigma'^2_t|\mathbf{S}')}{p(\sigma^2_t|\mathbf{S})} \frac{p(\sigma'^2_s|\mathbf{S}')}{p(\sigma^2_s|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})}, \qquad (4.33)$$

where the parameter prior distributions are defined in Equation 4.7 (p. 73), and if we assume a uniform prior on all possible states for the connection between any two trait nodes, $p(\mathbf{S}) = p(\mathbf{S}')$.

Following the general form of the acceptance probability given in Equation 4.2 (p. 70), and substituting in the proposal and prior ratios in Equations 4.32 and 4.33, the acceptance probability for a reverse step is:

$$
\begin{aligned}
\alpha_{reverse} &= min(1, r_{reverse}), \quad \text{where} \\
r_{reverse} &= \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})} \frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \frac{p(\mathbf{S}')}{p(\mathbf{S})} \frac{q(\mathbf{u}'_M|\mathbf{S}, x' = x^S, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x = x^S, \Omega, \mathbf{S})} \times \\
&\quad \frac{q(\mathbf{S}|x' = x^S, \Omega', \mathbf{S}')}{q(\mathbf{S}'|x = x^S, \Omega, \mathbf{S})} \frac{q(x' = x^S|\Omega', \mathbf{S}')}{q(x = x^S|\Omega, \mathbf{S})} \left| \frac{\partial(\Omega', \mathbf{u}'_M)}{\partial(\Omega, \mathbf{u}_{M'})} \right| \\
&= \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})} \frac{p(\beta'_{0t}|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})} \frac{p(\beta'_{0s}|\mathbf{S}')}{p(\beta_{0s}|\mathbf{S})} \frac{p(\beta'_{ts}|\mathbf{S}')}{p(\beta_{st}|\mathbf{S})} \frac{p(\sigma'^2_t|\mathbf{S}')}{p(\sigma^2_t|\mathbf{S})} \frac{p(\sigma'^2_s|\mathbf{S}')}{p(\sigma^2_s|\mathbf{S})} \times \\
&\quad \frac{p(\mathbf{S}')}{p(\mathbf{S})} \frac{q(\beta_{0t}|\widehat{\beta}_{0t}, \widehat{\sigma}'^2_t, V_{\beta_{0t}})}{q(\beta'_{0t}|\widehat{\beta}'_{0t}, \widehat{\sigma}'^2_t, V_{\beta'_{0t}})} \frac{q(\beta_{0s}|\widehat{\beta}_{0s}, \widehat{\sigma}^2_s, V_{\beta_{0s}})}{q(\beta'_{0s}|\widehat{\beta}'_{0s}, \widehat{\sigma}'^2_s, V_{\beta'_{0s}})} \frac{q(\beta_{st}|\widehat{\beta}_{st}, \widehat{\sigma}'^2_t, V_{\beta_{st}})}{q(\beta'_{ts}|\widehat{\beta}'_{ts}, \widehat{\sigma}'^2_s, V_{\beta'_{ts}})} \times \\
&\quad \frac{q(\sigma^2_t|\widehat{\sigma}^2_t)}{q(\sigma'^2_t|\widehat{\sigma}'^2_t)} \frac{q(\sigma^2_s|\widehat{\sigma}^2_s)}{q(\sigma'^2_s|\widehat{\sigma}'^2_s)} \frac{\sum_{g=1}^{E_M} I_a(S_g[M])}{\sum_{g=1}^{E_{M'}} I_a(S_g[M'])}. \qquad (4.34)
\end{aligned}
$$

The general expression of the likelihood is given in Equation 4.5 (p. 72). Assuming a uniform prior on the state of a connection between any two trait nodes means that $p(\mathbf{S}') = p(\mathbf{S})$; however, it does imply that an edge is more likely to be

included in the phenotype network structure (in one of two directions) than not.

The Jacobian is 1 (see Appendix C.1). Reversing an edge within a phenotype network structure can create cyclic graphs, therefore the terms $|Ne(M, x = x^S)|$ and $|Ne(M', x' = x^S)|$ may differ.

4.3.1 Example of a reverse step

Suppose that we want to carry out a reverse step to move from model M to M'; the example causal network structures are given in Figure 4.3. We propose to reverse the trait-to-trait effect of trait 2 on trait 3 (where this directed edge was selected randomly with probability $\frac{1}{3}$). We note that we cannot propose the reversal of β_{13} as this would create a cyclic model. However, we could have proposed to reverse β_{12}.

In this example we have a set of current parameters, Ω, and a set of candidate
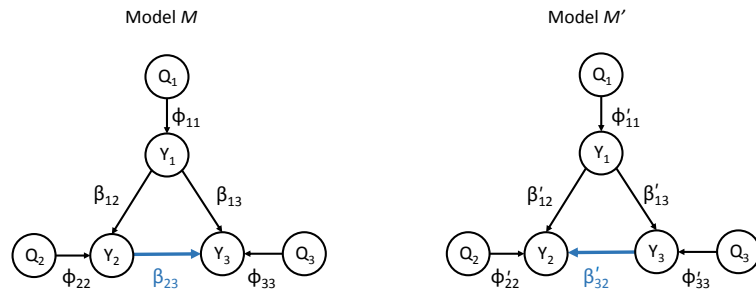
Figure 4.3: *Example causal network structures for both the current (M) and candidate (M') models, given a reverse step. Here, we propose to reverse the trait-to-trait effect of trait 2 on trait 3, i.e. delete the parameter β_{23} and add the parameter β'_{32}.*

parameters, Ω':

$$\Omega = \{\beta_{01}, \beta_{02}, \beta_{03}, \beta_{12}, \beta_{13}, \beta_{23}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma_1^2, \sigma_2^2, \sigma_3^2\}$$
$$\Omega' = \{\beta'_{01}, \beta'_{02}, \beta'_{03}, \beta'_{12}, \beta'_{13}, \beta'_{32}, \phi'_{11}, \phi'_{22}, \phi'_{33}, \sigma_1'^2, \sigma_2'^2, \sigma_3'^2\}.$$

Here we update the following parameters which are affected by the reversal of the edge corresponding to β_{23}: β'_{02}, β'_{03} β'_{32}, $\sigma_2'^2$, and $\sigma_3'^2$. It follows that

$$\{\beta'_{01}, \beta'_{12}, \beta'_{13}, \phi'_{11}, \phi'_{22}, \phi'_{33}, \sigma_1'^2\} = \{\beta_{01}, \beta_{12}, \beta_{13}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma_1^2\}.$$

To move from M to M', new parameters are generated from the following proposal distributions:

$$\beta'_{0t} | \widehat{\beta}'_{0t}, \widehat{\sigma}_t'^2, V_{\beta'_{0t}} \quad \sim \quad N\left(\widehat{\beta}'_{0t}, \frac{1}{\tau}\widehat{\sigma}_t'^2 V_{\beta'_{0t}}\right) \quad \text{for } t = 2, 3 \qquad \text{(see Eqn. 4.12)}$$

$$\beta'_{32} | \widehat{\beta}'_{32}, \widehat{\sigma}_2'^2, V_{\beta'_{32}} \quad \sim \quad N\left(\widehat{\beta}'_{32}, \frac{1}{\tau}\widehat{\sigma}_2'^2 V_{\beta'_{32}}\right) \qquad \text{(see Eqn. 4.13)}$$

$$\sigma_t'^2 | \widehat{\sigma}_t'^2 \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}_t'^2, 1.5\widehat{\sigma}_t'^2\right) \quad \text{for } t = 2, 3 \qquad \text{(see Eqn. 4.14)}.$$

where for parameter p, V_p is as defined in Equation 4.11 (p. 78).

To move from model M' to M, the parameter proposal distributions are:

$$\beta_{0t} | \widehat{\beta}_{0t}, \widehat{\sigma}_t^2, V_{\beta_{0t}} \quad \sim \quad N\left(\widehat{\beta}_{0t}, \frac{1}{\tau}\widehat{\sigma}_t^2 V_{\beta_{0t}}\right) \quad \text{for } t = 2, 3 \qquad \text{(see Eqn. 4.15)}$$

$$\beta_{23} | \widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}} \quad \sim \quad N\left(\widehat{\beta}_{23}, \frac{1}{\tau}\widehat{\sigma}_3^2 V_{\beta_{23}}\right) \qquad \text{(see Eqn. 4.16)}$$

$$\sigma_t^2 | \widehat{\sigma}_t^2 \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}_t^2, 1.5\widehat{\sigma}_t^2\right) \quad \text{for } t = 2, 3 \qquad \text{(see Eqn. 4.17)}.$$

The neighbourhoods for both models M and M', given a reverse step, are dis-

played in Figure 4.4, where

$$|Ne(M, x = x^S)| = |Ne(M', x' = x^S)| = 2.$$

As mentioned earlier, we cannot reverse the trait-to-trait effect β_{13} in model M, as it would create a cyclic model, so it is not included in the neighbourhood $Ne(M, x = x^S)$. Similarly, we cannot reverse the trait-to-trait effect β_{12} in model M'.

It follows that the proposal ratio for this example is:

$$\frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \times \frac{|Ne(M, x = x^S)|}{|Ne(M', x' = x^S)|} = \frac{q(\beta_{02}|\widehat{\beta}_{02}, \widehat{\sigma}_2^2, V_{\beta_{02}})}{q(\beta'_{02}|\widehat{\beta}'_{02}, \widehat{\sigma}_2'^2, V_{\beta'_{02}})} \frac{q(\beta_{03}|\widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}})}{q(\beta'_{03}|\widehat{\beta}'_{03}, \widehat{\sigma}_3'^2, V_{\beta'_{03}})} \times$$

$$\frac{q(\beta_{23}|\widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}})q(\sigma_2^2|\widehat{\sigma}_2^2)q(\sigma_3^2|\widehat{\sigma}_3^2)}{q(\beta'_{32}|\widehat{\beta}'_{32}, \widehat{\sigma}_2'^2, V_{\beta'_{32}})q(\sigma_2'^2|\widehat{\sigma}_2'^2)q(\sigma_3'^2|\widehat{\sigma}_3'^2)}.$$
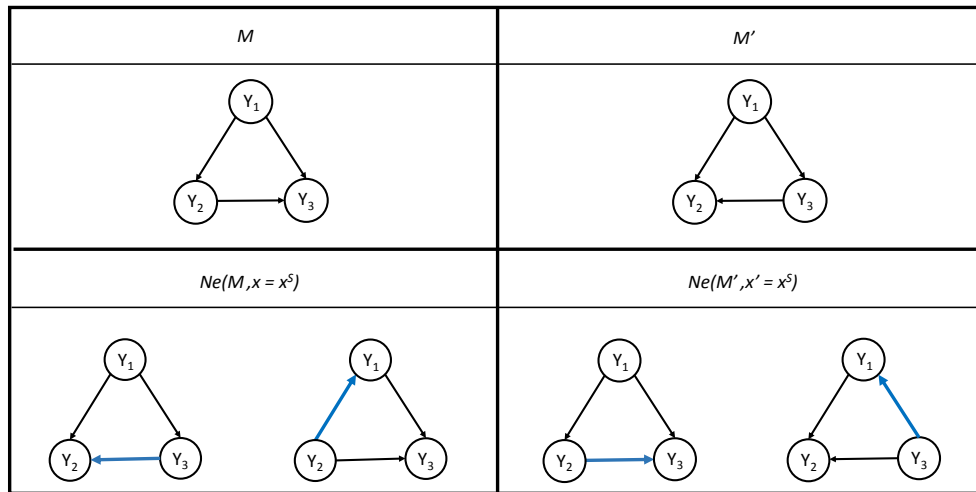
Figure 4.4: *The neighbourhoods for the current and candidate models, given a reverse step; denoted* $Ne(M, x = x^S)$ *and* $Ne(M', x' = x^S)$, *respectively.*

The prior ratio is:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{02}|\mathbf{S}')p(\beta'_{03}|\mathbf{S}')p(\beta'_{32}|\mathbf{S}')p(\sigma_2'^2|\mathbf{S}')p(\sigma_3'^2|\mathbf{S}')}{p(\beta_{02}|\mathbf{S})p(\beta_{03}|\mathbf{S})p(\beta_{23}|\mathbf{S})p(\sigma_2^2|\mathbf{S})p(\sigma_3^2|\mathbf{S})},$$

where the parameter prior distributions are defined in Equation 4.7 (p. 73), and we assume a uniform prior on all possible states for the connection between any two trait nodes, such that $p_{g1} = p_{g2} = p_{g3} = \frac{1}{3}$ and $p(\mathbf{S}) = p(\mathbf{S}')$.

It follows that the acceptance probability for this example reverse step is:

$$\alpha_{reverse} = min(1, r_{reverse}), \quad \text{where}$$

$$r_{reverse} = \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})} \frac{p(\beta'_{02}|\mathbf{S}')p(\beta'_{03}|\mathbf{S}')p(\beta'_{32}|\mathbf{S}')p(\sigma_2'^2|\mathbf{S}')p(\sigma_3'^2|\mathbf{S}')}{p(\beta_{02}|\mathbf{S})p(\beta_{03}|\mathbf{S})p(\beta_{23}|\mathbf{S})p(\sigma_2^2|\mathbf{S})p(\sigma_3^2|\mathbf{S})} \times$$

$$\frac{q(\beta_{02}|\widehat{\beta}_{02}, \widehat{\sigma}_2^2, V_{\beta_{02}})}{q(\beta'_{02}|\widehat{\beta}'_{02}, \widehat{\sigma}_2'^2, V_{\beta'_{02}})} \frac{q(\beta_{03}|\widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}})}{q(\beta'_{03}|\widehat{\beta}'_{03}, \widehat{\sigma}_3'^2, V_{\beta'_{03}})} \times$$

$$\frac{q(\beta_{23}|\widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}})q(\sigma_2^2|\widehat{\sigma}_2^2)q(\sigma_3^2|\widehat{\sigma}_3^2)}{q(\beta'_{32}|\widehat{\beta}'_{32}, \widehat{\sigma}_2'^2, V_{\beta'_{32}})q(\sigma_2'^2|\widehat{\sigma}_2'^2)q(\sigma_3'^2|\widehat{\sigma}_3'^2)}.$$

The general expression of the likelihood is given in Equation 4.5 (p. 72), and the Jacobian is 1 (see Appendix C.1).

4.4   The relocate step

A relocate step proposes the relocation of an existing directed edge, or trait-to-trait effect, within model M to form the model M'. The relocation of a directed edge involves the deletion of one of the existing edges within model M, and the addition of an edge in a new, distinct location such that both M and M' have phenotype network structures with the same number of directed edges; model M has E_M edges and model M' has $E_{M'}$ edges, and $E_M = E_{M'}$.

Given in Equation 4.18 (p. 80), the proposal ratio for the relocation of any trait-to-trait effect β_{st} to β'_{vw} is:

$$
\frac{q(\mathbf{u}'_M, \mathbf{S}, x'|\Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x|\Omega, \mathbf{S})} = \frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \times \frac{|Ne(M, x = x^C)|}{|Ne(M', x' = x^C)|}
$$

$$
= \frac{q(\beta_{0t}|\widehat{\beta}_{0t}, \widehat{\sigma}_t^2, V_{\beta_{0t}})}{q(\beta'_{0t}|\widehat{\beta}'_{0t}, \widehat{\sigma}_t'^2, V_{\beta'_{0t}})} \frac{q(\beta_{0w}|\widehat{\beta}_{0w}, \widehat{\sigma}_w^2, V_{\beta_{0w}})}{q(\beta'_{0w}|\widehat{\beta}'_{0w}, \widehat{\sigma}_w'^2, V_{\beta'_{0w}})} \frac{q(\beta_{st}|\widehat{\beta}_{st}, \widehat{\sigma}_t^2, V_{\beta_{st}})}{q(\beta'_{vw}|\widehat{\beta}'_{vw}, \widehat{\sigma}_w'^2, V_{\beta'_{vw}})} \times
$$

$$
\frac{q(\sigma_t^2|\widehat{\sigma}_t'^2)}{q(\sigma_t'^2|\widehat{\sigma}_t'^2)} \frac{q(\sigma_w^2|\widehat{\sigma}_w^2)}{q(\sigma_w'^2|\widehat{\sigma}_w'^2)} \frac{\displaystyle\sum_{g=1}^{E_M} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M]) I_R(g, f|\mathbf{S}) I_{\in}(f|\mathbf{S})}{\displaystyle\sum_{g=1}^{E_{M'}} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M']) I_R(g, f|\mathbf{S}') I_{\in}(f|\mathbf{S}')}
$$

$$(4.35)$$

Using the indicator functions are given in Equation 4.3 (p. 72), with $C_{g,f}[M]$ denotes the new model with edge g relocated to location f (in $\mathbf{S}$), as stated in Section 4.1.1, we define the size of the neighbourhoods as:

$$
|Ne(M, x = x^C)| = \sum_{g=1}^{E_M} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M]) I_R(g, f|\mathbf{S}) I_{\in}(f|\mathbf{S})
$$

and

$$
|Ne(M', x' = x^C)| = \sum_{g=1}^{E_{M'}} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M']) I_R(g, f|\mathbf{S}') I_{\in}(f|\mathbf{S}').
$$

It follows that the neighbourhood of a model, given a relocate step, includes only acyclic models where the relocated edge is not reversed, but relocated to a new, distinct position within the phenotype network.

The parameters which get updated differ depending on the relocated edge — both its position within model M and model M'. This is due to conditional independence, whereby trait t is independent of all other traits given the traits and QTL with direct effects on trait t ($\upsilon_Y(t)$ and $\upsilon_Q(t)$, respectively). Therefore, if we propose the relocation of β_{st} to β'_{vw}, we generate β'_{vw} in addition to β'_{0t}, β'_{0w}, $\sigma_t'^2$,

and $\sigma_w'^2$ for model M', provided $t \neq w$. We note that the edge β_{st} can be relocated to the location β_{vw} such that $t = w$, in which case we would only need to generate the parameter β_{wt}' in addition to β_{0t}' and $\sigma_t'^2$, for model M'. These proposal distributions are given in Equations 4.12 – 4.14 (p. 80), and the proposal distributions for $M' \to M$ are given in Equations 4.15 – 4.17 (p. 80). It follows that for the relocation of β_{st} to β_{vw}', where $t \neq w$, we have the following ratio of priors:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta_{0t}'|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})} \frac{p(\beta_{0w}'|\mathbf{S}')}{p(\beta_{0w}|\mathbf{S}} \frac{p(\beta_{vw}'|\mathbf{S}')}{p(\beta_{st}|\mathbf{S})} \frac{p(\sigma_t'^2|\mathbf{S}')}{p(\sigma_t^2|\mathbf{S})} \frac{p(\sigma_w'^2|\mathbf{S}')}{p(\sigma_w^2|\mathbf{S})} \frac{p(\mathbf{S}')}{p(\mathbf{S})} \qquad (4.36)$$

where the parameter prior distributions are defined in Equation 4.7 (p. 73) and if we assume a uniform prior on all possible states for the connection between any two trait nodes, $p(\mathbf{S}) = p(\mathbf{S}')$.

Following the general form of the acceptance probability in Equation 4.2 (p. 70), and substituting the proposal and prior ratios in equations 4.35 and 4.36, the acceptance probability for a relocate step is:

$$\alpha_{relocate} = min(1, r_{relocate}), \quad \text{where}$$

$$r_{relocate} = \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})} \frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \frac{p(\mathbf{S}')}{p(\mathbf{S})} \frac{q(\mathbf{u}_M'|\mathbf{S},x'=x^C,\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x=x^C,\Omega,\mathbf{S})} \times$$

$$\frac{q(\mathbf{S}|x'=x^C,\Omega',\mathbf{S}')}{q(\mathbf{S}'|x=x^C,\Omega,\mathbf{S})} \frac{q(x'=x^C|\Omega',\mathbf{S}')}{q(x=x^C|\Omega,\mathbf{S})} \left| \frac{\partial(\Omega',\mathbf{u}_M')}{\partial(\Omega,\mathbf{u}_{M'})} \right|$$

$$= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})} \frac{p(\beta_{0t}'|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})} \frac{p(\beta_{0w}'|\mathbf{S}')}{p(\beta_{0w}|\mathbf{S})} \frac{p(\beta_{vw}'|\mathbf{S}')}{p(\beta_{st}|\mathbf{S})} \frac{p(\sigma_t'^2|\mathbf{S}')}{p(\sigma_t^2|\mathbf{S})} \frac{p(\sigma_w'^2|\mathbf{S}')}{p(\sigma_w^2|\mathbf{S})} \times$$

$$\frac{p(\mathbf{S}')}{p(\mathbf{S})} \frac{q(\beta_{0t}|\widehat{\beta}_{0t},\widehat{\sigma}_t^2,V_{\beta_{0t}})}{q(\beta_{0t}'|\widehat{\beta}_{0t}',\widehat{\sigma}_t'^2,V_{\beta_{0t}'})} \frac{q(\beta_{0w}|\widehat{\beta}_{0w},\widehat{\sigma}_w^2,V_{\beta_{0w}})}{q(\beta_{0w}'|\widehat{\beta}_{0w}',\widehat{\sigma}_w'^2,V_{\beta_{0w}'})} \frac{q(\beta_{st}|\widehat{\beta}_{st},\widehat{\sigma}_t^2,V_{\beta_{st}})}{q(\beta_{vw}'|\widehat{\beta}_{vw}',\widehat{\sigma}_w'^2,V_{\beta_{vw}'})} \times$$

$$\frac{q(\sigma_t^2|\widehat{\sigma}_t'^2)}{q(\sigma_t'^2|\widehat{\sigma}_t'^2)} \frac{q(\sigma_w^2|\widehat{\sigma}_w^2)}{q(\sigma_w'^2|\widehat{\sigma}_w'^2)} \frac{\displaystyle\sum_{g=1}^{E_M} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M])I_R(g,f|\mathbf{S})I_{\in}(f|\mathbf{S})}{\displaystyle\sum_{g=1}^{E_{M'}} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M'])I_R(g,f|\mathbf{S}')I_{\in}(f|\mathbf{S}')}. \qquad (4.37)$$

 The general expression of the likelihood is given in Equation 4.5 (p. 72). Assuming a uniform prior on the state of a connection between any two trait nodes means that $p(\mathbf{S}) = p(\mathbf{S}')$; however, it does imply that an edge is more likely to be included in the phenotype network structure (in one of two directions) than not. The phenotype network prior is as defined in Equation 4.10 (p. 75), and the Jacobian is 1 (see Appendix C.2).

4.4.1 Example of a relocate step

Suppose that we want to carry out a relocation step to move from model M to M'; the example causal network structures are given in Figure 4.5. The edge β_{12} was randomly selected to be relocated with probability $\frac{1}{2}$. We could have also proposed the relocation of β_{13}. In this example we have a set of current parameters, Ω, and a set of candidate parameters, Ω':

$$
\begin{aligned}
\Omega &= \{\beta_{01}, \beta_{02}, \beta_{03}, \beta_{12}, \beta_{13}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma_1^2, \sigma_2^2, \sigma_3^2\} \\
\Omega' &= \{\beta_{01}', \beta_{02}', \beta_{03}', \beta_{13}', \beta_{23}', \phi_{11}', \phi_{22}', \phi_{33}', \sigma_1'^2, \sigma_2'^2, \sigma_3'^2\}.
\end{aligned}
$$
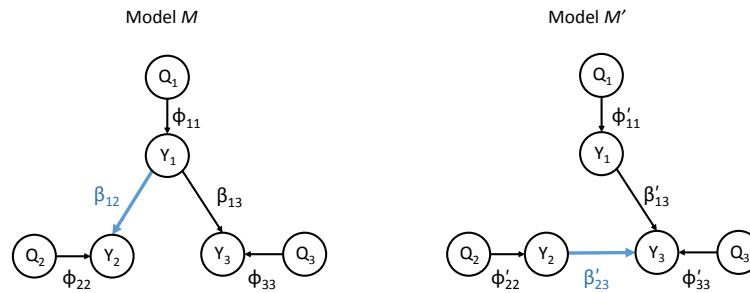
Figure 4.5: *Example causal network structures for both the current (M) and candidate (M′) models, given a relocate step. Here, we propose to remove the trait-to-trait effect β_{12} and "relocate" the directed edge such that we add the trait-to-trait effect β_{23}'.*

Both β_{12} and β'_{23} specify which parameters need to be updated in the candidate model. Here we propose new values for the following parameters; β'_{02}, β'_{03}, β'_{23}, σ'^2_2, σ'^2_3. It follows that

$$\{\beta'_{01}, \beta'_{13}, \phi'_{11}, \phi'_{22}, \phi'_{33}, \sigma'^2_1\} = \{\beta_{01}, \beta_{13}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma^2_1\}.$$

To move from model M to M', new parameters are generated from the following proposal distributions:

$$\beta'_{0t} | \widehat{\beta}'_{0t}, \sigma'^2_t, V_{\beta'_{0t}} \quad \sim \quad N\left(\widehat{\beta}'_{0t}, \frac{1}{\tau}\sigma'^2_t V_{\beta'_{0t}}\right) \quad \text{for } t = 2,3 \qquad \text{(see Eqn. 4.12)}$$

$$\beta'_{23} | \widehat{\beta}'_{23}, \sigma'^2_3, V_{\beta'_{23}} \quad \sim \quad N\left(\widehat{\beta}'_{23}, \frac{1}{\tau}\sigma'^2_3 V_{\beta'_{23}}\right) \qquad \text{(see Eqn. 4.13)}$$

$$\sigma'^2_t | \widehat{\sigma}'^2_t \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}'^2_t, 1.5\widehat{\sigma}'^2_t\right) \quad \text{for } t = 2,3 \qquad \text{(see Eqn. 4.14)},$$

where for parameter p, V_p is as defined in Equation 4.11 (p. 78).

To move from model M' to model M, the parameter proposal distributions are:

$$\beta_{0t} | \widehat{\beta}_{0t}, \sigma^2_t, V_{\beta_{0t}} \quad \sim \quad N\left(\widehat{\beta}_{0t}, \frac{1}{\tau}\sigma^2_t V_{\beta_{0t}}\right) \quad \text{for } t = 2,3 \qquad \text{(see Eqn. 4.15)}$$

$$\beta_{12} | \widehat{\beta}_{12}, \sigma^2_2, V_{\beta_{12}} \quad \sim \quad N\left(\widehat{\beta}_{12}, \frac{1}{\tau}\sigma^2_2 V_{\beta_{12}}\right) \qquad \text{(see Eqn. 4.16)}$$

$$\sigma^2_t | \widehat{\sigma}^2_t \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}^2_t, 1.5\widehat{\sigma}^2_t\right) \quad \text{for } t = 2,3 \qquad \text{(see Eqn. 4.17)}.$$

The neighbourhoods for both models M and M', given a relocation step, are displayed in Figure 4.6, where

$$|Ne(M, x = x^C)| = |Ne(M', x' = x^C)| = 4.$$
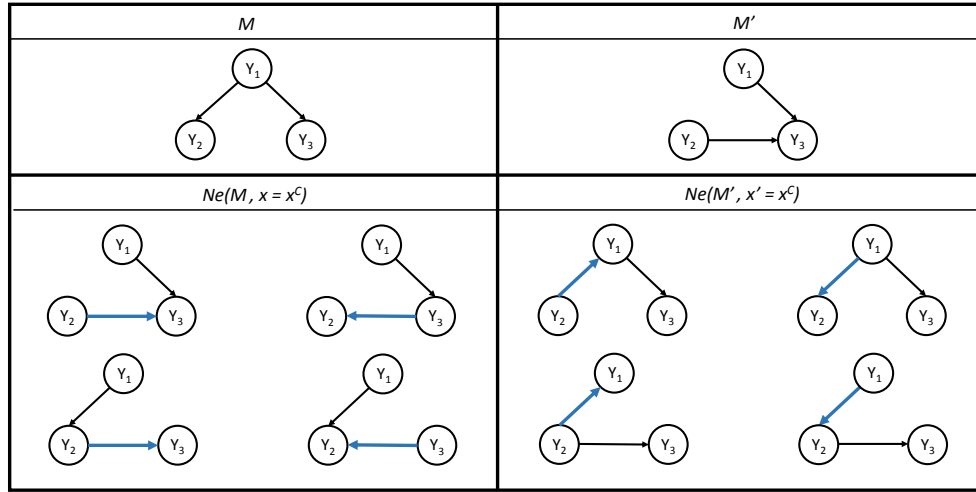
Figure 4.6: *The neighbourhoods for both the current and candidate models, given a relocation step; denoted $Ne(M, x = x^C)$ and $Ne(M', x' = x^C)$, respectively.*

It follows that the proposal ratio for this example is:

$$
\frac{q(\mathbf{u}'_M | \mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'} | \mathbf{S}', x, \Omega, \mathbf{S})} \times \frac{|Ne(M, x = x^C)|}{|Ne(M', x' = x^C)|} = \frac{q(\beta_{02} | \widehat{\beta}_{02}, \sigma_2^2, V_{\beta_{02}})}{q(\beta'_{02} | \widehat{\beta}'_{02}, \sigma_2'^2, V_{\beta'_{02}})} \frac{q(\beta_{03} | \widehat{\beta}_{03}, \sigma_3^2, V_{\beta_{03}})}{q(\beta'_{03} | \widehat{\beta}'_{03}, \sigma_3'^2, V_{\beta'_{03}})} \times
$$

$$
\frac{q(\beta_{12} | \widehat{\beta}_{12}, \sigma_2^2, V_{\beta_{12}})}{q(\beta'_{23} | \widehat{\beta}'_{23}, \sigma_3'^2, V_{\beta'_{23}})} \frac{q(\sigma_2^2 | \widehat{\sigma}_2^2)}{q(\sigma_2'^2 | \widehat{\sigma}_2'^2)} \frac{q(\sigma_3^2 | \widehat{\sigma}_3^2)}{q(\sigma_3'^2 | \widehat{\sigma}_3'^2)}.
$$

The prior ratio is:

$$
\frac{p(\Omega' | \mathbf{S}')}{p(\Omega | \mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{02} | \mathbf{S}') p(\beta'_{03} | \mathbf{S}') p(\beta'_{23} | \mathbf{S}') p(\sigma_2'^2 | \mathbf{S}') p(\sigma_3'^2 | \mathbf{S}')}{p(\beta_{02} | \mathbf{S}) p(\beta_{03} | \mathbf{S}) p(\beta_{12} | \mathbf{S}) p(\sigma_2^2 | \mathbf{S}) p(\sigma_3^2 | \mathbf{S})},
$$

where the parameter prior distributions are defined in Equation 4.7 (p. 73), and we assume a uniform prior on all possible states for the connection between any two trait nodes, such that $p_{g1} = p_{g2} = p_{g3} = \frac{1}{3}$ and $p(\mathbf{S}) = p(\mathbf{S}')$. It follows that

the acceptance probability for this example relocate step is:

$$
\begin{aligned}
\alpha_{relocate} &= min(1, r_{relocate}), \quad \text{where} \\
r_{relocate} &= \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})} \frac{p(\beta'_{02}|\mathbf{S}')p(\beta'_{03}|\mathbf{S}')p(\beta'_{23}|\mathbf{S}')p(\sigma_2'^2|\mathbf{S}')p(\sigma_3'^2|\mathbf{S}')}{p(\beta_{02}|\mathbf{S})p(\beta_{03}|\mathbf{S})p(\beta_{12}|\mathbf{S})p(\sigma_2^2|\mathbf{S})p(\sigma_3^2|\mathbf{S})} \times \\
&\quad \frac{q(\beta_{02}|\widehat{\beta}_{02}, \sigma_2^2, V_{\beta_{02}})}{q(\beta'_{02}|\widehat{\beta}'_{02}, \sigma_2'^2, V_{\beta'_{02}})} \frac{q(\beta_{03}|\widehat{\beta}_{03}, \sigma_3^2, V_{\beta_{03}})}{q(\beta'_{03}|\widehat{\beta}'_{03}, \sigma_3'^2, V_{\beta'_{03}})} \times \\
&\quad \frac{q(\beta_{12}|\widehat{\beta}_{12}, \sigma_2^2, V_{\beta_{12}})}{q(\beta'_{23}|\widehat{\beta}'_{23}, \sigma_3'^2, V_{\beta'_{23}})} \frac{q(\sigma_2^2|\widehat{\sigma}_2^2)}{q(\sigma_2'^2|\widehat{\sigma}_2'^2)} \frac{q(\sigma_3^2|\widehat{\sigma}_3^2)}{q(\sigma_3'^2|\widehat{\sigma}_3'^2)}
\end{aligned}
$$

The likelihood is given in Equation 4.5 (p. 72), and the Jacobian is 1 (see Appendix C.2).

4.5 Summary of the Bayesian approach to estimating an unknown causal structure within a fixed dimension

1. Initialise the chain by specifying an initial phenotype network structure by sampling from the edge priors and confirming acyclicity. Generate new parameter values using the proposal distributions in Equations 4.22 – 4.25 (pp. 84 – 85), using the LSE for the initial model as the current parameter estimates. The initial model is now referred to as the current model, $M = \{\Omega, \mathbf{S}\}$.

2. Select a move type x with probability $q(x|\Omega, \mathbf{S})$, given in Table 4.1 (p. 71), where

$$
\sum_{x \in \mathbb{X}} q(x|\Omega, \mathbf{S}) = 1 \quad \text{for all } (\Omega, \mathbf{S})
$$

and x is one of:

- **update** (x^U): update all parameters in the current model

- **reverse** (x^S): reverse the direction of one of the directed edges in the current phenotype network

- **relocate** (x^C): relocate one of the directed edges in the current pheno-type network, i.e. delete one edge and propose a new, distinct directed edge

3. Given move type x, propose a candidate phenotype network structure $\mathbf{S}'$ with probability

$$q(\mathbf{S}'|x,\Omega,\mathbf{S}) = \begin{cases} \frac{1}{|Ne(M,x)|}, & \text{if } \mathbf{S}' \in Ne(M,x) \\ 0, & \text{otherwise} \end{cases}$$

where

$$\sum_{\mathbf{S}'\in\mathbb{S}} q(\mathbf{S}'|x,\Omega,\mathbf{S}) = 1 \quad \text{for all } (x,\Omega,\mathbf{S}).$$

4. Generate a set of random variables $\mathbf{u}_{M'} \in \mathbb{U}_{M'}$, with probability density

$$q(\mathbf{u}_{M'}|x,\mathbf{S}',\Omega,\mathbf{S})$$

and the parameters of the new state of the chain, Ω', are generated by the deterministic function $g_{MM'}$:

$$\Omega' = g_{MM'}(\Omega,\mathbf{u}_{M'}).$$

- If $x = x^U$ then the proposal distributions are given in Equations 4.22 – 4.25 (p. 85), dependent on the current parameter estimates (Ω).

- If $x \in \{x^S, x^C\}$ then the phenotype network structure is altered and the parameter proposal distributions are given in Equations 4.12 – 4.14 (pp. 79 – 80).

5. Calculate the acceptance probability:

$$\alpha = min(1, r)$$

where

$$r = \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \times$$

$$\frac{q(\mathbf{S}|x',\Omega',\mathbf{S}')}{q(\mathbf{S}'|x,\Omega,\mathbf{S})} \frac{q(x'|\Omega',\mathbf{S}')}{q(x|\Omega,\mathbf{S})} \left| \frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})} \right|.$$

and

r_{update} is given in Equation 4.31 (p. 86)

$r_{reverse}$ is given in Equation 4.34 (p. 90)

$r_{relocate}$ is given in Equation 4.37 (p. 96).

6. Determine whether to accept or reject the candidate model by drawing a random number from a uniform distribution (between 0 and 1).

- If the random number is less than α, accept the candidate model. The candidate model is now the current model: $M = \{\Omega', \mathbf{S}'\}$.

- If the random number is greater than α, reject the candidate model. The current model remains unchanged: $M = \{\Omega, \mathbf{S}\}$.

7. Repeat steps 2 – 6 until the chain has converged to a stationary state and a large number of samples from the posterior have been drawn.

4.6 Simulation study I

Following Appendix B, data was simulated for five traits and five QTL with the true causal network structure given in Figure 4.7. The true parameter values were varied in order to determine how this approach is influenced by different combinations of QTL and trait-to-trait effects. Both small and large QTL effects and trait-to-trait effects are selected with values between 0.1 and 1.2. This range was based upon those values in published data sets; for example, Dhungana et al. (2007) and Mi et al. (2010). We note that it is possible for the effects to be positive or negative. The true parameter values for data sets 4.1 – 4.4 are the same as those listed in Table 3.4, but are restated here in Table 4.2 for convenience. We simulate $n = 500$ individuals for each data set.

Following the recommendation from (Neto et al. 2010) we ran long chains, with $N = 1,000,000$ iterations. Each had a burn-in period of 100,000 iterations
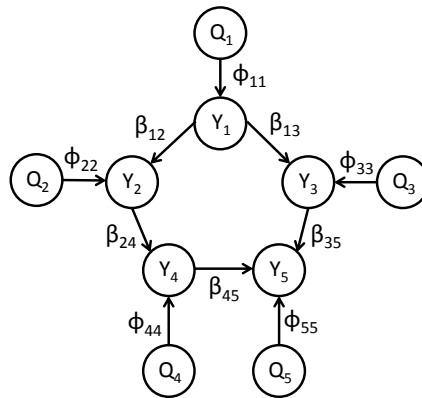
Figure 4.7: *The true causal network structure for the simulated data sets 4.1 – 4.4. The true parameter values for each data set are listed in Table 4.2.*

Table 4.2: *True parameter values for data sets 4.1 – 4.4.*

	data set			
parameter	4.1	4.2	4.3	4.4
β_{01}	0.5	0.5	0.5	0.5
β_{02}	0.5	0.5	0.5	0.5
β_{03}	0.5	0.5	0.5	0.5
β_{04}	0.5	0.5	0.5	0.5
β_{05}	0.5	0.5	0.5	0.5
β_{12}	0.2	1.1	0.2	1.1
β_{13}	0.2	1.1	0.2	1.1
β_{24}	0.2	1.1	0.2	1.1
β_{35}	0.2	1.1	0.2	1.1
β_{45}	0.2	1.1	0.2	1.1
ϕ_{11}	0.1	0.1	1.2	1.2
ϕ_{22}	0.1	0.1	1.2	1.2
ϕ_{33}	0.1	0.1	1.2	1.2
ϕ_{44}	0.1	0.1	1.2	1.2
ϕ_{55}	0.1	0.1	1.2	1.2
σ_1^2	1.0	1.0	1.0	1.0
σ_2^2	1.0	1.0	1.0	1.0
σ_3^2	1.0	1.0	1.0	1.0
σ_4^2	1.0	1.0	1.0	1.0
σ_5^2	1.0	1.0	1.0	1.0

removed and every 10th iteration was retained, so 90,000 states are stored. The results for each data set are given in Sections 4.6.1 – 4.6.4, and the overall results are summarized in Section 4.6.5. We run five Markov chains for each data set, initially presenting the marginal posterior probabilities for each directed edge, as well as the model joint posterior probabilities and model averaged parameter estimates for the first Markov chain run. The results from all five chains are then summarized in Figures 4.12, 4.17, 4.22 and 4.26 for data sets 4.1, 4.2, 4.3 and 4.4, respectively.

The marginal posterior probability of each of the directed edges is calculated

using the following steps, as used by Broom et al. (2012), for example.

1. Identify the K^* different models visited by the Markov chain.

2. For $k = 1, \cdots, K^*$, determine the posterior probability of model M_k given the data (D); denoted $p(M_k|D)$. In other words, determine the proportion of times model M_k appears in the chain.

3. Calculate the posterior probability of each directed edge, e_g, given the data, denoted $p(e_g|D)$:

$$p(e_g|D) = \sum_{k=1}^{K^*} p(e_g|M_k, D)p(M_k|D),$$

where

$$p(e_g|M_k, D) = \begin{cases} 1, & \text{if } e_g \text{ is an edge in model } M_k \\ 0, & \text{otherwise} \end{cases}$$

identifies the estimated models which include the directed edge e_g.

The model averaged estimates for each parameter, say ω, are also presented with their corresponding 95% credible intervals. The model averaged estimates are obtained using Bayesian model averaging, see Hoeting et al. (1999), for example.

1. Identify the K^* different models visited by the Markov chain.

2. For $k = 1, \cdots, K^*$, determine the posterior probability of model M_k given the data, denoted $p(M_k|D)$. That is, determine the proportion of times model M_k appears in the chain.

3. Let the model averaged estimate for parameter ω be denoted $E[\omega|D]$ such that

$$E[\omega|D] = \sum_{k=1}^{K^*} E[\omega|M_k,D]p(M_k|D).$$

where $E[\omega|M_k,D]$ is simply the mean of ω given model M_k, i.e.

$$E[\omega|M_k,D] = \frac{1}{n_k}\sum_{j=1}^{n_k} \omega_j^{(k)}.$$

The number of times model M_k was visited by the chain is denoted n_k, and $\omega_j^{(k)}$ is the estimate of ω for the jth time model M_k was visited by the chain.

The joint posterior probabilities of the models are used to determine the estimated phenotype network structure as we wish to make inference based on the entire phenotype network structure, not just the individual directed edges.

We ran a total of five chains for each data set so that we may comment on numerical convergence. As is standard in MCMC literature, we will refer to this simply as 'convergence'. Trace plots have been plotted for each chain, and for the model with the highest joint posterior probability within each chain. However, as our models contain a large number of parameters, and we are running many chains, we include only a small number of trace plots in Appendix D — selected to best illustrate those trends seen within all the trace plots. We also use Gelman and Rubin's convergence diagnostic, see Gelman & Rubin (1992), on the parameters that are independent of a changing phenotype network structure, as well as on the log likelihood, in order to comment on convergence. Gelman & Rubin (1992) utilize analysis of variance (ANOVA) methods by running many Markov chains and using these as the factor in a one-way analysis. This enables estimates of both

pooled variance and within-chain variance to be calculated, and by comparing the two it can be determined whether or not the chains are from the same (target) distribution.

4.6.1 Data set 4.1:
small QTL effects and small trait-to-trait effects

Data set 4.1 is comprised of small QTL effects and small trait-to-trait effects. As all the effects are small, the true model has a weak signal making it difficult to infer the true phenotype network structure. The marginal posterior probabilities of the directed edges are given in Figures 4.8 and 4.9 for chain 1. We note that with the dimension fixed to include five trait-to-trait effects, if we were to select the edges with the highest marginal posterior probability the estimated phenotype network would be comprised of: β_{21}, β_{24}, β_{31}, β_{35} and β_{54}, with the phenotype network structure given in Figure 4.10 (p. 110). The marginal posterior probabilities for all directed edges in both Figures 4.8 and 4.9 illustrate that with small parameter values for both the QTL and trait-to-trait effects, the true model has a weak signal and we have trouble identifying the direction of causality between traits. However, higher marginal posterior probabilities are associated with edges where a causal relationship does exist, i.e. between traits (Y_1 and Y_2), (Y_1 and Y_3), (Y_2 and Y_4), (Y_3 and Y_5), and (Y_4 and Y_5).

The model averaged parameter estimates, conditional on their existence within the model, are given in Table 4.3 (p. 109). We note that most of the 95% credible intervals contain the true parameter value; those trait-to-trait effects that are in the true model (β_{12}, β_{13}, β_{24}, β_{35}, and β_{45}) all contain the true value of 0.2 within their 95% credible intervals. Conversely, many of the trait-to-trait effect parameters
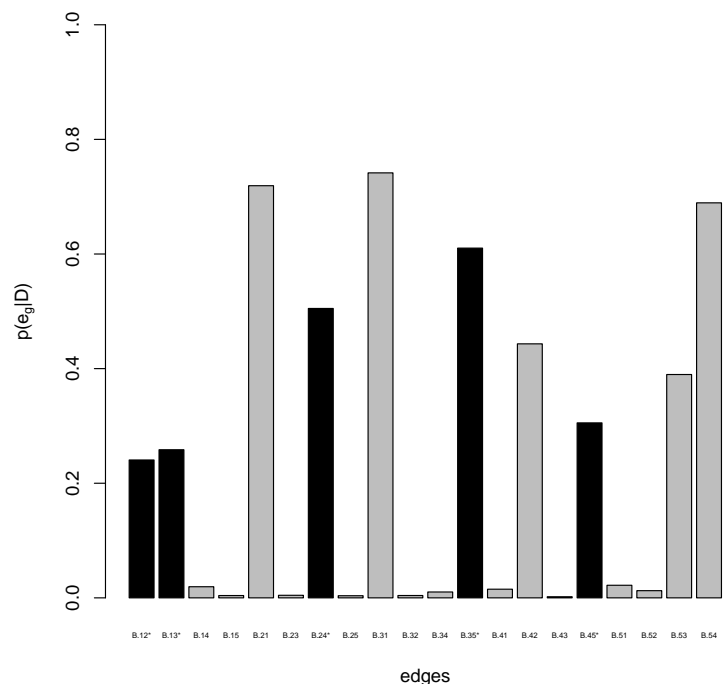
Figure 4.8: *The marginal posterior probability of each directed edge e_g, denoted $p(e_g|D)$ for data set 4.1 (chain 1). The terms "B.ij" denote the trait-to-trait effect β_{ij} corresponding to the directed edge, i.e. the effect of trait i on trait j. The edges comprising the true model are indicated by solid black bars and an "*" next to the label of the directed edge.*

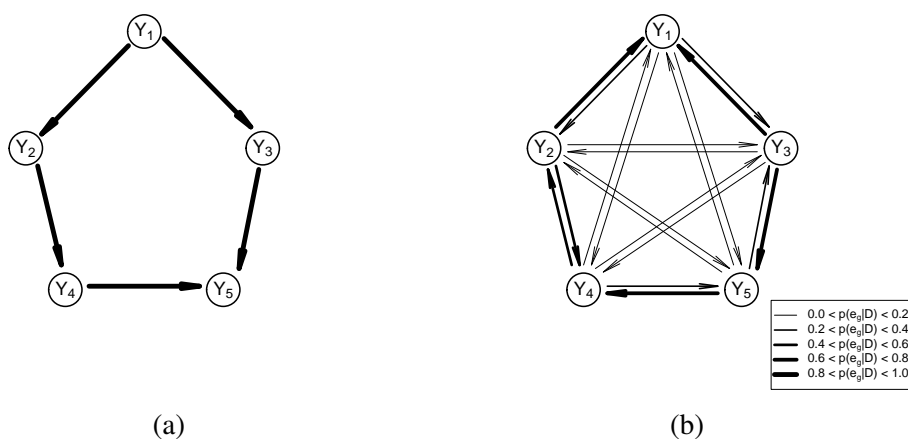Figure 4.9: *For data set 4.1 (chain 1): (a) the true phenotype network structure, and (b) the marginal posterior probability of each directed edge, denoted $p(e_g|D)$, displayed in the context of a phenotype network structure. The thicker the directed edge, the larger the marginal posterior probability.*

Table 4.3: *The model averaged parameter estimates for data set 4.1 (chain 1).*
n(parameter) denotes the number of times the parameter was included in the es-
timated causal network structure. Intervals including the true value are indicated
by an asterisk (*).

| parameter	true value	estimated value	95% CI	n(parameter)
β_{01}	0.5	0.223	$(0.037, 0.451)$	90000
β_{02}	0.5	0.536	$(0.347, 0.713)^*$	90000
β_{03}	0.5	0.551	$(0.350, 0.729)^*$	90000
β_{04}	0.5	0.609	$(0.413, 0.814)^*$	90000
β_{05}	0.5	0.592	$(0.385, 0.812)^*$	90000
β_{12}	0.2	0.133	$(0.051, 0.222)^*$	21652
β_{13}	0.2	0.202	$(0.122, 0.283)^*$	23258
β_{14}	0.0	0.081	$(0.001, 0.169)$	1746
β_{15}	0.0	0.073	$(-0.003, 0.143)^*$	363
β_{21}	0.0	0.144	$(0.060, 0.227)$	64727
β_{23}	0.0	0.004	$(-0.039, 0.059)^*$	401
β_{24}	0.2	0.127	$(0.045, 0.206)^*$	45447
β_{25}	0.0	0.002	$(-0.067, 0.058)^*$	334
β_{31}	0.0	0.232	$(0.139, 0.325)$	66742
β_{32}	0.0	0.047	$(-0.026, 0.127)^*$	368
β_{34}	0.0	0.059	$(-0.023, 0.162)^*$	921
β_{35}	0.2	0.233	$(0.139, 0.329)^*$	54926
β_{41}	0.0	0.078	$(-0.015, 0.182)^*$	1359
β_{42}	0.0	0.152	$(0.062, 0.242)$	39890
β_{43}	0.0	0.018	$(-0.037, 0.073)^*$	174
β_{45}	0.2	0.195	$(0.094, 0.292)^*$	27478
β_{51}	0.0	0.072	$(-0.048, 0.180)^*$	1976
β_{52}	0.0	0.007	$(-0.096, 0.130)^*$	1129
β_{53}	0.0	0.185	$(0.111, 0.263)$	35074
β_{54}	0.0	0.162	$(0.082, 0.246)$	62035
ϕ_{11}	0.1	0.260	$(0.087, 0.425)^*$	90000
ϕ_{22}	0.1	0.087	$(-0.095, 0.271)^*$	90000
ϕ_{33}	0.1	0.049	$(-0.109, 0.214)^*$	90000
ϕ_{44}	0.1	-0.040	$(-0.201, 0.121)^*$	90000
ϕ_{55}	0.1	0.258	$(0.084, 0.440)^*$	90000
σ_1^2	1.0	0.986	$(0.864, 1.125)^*$	90000
σ_2^2	1.0	1.004	$(0.885, 1.138)^*$	90000
σ_3^2	1.0	0.881	$(0.773, 1.004)^*$	90000
σ_4^2	1.0	0.858	$(0.755, 0.978)$	90000
σ_5^2	1.0	1.055	$(0.925, 1.205)^*$	90000
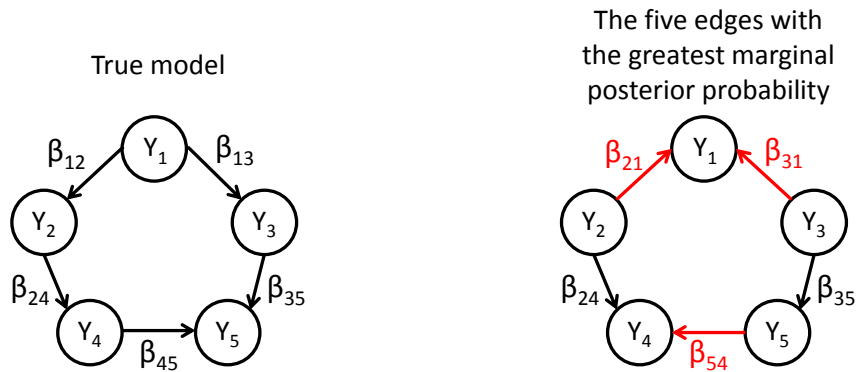
Figure 4.10: *The five edges with the greatest marginal posterior probability for data set 4.1 (chain 1). The edges highlighted in red differ from the true phenotype network structure.*

that are not included in the true model have credible intervals which include 0. However, parameters corresponding to edges which reverse the causal relationship observed in the true model, e.g. β_{21} and β_{42}, have 95% credible intervals that do not include 0.

121 different phenotype network structures were visited and the 10 models with the greatest joint posterior probabilities are given in Figure 4.11. Due to the small size of the effect used the true model has a weak signal, so many models were searched and the joint posterior probabilities are generally quite small (all less than 0.15). We note that the true phenotype network structure is not among the 10 models with the greatest joint posterior probabilities, having a posterior probability of just 0.0073. However, all of the top 10 models have edges where a causal relationship exists, so we may conclude that the direction of causality has not been clearly identified. Interestingly, the model estimated using the marginal posterior probabilities of each edge, given in Figure 4.8 (p. 108), is not the model with the greatest joint posterior probability. Meaning that while the parameter β_{24} was included in more models than β_{42}, we want to estimate the phenotype network as being the combination of edges that have the greatest probability of occurring
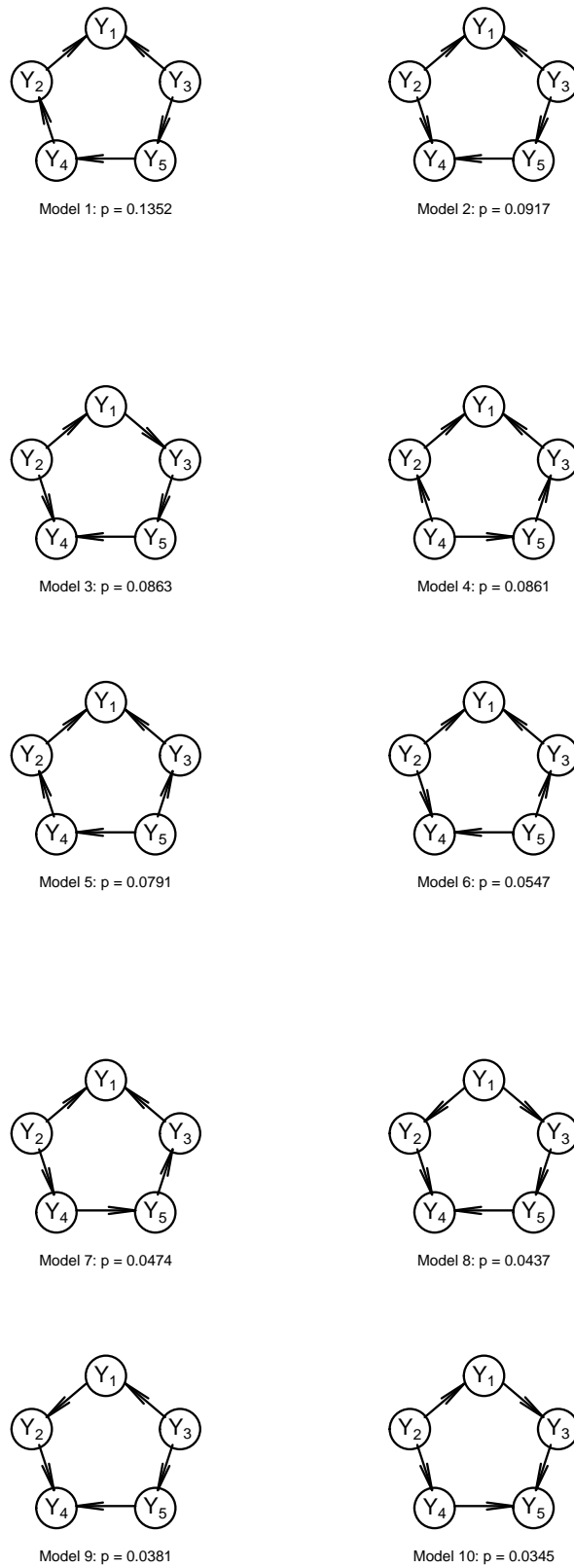
Figure 4.11: *The 10 phenotype networks with the greatest joint posterior probability, p, for data set 4.1 (chain 1). The true model is not included here.*

together, and so we select the model with the greatest joint posterior probability. The estimated phenotype network structure is given as Model 1 in Figure 4.11.

To assess convergence we ran multiple chains, summarized in Figure 4.12 by recording the initial model used by each chain, the number of models visited, the model with the greatest joint posterior probability and the joint posterior probability of the true model. We also include a graphical display of the marginal posterior probability of each directed edge to get an overview of the models visited by the chain, allowing us to get an impression of how the chains are moving.

Each chain has visited many models, incorporating all possible directed edges at some point, as shown by the graphical display of the marginal posterior probabilities of the directed edges, $p(e_g|D)$. The small QTL and trait-to-trait effects mean that the chains have moved between models easily. The true model has been visited in each chain; however, it is never the model with the greatest joint posterior probability, which is not surprising as with small effects the true model has a weak signal. In addition to Figure 4.12, Table 4.4 (p. 114) gives Gelman and Rubin's convergence diagnostic for the dimension independent parameters. The Gelman–Rubin diagnostic is close to 1 for all parameters tested, indicating the dimension independent parameters have converged. However, with such a weak signal there is very little to detect and the chain is just moving around.

The values of the log posterior density are plotted in Figure 4.13 (p. 114) for the five chains. It appears as though all the chains have converged to a stationary state; however, the size of the effects comprising the true model prevent the approach from identifying the correct directions of causality as a change in the phenotype network only results in a minor change in the log posterior density.

Figure 4.12: *A summary of the performance of each chain for data set 4.1. The number of models visited is denoted* $n(models)$, *and the marginal posterior probability of the directed edge* e_g *is denoted* $p(e_g|D)$. *The greatest joint posterior probability is denoted* p^*, *such that model* M_k *has the greatest joint posterior probability when* $p(M_k|D) = p^*$. *The posterior probability of the true model is denoted* $p(M_{TRUE}|D)$.

Table 4.4: *The Gelman–Rubin convergence diagnostic for data set 4.1, for all five chains summarized in Figure 4.12.*

	Gelman–Rubin diagnostic
log likelihood	1.01
β_{01}	1.05
β_{02}	1.03
β_{03}	1.02
β_{04}	1.03
β_{05}	1.04
ϕ_{11}	1.01
ϕ_{22}	1.01
ϕ_{33}	1.00
ϕ_{44}	1.00
ϕ_{55}	1.01
σ_1^2	1.01
σ_2^2	1.00
σ_3^2	1.01
σ_4^2	1.00
σ_5^2	1.01

Figure 4.13: *For data set 4.1; the log posterior values for all five chains.*

4.6.2 Data set 4.2:

small QTL effects and large trait-to-trait effects

Data set 4.2 is comprised of small QTL effects and large trait-to-trait effects. Increasing the size of the trait-to-trait effects from data set 4.1 should increase the signal of the true model and make it easier to detect the true phenotype network structure. The marginal posterior probabilities of the directed edges are given in Figures 4.14 and 4.15 for chain 1. We note that with the dimension fixed to include five trait-to-trait effects, we would include the following directed edges: β_{12}, β_{13}, β_{24}, β_{35}, β_{45}. This is the true phenotype network structure, given in Figure 4.15 (a).

The marginal posterior probabilities for all directed edges in both Figures 4.14 and 4.15 illustrate that with small QTL effects and large trait-to-trait effects, we easily identify the direction of causality between traits; however, the chain has not visited many models — only 4 different phenotype network structures. As a result, many directed edges have a marginal posterior probability of 0.

The model averaged parameter estimates, conditional on their existence within the model, are given in Table 4.5 (p. 118). All parameters in the true model have 95% credible intervals containing the true parameter value. Once again, we note that edges which reverse the true direction of causality, e.g. β_{21}, β_{31} and β_{42}, have 95% credible intervals that do not include 0.

Just 4 different phenotype network structures were visited, with the joint posterior probability of each given in Figure 4.16 (p. 117). All of the visited models have edges where a causal relationship exists, but the direction of causality varies between models. The true phenotype network structure is the model with the greatest joint posterior probability, $p(M_{TRUE}) = 0.5149$, and this is the same phe-
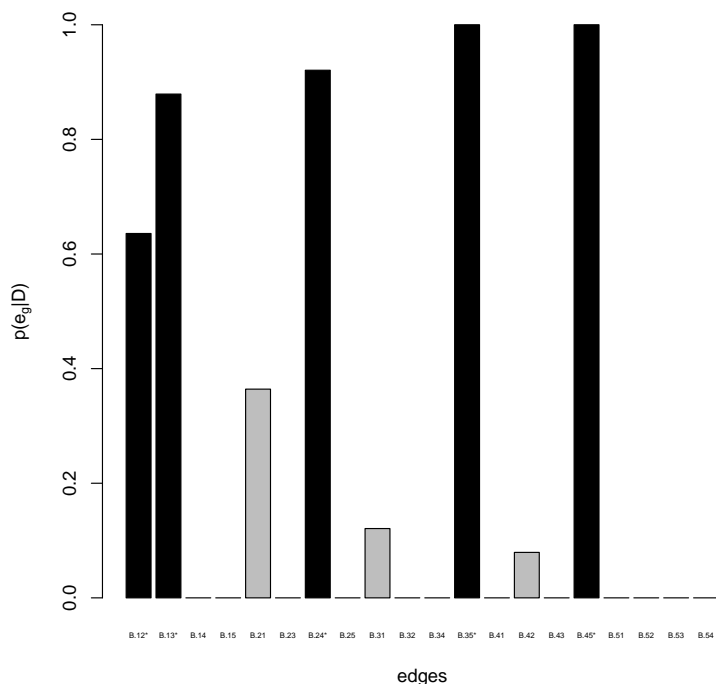
Figure 4.14: *The marginal posterior probability of each directed edge e_g, denoted $p(e_g|D)$, for data set 4.2 (chain 1). The terms "B.ij" denote the trait-to-trait effect β_{ij} corresponding to the directed edge, i.e. the effect of trait i on trait j. The edges comprising the true model are indicated by solid black bars and an "*" next to the label of the directed edge.*

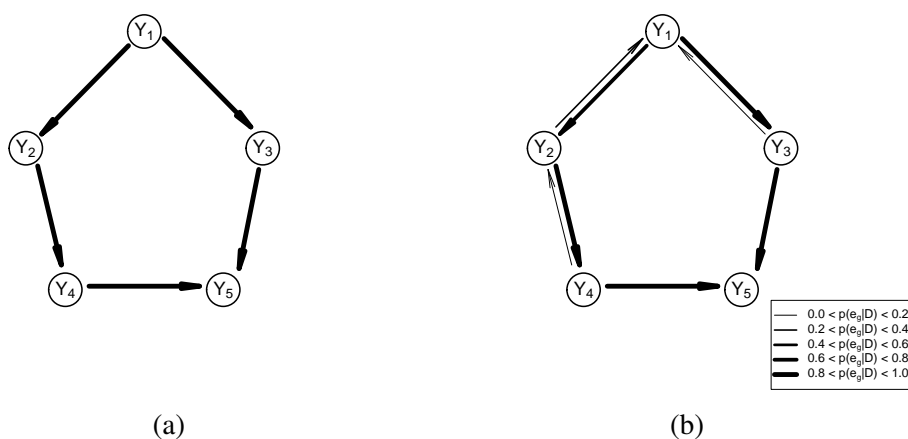Figure 4.15: *For data set 4.2 (chain 1): (a) the true phenotype network structure, and (b) the marginal posterior probability of each directed edge, denoted $p(e_g|D)$, displayed in the context of a causal network structure. The thicker the directed edge, the larger the marginal posterior probability.*

notype network structure as indicated by the marginal posterior probabilities for
the individual directed edges in Figures 4.14 and 4.15. Due to the increase in the
size of the trait-to-trait effects from data set 4.1, the chain does not move between
models as easily as it did with small QTL effects — just four causal network
structures were searched (down from 121 with small trait-to-trait effects). The
parameter estimates for the true model only are given in Table 4.6 (p. 119). These
parameter estimates are much closer to the true parameter values than the model
averaged parameter estimates in Table 4.5. The credible intervals are narrower,
and they all contain the true parameter values. However, due to the small size of
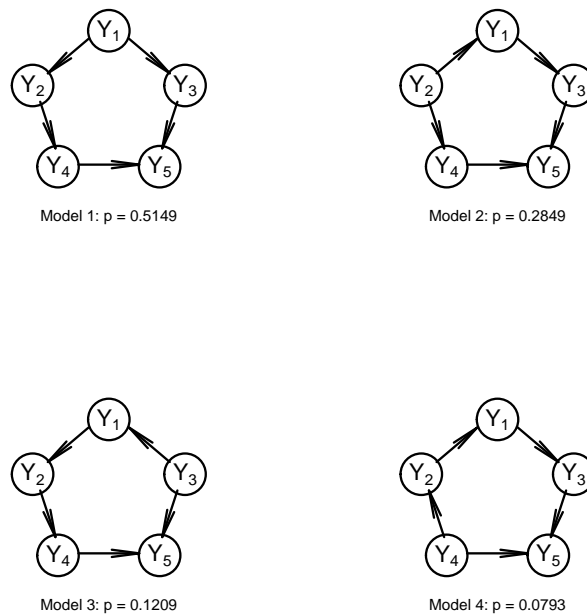the QTL effects, the 95% credible intervals for the QTL effects also include 0.

Figure 4.16: *The joint posterior probabilities, p, of the four models visited by
chain 1 for data set 4.2. The true model is the model with the greatest joint poste-
rior probability, Model 1.*

Table 4.5: *The model averaged parameter estimates for data set 4.2 (chain 1).*
n(parameter) denotes the number of times the parameter was included in the es-
timated causal network structure. Intervals including the true value are indicated
by an asterisk (*).

parameter	true value	estimated value	95% CI	n(parameter)
β_{01}	0.50	0.246	$(-0.117, 0.631)^*$	90000
β_{02}	0.50	0.588	$(-0.142, 1.116)^*$	90000
β_{03}	0.50	0.532	$(0.334, 1.135)^*$	90000
β_{04}	0.50	0.647	$(0.417, 1.737)^*$	90000
β_{05}	0.50	0.447	$(0.298, 0.595)^*$	90000
β_{12}	1.10	1.131	$(1.045, 1.218)^*$	57222
β_{13}	1.10	1.133	$(1.047, 1.219)^*$	79116
β_{14}	0.00	NA	(NA, NA)	0
β_{15}	0.00	NA	(NA, NA)	0
β_{21}	0.00	0.509	$(0.473, 0.545)$	32778
β_{23}	0.00	NA	(NA, NA)	0
β_{24}	1.10	1.071	$(1.014, 1.130)^*$	82860
β_{25}	0.00	NA	(NA, NA)	0
β_{31}	0.00	0.509	$(0.473, 0.546)$	10884
β_{32}	0.00	NA	(NA, NA)	0
β_{34}	0.00	NA	(NA, NA)	0
β_{35}	1.10	1.094	$(1.023, 1.164)^*$	90000
β_{41}	0.00	NA	(NA, NA)	0
β_{42}	0.00	0.693	$(0.660, 0.726)$	7140
β_{43}	0.00	NA	(NA, NA)	0
β_{45}	1.10	1.13	$(1.074, 1.187)^*$	90000
β_{51}	0.00	NA	(NA, NA)	0
β_{52}	0.00	NA	(NA, NA)	0
β_{53}	0.00	NA	(NA, NA)	0
β_{54}	0.00	NA	(NA, NA)	0
ϕ_{11}	0.10	-0.015	$(-0.182, 0.129)^*$	90000
ϕ_{22}	0.10	0.047	$(-0.164, 0.334)^*$	90000
ϕ_{33}	0.10	0.113	$(-0.076, 0.295)^*$	90000
ϕ_{44}	0.10	0.049	$(-0.134, 0.237)^*$	90000
ϕ_{55}	0.10	0.053	$(-0.124, 0.226)^*$	90000
σ_1^2	1.00	0.753	$(0.400, 1.160)^*$	90000
σ_2^2	1.00	1.334	$(0.588, 2.513)^*$	90000
σ_3^2	1.00	1.139	$(0.865, 2.438)^*$	90000
σ_4^2	1.00	1.155	$(0.835, 3.698)^*$	90000
σ_5^2	1.00	1.02	$(0.900, 1.154)^*$	90000

Table 4.6: *The parameter estimates for the true causal structure for data set 4.2 (chain 1). Intervals including the true value are indicated by an asterisk (*).*

parameter	true value	estimated value	95% CI
β_{01}	0.50	0.517	$(0.387, 0.651)^*$
β_{02}	0.50	0.493	$(0.357, 0.632)^*$
β_{03}	0.50	0.461	$(0.331, 0.589)^*$
β_{04}	0.50	0.556	$(0.411, 0.698)^*$
β_{05}	0.50	0.448	$(0.296, 0.597)^*$
β_{12}	1.10	1.132	$(1.047, 1.220)^*$
β_{13}	1.10	1.132	$(1.041, 1.219)^*$
β_{24}	1.10	1.072	$(1.015, 1.132)^*$
β_{35}	1.10	1.092	$(1.020, 1.159)^*$
β_{45}	1.10	1.131	$(1.077, 1.190)^*$
ϕ_{11}	0.10	-0.027	$(-0.206, 0.142)^*$
ϕ_{22}	0.10	0.012	$(-0.169, 0.199)^*$
ϕ_{33}	0.10	0.120	$(-0.042, 0.284)^*$
ϕ_{44}	0.10	0.050	$(-0.120, 0.229)^*$
ϕ_{55}	0.10	0.053	$(-0.126, 0.223)^*$
σ_1^2	1.00	1.044	$(0.922, 1.184)^*$
σ_2^2	1.00	0.987	$(0.871, 1.120)^*$
σ_3^2	1.00	0.976	$(0.862, 1.106)^*$
σ_4^2	1.00	0.944	$(0.834, 1.070)^*$
σ_5^2	1.00	1.021	$(0.900, 1.155)^*$

To assess convergence we ran multiple chains; summarized in Figure 4.17 by recording the initial model used by each chain, the number of models visited, the model with the greatest joint posterior probability and the joint posterior probability of the true model. We also include a graphical display of the marginal posterior probability of each directed edge to get an overview of the models visited by the chain, allowing us to get an impression of how the chains are moving. Each chain has visited very few models, appearing to get stuck at local maxima in chains 2, 3, and 5, where models with the greatest joint posterior probability were those which reversed some of the true causal relationships, or even added edges to imply a direct effect was present instead of the indirect effect of one trait on another

in the true model. For example, inferring that $Y_1 \rightarrow Y_5$, when in the true model
$Y_1 \rightarrow Y_2 \rightarrow Y_5$ and $Y_1 \rightarrow Y_3 \rightarrow Y_5$.

The inclusion of large trait-to-trait effects in our simulated data set has resulted
in chains that do not move between models easily, the true model was never vis-
ited by three of the five chains; however, the true model has the greatest joint
posterior probability in two of the chains, both with $p(M_{TRUE}|D) \approx 0.5$. In addi-
tion to Figure 4.17, Table 4.7 gives Gelman and Rubin's convergence diagnostic
for dimension independent parameters. The Gelman–Rubin Diagnostic is close to
1 for the QTL effect parameters; however, most dimension independent parame-
ters have not converged. When the trait-to-trait effects are large it is difficult to
propose a favourable move, fixing the number of edges is too restrictive. The log
likelihood has a particularly large potential scale reduction factor of 51.1, indicat-
ing that the chains have not converged. The values of the log posterior density are
plotted in Figure 4.18 for all five chains, and we observe what the Gelman–Rubin
diagnostic indicated — these chains have not converged. We note that chains 1
and 4 have similar log posterior densities and they are the smallest values within
the five chains, so chains 1 and 4 have the best estimate for the phenotype net-
work structure — the true model. Unfortunately, if there are small QTL effects
and large trait-to-trait effects the chain may get stuck in local maxima depending
on the initial model used. This illustrates why it is important to repeat the search
from a variety of different positions.

The trace plots for each chain do occasionally appear to get stuck in partic-
ular parts of the parameter space, this is the chain visiting different models with
different parameters. Appendix D contains the trace plots for all stored iterations
for chain 1; included to illustrate how the parameter values appear to be stuck in
areas of the parameter space as the chain visits different phenotype networks.

Figure 4.17: *A summary of the performance of each chain for data set 4.2. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Table 4.7: *The Gelman–Rubin convergence diagnostic for data set 4.2, for all five chains summarized in Figure 4.17.*

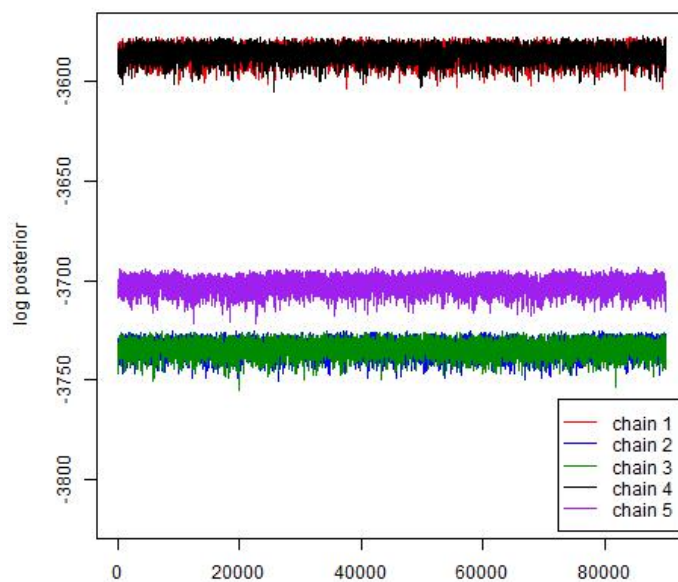|	Gelman–Rubin diagnostic
log likelihood	51.1
β_{01}	1.89
β_{02}	2.39
β_{03}	3.24
β_{04}	1.57
β_{05}	5.14
ϕ_{11}	1.03
ϕ_{22}	1.30
ϕ_{33}	1.00
ϕ_{44}	1.02
ϕ_{55}	1.21
σ_1^2	2.09
σ_2^2	1.92
σ_3^2	1.24
σ_4^2	1.08
σ_5^2	5.65

Figure 4.18: *For data set 4.2; the log posterior values for all five chains.*

4.6.3 Data set 4.3:

large QTL effects and small trait-to-trait effects

Data set 4.3 is comprised of large QTL effects and small trait-to-trait effects. The marginal posterior probabilities of the directed edges are given in Figures 4.19 and 4.20 for chain 1, illustrating that with large QTL effects and small trait-to-trait effects, we can identify the direction of causality between traits. However, as seen with data set 4.2, the chain has not visited many different phenotype network structures so is not mixing well. As a result, many directed edges have a marginal posterior probability of 0. With the dimension fixed to include five trait-to-trait effects, we would include the following edges: β_{12}, β_{13}, β_{24}, β_{35}, β_{45}. This is the true causal structure, given in Figure 4.20 (a).

The model averaged parameter estimates, conditional on their existence within the model, are given in Table 4.8 (p. 126). Those trait-to-trait effects that are in the true model all have the true value of 0.2 within their respective 95% credible interval. However, as seen in data sets 4.1 and 4.2, those edges which reverse the true direction of causality have 95% credible intervals that do not include 0 (here, β_{21}, β_{53} and β_{54}). 7 different phenotype network structures were visited, given in Figure 4.21 (p. 125). All of the models searched have edges where a causal relationship exists, but the direction of causality varies between the models. The true model has the greatest joint posterior probability, with $p(M_{TRUE}) = 0.8338$.

The parameter estimates for the true model only are given in Table 4.9. These parameter estimates are much closer to the true parameter values than the model averaged parameter estimates in Table 4.8. The credible intervals are narrower, all contain the true parameter values, and with larger QTL effects (than in data sets 4.1 and 4.2) the 95% credible intervals (for the QTL effects) no longer include 0.
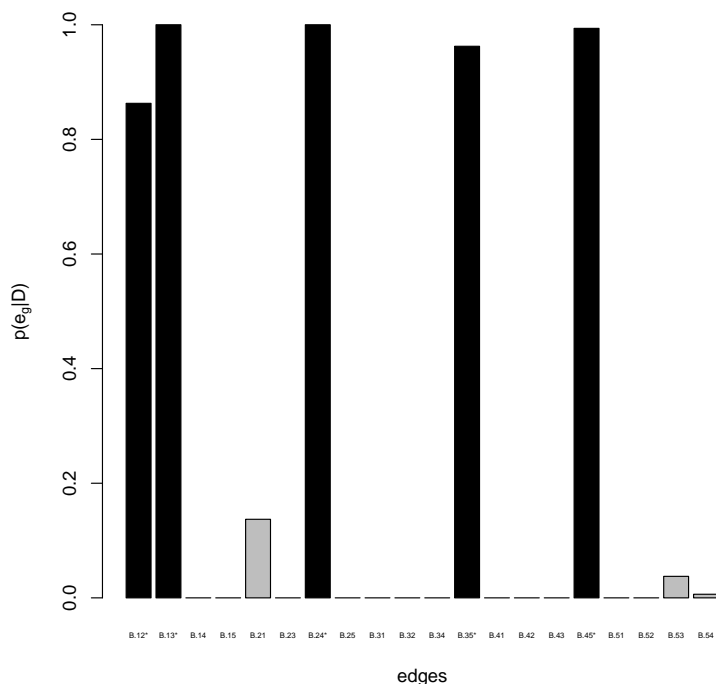
Figure 4.19: *The posterior probability of each directed edge e_g, denoted $p(e_g|D)$, for data set 4.3 (chain 1). The terms "B.ij" denote the trait-to-trait effect β_{ij} corresponding to the directed edge, i.e. the effect of trait i on trait j. The edges comprising the true model are indicated by solid black bars and an "*" next to the label of the directed edge.*
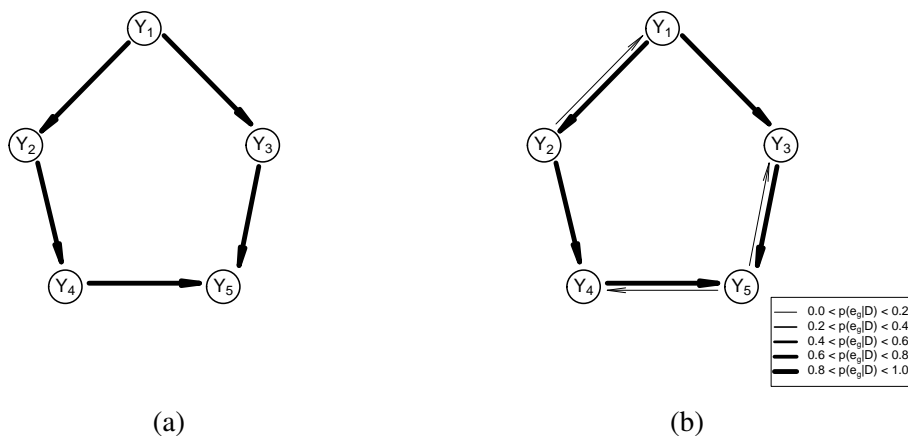

Figure 4.20: *For data set 4.3 (chain 1): (a) the true phenotype network structure, and (b) the marginal posterior probability of each directed edge, denoted $p(e_g|D)$, displayed in the context of a causal network structure. The thicker the directed edge, the larger the marginal posterior probability.*
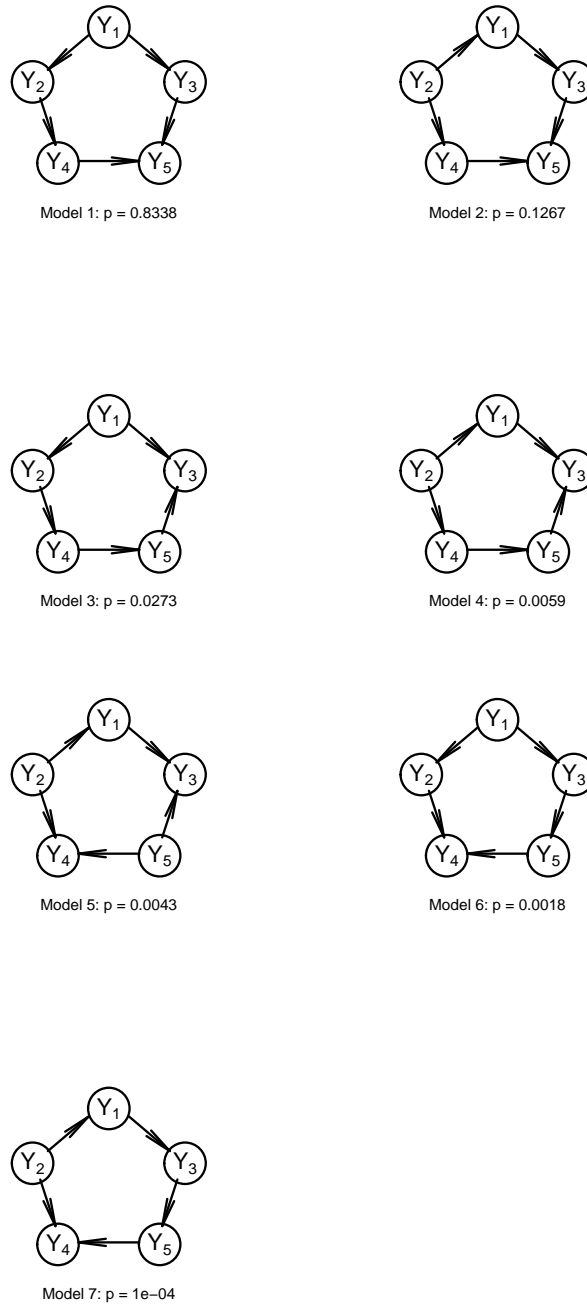
Figure 4.21: *The joint posterior probabilities, p, of the 7 models visited by chain 1 for data set 4.3. The true model is the model with the greatest joint posterior probability, Model 1.*

Table 4.8: *The model averaged parameter estimates for data set 4.3 (chain 1). $n(parameter)$ denotes the number of times the parameter was included in the estimated causal network structure. Intervals including the true value are indicated by an asterisk (*)..*

parameter	true value	estimated value	95% CI	n(parameter)
β_{01}	0.50	0.499	$(0.214, 0.662)^*$	90000
β_{02}	0.50	0.534	$(0.361, 0.794)^*$	90000
β_{03}	0.50	0.448	$(0.215, 0.600)^*$	90000
β_{04}	0.50	0.583	$(0.418, 0.739)^*$	90000
β_{05}	0.50	0.495	$(0.295, 0.744)^*$	90000
β_{12}	0.20	0.218	$(0.140, 0.296)^*$	77663
β_{13}	0.20	0.232	$(0.153, 0.307)^*$	90000
β_{14}	0.00	NA	(NA, NA)	0
β_{15}	0.00	NA	(NA, NA)	0
β_{21}	0.00	0.199	$(0.130, 0.275)$	12337
β_{23}	0.00	NA	(NA, NA)	0
β_{24}	0.20	0.172	$(0.096, 0.249)^*$	90000
β_{25}	0.00	NA	(NA, NA)	0
β_{31}	0.00	NA	(NA, NA)	0
β_{32}	0.00	NA	(NA, NA)	0
β_{34}	0.00	NA	(NA, NA)	0
β_{35}	0.20	0.191	$(0.114, 0.268)^*$	86624
β_{41}	0.00	NA	(NA, NA)	0
β_{42}	0.00	NA	(NA, NA)	0
β_{43}	0.00	NA	(NA, NA)	0
β_{45}	0.20	0.226	$(0.149, 0.306)^*$	89436
β_{51}	0.00	NA	(NA, NA)	0
β_{52}	0.00	NA	(NA, NA)	0
β_{53}	0.00	0.168	$(0.108, 0.236)$	3376
β_{54}	0.00	0.122	$(0.059, 0.193)$	564
ϕ_{11}	1.20	1.038	$(0.867, 1.214)^*$	90000
ϕ_{22}	1.20	1.085	$(0.907, 1.258)^*$	90000
ϕ_{33}	1.20	1.187	$(1.020, 1.354)^*$	90000
ϕ_{44}	1.20	1.110	$(0.944, 1.280)^*$	90000
ϕ_{55}	1.20	1.124	$(0.944, 1.298)^*$	90000
σ_1^2	1.00	1.036	$(0.908, 1.176)^*$	90000
σ_2^2	1.00	0.996	$(0.874, 1.136)^*$	90000
σ_3^2	1.00	0.976	$(0.861, 1.107)^*$	90000
σ_4^2	1.00	0.946	$(0.835, 1.071)^*$	90000
σ_5^2	1.00	1.025	$(0.903, 1.164)^*$	90000

Table 4.9: *The parameter estimates for the true causal structure for data set 4.3 (chain 1). Intervals including the true value are indicated by an asterisk (*).*

parameter	true value	estimated value	95% CI
β_{01}	0.50	0.534	$(0.405, 0.666)^*$
β_{02}	0.50	0.502	$(0.359, 0.662)^*$
β_{03}	0.50	0.459	$(0.306, 0.599)^*$
β_{04}	0.50	0.585	$(0.425, 0.740)^*$
β_{05}	0.50	0.486	$(0.290, 0.681)^*$
β_{12}	0.20	0.218	$(0.140, 0.295)^*$
β_{13}	0.20	0.232	$(0.155, 0.307)^*$
β_{24}	0.20	0.172	$(0.097, 0.249)^*$
β_{35}	0.20	0.191	$(0.114, 0.269)^*$
β_{45}	0.20	0.226	$(0.148, 0.307)^*$
ϕ_{11}	1.20	1.038	$(0.865, 1.215)^*$
ϕ_{22}	1.20	1.084	$(0.909, 1.257)^*$
ϕ_{33}	1.20	1.186	$(1.019, 1.355)^*$
ϕ_{44}	1.20	1.108	$(0.942, 1.281)^*$
ϕ_{55}	1.20	1.121	$(0.941, 1.295)^*$
σ_1^2	1.00	1.044	$(0.920, 1.180)^*$
σ_2^2	1.00	0.987	$(0.871, 1.119)^*$
σ_3^2	1.00	0.977	$(0.863, 1.108)^*$
σ_4^2	1.00	0.946	$(0.835, 1.071)^*$
σ_5^2	1.00	1.023	$(0.902, 1.159)^*$

To assess convergence we ran multiple chains, which we summarize in Figure 4.22 by recording the initial model used by each chain, the number of models visited, the model with the greatest joint posterior probability and the posterior probability of the true model. We also include a graphical display of the marginal posterior probability of each directed edge to get an overview of the models visited by the chain, allowing us to get an impression of how the chains are moving.

With large QTL effects and small trait-to-trait effects, the true phenotype network structure was consistently estimated with $p(M_{TRUE}|D) \geq 0.8000$ for all chains. Each chain visited between 9 and 13 models, slightly more than with data set 4.2, but much less than data set 4.1. The chains are not mixing well.

Figure 4.22: *A summary of the performance of each chain for data set 4.3. The number of models visited is denoted* $n(models)$, *and the marginal posterior probability of the directed edge* e_g *is denoted* $p(e_g|D)$. *The greatest joint posterior probability is denoted* p^*, *such that model* M_k *has the greatest posterior probability when* $p(M_k|D) = p^*$. *The posterior probability of the true model is denoted* $p(M_{TRUE}|D)$.

In addition to Figure 4.22, Table 4.10 gives Gelman and Rubin's convergence diagnostic for dimension independent parameters. The Gelman–Rubin Diagnostic is close to 1 for all parameters tested, indicating the dimension independent parameters have converged. With large QTL effects the true model has a strong signal; however, this makes it difficult for the chain to propose favourable moves.

The values of the log posterior density are plotted in Figure 4.23 for all five chains, and we observe what the Gelman–Rubin diagnostic indicated — these chains overlap and appear to have converged. Unfortunately the chain has visited few models so it is not mixing well.

Table 4.10: *The Gelman–Rubin convergence diagnostic for data set 4.3, for all five chains summarized in Figure 4.22.*

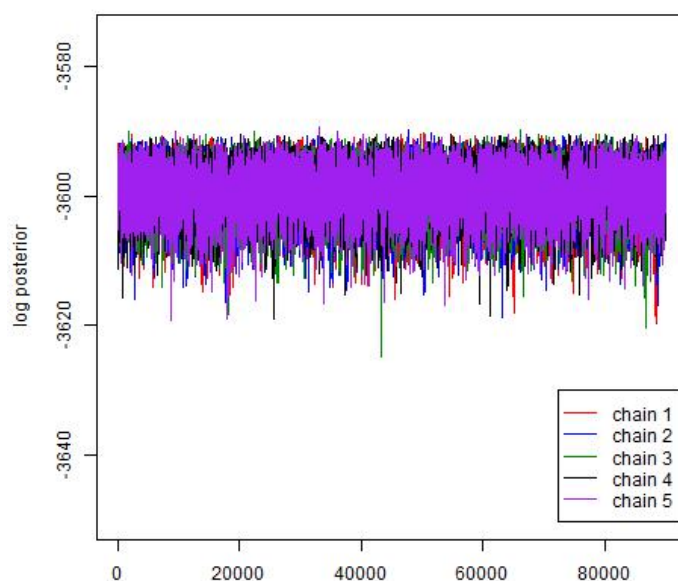|	Gelman–Rubin diagnostic
log likelihood	1.00
β_{01}	1.03
β_{02}	1.01
β_{03}	1.01
β_{04}	1.01
β_{05}	1.01
ϕ_{11}	1.01
ϕ_{22}	1.01
ϕ_{33}	1.00
ϕ_{44}	1.01
ϕ_{55}	1.01
σ_1^2	1.00
σ_2^2	1.00
σ_3^2	1.00
σ_4^2	1.00
σ_5^2	1.00

Figure 4.23: *For data set 4.3: the log posterior values for all five chains.*

4.6.4   Data set 4.4:
large QTL effects and large trait-to-trait effects

Data set 4.4 is comprised of large QTL effects and large trait-to-trait effects. The marginal posterior probabilities for the directed edges are given in Figures 4.24 and 4.25 for chain 1. We note that with large effects (both QTL and trait-to-trait) only the true phenotype network structure, given in Figure 4.25 (a), has been visited and therefore all directed edges comprising it have a posterior probability of 1, see Figure 4.25 (b).

The model averaged parameter estimates, here corresponding to the estimates for the true model, are given in Table 4.11 (p. 132). All of the 95% credible intervals contain the true parameter values. Due to the increase in the size of the QTL effects from data set 4.3, we have reduced the search from just four
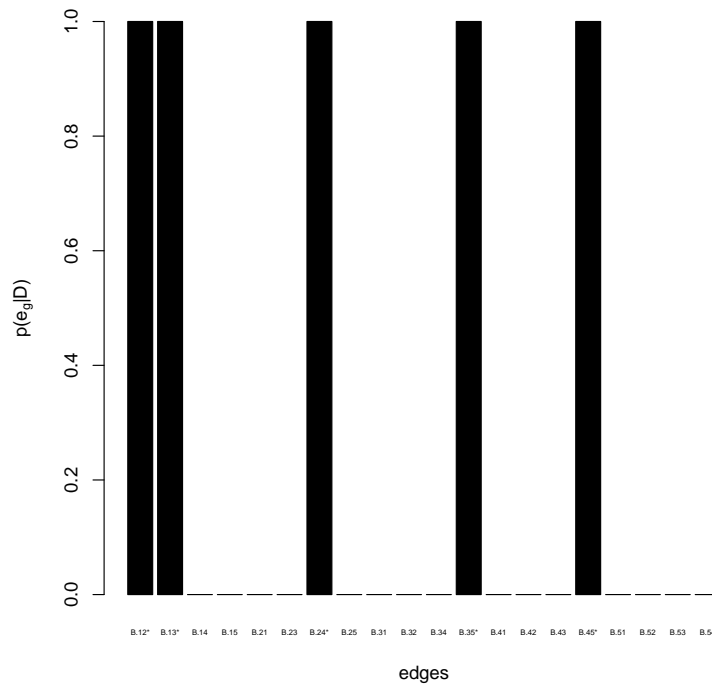
Figure 4.24: *The posterior probability of each directed edge e_g, denoted $p(e_g|D)$ for data set 4.4 (chain 1). The terms "B.ij" denote the trait-to-trait effect β_{ij} corresponding to the directed edge, i.e. the effect of trait i on trait j. The edges comprising the true model are indicated by solid black bars and an '*' next to the label of the directed edge.*
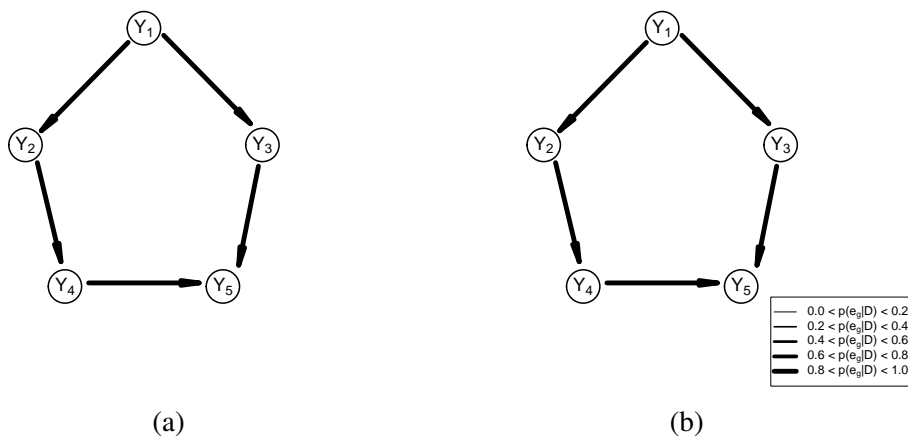

(a)   (b)

Figure 4.25: *For data set 4.4 (chain 1): (a) the true phenotype network structure, and (b) the marginal posterior probability of each directed edge, denoted $p(e_g|D)$, displayed in the context of a causal network structure. The thicker the directed edge, the larger the marginal posterior probability.*

Table 4.11: *The model averaged parameter estimates for data set 4.4 (chain 1). $n(parameter)$ denotes the number of times the parameter was included in the estimated causal network structure. Intervals including the true value are indicated by an asterisk (*).*

parameter	true value	estimated value	95% CI	n(parameter)
β_{01}	0.50	0.532	$(0.405, 0.659)^*$	90000
β_{02}	0.50	0.507	$(0.357, 0.670)^*$	90000
β_{03}	0.50	0.463	$(0.308, 0.611)^*$	90000
β_{04}	0.50	0.564	$(0.404, 0.733)^*$	90000
β_{05}	0.50	0.465	$(0.267, 0.658)^*$	90000
β_{12}	1.10	1.116	$(1.038, 1.193)^*$	90000
β_{13}	1.10	1.128	$(1.053, 1.202)^*$	90000
β_{14}	0.00	NA	(NA, NA)	0
β_{15}	0.00	NA	(NA, NA)	0
β_{21}	0.00	NA	(NA, NA)	0
β_{23}	0.00	NA	(NA, NA)	0
β_{24}	1.10	1.092	$(1.042, 1.142)^*$	90000
β_{25}	0.00	NA	(NA, NA)	0
β_{31}	0.00	NA	(NA, NA)	0
β_{32}	0.00	NA	(NA, NA)	0
β_{34}	0.00	NA	(NA, NA)	0
β_{35}	1.10	1.079	$(1.018, 1.140)^*$	90000
β_{41}	0.00	NA	(NA, NA)	0
β_{42}	0.00	NA	(NA, NA)	0
β_{43}	0.00	NA	(NA, NA)	0
β_{45}	1.10	1.126	$(1.077, 1.173)^*$	90000
β_{51}	0.00	NA	(NA, NA)	0
β_{52}	0.00	NA	(NA, NA)	0
β_{53}	0.00	NA	(NA, NA)	0
β_{54}	0.00	NA	(NA, NA)	0
ϕ_{11}	1.20	1.043	$(0.870, 1.211)^*$	90000
ϕ_{22}	1.20	1.081	$(0.899, 1.253)^*$	90000
ϕ_{33}	1.20	1.185	$(1.015, 1.355)^*$	90000
ϕ_{44}	1.20	1.114	$(0.944, 1.284)^*$	90000
ϕ_{55}	1.20	1.120	$(0.946, 1.300)^*$	90000
σ_1^2	1.00	1.044	$(0.922, 1.181)^*$	90000
σ_2^2	1.00	0.988	$(0.872, 1.120)^*$	90000
σ_3^2	1.00	0.978	$(0.864, 1.107)^*$	90000
σ_4^2	1.00	0.946	$(0.836, 1.071)^*$	90000
σ_5^2	1.00	1.021	$(0.901, 1.156)^*$	90000

phenotype network structures to only one phenotype network structure. This chain is not mixing well at all. To assess convergence we ran multiple chains, which we summarize in Figure 4.26 by recording the initial model used by each chain, the number of models visited, the model with the greatest joint posterior probability and the posterior probability of the true model. We also include a graphical display of the marginal posterior probability of each directed edge to get an overview of the models visited by the chain, allowing us to get an impression of how the chains are moving.

With large QTL effects and large trait-to-trait effects, the true phenotype network structure can be difficult to identify, depending on where in the model space the chain begins. These chains visited very few models, just 1 or 2, and just two chains found the true causal network structure, both with $p(M_{TRUE}|D) = 1$. Other chains failed to visit the true model at all ($p(M_{TRUE}|D) = 0$), instead finding models which reversed the true direction of causality (i.e. some edges were reversed) or even added edges between traits which have no causal relationship.

In addition to Figure 4.26, Table 4.12 gives Gelman and Rubin's convergence diagnostic for dimension independent parameters. The potential scale reduction factors from the Gelman–Rubin diagnostic are far from 1 for almost all parameters tested, indicating the dimension independent parameters have not converged. This is supported by the log posterior values plotted in Figure 4.27. We note that chains 1 and 3 have similar log posterior densities, the smallest seen within the five chains, indicating the best model is that estimated by these chains — the true model. Unfortunately, with large QTL effects and large trait-to-trait effects it is very difficult to propose favourable moves for the chain while keeping the number of directed edges fixed.

Figure 4.26: *A summary of the performance of each chain for data set 4.4. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Table 4.12: *The Gelman–Rubin convergence diagnostic for data set 4.4, for all five chains summarized in Figure 4.26.*

	Gelman–Rubin diagnostic
log likelihood	125.00
β_{01}	15.60
β_{02}	20.90
β_{03}	16.00
β_{04}	1.04
β_{05}	14.90
ϕ_{11}	7.52
ϕ_{22}	5.09
ϕ_{33}	5.33
ϕ_{44}	1.01
ϕ_{55}	1.25
σ_1^2	11.80
σ_2^2	25.20
σ_3^2	1.01
σ_4^2	1.00
σ_5^2	25.00

Figure 4.27: *For data set 4.4: the log posterior values for all five chains.*

4.6.5   Summary of results

Restricting the search for the true phenotype network structure to a fixed number of directed edges while allowing the phenotype network structure to change is in some sense artificial as we are working within a large dimension space, but selecting only models with a fixed number of parameters. Those simulated data sets with small effects (both QTL-to-trait and trait-to-trait) moved between phenotype networks easier than with large effects, as a change in the phenotype network structure had only a minor effect on the posterior density. Therefore more models with varying phenotype network structures were visited. In data set 4.1, the true model was visited by all five chains, but was not identified as the model with the greatest joint posterior probability due to it having a weak signal. The causal relationships were identified, but the direction of causality varied.

Large effects (both QTL and trait-to-trait) had more of an effect on the posterior density, with Figure 4.27 displaying a larger difference between the log posterior values for the five chains. With large effects the phenotype network was not altered very often, and few models were visited. However, in data set 4.4, the two chains that identified the true phenotype network structure had the smallest log posterior density among all the chains, indicating that they had estimated the best model (the true model). The models used to initialise the chain are therefore important, and multiple chains are recommended, starting at various phenotype network structures. The data set with large trait-to-trait effects and small QTL effects (data set 4.3), yielded similar results but the chain was able to visit different phenotype networks more easily than with large QTL effects.

The best results were seen in data set 4.2 with large QTL effects and small trait-to-trait effects; all chains estimated the true model with a joint posterior prob-

ability of at least 0.8, regardless of the initial model used.

Keeping this approach within a fixed dimension has restricted the way in which the sampler was able to propose new phenotype network structures, meaning that if a model had two or more badly placed edges it couldn't easily fix itself. This problem seems to be acute when the QTL effects and trait-to-trait effects are large, so the solution is to allow the number of edges to vary. With the inclusion of more move types in Chapter 5 we expect mixing to improve.

Chapter 5

Bayesian estimation of a phenotype network structure II

Chapter 4 introduced our Bayesian approach to estimating an unknown phenotype network structure. The update, reverse and relocation moves were used, and it was determined that this approach was able to estimate the phenotype network structure with the number of directed edges fixed; however, mixing was poor. There is a large dimension space, and if a model has two or more badly placed edges it can't easily fix itself if it is restricted to only reversing or relocating a single edge. The solution is to allow the dimension of the model to vary. Here, we extend the reversible jump sampler used in Chapter 4, utilizing the update, reverse and relocation steps, and introduce new 'add', 'remove' and 'double' steps which allow the number of directed edges in the phenotype network structure to vary.

This approach is based upon the QTLnet algorithm proposed by Neto et al. (2010), and the differences between this approach and the QTLnet algorithm are summarized below.

- Neto et al. (2010) propose the joint inference of the phenotype network structure and the genetic components; however, we assume that the genetic architecture is known and focus on the subproblem of estimating the unknown phenotype network structure and the QTL effects.

- Our approach makes use of LSE (equivalent to the maximum likelihood estimates) in the chain, obtained from separate regressions of each trait on the traits and QTL directly affecting it. This makes for a fast and efficient chain so we have an efficient sampler using well-tuned proposal distributions.

- Neto et al. (2010) experienced slow mixing of the Markov chain. This approach incorporates extra move types; in Chapter 4 we used the relocate step which relocates a directed edge, now we propose a double step to combine two steps that alter the phenotype network structure.

- We define the neighbourhood of a phenotype network structure as being conditional on the selected move type x; instead of containing all possible candidate phenotype network structures given all move types, as in Neto et al. (2010). This makes it faster to obtain the neighbourhood for a given model when the number of traits is large.

- Neto et al. (2010) assume uniform priors over all network structures; however, we introduce a prior on the number of directed edges in a model which causes simpler models, i.e. models with fewer directed edges, to be preferred.

The genetic components estimated by Neto et al. (2010) include the QTL locations, QTL genotypes and QTL-to-trait effects. We, however, assume that the number of QTL, QTL locations and QTL genotypes are known, and that we have

knowledge of which subset of QTL affect each trait (i.e. we know which QTL effect parameters require estimation). We refer to this as the genetic architecture (see Section 1.2), and focus on estimating the unknown phenotype network structure and the QTL effects. We may assume that the genetic architecture is known as many methods have been established to map the genetic architecture of a trait (or multiple traits). Neto et al. (2010), for example, recommend the use of the seemingly unrelated regression model to estimate the genetic components; see Banerjee et al. (2008). It follows that our approach is a subproblem of Neto et al. (2010). The model used is the same as given in Equation 3.4 (p. 38), restated here for convenience:

$$y_{it} = \beta_{0t} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik} + \varepsilon_{it}$$

where the model terms are as defined in Table 3.1. For the tth quantitative trait of the ith individual,

$$y_{it} = {\mu_{it}}^* + \varepsilon_{it}$$

where
$$ {\mu_{it}}^* = \beta_{0t} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{ik} $$
$$ \varepsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_t^2). $$

As explained in Chapters 3 and 4, we have a set of QTL parents

$$\upsilon_Q = (\upsilon_Q(1), \cdots, \upsilon_Q(N_T)),$$

and a set of trait parents

$$\upsilon_Y = (\upsilon_Y(1), \cdots, \upsilon_Y(N_T)),$$

where υ_Q is fixed as the genetic architecture is known, and the phenotype network

structure is estimated using RJMCMC, proposing changes to the components of υ_Y. In this chapter, the proposed change is the reversal or relocation of a directed edge, as detailed in Chapter 4, or the addition or removal of a directed edge as detailed in Sections 5.2 and 5.3. An additional step, the double step, is introduced later in Section 5.9.

An outline of the reversible jump algorithm for model updates is given in Section 5.1, detailing the selection of the move types (Section 5.1.1), the posterior distribution (Section 5.1.2), the proposal distributions for generating candidate parameters (Section 5.1.3), and the Jacobian for the transformation between models of different dimensions (Section 5.1.4). The add and remove steps used in this approach are explained in Sections 5.2 and 5.3, with small examples to demonstrate the calculation of the individual acceptance probabilities. The approach is summarized in Section 5.4, and a simulation study is included in Section 5.5 to demonstrate how the prior on the number of edges can be used to penalize models with too many edges. As this approach requires prior knowledge of the genetic architecture, we include simulated examples in Section 5.7 to demonstrate how the approach performs when the genetic architecture is incorrect. Section 5.8 uses a published data set, the winter wheat data set from Dhungana et al. (2007), to illustrate this reversible jump approach. Lastly, a new 'double' step is proposed in Section 5.9 complete with an illustrative example.

5.1 Reversible jump algorithm for model updates, for a varying number of edges

Following Section 4.1, we have a model space $\mathbb{M}$ with countable elements $M \in \mathbb{M}$, where model M is defined by its causal network structure $\mathbf{S}$ (comprised of a set

of nodes connected by a set of directed edges), and the parameters Ω. The model space is partitioned into the structure space $\mathbb{S}$ and the parameter space $\mathbb{O}$, such that $\mathbf{S} \in \mathbb{S}$ and $\Omega \in \mathbb{O}$. Let us propose a move from the current model, M, to a candidate model, M', by selecting a move type x from the move space $\mathbb{X}$, such that a unique move type x changes model M into M'. Here, x is one of

- **update** (x^U): update all parameters in the current model

- **reverse** (x^S): reverse the direction of one of the directed edges in the current phenotype network

- **relocate** (x^C): relocate one of the directed edges in the current phenotype network, i.e. delete one edge and propose a new, distinct directed edge

- **add** (x^A): add a directed edge to the current phenotype network

- **remove** (x^R): remove a directed edge from the current phenotype network.

A double step will also be introduced later as a way to propose larger changes in the phenotype network structure. It is excluded here so that we may include the simpler, and more conservative, relocate step, which is a possible type of double step.

Now that we are searching models of varying dimensions, the causal network structure for models M and M' has an extra parameter to denote the number of edges in the phenotype network structure: E_M and $E_{M'}$, respectively. For any current model M we have the parameter set Ω:

$$\Omega = \{E_M, \beta_{01}, \cdots, \beta_{0N_T}, \beta_{\upsilon_Y(1)1}, \cdots, \beta_{\upsilon_Y(N_T)N_T}, \phi_{\upsilon_Q(1)1}, \cdots, \phi_{\upsilon_Q(N_T)N_T}, \sigma_1^2, \cdots, \sigma_{N_T}^2\},$$

and the phenotype network structure $\mathbf{S}$, such that $M = \{\Omega, \mathbf{S}\}$. The dimension of the model M is denoted m,

$$m = E_M + 1 + N_\phi + 2N_T,$$

and is equivalent to the number of parameters in the current model, with E_M trait-to-trait effects (with E_M as a parameter as well), N_ϕ QTL effects, N_T intercepts and N_T residual variances. For any candidate model, M', we have the parameter set Ω':

$$\Omega' = \{E_{M'}, \beta'_{01}, \cdots, \beta'_{0N_T}, \beta'_{\upsilon_Y(1)1}, \cdots, \beta'_{\upsilon_Y(N_T)N_T}, \phi'_{\upsilon_Q(1)1}, \cdots, \phi'_{\upsilon_Q(N_T)N_T}, \sigma'^2_1, \cdots, \sigma'^2_{N_T}\},$$

and the phenotype network structure $\mathbf{S}'$, such that $M' = \{\Omega', \mathbf{S}'\}$. The dimension of the model M' is denoted m',

$$m' = E_{M'} + 1 + N_\phi + 2N_T$$

and is equivalent to the number of parameters in the candidate model, with $E_{M'}$ trait-to-trait effects (with E_M as a parameter as well), N_ϕ QTL effects, N_T intercepts and N_T residual variances. As we now propose the addition or removal of directed edges, m and m' may differ.

As in Chapter 4, whenever a Markov chain Monte Carlo step is carried out, conditional on the current state of the chain (i.e. model $M = \{\Omega, \mathbf{S}\}$), the following occur:

- select a move type x with probability $q(x|\Omega, \mathbf{S})$ where

$$\sum_{x \in \mathbb{X}} q(x|\Omega, \mathbf{S}) = 1 \quad \text{for all } (\Omega, \mathbf{S})$$

- given move type x, propose a candidate phenotype network structure $\mathbf{S}'$ with

probability $q(\mathbf{S}'|x,\Omega,\mathbf{S})$ where

$$\sum_{\mathbf{S}'\in\mathbb{S}} q(\mathbf{S}'|x,\Omega,\mathbf{S}) = 1 \quad \text{for all } (x,\Omega,\mathbf{S})$$

- generate a set of random variables $\mathbf{u}_{M'} \in \mathbb{U}_{M'}$, with probability density

$$q(\mathbf{u}_{M'}|x,\mathbf{S}',\Omega,\mathbf{S})$$

- the parameters of the new state of the chain, Ω' are generated by the deterministic function $g_{MM'}$, such that

$$\Omega' = g_{MM'}(\Omega,\mathbf{u}_{M'}).$$

The general form of the acceptance probability for the update, reverse, relocate, add and remove steps, is defined as $min(1,r)$, where

$$r = \frac{p(M'|\mathbf{Y},\Psi)}{p(M|\mathbf{Y},\Psi)} \times \frac{q(\mathbf{u}'_M,\mathbf{S},x'|\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'},\mathbf{S}',x|\Omega,\mathbf{S})} \times \left|\frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})}\right| \tag{5.1}$$

(Green 1995). The move type transforming model M into M' is denoted x, and x' is the reverse of move type x — transforming model M' into M. The posterior distributions are denoted by $p()$ for models M and M', and $q()$ denotes the proposal distributions. It follows that:

$$\begin{aligned} r &= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \times \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \times \\ &\quad \frac{q(\mathbf{S}|x',\Omega',\mathbf{S}')}{q(\mathbf{S}'|x,\Omega,\mathbf{S})} \times \frac{q(x'|\Omega',\mathbf{S}')}{q(x|\Omega,\mathbf{S})} \times \left|\frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})}\right| \end{aligned} \tag{5.2}$$

using the same notation as in Chapter 4.

5.1.1   Selecting the move type

At each step in the chain a move type is chosen at random from the update, reverse, relocate, add and remove steps, conditional on the current model, M. The probabilities associated with the selection of each move type are chosen to allow equal probability of model parameter updates and exploration of the various phenotype network structures.

The probability of each type of move is given in Table 5.1, where the indicator functions in Equation 4.3 (p. 72) are used in addition to the operators U, S_g and $C_{g,f}$ defined in Section 4.1.1. We define the operators A_g and R_g such that:

- $A_g[M] \in \mathbb{M}$ is a new model which adds edge g (to **S**) with an associated generation of new parameter values

- $R_g[M] \in \mathbb{M}$ is a new model which removes edge g (from **S**) with an associated generation of new parameter values.

The size of the resulting neighbourhoods, $|Ne(M,x)|$, are given in Table 5.1 using these operators and indicator functions.

The parameters associated with the addition or removal of an edge include the new trait-to-trait effect (for an add step), as well as select trait intercepts and residual variances. The exact parameters to be updated are stated in Sections 5.2 and 5.3 for the add and remove steps, respectively.

5.1.2   Posterior distribution

The formation of the posterior distribution is as described in Section 4.1.2. The QTL genotypes (**Q**), QTL locations (Λ), marker genotypes (**G**) and recombination fractions (**r**) are assumed to be known. Let Ψ denote the known parameters, such

Table 5.1: *The probability of each move type, as well as the number of candidate models for move type x, $|Ne(M,x)|$. There exist N_T traits, such that there are $N_T(N_T-1)$ unique directed edges possible; subsets of these edges form the current and candidate phenotype network structures. Model M has E_M edges, with the maximum number of directed edges denoted $max(E) = \frac{1}{2}N_T(N_T-1)$.*

Move Type	(x)	Number of edges in model M (E_M)		
		0	$1 \le E_M \le (max(E)-1)$	$max(E)$
Update	(x^U)	0.5	0.500	0.50
Add	(x^A)	0.5	0.125	0
Remove	(x^R)	0	0.125	0.25
Reverse	(x^S)	0	0.125	0.25
Relocate	(x^C)	0	0.125	0

Move Type	(x)	Size of the neighbourhood of model M $|Ne(M,x)|$
Update	(x^U)	1
Add	(x^A)	$\sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M])I_\in(g|\mathbf{S})$
Remove	(x^R)	E_M
Reverse	(x^S)	$\sum_{g=1}^{E_M} I_a(S_g[M])$
Relocate	(x^C)	$\sum_{g=1}^{E_M}\sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M])I_R(g,f|\mathbf{S})I_\in(f|\mathbf{S})$

that $\Psi = \{\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r}\}$. Given the data $\mathbf{Y}$ and Ψ, the posterior distribution for model M, $M = \{\Omega,\mathbf{S}\}$, is:

$$p(M|\mathbf{Y},\Psi) \propto p(\mathbf{Y}|\Omega,\mathbf{S},\Psi)p(\Omega|\mathbf{S})p(\mathbf{S})$$

as given in Equation 4.4. The expression of the likelihood is as given in Equation 4.5 (p. 72), restated here for convenience:

$$p(\mathbf{Y}|\Omega,\mathbf{S},\Psi) \propto \left(\prod_{t=1}^{N_T}\sigma_t\right)^{-n} \times$$
$$\prod_{i=1}^{n}\exp\left[-\frac{1}{2}\sum_{t=1}^{N_T}\frac{\left(y_{it}-\beta_{0t}-\sum_{\ell\in\upsilon_Q(t)}\phi_{\ell t}q_{i\ell}-\sum_{k\in\upsilon_Y(t)}\beta_{kt}y_{ik}\right)^2}{\sigma_t^2}\right].$$

(5.3)

The priors, however, must be expanded to accompany the fact that we now allow the number of directed edges in a model to vary, i.e. we have introduced a new parameter for the number of edges in model M, denoted E_M. The joint parameter prior is as defined as in Equations 4.6 and 4.7 (p. 73), for $\Omega \setminus E_M$ (Ω excluding E_M), and the prior for the number of edges is incorporated into the prior for the phenotype network structure of model M, $p(\mathbf{S})$. The graph structure is now defined by a set of nodes, which are known, the number of directed edges between traits (denoted E_M for model M) and the set of directed edges connecting nodes.

In absence of prior knowledge regarding the phenotype network structure, we assume independence among connections in the prior graph structure. That is, the state of the connection between nodes A and B is independent of the state of the connection between nodes A and C, for example. A connection has three possible states, illustrated in Figure 4.1 (p. 74). The prior on the state of the gth connection between two trait nodes in the phenotype network structure $\mathbf{S}$, denoted p_{gj}, is dependent on the directed edge present between these two nodes with assigned probabilities for $j \in \{1, 2, 3\}$, as given in Equation 4.8 (p. 75):

$$
p_{gj} = \begin{cases}
p_{g1}, & \text{if edge } e_g \text{ is reversed (e.g. } Y_1 \leftarrow Y_2) \\
p_{g2}, & \text{if edge } e_g \text{ is absent} \\
p_{g3}, & \text{if edge } e_g \text{ is forwards (e.g. } Y_1 \rightarrow Y_2),
\end{cases}
$$

where $\sum_{j=1}^{3} p_{gj} = 1 \ \forall g$. We may assign any values to p_{gj}, perhaps using prior knowledge of the phenotype network structure. For example, for edge g we could set the prior probability of the edge being absent as $p_{absent} = p_{g2}$, then the prior probability of the edge being present is $p_{present} = 1 - p_{absent}$ and the prior probability of the edge being forwards or reversed is then $p_{g1} = p_{g3} = p_{present}/2$.

The prior for the number of directed edges in the phenotype network structure of model *M* is a discretised exponential distribution, denoted $p(E_M)$:

$$p(E_M) = \xi e^{-\lambda E_M}$$

where

$$\sum_{E_M=0}^{\frac{1}{2}N_T(N_T-1)} e^{-\lambda E_M} = \frac{1}{\xi}$$

implying that

$$p(E_M) \propto e^{-\lambda E_M}. \tag{5.4}$$

Assuming *a priori* independence among connections, the joint probability for the phenotype network structure of model *M* is:

$$p^*(\mathbf{S}) = \left[\prod_{g=1}^{max(E)} p_{gj} \right] e^{-\lambda E_M}, \text{ for } j \in \{1,2,3\} \tag{5.5}$$

where there are $max(E) = \frac{1}{2}N_T(N_T-1)$ possible edge locations (or connections) in any model $M \in \mathbb{M}$ with N_T traits. Each connection can be reversed, absent or forwards, denoted p_{gj} for $j \in \{1,2,3\}$ (see above). It follows that there are $K = 3^{max(E)}$ possible directed phenotype network structures; however, this includes cyclic graphs. To exclude cyclic graphs we use the following indicator function:

$$I_a(\mathbf{S}) = \begin{cases} 1, & \text{if graph structure } \mathbf{S} \text{ is acyclic} \\ 0, & \text{otherwise.} \end{cases}$$

The prior probability distribution of all possible acyclic graph structures is then obtained by normalising these joint probabilities over all *K* acyclic graph struc-

tures to give:

$$p(\mathbf{S}) = \frac{p^*(\mathbf{S})I_a(\mathbf{S})}{\displaystyle\sum_{k=1}^{K} p^*(\mathbf{S}_k)I_a(\mathbf{S}_k)}. \tag{5.6}$$

In Chapter 3 we noted that overfitting was a concern when estimating an unknown phenotype network structure. Now that the chain is able to vary in dimension we have introduced a prior on the number of edges in a model that will help reduce overfitting. The hyperparameter λ is fixed; when equal to zero the prior does not differ between models of different dimensions, whereas a large value of λ causes simpler models, i.e. models with fewer directed edges, to be preferred.

5.1.3   Proposal distributions

The proposal distributions to move from model M to M', and from model M' to M remain the same as described in Section 4.1.3. The proposal ratio is:

$$\frac{q(\mathbf{u}_M',\mathbf{S},x'|\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'},\mathbf{S}',x|\Omega,\mathbf{S})} = \frac{q(\mathbf{u}_M'|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})}\frac{q(\mathbf{S}|x',\Omega',\mathbf{S}')}{q(\mathbf{S}'|x,\Omega,\mathbf{S})}\frac{q(x'|\Omega',\mathbf{S}')}{q(x|\Omega,\mathbf{S})} \tag{5.7}$$

where the probability of selecting move types x given M, and x' given M', are denoted $q(x|\Omega,\mathbf{S})$ and $q(x'|\Omega',\mathbf{S}')$, respectively, and are given in Table 5.1 (p. 147). The size of the neighbourhoods $Ne(M,x)$ and $Ne(M',x')$ are also given in Table 5.1, and the probabilities of selecting the phenotype network structures $\mathbf{S}'$ and $\mathbf{S}$, denoted $q(\mathbf{S}|x',\Omega',\mathbf{S}')$ and $q(\mathbf{S}'|x,\Omega,\mathbf{S})$, are as described in Section 4.1.3.

The proposal distributions for the parameters of the current and candidate models are denoted $q(\mathbf{u}_M'|\mathbf{S},x',\Omega',\mathbf{S}')$ and $q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})$, respectively. For the add and remove steps, the phenotype network is altered so we use an independence sampler to propose candidate parameter values. The proposal distributions

are based on the least squares estimates (LSE) of the parameters conditional on the proposed phenotype network structure. The LSE are effectively obtained from separate regressions of each trait on the traits QTL directly affecting it. This is both fast and efficient so we have an efficient sampler using well-tuned proposal distributions; however, once again, these could be improved by using multivariate normal proposal distributions.

Let us consider a move model M to M', the proposal distributions are:

$$\beta'_{0t}|\widehat{\beta}'_{0t}, \widehat{\sigma}'^2_t, V_{\beta'_{0t}} \quad \sim \quad N\left(\widehat{\beta}'_{0t}, \frac{1}{\tau}\widehat{\sigma}'^2_t V_{\beta'_{0t}}\right)$$

$$\beta'_{st}|\widehat{\beta}'_{st}, \widehat{\sigma}'^2_t, V_{\beta'_{st}} \quad \sim \quad N\left(\widehat{\beta}'_{st}, \frac{1}{\tau}\widehat{\sigma}'^2_t V_{\beta'_{st}}\right)$$

$$\sigma'^2_t|\widehat{\sigma}'^2_t \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}'^2_t, 1.5\widehat{\sigma}'^2_t\right),$$

as given in Equations 4.12 – 4.14 (p. 79). The expected values of the proposal distributions are conditional means: $\text{E}(\beta'_{0t}|\widehat{\beta}'_{0t}, \widehat{\sigma}'^2_t, V_{\beta'_{0t}}) = \widehat{\beta}'_{0t}$, $\text{E}(\beta'_{st}|\widehat{\beta}'_{st}, \widehat{\sigma}'^2_t, V_{\beta'_{st}}) = \widehat{\beta}'_{st}$, and $\text{E}(\sigma'^2_t|\widehat{\sigma}'^2_t) = \widehat{\sigma}'^2_t$. Let parameter p be any of $\beta'_{0t}, \beta'_{st}, \phi'_{\ell t}$ for $\ell \in \{1, \cdots, N_{qtl}\}$, $t \in \{1, \cdots, N_T\}$ and $s \in \{1, \cdots, N_T\}$ where $s \neq t$; for parameter p the term V_p is defined in Equation 4.11 (p. 78).

The number of edges in the candidate phenotype network structure, $E_{M'}$, is updated according to the move type, x, as follows:

$$E_{M'} = E_M + \eta \tag{5.8}$$

where

$$\eta = \begin{cases} 0, & \text{if } x = x^U, x^S, x^C \\ 1, & \text{if } x = x^A \\ -1, & \text{if } x = x^R, \end{cases}$$

and x^U, x^S, x^C, x^A and x^R denote the move types as defined in Table 5.1 (p. 147).

The proposal distributions to move from model M' back to model M are the same as Equations 4.15 – 4.17 (p. 80):

$$\beta_{0t}|\widehat{\beta}_{0t}, \widehat{\sigma}_t'^2, V_{\beta_{0t}} \quad \sim \quad N\left(\widehat{\beta}_{0t}, \frac{1}{\tau}\widehat{\sigma}_t'^2 V_{\beta_{0t}}\right)$$

$$\beta_{st}|\widehat{\beta}_{st}, \widehat{\sigma}_t'^2, V_{\beta_{st}} \quad \sim \quad N\left(\widehat{\beta}_{st}, \frac{1}{\tau}\widehat{\sigma}_t'^2 V_{\beta_{st}}\right)$$

$$\sigma_t^2|\widehat{\sigma}_t^2 \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}_t^2, 1.5\widehat{\sigma}_t^2\right).$$

Let parameter p be any of $\beta_{0t}, \beta_{st}, \phi_{\ell t}$ for $\ell \in \{1, \cdots, N_{qtl}\}$, $t \in \{1, \cdots, N_T\}$ and $s \in \{1, \cdots, N_T\}$ where $s \neq t$; for parameter p the term V_p is defined in Equation 4.11 (p. 78), and

$$E_M = E_{M'} + \eta' \tag{5.9}$$

where the move type x' is the reverse of move type x:

$$x' = \begin{cases} x, & \text{if } x = x^U, x^S, x^C \\ x^R, & \text{if } x = x^A \\ x^A, & \text{if } x = x^R, \end{cases}$$

and η' is conditional on the move type x':

$$\eta' = \begin{cases} 0, & \text{if } x' = x^U, x^S, x^C \\ 1, & \text{if } x' = x^A \\ -1, & \text{if } x' = x^R. \end{cases}$$

To summarize, the proposal ratio for the add and remove steps can be written as:

$$\frac{q(\mathbf{u}'_M, \mathbf{S}, x'|\Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x|\Omega, \mathbf{S})} = \frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \frac{q(\mathbf{S}|x', \Omega', \mathbf{S}')}{q(\mathbf{S}'|x, \Omega, \mathbf{S})} \frac{q(x'|\Omega', \mathbf{S}')}{q(x|\Omega, \mathbf{S})}$$

$$= \frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \frac{|Ne(M, x)|}{|Ne(M', x')|} \frac{q(x'|\Omega', \mathbf{S}')}{q(x|\Omega, \mathbf{S})}. \tag{5.10}$$

5.1.4 The Jacobian

The general expression of the Jacobian for a transition from model M to M' remains unchanged from Section 4.1.4, and the Jacobians for the add, remove and double steps are 1; see Appendix C.

5.2 The add step

An add step, also commonly referred to as a birth step in other RJMCMC literature, proposes the addition of a new directed edge (or trait-to-trait effect) to model M to form model M'. The proposed addition of a directed edge to model M means that the candidate model is of a higher dimension than the current model, i.e. $E_{M'} = E_M + 1$ and $m' = m + 1$. Given in Equation 5.10, the proposal ratio for the addition of any edge $Y_s \to Y_t$ (β_{st}) is:

$$
\begin{aligned}
\frac{q(\mathbf{u}'_M, \mathbf{S}, x' = x^R | \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x = x^A | \Omega, \mathbf{S})} &= \frac{q(\mathbf{u}'_M | \mathbf{S}, x' = x^R, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'} | \mathbf{S}', x = x^A, \Omega, \mathbf{S})} \frac{|Ne(M, x = x^A)|}{|Ne(M', x' = x^R)|} \frac{q(x' = x^R | \Omega', \mathbf{S}')}{q(x = x^A | \Omega, \mathbf{S})} \\
&= \frac{q(\beta_{0t} | \widehat{\beta}_{0t}, \widehat{\sigma}'^2_t, V_{\beta_{0t}})}{q(\beta'_{0t} | \widehat{\beta}'_{0t}, \widehat{\sigma}'^2_t, V_{\beta'_{0t}})} \frac{1}{q(\beta'_{st} | \widehat{\beta}'_{st}, \widehat{\sigma}'^2_t, V_{\beta'_{st}})} \frac{q(\sigma^2_t | \widehat{\sigma}^2_t)}{q(\sigma'^2_t | \widehat{\sigma}'^2_t)} \times \\
&\qquad \frac{\sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M]) I_{\in}(g|\mathbf{S})}{E_{M'}} \times \frac{q(x' = x^R | \Omega', \mathbf{S}')}{q(x = x^A | \Omega, \mathbf{S})}, \quad (5.11)
\end{aligned}
$$

as

$$
|Ne(M, x = x^A)| = \sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M]) I_{\in}(g|\mathbf{S})
$$

and

$$
|Ne(M', x' = x^R)| = E_{M'}
$$

with the operator A_g as described in Section 5.1.1 and the indicator functions $I_a(M)$ and $I_{\in}(g|\mathbf{S})$ as defined in Equation 4.3 (p. 72). The probability of selecting an add step, given M, is denoted $q(x = x^A | \Omega, \mathbf{S})$, and $q(x' = x^A | \Omega', \mathbf{S}')$ denotes the

probability of selecting a remove step, given M'. Both are given in Table 5.1 (p. 147).

The add step uses an independence sampler to generate new candidate parameter values. As the phenotype network structure is changed, the proposal distributions are based on the corresponding MLE. The parameters which get updated differ depending on the edge added to $\mathbf{S}$. This is due to conditional independence, whereby a trait is independent of all other traits given its parents. Therefore, if we add the edge corresponding to β_{st}, we generate β'_{st}, in addition to β'_{0t} and σ'^2_t for model M'. These proposal distributions are given in Equations 4.12 – 4.14 (pp. 79 – 80), and the proposal distributions for $M' \to M$ are given in Equations 4.15 – 4.17 (p. 80). It follows that for the addition of any edge $Y_s \to Y_t$ (β'_{st}) we have the following ratio of parameter proposal distributions:

$$\frac{q(\mathbf{u}'_M|\mathbf{S}, x' = x^R, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x = x^A, \Omega, \mathbf{S})} = \frac{q(\beta_{0t}|\widehat{\beta}_{0t}, \widehat{\sigma}'^2_t, V_{\beta_{0t}})}{q(\beta'_{0t}|\widehat{\beta}'_{0t}, \widehat{\sigma}'^2_t, V_{\beta'_{0t}})} \frac{1}{q(\beta'_{st}|\widehat{\beta}'_{st}, \widehat{\sigma}'^2_t, V_{\beta'_{st}})} \frac{q(\sigma^2_t|\widehat{\sigma}^2_t)}{q(\sigma'^2_t|\widehat{\sigma}'^2_t)}, \quad (5.12)$$

and the ratio of priors:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{0t}|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})} p(\beta'_{st}|\mathbf{S}') \frac{p(\sigma'^2_t|\mathbf{S}')}{p(\sigma^2_t|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})}, \quad (5.13)$$

where the parameter prior distributions are defined in Equation 4.7 (p. 73), and the phenotype network prior is as defined in Equation 5.6 (p. 150).

Following the general form of the acceptance probability given in Equation 5.2 (p. 145), and substituting in the proposal and prior ratios in Equations 5.11 and 5.13, the acceptance probability for adding β'_{st} is:

$$\begin{aligned} \alpha_{add} &= min(1, r_{add}), \quad \text{where} \\ r_{add} &= \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) p(\Omega'|\mathbf{S}') p(\mathbf{S}')}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r}) p(\Omega|\mathbf{S}) p(\mathbf{S})} \frac{q(\mathbf{u}'_M|\mathbf{S}, x' = x^R, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x = x^A, \Omega, \mathbf{S})} \times \\ & \qquad \frac{q(\mathbf{S}|x' = x^R, \Omega', \mathbf{S}')}{q(\mathbf{S}'|x = x^A, \Omega, \mathbf{S})} \frac{q(x' = x^R|\Omega', \mathbf{S}')}{q(x = x^A|\Omega, \mathbf{S})} \left| \frac{\partial(\Omega', \mathbf{u}'_M)}{\partial(\Omega, \mathbf{u}_{M'})} \right| \end{aligned}$$

$$
\begin{aligned}
r_{add} \;=\;& \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}\frac{p(\beta'_{0t}|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})}p(\beta'_{st}|\mathbf{S}')\frac{p(\sigma'^2_t|\mathbf{S}')}{p(\sigma^2_t|\mathbf{S})}\frac{p(\mathbf{S}')}{p(\mathbf{S})}\times \\[2mm]
& \frac{q(\beta_{0t}|\widehat{\beta}_{0t},\widehat{\sigma}'^2_t,V_{\beta_{0t}})}{q(\beta'_{0t}|\widehat{\beta}'_{0t},\widehat{\sigma}'^2_t,V_{\beta'_{0t}})}\frac{1}{q(\beta'_{st}|\widehat{\beta}'_{st},\widehat{\sigma}'^2_t,V_{\beta'_{st}})}\frac{q(\sigma^2_t|\widehat{\sigma}^2_t)}{q(\sigma'^2_t|\widehat{\sigma}'^2_t)}\times \\[2mm]
& \frac{\sum_{g=1}^{N_T(N_T-1)}I_a(A_g[M])I_\in(g|\mathbf{S})}{E_{M'}}\frac{q(x'=x^R|\Omega',\mathbf{S}')}{q(x=x^A|\Omega,\mathbf{S})}.
\end{aligned}
\tag{5.14}
$$

The general expression of the likelihood is given in Equation 5.3 (p. 147). If we assume a uniform prior on the state of a connection between any two trait nodes, the ratio of phenotype network priors simplifies to $\frac{p(\mathbf{S}')}{p(\mathbf{S})}=\exp(-\lambda)$. However, this then implies that an edge is more likely to be included in the phenotype network structure (in one of two directions) than not. The Jacobian is 1 (see Appendix C.3).

5.2.1 Example of an add step

Suppose that we want to carry out an add step to move from model M to model M'; the example causal network structures are given in Figure 5.1. Let us propose the addition of the trait-to-trait effect of trait 2 on trait 3 (β'_{23}) selected randomly from $Ne(M,x=x^A)$ with $q(\mathbf{S}'|x=x^A,\Omega,\mathbf{S})=\frac{1}{2}$. We could have also proposed the addition of β'_{32} as this would also create an acyclic model.

In this example we have a set of current parameters, Ω, and a set of candidate parameters, Ω':

$$
\begin{aligned}
\Omega \;=\;& \{E_M,\beta_{01},\beta_{02},\beta_{03},\beta_{12},\beta_{13},\phi_{11},\phi_{22},\phi_{33},\sigma^2_1,\sigma^2_2,\sigma^2_3\} \\
\Omega' \;=\;& \{E_{M'},\beta'_{01},\beta'_{02},\beta'_{03},\beta'_{12},\beta'_{13},\beta'_{23},\phi'_{11},\phi'_{22},\phi'_{33},\sigma'^2_1,\sigma'^2_2,\sigma'^2_3\}.
\end{aligned}
$$

The new directed edge (β'_{23}) specifies which parameters need to be updated in
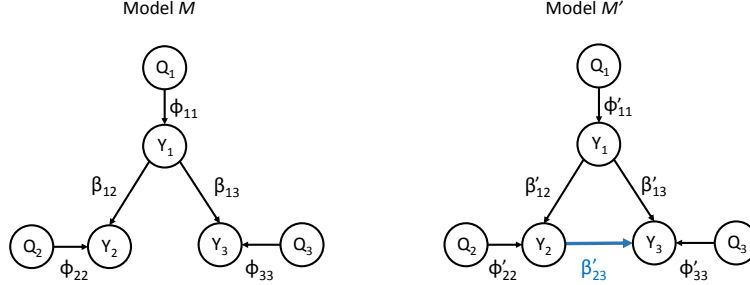
Figure 5.1: *Example causal network structures for both the current (M) and candidate (M′) models, given an add step proposing the addition of the trait-to-trait effect of trait 2 on trait 3 (β'_{23}).*

the candidate model. Here we generate the following parameters for model M': β'_{23}, β'_{03} and σ'^2_3. Furthermore, the number of directed edges in the candidate phenotype network structure increases by one, i.e. $E_{M'} = E_M + 1$. It follows that:

$$\{\beta'_{01}, \beta'_{02}, \beta'_{12}, \beta'_{13}, \phi'_{11}, \phi'_{22}, \phi'_{33}, \sigma'^2_1, \sigma'^2_2\} = \{\beta_{01}, \beta_{02}, \beta_{12}, \beta_{13}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma^2_1, \sigma^2_2\}.$$

To move from model M to M', new parameters are generated from the following proposal distributions:

$$
\begin{aligned}
E_{M'} &= E_M + 1 &&\text{(see Eqn. 5.8)} \\
\beta'_{03} | \widehat{\beta}'_{03}, \widehat{\sigma}'^2_3, V_{\beta'_{03}} &\sim N\left(\widehat{\beta}'_{03}, \frac{1}{\tau}\widehat{\sigma}'^2_3 V_{\beta'_{03}}\right) &&\text{(see Eqn. 4.12)} \\
\beta'_{23} | \widehat{\beta}'_{23}, \widehat{\sigma}'^2_3, V_{\beta'_{23}} &\sim N\left(\widehat{\beta}'_{23}, \frac{1}{\tau}\widehat{\sigma}'^2_3 V_{\beta'_{23}}\right) &&\text{(see Eqn. 4.13)} \\
\sigma'^2_3 | \widehat{\sigma}'^2_3 &\sim \text{Unif}\left(0.5\widehat{\sigma}'^2_3, 1.5\widehat{\sigma}'^2_3\right) &&\text{(see Eqn. 4.14).}
\end{aligned}
$$

To move from model M' to M, the parameter proposal distributions are:

$$
\begin{aligned}
E_M &= E_{M'} - 1 && \text{(see Eqn. 5.9)} \\
\beta_{03}|\widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}} &\sim N\left(\widehat{\beta}_{03}, \frac{1}{\tau}\widehat{\sigma}_3^2 V_{\beta_{03}}\right) && \text{(see Eqn. 4.15)} \\
\sigma_3^2|\widehat{\sigma}_3^2 &\sim \text{Unif}\left(0.5\widehat{\sigma}_3^2, 1.5\widehat{\sigma}_3^2\right) && \text{(see Eqn. 4.17)}
\end{aligned}
$$

where for parameter p, V_p is as defined in Equation 4.11 (p. 78).

The neighbourhood for model M given an add step, $|Ne(M, x = x^A)|$, and for model M' given a remove step, $|Ne(M', x' = x^R)|$, are given in Figure 5.2, such that:

$$
\frac{|Ne(M, x = x^A)|}{|Ne(M', x' = x^R)|} = \frac{2}{3}.
$$
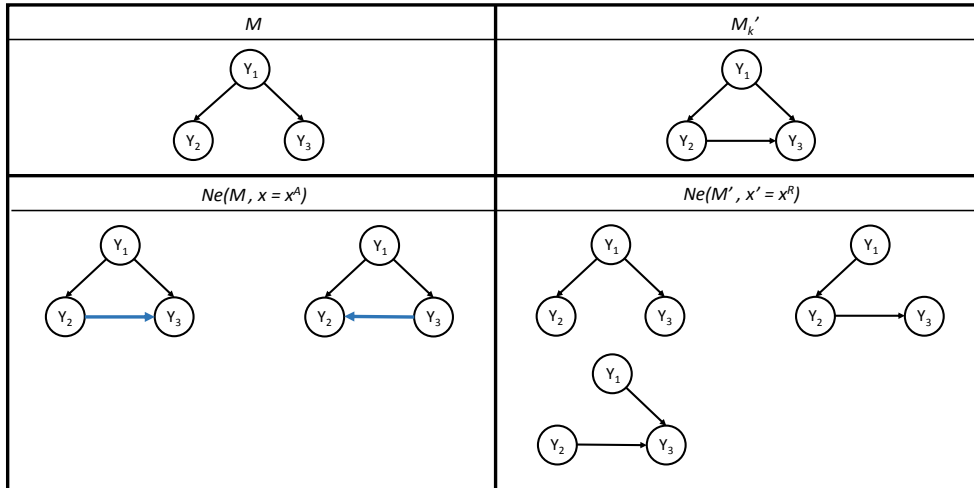
Figure 5.2: *The neighbourhood for the current model, given an add step, and for the candidate model, given a remove step; denoted $Ne(M, x = x^A)$ and $Ne(M', x' = x^R)$, respectively.*

Here, $q(x' = x^R | \Omega', \mathbf{S}')$ is the probability of a remove step given $E_{M'} = max(E) = 3$ edges, given in Table 5.1 (p. 147) as 0.25. Similarly, $q(x = x^A | \Omega, \mathbf{S}) = 0.125$ such that:

$$\frac{q(x' = x^R | \Omega', \mathbf{S}')}{q(x = x^A | \Omega, \mathbf{S})} = \frac{0.25}{0.125} = 2.$$

It follows that the proposal ratio for this example is:

$$\frac{q(\mathbf{u}'_M | \mathbf{S}, x' = x^R, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'} | \mathbf{S}', x = x^A, \Omega, \mathbf{S})} \frac{q(\mathbf{S} | x' = x^R, \Omega', \mathbf{S}')}{q(\mathbf{S}' | x = x^A, \Omega, \mathbf{S})} \frac{q(x' = x^R | \Omega', \mathbf{S}')}{q(x = x^A | \Omega, \mathbf{S})}$$

$$= \frac{q(\beta_{03} | \widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}}) q(\sigma_3^2 | \widehat{\sigma}_3^2)}{q(\beta'_{03} | \widehat{\beta}'_{03}, \widehat{\sigma}_3'^2, V_{\beta'_{03}}) q(\beta'_{23} | \widehat{\beta}'_{23}, \widehat{\sigma}_3'^2, V_{\beta'_{23}}) q(\sigma_3'^2 | \widehat{\sigma}_3'^2)} \times \frac{4}{3}.$$

Similarly, the prior ratio is:

$$\frac{p(\Omega' | \mathbf{S}')}{p(\Omega | \mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{03} | \mathbf{S}') p(\beta'_{23} | \mathbf{S}') p(\sigma_3'^2 | \mathbf{S}')}{p(\beta_{03} | \mathbf{S}) p(\sigma_3^2 | \mathbf{S})} \times \exp(-\lambda),$$

assuming the prior on the state of a connection between nodes is uniform, i.e. $p_{1g} = p_{2g} = p_{3g} = 1/3$. It follows that the acceptance probability for this example add step is:

$$\alpha_{add} = min(1, r_{add}), \quad \text{where}$$

$$r_{add} = \frac{p(\mathbf{Y} | \Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})}{p(\mathbf{Y} | \Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})} \frac{p(\beta'_{03} | \mathbf{S}') p(\beta'_{23} | \mathbf{S}') p(\sigma_3'^2 | \mathbf{S}')}{p(\beta_{03} | \mathbf{S}) p(\sigma_3^2 | \mathbf{S})} \exp(-\lambda) \times$$

$$\frac{q(\beta_{03} | \widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}}) q(\sigma_3^2 | \widehat{\sigma}_3^2)}{q(\beta_{03}' | \widehat{\beta}'_{03}, \widehat{\sigma}_3'^2, V_{\beta'_{03}}) q(\beta_{23}' | \widehat{\beta}'_{23}, \widehat{\sigma}_3'^2, V_{\beta'_{23}}) q(\sigma_3'^2 | \widehat{\sigma}_3'^2)} \times \frac{4}{3}.$$

The general expression of the likelihood is given in Equation 5.3 (p. 147), and the Jacobian is 1 (see Appendix C.3).

5.3 The remove step

A remove step, also commonly referred to as a death step in other RJMCMC literature, proposes the removal of an existing directed edge (or trait-to-trait effect) from model M to form model M'. The proposed removal of a directed edge from model M means that the candidate model is of a lower dimension than the current model, i.e. $E_{M'} = E_M - 1$ and $m' = m - 1$.

Given in Equation 5.10 (p. 152), the proposal ratio for the removal of any edge $Y_s \to Y_t$ (β_{st}) is:

$$
\begin{aligned}
\frac{q(\mathbf{u}'_M, \mathbf{S}, x' = x^A | \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x = x^R | \Omega, \mathbf{S})} &= \frac{q(\mathbf{u}'_M | \mathbf{S}, x' = x^A, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'} | \mathbf{S}', x = x^R, \Omega, \mathbf{S})} \frac{|Ne(M, x = x^R)|}{|Ne(M', x' = x^A)|} \frac{q(x' = x^A | \Omega', \mathbf{S}')}{q(x = x^R | \Omega, \mathbf{S})} \\
&= \frac{q(\beta_{0t} | \widehat{\beta}_{0t}, \widehat{\sigma}_t^2, V_{\beta_{0t}})}{q(\beta'_{0t} | \widehat{\beta}'_{0t}, \widehat{\sigma}_t'^2, V_{\beta'_{0t}})} q(\beta_{st} | \widehat{\beta}_{st}, \widehat{\sigma}_t^2, V_{\beta_{st}}) \frac{q(\sigma_t^2 | \widehat{\sigma}_t^2)}{q(\sigma_t'^2 | \widehat{\sigma}_t'^2)} \times \\
& \quad \frac{E_M}{\sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M']) I_{\in}(g | \mathbf{S}')} \frac{q(x' = x^A | \Omega', \mathbf{S}')}{q(x = x^R | \Omega, \mathbf{S})}, \quad (5.15)
\end{aligned}
$$

as

$$
|Ne(M, x = x^R)| = E_M
$$

and

$$
|Ne(M', x' = x^A)| = \sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M']) I_{\in}(g | \mathbf{S}')
$$

with the operator A_g as described in Section 5.1.1 and the indicator functions $I_a(M)$ and $I_{\in}(g | \mathbf{S})$ as defined in Equation 4.3 (p. 72). The probability of selecting a remove step, given M, is denoted $q(x = x^R | \Omega, \mathbf{S})$, and $q(x' = x^A | \Omega', \mathbf{S}')$ denotes the probability of selecting an add step, given M'. Both are given in Table 5.1 (p. 147).

Just like the add step, the remove step uses an independence sampler to generate new candidate parameter values. As the phenotype network structure is

changed, the proposal distributions are based on the corresponding LSE. The parameters which get updated differ depending on the edge removed from $\mathbf{S}$. This is due to conditional independence, whereby a trait is independent of all other traits given its parents. Therefore, if we remove the directed edge corresponding to β_{st}, we would only need to generate β'_{0t} and σ'^2_t for M'. These proposal distributions are given in Equations 4.12 – 4.14 (p. 79), and the proposal distributions for $M' \rightarrow M$ are given in Equations 4.15 – 4.17 (p. 80). It follows that for the removal of β_{st} we have the following ratio of parameter proposal distributions:

$$\frac{q(\mathbf{u}'_M|\mathbf{S},x'=x^A,\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x=x^R,\Omega,\mathbf{S})} = \frac{q(\beta_{0t}|\widehat{\beta}_{0t},\widehat{\sigma}_t^2,V_{\beta_{0t}})}{q(\beta'_{0t}|\widehat{\beta}'_{0t},\widehat{\sigma}'^2_t,V_{\beta'_{0t}})}q(\beta_{st}|\widehat{\beta}_{st},\widehat{\sigma}'^2_t,V_{\beta_{st}})\frac{q(\sigma_t^2|\widehat{\sigma}_t^2)}{q(\sigma'^2_t|\widehat{\sigma}'^2_t)},$$

and the ratio of priors:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{0t}|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})}\frac{1}{p(\beta_{st}|\mathbf{S})}\frac{p(\sigma'^2_t|\mathbf{S}')}{p(\sigma_t^2|\mathbf{S})}\frac{p(\mathbf{S}')}{p(\mathbf{S})}, \qquad (5.16)$$

where the parameter prior distributions are defined in Equation 4.7 (p. 73), and the phenotype network prior is as defined in Equation 5.6 (p. 150).

Following the general form of the acceptance probability given in Equation 5.2 (p. 145), and substituting in the proposal and prior ratios in Equations 5.15 and 5.16, the acceptance probability for removing β_{st} is:

$$
\begin{aligned}
\alpha_{remove} &= min(1,r_{remove}), \quad \text{where} \\
r_{remove} &= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})}\frac{q(\mathbf{u}'_M|\mathbf{S},x'=x^A,\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x=x^R,\Omega,\mathbf{S})} \times \\
&\quad \frac{q(\mathbf{S}|x'=x^A,\Omega',\mathbf{S}')}{q(\mathbf{S}'|x=x^R,\Omega,\mathbf{S})}\frac{q(x'=x^A|\Omega',\mathbf{S}')}{q(x=x^R|\Omega,\mathbf{S})}\left|\frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})}\right|
\end{aligned}
$$

$$r_{remove} = \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})} \frac{p(\beta'_{0t}|\mathbf{S}')}{p(\beta_{0t}|\mathbf{S})} \frac{1}{p(\beta_{st}|\mathbf{S})} \frac{p(\sigma_t'^2|\mathbf{S}')}{p(\sigma_t^2|\mathbf{S})} \frac{p(\mathbf{S}')}{p(\mathbf{S})} \times$$

$$\frac{q(\beta_{0t}|\widehat{\beta}_{0t},\widehat{\sigma}_t^2,V_{\beta_{0t}})}{q(\beta'_{0t}|\widehat{\beta}'_{0t},\widehat{\sigma}_t'^2,V_{\beta'_{0t}})} q(\beta_{st}|\widehat{\beta}_{st},\widehat{\sigma}_t'^2,V_{\beta_{st}}) \frac{q(\sigma_t^2|\widehat{\sigma}_t^2)}{q(\sigma_t'^2|\widehat{\sigma}_t'^2)} \times$$

$$\frac{E_M}{\sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M'])I_\in(g|\mathbf{S}')} \frac{q(x'=x^A|\Omega',\mathbf{S}')}{q(x=x^R|\Omega,\mathbf{S})}. \tag{5.17}$$

The general expression of the likelihood is given in Equation 5.3 (p. 147). If we assume a uniform prior on the state of a connection between any two trait nodes, then the ratio of phenotype network priors simplifies to $\frac{p(\mathbf{S}')}{p(\mathbf{S})} = \exp(\lambda)$. However, this then implies that an edge is more likely to be included in the phenotype network structure (in one of two directions) than not. The Jacobian is 1 (see Appendix C.4).

5.3.1 Example of a remove step

Suppose that we want to carry out a removal step to move from model M to model M'; the example causal network structures are given in Figure 5.3. We can propose the removal of any of the three directed edges comprising $\mathbf{S}$. Let us propose the removal of the trait-to-trait effect of trait 2 on trait 3 (β_{23}) selected randomly from $Ne(M, x = x^R)$ with $q(\mathbf{S}'|x = x^R, \Omega, \mathbf{S}) = \frac{1}{3}$. In this example we have a set of current parameters, Ω, and a set of candidate parameters, Ω':

$$\Omega = \{E_M, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{12}, \beta_{13}, \beta_{23}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma_1^2, \sigma_2^2, \sigma_3^2\}$$
$$\Omega' = \{E_{M'}, \beta'_{01}, \beta'_{02}, \beta'_{03}, \beta'_{12}, \beta'_{13}, \phi'_{11}, \phi'_{22}, \phi'_{33}, \sigma_1'^2, \sigma_2'^2, \sigma_3'^2\}.$$

The removed edge (β_{23}) specifies which parameters need to be updated in the candidate model. Here we generate the following parameters for model M': β'_{03}
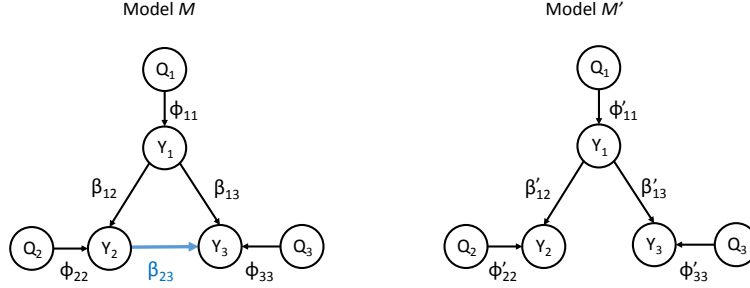
Figure 5.3: *Example causal network structures for both the current (M) and candidate (M′) models, given a removal step proposing the removal of the trait-to-trait effect of trait 2 on trait 3 (β_{23}).*

and $\sigma_3'^2$. Furthermore, the number of directed edges in the candidate phenotype network structure decreases by one, i.e. $E_{M'} = E_M - 1$. It follows that:

$$\{\beta_{01}', \beta_{02}', \beta_{12}', \beta_{13}', \phi_{11}', \phi_{22}', \phi_{33}', \sigma_1'^2, \sigma_2'^2\} = \{\beta_{01}, \beta_{02}, \beta_{12}, \beta_{13}, \phi_{11}, \phi_{22}, \phi_{33}, \sigma_1^2, \sigma_2^2\}.$$

To move from model M to M', new parameters are generated from the following proposal distributions:

$$
\begin{aligned}
E_{M'} &= E_M - 1 \qquad \text{(see Eqn. 5.8)} \\
\beta_{03}' | \widehat{\beta}_{03}', \widehat{\sigma}_3'^2, V_{\beta_{03}'} &\sim N\left(\widehat{\beta}_{03}', \frac{1}{\tau}\widehat{\sigma}_3'^2 V_{\beta_{03}'}\right) \qquad \text{(see Eqn. 4.12)} \\
\sigma_3'^2 | \widehat{\sigma}_3'^2 &\sim \text{Unif}\left(0.5\widehat{\sigma}_3'^2, 1.5\widehat{\sigma}_3'^2\right) \qquad \text{(see Eqn. 4.14).}
\end{aligned}
$$

To move from model M' to M, the parameter proposal distributions are:

$$
\begin{aligned}
E_M &= E_{M'} + 1 \qquad \text{(see Eqn. 5.9)} \\
\beta_{03} | \widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}} &\sim N\left(\widehat{\beta}_{03}, \frac{1}{\tau}\widehat{\sigma}_3^2 V_{\beta_{03}}\right) \qquad \text{(see Eqn. 4.15)}
\end{aligned}
$$

$$\beta_{23}|\widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}} \quad \sim \quad N\left(\widehat{\beta}_{23}, \frac{1}{\tau}\widehat{\sigma}_3^2 V_{\beta_{23}}\right) \qquad \text{(see Eqn. 4.16)}$$

$$\sigma_3^2|\widehat{\sigma}_3^2 \quad \sim \quad \text{Unif}\left(0.5\widehat{\sigma}_3^2, 1.5\widehat{\sigma}_3^2\right) \qquad \text{(see Eqn. 4.17)},$$

where for parameter p, V_p is as defined in Equation 4.11 (p. 78).

The neighbourhood for model M given a remove step, $Ne(M, x = x^R)$, and for model M' given an add step, $Ne(M', x' = x^A)$, are given in Figure 5.4, such that:

$$\frac{|Ne(M, x = x^R)|}{|Ne(M', x' = x^A)|} = \frac{3}{2}.$$

Here, $q(x' = x^A|\Omega', \mathbf{S}')$ is the probability of an add step given $E_{M'} = 2$ edges,
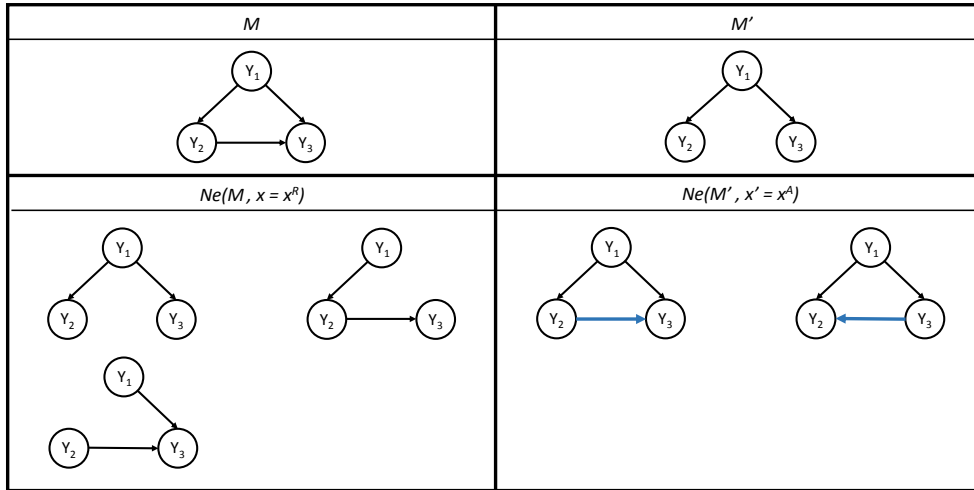
Figure 5.4: *The neighbourhood for the current model, given a remove step, and for the candidate model, given an add step; denoted $Ne(M, x = x^R)$ and $Ne(M', x' = x^A)$, respectively.*

given in Table 5.1 (p. 147) as 0.125. Similarly, $q(x = x^R|\Omega, \mathbf{S}) = 0.25$ such that

$$\frac{q(x' = x^A|\Omega', \mathbf{S}')}{q(x = x^R|\Omega, \mathbf{S})} = \frac{0.125}{0.25} = \frac{1}{2}.$$

It follows that the proposal ratio for this example is:

$$\frac{q(\mathbf{u}'_M|\mathbf{S}, x' = x^A, \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x = x^R, \Omega, \mathbf{S})} \frac{q(\mathbf{S}|x', \Omega', \mathbf{S}')}{q(\mathbf{S}'|x, \Omega, \mathbf{S})} \frac{q(x' = x^A|\Omega', \mathbf{S}')}{q(x = x^R|\Omega, \mathbf{S})}$$

$$= \frac{q(\beta_{03}|\widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}}) q(\beta_{23}|\widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}}) q(\sigma_3^2|\widehat{\sigma}_3^2)}{q(\beta'_{03}|\widehat{\beta}'_{03}, \widehat{\sigma}_3'^2, V_{\beta'_{03}}) q(\sigma_3'^2|\widehat{\sigma}_3'^2)} \times \frac{3}{4}.$$

Similarly, the prior ratio is:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta'_{03}|\mathbf{S}')p(\sigma_3'^2|\mathbf{S}')}{p(\beta_{03}|\mathbf{S})p(\beta_{23}|\mathbf{S})p(\sigma_3^2|\mathbf{S})} \times \exp(\lambda),$$

assuming the prior on the state of a connection between nodes is uniform, i.e. $p_{1g} = p_{2g} = p_{3g} = 1/3$. It follows that the acceptance probability for this example remove step is:

$$\alpha_{remove} = min(1, r_{remove})$$

$$r_{remove} = \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})} \frac{p(\beta'_{03}|\mathbf{S}')p(\sigma_3'^2|\mathbf{S}')}{p(\beta_{03}|\mathbf{S})p(\beta_{23}|\mathbf{S})p(\sigma_3^2|\mathbf{S})} \exp(\lambda) \times$$

$$\frac{q(\beta_{03}|\widehat{\beta}_{03}, \widehat{\sigma}_3^2, V_{\beta_{03}}) q(\beta_{23}|\widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}}) q(\sigma_3^2|\widehat{\sigma}_3^2)}{q(\beta'_{03}|\widehat{\beta}'_{03}, \widehat{\sigma}_3'^2, V_{\beta'_{03}}) q(\sigma_3'^2|\widehat{\sigma}_3'^2)} \times \frac{3}{4}.$$

The general expression of the likelihood is given in Equation 5.3 (p. 147), and the Jacobian is 1 (see Appendix C.4).

5.4 Summary of the Bayesian approach to estimating an unknown phenotype structure

1. Initialise the chain by specifying an initial causal network structure by sampling from the edge priors and confirming acyclicity. Generate new parameter values using the proposal distributions in Equations 4.22 – 4.25 (pp. 84 – 85), using the MLE for the initial model as the current parameter estimates. The initial model is now referred to as the current model, $M = \{\Omega, \mathbf{S}\}$.

2. Select a move type x with probability $q(x|\Omega, \mathbf{S})$, given in Table 5.1 (p. 147), where

$$\sum_{x \in \mathbb{X}} q(x|\Omega, \mathbf{S}) = 1 \quad \text{for all } (\Omega, \mathbf{S})$$

and x is one of

- **update** (x^U): update all parameters in the current model

- **add** (x^A): add a directed edge to the current phenotype network

- **remove** (x^R): remove a directed edge from the current phenotype network

- **reverse** (x^S): reverse the direction of one of the directed edges in the current phenotype network

- **relocate** (x^C): relocate one of the directed edges in the current phenotype network, i.e. delete one edge and propose a new, distinct directed edge.

3. Given move type x, propose a candidate phenotype network structure $\mathbf{S}'$ with

probability

$$q(\mathbf{S}'|x,\Omega,\mathbf{S}) = \begin{cases} \frac{1}{|Ne(M,x)|}, & \text{if } \mathbf{S}' \in Ne(M,x) \\ 0, & \text{otherwise} \end{cases}$$

where

$$\sum_{\mathbf{S}'\in\mathbb{S}} q(\mathbf{S}'|x,\Omega,\mathbf{S}) = 1 \quad \text{for all } (x,\Omega,\mathbf{S}).$$

4. Generate a set of random variables $\mathbf{u}_{M'} \in \mathbb{U}_{M'}$, with probability density

$$q(\mathbf{u}_{M'}|x,\mathbf{S}',\Omega,\mathbf{S})$$

and the parameters of the new state of the chain, Ω' are generated by the deterministic function $g_{MM'}$, such that

$$\Omega' = g_{MM'}(\Omega, \mathbf{u}_{M'}).$$

- If $x = x^U$ then the proposal distributions are given in Equations 4.22 – 4.25 (pp. 84 – 85), dependent on the current parameter estimates (Ω).

- If $x \in \{x^A, x^R, x^S, x^C\}$ then the phenotype network structure is altered and the parameter proposal distributions are given in Equations 4.12 – 4.14 (pp. 79 – 80).

5. Calculate the acceptance probability:

$$\alpha = min(1,r)$$

where

$$r = \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \times$$
$$\frac{q(\mathbf{S}|x',\Omega',\mathbf{S}')}{q(\mathbf{S}'|x,\Omega,\mathbf{S})} \frac{q(x'|\Omega',\mathbf{S}')}{q(x|\Omega,\mathbf{S})} \left| \frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})} \right|$$

and

r_{update} is given in Equation 4.31 (p. 86)

r_{add} is given in Equation 5.14 (p. 155)

r_{remove} is given in Equation 5.17 (p. 161)

$r_{reverse}$ is given in Equation 4.34 (p. 90)

$r_{relocate}$ is given in Equation 4.37 (p. 96).

6. Determine whether to accept or reject the candidate model by drawing a random number from a uniform distribution (between 0 and 1).

- If the random number is less than the α, accept the candidate model. The candidate model is now the current model: $M = \{\Omega',\mathbf{S}'\}$.

- If the random number is greater than α, reject the candidate model. The current model remains unchanged: $M = \{\Omega,\mathbf{S}\}$.

7. Repeat steps $2 - 6$ until the chain has converged to a stationary state and a large number of samples from the posterior have been drawn.

5.5 Simulation study II

With the extension of our model search approach to include models of varying dimensions, we incorporated a prior for the number of edges in a phenotype net-

work structure, designed to prefer simpler models over more complex ones, given the hyperparameter $\lambda > 0$. This simulation study demonstrates how the value of λ influences the performance of the reversible jump algorithm, and how this may vary for different sized QTL and trait-to-trait effects.

Following Appendix B, data was simulated for five traits and five QTL with the true causal network structure given in Figure 4.7, included below for convenience. The true parameter values were varied in order to determine how this approach is influenced by different combinations of QTL and trait-to-trait effects. Both small and large QTL effects and trait-to-trait effects are selected with values between 0.1 and 1.2. This range was based upon those values in published data sets; for example, Dhungana et al. (2007) and Mi et al. (2010). We note that it is possible for the effects to be positive or negative. We simulate six data sets, one for each of the parameter sets listed in Table 5.2, each with $n = 500$ individuals, labelled 5.1 − 5.6 to indicate that they relate to this chapter.
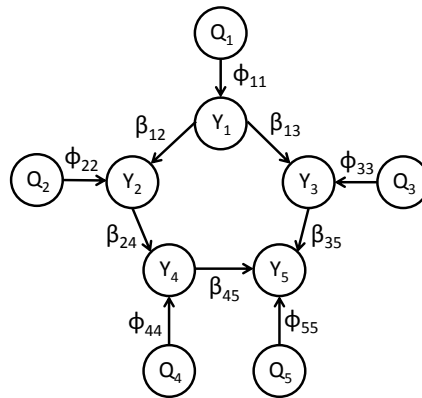
The true causal network structure for the simulated data sets 5.1 − 5.6. The true parameter values for each data set are listed in Table 5.2.

Table 5.2: *True parameter values for data sets 5.1 – 5.6.*

parameter	5.1	5.2	5.3	5.4	5.5	5.6
			data set			
β_{01}	0.5	0.5	0.5	0.5	0.5	0.5
β_{02}	0.5	0.5	0.5	0.5	0.5	0.5
β_{03}	0.5	0.5	0.5	0.5	0.5	0.5
β_{04}	0.5	0.5	0.5	0.5	0.5	0.5
β_{05}	0.5	0.5	0.5	0.5	0.5	0.5
β_{12}	0.2	0.6	1.1	0.2	0.6	1.1
β_{13}	0.2	0.6	1.1	0.2	0.6	1.1
β_{24}	0.2	0.6	1.1	0.2	0.6	1.1
β_{35}	0.2	0.6	1.1	0.2	0.6	1.1
β_{45}	0.2	0.6	1.1	0.2	0.6	1.1
ϕ_{11}	0.1	0.1	0.1	1.2	1.2	1.2
ϕ_{22}	0.1	0.1	0.1	1.2	1.2	1.2
ϕ_{33}	0.1	0.1	0.1	1.2	1.2	1.2
ϕ_{44}	0.1	0.1	0.1	1.2	1.2	1.2
ϕ_{55}	0.1	0.1	0.1	1.2	1.2	1.2
σ_1^2	1.0	1.0	1.0	1.0	1.0	1.0
σ_2^2	1.0	1.0	1.0	1.0	1.0	1.0
σ_3^2	1.0	1.0	1.0	1.0	1.0	1.0
σ_4^2	1.0	1.0	1.0	1.0	1.0	1.0
σ_5^2	1.0	1.0	1.0	1.0	1.0	1.0

- **Data set 5.1**: small QTL effects and small trait-to-trait effects

- **Data set 5.2**: small QTL effects and medium trait-to-trait effects

- **Data set 5.3**: small QTL effects and large trait-to-trait effects

- **Data set 5.4**: large QTL effects and small trait-to-trait effects

- **Data set 5.5**: large QTL effects and medium trait-to-trait effects

- **Data set 5.6**: large QTL effects and large trait-to-trait effects.

For each of the data sets 5.1 – 5.6 we consider different values of λ in the prior for the number of edges in the phenotype network structure, in order to determine what effect this has on the reversible jump MCMC algorithm for each data set. We consider values of λ corresponding to no complexity penalty ($\lambda = 0$) up to a high complexity penalty of $\lambda = 10$, which spans a large enough range for us to determine the effect of the prior for each simulated data set. The results are separated into two sections, 5.5.1 and 5.5.2, for small and large QTL effects, respectively.

Following the recommendation from (Neto et al. 2010) we ran long chains, with $N = 1,000,000$ iterations. Each had a burn-in period of 100,000 iterations removed and every 10th iteration was retained, so 90,000 states are stored. Five chains were run for each data set. Figures 5.5 and 5.9 were created using the first chain run for data sets 5.1 – 5.3 (small QTL effects) and 5.4 – 5.6 (large QTL effects), respectively. They present three key features to illustrate the performance of each chain; the mean number of edges in the visited models, the number of different models visited, and the joint posterior probability of the true model (denoted $p(M_{TRUE}|D)$), calculated as in Section 4.6.

Here we use the terms 'causal network' and 'phenotype network' interchangeably, as the genetic architecture is assumed known and therefore only the phenotype network structure can vary — the estimation of the phenotype network structure then implies the estimation of the causal network structure and vice versa.

5.5.1 Small QTL effects

Data sets 5.1 – 5.3 all have small QTL effects, with the trait-to-trait effects increasing in size from small (data set 5.1), to medium (data set 5.2), to large (data set 5.3). The results are displayed in Figure 5.5 and are also tabulated in Appendix
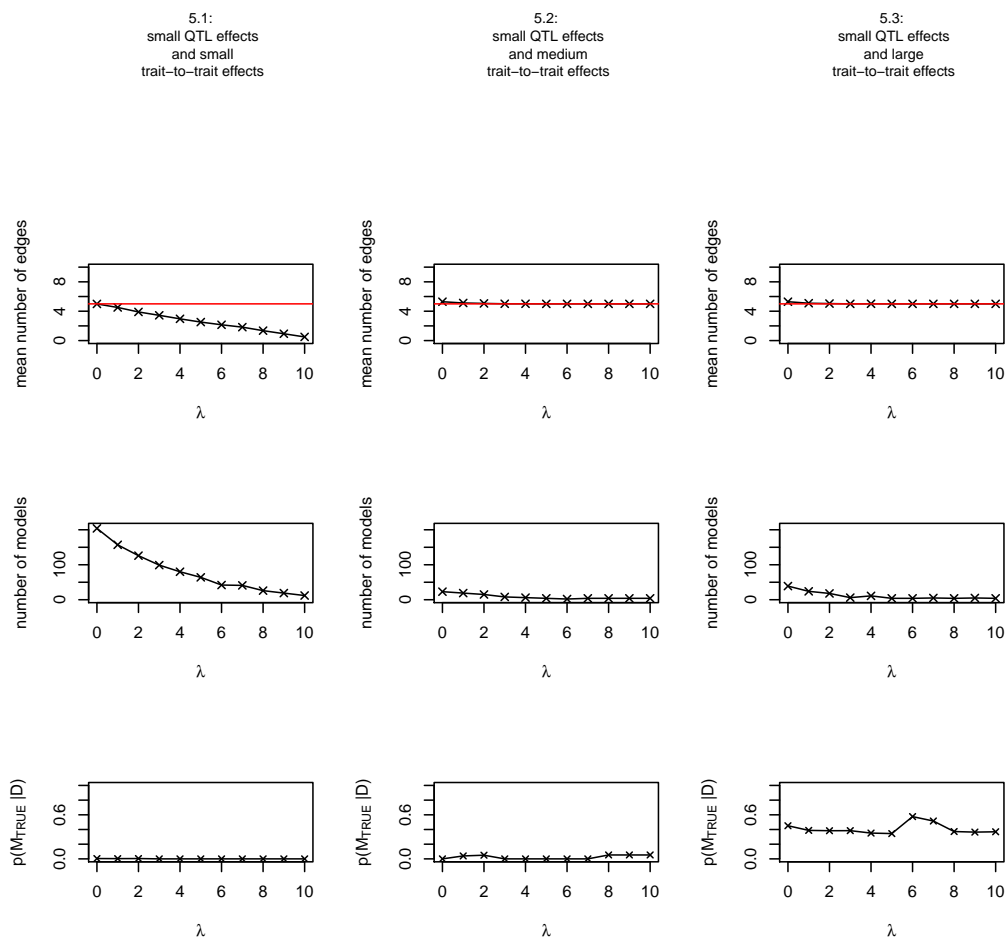
Figure 5.5: *Summary of the results from the first chain run for simulated data sets with small QTL effects, for $\lambda = 0, \cdots, 10$. The joint posterior probability of the true model is denoted $p(M_{TRUE}|D)$, and the true model has five directed edges, indicated by the red horizontal line in the plots for the mean number of edges.*

E (Table E.1, p. 276).

The combination of small trait-to-trait effects and small QTL effects (data set 5.1) was unable to identify the true causal network structure — the joint posterior probability was never greater than 0.0044. However, it did visit the most models,
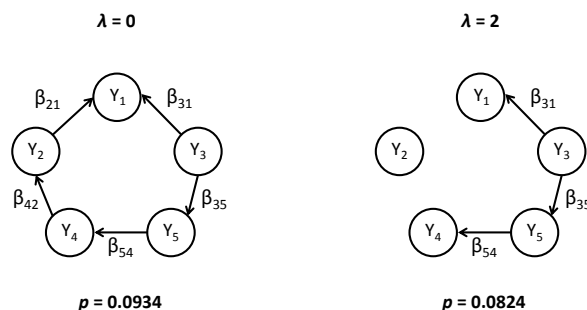
Figure 5.6: *The models with the greatest joint posterior probabilities (denoted p) for the first chain run for data set 5.1, when λ is equal to 0 and 2.*

the number of which decreased as the value of λ increased. The model with the greatest joint posterior probability differed depending on the value of λ; two example phenotype network structures are given in Figure 5.6 (p. 172). The most noticeable difference is that when $\lambda = 0$, the estimated model has five directed edges; however, increasing λ to 2 has removed two of the directed edges. This is to be expected as a large value of λ causes simpler models, i.e. models with fewer directed edges, to be preferred. With small effects, the true model has such a small signal that the chain is moving around easily, and can be influenced by the prior more than other data sets with larger effects: increasing λ resulted in the mean number of edges in the visited models to steadily decrease, as the prior caused simpler models (with fewer directed edges) to be preferred.

Keeping the QTL effects small and increasing the trait-to-trait effects to a 'medium' size (as in data set 5.2) does not significantly change the joint posterior probability of the true model (the greatest value of the posterior probability was 0.0536, when $\lambda = 8, 9, 10$; see Table E.1, p. 276). However, the trait-to-trait
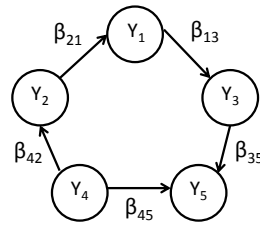
Figure 5.7: *The model with the greatest joint posterior probability for all values of λ for the first chain run for data set 5.2.*

effects have increased enough to improve the signal of the true model such that the mean number of edges in the visited models is now constant around 5, regardless of the size of λ used in the prior. We also note that the number of models searched has significantly decreased from when we had small trait-to-trait effects. At most, the chain visited 38 different models; we were able to estimate the number of edges correctly, but could not estimate the true model. The model with the greatest joint posterior probability is the same regardless of the size of λ used here, and is given in Figure 5.7. This phenotype network structure has 5 directed edges, as does the true model, although two of the directed edges have been reversed: including β_{21} and β_{42} instead of β_{12} and β_{24}. The chain is detecting a causal relationship, but it is unable to determine the direction of causality.

Further increasing the trait-to-trait effects to be 'large' (data set 5.3) resulted in the joint posterior probability of the true model increasing to around 0.4 for most values of λ. Interestingly, the relationship between the value of λ and the posterior probability of the true model is not monotonic, but instead there is some interplay between the two, as detailed in Figure 5.8. Figure 5.8 includes the two models with the greatest joint posterior probabilities (denoted M_{TRUE} and M_A), and the joint posterior probability for each, given λ. The difference between

the phenotype network structures of the two models is the reversal of one edge, with M_A including β_{21} instead of β_{12} as in M_{TRUE}. As both models include five directed edges, the prior would not favour one model over the other; however, it would alter the moves accepted by the chain, with a large value of λ in the prior encouraging the acceptance of moves to models with fewer directed edges — changing the set of models visited by each chain. We note that the joint posterior probability of the true model was gradually decreasing as λ was increased from 0 to 5, and the posterior probability of model M_A increases until it becomes the model with the greatest joint posterior probability. However, when $\lambda = 6, 7$ the posterior probability of the true model peaks before decreasing back to around 0.36 for larger values of λ.

For data set 5.3 the mean number of edges in the visited model was around the

	M_{TRUE}	M_A
$\lambda = 0$	0.4481	0.1803
$\lambda = 1$	0.3821	0.3726
$\lambda = 2$	0.3807	0.3486
$\lambda = 3$	0.3859	0.3928
$\lambda = 4$	0.3462	0.4924
$\lambda = 5$	0.3435	0.3968
$\lambda = 6$	0.5755	0.2800
$\lambda = 7$	0.5140	0.3175
$\lambda = 8$	0.3710	0.4391
$\lambda = 9$	0.3623	0.4136
$\lambda = 10$	0.3671	0.4049

Figure 5.8: *Models with the greatest joint posterior probability for different values of λ, for data set 5.3. The posterior probabilities in the blue boxes highlight the model with the greatest joint posterior probability.*

true number of edges in the model (5) for all values of λ used. The increase in the size of the trait-to-trait effects (from medium to large) resulted in more models being visited when $\lambda = 0$; however, as λ increased, this number decreased to just 4 models visited.

5.5.2 Large QTL effects

Data sets 5.4 – 5.6 all have large QTL effects, with the trait-to-trait effects increasing in size from small (data set 5.4), to medium (data set 5.5), to large (data set 5.5). The results are displayed in Figure 5.9 and are also tabulated in Appendix E (Table E.2, p. 277).

Data set 5.4 combined small trait-to-trait effects with large QTL effects, and we note that for $\lambda = 0$ fewer models were searched than in data set 5.1 (with small QTL effects). Furthermore, in data set 5.1 with small QTL effects, the mean number of edges in the estimated models continuously decreased as λ increased (see Figure 5.5, p. 171), but in data set 5.4 (with large QTL effects) the mean number of edges in the models visited by the chain appears more stable, equal to the number of edges in the true model until $\lambda = 4$ when it begins to decline. The joint posterior probability of the true model has greatly improved with larger QTL effects: going from 0.0044 (in data set 5.1 with small QTL effects) to 0.7239 (with large QTL effects) when $\lambda = 0$; see Table E.2 (p. 277). The joint posterior probability of the true model slightly increases as λ is increased from 0, until λ becomes too large ($\lambda = 4$) and it begins to decline. As the joint posterior probability of the true model decreases, the number of models visited by the chain slightly increases. Figure 5.10 gives a good example of our prior on the number of edges in the phenotype network structure prefers simple models with fewer directed edges; the true phenotype network structure is the model with the highest joint posterior
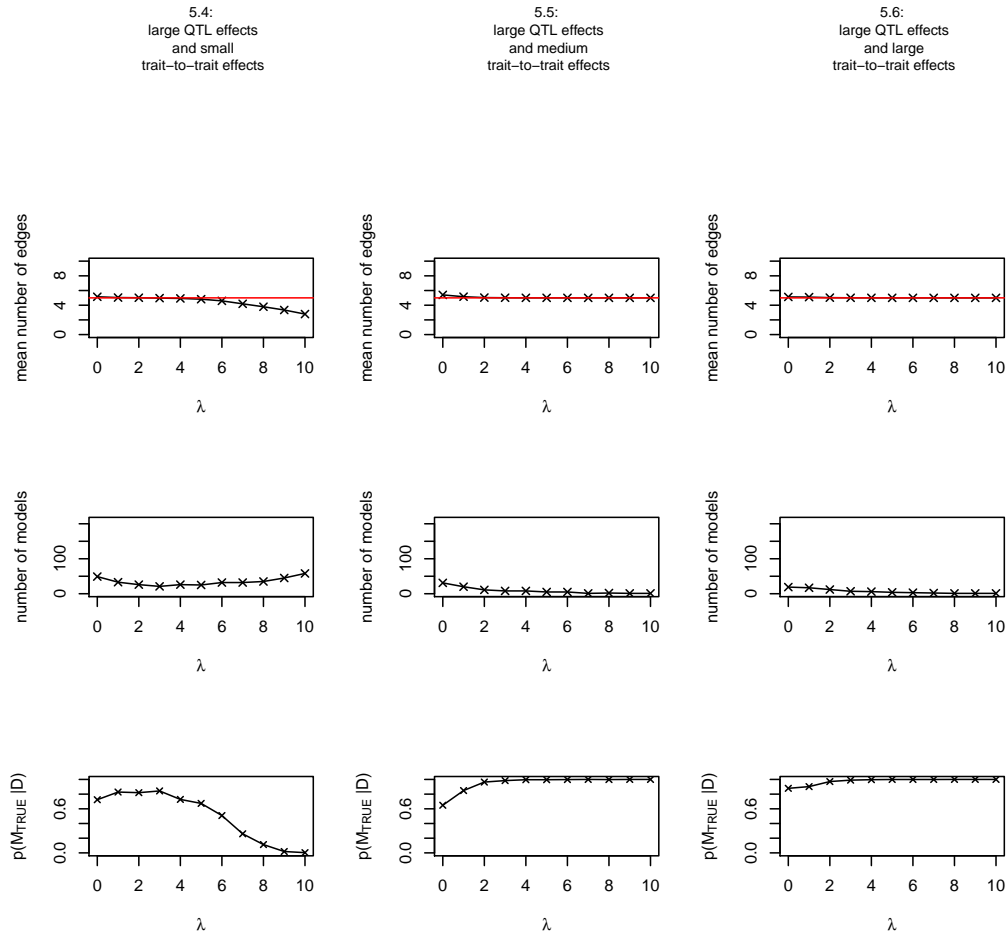
Figure 5.9: *Summary of results from the first chain run for simulated data sets with large QTL effects, for $\lambda = 0, \cdots, 10$. The joint posterior probability of the true model is denoted $p(M_{TRUE}|D)$, and the true model has five directed edges, indicated by the red horizontal line in the plots for the mean number of edges.*

probability for $0 \leq \lambda \leq 6$, then the edge corresponding to β_{24} is removed when $\lambda = 7$ and 8 (model M_A), and a second edge corresponding to β_{35} is removed for $\lambda = 9$ and 10 (model M_B). The model with the greatest joint posterior probability has fewer edges as λ increases. This is expected, as larger values of λ cause

simpler models to be preferred.

When small QTL effects were combined with medium-sized trait-to-trait effects in data set 5.2, the true model had a low joint posterior probability regardless of the value of λ (see Figure 5.9, p. 176, and Table E.2, p. 277). Now, with large QTL effects as in data set 5.5, we note that there has been a significant increase in the joint posterior probability of the true model: going from 0 (in data set 5.2 with small QTL effects) to 0.6486 (in data set 5.5 with large QTL effects) when $\lambda = 0$. Furthermore, the posterior probability increases as λ increases — the increase in the size of the trait-to-trait effects (from small to medium) has made the analysis less sensitive to the value of λ. The mean number of edges in the visited

	M_{TRUE}	M_A	M_A
$\lambda = 0$	0.7239		
$\lambda = 1$	0.8282	0.0066	
$\lambda = 2$	0.8199	0.0062	
$\lambda = 3$	0.8434	0.0306	
$\lambda = 4$	0.7280	0.0607	
$\lambda = 5$	0.6736	0.1364	0.0024
$\lambda = 6$	0.5083	0.2643	0.0171
$\lambda = 7$	0.2597	0.4147	0.0960
$\lambda = 8$	0.1130	0.4137	0.2316
$\lambda = 9$	0.0185	0.2821	0.4008
$\lambda = 10$	0.0031	0.1130	0.3710

Figure 5.10: *The joint posterior probabilities of models visited by the chain, for different values of λ, for data set 5.4. M_{TRUE} is the true model. The joint posterior probabilities in boxes highlight the model with the greatest joint posterior probability. If a posterior probability is not listed, then the model was not one of the top 10 models with the greatest posterior probabilities for the given value of λ.*

models and the number of models visited have changed very little from data set 5.2; however, fewer models have been visited than when small trait-to-trait effects are present (data sets 5.1 and 5.4). With large QTL effects and medium-sized trait-to-trait effects, the chain does not mix well, especially with large values of λ.

In data set 5.3, with small QTL effects and large trait-to-trait effects, we saw some interplay between the value of λ and the joint posterior probability of the true model (see Figure 5.5, p. 171); however, once we increased the size of the QTL effects (as in data set 5.6, with large QTL effects) we no longer see this relationship. When the trait-to-trait effects are large, increasing the size of the QTL effects has resulted in little change in the mean number of edges in the estimated models, and in the number of models visited. However, the posterior probability of the true model has greatly increased: beginning at 0.8789 when $\lambda = 0$ and going to 1 when $\lambda = 10$ (see Table E.2 in Appendix E, p. 277). Unfortunately, few models were searched for large values of λ, so mixing is poor for large QTL effects and large trait-to-trait effects.

Overall, our simulation study investigating the effects of our prior on the number of edges in a phenotype network structure, has revealed the following about the simulated data sets 5.1 – 5.6:

1. it is possible to identify the true causal network structure when there are small QTL effects; however, the trait-to-trait effects must be large (i.e. around 1.1)

2. there is not necessarily a monotonic relationship existing between the joint posterior probability of the true model and the value of λ in the prior for the

number of edges in a phenotype network structure

3. large QTL effects will, in general, lead to a greater joint posterior probability of the true model

4. large trait-to-trait effects will, in general, lead to a greater joint posterior probability of the true model

5. large QTL effects will, in general, make the RJMCMC sampler less sensitive to large values of λ in the prior for the number of edges in a phenotype network structure

6. the presence of small QTL effects and small trait-to-trait effects will mean the chain searches the model space well; however, the true model has a small signal and it is therefore difficult to determine the direction of causality between traits

7. large effects make it difficult for the chain to propose favourable moves, therefore mixing is slow and long chains are required.

In Chapter 4 we restricted the number of directed edges in the phenotype network structure; consequently observing poor mixing and many of the dimension independent parameters failed the Gelman–Rubin convergence diagnostic. Next, we briefly consider the convergence of the reversible jump approach where the number of directed edges in the model can vary.

5.5.3 Convergence assessment for data sets 5.1 – 5.6

Note that in Chapter 5 we have used the same data sets as in Chapter 4, and added two extra data sets generated from parameter sets with medium sized trait-to-trait

effects. This means that by setting $\lambda = 0$, we can compare the convergence of data sets 5.1, 5.3, 5.4 and 5.6 to data sets 4.1, 4.2, 4.3 and 4.4, respectively. However, as the number of directed edges can now vary, the chains use different initial models than those in Chapter 4, apart from chain 1, which is the same for all data sets. As the initial models have changed, we do not expect to see the same results as in Chapter 4.

To assess convergence for data sets 5.1 – 5.6 we ran five chains for each data set, given $\lambda = 0$. Each set of chains is summarized in a figure in a similar manner to the simulation study in Chapter 4. We note the initial model used, the number of models visited by each chain, and the joint posterior probability of the true model, as well as providing a graphical display of the marginal posterior probability of each directed edge and presenting the model with the greatest joint posterior probability. This allows us to get an impression of how the chains are moving. In addition, Table 5.3 (p. 190) gives the Gelman–Rubin convergence diagnostic for the dimension independent parameters for all six data sets (5.1 – 5.6).

Data set 5.1

Data set 5.1 is comprised of small QTL effects and small trait-to-trait effects, and is the same as data set 4.1 (see Figure 4.12, p. 113). Allowing the number of edges in the phenotype network structure to vary has significantly increased the number of models visited by the chain. In data set 4.1 each chain visited around 100 models (see Figure 4.12, p. 113), and by introducing add and remove steps most chains visited around 500 models (see Figure 5.11). The true model was visited in four of the five chains, with joint posterior probabilities similar to those for data set 4.1, $p(M_{TRUE}|D) < 0.01$ for all chains. Chain number 4 searched fewer models than the other chains; however, the initial model contained the maximum

number of edges possible and not a single directed edge in common with the true model. As a result, this chain did not visit the true model and got stuck at a local maxima as the minor changes able to be proposed to the phenotype network structure (using add, remove, reverse and relocate steps) were not large enough for a favourable model to be proposed. According to the Gelman–Rubin diagnostic in Table 5.3 (p. 190), the dimension independent parameters have all converged as the values are all close to 1.

Data set 5.2

Data 5.2 combines small QTL effects and medium trait-to-trait effects, and so there is no corresponding data set in Chapter 4. However, we note that as the trait-to-trait effects have increased from those used in data set 5.1, the number of models visited has decreased to about 30 models (see Figure 5.12, p. 183). The true model was not visited in any of the five chains, even though it was visited in four of the chains for data set 5.1 in Figure 5.11. Increasing the size of the trait-to-trait effects has decreased mixing. Once again chain number 4 searched fewer models than the other chains, due to the unfavourable initial model. The Gelman–Rubin diagnostic for the dimension independent parameters, given in Table 5.3 (p. 190), indicates that the log likelihood and the QTL effects appear to have converged, although longer chains are recommended. Unfortunately, the other dimension independent parameters have not converged.

Data set 5.3

Data set 5.3 is comprised of small QTL effects and large trait-to-trait effects, and is the same as data set 4.2 (see Figure 4.17, p. 121). Allowing the number of edges in the phenotype network structure to vary has significantly increased the number of models visited by the chain. In data set 4.2 each chain visited just 3 or

Figure 5.11: *A summary of the performance of each chain for data set 5.1 with $\lambda = 0$. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Figure 5.12: *A summary of the performance of each chain for data set 5.2 with $\lambda = 0$. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

4 models (see Figure 4.17, p. 121), and by introducing add and remove steps most chains visited up to 63 models (see Figure 5.13). In Chapter 4 the true model is visited by two chains; here we have proposed different initial models, and the true model was visited by just one chain, with a joint posterior probability of 0.4481. Again, chain number 4 searched fewer models than the other chains, due to the unfavourable initial model. The Gelman–Rubin diagnostic for the dimension independent parameters, given in Table 5.3 (p. 190), indicates that the QTL effects appear to have converged as values are close to 1, although longer chains are recommended. Unfortunately, the other dimension independent parameters have not converged.

Data set 5.4

Data set 5.4 is comprised of large QTL effects and small trait-to-trait effects, and is the same as data set 4.3 (see Figure 4.22, p. 128). Allowing the number of edges in the phenotype network structure to vary has increased the number of models visited by the chain. In data set 4.3 each chain visited about 5 models (see Figure 4.22, p. 128), and by introducing add and remove steps most chains visited around 50 models (see Figure 5.13). In Chapter 4 the true model was estimated by all chains; here we have proposed different initial models and the true model was the model with the greatest joint posterior probability for four chains, with $p(M_{TRUE}|D) > 0.5$ in each case. Again, chain number 4 searched fewer models than the other chains, due to the unfavourable initial model. These chains have converged according to the Gelman–Rubin diagnostic for the dimension independent parameters in Table 5.3 (p. 190).

Data set 5.5

Data 5.5 combines large QTL effects and medium trait-to-trait effects, and so there

Figure 5.13: *A summary of the performance of each chain for data set 5.3 with $\lambda = 0$. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Figure 5.14: *A summary of the performance of each chain for data set 5.4 with $\lambda = 0$. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

is no corresponding data set in Chapter 4. However, we note that as the size of the trait-to-trait effects have increased from those used in data set 5.4, the number of models visited has decreased to about $20 - 30$ models (see Figure 5.14). The true model was the model with the highest joint posterior probability in two of the five chains, but it was not visited in the remaining three chains. This is an improvement from data set 5.2, with small QTL effects and medium trait-to-trait effects, where the true model was not visited by any of the five chains. Increasing the size of the QTL effects has increased the signal of the true model. Increasing the size of the trait-to-trait effects (from data set 5.4) has decreased mixing, with the chains searching fewer models. Again, chain number 4 searched fewer models than the other chains, due to the unfavourable initial model. The Gelman–Rubin diagnostic for the dimension independent parameters, given in Table 5.3 (p. 190), indicates that none of the dimension independent parameters have converged.

Data set 5.6

Data set 5.6 is comprised of large QTL effects and large trait-to-trait effects, and is the same as data set 4.4 (see Figure 4.26, p. 134). Allowing the number of edges in the phenotype network structure to vary has generally increased the number of models visited by the chain. In data set 4.4 each chain visited just $1 - 2$ models (see Figure 4.22, p. 128), and by introducing add and remove steps chains visited up to 29 models (see Figure 5.13). Unfortunately, chains 2, 3 and 4 in Figure 5.14 (p. 186) do not visit many models, so mixing is poor. As a result, these chains have not converged according to the Gelman–Rubin diagnostic for the dimension independent parameters, and have the largest potential scale reduction factors observed for all data sets $5.1 - 5.6$; see Table 5.3 (p. 190).

In general, allowing the number of edges in the phenotype network to vary has

Figure 5.15: *A summary of the performance of each chain for data set 5.5 with $\lambda = 0$. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Figure 5.16: *A summary of the performance of each chain for data set 5.6 with $\lambda = 0$. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Table 5.3: *The Gelman–Rubin convergence diagnostic for data sets 5.1 – 5.6, corresponding to the chains summarized in Figures 5.11 – 5.16.*

data set	5.1	5.2	5.3	5.4	5.5	5.6
likelihood	1.02	1.10	42.3	1.04	26.3	114.00
β_{01}	1.12	7.31	5.09	1.03	13.3	19.60
β_{02}	1.01	1.68	13.7	1.05	10.6	14.30
β_{03}	1.04	5.03	3.92	1.02	7.53	12.10
β_{04}	1.01	3.26	2.04	1.03	9.45	29.90
β_{05}	1.17	7.18	3.13	1.03	7.18	27.00
ϕ_{11}	1.01	1.25	1.07	1.01	5.24	9.60
ϕ_{22}	1.00	1.04	1.72	1.01	2.46	9.19
ϕ_{33}	1.01	1.35	1.04	1.01	1.52	7.32
ϕ_{44}	1.00	1.01	1.04	1.01	2.18	1.12
ϕ_{55}	1.01	1.23	1.24	1.00	1.70	1.31
σ_1^2	1.01	8.14	6.38	1.00	6.12	16.8
σ_2^2	1.00	2.23	11.3	1.00	7.17	8.12
σ_3^2	1.02	4.70	4.53	1.00	1.52	15.7
σ_4^2	1.01	2.88	2.44	1.00	7.46	32.1
σ_5^2	1.02	7.49	3.20	1.00	8.28	39.3

improved mixing. That said, mixing is still slow and long chains are required. For data sets simulated with large effects (QTL-to-trait and trait-to-trait) the chains do not tend to visit many models, meaning that convergence and mixing are still a concern and an area for future research. We will discuss the consequences of assuming an incorrect genetic architecture in Section 5.7, and then we use this reversible jump algorithm on the published winter wheat data set in Section 5.8. Later, in Section 5.9, we propose the double step which will allow the chain to propose slightly larger steps in the hope that if a model has two or more badly placed edges, a double step will help it fix itself.

Next, we compare the estimated phenotype network structures from simulated data sets 5.1 – 5.6 to those obtained by the QTLnet algorithm by Neto et al. 2010.

5.6 Comparison to the QTLnet algorithm

In this section we briefly compare our RJMCMC II algorithm (including the add and remove steps) to the QTLnet algorithm from Neto et al. (2010). The QTLnet algorithm has been implemented in R using the 'qtlnet' package (Neto & Yandell 2014).

The RJMCMC II algorithm was implemented in R (Version 3.2.0), and generally took around 52 hours to complete 1,000,000 iterations on a 64 bit Intel(R) Core(TM) i5-4440 Quad 3.1GHz machine with 8GB RAM, running up to four chains in parallel. As the code was written from scratch, we note that the run time could be significantly improved by optimising the R code and converting key functions to C. Once the code has been optimised and key functions are converted to C, we expect the computing time required to drastically decrease. The contribution of this thesis is in the approach taken, not the efficiency of the R code; however, more details on the R code used are given in Appendix F. The QTLnet code has been optimised, with an updated version of the 'qtlnet' package released in February 2015. As a result, the QTLnet algorithm took just 1 minute to complete 1,000,000 iterations.

The models with the greatest joint posterior probability are given in Figure 5.17 for data sets 5.1 – 5.3, and in Figure 5.18 for data sets 5.4 – 5.6. For each data set we have run 5 chains, with the initial models given in Figures 5.11 – 5.16.

For data sets 5.1 – 5.3, with small QTL effects, the small joint posterior probabilities for both the RJMCMC II and QTLnet algorithm indicate that the chains are moving between many different phenotype network structures; however, the QTLnet algorithm is the first to identify the true phenotype network structure,

indicated by an asterisk ($*$), in data set 5.2 (with small QTL effects and medium-sized trait-to-trait effects). For data set 5.3, with small QTL effects and large trait-to-trait effects, the QTLnet algorithm also estimates the true phenotype network structure in more chains than the RJMCMC II algorithm, and appears to be moving (within the model space) better than the RJMCMC II approach proposed in this thesis. This is confirmed by Table 5.4, which gives the joint posterior probability of the true phenotype network structure, as estimated by the RJMCMC II and QTLnet algorithms, for data sets 5.1 – 5.6. The QTLnet algorithm visits the true model in every chain for data sets 5.1 – 5.3. The QTLnet algorithm should also estimate the genetic architecture; however, the QTL effects used in data sets 5.1 – 5.3 were too small, and the QTLnet algorithm did not identify any QTL that had an effect on the traits Y_1, Y_2, Y_3, Y_4 and Y_5.

For data sets 5.4 – 5.6 with large QTL effects, the larger joint posterior probabilities for both the RJMCMC II and QTLnet algorithm indicate that the chains are moving between few phenotype network structures; however, the RJMCMC II algorithm identifies the true phenotype network structure, indicated by an asterisk ($*$), in all three data sets. Here, for large QTL effects, the QTLnet algorithm did identify the approximate QTL locations, but it failed to estimate the true phenotype network structure. For large QTL effects, the RJMCMC II algorithm, appears to be moving (within the model space) better than the QTLnet algorithm, which appears to have problems with overfitting; consistently estimating a model with extra edges. Table 5.4 gives the joint posterior probability of the true phenotype network structure, as estimated by the RJMCMC II and QTLnet algorithms, and we note that the QTLnet algorithm did not estimate the true phenotype network structure when the QTL effects were large.
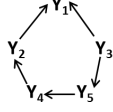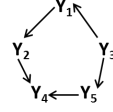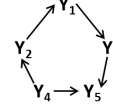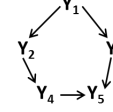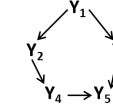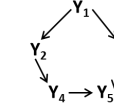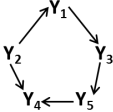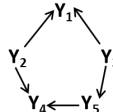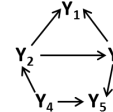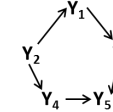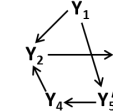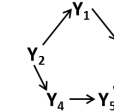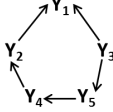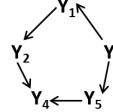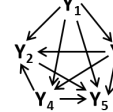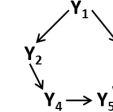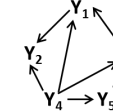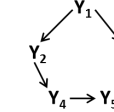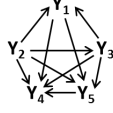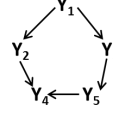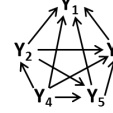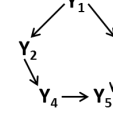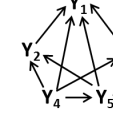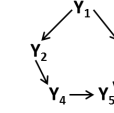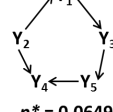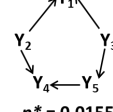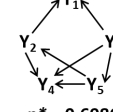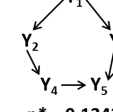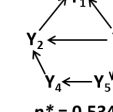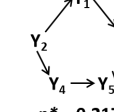
Figure 5.17: *The phenotype network structures with the greatest joint posterior probability (p^*) for each chain, for data sets 5.1 − 5.3. An asterisk (∗) in the upper right corner indicates that the true phenotype network structure has been estimated.*
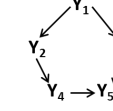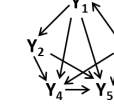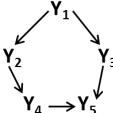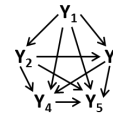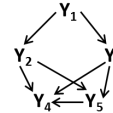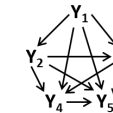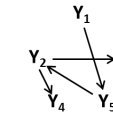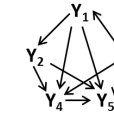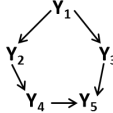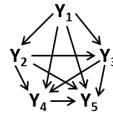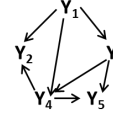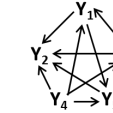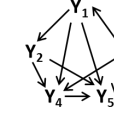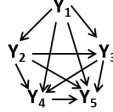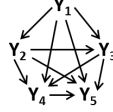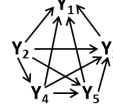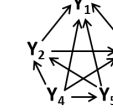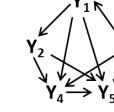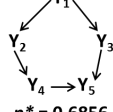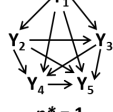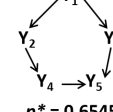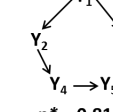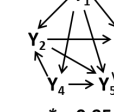
Figure 5.18: *The phenotype network structures with the greatest joint posterior probability (p^*) for each chain, for data sets 5.4 – 5.6. An asterisk (∗) in the upper right corner indicates that the true phenotype network structure has been estimated.*
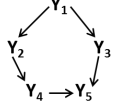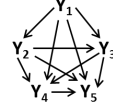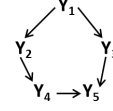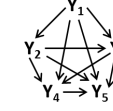
Table 5.4: *The joint posterior probability of the true phenotype network structure, as estimated by the RJMCMC II and QTLnet algorithms, for data sets 5.1 – 5.6.*

data set	5.1		5.2		5.3	
chain	RJMCMC II	QTLnet	RJMCMC II	QTLnet	RJMCMC II	QTLnet
1	0.0044	0.0211	0	0.1432	0.4481	0.2031
2	0.0045	0.0189	0	0.1276	0	0.1698
3	0.0079	0.0100	0	0.1543	0	0.2098
4	0	0.0189	0	0.1643	0	0.2131
5	0.0058	0.0155	0	0.1343	0	0.2175
data set	5.4		5.5		5.6	
chain	RJMCMC II	QTLnet	RJMCMC II	QTLnet	RJMCMC II	QTLnet
1	0.7239	0	0.6486	0	0.8789	0
2	0.7279	0	0	0	0	0
3	0.5643	0	0	0	0	0
4	0	0	0	0	0	0
5	0.6856	0	0.6545	0	0.8122	0

By comparing the RJMCMC II approach proposed in this thesis to the established QTLnet algorithm we have identified the following issues:

1. The QTLnet algorithm is much faster to run in R. We need to optimise our R code by exporting key functions to C.

2. The QTLnet algorithm is moving between models better than the RJMCMC II algorithm when the QTL effects are small. In order to improve mixing, the proposal distributions we use need tuning.

3. Both the RJMCMC II and the QTLnet algorithms have trouble moving between models when the effect sizes are large. However the QTLnet algorithm has a problem with overfitting. Therefore, the RJMCMC II approach proposed in this thesis is more likely to estimate the phenotype network structure when the QTL effects are large.

5.7   Incorrect genetic architecture

The extension to the RJMCMC approach proposed in this thesis assumes that the genetic architecture is known. This section provides two small simulated examples to demonstrate how the approach performs if the "known" genetic architecture is actually incorrect.

First, suppose that there is a minor mistake in the genetic architecture: one QTL is incorrect. Let the true causal network structure be that which is given in Figure 4.7 (p. 103); however, suppose that instead of QTL 4 affecting trait 4 ($Q_4 \rightarrow Y_4$) we have incorrectly specified that a new QTL, Q_6, is affecting trait 4 ($Q_6 \rightarrow Y_4$). This incorrect genetic architecture is illustrated in Figure 5.19. Using data set 5.4, generated using the parameter set given in Table 5.2 (p. 169), we simulate QTL genotypes for Q_6 for all $n = 500$ individuals and run five chains with the same initial models as those used in Figure 5.14 (p. 186). Assuming the correct genetic architecture, the true phenotype network structure was the model with the highest joint posterior probability in four of the five chains (see Figure 5.14, p. 186). Each chain searched between $14 - 51$ models, and mixing was poor. The chains generally visited more models when the genetic architecture assumed
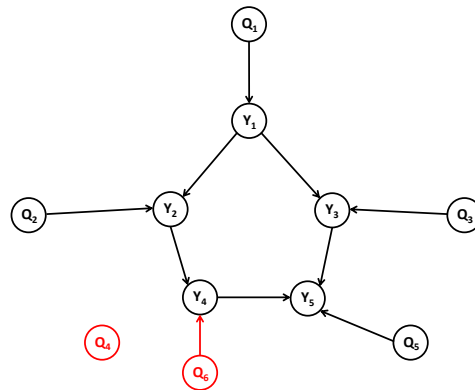
Figure 5.19: *Assuming an incorrect genetic architecture: Q_6 has a direct effect on trait Y_4 and the QTL Q_4 has been excluded from the causal network structure.*

is incorrect, visiting between 8 and 739 different phenotype network structures. The second chain visited the most models, and the true phenotype network structure had the highest joint posterior probability; $p(M_{TRUE}|D) = 0.1887$. However, this was the only chain to visit the true phenotype network structure. Other chains estimated phenotype network structures in which one or both the true causal relationships connected to Y_4 were reversed, and an extra trait-to-trait effect was estimated. We note that even causal relationships upstream of the incorrect QTL effect were affected.

Next, let us assume that there is a larger mistake in the genetic architecture. Suppose that QTL Q_4 has not been identified as having an effect on trait 4 ($Q_4 \nrightarrow Y_4$) and an additional QTL, Q_7, has been identified as having an effect on trait 1 ($Q_7 \rightarrow Y_1$). This incorrect genetic architecture is illustrated in Figure 5.20. Again, using data set 5.4, generated using the parameter values given in Table 5.2 (p. 169), we simulate QTL genotypes for Q_7 for all $n = 500$ individuals and run five chains with the same initial models as those used in Figure 5.14 (p. 186). The results are summarized in Figure 5.22.
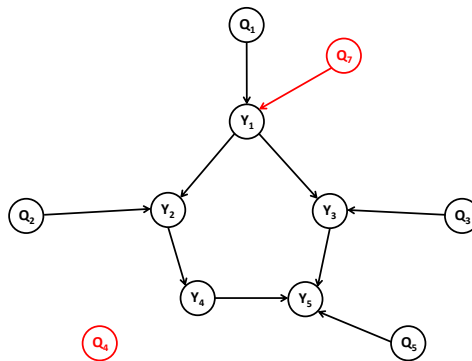
Figure 5.20: *Assuming an incorrect genetic architecture: Q_7, has been identified as having an effect on trait 1 and the QTL Q_4 has been excluded from the causal network structure.*

DATA SET 5.4 assuming an incorrect genetic architecture (I)					
Initial model:					
n(models)	**206**	**739**	**236**	**8**	**269**
$p(e_g\|D)$					
$p(M_k\|D) = p^*$	$p^* = 0.3906$	$p^* = 0.1887$	$p^* = 0.2465$	$p^* = 0.2393$	$p^* = 0.2603$
$p(M_{TRUE}\|D)$	**0**	**0.1887**	**0**	**0**	**0**

Figure 5.21: *A summary of the performance of each chain for data set 5.4 assuming an incorrect genetic architecture (I). The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Figure 5.22: *A summary of the performance of each chain for data set 5.4 assuming an incorrect genetic architecture (II). The number of models visited is denoted* n(models), *and the marginal posterior probability of the directed edge* e_g *is denoted* $p(e_g|D)$. *The greatest joint posterior probability is denoted* p^*, *such that model* M_k *has the greatest posterior probability when* $p(M_k|D) = p^*$. *The posterior probability of the true model is denoted* $p(M_{TRUE}|D)$.

The chains have generally visited more models with an incorrect genetic architecture, visiting between 8 and 370 different phenotype network structures. However, none of the chains visit the true phenotype network structure. The phenotype network structures estimated typically have additional edges and reverse the true direction of causality.

In conclusion, it is possible to for the RJMCMC II approach to estimate the true causal network structure if the assumed genetic architecture is not too different from the true causal network structure; however, in general, estimated phenotype network structures have additional edges and reverse the true direction of causality. Therefore we require that the genetic architecture be correct and known *a priori*.

5.8   Winter wheat data example

Dhungana et al. (2007) use a Structural Equation Model (SEM) to analyse genotype-by-environmental interactions in wheat. This example is included here to demonstrate that the extension to the reversible jump Markov chain Monte Carlo algorithm proposed in this thesis is capable of estimating the phenotype network structure of a real data set, not to do a comprehensive analysis on this data set.

A detailed description of the data is given in Campbell et al. (2003). A total of 2,268 observations from a population of recombinant inbred chromosomes lines (RICLs-3A) was used, created by crossing cv. Cheyenne (CNN) and the chromosome substitution line CNN (Whichita 3A). The set of quantitative traits used includes the grain yield (YLD) and the yield-component traits 1000 kernel weight (TKW), kernels per spike (KPS), and spikes per square metre ($SPSM$). Campbell et al. (2003) performed linkage analysis on these traits and identified six genetic markers linked to QTL: $Xtam055$, $Xbarc86$, $Xbarc67$, $Xksua6$, $Xbcd1555$, and

*Xbcd*361. Just as in Dhungana et al. (2007) we will use these markers as proxy QTL, so that we may consider the QTL genotypes to be known.

Dhungana et al. (2007) model genotype-by-environment interactions; they have climate data for daily minimum and maximum temperature ($^{\circ}C$), daily solar irradiance ($W\ m^{-2}$), and daily precipitation (mm) at each of the experimental sites at the time of each experiment. A total of 10 environmental covariates were considered by Dhungana et al. (2007); the mean daily temperature ($T1$, $T2$, $T3$), total solar radiance ($SR1$, $SR2$, $SR3$), and total precipitation ($P0$, $P1$, $P2$ and $P3$), where the suffix used denotes the period of the winter wheat growing season, as specified by the winter wheat development model developed by Streck et al. (2003):

1. seedling emergence to terminal spikelet initiation

2. terminal spikelet initiation to anthesis

3. anthesis to physiological maturity.

The total precipitation also includes the total precipitation during the three months before sowing ($P0$) so as to account for soil moisture before sowing. Using prior knowledge regarding wheat development, Dhungana et al. (2007) state that the environmental covariates from periods 1 and 2 were used to model *SPSM* and *KPS* genotype-by-environment interactions ($SPSM_{GEI}$ and KPS_{GEI}, respectively), those from periods 2 and 3 were used to model *TKW* genotype-by-environment interactions (TKW_{GEI}), and as SPSM, KPS and TKW all affect YLD, all ten environmental effects were used to model YLD genotype-by-environment interactions (YLD_{GEI}) (Dhungana et al. 2007, Donmez et al. 2001). Furthermore, the covariate $P0$ was also used for modeling all yield and yield component genotype-by-environment interactions. The causal network structure estimated by Dhungana et al. (2007) is given in Figure 5.23.
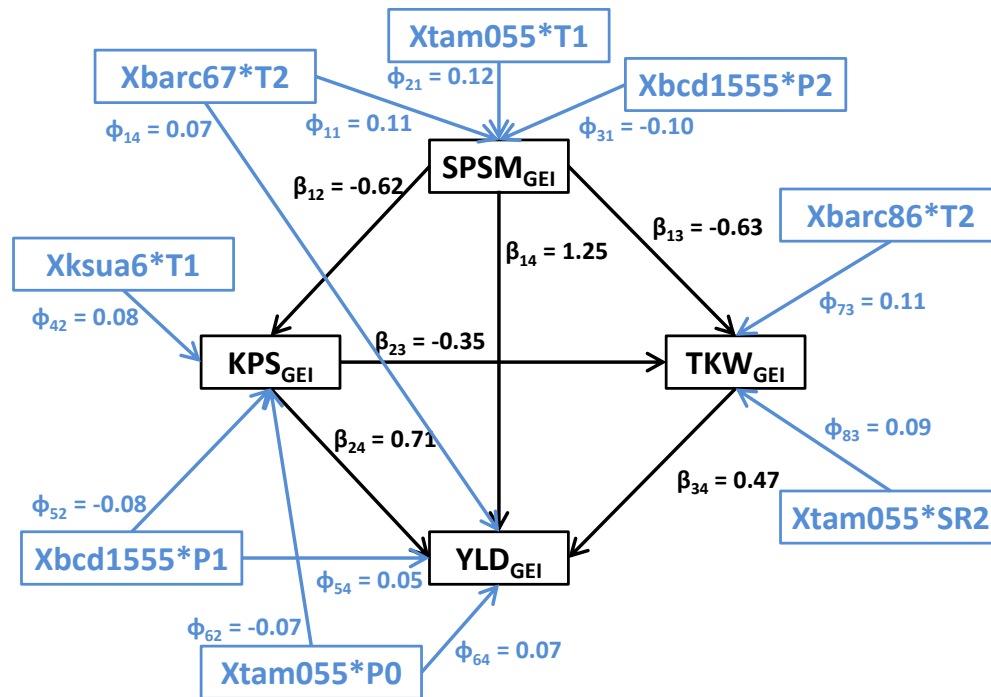
Figure 5.23: *The final causal network structure as estimated by Dhungana et al. (2007).*

Unfortunately we were unable to obtain the data required to define the three periods of the winter wheat growing season and so we may only model the QTL effects. We therefore expect to see some difference in the final parameter estimates. Dhungana et al. (2007) found that all genotype-by-environmental interactions involving the proxy QTL *Xbcd*361 resulted in non-significant path coefficients, so this proxy QTL has been excluded from the genetic architecture used in this analysis. The genetic architecture that we assume here is given in Figure 5.24, where we simplify the QTL and quantitative traits to the following notation:

Y_1: spikes per square metre (SPSM) Y_3: 1000 kernel weight (TKW)

Y_2: kernels per spike (KPS) Y_4: grain yield (YLD)

Q_1: *Xbarc*67 Q_4: *Xtam*055

Q_2: *Xbcd*1555 Q_5: *Xbarc*86.

Q_3: *Xksua*6

Following Dhungana et al. (2007), before inferring the phenotype network structure we fitted a randomized block design to each of the quantitative traits, and the residuals were then used in our analysis. This is to remove any variability due to the variety of wheat, and the replications done over different locations throughout a number of years.

5.8.1 Wheat data without a complexity prior

We set $\lambda = 0$, such that the prior on the number of edges in the phenotype network structure will not favour models with fewer directed edges. We ran very long
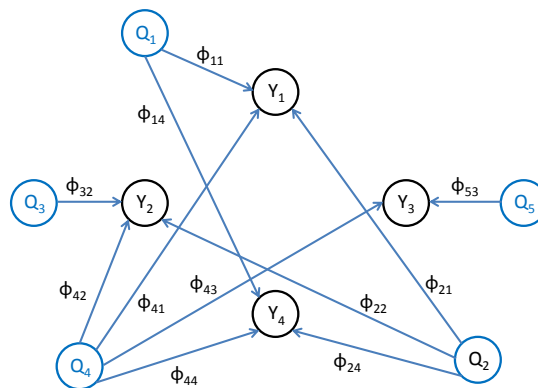
Figure 5.24: *The genetic architecture based on that used by Dhungana et al. (2007) who modeled the genotype-by-environment interactions, instead of the direct QTL effects we use here.*

chains, with $N = 2,000,000$ iterations. Applying a burn-in period of $100,000$ iterations and retaining every 10th iteration left $190,000$ states stored.

The marginal posterior probability of each directed edge, and the model averaged parameter estimates are calculated as outlined in Section 4.6. The marginal posterior probabilities for each directed edge are displayed in Figures 5.25 and 5.26, where the six directed edges with the greatest posterior probabilities are: β_{12}, β_{13}, β_{14}, β_{23}, β_{24} and β_{34}. Edges β_{12}, β_{13}, β_{14}, β_{24} and β_{34} have a posterior probability of 1, meaning that they were included in all of the causal network structures visited by the chain. Despite the large number of iterations, only 2 models were visited by the chain (each with 6 directed edges) indicating mixing was poor. The two models visited by the chain are given in Figure 5.27 along with the joint posterior probability of each. The true model, as published by Dhungana et al. (2007), is the model we have estimated with the greatest joint posterior probability: Model 1 with $p(M_{TRUE}|D) = 0.8133$. The other model searched has reversed one of the edges: the edge between traits Y_2 and Y_3 (KPS and TKW).

The parameter estimates for the true model only are given in Tables 5.5 and 5.6; Dhungana et al. (2007) did not present their estimates for the trait intercepts and residual variances, so we present them in a separate table. As expected, our estimates are slightly different from those obtained by Dhungana et al. (2007), with the 95% credible intervals not containing the true parameter values published by Dhungana et al. (2007). We note that many estimates are close, however, so we calculate the direct and indirect effects of each trait or QTL on the primary trait of interest, grain yield (YLD, or Y_4), to compare our results to those obtained by Dhungana et al. (2007). The causal network structure complete with the estimated QTL and trait-to-trait effects is given in Figure 5.28, and the corresponding direct, indirect and total effects on yield are given in Table 5.7 (p. 209).
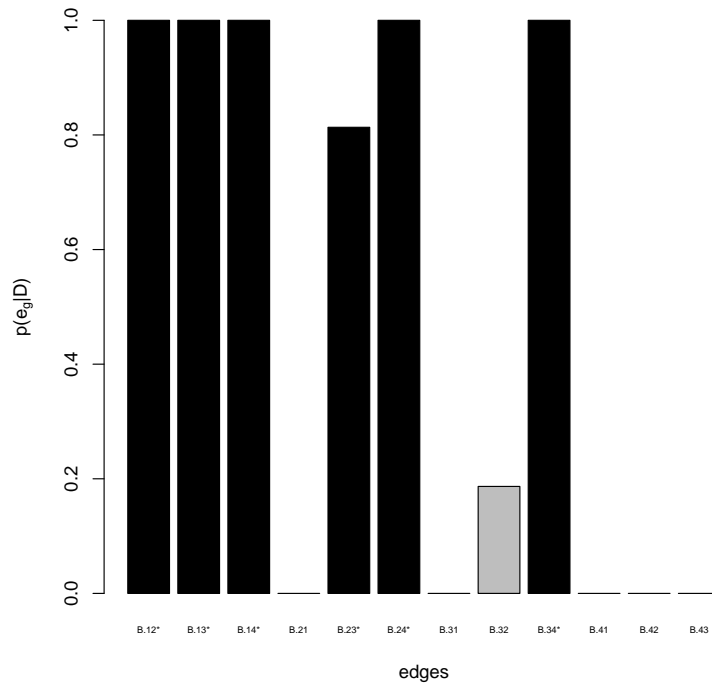
Figure 5.25: *The marginal posterior probability of each directed edge e_g, denoted $p(e_g|D)$ for the winter wheat data set with $\lambda = 0$. The terms "B.ij" denote the trait-to-trait effect β_{ij}, i.e. the effect of trait i on trait j. The edges comprising the true model are indicated by solid black bars and an "*" next to the label of the directed edge.*
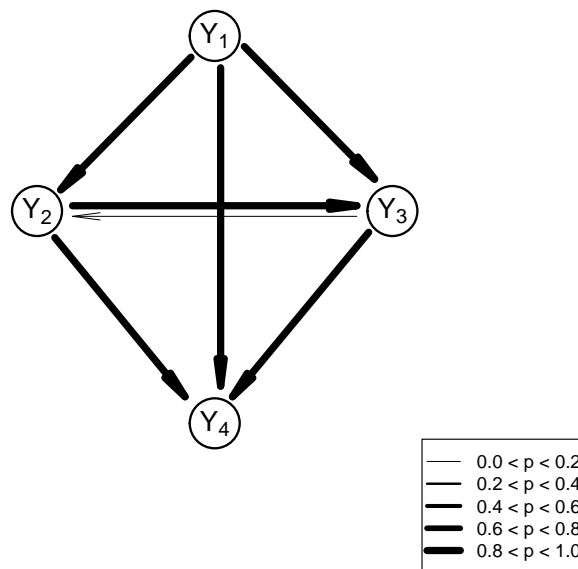
Figure 5.26: *For the winter wheat data set with $\lambda = 0$: the marginal posterior probability of each directed edge displayed in the context of a phenotype network structure. The thicker the directed edge, the larger the marginal posterior probability.*
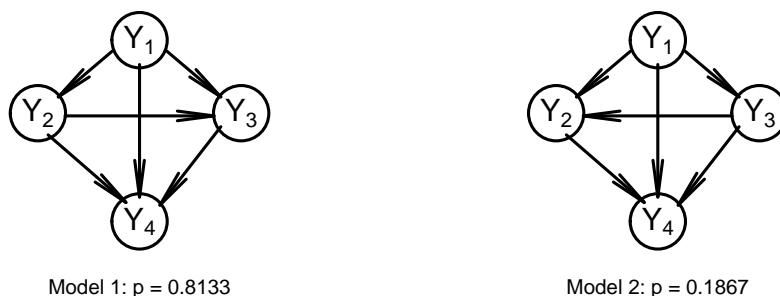
Figure 5.27: *The two estimated phenotype network structures, with the joint posterior probabilities denoted p. The true model is Model 1.*

Table 5.5: *Parameter estimates for the true model. Note that the true values for the QTL effects are taken from Dhungana et al. (2007) so they actually refer to the genotype-by-environment interaction.*

Parameter	True Value	Estimated Value	95% CI
β_{12}	-0.62	-0.542	(-0.579, -0.507)
β_{13}	-0.63	-0.236	(-0.286, -0.186)
β_{14}	1.25	1.129	(1.116, 1.142)
β_{23}	-0.35	-0.170	(-0.218, -0.123)
β_{24}	0.71	0.609	(0.597, 0.623)
β_{34}	0.47	0.450	(0.440, 0.462)
ϕ_{11}	0.11	0.116	(0.009, 0.222)
ϕ_{14}	0.07	0.041	(0.014, 0.067)
ϕ_{21}	-0.10	0.169	(0.069, 0.272)
ϕ_{22}	-0.08	-0.093	(-0.171, -0.011)
ϕ_{24}	0.05	-0.003	(-0.030, 0.024)
ϕ_{32}	0.80	0.115	(0.038, 0.194)
ϕ_{41}	0.12	0.088	(0.001, 0.175)
ϕ_{42}	-0.07	0.207	(0.138, 0.277)
ϕ_{43}	0.09	-0.099	(-0.195, -0.027)
ϕ_{44}	0.07	-0.022	(-0.043, 0.000)
ϕ_{53}	0.11	-0.025	(-0.100, 0.052)

Table 5.6: *Parameter estimates for the trait intercepts and residual variances for the true model.*

Parameter	Estimated Value	95% CI
β_{01}	-0.171	(-0.249, -0.094)
β_{02}	-0.141	(-0.208, -0.078)
β_{03}	0.073	(0.000, 0.155)
β_{04}	-0.003	(-0.023, 0.017)
σ_1^2	0.983	(0.927, 1.043)
σ_2^2	0.695	(0.656, 0.738)
σ_3^2	0.953	(0.898, 1.011)
σ_4^2	0.065	(0.061, 0.069)

The calculation of the indirect and total effects were outlined in Section 2.1, but we also give the following example. There exist four different paths from *Xbarc*67 (Q_1) to *YLD* (Y_4), see Figure 5.28. For each path, we calculate the product of the path coefficients, and the indirect effect of Q_1 on Y_4 is the sum of these;

$$
\begin{aligned}
& \phi_{11}\beta_{14} + \phi_{11}\beta_{13}\beta_{34} + \phi_{11}\beta_{12}\beta_{24} + \phi_{11}\beta_{12}\beta_{23}\beta_{34} \\
= \; & (0.116)(1.129) + (0.116)(-0.236)(0.45) + (0.116)(-0.542)(0.609) + \\
& (0.116)(-0.542)(-0.170)(0.45) \\
= \; & 0.1310 + (-0.0123) + (-0.0383) + 0.0048 \\
= \; & 0.0852
\end{aligned}
$$

The total effect is the sum of the direct and indirect effects. The direct effect of Q_1 on Y_4 is 0.041, and so the total effect of Q_1 on Y_4 is 0.126 as given in Table 5.7.

Let us briefly interpret the results by first focusing on the effects of the quantitative traits on yield.

- Spikes per square metre (SPSM, or Y_1) has the strongest positive total effect on yield, even including the negative indirect effect. This implies that, in

general, the more spikes per square metre, the higher the grain yield. Both the direct and indirect effects of SPSM on YLD are similar to what was published by Dhungana et al. (2007), with a large positive direct effect of SPSM on YLD estimated, and a negative indirect effect.

- The kernels per spike (KPS, or Y_2) has a large positive total effect on grain yield, but the direct and indirect effects are smaller than those estimated for the spikes per square metre. These estimates are still consistent with Dhungana et al. (2007), with a medium sized positive total effect on yield, and a small negative indirect effect. With an overall positive direct effect of KPS on YLD, this implies that, in general, the more kernels per spike, the higher the grain yield.

- Thousand kernel weight (TKW, or Y_3) only has a direct effect on grain yield. This effect is estimated as being 0.450 using our approach, and as 0.47 by Dhungana et al. (2007) using environmental data. With a positive total effect of TKW on YLD, this implies that, in general, the higher the thousand kernel weight, the higher the grain yield.

Interestingly, even without including the environmental data the relative scale of the total effects are the same as published by Dhungana et al. (2007); the spikes per square metre have the strongest positive effect on the grain yield, followed by the kernels per spike, and then the thousand kernel weight.

The environmental effects have had to be excluded from our analysis, so we cannot compare the QTL effects estimated here to the genotype-by-environment interactions used in Dhungana et al. (2007), but we will comment on the comparative size of the total effects. Our study included the five QTL that Dhungana et al.

Table 5.7: *Direct and indirect effects on grain yield for the true model (parameter estimates given in Table 5.5 and Figure 5.28). The effects published in Dhungana et al. (2007) are given in red.*

	direct effect	indirect effect	total effects
Y_1	1.129	-0.395	0.734
(SPSM)	1.25	-0.64	0.61
Y_2	0.609	-0.077	0.532
(KPS)	0.71	-0.16	0.55
Y_3	0.450	0	0.450
(TKW)	0.47	0	0.47
Q_1	0.041	0.085	0.126
$Xbarc67 \times T2$	0.07	0.07	0.14
Q_2	-0.003	0.075	0.072
$Xbcd1555 \times P1$	0.05	-0.05	0
$Xbcd1555 \times P2$	0	-0.07	-0.07
Q_3	0	0.061	0.061
$Xksua6 \times T1$	0	0.05	0.05
Q_4	-0.022	0.130	0.108
$Xtam055 \times T1$	0	0.07	0.07
$Xtam055 \times SR2$	0	0.04	0.04
$Xtam055 \times P0$	0.07	-0.04	0.03
Q_5	0	-0.011	-0.011
$Xbarc86 \times T2$	0	0.05	0.05

Figure 5.28: *The estimated effects for the model with the highest joint posterior probability (the true model).*

(2007) identified as being significant, because we assume that the genetic architecture is known. The estimates of the QTL direct and indirect effects for the true model are listed in Table 5.7, and we have included the genotype-by-environment interactions identified by Dhungana et al. (2007) with the corresponding QTL included in the interaction term.

- $Xbarc67$ (Q_1) has small positive direct and indirect effects on grain yield, very close to those estimated by Dhungana et al. (2007) for the $Xbarc67 \times T2$ interaction. As a result, $Xbarc67$ is positively related to grain yield with a total effect of 0.126 compared to 0.14 for $Xbarc67 \times T2$ as estimated by Dhungana et al. (2007).

- $Xbcd1555$ (Q_2) has very small negative direct effect and a very small positive indirect effect on grain yield. However, Dhungana et al. (2007) identified a negative indirect effect of the interaction terms $Xbcd155 \times P1$ and $Xbcd1555 \times P2$. The size of the indirect effects are similar, but the sign ($+/-$) differs because we estimate the direct QTL effect, not the genotype-by-environment interaction.

- $Xksua6$ (Q_3) has only an indirect effect on the grain yield; a small positive total effect of 0.061 which is similar to the value of 0.05 estimated by Dhungana et al. (2007) for the term $Xksua6 \times T1$.

- $Xtam055$ (Q_4) has a very small negative direct effect and a larger (but still small) positive indirect effect on grain yield, resulting in a positive total effect of 0.108, not too dissimilar in sign and size from the genotype-by-environment interactions published in Dhungana et al. (2007).

- $Xbarc86$ (Q_5) only has an indirect effect on grain yield, -0.011, so it does

differ in sign from the $Xbarc86 \times T2$ interaction used by Dhungana et al. (2007); however, both are close to zero. $Xbarc86$ is negatively associated with grain yield.

Based on our estimates, QTL $1-4$ are positively related to grain yield. A RIL population has two possible genotypes at any given marker, so suppose 0 denotes a genotype corresponding to the parental type cv. Cheyenne (CNN) and 1 corresponds to the parental type Whichita 3A. Then the parental type Whichita 3A at QTL $1-4$ is more favourably associated with higher grain yield. On the other hand, the parental type Whichita 3A at QTL 5 is less favourable for the grain yield.

5.8.2 Wheat data with a complexity prior

Including a strict complexity penalty where $\lambda = 10$ has not changed the results from Section 5.8.1 where $\lambda = 0$. The true model is still the model with the greatest joint posterior probability; however, this has decreased to $p(M_{TRUE}|D) = 0.6653$, with the chain now visiting 5 models. Mixing is still poor.

Figure 5.29 illustrates all of the models visited, both with no complexity penalty ($\lambda = 0$) and a large complexity penalty ($\lambda = 10$). We note that the chain is slow moving; however, in each case the true phenotype network structure has been identified with a large joint posterior probability. With a larger λ the chain prefers models with fewer edges, here the chain has searched one model with five edges (instead of six).
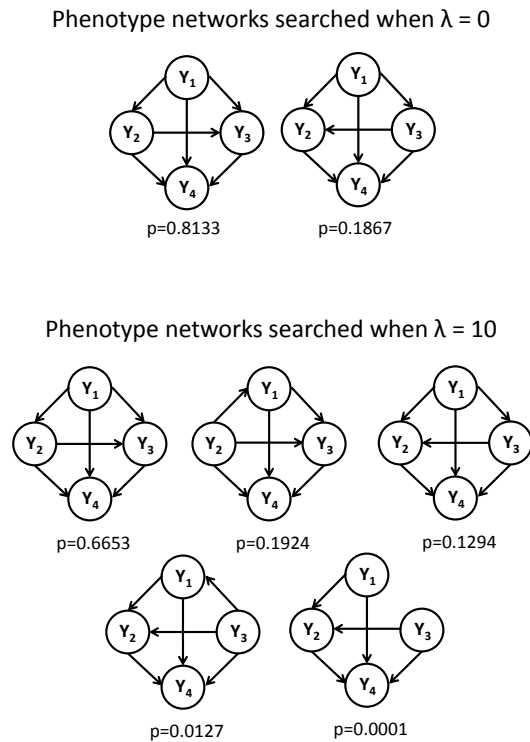
Figure 5.29: *Phenotype networks visited for the winter wheat data set, with* $\lambda = 0$ *and* $\lambda = 10$. *The joint posterior probability, denoted p, is given for each phenotype network structure.*

5.9   The double step

In addition to the standard add (or birth), remove (or death) and reverse steps in RJMCMC methods, Lunn et al. (2009) propose the 'replace' step; deleting a number of parameters, say δ, and then replacing them with the same number of new parameters — distinct from those that were deleted. This type of step allows the sampler to make larger steps between models, allowing more of the parameter space to be searched (Lunn et al. 2009). Here in Chapter 5 we have already proposed a relocate step, which removes one directed edge and relocates

it to a new (distinct) position, performing the replace step proposed by Lunn et al. (2009) with $\delta = 1$.

Now, we extend on the idea of the replace step proposed by Lunn et al. (2009), and create our own trans-dimensional version which we refer to as the 'double' step. Our double step will combine two of the moves used to alter the phenotype network structure. We consider combinations of the following move types: add, remove, and reverse. Six different types of double steps are then possible, listed in Section 5.9.1. We exclude the relocation step as it can already be considered a type of double step — a simultaneous add and remove step. Following Lunn et al. (2009), we include the constraint that a newly added edge must be placed in a new distinct position, i.e. if we propose to add one directed edge and remove another we cannot add an edge between the same two nodes from which we remove an edge. This increases mixing by reducing the number of redundant moves attempted.

The acceptance probability is detailed in Section 5.9.2 and an example of a double step is given in Section 5.9.3. In Section 5.9.4 we use simulated data from Section 5.5 to implement the double step, and briefly compare the results to those previously obtained.

5.9.1   Different types of double step:

The double step, denoted by $x^{D,d}$, selects two steps that alter the phenotype network structure and combines them into one step, so the modifications to the phenotype network structure will be carried out simultaneously. For this reason, the double step, x^D, has type d, and is denoted $x^{D,d}$. The six different double steps are:

Table 5.8: *The probability of each move type, as well as the number of candidate models for move type x, $|Ne(M,x)|$. There exist N_T traits, such that there are $N_T(N_T-1)$ unique directed edges possible; subsets of these edges form the current and candidate phenotype network structures. Model M has E_M edges, with the maximum number of directed edges denoted $max(E) = \frac{1}{2}N_T(N_T-1)$.*

Move type	x	Number of edges in model M		
		0	$1 \leq E_k \leq (max(E)-1)$	$max(E)$
Update	(x^U)	0.5	0.5	0.5
Add	(x^A)	0.4	0.1	0.0
Remove	(x^R)	0.0	0.1	0.2
Reverse	(x^S)	0.0	0.1	0.2
Double	(x^D)	0.1	0.2	0.1

Move Type	(x)	Size of the neighbourhood of model M $|Ne(M,x)|$
Update	(x^U)	1
Add	(x^A)	$\sum_{g=1}^{N_T(N_T-1)} I_a(A_g[M])I_\in(g|\mathbf{S})$
Remove	(x^R)	E_M
Reverse	(x^S)	$\sum_{g=1}^{E_M} I_a(S_g[M])$
Double	(x^D)	$|Ne(M,x^{D,d})|$

$d = 1$: add two edges $(x^{D,1})$

$d = 2$: add an edge and remove an existing edge $(x^{D,2})$;

      this is the same as the relocate step detailed in Section 4.4

$d = 3$: add an edge and reverse an existing edge $(x^{D,3})$

$d = 4$: remove two existing edges $(x^{D,4})$

$d = 5$: remove one edge and reverse another $(x^{D,5})$

$d = 6$: reverse two existing edges $(x^{D,6})$.

The probabilities for the selection of a move type, given model M, are denoted $q(x|\Omega,\mathbf{S})$ and are given in Table 5.8. Note that the size of the neighbourhood for

a double step is denoted $|Ne(M, x^{D,d})|$ in Table 5.8 as it differs depending on the type of double step d selected. In order to specify $|Ne(M, x^{D,d})|$ for $d = 1, \cdots, 6$, we use the indicator functions given in Equation 4.3 (p. 72) in addition to the newly defined operators D_{AA,g^*}, $D_{AS,g,f}$, $D_{RS,g,f}$, and $D_{RR,g^a st}$ such that:

- $D_{AA,g^*}[M] \in \mathbb{M}$ is a new model which adds the pair of directed egdes g^* (to **S**) with an associated generation of new parameter values

- $D_{AS,g,f}[M] \in \mathbb{M}$ is a new model which adds the g and reverses the edge f (in **S**) with an associated generation of new parameter values

- $D_{RS,g,f}[M] \in \mathbb{M}$ is a new model which removes the edge g and reverses the edge f (in **S**) with an associated generation of new parameter values

- $D_{SS,g^*}[M] \in \mathbb{M}$ is a new model which reverses the pair of directed egdes g^* (in **S**) with an associated generation of new parameter values

For N_T traits, the model M has E_M edges, and let $max(E)$ denote the maximum number of directed edges in a phenotype network structure, $max(E) = \frac{1}{2}N_T(N_T - 1)$.

1. $|Ne(M, x^{D,1})| = \displaystyle\sum_{g=1}^{2(max(E)-E_M)C_2} I_a(D_{AA,g^*}[M])$

   where the number of different combinations of two edges that can be added with $2(max(E) - E_M)$ possible directed edges is equal to $^{2(max(E)-E_M)}C_2 = \frac{(2(max(E)-E_M))!}{2!(2(max(E)-E_M)-2)!}$.

2. $|Ne(M, x^{D,2})| = \displaystyle\sum_{g=1}^{E_M} \sum_{f=1}^{N_T(N_T-1)} I_a(C_{g,f}[M])I_R(f,g|\mathbf{S})I_\in(f|\mathbf{S})$

   where this is the same as the neighbourhood for a relocate step described in Section 4.4.

3. $|Ne(M, x^{D,3})| = \sum_{g=1}^{2(max(E)-E_M)} \sum_{f=1}^{E_M} I_a(D_{AS,g,f}[M])$

   where one of $2(max(E) - E_M)$ directed edges can be added to model M and one of E_M edges can be reversed.

4. $|Ne(M, x^{D,4})| = {}^{E_M}C_2$

   where the number of pairs of edges that can be deleted with E_M edges in the current model is equal to ${}^{E_M}C_2 = \frac{E_M!}{2!(E_M-2)!}$.

5. $|Ne(M, x^{D,5})| = \sum_{g=1}^{E_M} \sum_{f=1;f\neq g}^{E_M} I_a(D_{RS,g,f}[M])$

   where one of E_M edges can be removed and given the removal of the gth edge, one of $E_M - 1$ edges can be reversed (i.e. $f \neq g$).

6. $|Ne(M, x^{D,6})| = \sum_{g=1}^{{}^{E_M}C_2} I_a(D_{SS,g^*}[M])$

   where the number of pairs of edges that can be reversed with E_M possible edges is equal to ${}^{E_M}C_2 = \frac{E_M!}{2!(E_M-2)!}$.

5.9.2   Acceptance probability for a double step

The proposal of model M' given model M can be split into three steps:

- the selection of the move type,

- the selection of the candidate phenotype network structure (given the move type),

- the generation of new candidate parameters (given the move type, and the candidate phenotype network structure).

Given in Equation 5.10 (p. 152), the proposal ratio for a move from model M to M' is therefore:

$$\frac{q(\mathbf{u}'_M,\mathbf{S},x'|\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'},\mathbf{S}',x|\Omega,\mathbf{S})} = \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \frac{|Ne(M,x)|}{|Ne(M',x')|} \frac{q(x'|\Omega',\mathbf{S}')}{q(x|\Omega,\mathbf{S})}.$$

The double step has six substeps and so the probability of selecting a double step of type d, given model M, is denoted $q(x=x^{D,d}|\Omega,\mathbf{S})$ and has two components;

$$q(x=x^{D,d}|\Omega,\mathbf{S}) = q(x=x^D|\Omega,\mathbf{S})q(d|x^D,\Omega,\mathbf{S}). \tag{5.18}$$

The probability of selecting a double step, given model M, is denoted $q(x=x^D|\Omega,\mathbf{S})$ and is given in Table 5.8 (p. 214); the probability of selecting the substep d, given a double step, is denoted $q(d|x^D,\Omega,\mathbf{S})$. Not all of the different types of replace step can be carried out on all models, so

$$q(d|x^D,\Omega,\mathbf{S}) = \frac{1}{\sum_d I_D(M,d)}$$

where $I_D(M,d)$ is an indicator function:

$$I_D(M,d) = \begin{cases} 1, & \text{if } |Ne(M,x=x^{D,d})| \geq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5.19}$$

It follows that the proposal ratio for a move from model M to M', given a double step, is therefore:

$$\begin{aligned} &\frac{q(\mathbf{u}'_M,\mathbf{S},x'|\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'},\mathbf{S}',x|\Omega,\mathbf{S})} \\ &= \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \frac{|Ne(M,x=x^{D,d})|}{|Ne(M',x'=x^{D,d'})|} \frac{q(x'=x^D|\Omega',\mathbf{S}')q(d'|x^D,\Omega',\mathbf{S}')}{q(x=x^D|\Omega,\mathbf{S})q(d|x^D,\Omega,\mathbf{S})}, \end{aligned} \tag{5.20}$$

where $|Ne(M, x = x^{D,d})|$ is the number of candidate (acyclic) models comprising the neighbourhood of the current model, given a double step of type d (denoted $x^{D,d}$). The reverse of move type $x^{D,d}$ is denoted $x^{D,d'}$, where the substep d' is the reverse of d:

$$x^{D,d'} = \begin{cases} x^{D,1}, & \text{if } x = x^{D,4} \\ x^{D,2}, & \text{if } x = x^{D,2} \\ x^{D,3}, & \text{if } x = x^{D,5} \\ x^{D,4}, & \text{if } x = x^{D,1} \\ x^{D,5}, & \text{if } x = x^{D,3} \\ x^{D,6}, & \text{if } x = x^{D,6}. \end{cases} \tag{5.21}$$

The product of the candidate parameter proposal densities for model M' is denoted by $q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})$ in Equation 5.20 where the candidate parameter proposal distributions for a move from model M to model M' are given in Equations 4.12 – 4.14 (p. 79). These are the same for all move types that alter the structure of the phenotype network, so will not be re-stated here. The number of edges in model M' will differ depending on the type of double step proposed. Suppose the current model has E_M edges, where $2 \leq E_M \leq (max(E) - 2)$. The number of edges in the candidate model, $E_{M'}$, is:

$$E_{M'} = E_M + \eta \tag{5.22}$$

where

$$\eta = \begin{cases} -2, & \text{if } x = x^{D,4} \\ -1, & \text{if } x = x^{D,5} \\ 0, & \text{if } x = x^{D,2}, x^{D,6} \\ 1, & \text{if } x = x^{D,3} \\ 2, & \text{if } x = x^{D,1}. \end{cases}$$

The product of the current parameter generating densities for model M is denoted by $q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')$ in Equation 5.20 where the current parameter generating distributions for a move from model M' to M are given in Equations 4.15 – 4.17 (p. 80). The number of edges in model M is:

$$E_M = E_{M'} + \eta' \tag{5.23}$$

where

$$\eta' = \begin{cases} -2, & \text{if } x' = x^{D,4} \\ -1, & \text{if } x' = x^{D,5} \\ 0, & \text{if } x' = x^{D,2}, x^{D,6} \\ 1, & \text{if } x' = x^{D,3} \\ 2, & \text{if } x' = x^{D,1}. \end{cases}$$

By combining two steps that alter the structure of the phenotype network we are likely to have an overlap in the parameters which need to be updated; if this occurs the parameters should only be updated once.

The parameter prior distributions are defined in Equation 4.7 (p. 73), and the phenotype network prior is as defined in Equation 5.6 (p. 150), such that the ratio of priors is:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \left[\frac{\prod_{t=1}^{N_T} \left\{ p(\beta'_{0t}|\mathbf{S}') p(\sigma_t'^2|\mathbf{S}') \right\}}{\prod_{t=1}^{N_T} \left\{ p(\beta_{0t}|\mathbf{S}) p(\sigma_t^2|\mathbf{S}) \right\}} \right] \left[\frac{\prod_{t=1}^{N_T} \prod_{s \in \upsilon_Y(t)} p(\beta'_{st}|\mathbf{S}')}{\prod_{t=1}^{N_T} \prod_{s \in \upsilon_Y(t)} p(\beta_{st}|\mathbf{S})} \right] \times$$
$$\left[\frac{\prod_{t=1}^{N_T} \prod_{\ell \in \upsilon_Q(t)} p(\phi'_{\ell t}|\mathbf{S}')}{\prod_{t=1}^{N_T} \prod_{\ell \in \upsilon_Q(t)} p(\phi_{\ell t}|\mathbf{S})} \right] \frac{p(\mathbf{S}')}{p(\mathbf{S})}. \tag{5.24}$$

Following the general form of the acceptance probability given in Equation 5.2 (p. 145), and substituting in the proposal and prior ratios in Equations 5.20

and 5.24, the acceptance probability for a double step is:

$$
\begin{aligned}
\alpha_{double} &= min(1, r_{double}) \quad \text{where} \\
r_{double} &= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \frac{q(\mathbf{u}'_M|\mathbf{S},x'=x^{D,d'},\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x=x^{D,d},\Omega,\mathbf{S})} \times \\
&\quad \frac{q(\mathbf{S}|x'=x^{D,d'},\Omega',\mathbf{S}')}{q(\mathbf{S}'|x=x^{D,d},\Omega,\mathbf{S})} \frac{q(x'=x^{D,d'}|\Omega',\mathbf{S}')}{q(x=x^{D,d'}|\Omega,\mathbf{S})} \left| \frac{\partial(\Omega',\mathbf{u}'_M)}{\partial(\Omega,\mathbf{u}_{M'})} \right| \\
&= \frac{p(\mathbf{Y}|\Omega',\mathbf{S}',\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega,\mathbf{S},\mathbf{Q},\Lambda,\mathbf{G},\mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \left[\frac{\prod_{t=1}^{N_T}\left\{p(\beta'_{0t}|\mathbf{S}')p(\sigma'^2_t|\mathbf{S}')\right\}}{\prod_{t=1}^{N_T}\left\{p(\beta_{0t}|\mathbf{S})p(\sigma^2_t|\mathbf{S})\right\}} \right] \times \\
&\quad \left[\frac{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}p(\beta'_{st}|\mathbf{S}')}{\prod_{t=1}^{N_T}\prod_{s\in\upsilon_Y(t)}p(\beta_{st}|\mathbf{S})} \right] \left[\frac{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}p(\phi'_{\ell t}|\mathbf{S}')}{\prod_{t=1}^{N_T}\prod_{\ell\in\upsilon_Q(t)}p(\phi_{\ell t}|\mathbf{S})} \right] \frac{p(\mathbf{S}')}{p(\mathbf{S})} \times \\
&\quad \frac{q(\mathbf{u}'_M|\mathbf{S},x',\Omega',\mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}',x,\Omega,\mathbf{S})} \frac{|Ne(M,x=x^{D,d})|}{|Ne(M',x'=x^{D,d'})|} \frac{q(x'=x^D|\Omega,\mathbf{S})q(d'|x^D,\Omega,\mathbf{S})}{q(x=x^D|\Omega,\mathbf{S})q(d|x^D,\Omega,\mathbf{S})}
\end{aligned}
$$
$$(5.25)$$

The general expression of the likelihood is given in Equation 5.3 (p. 147), and the Jacobian is 1 (see Appendix C.5). If we assume a uniform prior on the state of a connection between any two trait nodes, then the ratio of phenotype network priors simplifies to $\frac{p(\mathbf{S}')}{p(\mathbf{S})} = \exp(-\lambda(E_{M'}-E_M))$. However, this implies that an edge is more likely to be included in the phenotype network structure (in one of two directions) than not.

5.9.3   Example of a double step

Suppose that we want to carry out a double step to move from model M to M'; the example causal network structures are given in Figure 5.30. Let us select a double step with $d = 3$, where we propose the reversal of an existing edge and the addition of a new edge.

In this example we have a set of parameters for the current model, Ω, and a

set of parameters for the candidate model, Ω':

$$\Omega = \{E_M, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{23}, \beta_{24}, \phi_{11}, \phi_{22}, \phi_{33}, \phi_{44}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\}$$
$$\Omega' = \{E_{M'}, \beta'_{01}, \beta'_{02}, \beta'_{03}, \beta'_{04}, \beta'_{12}, \beta'_{13}, \beta'_{14}, \beta'_{24}, \beta'_{32}, \beta'_{34}, \phi'_{11}, \phi'_{22}, \phi'_{33}, \phi'_{44}, \sigma_1'^2, \sigma_2'^2, \sigma_3'^2, \sigma_4'^2\}$$

Only those parameters affected by the addition of β'_{34} and the reversal of β_{23} are updated, such that:

$$\{\beta'_{01}, \beta'_{12}, \beta'_{13}, \beta'_{14}, \beta'_{24}, \sigma_1'^2\} = \{\beta_{01}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{24}, \sigma_1^2\}.$$

To move from M to M', new parameters are generated from the following
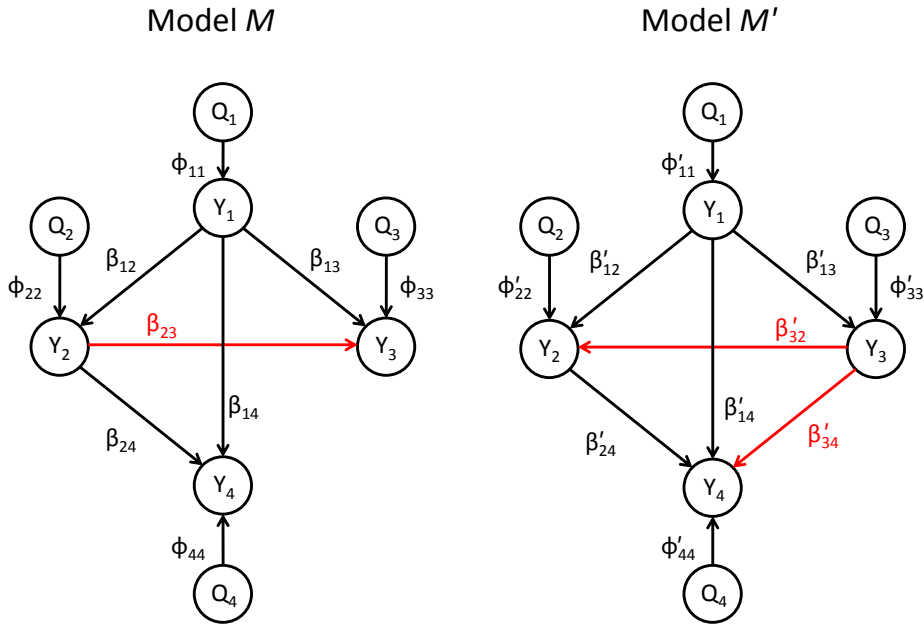
Figure 5.30: *Example causal network structures for both the current (M) and candidate (M') models, given a double step with $d = 3$. Here, we propose the addition of β'_{34} and the reversal of β_{23}.*

proposal distributions:

$$
\begin{aligned}
E_{M'} &= E_M + 1 &&\text{(see Eqn. 5.22)} \\
\beta'_{0t}|\widehat{\beta}'_{0t}, \widehat{\sigma}_t'^2, V_{\beta'_{0t}} &\sim N\left(\widehat{\beta}'_{0t}, \frac{1}{\tau}\widehat{\sigma}_t'^2 V_{\beta'_{0t}}\right) \text{ for } t = 2,3,4 &&\text{(see Eqn. 4.12)} \\
\beta'_{st}|\widehat{\beta}'_{st}, \widehat{\sigma}_t'^2, V_{\beta'_{st}} &\sim N\left(\widehat{\beta}'_{st}, \frac{1}{\tau}\widehat{\sigma}_t'^2 V_{\beta'_{st}}\right) \text{ for } \{st\} \in \{32,34\} &&\text{(see Eqn. 4.13)} \\
\sigma_t'^2|\widehat{\sigma}_t'^2 &\sim \text{Unif}\left(0.5\widehat{\sigma}_t'^2, 1.5\widehat{\sigma}_t'^2\right) \text{ for } t = 2,3,4 &&\text{(see Eqn. 4.14)}
\end{aligned}
$$

where for parameter p, V_p is as defined in Equation 4.11 (p. 78). To move from model M' to M, the parameter proposals are:

$$
\begin{aligned}
E_M &= E_{M'} - 1 &&\text{(see Eqn. 5.23)} \\
\beta_{0t}|\widehat{\beta}_{0t}, \widehat{\sigma}_t^2, V_{\beta_{0t}} &\sim N\left(\widehat{\beta}_{0t}, \frac{1}{\tau}\widehat{\sigma}_t^2 V_{\beta_{0t}}\right) \text{ for } t = 2,3,4 &&\text{(see Eqn. 4.15)} \\
\beta_{23}|\widehat{\beta}_{23}, \widehat{\sigma}_3^2, V_{\beta_{23}} &\sim N\left(\widehat{\beta}_{23}, \frac{1}{\tau}\widehat{\sigma}_3^2 V_{\beta_{23}}\right) &&\text{(see Eqn. 4.16)} \\
\sigma_t^2|\widehat{\sigma}_t^2 &\sim \text{Unif}\left(0.5\widehat{\sigma}_t^2, 1.5\widehat{\sigma}_t^2\right) \text{ for } t = 2,3,4 &&\text{(see Eqn. 4.17)}.
\end{aligned}
$$

In this example, $|Ne(M', x' = x^{D,d'})|$ and $|Ne(M, x = x^{D,d})|$ are of different sizes; the neighbourhoods for the candidate and current models are given in Figures 5.31 and 5.32, respectively. It follows that:

$$
\frac{|Ne(M, x^{D,d})|}{|Ne(M', x' = x^{D,d'})|} = \frac{4}{19}.
$$

The proposal ratio for a move from model M to M', given a double step, is therefore:

$$
\frac{q(\mathbf{u}'_M, \mathbf{S}, x'|\Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x|\Omega, \mathbf{S})} = \frac{q(\mathbf{u}'_M|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \frac{|Ne(M, x = x^{D,d})|}{|Ne(M', x' = x^{D,d'})|} \frac{q(x' = x^D|\Omega', \mathbf{S}')q(d'|x^D, \Omega', \mathbf{S}')}{q(x = x^D|\Omega, \mathbf{S})q(d|x^D, \Omega, \mathbf{S})}.
$$
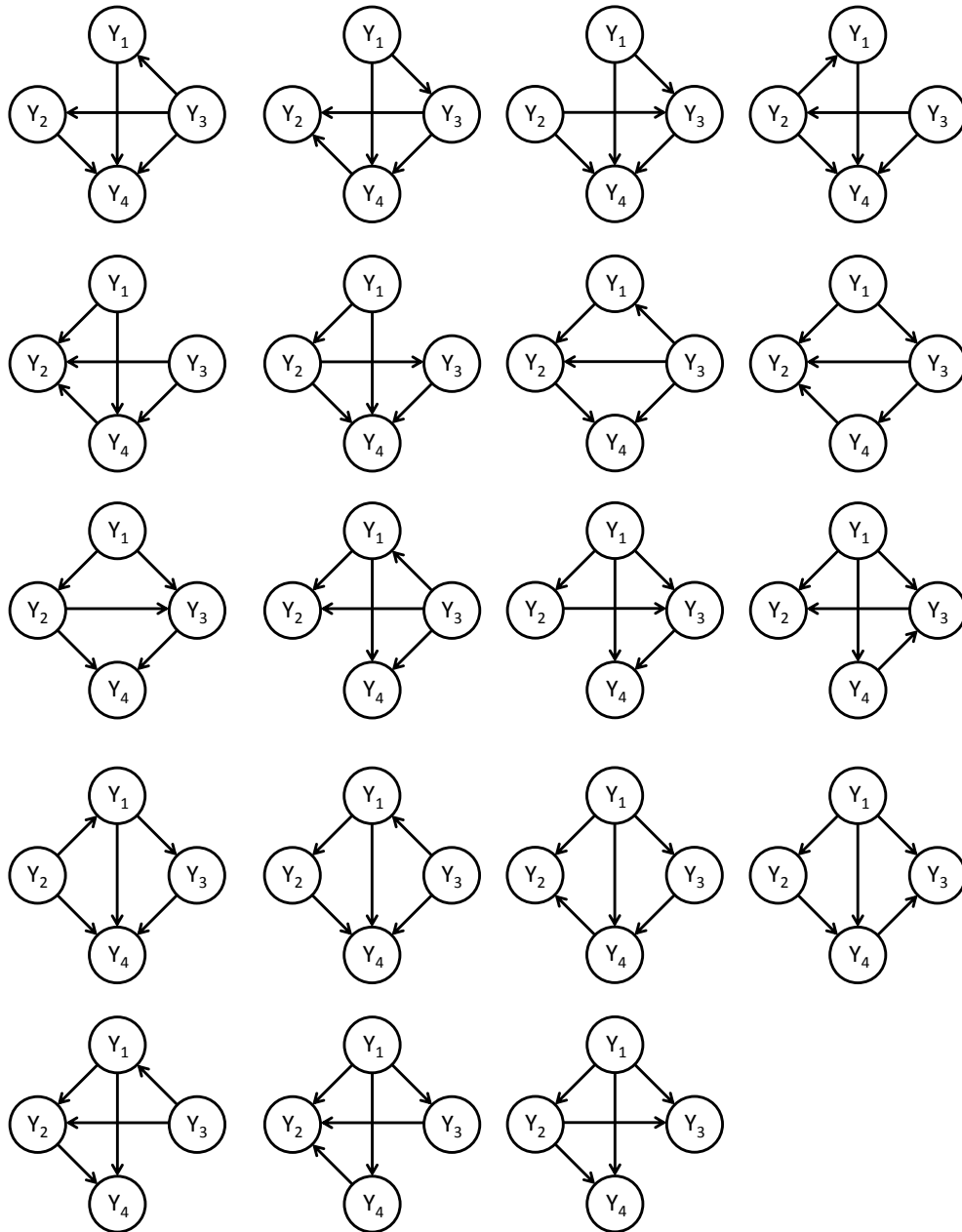
Figure 5.31: *The neighbourhood for the candidate model, model M' in Figure 5.30, given a double step of type $d = 5$; $Ne(M, x' = x^{D,5})$. That is, the reversal of an existing edge and the removal of another existing edge.*
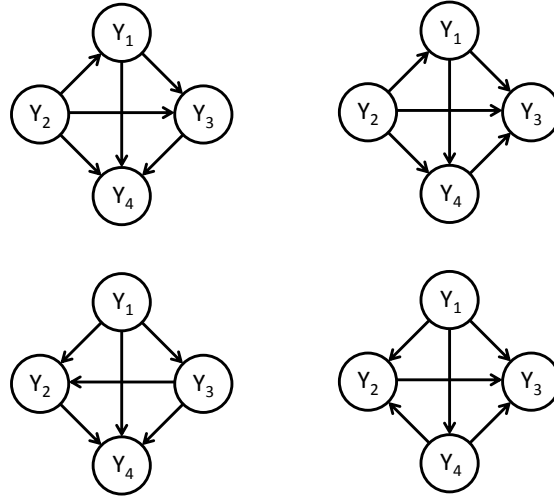
Figure 5.32: *The neighbourhood for the current model, model M in Figure 5.30, given a double step of type $d = 3$; $Ne(M, x = x^{D,3})$. That is, the reversal of an existing edge and the addition of a new distinct edge.*

Here $q(x' = x^D | \Omega', \mathbf{S}')$ is the probability of selecting a double step given $E_{M'} = 6$, given in Table 5.8 (p. 214) as 0.1. Similarly, $q(x = x^D | \Omega, \mathbf{S}) = 0.2$, such that

$$\frac{q(x' = x^D | \Omega', \mathbf{S}')}{q(x = x^D | \Omega, \mathbf{S}))} = \frac{0.1}{0.2} = \frac{1}{2}$$

Given model M and $x = x^D$, the probability of selecting a double step of type $d = 3$ is:

$$q(d = 3 | x^D, \Omega, \mathbf{S}) = \frac{1}{5}$$

as a double step of type $D = 1$ cannot be selected, as $E_M + 2 > max(E)$. Given model M_k' and $x = x^D$, the probability of selecting a double step of type $d' = 5$ is:

$$q(d' = 5 | x^D, \Omega', \mathbf{S}') = \frac{1}{3}$$

as a double step of any type including an add step (i.e. $d = 1, 2, 3$) cannot be selected, as $E_M + 1 > max(E)$. It follows that

$$\frac{q(d'|x^D, \Omega', \mathbf{S}')}{q(d|x^D, \Omega, \mathbf{S})} = \frac{5}{3}.$$

Following Equation 5.20 (p. 217), the proposal ratio for this example is:

$$\frac{q(\mathbf{u}_M', \mathbf{S}, x'|\Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}, \mathbf{S}', x|\Omega, \mathbf{S})} = \frac{q(\mathbf{u}_M'|\mathbf{S}, x', \Omega', \mathbf{S}')}{q(\mathbf{u}_{M'}|\mathbf{S}', x, \Omega, \mathbf{S})} \frac{|Ne(M, x = x^{D,d})|}{|Ne(M', x' = x^{D,d'})|} \frac{q(x' = x^D|\Omega', \mathbf{S}')q(d'|x^D, \Omega', \mathbf{S}')}{q(x = x^D|\Omega, \mathbf{S})q(d|x^D, \Omega, \mathbf{S})}$$

$$= \frac{q(\beta_{02})q(\beta_{03})q(\beta_{04})q(\beta_{23})q(\sigma_2^2)q(\sigma_3^2)q(\sigma_4^2)}{q(\beta_{02}')q(\beta_{03}')q(\beta_{04}')q(\beta_{32}')q(\beta_{34}')q(\sigma_2'^2)q(\sigma_3'^2)q(\sigma_4'^2)} \times \frac{10}{57}.$$

Similarly, the prior ratio is:

$$\frac{p(\Omega'|\mathbf{S}')}{p(\Omega|\mathbf{S})} \times \frac{p(\mathbf{S}')}{p(\mathbf{S})} = \frac{p(\beta_{02}')p(\beta_{03}')p(\beta_{04}')p(\beta_{32}')p(\beta_{34}')p(\sigma_2'^2)p(\sigma_3'^2)p(\sigma_4'^2)}{p(\beta_{02})p(\beta_{03})p(\beta_{04})p(\beta_{23})p(\sigma_2^2)p(\sigma_3^2)p(\sigma_4^2)} \times \exp(-\lambda),$$

assuming the prior on the state of a connection between any two trait nodes is uniform, i.e. $p_{1g} = p_{2g} = p_{3g} = 1/3$. It follows that the acceptance probability for this example double step is then

$$
\begin{aligned}
\alpha_{double} &= min(1, r_{double}) \quad \text{where} \\
r_{double} &= \frac{p(\mathbf{Y}|\Omega', \mathbf{S}', \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})p(\Omega'|\mathbf{S}')p(\mathbf{S}')}{p(\mathbf{Y}|\Omega, \mathbf{S}, \mathbf{Q}, \Lambda, \mathbf{G}, \mathbf{r})p(\Omega|\mathbf{S})p(\mathbf{S})} \times \\
&\qquad \frac{p(\beta_{02}')p(\beta_{03}')p(\beta_{04}')p(\beta_{32}')p(\beta_{34}')p(\sigma_2'^2)p(\sigma_3'^2)p(\sigma_4'^2)}{p(\beta_{02})p(\beta_{03})p(\beta_{04})p(\beta_{23})p(\sigma_2^2)p(\sigma_3^2)p(\sigma_4^2)} \times \\
&\qquad \exp(-\lambda)\frac{q(\beta_{02})q(\beta_{03})q(\beta_{04})q(\beta_{23})q(\sigma_2^2)q(\sigma_3^2)q(\sigma_4^2)}{q(\beta_{02}')q(\beta_{03}')q(\beta_{04}')q(\beta_{32}')q(\beta_{34}')q(\sigma_2'^2)q(\sigma_3'^2)q(\sigma_4'^2)} \times \frac{10}{57}.
\end{aligned}
$$

The general expression of the likelihood is given in Equation 5.3 (p. 72), and the Jacobian is 1 (see Appendix C.5).

5.9.4   Example with simulated data set 5.6

Here we run our reversible jump MCMC approach to estimate an unknown phenotype network structure with a double step in the move space $\mathbb{X}$. We use the data set 5.6, with large QTL effects and large trait-to-trait effects, to illustrate this new step by comparing the results to those in Section 5.5. Previously, this data set correctly identified the true causal network structure with a high posterior probability, although there was poor mixing. Therefore it will provide a good example for the double step. We set $\lambda = 0$ so that any changes will be due to the inclusion of the double step. The results from the five chains repeated with the inclusion of a double step (excluding the relocation step) are presented in Figure 5.33 and the differences resulting from the inclusion of the double step are outlined in Table 5.9. By comparing Figure 5.33 to Figure 5.16 (p. 189) and looking at the summary in Table 5.9, we determine that the removal of the relocate step used in the

Table 5.9: *Comparison of simulation results for data set 5.6 with $\lambda = 0$, including the number of models searched and the joint posterior probability of the true model, $p(M_{TRUE}|D)$.*

		first approach (with relocate step)	second approach (with double step)
Chain 1	models searched	19	22
	$p(M_{TRUE}|D)$	0.8789	0.7894
Chain 2	models searched	4	20
	$p(M_{TRUE}|D)$	0	0.8363
Chain 3	models searched	10	20
	$p(M_{TRUE}|D)$	0	0
Chain 4	models searched	1	3
	$p(M_{TRUE}|D)$	0	0
Chain 5	models searched	29	31
	$p(M_{TRUE}|D)$	0.8122	0.8179

first approach and the introduction of the double step described in Section 5.9 has resulted in the chain visiting additional models compared to the more conservative approach in Section 5.5. Furthermore, the true phenotype network structure has been identified in chain 2, where the approach previously got stuck at a local maxima. The joint posterior probability of the true model did decrease slightly in the the first chain; however, at $p(M_{TRUE}|D) = 0.7894$ the joint posterior probability of the true model is still large, and indicates the true model is the model that best fits the data. The parameter estimates for the true model are given in Table 5.10 for chain 1, and we note that the introduction of the double step (and the removal of the relocation step) has had little effect on the parameter estimates, nor has it affected the size of the 95% credible intervals, as expected. All credible intervals for both approaches contain the true parameter values.

DATA SET 5.6, with a double step					
Chain:	1	2	3	4	5
Initial model:	(diagram)	(diagram)	(diagram)	(diagram)	(diagram)
n(models)	22	20	20	3	31
$p(e_g\|D)$	(diagram)	(diagram)	(diagram)	(diagram)	(diagram)
$p(M_k\|D) = p^*$	(diagram) $p^* = 0.7894$	(diagram) $p^* = 0.8363$	(diagram) $p^* = 0.8147$	(diagram) $p^* = 0.9574$	(diagram) $p^* = 0.8179$
$p(M_{TRUE}\|D)$	0.7894	0.8363	0	0	0.8179

Figure 5.33: *A summary of the performance of each chain for data set 5.6 with a double step. The number of models visited is denoted n(models), and the marginal posterior probability of the directed edge e_g is denoted $p(e_g|D)$. The greatest joint posterior probability is denoted p^*, such that model M_k has the greatest posterior probability when $p(M_k|D) = p^*$. The posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

Table 5.10: *Parameter estimates for the true model for data set 5.6; comparing parameter estimates obtained with and without the double step. Intervals including the true value are indicated by an asterisk (*).*

parameter	true value	first approach (with relocation step)		second approach (with double step)	
		estimated value	95% CI	estimated value	95% CI
β_{01}	0.5	0.532	$(0.397, 0.665)^*$	0.532	$(0.401, 0.664)^*$
β_{02}	0.5	0.504	$(0.361, 0.645)^*$	0.505	$(0.361, 0.649)^*$
β_{03}	0.5	0.451	$(0.304, 0.591)^*$	0.447	$(0.300, 0.588)^*$
β_{04}	0.5	0.562	$(0.392, 0.733)^*$	0.568	$(0.408, 0.729)^*$
β_{05}	0.5	0.473	$(0.276, 0.672)^*$	0.452	$(0.258, 0.655)^*$
β_{12}	1.1	1.116	$(1.044, 1.190)^*$	1.117	$(1.043, 1.195)^*$
β_{13}	1.1	1.135	$(1.062, 1.209)^*$	1.135	$(1.059, 1.210)^*$
β_{24}	1.1	1.092	$(1.043, 1.142)^*$	1.091	$(1.041, 1.142)^*$
β_{35}	1.1	1.081	$(1.022, 1.140)^*$	1.080	$(1.023, 1.139)^*$
β_{45}	1.1	1.124	$(1.078, 1.169)^*$	1.128	$(1.080, 1.176)^*$
ϕ_{11}	1.2	1.042	$(0.862, 1.223)^*$	1.044	$(0.865, 1.226)^*$
ϕ_{22}	1.2	1.086	$(0.922, 1.254)^*$	1.085	$(0.917, 1.261)^*$
ϕ_{33}	1.2	1.19	$(1.018, 1.356)^*$	1.198	$(1.023, 1.368)^*$
ϕ_{44}	1.2	1.115	$(0.946, 1.281)^*$	1.110	$(0.942, 1.270)^*$
ϕ_{55}	1.2	1.113	$(0.935, 1.283)^*$	1.123	$(0.951, 1.301)^*$
σ_1^2	1.0	1.044	$(0.922, 1.182)^*$	1.044	$(0.922, 1.182)^*$
σ_2^2	1.0	0.989	$(0.873, 1.120)^*$	0.988	$(0.872, 1.119)^*$
σ_3^2	1.0	0.978	$(0.863, 1.107)^*$	0.977	$(0.863, 1.104)^*$
σ_4^2	1.0	0.947	$(0.836, 1.073)^*$	0.947	$(0.837, 1.072)^*$
σ_5^2	1.0	1.021	$(0.901, 1.157)^*$	1.022	$(0.901, 1.158)^*$

Chapter 6

Discussion

This thesis has extended the trans-dimensional reversible jump algorithm to develop a Bayesian approach to estimate an unknown phenotype network structure existing among multiple interacting traits. Our approach is a subproblem of causal network structure estimation because we assume that the genetic architecture underlying the given set of traits is known.

Recent QTL mapping methods incorporate multiple traits, and focus on the estimation of an unknown causal network structure existing among a set of quantitative traits, referred to here as the phenotype network structure. The number of possible structures is known to increase rapidly as the number of traits increases, therefore an efficient trans-dimensional search algorithm is required to search both within and between models. It follows that the approach taken in this thesis is in line with current causal network structure determination methods. This extension to standard reversible jump MCMC methods is more flexible than traditional structural equation modeling, allowing the entire model space to be searched without restriction. It has similarities to the QTLnet algorithm proposed by Neto et al. (2010), which uses reversible jump MCMC to jointly infer the phenotype network

structure and genetic components.

The QTLnet algorithm proposes the joint inference of the phenotype network structure and the genetic components, whereas we assume that the genetic architecture is known and focus on the subproblem of estimating the unknown phenotype network and the QTL effects. This has simplified our approach, reducing the number of parameters and allowing us to focus on the estimation of the phenotype network structure. It is not unreasonable to assume that the QTL genotypes and locations are known: Dhungana et al. (2007) used known genetic markers as proxy QTL, for example. Any established QTL mapping method for multiple traits could be used to estimate the genetic architecture prior to the analysis on the phenotype network structure. If the assumed genetic architecture is incorrect, it is likely that our approach will estimate an incorrect phenotype network structure. However, if the genetic architecture is mostly correct we may still estimate the true phenotype network structure provided the search is repeated from different positions in the model space. Future research could remove the assumption of the genetic architecture being known, adding an extra layer to the approach so that it may jointly infer the genetic architecture and the phenotype network structure. Such an extension would be straightforward as there are many existing methods for single- or multiple-trait QTL mapping. However, this would increase the number of parameters, and would require more computing time to obtain results.

Another way in which our approach differs from the QTLnet algorithm is with the proposal distributions used. Our approach makes use of maximum likelihood estimates in the chain. We note that the use of the multivariate normal distribution would be more efficient; however, we obtain the MLE from separate regressions of each trait on the traits and QTL directly affecting it. This still makes for an efficient sampler using well-tuned proposal distributions.

Chapter 3 illustrated that a score-based greedy hill-climbing search strategy using AICc is capable of estimating an unknown phenotype network structure; however, there was a tendency to select models of a higher dimension than the true model, i.e. extra parameters were estimated. We introduce a prior on the number of edges in a phenotype network structure which causes simpler models, models with fewer directed edges, to be preferred. We also have a prior on the graph structure which considers the state of a connection between each pair of trait nodes (here referred to as reversed, absent or forwards). Assuming *a priori* independence among connections, the joint probability for the phenotype network structure of a model is the product over all connections between nodes. This seems to work well; however, we note that if we are to assume a uniform prior on the state of a connection between any two trait nodes, it implies that an edge is more likely to be included in the phenotype network structure (in one of two directions) than not. The prior on a graph structure is a current area of research. Neto et al. (2010) simply assumed a uniform prior over all phenotype network structures, which is the common approach. Recently, however, Sheridan et al. (2010) have proposed priors that consider the degree of a node, i.e. the number of edges connecting to other nodes, whereas Scutari (2013) focuses on the edges. Scutari (2013) uses a multivariate trinomial random variable to model the connections between nodes, focusing on the set of possible edges instead of the set of possible network structures. This reduces the dimension of the sample space from super-exponential to quadratic in the number of variables (Scutari 2013).

In our simulation study in Chapter 5, the prior on the number of edges in the phenotype network structure was found to affect the posterior probability of the true model in various ways, depending on the true parameter values. Most notably, the relationship between the size of the hyperparameter λ and the joint posterior

probability of the true model is not monotonic; there is some interplay between the two. For example, in data set 5.4, with large QTL effects and small trait-to-trait effects, the posterior probability of the true model was maximised when $\lambda = 3$, but further increasing λ only decreased the joint posterior probability, with the sampler searching more models with a lower dimension than the true model. The interplay between the prior and the posterior probability of the true model was more prominent in the data sets simulated with small QTL effects, where the true model isn't easily identifiable. This interplay is unlikely to occur with large effects as they make the model easier to identify. With current research in the area, there is a demand for future research regarding the graph prior. For future research we would like to explore the use of a prior on only those directed edges included in the given causal network structure, instead of a prior on the state of a connection between two nodes. Furthermore, we would like to expand on the prior for the number of edges in a phenotype network structure, specifically focusing on the hyperparameter λ which defines how strict the penalty is, and treating it as an unknown parameter.

Neto et al. (2010) experienced slow mixing of the Markov chain with their QTLnet algorithm, and we have had concerns regarding mixing and convergence in our simulation studies as well. In Chapter 4 we restricted the number of edges within a model and determined that both mixing and convergence were typically poor. However, this was expected as fixing the number of edges restricted the way in which the sampler was able to propose new phenotype network structures. This means that if a model had two or more badly placed edges it couldn't easily fix itself. The solution was to allow the number of edges to vary in Chapter 5. This did improve mixing from Chapter 4, although mixing was still poor.

The QTLnet algorithm searches the model space using standard trans-dimensional

steps, proposing the single addition, removal and reversal of trait-to-trait relationships. Our approach incorporates additional move types from those used in Neto et al. (2010); in Chapter 4 the relocate step was used, equivalent to a simultaneous add and remove (or birth and death) step, with the specification that a new edge cannot be added between the same two nodes from which an edge was removed. This relocation step was then developed into a double step in Chapter 5, combining two steps that alter the structure of the phenotype network, possibly changing the dimension of the model by more than one. Reversible jump Markov chain Monte Carlo methods are known to be very slow moving and generally require long chains with large thinning windows, so combining two steps that alter the causal network structure will help to prevent the chain from getting stuck at local maxima. The inclusion of the double step for data set 5.6 yielded very similar results to the traditional add, remove, reverse and relocate steps, although it helped the sampler move between different phenotype networks — even estimating the true model in a chain previously stuck at a local maxima. The double step has potential for future research, creating different ways to move between models of a different dimension. For example, it could expand on the replace step proposed by Lunn et al. (2009). Lunn et al. (2009) proposed a step to alter the causal network structure by deleting δ edges and adding δ new (distinct) edges. This could be extended to delete δ edges and add γ new (distinct) edges, where δ and γ need not be equal. This type of step proposes candidate models which may differ greatly from the current model, and so the more conservative double step was proposed in this thesis.

The sampler used in this thesis selected candidate models in a different way to the QTLnet algorithm. We define the neighbourhood of a phenotype network structure as being conditional on the selected move type x; instead of containing

all possible candidate phenotype network structures given all move types, as in Neto et al. (2010). This makes it faster to obtain the neighbourhood for a given model when the number of traits is large.

Our trans-dimensional Bayesian approach was applied to the winter wheat data set published in Dhungana et al. (2007). We were not able to obtain the data on the environmental factors used, but our analysis incorporating the direct effects of QTL produced results comparable to those published by Dhungana et al. (2007). Although mixing was once again poor, the phenotype network structure published in Dhungana et al. (2007) was identified as the model with the greatest joint posterior probability. Even though the true model contained the maximum number of directed edges possible for the number of quantitative traits, the true model was estimated when a large value of λ was used in the prior. In other words, the prior for the number of edges in a phenotype network structure is designed to favour moves to phenotype networks with fewer edges, but if the true model has a large number of edges then it can still be estimated, even with a strict prior. The causal network structure of wheat contained effects of all sizes, and generally speaking, the estimated direct and indirect effects of each QTL, or quantitative trait, on grain yield were similar to those published by Dhungana et al. (2007) both in size and sign (i.e. positive or negative).

The example causal networks used in this thesis incorporate only a few trait nodes. In practice, the number of nodes in the causal network can vary greatly. Published data sets such as the winter wheat data set (see for example Dhungana et al. 2007) incorporates only four trait nodes, where as (Neto et al. 2010), for example, use a liver hot spot data set that incorporates 16 trait nodes. Currently, this approach has had to be restricted to small causal network structures because of the R code used to implement the algorithm. Once the code has been optimised we

can extend the approach to accomodate larger networks. This must be done with caution as the size of the model space increases super-exponentially as the number of traits increases. We therefore strongly recommend repeating the search from a variety of different positions to ensure that the model space has been searched appropriately. The extension of the double step will allow for the proposal of larger steps between models, and should increase mixing for larger networks which may require larger steps to escape local maxima.

The trans-dimensional Bayesian approach taken in this thesis addresses three of the six QTL mapping objectives outlined in Chapter 1; it is able to estimate the size of QTL effects on the trait(s) of interest, it can estimate the size of the trait-to-trait effects, and most importantly, it can infer the phenotype network structures existing among multiple interacting traits, thus enabling more accurate estimation of QTL effects on related traits than previous approaches.

Bibliography

Akaike, H. (1973), 'Information theory and an extension of the maximum likelihood principle', Pages 267-281 in Second International Symposium on Information Theory. Akademiai Kaido, Budapest.

Anderson, D. R. (2008), Model Based Inference in the Life Sciences, Springer.

Banerjee, S., Yandell, B. S. & Yi, N. (2008), 'Bayesian Quantitative Trait Loci Mapping for Multiple Traits', Genetics **179**, 2275–2289.

Bentler, P. M. & Bonett, D. G. (1980), 'Significance tests and goodness-of-fit in the analysis of covariance structures', Psychological Bulletin **88**(3), 588–606.

Broman, K. W. (2005), 'The Genomes of Recombinant Inbred Lines', Genetics **169**, 1133–1146.

Broom, B. M., Do, K.-A. & Subramanian (2012), 'Model averaging strategies for structure learning in Bayesian networks with limited data', BMC Bioinformatics **13:S10**.

Burnham, K. P. & Anderson, D. R. (2002), Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second edn, Springer.

Campbell, D. T., Baenziger, P. S., Eskridge, K. M., Budak, H., Erayman, M. & Yen, Y. (2003), 'Identification of QTL and environmental interaction associated with agronomic traits on chromosomes 3A of wheat', Crop Science Society of America **43**, 1493–1505.

Cockerham, C. C. (1954), 'An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present', Genetics **39**, 859–882.

Dhungana, P., Eskridge, K. M., Baenziger, P. S., Campbell, B. T., Gill, K. S. & Dweikat, I. (2007), 'Analysis of Genotype-by-Environment Interaction in Wheat Using a Structual Equation Model and Chromosome Substitution Lines', Crop Science **47**(2), 477–484.

Donmez, E., Sears, R. G., Shroyer, J. P. & Paulsen, G. M. (2001), 'Genetic gain in yield attributes of winter wheat in the Great Plains', Crop Science Society of America **41**, 1412–1419.

Fisher, R. A. (1918), 'The Correlation between Relatives on the Supposition of Mendelian Inheritance', Philosophical Transactions of the Royal Society of Edinburgh **52**, 399–433.

Friedman, N., Nachman, I. & Pe'er, D. (1999), 'Learning bayesian network structure from massive datasets: The "sparse candidate" algorithm', Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc.

Fujikoshi, Y. & Satoh, K. (1997), 'Modified AIC and Cp in multivariate linear regression', Biometrika **84**, 707–716.

Gelman, A. & Rubin, D. B. (1992), 'Inference from iterative simulations using multiple sequences', Statistical Science **7**, 457–511.

Green, P. J. (1995), 'Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination', Biometrika **82 No. 4**, 711–732.

Haavelmo, T. (1943), 'The statistical implications of a system of simultaneous equations', Econometrica **11**, 1–12.

Hackett, C. A., Meyer, R. & Thomas, W. T. B. (2001), 'Multi-trait QTL mapping in barley using multivariate regression', Genetical Research Cambridge **77**, 95–106.

Haldane, J. (1919), 'The combination of linkage values, and the calculation of distances between the loci of linked factors', Journal of Genetics **8**, 299–309.

Heath, S. C. (1997), 'Markov chain Monte Carlo segregation and linkage analysis for oligogenic models', American Journal of Human Genetics **61**, 748–760.

Heckerman, D., Geiger, D. & Chickering, D. (1995), 'Learning Bayesian Networks: The combination of knowledge and statistical data', Machine Learning **20**, 197–243.

Hoeting, J. A., Madigan, D., Raftery, E. & Volinsky, C. T. (1999), 'Bayesian model averaging: a tutorial', Statistical Science **14**(4), 382–417.

Hurvich, C. M. & Tsai, C.-L. (1989), 'Regression and time series model selection in small samples', Biometrika **76**, 297307.

Husmeier, D. (2003), 'Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks', Bioinformatics **19 No. 17**, 2271–2282.

Jackman, S. (2015), pscl: Classes and Methods for R, Political Science Computational Laboratory, Stanford University, California. R package version 1.4.9.
URL: *http://pscl.stanford.edu/*

Jiang, C. & Zeng, Z.-B. (1995), 'Multiple trait analysis of genetic mapping for quantitative trait loci', Genetics **140**, 1111–1127.

Joreskog, K. G. (1970), 'A general method for analysis of covariance structures', Biometrika **57**, 239–251.

Kao, D.-H. & Zeng, Z.-B. (2002), 'Modeling epistasis of quantitative trait loci using Cockerham's model', Genetics **160**, 1243–1261.

Koopmans, T. C. (1953), 'Identification problems in econometric model construction', pp. 27–48.

Koopmans, T. C., Rubin, H. & Leipnik, R. B. (1950), 'Measuring the equation systems of dynamic economics', pp. 53–237.

Kosambi, D. D. (1943), 'The estimation of map distances from recombination values', Annals of Eugenics **2**, 172–175.

Lander, E. S. & Botstein, D. (1989), 'Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps', Genetics **121**, 185–199.

Lauritzen, S. (1996), Graphical Models, Oxford University Press, USA.

Li, R., Tsaih, S. W., Shockley, K., Stylianou, I. M., Wergedal, J., Paigen, B. & Churchill, G. A. (2006), 'Structural model analysis of multiple quantitative traits', PLoS Genetics **2**(7), e114.

Lunn, D. J., Best, N. & Whittaker, J. C. (2009), 'Generic reversible jump MCMC using graphical models', Statistics and Computing **19**(4), 395–408.

Lynch, M. & Walsh, B. (1998), Genetics and Analysis of Quantitative Traits, Sinauer Associates, Inc.

Margaritis, D. (2003), Learning Bayesian network model structure from data, PhD thesis, Carnegie Mellon University.

Mi, X., Eskridge, K., Wang, D., Baenziger, P. S., Campbell, B. T., Gill, K. S. & Dweikat, I. (2010), 'Bayesian mixture structural equation modelling in multiple-trait QTL mapping', Genetical Research **92**, 239–250.

Neto, E. C., Ferrara, C. T., Attie, A. D. & Yandell, B. S. (2008), 'Inferring Causal Phenotype Networks From Segregating Populations', Genetics Society of America **179**, 1089–1100.

Neto, E. C., Keller, M. P., Attie, A. D. & Yandell, B. S. (2010), 'Causal graphical models in systems genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes', Annals of Applied Statistics **4 No. 1**, 320–339.

Neto, E. C. & Yandell, B. S. (2014), qtlnet: Causal Inference of QTL Networks. R package version 1.3.6.
URL: *http://CRAN.R-project.org/package=qtlnet*

Ouyang, J.-W., Liang, H., Jia, S.-E., Zhao, T.-H., He, L.-Z. & Jia, X. (1994), 'Studies on the chromosome doubling of wheat pollen plants', Plant Science **98**(2), 209–214.

Pearl, J. (1988), Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Kaufmann; San Mateo, CA.

Pearl, J. (2000), Causality: Models, Reasoning and Inference, Cambridge Univ. Press, New York.

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006), 'Coda: Convergence diagnosis and output analysis for mcmc', R News **6**(1), 7–11.
URL: *http://CRAN.R-project.org/doc/Rnews/*

R Core Team (2014), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Robinson, R. (1977), Counting unlabeled acyclic digraphs, in 'Combinatorial Mathematics V', Vol. 622 of Lecture Notes in Mathematics, Springer Berlin Heidelberg, pp. 28–43.

Russell, P. (2006), iGenetics: A Molecular Approach, second edn, Pearson/Benjamin Cummings.

Satagopan, J. M., Yandell, B. S., Newton, M. A. & Osborn, T. C. (1996), 'A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo', Genetics **144**, 805–816.

Scutari, M. (2013), 'On the prior and posterior distributions used in graphical modelling', Bayesian Analysis **8**(3), 505–532.

Sen, S. & Churchill, G. A. (2001), 'A Statistical Framework for Quantitative Trait Mapping', Genetics **159**, 371–387.

Sheridan, P., Kamimura, T. & Shimodaira, H. (2010), 'A scale-free structure prior for graphical models with applications in functional genomics', PLoS ONE **5**(11), e13580.

Sibanda, N. (2002), Bayesian Analysis of Genetic Mapping and Related Problems via Simulation Based Techniques, PhD thesis, Imperial College, University of London - London.

Soller, M., Brody, T. & Genizi, A. (1976), 'On the power of experimental design for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines', Theor. Appl. Genet. **47**, 35–39.

Spirtes, P., Glymour, C. & Scheines, R. (2000), Causation, Prediction and Search, 2nd edn, MIT Press; Cambridge, MA.

Stephens, D. & Fisch, R. D. (1998), 'Bayesian Analysis of Quantitative Trait Locus Data Using Reversible Jump Markov Chain Monte Carlo', Biometrics **54**, 1334–1347.

Streck, N. A., Weiss, A. & Baenziger, P. S. (2003), 'A generalized vernalization response function for winter wheat', Agronomy Journal **95**, 155–159.

Tierney, L., Rossini, A. J., Li, N. & Sevcikova, H. (2013), snow: Simple Network of Workstations. R package version 0.3-13.
URL: *http://CRAN.R-project.org/package=snow*

Trow, A. H. (1913), 'Forms of reduplication: primary and secondary', Journal of Genetics **12**, 313–324.

Tsamardinos, I., Brown, L. E. & Aliferis, C. F. (2002), 'Optimal Structure Identification with Greedy Search', Journal of Machine Learning Research **3**, 507–554.

Tsamardinos, I., Brown, L. E. & Aliferis, C. F. (2006), 'The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm', Machine Learning **65**(1), 31–78.

Wright, S. (1921), 'Correlation and causation', Journal of Agricultural Research **20**, 557–585.

Xu, S. (2013), Principles of Statistical Genomics, Springer.

Zeng, Z.-B. (1994), 'Precision mapping of quantitative trait loci', Genetics **136**, 1457–1468.

Zheng, Z. P., Liu, X. H., Huang, Y. B., Wu, X., He, C. & Li, Z. (2012), 'QTLs for days to silking in a recombinant inbred line maize population subjected to high and low nitrogen regimes', Genetics and Molecular Research **11**(2), 790–798.

Zhou, Y. (2011), 'Structure Learning of Probabilistic Graphical Models: A Comprehensive Survey', ArXiv e-prints .

Appendix A

Biological Principles

A.1 Mendelian inheritance

Mendelian inheritance originates from a 1865 paper written by Gregor Mendel, and has been included in many published articles and textbooks since; see for example Russell (2006). Mendel summarized his findings into three laws:

1. Law of Segregation

2. Law of Independent Assortment

3. Law of Dominance

(Russell 2006). These laws are discussed in Sections A.1.1-A.1.3.

A.1.1 Law of Segregation:

During gamete formation, the alleles for each gene segregate (separate) from each other so that each gamete carries only one

allele for each gene (half the gametes carry one allele, and the other half carry the other allele).

Let us focus on a specific individual with the genotype *Aa* at a given locus. During the reproductive process known as meiosis gametes (reproductive cells) are produced, half carrying the allele *A*, half *a*, at the specified locus. A gamete is either an egg (for females) or a sperm (for males) such that the two combine during sexual reproduction and the resulting offspring will inherit one allele from each parent via the gametes.

A.1.2   Law of Independent Assortment:

Genes for different traits can segregate independently of one another in gamete production.

Prior to meiosis the genetic material is replicated and each chromosome is then comprised of two sister chromatids (i.e. two identical copies). When gametes are formed (during meiosis) homologous chromosomes, i.e. pairs of chromosomes with the same arrangement of genetic loci, are known to crossover — transferring genetic material from one chromosome to another. This is recombination and is a feature of independent assortment; see the example in Figure A.1, focusing on a single pair of chromosomes. Independent assortment is the random assortment of homologous chromosomes such that one will be randomly selected for each gamete (i.e. each gamete will contain a mixture of maternal and paternal genes).

Genes located close together on the same chromosome are more likely to be inherited together. The distance between two loci within the genome can be determined using the recombination fraction (also referred to as the recombination frequency) (Xu 2013). The recombination fraction is the frequency with which
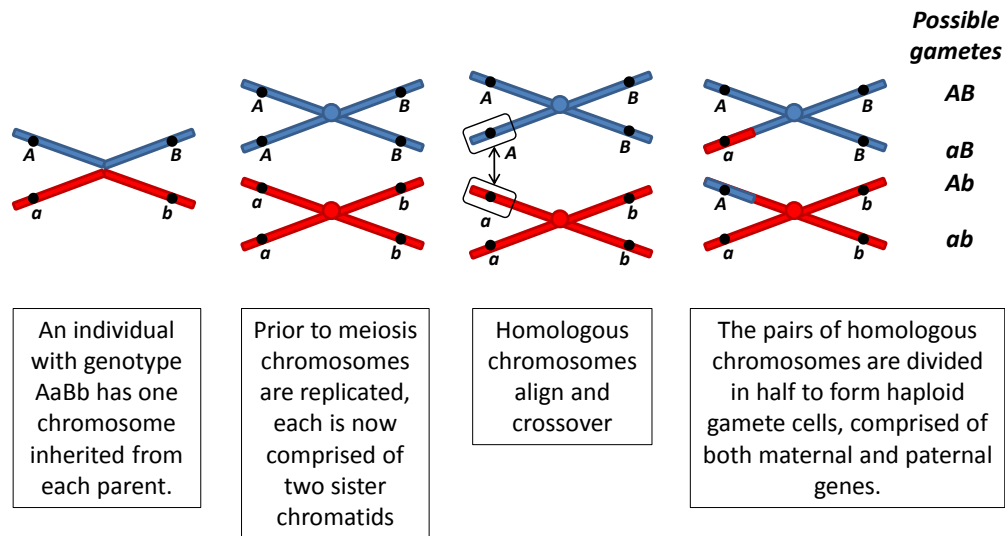
Figure A.1: *An example of independent assortment occurring during gamete production: assume that we have an individual from an F_2 population, with genotype AaBb (with AB on one chromosome, ab on another), where A and a are alleles at locus 1, and B and b are alleles at locus 2. Here the resulting gametes are of genotypes AB, aB, Ab and ab.*

a single chromosomal crossover will take place between two genes during gamete production and has a maximum value of 0.5 (or 50%), which would indicate that the genes are on opposite ends of the chromosome, or perhaps on different chromosomes. If the number of crossovers is odd then the allelic composition of the chromosome now differs from the parental chromosomes and is termed a 'recombinant chromosome' (Russell 2006).

A recombination fraction (θ) between two loci, is used to obtain the map distance via a mapping function. In this thesis we use Haldane's map distances obtained via Haldane's mapping function, although other mapping functions do exist, e.g. Kosambi's mapping function (Kosambi 1943). Given a recombination fraction, θ, the map distance, h, is expressed in Morgans (where 100 centiMorgans (cM) = 1 Morgan), also called map units, where 1cM is approximately equal to a

recombination fraction of 0.01:

$$h = \frac{1}{2}\ln(1 - 2\theta) \tag{A.1}$$

(Haldane 1919, Lynch & Walsh 1998).

Consider the example given in Figure A.2, where we have three genetic makers (labelled A, B, and C) and the recombination fraction between loci j and k is denoted θ_{jk}.

Mapping functions differ by the assumptions made about where crossovers occur. Haldane's mapping function assumes that crossovers occur at any point on the chromosome, randomly and independently, such that Trow's formula holds:

$$\theta_{AC} = \theta_{AB} + \theta_{BC} - 2\theta_{AB}\theta_{BC} \tag{A.2}$$

(Trow 1913). In other words, the recombination fractions are not additive, but the map distances are. Let the map distance between loci j and k be denoted h_{jk}, using Haldane's mapping function given in Equation A.1 we can show that the map distances are additive assuming that crossovers occur at any point on the
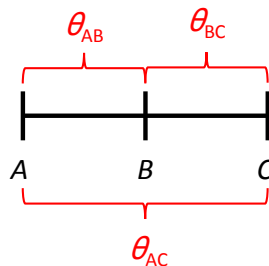
Figure A.2: *Recombination fractions θ_{AB}, θ_{BC} and θ_{AC} between genetic markers A, B and C.*

chromosome (i.e. Trow's formula holds):

$$
\begin{aligned}
h_{AB} + h_{BC} &= \frac{1}{2}\ln(1 - 2\theta_{AB}) + \frac{1}{2}\ln(1 - 2\theta_{BC}) \\
&= \frac{1}{2}\ln\left((1 - 2\theta_{AB})(1 - 2\theta_{BC})\right) \\
&= \frac{1}{2}\ln\left(1 - 2\theta_{AB} - 2\theta_{BC} + 4\theta_{AB}\theta_{BC}\right) \\
&= \frac{1}{2}\ln\left(1 - 2(\theta_{AB} + \theta_{BC} - 2\theta_{AB}\theta_{BC})\right) \\
&= \frac{1}{2}\ln\left(1 - 2\theta_{AC}\right) \quad \text{(see Eqn. A.2)} \\
&= h_{AC}.
\end{aligned}
$$

Note that Haldane's mapping function can be rearranged to calculate the recombination fraction given the distance, h.

A.1.3 Law of Dominance:

Some alleles are dominant while others are recessive; an organism with at least one dominant allele will display the effect of the dominant allele.

Alleles can be either dominant or recessive and are denoted, for example, *A* and *a* respectively. If the two alleles are the same, e.g. *AA* or *aa*, the QTL genotype is said to be homozygous: homozygous dominant for *AA*, and homozygous recessive for *aa*. A heterozygous QTL genotype refers to the presence of one dominant and one recessive allele, e.g. *Aa*. A recessive allele is only fully expressed in the phenotype when the genotype is homozygous recessive. However the dominant allele is expressed in the phenotype with a homozygous dominant genotype, or with a heterozygous genotype where it may mask the effect of the recessive allele.

A.2   Populations used in QTL mapping studies

There are many different population types which are used in QTL mapping stud-
ies. Here, in Sections A.2.1 – A.2.4, we introduce four of the population types
commonly used in QTL mapping literature: F_2, backcross, doubled haploid and a
recombinant inbred line (RIL), respectively.

A.2.1 F_2 population

An F_2 population is obtained by crossing (mating) two different homozygous in-
bred lines to form an F_1 (first filial) generation. This F_1 population is comprised
entirely of heterozygotes - with one allele inherited from each parent. Individuals
from the F_1 generation are then crossed to obtain the F_2 (second filial) generation.
See Figure A.3 for an example cross.
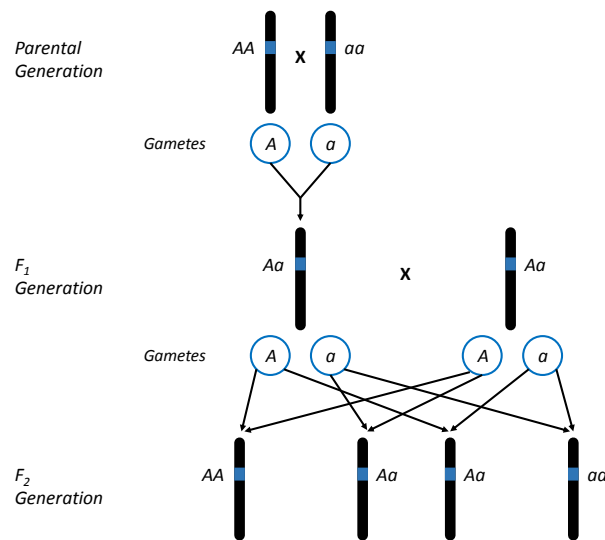
Figure A.3: *An example of the creation of an F_2 population, focusing on one locus
with alleles A and a.*

Individuals from an F_2 population have one of three genotypes at each locus (Russell 2006). These three genotypes are homozygous recessive, heterozygous, and homozygous dominant; for example, *aa*, *Aa* and *AA*, respectively. Consider the following example incorporating one locus from an F_2 population. Let the parents have genotypes *Aa* and *Aa* (both from the F_1 generation), the offspring resulting from this cross then have one of three genotypes at the given locus, shown in the following Punnett square:

	A	*a*
A	*AA*	*Aa*
a	*Aa*	*aa*

With three possible genotypes at a locus, both additive and dominance effects can be estimated for an F_2 population. This is achieved by using the Cockerham genetic model; see Cockerham (1954), Kao & Zeng (2002). Building on the Fisher genetic model which uses the least squares principle to partition the conditional phenotypic value into three types of QTL effects additive, dominance, and epistatic (Fisher 1918), the Cockerham genetic model further partitions the epistatic variance into components using orthogonal contrasts.

A.2.2 Backcross population

Individuals from a backcross population have one of two genotypes (Russell 2006), which makes them simpler to analyse than more complicated populations such as an F_2 population. A backcross population is obtained by crossing two inbred lines to form an F_1 generation, then individuals from the F_1 generation are crossed with

one of the parents to obtain the backcross population, see Figure A.4 for an example cross.

Consider the following example incorporating one locus from a backcross population. Let the parents have genotypes *Aa* (from the F_1 generation) and *aa* (from the parental generation), the possible genotypes at this locus for offspring resulting from this cross are shown in the following Punnett square:

	a	*a*
A	*Aa*	*Aa*
a	*aa*	*aa*
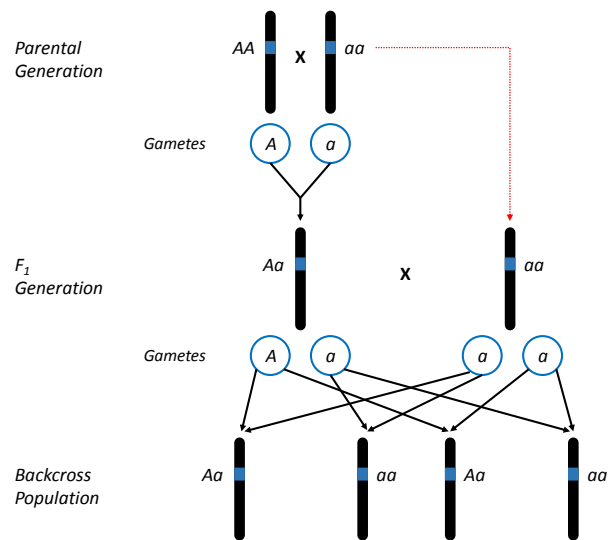
Figure A.4: *An example of the creation of a backcross population, focusing on one locus with alleles A and a.*

A.2.3 Doubled Haploid Population

Individuals with one copy of each chromosome are said to be haploid (Russell 2006). Doubled haploid populations are often used in studies involving plants as haploid plants are sterile so the chromosomes must be doubled to make them fertile. A popular method of obtaining a doubled haploid population is to apply the drug colchicine on the roots of haploid plants to double the chromosomal information and produce doubled haploid shoots (Ouyang et al. 1994). Figure A.5 illustrates that as the chromosomes are doubled, the resulting genotypes are homozygous only — no heterozygotes exist within a doubled haploid population. An advantage of a doubled haploid population is that the production of doubled haploids only requires one generation, and is much faster than other breeding procedures required to obtain homozygotes, such as a recombinant inbred line (Ouyang et al. 1994).

The analysis for a doubled haploid population is similar to that for a backcross population as there are only two possible combinations of alleles at each locus.
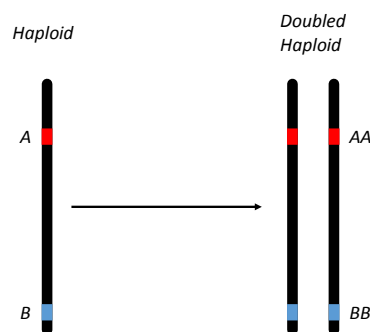
Figure A.5: *An example of the production of a doubled haploid population with alleles A and B at two loci.*

A.2.4   Recombinant Inbred Line

A Recombinant Inbred Line (RIL) is the result of crossing two inbred strains then continuously selfing or sibling mating to create inbred lines. This process takes time, but the advantage is that each strain can be considered eternal because recombination will not change the homozygous individuals, proving an unlimited source from which data can be obtained (Zheng et al. 2012). Figure A.6 illustrates the formation of recombinant inbred lines by selfing, adapted from Broman (2005).
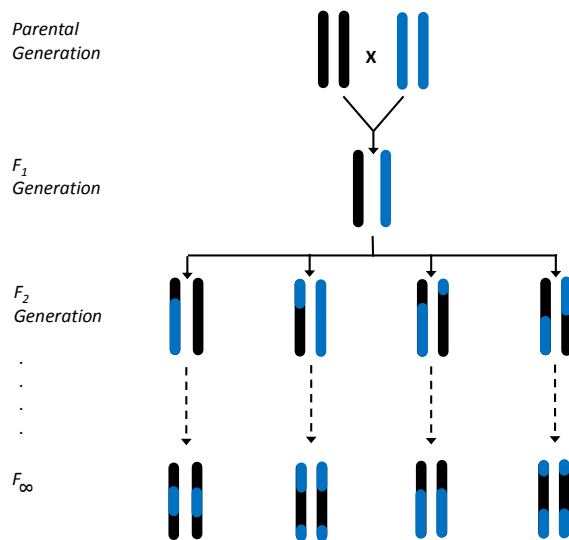
Figure A.6: *An example of the formation of recombinant inbred lines by selfing.*

Appendix B

Data simulation

Chapters 3, 4, and 5 all simulate data as described in this section, and the notation used here in summarized in Table 3.1 (p. 37). In order to simulate data for multiple traits, a true causal network must be decided upon. That is, the genetic architecture and the phenotype network structure are set. From Section 1.2, the genetic architecture includes:

- the number of QTL

- the QTL locations

- the QTL genotypes

- the knowledge of which QTL affect which traits.

Knowing which QTL affect which traits means that we know which QTL parameters we need to estimate, as in a multiple trait analysis not all QTL affect all traits, i.e. the set of QTL parents for each trait are known

$$\upsilon_Q = (\upsilon_Q(1), \cdots, \upsilon_Q(N_T)).$$

257

The phenotype network is estimated using a model selection procedure, such that the set of trait parents for each trait will vary. The set of parent traits is denoted by υ_Y;

$$\upsilon_Y = (\upsilon_Y(1), \cdots, \upsilon_Y(N_T)).$$

This explanation of data simulation will be further simplified by using a doubled haploid population, where the genotypes are recorded as "1" or "0" (homozygous dominant and recessive, respectively), but can easily be extended to model other populations.

Summary of how data is simulated:

1. Randomly sample N_M marker genotypes for *n* individuals according to the chosen breeding design.

   Once the marker genotypes have been simulated the locations of N_{qtl} QTL are specified by the user, where the location of the ℓth QTL is denoted $\lambda_\ell = (\xi_\ell, \gamma_\ell, \theta_\ell)$ as described in Section 1.1 and illustrated in Figure 1.1, included here for convenience. We assume that only one QTL can be located
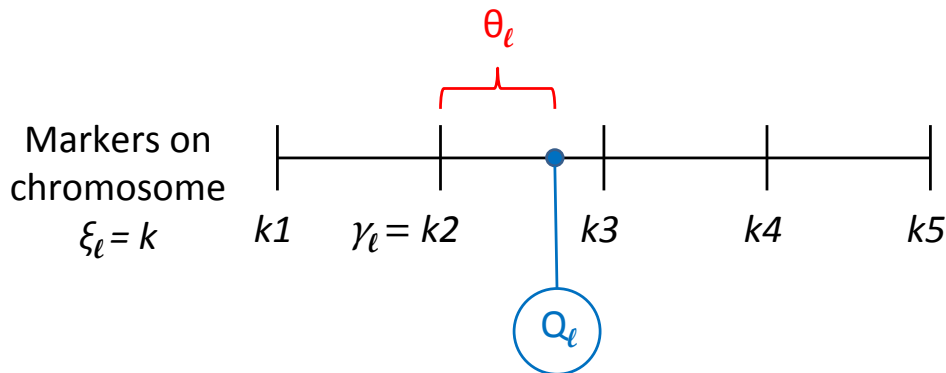
The location of the ℓth QTL is defined by the chromosome it is located on ($\xi_\ell = k$), the left flanking marker ($\gamma_\ell = k2$) and the recombination fraction (θ_ℓ).

within each marker interval (see for example, Zeng 1994).

2. For each of the N_{qtl} QTL, generate QTL genotypes for all $i = 1, \cdots, n$ individuals, denoted $q_{i\ell}$ for $\ell = 1, \cdots, N_{qtl}$, given the flanking marker genotypes of each individual and the position of the QTL within the interval using conditional probabilities as given in Table B.1 (for a doubled haploid population).

3. Decide upon the causal network structure; specify the set of QTL and trait parent nodes, denoted $v_Q(t)$ and $v_Y(t)$ for each of the $t = 1, \cdots, N_T$ quantitative traits. Also specify the values for the corresponding QTL-to-trait and trait-to-trait effect parameters. An example causal network structure is given in Figure 1.2, with three quantitative traits, denoted Y_1, Y_2 and Y_3, and three QTL, denoted Q_1, Q_2 and Q_3, given here for convenience.

4. Given the causal network structure (i.e. all necessary QTL effects and trait-to-trait effects), generate quantitative trait values via equations B.1 and B.2. Traits which are the furthest upstream (i.e. those with no parent traits) are

Table B.1: *The probability of the ℓth QTL genotype for individual i from a doubled haploid population, given the left- and right-flanking marker genotypes (g_{i1} and g_{i2}, respectively). Let $p = \frac{r_{G_1 Q_\ell}}{r_{G_1 G_2}}$, where $r_{G_1 Q_\ell}$ is the recombination fraction between marker G_1 and the ℓth QTL (Q_ℓ), and $r_{G_1 G_2}$ is the recombination fraction between markers G_1 and G_2.*

Marker Genotype		Genotype of ℓth QTL, $q_{i\ell}$	
g_{i1}	g_{i2}	*AA* (1)	*aa* (0)
0	0	0	1
0	1	p	$1 - p$
1	0	$1 - p$	p
1	1	1	0
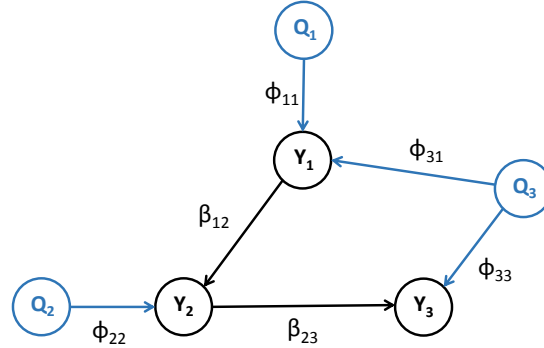
An example causal network structure with three quantitative traits (Y_1, Y_2, Y_3), three QTL (Q_1, Q_2, Q_3), two trait-to trait effects (β_{12}, β_{23}) and four QTL effects (ϕ_{11}, ϕ_{22}, ϕ_{31}, ϕ_{33}).

simulated first as they are only affected by QTL. The quantitative trait values for *n* individuals are drawn from a normal distribution as follows:

$$Y_{it} \sim N\left(\mu_{it} = \beta_0 + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell}, \sigma_t^2\right) \tag{B.1}$$

given the specified QTL effects of each ℓth QTL on the *t*th trait, denoted $\phi_{\ell t}$. The intercept, β_0, and the residual variance for trait *t*, σ_t^2, are also specified by the user. Equation B.1 should be used if $\upsilon_Y(t) = \emptyset$, i.e. no other quantitative traits have a direct effect on trait Y_t. Those traits with parent traits must incorporate the additive effects of the traits directly upstream, denoted $\upsilon_Y(t)$ for trait *t*:

$$y_{it} \sim N\left(\mu_{it} = \beta_{0t} + \sum_{\ell \in \upsilon_Q(t)} \phi_{\ell t} q_{i\ell} + \sum_{k \in \upsilon_Y(t)} \beta_{kt} y_{it}, \sigma_t^2\right) \tag{B.2}$$

where β_{kt} is a trait-to-trait effect, specifically the effect of trait *k* on trait *t* as specified by the user in the causal network structure.

Appendix C

The Jacobian

This Appendix supplements Chapters 4 and 5, following Section 4.1.4 to compute the Jacobian for the move types used in our reversible jump Markov chain Monte Carlo approach. The Jacobians for the reverse, and relocation steps are used in Chapter 4, and are computed in Sections C.1 – C.2. The Jacobians for the add, remove and double steps are used in Chapter 5, and are computed in Sections C.3 – C.5.

Most parameters in Ω and Ω' are in the same space and so we partition Ω and Ω' to reflect the fact that they share a set of parameters, with $\Omega_{MM'}$ and $\Omega_{M'M}$ denoting those parameters in both models M and M', and $\Omega_{M\setminus M'}$ denoting those parameters in model M, but not M'. It follows that:

$$\Omega = (\Omega_{MM'}, \Omega_{M\setminus M'})$$
$$\Omega' = (\Omega'_{M'M}, \Omega'_{M'\setminus M})$$

C.1   Jacobian for the reverse step

A reverse step alters the phenotype network structure, removing a directed edge from model M and adding the reversed directed edge to form $\mathbf{S}'$, the phenotype network structure for model M' ($M' = \{\Omega', \mathbf{S}'\}$). Consider the example given in Section 4.3.1, using models M and M' shown in Figure 4.3 (p. 91). We are reversing β_{23}, such that:

$$
\begin{aligned}
\Omega &= (\Omega \setminus \beta_{23}, \beta_{23}) \\
\Omega' &= (\Omega' \setminus \beta'_{32}, \beta'_{32}) \\
\mathbf{u}_{M'} &= \beta'_{32} \\
\mathbf{u}'_M &= \beta_{23},
\end{aligned}
$$

and

$$
\begin{aligned}
(\Omega, \mathbf{u}_{M'}) &= (\Omega \setminus \beta_{23}, \beta_{23}, \beta'_{32}) \\
(\Omega', \mathbf{u}'_M) &= (\Omega' \setminus \beta'_{32}, \beta'_{32}, \beta_{23}).
\end{aligned}
$$

The Jacobian is calculated according to Equation 4.20 (p. 83):

$$
\begin{aligned}
J_S &= \left| \frac{\partial (\Omega', \mathbf{u}'_M)}{\partial (\Omega, \mathbf{u}_{M'})} \right| && \text{(C.1)} \\[2ex]
&= \left| \frac{\partial (\Omega' \setminus \beta'_{32}, \beta'_{32}, \beta_{23})}{\partial (\Omega \setminus \beta_{23}, \beta_{23}, \beta'_{32})} \right| \\[2ex]
&= \begin{vmatrix}
\dfrac{\partial \Omega' \setminus \beta'_{32}}{\partial \Omega \setminus \beta_{23}} & \dfrac{\partial \Omega' \setminus \beta'_{32}}{\partial \beta_{23}} & \dfrac{\partial \Omega' \setminus \beta'_{32}}{\partial \beta'_{32}} \\[2ex]
\dfrac{\partial \beta'_{32}}{\partial \Omega \setminus \beta_{23}} & \dfrac{\partial \beta'_{32}}{\partial \beta_{23}} & \dfrac{\partial \beta'_{32}}{\partial \beta'_{32}} \\[2ex]
\dfrac{\partial \beta_{23}}{\partial \Omega \setminus \beta_{23}} & \dfrac{\partial \beta_{23}}{\partial \beta_{23}} & \dfrac{\partial \beta_{23}}{\partial \beta'_{32}}
\end{vmatrix}
\end{aligned}
$$

$$J_S = \begin{vmatrix} I_{11,11} & \underline{0}_{11,1} & \underline{0}_{11,1} \\ \underline{0}_{1,11} & 0 & 1 \\ \underline{0}_{1,11} & 1 & 0 \end{vmatrix}$$

$$= -1 \qquad\qquad (C.2)$$

where $\underline{0}_{n,p}$ denotes a matrix of zeros with n rows and p columns, $I_{n,n}$ denotes an identity matrix of size n, and $|\Omega| = |\Omega'| = 12$. We take the absolute value of the Jacobian, so $J_S = 1$. This result from the example reverse move extends to the reversal of any edge.

C.2 Jacobian for the relocate step

A relocate step alters the phenotype network structure, removing edge g from model M and relocating it at location f such that edge g is removed, and the new distinct edge f is added to form the acyclic phenotype network structure $\mathbf{S}'$ for model M' ($M' = \{\Omega', \mathbf{S}'\}$).

Consider the example given in Section 4.4.1 (p. 97), using models M and M' shown in Figure 4.5 (p. 97). We are relocating β_{12} to β'_{23}, it follows that:

$$\Omega = (\Omega \setminus \beta_{12}, \beta_{12})$$
$$\Omega' = (\Omega' \setminus \beta'_{23}, \beta'_{23})$$
$$\mathbf{u}_{M'} = \beta'_{23}$$
$$\mathbf{u}'_M = \beta_{12},$$

such that

$$(\Omega, \mathbf{u}_{M'}) = (\Omega \setminus \beta_{12}, \beta_{12}, \beta'_{23})$$
$$(\Omega', \mathbf{u}'_M) = (\Omega' \setminus \beta'_{23}, \beta'_{23}, \beta_{12}).$$

The Jacobian is calculated according to Equation 4.20 (p. 83):

$$
\begin{aligned}
J_C &= \left| \frac{\partial(\Omega', \mathbf{u}'_M)}{\partial(\Omega, \mathbf{u}_{M'})} \right| \\[2mm]
&= \left| \frac{\partial(\Omega' \setminus \beta'_{23}, \beta'_{23}, \beta_{12})}{\partial(\Omega \setminus \beta_{12}, \beta_{12}, \beta'_{23})} \right| \\[2mm]
&= \left|
\begin{array}{ccc}
\dfrac{\partial \Omega \setminus \beta'_{23}}{\partial \Omega \setminus \beta_{12}} & \dfrac{\partial \Omega \setminus \beta'_{23}}{\partial \beta_{12}} & \dfrac{\partial \Omega \setminus \beta'_{23}}{\partial \beta'_{23}} \\[3mm]
\dfrac{\partial \beta'_{23}}{\partial \Omega \setminus \beta_{12}} & \dfrac{\partial \beta'_{23}}{\partial \beta_{12}} & \dfrac{\partial \beta'_{23}}{\partial \beta'_{23}} \\[3mm]
\dfrac{\partial \beta_{12}}{\partial \Omega \setminus \beta_{12}} & \dfrac{\partial \beta_{12}}{\partial \beta_{12}} & \dfrac{\partial \beta_{12}}{\partial \beta'_{23}}
\end{array}
\right| \\[2mm]
&= \left|
\begin{array}{ccc}
I_{10,10} & \underline{0}_{10,1} & \underline{0}_{10,1} \\[2mm]
\underline{0}_{1,10} & 0 & 1 \\[2mm]
\underline{0}_{1,10} & 1 & 0
\end{array}
\right| \\[2mm]
&= -1 \qquad\qquad\qquad\qquad\qquad\text{(C.3)}
\end{aligned}
$$

where $\underline{0}_{n,p}$ denotes a matrix of zeros with n rows and p columns, $I_{n,n}$ denotes an identity matrix of size n, and $|\Omega| = |\Omega'| = 11$. We take the absolute value of the Jacobian, so $J_C = 1$. This result from the example relocate step extends to the relocation of any edge.

C.3   Jacobian for the add step

An add step alters the phenotype network structure, adding an edge to model M.

Consider the example given in Section 5.2.1 (p. 155), using models M and M' shown in Figure 5.1 (p. 156). We are adding β'_{23}, it follows that:

$$
\begin{aligned}
\Omega &= \Omega \\
\Omega' &= (\Omega' \setminus \beta'_{23}, \beta'_{23})
\end{aligned}
$$

$$\mathbf{u}_{M'} = \beta'_{23}$$
$$\mathbf{u}'_M = \emptyset$$

such that

$$(\Omega, \mathbf{u}_{M'}) = (\Omega, \beta'_{23})$$
$$(\Omega', \mathbf{u}'_M) = (\Omega' \setminus \beta'_{23}, \beta'_{23}).$$

The Jacobian is calculated according to Equation 4.20 (p. 83):

$$
\begin{aligned}
J_A &= \left| \frac{\partial(\Omega', \mathbf{u}'_M)}{\partial(\Omega, \mathbf{u}_{M'})} \right| \\[2mm]
&= \left| \frac{\partial(\Omega' \setminus \beta'_{23}, \beta'_{23})}{\partial(\Omega, \beta'_{23})} \right| \\[2mm]
&= \left| \begin{matrix} \frac{\partial \Omega' \setminus \beta'_{23}}{\partial \Omega} & \frac{\partial \Omega' \setminus \beta'_{23}}{\partial \beta'_{23}} \\[2mm] \frac{\partial \beta'_{23}}{\partial \Omega} & \frac{\partial \beta'_{23}}{\partial \beta'_{23}} \end{matrix} \right| \\[2mm]
&= \left| \begin{matrix} I_{12,12} & \underline{0}_{12,1} \\[1mm] \underline{0}_{1,12} & 1 \end{matrix} \right| \\[2mm]
&= 1
\end{aligned}
\qquad (\text{C.4})
$$

where $\underline{0}_{n,p}$ denotes a matrix of zeros with n rows and p columns, $I_{n,n}$ denotes an identity matrix of size n, and $|\Omega| = 12$ and $|\Omega'| = 13$. This result from the example add step extends to the addition of any edge.

C.4   Jacobian for the remove step

A remove step alters the phenotype network structure, removing an edge from model M.

Consider the example given in Section 5.3.1 (p. 161), using models M and M' shown in Figure 5.3 (p. 162).

We are removing β_{23}, it follows that:

$$
\begin{aligned}
\Omega &= (\Omega \setminus \beta_{23}, \beta_{23}) \\
\Omega' &= \Omega' \\
\mathbf{u}_{M'} &= \emptyset \\
\mathbf{u}'_M &= \beta_{23}
\end{aligned}
$$

such that

$$
\begin{aligned}
(\Omega, \mathbf{u}_{M'}) &= (\Omega \setminus \beta_{23}, \beta_{23}) \\
(\Omega', \mathbf{u}'_M) &= (\Omega', \beta_{23}).
\end{aligned}
$$

For this example of the remove step the Jacobian is calculated according to Equation 4.20 (p. 83):

$$
\begin{aligned}
J_R &= \left| \frac{\partial (\Omega', \mathbf{u}'_M)}{\partial (\Omega, \mathbf{u}_{M'})} \right| \\
&= \left| \frac{\partial (\Omega', \beta_{23})}{\partial (\Omega \setminus \beta_{23}, \beta_{23})} \right| \\
&= \left| \begin{matrix} \frac{\partial \Omega'}{\partial \Omega \setminus \beta_{23}} & \frac{\partial \Omega'}{\partial \beta_{23}} \\ \frac{\partial \beta_{23}}{\partial \Omega \setminus \beta_{23}} & \frac{\partial \beta_{23}}{\partial \beta_{23}} \end{matrix} \right| \\
&= \left| \begin{matrix} I_{12,12} & \underline{0}_{12,1} \\ \underline{0}_{1,12} & 1 \end{matrix} \right| \\
&= 1 \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{(C.5)}
\end{aligned}
$$

where $\underline{0}_{n,p}$ denotes a matrix of zeros with n rows and p columns, $I_{n,n}$ denotes an identity matrix of size n, and $|\Omega| = 13$ and $|\Omega'| = 12$. This result from the example remove step extends to the removal of any edge.

C.5 Jacobian for the double step

A double step alters the phenotype network structure in one of six ways, combining pairs of add, remove and reverse steps to propose changes to model M.

Consider the example given in Section 5.9.3 (p. 220), we propose a double step with $d = 3$, such that we add an edge, and reverse an edge. We use models M and M' shown in Figure 5.30 (p. 221). We are therefore removing β_{23} from model M and adding β'_{32} and β'_{34}, it follows that:

$$
\begin{aligned}
\Omega &= (\Omega \setminus \beta_{23}, \beta_{23}) \\
\Omega' &= (\Omega' \setminus \{\beta'_{32}, \beta'_{34}\}, \beta'_{32}, \beta'_{34}) \\
\mathbf{u}_{M'} &= (\beta'_{32}, \beta'_{34}) \\
\mathbf{u}'_M &= \beta_{23}
\end{aligned}
$$

such that

$$
\begin{aligned}
(\Omega, \mathbf{u}_{M'}) &= (\Omega \setminus \beta_{23}, \beta_{23}, \beta'_{32}, \beta'_{34}) \\
(\Omega', \mathbf{u}'_M) &= (\Omega' \setminus \{\beta'_{32}, \beta'_{34}\}, \beta'_{32}, \beta'_{34}, \beta_{23}).
\end{aligned}
$$

The Jacobian is calculated according to Equation 4.20 (p. 83):

$$
\begin{aligned}
J_D &= \left| \frac{\partial (\Omega', \mathbf{u}'_M)}{\partial (\Omega, \mathbf{u}_{M'})} \right| \\
&= \left| \frac{\partial (\Omega' \setminus \{\beta'_{32}, \beta'_{34}\}, \beta'_{32}, \beta'_{34}, \beta_{23})}{\partial (\Omega \setminus \beta_{23}, \beta_{23}, \beta'_{32}, \beta'_{34})} \right|
\end{aligned}
$$

$$J_D \;=\; \left| \begin{array}{cccc}
\dfrac{\partial \Omega' \backslash \{\beta'_{32}, \beta'_{34}\}}{\partial \Omega \backslash \beta_{23}} & \dfrac{\partial \Omega' \backslash \{\beta'_{32}, \beta'_{34}\}}{\partial \beta_{23}} & \dfrac{\partial \Omega' \backslash \{\beta'_{32}, \beta'_{34}\}}{\partial \beta'_{32}} & \dfrac{\partial \Omega' \backslash \{\beta'_{32}, \beta'_{34}\}}{\partial \beta'_{34}} \\[1.2em]
\dfrac{\partial \beta'_{32}}{\partial \Omega \backslash \beta_{23}} & \dfrac{\partial \beta'_{32}}{\partial \beta_{23}} & \dfrac{\partial \beta'_{32}}{\partial \beta'_{32}} & \dfrac{\partial \beta'_{32}}{\partial \beta'_{34}} \\[1.2em]
\dfrac{\partial \beta'_{34}}{\partial \Omega \backslash \beta_{23}} & \dfrac{\partial \beta'_{34}}{\partial \beta_{23}} & \dfrac{\partial \beta'_{34}}{\partial \beta'_{32}} & \dfrac{\partial \beta'_{34}}{\partial \beta'_{34}} \\[1.2em]
\dfrac{\partial \beta_{23}}{\partial \Omega \backslash \beta_{23}} & \dfrac{\partial \beta_{23}}{\partial \beta_{23}} & \dfrac{\partial \beta_{23}}{\partial \beta'_{32}} & \dfrac{\partial \beta_{23}}{\partial \beta'_{34}}
\end{array} \right|$$

$$=\; \left| \begin{array}{cccc}
I_{17,17} & \underline{0}_{17,1} & \underline{0}_{17,1} & \underline{0}_{17,1} \\[0.8em]
\underline{0}_{1,17} & 0 & 1 & 0 \\[0.8em]
\underline{0}_{1,17} & 0 & 0 & 1 \\[0.8em]
\underline{0}_{1,17} & 1 & 0 & 0
\end{array} \right|$$

$$=\; -1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(C.6)}$$

where $\underline{0}_{n,p}$ denotes a matrix of zeros with n rows and p columns, $I_{n,n}$ denotes an identity matrix of size n, and $|\Omega| = 18$ and $|\Omega'| = 19$. We take the absolute value of the Jacobian, so $J_D = 1$. For all different types of double step, the Jacobian is 1.

Appendix D

Trace plots for data set 4.2

Included in this appendix are the trace plots and corresponding densities for all parameters estimated by the chain starting with initial model 1 in Figure 4.17 for data set 4.2, to illustrate how the chain is mixing. We include all stored states of the chain, and if a parameter is not included here, then it was not estimated at all in the chain.

These trace plots were created in R using the $traceplot()$ function in the 'coda' package with the default bandwidth; see Plummer et al. (2006).
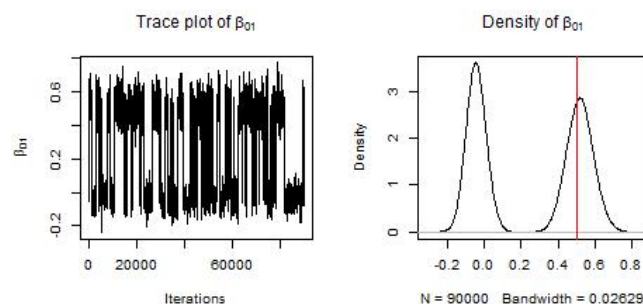
Figure D.1: *Trace plot and density for parameter* β_{01}.
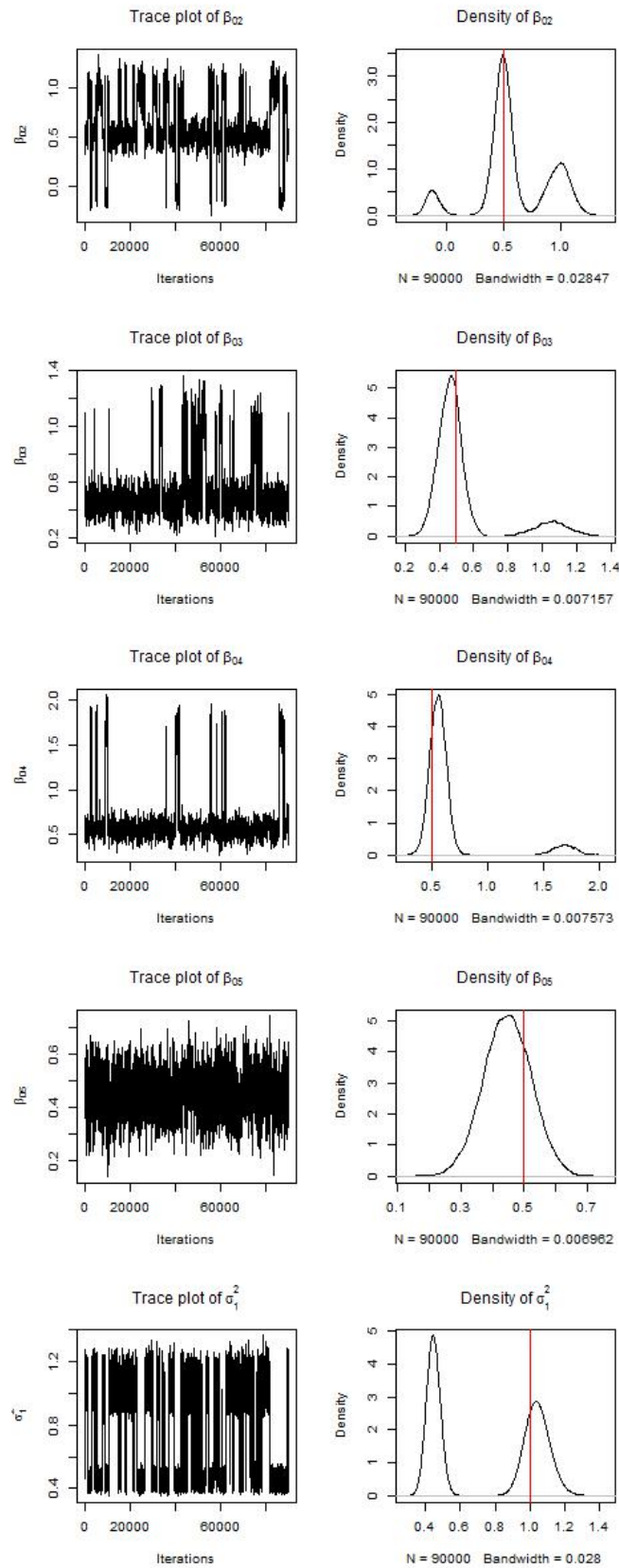
Figure D.2: *Trace plots and densities for parameters β_{02}, β_{03}, β_{04}, β_{05} and σ_1^2.*
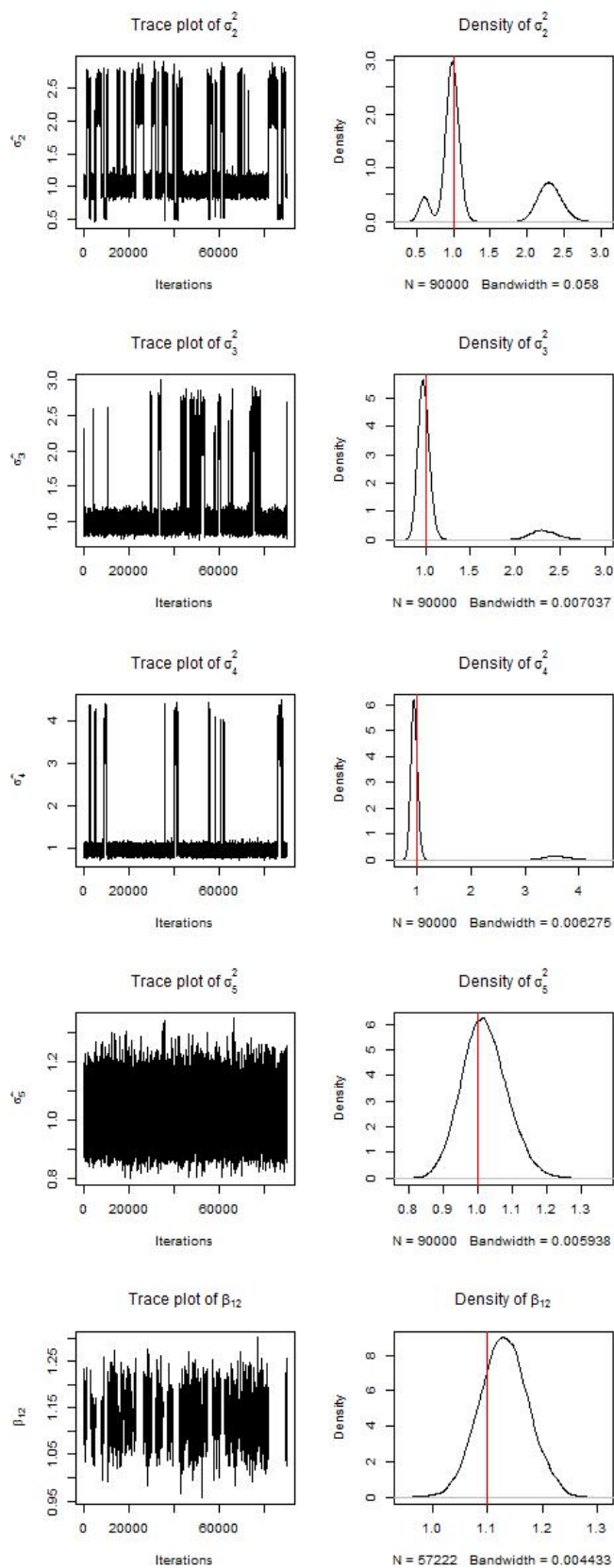
Figure D.3: *Trace plots and densities for parameters* σ_2^2, σ_3^2, σ_4^2, σ_5^2 *and* β_{12}.
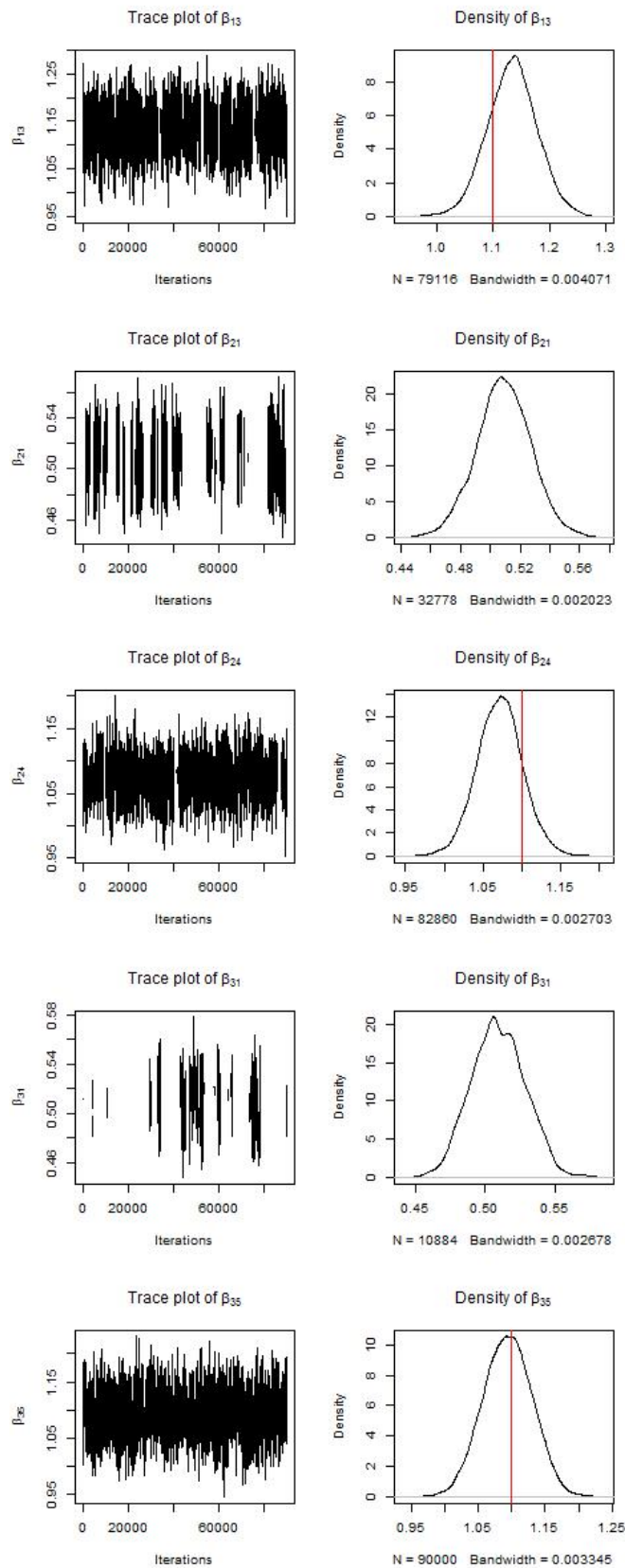
Figure D.4: *Trace plots and densities for parameters β_{13}, β_{21}, β_{24}, β_{31}, and β_{35}.*
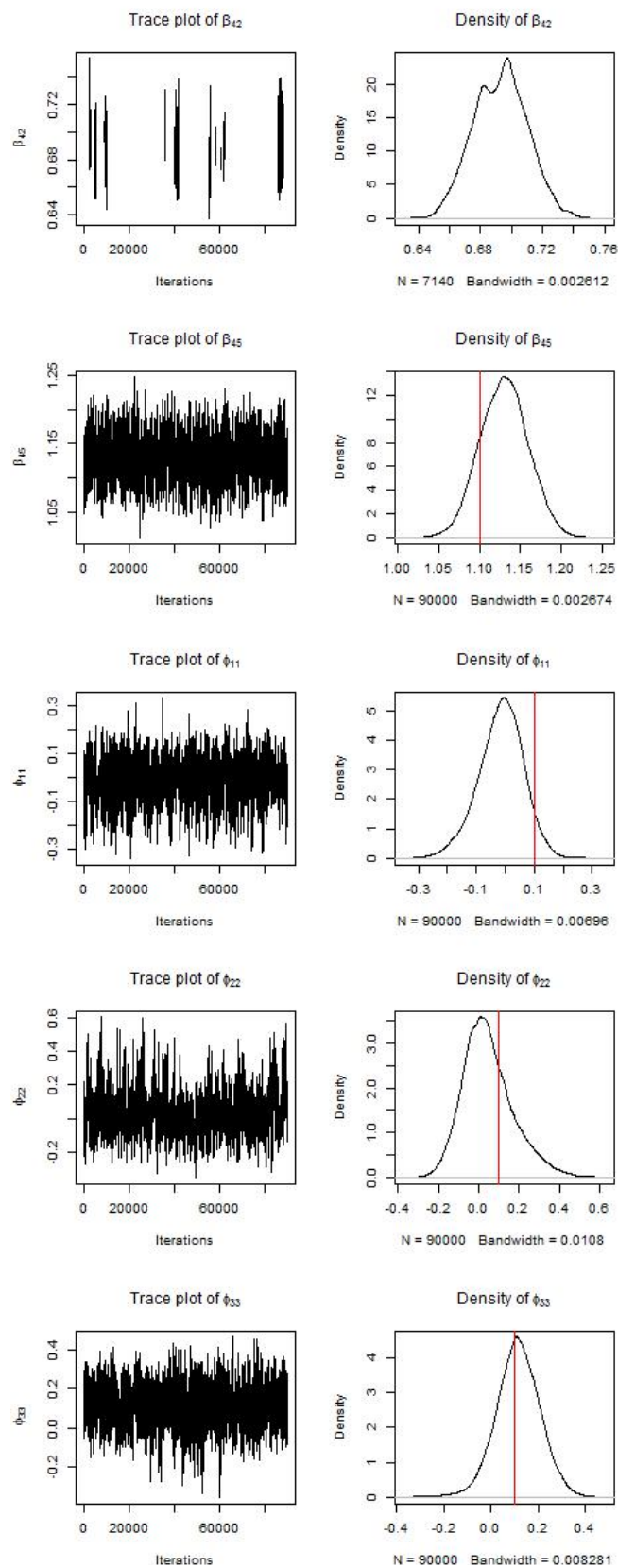
Figure D.5: *Trace plots and densities for parameters β_{42}, β_{45}, ϕ_{11}, ϕ_{22}, and ϕ_{33}.*
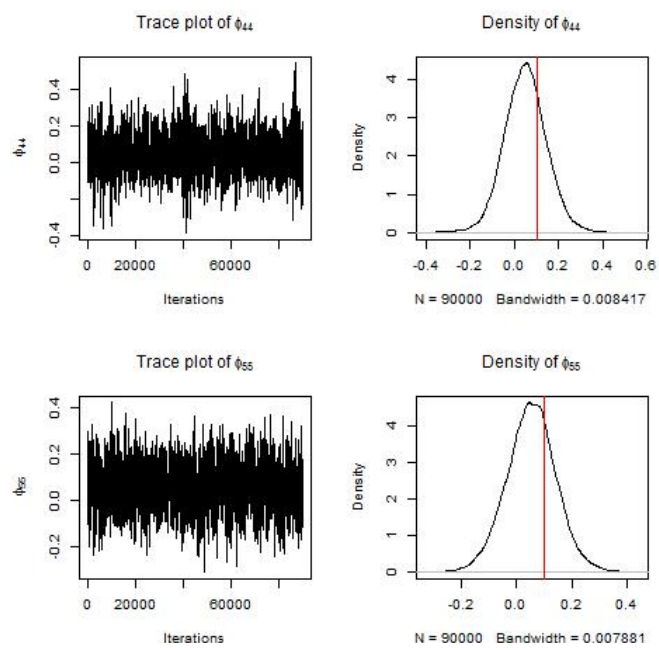
Figure D.6: *Trace plots and densities for parameters ϕ_{44} and ϕ_{55}.*

Appendix E

Simulation II results: for varying λ

This Appendix supplements the simulation study in Chapter 5, demonstrating how the prior on the number of edges can be used to penalize models with too many edges. These tables contain the estimates used to plot Figures E.1 and E.2.

Table E.1: *Tabulated results for simulated data sets 5.1 – 5.3, for $\lambda = 0, \cdots, 10$. The mean number of edges in the visited models is denoted $E[E_M]$, the number of models visited by the chain is denoted n_{models}, and the joint posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

λ	5.1			5.2			5.3		
	$E[E_M]$	n_{models}	$p(M_{TRUE}|D)$	$E[E_M]$	n_{models}	$p(M_{TRUE}|D)$	$E[E_M]$	n_{models}	$p(M_{TRUE}|D)$
0	5.01	509	0.0044	5.30	38	0.0000	5.29	63	0.4481
1	4.51	436	0.0025	5.12	31	0.0370	5.09	38	0.3821
2	3.91	295	0.0038	5.06	23	0.0478	5.04	25	0.3807
3	4.51	216	0.0002	5.01	13	0.0000	5.01	21	0.3859
4	2.96	170	0.0000	5.01	12	0.0000	5.01	17	0.3462
5	2.54	125	0.0000	5.00	8	0.0000	5.00	8	0.3435
6	2.15	97	0.0000	5.00	3	0.0000	5.00	5	0.5755
7	1.84	84	0.0000	5.00	4	0.0000	5.00	6	0.5140
8	1.37	54	0.0000	5.00	4	0.0536	5.00	4	0.3710
9	0.92	35	0.0000	5.00	4	0.0536	5.00	5	0.3623
10	0.50	26	0.0000	5.00	4	0.0536	5.00	4	0.3671

Table E.2: *Tabulated results for simulated data sets 5.4 – 5.6, for $\lambda = 0, \cdots, 10$. The mean number of edges in the visited models is denoted $E[E_M]$, the number of models visited by the chain is denoted n_{models}, and the joint posterior probability of the true model is denoted $p(M_{TRUE}|D)$.*

	5.4			5.5			5.6		
λ	$E[E_M]$	n_{models}	$p(M_{TRUE}|D)$	$E[E_M]$	n_{models}	$p(M_{TRUE}|D)$	$E[E_M]$	n_{models}	$p(M_{TRUE}|D)$
0	5.15	49	0.7239	5.40	31	0.6486	5.12	19	0.8789
1	5.04	33	0.8282	5.16	20	0.8486	5.10	17	0.9023
2	5.02	26	0.8199	5.04	11	0.9642	5.03	12	0.9712
3	4.97	21	0.8434	5.02	8	0.9841	5.01	7	0.9903
4	4.92	26	0.7280	5.00	8	0.9962	5.00	6	0.9968
5	4.82	25	0.6736	5.00	5	0.9961	5.00	4	0.9994
6	4.59	32	0.5083	5.00	5	0.9983	5.00	3	0.9994
7	4.19	32	0.2597	5.00	1	1.0000	5.00	2	0.9994
8	3.79	35	0.1130	5.00	2	0.9984	5.00	1	1.0000
9	3.35	45	0.0185	5.00	1	1.0000	5.00	1	1.0000
10	2.79	58	0.0031	5.00	1	1.0000	5.00	1	1.0000

Appendix F

Technical Appendix

The extension to the traditional trans-dimensional reversible jump algorithm as detailed in Chapters 4 and 5 is currently implemented in R (Version 3.2.0; see R Core Team 2014). The analysis on our simulated data in Chapters 4 and 5 (with data on 5 QTL and 5 traits for 500 individuals) generally took around 52 hours to complete 1,000,000 iterations on a 64 bit Intel(R) Core(TM) i5-4440 Quad 3.1GHz machine with 8GB RAM, running up to four chains in parallel.

Almost all of the code was written from scratch, making use of base R functions such as *rnorm* and *lm* to obtain the least squares estimates where necessary (R Core Team 2014). The 'pscl' package was used so that the inverse gamma (*igamma*) function was defined, and the 'parallel' and 'snow' packages were used to run chains in parallel (Jackman 2015, R Core Team 2014, Tierney et al. 2013).

As the code was written from scratch, we note that the run time could be significantly improved by optimising the R code and converting functions to C. The contribution of this thesis is in the approach taken, not the efficiency of the R code. Increasing the size of the causal networks to be estimated greatly increases the amount of computing time required, requiring weeks to run long chains. This

is the result of coding in R, once code has been optimised and key functions are converted to C, we expect the computing time required for large causal network structures to drastically decrease.

A brief outline of the R code used in this thesis is given below, including a description of the code used to implement the RJMCMC algorithm, as well as the code used to generate the results.

F.1   The RJMCMC algorithm

In order to run the RJMCMC algorithm given in Chapters 4 and 5, we require a data set with N_T trait nodes and a known genetic architecture comprised of N_{qtl} QTL with N_ϕ QTL effects on our N_T traits.

Consider the example causal network structure given in Figure F.1, with three trait nodes and two QTL. The genetic architecture is assumed to be known, so we know that QTL 1 affects trait 1 (ϕ_{11}) and QTL 2 affects trait 2 (ϕ_{22}). We also know that with three trait nodes, there are six unique directed edges existing between traits (β_{12}, β_{13}, β_{21}, β_{23}, β_{31}, β_{32}).

In order to initialise the chain, we are required to input the following:

- A list of possible parameters; the trait-to-trait effects, trait intercepts, QTL effects and residual variances for each trait

```
par.order.names <- c("B.12","B.13","B.21","B.23","B.31","B.32",
    "B.01","B.02","B.03","phi.11","phi.22","sigma.1","sigma.2",
    "sigma.3")
```

- The initial model (where the chain starts) is specified by a vector specifying the inclusion or exclusion of parameters, corresponding to the list of possible parameters above. We could choose to randomly select a start the state
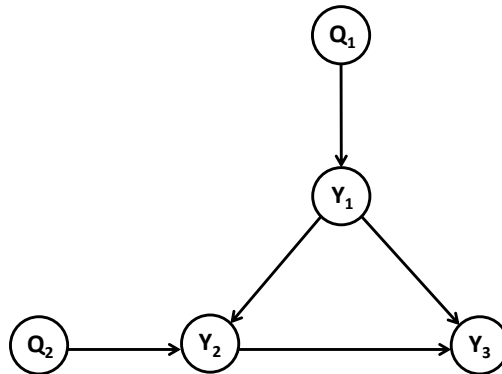
Figure F.1: *Example causal network structure to demonstrate the input and output of the R code.*

of each edge (reversed, absent, or exists) until an acyclic model is obtained, or we could start the chain at a specific model. See Section F.2 to see how a cyclic model is identified. Let us select an initial model with only one trait-to-trait effect, β_{12}, such that we have:

```
initial.model <- c(1,0,0,0,0,0,1,1,1,1,1,1,1,1)
```

- In order to initialise the chain we also need to specify the sample size ("nind"), the number of traits ("ntrait"), the number of QTL effects ("nqtl"), the number of iterations to run ("niter") and the thinning window we wish to use (thin). For example:

```
nind <- 500;  ntrait <- 5;  nqtl <- 5
niter <- 1000000;  thin <- 10
```

Once the chain has been initialised then a move type is selected randomly using the *runif* function and the move type probabilities given in Table 5.8 (p.

214). Given the selected move type, a candidate phenotype network structure is randomly selected from the neighbourhood of all possible acyclic models (given the selected move type).We use the *lm* function in R to obtain the LSE and from these we propose new candidate values and calculate the acceptance probabilities

$$\alpha = min(1, r)$$

where *r* is given in Equations 4.31 (p. 86), 4.34 (p. 90), 4.37 (p. 96), 5.14 (p. 155) and 5.17 (p. 161) for the update, reverse, relocate, add and remove steps, respectively. Note that if the double step is included, then we exclude the relocate step and use r_{double} given in Equation 5.25 (p. 220). If the move is accepted the candidate model becomes the current model, otherwise the current model remains unchanged.

F.2 Identifying cyclic models

Code was created (from scratch) to determine if a given phenotype network structure was acyclic. The algorithm iteratively follows each directed path from each trait node Y_t ($t = 1, \cdots, N_T$) and stops once all directed paths have been checked, or if a cycle is found (i.e. if it identifies a directed path back to Y_t). The R code is given below and it requires the vector cyc.mat — a causal network structure given in vector form (specifying the inclusion or exclusion of parameters). The R code is followed by an example to illustrate how the code works.

```
cyclic.model <- function(cyc.mat){
  CYCLE <- 0
  ## With ntrait trait nodes and nqtl QTL effects, a model
  ## requires at least 2 edges to create
```

```
## a directed path.
if(sum(cyc.mat)>=( 2*ntrait+nqtl + 2)){
## Get a list of edges in the model and determine which
## traits they originate from (from.list) and which traits
## they go to (to.list).
  edge.names <- cyc.mat[1:(ntrait^2-ntrait)]
  exist.edges <- par.order.names[1:(ntrait^2-ntrait)][
      edge.names==1]
  from.list <- 0
  to.list <- 0
  for(ee in 1:length(exist.edges)){
    edge <- strsplit(exist.edges[ee],split="")
    from.list <- c(from.list,as.numeric(edge[[1]][3]))
    to.list <- c(to.list,as.numeric(edge[[1]][4]))
  }
  from.list <- from.list[-1]
  to.list <- to.list[-1]

  ## For each trait with an edge originating from it,
  ## follow all directed paths and determine
  ## if there is a directed path back to that trait.
  ## CYCLE becomes 1 if a cycle is found.
  for(ll in 1:length(from.list)){
    from <- from.list[ll]
    start <- from
    to <- to.list[from.list==from]
    from <- to
    count <- 1
    while(CYCLE==0 && count<=length(from.list)){
      if(!is.na(sum(from))){
        to <- 0
```

```
      for(pp in 1:length(from)){

        to <- c(to,to.list[from.list==from[pp]])

      }

      to <- as.numeric(levels(as.factor(to)))

      if(length(to)>1){

        to <- to[2:length(to)]

      }

      from <- to

      for(pp in 1:length(to)){

        if(to[pp]==start){

          CYCLE <- 1

        }

      }

    }

    count <- count+1

  }

}


}
## return CYCLE, which is 1 is a cycle was found,
## and 0 otherwise.
return(CYCLE)
}
```

Let us consider the example phenotype network structure in Figure F.2. Starting with node Y_1, we identify the two directed paths given in Figure F.3 (a), but no cycle has been identified yet as there is no directed path back to Y_1. Moving to node Y_2, we identify the directed path given in Figure F.3 (b), and this time we have identified a cycle. The algorithm terminates and indicates this model is cyclic.
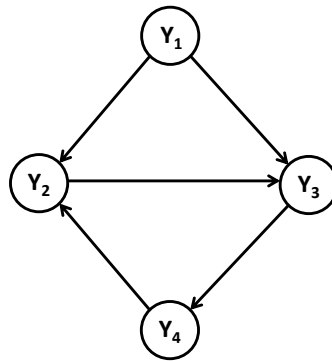
Figure F.2: *Example causal network structure to demonstrate the identification of a cyclic graph.*
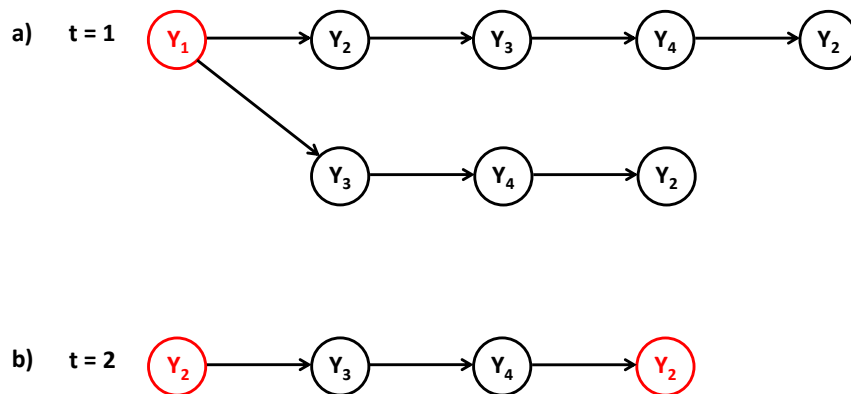

Figure F.3: *Using directed paths to identify cycles within a graph.*

Following from the example given above in Figure F.1 (with $N_T = 3$ trait nodes), an output file is created to store information regarding the (thinned) chain. The output file is set out as follows:

```
"step" "move" "Nedges" "B12" ... "s3" "llike" "lpost" "a" "acc"
"1" "initial" "1" "0.53" ... "1.27" "-3617.39" "-3663.25" "1" "Y"
"10" "reloc" "2" NA ... "1.05" "-3696.07" "-3741.06" "0.31" "N"
```

where the parameters β_{12}, β_{13}, β_{21}, β_{23}, β_{31}, β_{32}, β_{01}, β_{02}, β_{03}, σ_1^2, σ_2^2, and σ_3^2 are represented by

```
"B12" ... "s3"
```

to reduce space here. Recording the following information about the current state of the chain:

- Iteration number (step).

- Proposed move type (move). Note that the initial model has the move type recorded as "initial".

- Number of edges in the current model (Nedges).

- Current parameter estimates (corresponding to β_{12}, β_{13}, β_{21}, β_{23}, β_{31}, β_{32}, β_{01}, β_{02}, β_{03}, σ_1^2, σ_2^2, and σ_3^2) which are NA if the parameter is not in the current model.

- Log likelihood (llike).

- Log posterior density (lpost).

- The acceptance probability (a).

- Whether or not this was an accepted candidate model (acc = "Y") or if the candidate model was rejected and this is the previous state of the chain (acc = "N").

F.3   Presenting the results

The simulation results in this thesis focus on presenting the joint posterior probability of the models visited by the chain, the marginal posterior probabilities of the trait-to-trait effects (directed edges), and the comparison of the log posterior values. Continuing with the example above, we identify the different phenotype network structures visited by the chain using the following R code:

```
## Load the results file for a particular chain.
data <- read.table(file.choose(),header=T)
## discard a burn in of 100,000.
data <- data[data$step>100000,]


## Trait-to-trait effect parameters excluded from the estimated
## model will be recorded as NA. Create a matrix, B, to specify
## which parameters were present/absent in the model.
## 1 = present and 0 = absent.
B <- data[,4:9]
B[!is.na(B)] <- 1
B[is.na(B)] <- 0


## Determine how many different models were visited by the chain
## by combining the presence/absence indicators for each of the
## trait-to-trait effect parameters. Each different sequence of 0s
## and 1s specifies a different phenotype network structure.
models <- rep(NA,nrow(B))
for(i in 1:nrow(B)){
  models[i] <- paste(B[i,1], B[i,2], B[i,3], B[i,4], B[i,5],
      B[i,6], sep="")
}
```

The joint posterior probability for each model, here called 'model.jp, is then:

```
## Record the models visited.
models.factor <- levels(as.factor(models))
## Determine the joint posterior probability of each model visited.
model.jp <- rep(NA,length(models.factor))
for(i in 1:length(models.factor)){
  model.jp[i] <- sum(models==models.factor[i])/length(models)
}
```

To calculate the marginal posterior probability for each directed edge, find the models incorporating that edge, and sum up their joint posterior probabilities. The resulting marginal posterior probabilities for each directed edge can be presented in a bar chart, or graphically in the context of the phenotype network structure.

```
## List trait-to-trait effects.
trait.effects <- c("B.12","B.13","B.21","B.23","B.31","B.32")
## Record joint posterior probabilities.
edges.jp <- rep(0,length(trait.effects))
for(i in 1:length(trait.effects)){
  ## List models in which the edge exists.
  exists <- levels(as.factor(models[!is.na(eval(parse(
    text=sprintf("data$%s",trait.effects[i]))))]))
  ## If the edge was included in at least one of the models
  ## visited, add the joint posterior probabilities.
  if(length(exists)>0){
    for(j in 1:length(exists)){
      edges.jp[i] <- edges.jp[i] +model.jp[models.factor==exists[j]]
    }
  }
}


## Plot the marginal posterior probabilities in a bar chart.
par(mfrow=c(1,1))
```

```
plot.new()
## Use an asterisk to identify directed edges in the true model.
## Here we assume that the true model has three edges: B.12,
## B.13 and B.23.
ejp.names <- c("B.12*","B.13*","B.21","B.23*","B.31","B.32")
names(edges.jp) <- ejp.names
barplot(edges.jp,cex.names=0.45,names=ejp.names,xlab="edges",
    col=c("black","black","grey","black","grey","grey"),
    ylab=expression('p(e'[g]*'|D)'),ylim=c(0,1.0))
dev.copy(pdf,file=" edgemarginalpost_bar.pdf)
dev.off()


## Plot the marginal posterior probabilities in the context
## of a phenotype network structure.
par(mfrow=c(1,1))
plot.new()
## Plot trait nodes:
text(.4, .98, substitute(Y[1]),cex=1.5)
text(.1, .65, substitute(Y[2]),cex=1.5)
text(.7, .65, substitute(Y[3]),cex=1.5)
points(x =.4, y =.98, cex = 6, lwd=2)
points(x =.1, y =.65, cex = 6, lwd=2)
points(x =.7, y =.65, cex = 6, lwd=2)
## Plot the directed edges, where the edge thickness is
## dependent upon the marginal posterior probability.


## Determine the line thickness for each directed edge.
thick <- rep(0.1,length(edges.jp))
thick[(edges.jp[1] >=0.2)&&(edges.jp[1]<0.4)] <- 0.3
thick[(edges.jp[1] >=0.4)&&(edges.jp[1]<0.6)] <- 0.5
thick[(edges.jp[1] >=0.6)&&(edges.jp[1]<0.8)] <- 0.7
```

```
thick[edges.jp>=0.8] <- 0.9


## Plot the directed edges.
arrows(.36,.94,.14,.69,angle=7.5,lwd=(7*thick[1]))  # B.12
arrows(.45,.96,.67,.71,angle=7.5,lwd=(7*thick[2]))  # B.13
arrows(.13,.71,.35,.96,angle=7.5,lwd=(7*thick[3]))  # B.21
arrows(.16,.66,.64,.66,angle=7.5,lwd=(7*thick[4]))  # B.23
arrows(.66,.69,.44,.94,angle=7.5,lwd=(7*thick[5]))  # B.31
arrows(.64,.64,.16,.64,angle=7.5,lwd=(7*thick[6]))  # B.32
## Add a legend:
leg.txt <- c(expression('0.0 < p(e'[g]*'|D) < 0.2'), expression(
    '0.2 < p(e'[g]*'|D) < 0.4'),
    expression('0.4 < p(e'[g]*'|D) < 0.6'),
    expression('0.6 < p(e'[g]*'|D) < 0.8'),
    expression('0.8 < p(e'[g]*'|D) < 1.0'))
legend(0.675, 0.22, leg.txt, lwd=c(7*0.1,7*0.3,7*0.5,7*0.7,7*0.9),
    merge = TRUE)


dev.copy(pdf,file="edgemarginalpost_network.pdf")
dev.off()
```

The phenotype network structures with the highest joint posterior probabilities can be plotted in groups of up to four. The following R code plots the four phenotype network structures with the greatest joint posterior probabilities.

```
## Order the models according to the joint posterior probabilities.
models.order.prob <- model.jp[order(model.jp)][length(model.jp):1]
models.order.factor <- models.factor[order(model.jp)][
    length(model.jp):1]


par(mfrow=c(2,2))
for(i in 1:4){
```

```
## Select the model and identify the edges comprising it.
c.model <- models.order.factor[i]
c.model <- as.numeric(strsplit(c.model,"")[[1]])
pon.incl <- par.order.names[c.model==1]


## Plot the trait nodes.
plot.new()
text(.5, .92, substitute(Y[1]),cex=1.5)
text(.2, .65, substitute(Y[2]),cex=1.5)
text(.8, .65, substitute(Y[3]),cex=1.5)
points(x =.5, y =.92, cex = 6, lwd=2)
points(x =.2, y =.65, cex = 6, lwd=2)
points(x =.8, y =.65, cex = 6, lwd=2)


## Plot the edges comprising the selected model.
## B.12
if(sum(pon.incl=="B.12")==1){
  arrows(.42,.88,.26,.71,angle=7.5,lwd=2)
}
## B.21
if(sum(pon.incl=="B.21")==1){
  arrows(.26,.71,.42,.88,angle=7.5,lwd=2)
}
## B.13
if(sum(pon.incl=="B.13")==1){
  arrows(.58,.88,.74,.71,angle=7.5,lwd=2)
}
## B.31
if(sum(pon.incl=="B.31")==1){
  arrows(.74,.71,.58,.88,angle=7.5,lwd=2)
}
```

```
  ## B.32
  if(sum(pon.incl=="B.32")==1){
    arrows(.72,.64,.28,.64,angle=7.5,lwd=2)
  }
  ## B.23
  if(sum(pon.incl=="B.23")==1){
    arrows(.28,.64,.72,.64,angle=7.5,lwd=2)
  }
  ## Label the plot according to where the model joint
  ## posterior probability fals in the top 10.
  ## E.g. "Model 1" is the model with the greatest
  ## joint posterior probability.
  text(sprintf("Model %s: p = %s", as.character(i),
       as.character(round(models.order.prob[i],4))),x=.5,y=0)

}
dev.copy(pdf,file="top4models_graph.pdf")
dev.off()
```