(Waddell and Kishino, 2000; Toh and Horimoto, 2002; Stingo et al., 2010), to determine the dependence structure in each triplet. However, there are several differences between studying gene networks and the biological relationships for multiple platforms. (1) The purpose of network studies is to determine the relationship of a large number of features from one or two platforms in a single graph, whereas we are interested in investigating the relationships among one feature from each platform for multiple platforms. (2) In the network models mentioned previously, one high-dimensional graph is the typical "object" of estimation. Conversely, in our study, numerous low-dimensional graphs need to be investigated. (3) The aim of network approaches is to estimate the strength of the edges in one large-scale network, but the intent of our study is to determine the dependence structure among multiple platforms. For example, both Stingo et al. (2010) and our study example seek to integrate microRNA and mRNA expression. In the Stingo et al. (2010) study, 23 mouse microRNAs and 1,297 potential target genes were analyzed in one graphical model. That study determined the estimations of the big regulatory network for all these features under two different treatment conditions. However, our study is focused on investigating the dependence of each "microRNA–gene–clinical outcome" triplet. We investigate all the possible combinations to determine the microRNAs and genes that have relationships consistent with the real biological process, and we cast this as a model selection problem.

We apply an approach using the Bayesian GGM (Carvalho and Scott, 2009) to determine the biological relationships among different platforms for two reasons. First, based on this approach, the Bayes factor, which is used for model selection, has a closed-form expression. This greatly reduces computational time and cost in the real analysis. Second, our proposed Bayesian GGM approach utilizes an objective Bayesian prior for model selections; thus it eliminates the biases caused by choosing too strong or too vague priors (see the detailed explanations in Section 14.3).

14.3 Objective Bayesian Model Selection for GGM

The GGM (Dempster, 1972) is a class of models for measuring the dependency structure for multivariate normal distributions. In general, let $X = (x_1, \ldots, x_p)'$ be a p-dimensional normal random vector with mean μ and covariance matrix Σ. In our application, $p = 3$. For simplicity, we assume throughout that $\mu = 0$. Let $\Omega = \Sigma^{-1}$ denote the inverse covariance, also known as the concentration matrix with elements (ω_{ij}). Then the partial correlation between x_i and x_j given all the other variables is $\rho_{ij} = -\omega_{ij}/\sqrt{\omega_{ii}\omega_{jj}}$. Thus $\omega_{ij} = 0$ if and only if x_i and x_j are conditionally independent given all other variables.

The GGM can be represented by an undirected graph $G = (V, E)$, where V is a set of vertices representing the variables and E is a set of undirected edges indicating the relationships among the variables. The graph represents the model by drawing an edge between vertices i and j when $\omega_{ij} \neq 0$. Complete graphs are defined as graphs having $(i, j) \in E$ for every $i, j \in V$. A graph C is called a clique if it is a maximal complete subgraph. A graph S is called a separator if it is the overlap of two cliques. We denote the sets of cliques and separators of a graph by $\mathcal{C}$ and $\mathcal{S}$, respectively. A graph $G = (V, E)$ is called a decomposable graph if either G is complete, or we can reexpress $V = A \cup S \cup B$, where S is complete and separates A and B and both $A \cup S$ and $B \cup S$ are decomposable. The eight models we introduced in Figure 14.2 are all decomposable graphs according to the definition.

We can reformulate our research goal as a model selection problem in the GGM. There are various approaches to select models in the GGM. The traditional approaches include Whittaker (1990) for small GGMs and Bayesian shrinkage approaches for large-scale gene networks (Dobra et al., 2004; Schafer and Strimmer, 2005; Opgen-Rhein and Strimmer, 2007). Here, we follow a Bayesian approach to solve this problem.

Suppose we observe n samples $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ of p-dimensional vectors from an unknown decomposable graph G, where each $\boldsymbol{x}_i \sim N(0, \Sigma)$, with unknown covariance matrix Σ. Let X be the $n \times p$ matrix of observed data. The posterior probability of the graph G given X can be expressed as

$$p(G|X) \propto p(G) \int p(X|\Sigma, G) p(\Sigma|G) d\Sigma, \tag{14.1}$$

where $p(G)$ is the prior probability of the graph G, and $p(\Sigma|G)$ is the prior for Σ given G.

The expression in Equation (14.1) involves an integral over the prior for Σ under the graph G. One problem in model selections is that the integral is very sensitive to different choices of the prior (Berger and Pericchi, 2001; Jones et al., 2005). Unlike estimation problems, this sensitivity does not decrease as the number of observations increases. Therefore, making an intelligent choice of $p(\Sigma|G)$ for each graph G is very critical in Bayesian GGM selection. An appropriate prior should at least have the following properties: (1) $p(\Sigma|G)$ should be a conjugate prior for practical reasons, and (2) neither improper priors nor vague priors can be used.

The hyper-inverse Wishart (HIW) distribution was first introduced by Dawid and Lauritzen (1993). For a decomposable graph G with covariance matrix Σ, it can be expressed as $p(\Sigma|G) \sim \mathrm{HIW}_G(b, D)$, where b is a positive parameter

for degrees of freedom, and D is a symmetric positive-definite matrix for scale. One of the important properties of the HIW distribution is that it is a conjugate prior for the covariance matrix. Berger and Pericchi (2001) proposed to set the prior as $p(\Sigma|G) \sim \text{HIW}_G(\sigma, \tau I)$, where the scale matrix $D = \tau I$ is proportional to the identity matrix. For this class of priors, we need to choose the scale parameter close to the real scale of the data. A too small-scale will cause a lack of discrimination among the graphs we compare, and a too large-scale will overpower the likelihood.

Carvalho and Scott (2009) suggested another prior for Σ in Bayesian GGM selections, and $p(\Sigma|G) \sim \text{HIW}_G(gn, gX'X)$, where g is between 0 and 1. This prior is referred as an HIW g-prior similar to Zellner's g-prior in linear regression (Zellner, 1986). Direct use of the HIW g-prior seems inappropriate because it involves a double use of the data. However, Carvalho and Scott (2009) showed that this prior corresponds to the implied fractional prior for selecting a graph using fractional Bayes factors.

The fractional Bayes factor is a Bayesian model selection technique proposed by O'Hagan (1995) when prior information is weak. It is motivated by the partial Bayes factor, which uses part of the samples to train the noninformative priors as proper priors and then uses the remaining data to perform model comparisons. The parameter g in the HIW g-prior can be viewed as the fraction of the likelihood used for training the noninformative prior.

However, if we interpret the HIW g-prior as the fractional Bayes factor, we cannot put a hyper-prior on g because g is no longer a model parameter. Instead, it represents the fractional power of the likelihood that is used for training the noninformative prior. Intuitively, we want to save as much of the data as possible to choose between models. Hence we simply choose $g = 1/n$, which means that the training sample size is equal to 1.

In summary, there are several advantages to using this HIW g-prior. (1) The conjugate property of the prior produces closed-form Bayes factors, which makes the selection of millions of models become feasible. (2) Compared with using the conventional prior, whose results depend largely on the arbitrary choice of a constant, using the HIW g-prior, which corresponds to the implied fractional prior for $(\Sigma|G)$, automatically provides us with an objective Bayesian approach for the model selection. (3) The consistency of fractional Bayes factors as $n \to \infty$ has been shown by O'Hagan (1995). In addition, Carvalho and Scott (2009) showed that the Bayes factors based on the HIW g-prior is information consistent.

Let G_0 denote the null graph corresponding to model 0, which has no edges, and let G_i denote the graph corresponding to model i to be compared with the