Investigation of patient-sharing networks using a Bayesian network model selection approach for congruence class models

Ravi Goyal1 | Victor De Gruttola2

1Health Unit, Mathematica, Princeton, New Jersey
2Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

Correspondence
Ravi Goyal, Health Unit, Mathematica, P.O. Box 2393, Princeton, NJ 08543-2393, USA.
Email: rgoyal@mathematica-mpr.com

Funding information
Division of Intramural Research, National Institute of Allergy and Infectious Diseases, Grant/Award Numbers: R24 AI-106039, R37 AI-51164

A Bayesian approach to conduct network model selection is presented for a general class of network models referred to as the congruence class models (CCMs). CCMs form a broad class that includes as special cases several common network models, such as the Erdős-Rényi-Gilbert model, stochastic block model, and many exponential random graph models. Due to the range of models that can be specified as CCMs, our proposed method is better able to select models consistent with generative mechanisms associated with observed networks than are current approaches. In addition, our approach allows for incorporation of prior information. We illustrate the use of this approach to select among several different proposed mechanisms for the structure of patient-sharing networks; such networks have been found to be associated with the cost and quality of medical care. We found evidence in support of heterogeneity in sociality but not selective mixing by provider type or degree.

KEYWORDS
Bayesian, network model selection, patient-sharing networks

1 INTRODUCTION

There is a growing body of research that leverages administrative claims data to identify connections among medical providers; a link between two providers arises when both treat the same patient as indicated by medical claims. The collection of these connections constitute patient-sharing networks.1,2 Such connections have been shown to imply professional relationships among providers.3,4 Investigating the relationship between such networks and patient health outcomes is an emerging area of research; see DuGoff et al for a systematic review of the literature.2 The most common theoretical explanation for this association is that the networks represent aspects of collaboration, continuity, and care coordination;3,5 these aspects may be especially important for patients with multiple chronic conditions, who account for a high percentage of health care costs.6 Health outcomes that have been studied in this regard include cost,7-16 utilization,15,17 patient-reported outcomes,18,19 quality of care,4,20 and mortality.14,20 Even as such evidence grows, there remain gaps regarding how to make use of knowledge about the patient-sharing networks to improve health care. In particular, neither the evolution of patient-sharing networks in response to different incentives nor their best use for the development or evaluation of interventions is well understood.2 Addressing these gaps requires knowledge of the generative mechanisms governing the evolution of the patient-sharing network.

We present a Bayesian model selection approach to evaluate the evidence supporting hypotheses that observed networks were generated by specified mechanisms. Bayesian model selection identifies the model with the highest posterior
probability among a set of candidate models, which we denote as \( \{ m_1, \ldots, m_k \} \); each model is associated with a specific mechanism. In particular, we specify models associated with heterogeneity in sociality and with selective mixing; details are provided in Section 4.

In order to present the posterior probability for each model, we introduce some terminology. Let \( I_{m_h} \) be an indicator variable that \( m_h \) is the correct model for the observed network. Equation (1) shows the posterior probability for model \( m_h \) given an observed network \( g \):

\[
P(I_{m_h} = 1 | g) = \frac{p(g | I_{m_h} = 1) p(I_{m_h} = 1)}{\sum_{j=1}^{k} p(g | I_{m_j} = 1) p(I_{m_j} = 1)},
\]

where \( p(g | I_{m_h} = 1) \) and \( p(I_{m_h} = 1) \) are the model evidence (shown below) and prior probability, respectively, for model \( m_h \). Equation (2) shows the model evidence for model \( m_h \):

\[
P(g | I_{m_h} = 1) = \int_{\theta_h} p(g | \theta_h, I_{m_h} = 1) p(\theta_h | I_{m_h} = 1) d\theta_h.
\]

where \( \theta_h \) are the parameters for model \( m_h \), \( p(g | \theta_h, I_{m_h} = 1) \) is the likelihood, and \( p(\theta_h | I_{m_h} = 1) \) is the prior distribution of the parameters of model \( m_h \).

As discussed by Holder and Lewis, Bayesian approaches to model selection differ from maximum likelihood (ML) approaches in the way in which they treat parameters (\( \theta \)) in the model.\(^{21}\) These parameters are typically not of primary interest when selecting a model, but allow for calculation of the probability of the data (D) given a model (M), that is, when computing \( P(D | M, \theta) \). ML approaches identify the parameter values that maximizes the likelihood. In contrast, Bayesian approaches calculate the volume under the posterior probability surface by integrating out the model parameters to evaluate the model evidence; the equation for the model evidence in our setting is provided in Equation (2). See Holder and Lewis for further explanation and illustrative examples of the distinction between the approaches.\(^{21}\)

Bayesian network model selection has shown to be challenging because of the computational burden of estimating the likelihood—and therefore the posterior probability—for each candidate network model.\(^{22}\) There have been recent advances in Bayesian model selection for a class of network models called exponential random graph models (ERGMs)—one of the most commonly used network models in social network analysis. Caimo and Friel developed a method based on an extension of their reversible jump Markov chain Monte Carlo algorithm that estimates the posterior probability for each model.\(^{22,23}\) Thiemichen et al proposed a method that applies a Laplace approximation to estimate the Bayes factor, which allows for model selection in the Bayesian paradigm.\(^{24}\) However, these approaches for ERGMs selection have a high complexity cost. In order to reduce the computational burden, Bouranis et al proposed alternative approaches to Bayesian inference for ERGMs based on adjusting the pseudo-posterior distribution or pseudo-likelihood, respectively.\(^{25,26}\) However, they note that implementing procedures for approximating a solution to the likelihood is more challenging for larger datasets.\(^{26}\) Even if these computational issues can be addressed, ERGMs are nonetheless severely limited in the types of generative mechanisms that can be modeled.\(^{27}\) The primary contribution of our approach is to develop a Bayesian model selection approach for networks that broadens the range of potential mechanisms whose role in generating an observed network can be investigated.

We base our approach to estimating posterior probabilities on a class of network models denoted congruence class models (CCMs);\(^{28}\) this class includes as special cases several common network models, such as the Erdős-Rényi-Gilbert (ER) model, stochastic block (SB) model, and ERGMs. See Goldenberg et al for a review of these models.\(^{29}\) The ability of CCMs to generalize such a broad set of network models arises from its flexibility in specifying the probability distribution associated with network properties included in a model. This flexibility enables investigators to develop network models that more closely correspond to potential network generation mechanisms than do current classes of network models. In this regard, it allows investigators to make better use of the totality of their knowledge regarding the generative mechanism. Computing the likelihood of a CCM requires the enumeration the graphs associated with the observed values of network properties—a problem that is referred to as graph enumeration. In this manuscript, we use a general recursive formula to estimate the number of labeled graphs for given values of network properties.

The next section provides background details on CCMs, their relationship to other network models, and the general graph enumeration recursive formula. Section 3 provides a simulation study demonstrating the proposed Bayesian model selection approach. Section 4 introduces patient-sharing networks and a description of several competing network models representing distinct generative mechanisms. In addition, the evaluation of the models and results
of the model selection approach is presented in Section 4. This article concludes with a discussion and further research.

2 | CONGRUENCE CLASS MODELS FOR NETWORKS

2.1 | Overview

We denote a network as \( g = (V_g, E_g) \), where \( V_g \) and \( E_g \) are the vertex and edge sets of \( g \). Let \( n \) represent the number of vertices in \( g \) and \( |Z| \) denote the size of set \( Z \); therefore, \( n = |V_g| \). We represent a network \( g \) as an adjacency matrix. Let \( g[i, j] = 1 \) indicate that there is an edge between \( i \) and \( j \), where \( i, j \in V_g \), whereas \( g[i, j] = 0 \) indicates that there is no edge.

Let \( \mathcal{G}_n \) denote the space of all potential networks with \( n \) vertices. Let \( \phi \) denote an algebraic map from \( \mathcal{G}_n \) to network summary statistics (eg, degree distribution or degree mixing) and let \( c_\phi(x) = \{ g : \phi(g) = x, g \in \mathcal{G}_n \} \) denote the inverse image associated with \( \phi \). We refer to these inverse images of singleton sets as congruence classes;\(^{28}\) they also have been referred to as fibers in algebraic statistics literature.\(^{30}\) We use \( \phi \) to represent the mapping as well as the associated network property being calculated. Let \( |c_\phi(x)| \) denote the number of graphs for which network property \( \phi \) equals \( x \); this quantity has been referred to as a volume factor.\(^{31}\)

The probability distribution on \( \mathcal{G}_n \) for the CCM is specified by the probability mass function (PMF) on the congruence classes defined by \( \phi \); we denote this PMF as \( P_\phi \). \( P_\phi(x|\theta) \) denotes the total probability of all networks that are elements in \( c_\phi(x) \) given \( \theta \), that is,

\[
P_\phi(x|\theta) = \sum_{g \in c_\phi(x)} P_{\mathcal{G}_n}(g|\theta),
\]

where \( P_{\mathcal{G}_n}(g|\theta) \) is the probability of graph \( g \). CCMs assume that two networks within a congruence class have the same probability of being observed; this assumption is also made in common network models including the ER model, SB model, and ERGM. This assumption leads to the following probability distribution on \( \mathcal{G}_n \) for a CCM:

\[
P_{\mathcal{G}_n}(g|\theta) = \left( \frac{1}{|c_\phi(\phi(g))|} \right) P_\phi(\phi(g)|\theta).
\]

CCMs can represent several common network models because of their flexibility in the specification of \( P_\phi \). To illustrate, we consider as an example the specification of a probability distribution identical to that of the ER model. We let \( \phi_1 \) be the mapping from a network \( g \) to its number of edges, that is, \( |E_g| \), and let \( P_{\phi_1}(\phi_1(g)|\theta = p) \) equal the following:

\[
P_{\phi_1}(\phi_1(g)|\theta = p) = p^{\phi_1(g)}(1 - p)^{\tbinom{n}{2} - \phi_1(g)},
\]

where \( \phi_1(g) = \tbinom{n}{2} \). The ER model, which assumes each edge forms with a fixed probability and its formation is independent of all other edges, is one specification of a model that includes only the number of edges as a network property. However other network generative mechanisms can result in a different distribution for the number of edges, but still be fully categorized by a distribution that only includes the number of edges. For instance, one can draw the number of edges from a uniform distribution and then select a network at random with the selected number; the \( P_{\phi_1}(\phi_1(g)) \) associated with is mechanism is shown below:

\[
P_{\phi_1}(\phi_1(g)) = \left( \frac{n(n - 1)}{2} + 1 \right)^{-1}.
\]

Another such process is described by Kretzschmar and Morris (1996),\(^{32,33}\) which can be modeled as a CCM.\(^{34}\) As shown, CCMs can be used to specify various distributions for the same network property—in this case number of edges; Section 3 provides examples based on degree distribution.

For ERGMs, networks in the same class have the same probability,\(^{30}\) hence there exists a \( P_\phi \) such that a CCM and an ERGM assign the same probability distribution on \( \mathcal{G}_n \). However, CCMs provide additional flexibility in modeling \( P_{\mathcal{G}_n} \) compared to ERGMs. ERGMs require \( P_\phi \) to have a specific functional form, whereas CCMs do not place restrictions on this
form. For instance, ERGMs require that the number of edges follow a binomial distribution when it is the sole network property; therefore, the distribution specified in Equation (6) is not feasible within the ERGM framework.

For clarity of presentation and computation, we selected network models that focus on a particular network property. However, a CCM can be defined based on several network properties. For instance, Goyal et al and Goyal and De Gruttola specified a CCM with degree distribution and mixing patterns.28,35

2.2 Bayesian model selection with CCMs

The model evidence for a CCM can be derived by substituting the CCM likelihood (Equation (4)) into the general equation for model evidence (Equation (2)) as shown in Equation (7):

\[ P(g|I_{m_0} = 1) = \int_{\theta_h} \left( \frac{1}{|c_{\phi_h}(\phi_h(g))|} \right) P_{\phi_h}(\phi_h(g)|\theta_h) \ast p(\theta_h|I_{m_0} = 1) d\theta_h. \]  

(7)

The volume factor is not dependent on \( \theta_h \) and, therefore can be brought outside of the integral as shown below in Equation (8):

\[ P(g|I_{m_0} = 1) = \left( \frac{1}{|c_{\phi_h}(\phi_h(g))|} \right) \int_{\theta_h} P_{\phi_h}(\phi_h(g)|\theta_h) \ast p(\theta_h|I_{m_0} = 1) d\theta_h. \]  

(8)

This integral can be computed using standard approaches,36 however, our computation makes use of a recursive formula to estimate the volume factor.

2.3 Graph enumeration

Equation (9) provides a recursive formula to estimate the number of graphs, \( |c_{\phi}(x_k)| \), with specific value(s), \( x_k \), for particular network properties, \( \phi \):

\[ |c_{\phi}(x_k)| = r_{\phi}(x_k, x_{k-1}) \ast |c_{\phi}(x_{k-1})|, \]  

(9)

where \( r_{\phi}(x_k, x_{k-1}) \) is the ratio between the sizes of congruence class \( c_{\phi}(x_k) \) and \( c_{\phi}(x_{k-1}) \), that is,

\[ r_{\phi}(x_k, x_{k-1}) = \frac{|c_{\phi}(x_k)|}{|c_{\phi}(x_{k-1})|}. \]  

(10)

Goyal et al presents methods for evaluation of the ratio, Equation (10), for a range of network properties of interest to social network analysis, including number of edges, mixing by nodal covariates, degree distribution, degree mixing, and number of triangles.28 An algorithm to compute the recursive formula and validation of the algorithm is presented in Goyal and De Gruttola.37

3 SIMULATION STUDY

We conducted a simulation study to assess the performance of the proposed Bayesian model selection approach to select a network model—among a set of candidate models—that generated a given network, \( g \). For this study, we generated \( g \) from one of three network models, denoted as \( m_1^1, m_2^1, \) and \( m_3^1 \). In order to specify \( m_1^1, m_2^1, \) and \( m_3^1 \), we introduce some notation. Let the degree of vertex \( i \), denoted as \( d(i) \), be the number of edges between that vertex and others; hence, \( d(i) = \sum g(i, j) \). Let \( d(g) = (d_1(g), \ldots, d_n(g)) \) represent the vector of degrees of nodes in set \( V_g \), commonly referred to as a degree sequence. The degree distribution, denoted as \( D(g) \), is a vector such that the \( k^{th} \) entry represents the number of vertices having degree \( k \), that is, \( D_k(g) = \sum_{i=1}^{n} I_{(d(i)=k)} \). Let \( \phi_2 \) be the mapping from a network \( g \) to its degree distribution, that is, \( \phi_2(g) = D(g) \).
Selecting among three models that all focus on degree distribution illustrates the flexibility of the proposed approach; as we show below, CCMs can be used to specify various probability distributions for the degree distribution. By contrast, ERGMs allow for inclusion of terms for degree distribution (such as k-stars or altering stars), but not for specification of their functional form. Therefore, ERGMs have only limited ability to assess whether observed network data support a power-law or an exponential distribution; one could only conclude that the observed network is—or is not—consistent with a model that does includes or does not include terms associated with degree distribution. CCMs allow investigators to model not only traditional degree distributions, such as exponential, power-law, and binomial, but also more complex distributions, such as those that are bi- or tri-modal. One can actually specify a different probability for each possible congruence class without making parametric assumptions. The ability to investigate the functional form of the degree distribution has important implications; for example, in the investigation of the spread of disease. In the setting of sexually transmitted infections, the degree distribution of a sexual contact network—the set of sexual relationships within a population—can have profound effects on both the spread of infection and the effectiveness of control programs.\(^{38}\)

The choice of appropriate probability distribution associated with degree distribution depends on the specific mechanisms hypothesized to generate the network. For example, a common feature in social systems is the concentration of influence among a few individuals that arises through mechanisms that encourage preferential attachment—the tendency of connections to form between nodes with probabilities proportional to the number of connections the nodes already have (ie, degree).\(^{39,40}\) This mechanism generates networks with degrees following a fat-tailed distribution—specifically a power-law distribution. Many real-world networks can be modeled using this mechanism, which leads to a network wherein many nodes have a moderate number of edges and fewer nodes have a large number. Other mechanisms may result in different distributions. For example, mechanisms based on the non-equilibrium theory can result in an exponential distribution for degree.\(^{41}\) Each model, \(m_1^\text{ss}\), \(m_2^\text{ss}\), and \(m_3^\text{ss}\), generates networks following a specified degree distribution. Details are provided below for each of the three models.

### 3.1 Candidate models

Model \(m_1^\text{ss}\) generates networks such that the degrees follow a power-law distribution, that is, the probability that a vertex in the graph has degree \(k\) decays as a power-law, that is, \(P(k) \sim k^{-\gamma}\), where \(P(k)\) is the probability that a vertex has degree \(k\) and \(\gamma \in \{1, \ldots, \infty\}\). We generated networks from \(m_1\) using a two-step process. First, we used the Barabasi-Albert (BA) model to generate an initial network with a power-law degree distribution.\(^{40}\) The BA model generation process can be initiated with a small seed graph that grows by the addition of new vertices one at a time. Each new vertex forms a new edge with an existing vertex based on preferential attachment rules. Vertices and edges, once introduced, are never deleted. Second, we used an edge swapping procedure to generate networks at random with the same degree distribution.

Model \(m_2^\text{ss}\) generates networks in which degrees follow an exponential distribution with rate parameter \(\lambda\) that provides the same mean degree as \(m_1^\text{ss}\). In order to generate networks under \(m_2^\text{ss}\), we first sampled degrees from the exponential distribution and then used the configuration model to construct a network based on the sampled degree sequence.\(^{38}\) The configuration model is a method for generating random networks from a given degree sequence. The model first assigns a degree to each vertex based on a given degree sequence. The degrees of the vertices are represented as half-links or stubs. Next, the configuration model chooses two stubs uniformly at random and connects them to form an edge. The model continues selecting pairs of stubs at random connecting them to form edges until all of the stubs have been joined. The result is a network with the pre-defined degree sequence. The realization of the network might potentially include self-loops or multi-links, which are removed as they arise.

Model \(m_3^\text{ss}\) is the Erdos-Renyi (ER) model, wherein networks are generated such that degrees follow a binomial distribution. This is done by selecting \(m\) edges at random; we set \(m\) so that networks generated from \(m_3^\text{ss}\) have the same number of edges as \(m_1^\text{ss}\). In the ER model, all graphs on a fixed vertex set with a fixed number of edges are equally likely.

For each of these models, if two networks \(g\) and \(h\) have the same degree distribution, that is, \(D(g) = D(h)\), then the probability of generating two networks is identical, that is, \(P(g|I_m^i = 1) = P(h|I_m^i = 1)\) for \(i \in \{1, 2, 3\}\). Therefore, each of these three models can be represented as a CCM where \(P_{I_m^i}(h^i(g)|\theta)\) equals a power-law distribution for \(m_1^\text{ss}\), an exponential distribution for \(m_2^\text{ss}\), and a binomial distribution for \(m_3^\text{ss}\).

### 3.2 Procedure

Given a network \(g\), the Bayesian model selection approach chooses one of the three models \((m_1^\text{ss}, m_2^\text{ss}, m_3^\text{ss})\) based on which model has the highest posterior probability. Therefore, for each network, we must calculate \(P(I_{m_i^i} = 1|g)\), where \(i\) ranges
from 1 to 3; that is, we need to calculate Equation (1). In order to calculate $P(I_{m_i} = 1 | g)$, we need to select a prior distribution for the model parameter(s). Each of the models, $m^i_1$, $m^i_2$, and $m^i_3$, has a single parameter describing its degree distribution. For $m^i_1$, it is the power-law parameter $\gamma$; for $m^i_2$ and $m^i_3$, the parameters are the rate parameter $\lambda$ and the binomial parameter $p$, respectively.

We created a prior distribution for each of these parameters by sampling 50 networks from model $m^i_1$. Using the degree sequence for these 50 networks, we estimated $\gamma$ by maximum likelihood. We assume the prior distribution of $\gamma$ follows a normal distribution with mean equal to its maximum likelihood estimate (MLE) and variance equal to the estimated variance of the MLE. The prior distributions for $\lambda$ and $p$ are created such that the means were equal to the mean of the 50 networks, and the variances were scaled so as to be comparable to the prior distribution for $\gamma$. We do not adjust the model evidence to account for the fact that not all degree sequences are graphical as the PMF on the congruence classes should only be defined on graphical classes.\(^{42}\)

### 3.3 Results

We present results for the Bayesian model selection approach for networks of sizes 25, 50, 100, 1000, and 5000. For each network size, we generated 300 networks, 100 from each $m^i_1$, $m^i_2$, and $m^i_3$. For each of the 300 networks, we conducted the proposed Bayesian model selection approach to select a network model—among a set of candidate models ($m^i_1$, $m^i_2$, and $m^i_3$)—that generated the network.

Table 1 presents the results of the simulation study. The first column indicates the size of the network (25, 50, 100, 1000, or 5000). The second column indicates the distribution under the network was generated (exponential, power-law, or binomial distribution). The third to fifth columns provide the mean of the posterior distribution (over 100 simulations) that the network was generated from a model with degrees following exponential (column 3), power-law (column 4), or binomial (column 5) distribution. The sixth column indicates the percent of times that the proposed Bayesian model selection approach was correct.

For networks of size $n = 5000$, our approach selected the correct model for each of the 300 graphs. In addition, the posterior distribution was near 1 for the correct model (and near 0 for the incorrect models). The same held for networks of sizes $n = 1000$ and $n = 100$. For $n = 50$, the Bayesian model selection approach correctly specified 99 of the 100 networks generated under the binomial distribution. For $n = 25$, the Bayesian model selection approach correctly specified 95 of the 100 networks generated under the exponential distribution and 99 of the 100 networks generated under the binomial distribution.
distribution. For the other specifications associated with \( n = 50 \) and \( n = 25 \), the approach selected the correct model. The average time to compute the model evidence for a single model for networks of sizes \( n = 25, 50, \) or 100 was 0.29 seconds; this time did not greatly vary with this range of size (0.30 second for \( n = 25 \) to 0.31 seconds for \( n = 100 \)). The time increased to 0.43 and 2.51 seconds for networks of sizes 1000 and 5000, respectively.

4 | PATIENT-SHARING NETWORKS

For each US state, we analyzed the patient-sharing network in order to investigate the generative mechanisms underlying this network. State networks include all resident providers who share Medicare patients. The next section (4.1) provides a description of the data used to generate the patient-sharing network for each state. The subsequent section (4.2) provides details of generative mechanisms we investigated and their associated CCM. Section 4.2.1 presents findings from the Bayesian model selection approach.

4.1 | Data

Our analysis uses two publicly available data sets from the Center for Medicare and Medicaid Services (CMS). The first identifies edges between providers; we define two providers as connected if a Medicare patient encounters both within a 30-day interval; our analysis investigates networks that arose in the first three quarters of 2015—the most recent publicly available data set. Therefore, two providers are connected if at any point during the first three quarters of 2015 there was a 30-day interval in which they both saw the same patient. CMS provides variations of this data set for which the time interval between encounters can be set at 30-, 60-, 90-, or 180-days. We used the 30-day interval, as it is the most conservative choice with regard to the implication that two providers are coordinating care for a patient; this choice is intended to reduce the number of “spurious” edges that arise when providers treat distinct ailments and do not need to coordinate care. Our analysis assumes that providers are connected if they share at least one patient; previous studies have used this threshold as well as others. In our analysis, connections are binary, that is, either present or not, and are undirected.

The second data set, called Medicare Provider Utilization and Payment Data for 2015, lists the geographic location and medical specialty of the provider. This information was used to filter our patient-sharing network based on the state the of residence of the provider as well as to label providers as either primary or specialty care.

4.2 | Network mechanisms and models

We investigated several CCMs, denoted as \( m_1 - m_4 \), that model mechanisms that may be responsible for the resultant patient-sharing network. In this section, we introduce the models, and in Section 4.2.1, present the results of selecting among the models. The first model (\( m_1 \)) is associated with the mechanism in which edges form at random and is presented as a null model. Model \( m_2 \) is associated with heterogeneity in sociality, whereas models \( m_3 \) and \( m_4 \) are associated with two different types of selective mixing.

For each candidate network model \( m_h \), the proposed Bayesian model selection approach requires specification of the likelihood for each model, \( p(g|\theta_h, I_m = 1) \), and the prior distribution of the parameters of the model, \( p(\theta_h|I_m = 1) \). Specifying the likelihood requires associating a probabilistic model for generating networks consistent with the network model \( m_h \). In the network science literature, many network models and mechanisms have been proposed. These mechanisms provide a basis for the generation of a range of types of networks, including sexual contact networks and gene regulatory networks. In Section 3, we described several network models, such as BA and ER, and mechanisms, such as preferential attachment. The flexibility of CCMs in specifying \( p(g|\theta_h) \) provides an opportunity to evaluate the evidence of a range of network models regarding whether they are consistent with our observed network. To do so requires translating potentially complex network models into probabilistic models—which can be challenging. Furthermore, a given network model can be associated with a probabilistic model that requires a large number of parameters, making the integration needed for evaluating the model evidence intractable. For clarity of presentation and computation, we selected network
models that focus on a particular network property; the proposed approach could also be used to evaluate networks models incorporating multiple properties simultaneously.

Regarding specification of the parameter priors, \( p(\theta_h | I_{m_h} = 1) \), priors can either draw on subject-matter knowledge or be essentially non-informative; there have been arguments for each of these two general types of priors. In our investigation of a patient-sharing networks for a particular state, we choose to include an informative prior based on patient-sharing networks from states other than the one of focus.

4.2.1 | Null model \((m_1)\)

We investigate a network model using a simple mechanism where each pair of providers form a connection with a fixed probability, \( p \), that is independent of all other edges. This mechanism corresponds to the ER model, which is commonly used as a null network model. Therefore, for \( m_1 \), \( \phi_1(g) = |E_g| \) and \( P_{\phi_1}(\phi_1(g) | \theta_h) \) is equal to Equation (5). We assume that the prior information for \( p \) follows a beta distribution, \( \text{Beta}(\alpha_1, \beta_1) \). The model evidence for \( m_1 \) is shown below in Equation (11):

\[
P(g | I_{m_1} = 1) = \frac{1}{|c_{\phi_1}(\phi_1(g))|} * \int_{p}^{p} p^{\phi_1(g)} (1 - p)^{n - \phi_1(g)} \frac{n(n - 1)/2}{\phi_1(g)} dp. \tag{11}
\]

4.2.2 | Sociality \((m_2)\)

Sociality is defined as the propensity for an individual to create connections; our goal is to evaluate whether there is heterogeneity among individuals in sociality. In any realization (whether observed or simulated) of a network, heterogeneity in sociality would be reflected in the degree distribution. To investigate this issue in the patient-sharing network, we consider a model \((m_2)\) that includes degree distribution as the sole network property. As Goodreau et al note, sociality is not synonymous with degree; the former is a feature of the network generating mechanisms, whereas the latter is a feature of any given realization of the mechanism.

Model \( m_2 \) represents networks generated such that the degrees follow an exponential distribution with scale parameter \( \lambda \); that is \( d_i \sim \exp(\lambda) \); one could also investigate alternative models, such as where the degrees follow a power-law distribution (discussed in Section 3). Again, we include prior information based on the patient-sharing networks from the states other than the one of focus. We assume the prior information for \( \lambda \) follows a normal distribution. The model evidence for \( m_2 \) is shown below in Equation (12):

\[
P(g | I_{m_2} = 1) = \frac{1}{|c_{\phi_2}(\phi_2(g))|} * \prod_{i=1}^{n} \prod_{j=0}^{D_i(g)} \lambda \exp^{-\lambda d_i} \exp \left[ -\frac{1}{2} \left( \frac{\lambda - \mu}{\sigma} \right)^2 \right] d\lambda; \tag{12}
\]

as with our simulation study, we do not adjust the model evidence due to not all degree sequences are graphical.

4.2.3 | Selective mixing \((m_3 \text{ and } m_4)\)

The resultant patient-sharing network may also be influenced by the presence of mechanisms by which providers form connections based on one or more of their individual characteristics. Often mixing is assortative, but it can also be disassortative; the former implies preferential formation of connections between individuals with similar characteristics, and the latter, between individuals with contrasting characteristics. Model \( m_3 \) investigates selective mixing by specialty, whereas model \( m_4 \) investigates mixing by the number of connections, that is, degree of a provider.
For model $m_3$, we consider mixing between primary and specialty care providers. Let $MM(g)$ be a matrix representing the mixing pattern of network $g$, where the entry $MM_{k,l}(g)$ is the total number of edges between a vertex with characteristic $k$ and vertex with characteristic $l$; in our application, we are interested in the characteristic indicating the provider type (primary or specialty care). Let $\phi_3(g) = MM(g)$. To investigate mixing by provider type in the patient-sharing network, we consider a model $(m_3)$ that includes mixing matrix as the sole network property. To develop a probability distribution for the mixing matrix, we use three independent variables representing the proportion of possible edges that link: (i) a primary care provider to another primary care primary, (ii) primary care provider to a specialty provider, and (iii) specialty provider to another specialty provider. Each of these variables follow a binomial distribution with parameters, $p_{3,pp}, p_{3,ps}$, and $p_{3,ss}$, respectively. We assume the prior information for these parameters, based on the patient-sharing networks from the states other than the one of focus, follow beta distributions, denoted as Beta($\alpha_{3,pp}, \beta_{3,pp}$), Beta($\alpha_{3,ps}, \beta_{3,ps}$), and Beta($\alpha_{3,ss}, \beta_{3,ss}$). Therefore, the model evidence for $m_3$ is as follows:

$$P(g|m_3 = 1) = \left(\frac{1}{|\phi_3(g)|}\right) \int_{p_{3,pp}} \int_{p_{3,ps}} \int_{p_{3,ss}} f_{3,pp} \cdot f_{3,ps} \cdot f_{3,ss} dp_{3,pp} dp_{3,ps} dp_{3,ss},$$

where,

$$f_{3,kl} = p_{3,kl}^{\phi_{3,kl}(g)} (1 - p_{3,kl})^{\psi_{3,kl}(g)} \cdot \frac{\phi_{3,kl}(g)^{\psi_{3,kl}(g)-1} \cdot (1 - \phi_{3,kl}(g))^{\psi_{3,kl}(g)-1}}{B(\alpha_{3,kl}, \beta_{3,kl})};$$

$\phi_{3,kl}(g)$ is the entry $MM_{k,l}(g)$, the number of edges in between providers specified by $k$ and $l$; and $\psi_{3,kl}$ is the number of possible edges between providers of type $k$ and $l$. As CCMs places only minimal restrictions on the selection of model for $P_{\phi_3}(\phi_3(g))$, an investigator could select the most appropriate model; for example, a model that allows for dependence in the variables representing the proportion of possible edges. We selected an model in which they are independent for clarity of presentation and computation. As discussed earlier, previous network models, such as ERGMs, do not enable an investigator to select the probability distribution for a network property; hence the limited literature on approaches to conduct such as selection. In the conclusion section, we discuss the potential for future research in this area.

Model $m_4$ represents networks that are generated based on the mechanism of selective mixing by degree. We evaluated the presence of this mechanism by modeling the degree mixing matrix, denoted as $DMM(g)$; the entry $DMM_{k,l}(g)$ is the total number of edges between vertices of degrees $k$ and $l$ and let $\phi_4(g) = DMM(g)$. As with the previous models, we need to select a probability distribution for observing a given degree mixing matrix. Let $p_{k,l}$ denote the expected proportion of edges between nodes of degrees $k$ and $l$, that is, the expected probability that an edge in a graph has endpoint degrees $k$ and $l$. Unlike the 2x2 mixing matrix investigated for selective mixing by provider type, a degree mixing matrix can be large. Therefore, it would not be possible to base our probability distribution for degree mixing matrices using a parameter for each upper triangular matrix entry (including diagonal entries) as we did for selective mixing by provider type. To reduce the number of parameters, we assume that the proportion of edges between nodes of degrees $k$ and $l$ ($k \leq l$), that is, $p_{k,l}$, is based on the following logistic model:

$$\text{logit}(p_{k,l}) = \beta_0 + \beta_1 \cdot k + \beta_2 \cdot l;$$

therefore, our probability distribution for degree mixing matrices is only based on three parameters ($\beta_0, \beta_1$, and $\beta_2$).

We assume the prior information for $\beta_0, \beta_1$, and $\beta_2$ follows a multivariate normal distribution; to minimize the effects of noise, our estimation of the prior only includes degrees up to 300. Therefore, the model evidence for $m_4$, assuming that the entries of the degree mixing matrix follow a multinomial distribution, is as follows:

$$P(g|m_4 = 1) \approx \left(\frac{1}{|\phi_4(g)|}\right) \int_{\beta} \frac{n!}{\prod_{k=1}^{n} \prod_{l \leq k} DMM_{k,l}(g)} \prod_{k=1}^{n} \prod_{l \leq k} p_{k,l}^{DMM_{k,l}(g)}$$

$$\cdot \frac{(2\pi)^{\frac{3}{2}}}{\det(\Sigma)} \exp \left[ -\frac{1}{2} (\beta - \mu)^T \Sigma^{-1} (\beta - \mu) \right] \, d\beta.$$
Once again, we do not adjust the model evidence for the fact that not all degree mixing matrices are graphical. As with mixing matrices themselves, an investigator could select the most appropriate model for the probability distribution for matrices that characterize degree mixing.

4.3 Investigation of patient-sharing networks

In this section, we present findings on whether there is evidence of sociality and selective mixing. We investigate these questions for each of the 50 states. However, we first present our findings for the state of Wyoming; the small number of providers reside in the state makes results easier to visualize compared to those of other states.
In 2015, the state of Wyoming had 1283 medical providers that share Medicare patients; we designated 412 and 871 as primary care and as specialty providers, respectively, based on their provider type in the Medicare Provider Utilization and Payment Data. In the first three quarters of 2015, there were 12,749 connections among these providers based on shared patients. Figure 1 presents a visualization of the patient-sharing network for these providers. The nodes represent providers and colored based on whether they are designated as primary (blue) or specialty (red) care. The edges between the nodes represent that the providers have at least one shared patient; we denote the patient-sharing network for Wyoming as \( g_{\text{WY}} \).

To investigate evidence for heterogeneity among individuals in their propensity to create connections, one could assess whether there is support for a model in which degrees of providers are more dispersed than would arise by chance; Figure 2 shows the observed degree distribution. We conducted this investigation by comparing a model in which the degrees follow a binomial distribution \( (m_1) \) to one with degrees that have a right-skewed distribution—specifically an exponential distribution \( (m_2) \). As noted above, a binomial distribution in degrees implies that connections form with a fixed probability, \( p \), and their formation is independent of all other edges.

The Bayesian model selection procedure requires evaluation of the model evidence for \( m_1 \) and \( m_2 \). As shown in Equation (8), a component of the model evidence is the inverse of the size of congruence class. Therefore, we must calculate this quantity associated with each model, that is, \( c_{\phi_h}(\phi_h(g_{\text{WY}})) \) for \( h \in \{1, 2\} \); Table 2 presents the model evidence as well as the natural logarithm for the size of congruence classes. The calculations of the size of congruence classes are based on the recursive algorithm described in Section 2.3. This results in estimated posterior probabilities of \( P(I_{m_1} = 1|g_{\text{WY}}) \leq 0.01\% \) and \( P(I_{m_1} = 1|g_{\text{WY}}) \geq 99.9\% \) when comparing only models \( m_1 \) and \( m_2 \). These results imply strong evidence of heterogeneity in sociality in the Wyoming patient-sharing network.

To investigate whether there is evidence of selective mixing by provider type, we compared a model in which connections are formed without consideration for the provider type \( (m_1) \) to a model in which the probability of forming a connection can vary by the provider types of the two providers associated with the edge \( (m_3) \). The model \( m_3 \) controls for density of the network as the entries in the mixing matrix denote the number of edges between the provider types. Therefore, the sum of the values in the upper triangular portion of the matrix provides the total number of edges in the network. The estimated size of congruence classes and model evidence for \( m_1 \) and \( m_3 \) are presented in Table 2. This results in posterior probability estimates of \( P(I_{m_1} = 1|g_{\text{WY}}) \geq 99.9\% \) and \( P(I_{m_3} = 1|g_{\text{WY}}) \leq 0.01\% \). Therefore, there is little evidence for selective mixing by provider type for Wyoming given our choice of \( m_3 \)—as specified by our choice of \( P_{\phi} \). There is the possibility that an alternative specification of \( m_3 \), which CCMs enables investigators to explore, would result in a different conclusion regarding selective mixing.
To assess evidence for selective mixing by degree, we must control for the distribution of degrees. Therefore, we compared models $m_2$ and $m_4$; $m_2$ allows the probability of forming a connection to vary by the degrees of the two providers associated with the edge; model $m_4$ controls for both degree distribution and number of edges, while model $m_2$ controls for the number of edges. The estimated size of congruence classes and model evidence for $m_2$ and $m_4$ are presented in Table 2. This results in posterior probabilities of $P(I_{m_2}=1|g_{WY}) \geq 99.9\%$ and $P(I_{m_4}=1|g_{WY}) \leq 0.01\%$. The posterior probability estimates for $m_2$ and $m_4$ suggest that there is strong evidence in favor of $m_2$, the model which includes only degree distribution. Again, we cannot conclude that degree mixing is not an influence on the network structure, but only assert that $m_4$—as specified by our choice of $P_{\Phi_4}$—did not lead to a better model compared to our choice of model for degree distribution; CCMs enable investigation of alternative models for specifying degree mixing.

Across all states, we see strong evidence of heterogeneity in sociality. However, we see little evidence of selective mixing by provider type and degree; for computational reasons, we only investigated 10 states for selective mixing by degree.

## DISCUSSION

Two factors that allow investigators to make best use of their prior knowledge about networks and of observed data are: (i) the ability to select the probability distribution for network properties, which enables investigators to evaluate the appropriateness of different network models that correspond to mechanisms that they hypothesize lead to generation of an observed network, and (ii) the ability to incorporate prior information into network analyses. We demonstrate these capabilities in our investigation of generative mechanisms associated with patient-sharing networks. In particular, we investigated heterogeneity in sociality as well as selective mixing by provider type and degree. To do so, we developed CCMs corresponding to these network properties. Using our Bayesian model selection approach, we found evidence in support of heterogeneity in sociality but not selective mixing.

There are many potential mechanisms that can generate patient-sharing networks; therefore, potentially many CCMs (beyond the ones we selected) that can be used to investigate heterogeneity in sociality and selective mixing. A strength of CCMs is that they enable investigation of alternative models, which could result in a different conclusion regarding selective mixing. More work is required to develop probability distributions for network properties that correspond to distinct generative network mechanisms.

There are several limitations to our analysis. First, our analysis did not account for heterogeneity in the number of patients treated by providers; developing methods to incorporate this information into analyses is another promising area of research. Second, we investigated patient-sharing networks based on the representation of connections between providers as being binary (present or absent); future research is needed to expand CCMs to account for weighted edges.

ACKNOWLEDGMENTS

This research is supported by grants from the National Institute of Health (R37 AI-51164 and R24 AI-106039).

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in CMS Data portal at https://www.cms.gov/Regulations-and-Guidance/Legislation/FOIA/Referral-Data-FAQs and https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier2015.
REFERENCES

1. Landon B, Keating N, Barnett M, et al. Variation in patient-sharing networks of physicians across the United States. *JAMA*. 2012;308(3):265-273.

2. DuGoff E, Fernandes-Taylor S, Weissman G, Huntley J, Pollack C. A scoping review of patient-sharing network studies using administrative data. *Transl Behav Med*. 2018;8(4):598-625.

3. Barnett M, Landon B, O’Malley AJ, Keating N, Christakis N. Mapping physician networks with self-reported and administrative data. *Health Serv Res*. 2011;46(5):1592-1609.

4. DuGoff E, Cho J, Shi Y, Pollack C. Geographic variations in physician relationships over time: implications for care coordination. *Med Care Res Rev*. 2018;75(5):586-611.

5. Pollack C, Lemke K, Roberts E, Weiner J. Patient sharing and quality of care: measuring outcomes of care coordination using claims data. *Med Care*. 2015;53(4):317.

6. Bodenheimer T, Berry-Millett R. Care management of patients with complex health care needs. Princeton, NJ: Robert Wood Johnson Foundation; 2009.

7. Barnett M, Christakis N, O’Malley AJ, Onnela J-P, Keating N, Landon B. Physician patient-sharing networks and the cost and intensity of care in US hospitals. *Med Care*. 2012;50(2):152.

8. Landon B, Keating N, Onnela J-P, Zaslavsky A, Christakis N, O’Malley AJ. Patient-sharing networks of physicians and health care utilization and spending among medicare beneficiaries. *JAMA Intern Med*. 2018;178(1):66-73.

9. Uddin S, Hamra J, Hossain L. Mapping and modeling of physician collaboration network. *Stat Med*. 2013;32(20):3539-3551.

10. Pollack C, Weissman G, Lemke K, Hussey P, Weiner J. Patient sharing among physicians and costs of care: a network analytic approach to care coordination using claims data. *J Gen Int Med*. 2013;28(3):459-465.

11. Uddin S, Hossain L, Kelaher M. Effect of physician community structure on healthcare outcomes. *Stud Health Technol Inform*. 2015;214:152-158.

12. Hussain T, Chang H-Y, Veenstra C, Pollack C. Collaboration between surgeons and medical oncologists and outcomes for patients with stage III colon cancer. *J Oncol Pract*. 2015;11(3):e388-e397.

13. Pollack C, Frick K, Herbert R, et al. It’s who you know: patient-sharing, quality, and costs of cancer survivorship care. *J Cancer Surviv*. 2014;8(2):156-166.

14. Carson M, Scholtens D, Frailey C, et al. Characterizing teamwork in cardiovascular care outcomes: a network analytics approach. *Circ Cardiovasc Qual Outcomes*. 2016;9(6):670-678.

15. Carson M, Scholtens D, Frailey C, Gravenor S, Kricke G, Soulakis N. An outcome-weighted network model for characterizing collaboration. *PLoS One*. 2016;11(10):e0163861.

16. Holder M, Lewis P. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet*. 2003;4(4):275-284.

17. Caimo A, Friel N. Bayesian model selection for exponential random graph models. *Soc Netw*. 2013;35(1):11-24.

18. Caimo A, Friel N. Bayesian inference for exponential random graph models. *Soc Netw*. 2011;33(1):41-55.

19. Thiemichen S, Friel N, Caimo A, Kauermann G. Bayesian exponential random graph models with nodal random effects. *Soc Netw*. 2016;46:11-28.

20. Bouranis L, Friel N, Maire F. Efficient Bayesian inference for exponential random graph models by correcting the pseudo-posterior distribution. *Soc Netw*. 2017;50:98-108.

21. Goyal R, De Gruttola V. Dynamic network prediction. *Netw Sci*. 2020;8(4):574-595.

22. Goyal R, Blitzstein J, De Gruttola V. Sampling networks from their posterior predictive distribution. *Netw Sci*. 2014;2(01):107-131.

23. Goldenberg A, Zheng A, Fienberg S, Airoldi E. A survey of statistical network models. *Found Trends Mach Learn*. 2010;2(2):129-233.

24. Petrovic S. A survey of discrete methods in (algebraic) statistics for networks. *Algebr Geometr Methods Discr Math*. 2017;685:260-281.

25. ShaCli C, Rinaldo A. Consistency under sampling of exponential random graph models. *Ann Stat*. 2013;41(2):508-535.

26. Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of infectious disease. *Math Biosci*. 1996;133:165-195.
How to cite this article: Goyal R, De Gruttola V. Investigation of patient-sharing networks using a Bayesian network model selection approach for congruence class models. *Statistics in Medicine*. 2021;40:3167–3180. https://doi.org/10.1002/sim.8969