Tilburg University

Network autocorrelation modeling

Dittrich, Dino; Leenders, Roger Th A.J.; Mulder, Joris

&ᴈ 4 ᴈ&

NETWORK AUTOCORRELATION MODELING: BAYESIAN TECHNIQUES FOR ESTIMATING AND TESTING MULTIPLE NETWORK AUTOCORRELATIONS

Dino Dittrich* [iD]
Roger Th. A. J. Leenders[†]*
Joris Mulder*[†]

Abstract

The network autocorrelation model has been the workhorse for estimating and testing the strength of theories of social influence in a network. In many network studies, different types of social influence are present simultaneously and can be modeled using various connectivity matrices. Often, researchers have expectations about the order of strength of these different influence mechanisms. However, currently available methods cannot be applied to test a specific order of social influence in a network. In this article, the authors first present flexible Bayesian techniques for estimating network autocorrelation models with multiple network autocorrelation parameters. Second, they develop new Bayes factors that allow researchers to test hypotheses with order constraints on the network autocorrelation parameters in a direct

*Tilburg University, Tilburg, the Netherlands
[†]Jheronimus Academy of Data Science, 's-Hertogenbosch, the Netherlands

Corresponding Author:
Dino Dittrich, Tilburg University, Tilburg School of Social and Behavioral Sciences, Department of Methodology and Statistics, Warandelaan 2, 5037 AB Tilburg, the Netherlands.
Email: dino.dittrich@gmail.com

manner. Concomitantly, the authors give efficient algorithms for sampling from the posterior distributions and for computing the Bayes factors. Simulation results suggest that frequentist properties of Bayesian estimators on the basis of noninformative priors for the network autocorrelation parameters are overall slightly superior to those based on maximum likelihood estimation. Furthermore, when testing statistical hypotheses, the Bayes factors show consistent behavior with evidence for a true data-generating hypothesis increasing with the sample size. Finally, the authors illustrate their methods using a data set from economic growth theory.

Keywords

network autocorrelation model, hypothesis testing, Bayes factor, order hypotheses, empirical Bayes

1. INTRODUCTION

Social network research plays an important role in understanding how countries, organizations, and persons influence one another's behavior, decision making, and well-being. The network autocorrelation model (Doreian 1981; Ord 1975) has been the workhorse for estimating and testing the strength of social influence with respect to a variable of interest in a given network (Fujimoto, Chou, and Valente 2011). In the network autocorrelation model, actors' behavior, decision making, or well-being is assumed to be correlated, and a *network autocorrelation parameter ρ* is estimated, representing the strength of a social influence mechanism in the network. The network autocorrelation model has been used to analyze network influence on individual behavior across many different fields, such as criminology (Tita and Radil 2011), ecology (McPherson and Nieswiadomy 2005), economics (Kalenkoski and Lacombe 2008), geography (Mur, López, and Angulo 2008), organization studies (Mizruchi and Stearns 2006), political science (Gimpel and Schuknecht 2003), and sociology (Burt and Doreian 1982).

Even though the network autocorrelation model has yielded many useful findings, the standard, or *first-order*, specification of the model implicitly assumes the presence of only a single network influence mechanism. However, this may be too restrictive in many cases, as different types of social influence are likely to be present simultaneously. For example, an actor is often a member of multiple distinct but potentially overlapping networks, such as a collaboration network, a friendship network, or an information-sharing network. Similarly, ties need

not only be defined by social interaction but can also refer to geographical proximity, joint memberships, or money flows. Each of these networks may have some connection to the variable of interest; hence, a model that ignores multiple influence mechanisms might be overly simplistic. Besides the fact that individuals are often members of multiple, potentially overlapping, networks, it is also the case that many networks are characterized by subgroups. For example, children in school classes may belong to different social classes. We might ask if, with respect to school performance, children of socially disadvantaged backgrounds influence one another on the basis of the same influence mechanism, say friendship, stronger than do children from more privileged backgrounds. Another example of grouping can be found in economic growth theory: with respect to economic growth, central nations are expected to be subject to different processes than are peripheral developing nations (Dall'erba, Percoco, and Piras 2009; Leenders 1995).

The network autocorrelation model can be straightforwardly extended to include multiple influence mechanisms and different subgroups within a network. Anselin (2001), Badinger and Egger (2013), Elhorst, Lacombe, and Piras (2012), Gupta and Robinson (2015), Hepple (1995b), Lee and Liu (2010), and Leenders (1995) provided theoretical discussions of and frequentist estimation procedures for these *higher-order* network autocorrelation models, and empirical applications can be found in, for example, Beck, Gleditsch, and Beardsley (2006), Dall'erba et al. (2009), Lacombe (2004), LeSage and Pace (2008), Lin, Wu, and Lee (2006), and Tita and Radil (2011). There has been a steady evolution of Bayesian estimation and model selection techniques for the first-order network autocorrelation model as well as the encompassing spatial Durbin model over the past 25 years (for a taxonomy of spatial regression models, see Elhorst 2010), most notably in Dittrich, Leenders, and Mulder (2017a, 2017b), Hepple (1995a, 1995b), LeSage (1997a), LeSage and Pace (2007), and LeSage and Parent (2007), but comparable advances for higher-order models are nearly nonexistent. Only recently, Han, Hsieh, and Lee (2017) developed an efficient Bayesian sampler for higher-order network autocorrelation models on the basis of the exchange algorithm in Murray, Ghahramani, and MacKay (2006), providing a fast Bayesian estimation of the model when networks are large. However, despite the usefulness of the sampler in Han et al. (2017), their approach has two important limitations: first, it does not permit researchers to include substantial prior information about the network

autocorrelation parameters to their analyses; second, it is restricted to testing nested hypotheses on these parameters only, which prohibits researchers from testing more intricate hypotheses on the network effects.

In this article, we develop a fully Bayesian framework that allows the inclusion of external prior information for estimating higher-order network autocorrelation models and for simultaneously testing multiple non-nested constraints on the relative order of network effects, such as $H_0 : \rho_1 = \rho_2 = 0$, $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, or $H_3 : \rho_1 = \rho_2 > 0$, where ρ_1 and ρ_2 quantify the strength of different influence mechanisms, respectively. Using a Bayesian approach for estimating and testing higher-order network autocorrelation models has several advantages compared with classical methods such as maximum likelihood estimation and null hypothesis significance testing. First, in contrast to maximum likelihood estimation of higher-order models, Bayesian estimation does not rely on asymptotic theory for computing standard errors of the network autocorrelation parameters, potentially resulting in a more accurate quantification of parameter uncertainty in small networks. Second, unlike null hypothesis significance testing, *Bayes factors* allow researchers to quantify relative evidence in the data in favor of the null, or any other, hypothesis against another hypothesis (Kass and Raftery 1995), and Bayes factors can be easily extended to test more than two hypotheses against each other simultaneously (Raftery, Madigan, and Hoeting 1997). Hence, this enables researchers to precisely test multiple network operationalizations against one another. Third, Bayes factors have been proven to be very effective for testing hypotheses with order constraints on the parameters of interest (Braeken, Mulder, and Wood 2015; Klugkist, Laudy, and Hoijtink 2005; Mulder 2016; Mulder and Wagenmakers 2016).

For example, in a simple research design in which regions are divided into higher-productivity and lower-productivity regions, it would be reasonable to expect differing levels of influence within and between the two sets of regions. In such a setup, one could hypothesize that higher-productivity regions might influence one another's policies more strongly than lower-productivity regions would influence one another. Moreover, one could argue that the influence of higher-productivity regions on lower-productivity ones is likely higher than that of lower-productivity regions on higher-productivity regions. Of course, one

could formulate competing hypotheses, such as that all network autocorrelations are zero, that they are nonzero but equal to one another, or that there is neither influence of lower-productivity regions on higher-productivity ones nor influence among lower-productivity regions themselves. This cannot be done using classical tests and is of particular importance in higher-order network autocorrelation models, as in this setting, researchers often have expectations about the order of strength of different network effects. Such expectations are implicit in most research and Bayes factors permit researchers to state them as actual hypotheses and then test them in a precise and straightforward manner. Beyond allowing researchers to test the more interesting hypotheses they may already have, we hope the availability of this approach will also stimulate researchers to theorize more creatively and more precisely about social influence phenomena, knowing their hypotheses can be easily and correctly tested against one another.

Thus, we propose Bayes factors for testing multiple hypotheses on the relative importance of network influence in a given network. The presented methodology not only allows a researcher to conclude if there is evidence in the data for, or against, nonzero network autocorrelations in the network, but it also grants the researcher the opportunity to simultaneously test any number of competing hypotheses on the relative strength of the network effects against one another. Subsequently, we conduct an extensive simulation study to investigate and show the desirable numerical properties of the new procedures, which we then use to reanalyze a data set from the economic growth literature. In addition to motivating and introducing new methodology, another main goal of this article is to make the methods easily available to researchers by providing ready-to-use R code.

We proceed as follows. In the next section, we present higher-order network autocorrelation models in detail before introducing Bayesian estimation and hypothesis testing techniques for the model in Sections 3 and 4. Concomitantly, we provide efficient implementations for estimating higher-order network autocorrelation models and for computing Bayes factors involving order hypotheses on the network autocorrelation parameters. We assess the numerical behavior of the proposed methods in Section 5. In Section 6, we illustrate our approaches with an empirical example, and Section 7 concludes.

2. THE NETWORK AUTOCORRELATION MODEL

2.1. *The First-Order Network Autocorrelation Model*

Building on a standard linear regression model, the network autocorrelation model relaxes the assumption of independence of observations and allows correlation between them by explicitly using the underlying network structure. More precisely, an actor's response is modeled as the weighted sum of the actor's neighbor responses and a linear combination of actor attributes. In mathematical notation, the first-order network autocorrelation model is given by

$$\boldsymbol{y} = \rho W \boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}_g, \sigma^2 I_g\right), \tag{1}$$

where $\boldsymbol{y}$ is a vector of length g containing the observations for a variable of interest for the g actors in a network, X is a standard $(g \times k)$ design matrix (possibly including a vector of ones in the first column for an intercept term), $\boldsymbol{\beta} \in \mathbb{R}^k$ is a vector of k regression coefficients as in standard linear regression, $\boldsymbol{\varepsilon} \in \mathbb{R}^g$ comprises the error terms that are assumed to be independent and identically normally distributed with zero mean and variance of σ^2, $\mathbf{0}_g$ is a vector of zeros of length g, and I_g denotes the $(g \times g)$ identity matrix. Furthermore, W is a $(g \times g)$ *connectivity matrix*, where a nonzero entry W_{ij} amounts to the influence of actor j on actor i and $W_{ii} = 0$ for all $i \in \{1, ..., g\}$. Typically, W is *row-standardized*; that is, all rows sum to 1, which in this case means that the term $W\boldsymbol{y}$ represents the vector of the actors' neighbor average responses. Finally, ρ is the network autocorrelation parameter and quantifies the magnitude of the network influence with respect to a variable of interest in a given network as induced by W. For a substantive interpretation of the model, see Leenders (1995, 2002).

The model's likelihood is multivariate normal and can be written as

$$f(\boldsymbol{y}|\rho, \sigma^2, \boldsymbol{\beta}) = \left|\det\left(A_\rho\right)\right| (2\pi\sigma^2)^{-\frac{g}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)^T \left(A_\rho \boldsymbol{y} - X\boldsymbol{\beta}\right)\right), \tag{2}$$

where $A_\rho := I_g - \rho W$ (see, e.g., Doreian 1981). Usually, the parameter space of ρ is chosen as the interval around $\rho = 0$ for which A_ρ is nonsingular (Hepple 1995a; LeSage and Parent 2007; Smith 2009). The bounds of this feasible range of ρ are determined by the eigenvalues of W with the smallest and largest real part, respectively, which means that ρ must be contained in $\left(1/Re(\lambda_g[W]), 1/Re(\lambda_1[W])\right)$, where $\lambda_1[W], ..., \lambda_g[W]$ denote the eigenvalues of W with $Re(\lambda_1[W]) \geq ... \geq Re(\lambda_g[W])$

(Hepple 1995a). The model's overall parameter space of $\boldsymbol{\theta} := (\rho, \sigma^2, \boldsymbol{\beta})$ is then given by $\Theta := \Theta_\rho \times \Theta_{\sigma^2} \times \Theta_{\boldsymbol{\beta}} = \left(1/Re(\lambda_g[W]), 1/Re(\lambda_1[W])\right) \times (0, \infty) \times \mathbb{R}^k.$[1]

2.2. *Higher-Order Network Autocorrelation Models*

The standard, or first-order, network autocorrelation model in equation (1) is limited to a single network autocorrelation parameter ρ and a single connectivity matrix W. Hence, in this model the network influence is assumed to be homogeneously distributed across the network on the basis of a single influence mechanism. Extending the first-order model to higher-order network autocorrelation models allows a richer dependence structure by including multiple connectivity matrices, representing different influence mechanisms (e.g., geographic adjacency and social similarity).[2] This amounts to the functional form

$$y = \sum_{r=1}^{R} \rho_r W_r y + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}_g, \sigma^2 I_g\right), \tag{3}$$

where $\{W_r\}_r$ are distinct connectivity matrices, and the corresponding network autocorrelation parameters $\{\rho_r\}_r$ denote the strength of the different influence mechanisms.

In practice, there can be overlap between connectivity matrices; that is, different connectivity matrices may share common ties. Partially overlapping connectivity matrices do not pose identification problems as long as there is no complete overlap (Elhorst et al. 2012), but overlap does make interpretability of the network autocorrelation parameters more difficult (Elhorst et al. 2012; LeSage and Pace 2011). In particular, partial overlap may result in empirically unlikely negative network autocorrelations (Dittrich et al. 2017a; Elhorst et al. 2012). We analyze the numerical effect of overlapping connectivity matrices on the estimation of and hypothesis tests on $\boldsymbol{\rho} := (\rho_1, ..., \rho_R)$ in more detail in a simulation study in Section 5.

Higher-order network autocorrelation models not only allow one to consider multiple influence mechanisms, but they also allow researchers to partition a network into several subgroups. In the latter case, possible heterogeneity in network influence strength is included in the model by allowing for different levels of network autocorrelation within and between subgroups for a given influence mechanism (e.g., geographic adjacency). Dividing actors in a network into S subgroups, with sizes

$g_1, ..., g_S$ and $\sum_{s=1}^{S} g_s = g$, a model with multiple subgroups can be expressed using the representation in equation (3) by writing

$$y = \begin{bmatrix} y_1 \\ \cdots \\ y_S \end{bmatrix} = \begin{bmatrix} \rho_{11} W_{11} & \cdots & \rho_{1S} W_{1S} \\ \cdots & \cdots & \cdots \\ \rho_{S1} W_{S1} & \cdots & \rho_{SS} W_{SS} \end{bmatrix} \begin{bmatrix} y_1 \\ \cdots \\ y_S \end{bmatrix} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$= \left(\rho_{11} \begin{bmatrix} W_{11} & 0 & \cdots \\ 0 & \cdots & 0 \end{bmatrix} + \cdots + \rho_{SS} \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & W_{SS} \end{bmatrix} \right) \begin{bmatrix} y_1 \\ \cdots \\ y_S \end{bmatrix} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where y_s is a vector of length g_s containing the observations for the g_s actors in the sth subgroup of the network, $W_{ss'}$ is a $(g_s \times g_{s'})$ connectivity matrix defining the influence relationships between members of subgroup s' and members of subgroup s, and $\rho_{ss'}$ is a network autocorrelation parameter representing the strength of the network influence of the actors in subgroup s' on the actors in subgroup s. Because the sizes of the S subgroups potentially differ, each $W_{ss'}$ is typically row-standardized separately, which removes scale effects and eases direct comparison between the network autocorrelation parameters (McMillen, Singell, and Waddell 2007).

The structure of the likelihood function of higher-order network autocorrelation models remains the same as in the first-order model in equation (2), with A_ρ being replaced by $A_{\boldsymbol{\rho}} := I_g - \sum_{r=1}^{R} \rho_r W_r$. As in the first-order model, we define the R-dimensional parameter space of $\boldsymbol{\rho} = (\rho_1, ..., \rho_R)$ as the space containing the origin for which $A_{\boldsymbol{\rho}}$ is nonsingular. Elhorst et al. (2012) provided a simple procedure for checking if a point $\boldsymbol{\rho}^* \in \mathbb{R}^R$, given $W_1, ..., W_R$, lies in the corresponding feasible parameter space $\Theta_{\boldsymbol{\rho}}$.[3] Figure 1 shows two exemplary feasible parameter spaces of (ρ_1, ρ_2) in a second-order network autocorrelation model for simulated data based on nonoverlapping matrices (left) and with 40 percent overlap (right).

2.3. *Application of a Higher-Order Network Autocorrelation Model: Economic Growth of Labor Productivity*

In this subsection, we introduce a data set from the economic growth literature that prompts questions that can readily be answered using the proposed Bayes factor approach. Here, we merely describe the data set and the research questions; in Section 6, we provide solutions to the posed questions.
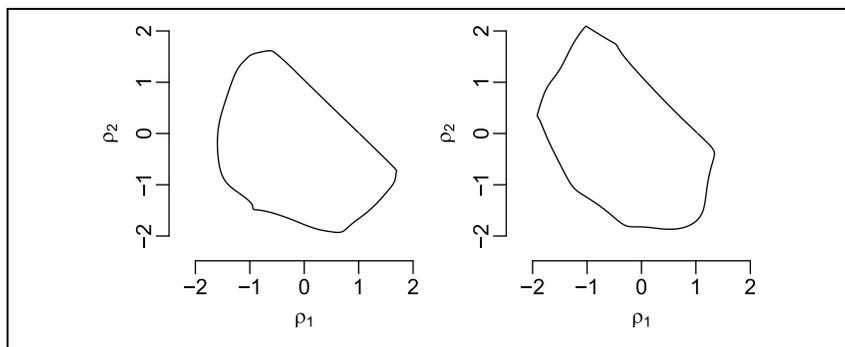
Figure 1. Feasible two-dimensional parameter space $\Theta_{(\rho_1, \rho_2)}$ for simulated data based on nonoverlapping connectivity matrices (left) and connectivity matrices with a 40 percent overlap (right).

Dall'erba et al. (2009) used a second-order network autocorrelation model to explain the growth rates of labor productivity in the service industry across 188 European regions in 12 countries from 1980 to 2003. To adequately deal with interregional spillovers, the authors introduced two different spatial weight matrices, W_1 and W_2, "under the assumption that economic interactions decrease very substantially when a national border is passed" (p. 337). They constructed W_1 using a region's three nearest neighbors within the same country, and W_2 was based on the three nearest neighbors in bordering countries. The authors then row-normalized these raw binary connectivity matrices. In addition to an intercept term, Dall'erba et al. (2009) considered four more explanatory variables: the growth rate of market service output in a region, the initial labor productivity gap between the region and the leading region, a measure of the region's urbanization, and a measure of the region's accessibility. Thus, their model is given by

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \beta_1 X_{\cdot 1} + \beta_2 X_{\cdot 2} + \beta_3 X_{\cdot 3} + \beta_4 X_{\cdot 4} + \beta_5 X_{\cdot 5} + \varepsilon, \quad \varepsilon \sim N(\mathbf{0}, \sigma^2 I), \tag{4}$$

where $y \in \mathbb{R}^{188}$ is the vector of growth rates of labor productivity in the service industry across the 188 regions, $\boldsymbol{\beta} \in \mathbb{R}^5$ represents the vector of the four regression coefficients plus an intercept term, $X \in \mathbb{R}^{188 \times 5}$ contains the values for the explanatory variables for the 188 regions, where

$X_{.i}, i \in \{1, ..., 5\}$, denotes the ith column of X, $\mathbf{0} \in \mathbb{R}^{188}$ is a vector of zeros, and $I \in \mathbb{R}^{188 \times 188}$ represents the corresponding identity matrix.

The authors found that the estimate of ρ_1, reflecting interactions within the same country, is positive and statistically significant, indicating the presence of positive spatial within-country spillover effects. On the other hand, the estimate of ρ_2 is very close to zero and statistically not significant. Dall'erba et al. (2009) concluded by saying that "the results obtained also confirm the hypothesis that economic interactions decrease very substantially when a national border is passed (indeed, the coefficient reflecting external spillovers is not statistically significant)" (p. 342). However, to draw this conclusion, one needs to directly test a corresponding hypothesis, for example, $H_1 : \rho_1 > \rho_2 = 0$, against a (set of) competing hypothesis (hypotheses), such as $H_0 : \rho_1 = \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, or (and) $H_3 : \rho_1 = \rho_2 > 0$. These four hypotheses correspond to the notion of "no network effects" (H_0), "a positive within-country network effect only" (H_1), "positive but decreasing network effects after a national border is passed" (H_2), and "positive and equally strong within-country and between-country network effects" (H_3). Currently, no formal statistical method is available to directly test such hypotheses on multiple network autocorrelations. In the remainder of this article, we develop a Bayesian framework for testing and quantifying the evidence in the data for such hypotheses involving equality as well as order constraints on the network effects. We come back to this empirical example and test these hypotheses against one another using Bayes factors in Section 6. Finally, Dall'erba et al. (2009) stated that "there is evidence that the coefficients in a growth model are potentially varying for different subsets of the total sample" (p. 342). In Section 6, we investigate if there is such evidence in this data set by considering a network autocorrelation model with two subgroups, allowing for differing levels of network autocorrelation within and between the two subgroups.

3. BAYESIAN ESTIMATION OF HIGHER-ORDER NETWORK AUTOCORRELATION MODELS

3.1. *Prior Specification*

Bayesian estimation starts with formulating prior expectations about the parameters in a model in terms of *prior distributions*, or *priors*. These priors summarize the (lack of) information about the model parameters

before observing the data. If such prior information is available (e.g., on the basis of previous literature), *informative priors* for the parameters of interest can be formulated. For the first-order network autocorrelation model, Dittrich et al. (2017a) performed a literature study, looking at the distribution of reported network autocorrelations across many different fields. On the basis of their results, most of the analyzed data in the literature exhibit positive network autocorrelation between 0 and 0.5, and it seems highly unlikely to observe negative network autocorrelation estimates (as previously noted by, e.g., Neuman and Mizruchi 2010). This information could then be used to formulate an informative prior for ρ in a first-order network autocorrelation model, as Dittrich et al. (2017a, 2017b) did. On the other hand, if such prior information is missing, or a researcher deliberately refrains from adding additional information to the model through the prior, *noninformative priors* are often used (Gelman et al. 2003). In the network autocorrelation model, σ^2 and $\boldsymbol{\beta}$ are commonly assigned the standard noninformative priors $p(\sigma^2) \propto 1/\sigma^2$ and $p(\boldsymbol{\beta}) \propto 1$ (Hepple 1995a; Holloway, Shankar, and Rahman 2002; LeSage 1997b). These priors assume that all possible values for $\log(\sigma^2)$ and $\boldsymbol{\beta}$ are equally likely a priori. We also do so throughout this article. Note that these priors are not *proper* in the sense that they do not integrate to a finite value, which does not affect estimation of the model.

We use a general R-variate normal prior for $\boldsymbol{\rho}$, $p(\boldsymbol{\rho}) = \phi_{\boldsymbol{\mu}, \Sigma}(\boldsymbol{\rho}) \mathbb{1}_{\Theta_{\rho}}(\boldsymbol{\rho}) c^{-1}$, where $\phi_{\boldsymbol{\mu}, \Sigma}(\cdot)$ denotes the probability density function of a multivariate normal distribution with prior mean $\boldsymbol{\mu}$ and prior covariance matrix Σ, $\mathbb{1}.(\cdot)$ is the standard indicator function, and $c := \int_{\Theta_{\rho}} \phi_{\boldsymbol{\mu}, \Sigma}(\boldsymbol{\rho}) \mathrm{d}\boldsymbol{\rho}$ is a normalizing constant representing the probability mass of $\phi_{\boldsymbol{\mu}, \Sigma}(\cdot)$ contained in the network autocorrelation parameters' space Θ_{ρ}. If researchers have sufficient prior information about the network autocorrelations, they can specify $\boldsymbol{\mu}$ and Σ directly. Alternatively, when specifying Σ vaguely enough, that is, with very large diagonal elements, the prior becomes essentially identical to a proper uniform distribution for $\boldsymbol{\rho}$ on the bounded parameter space Θ_{ρ}.[4]

In summary, we use the following priors for the model parameters, which we assume to be a priori independent from one another:

$$p(\boldsymbol{\rho}) = \phi_{\boldsymbol{\mu}, \Sigma}(\boldsymbol{\rho}) \mathbb{1}_{\Theta_{\rho}}(\boldsymbol{\rho}) c^{-1}, \tag{5}$$

$$p(\sigma^2) \propto 1/\sigma^2, \tag{6}$$

$$p(\boldsymbol{\beta}) \propto 1, \tag{7}$$
$$p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}) = p(\boldsymbol{\rho}) \times p(\sigma^2) \times p(\boldsymbol{\beta}).$$

3.2. *Posterior Computation*

After having specified a prior distribution for the model parameters, the information contained in the observed data y is used to update the prior distribution and to arrive at the *posterior distribution*, or simply *posterior*. The posterior is used for all Bayesian inference in the model, for example, to obtain point estimates of model parameters (the posterior mean or the posterior median), to construct Bayesian credible intervals (i.e., intervals in the domain of the posterior), or to determine other statistics of interest, such as the probability that one network effect is stronger than another one for given data, $p(\rho_1 > \rho_2 | y)$. In this subsection, we specify the posterior in higher-order network autocorrelation models on the basis of the priors from Section 3.1, and we provide an automatic and efficient scheme to sample from this posterior.

First, Bayes' theorem gives that the posterior is proportional to the prior multiplied by the likelihood, more precisely

$$p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta} | y) = \frac{p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}) f(y | \boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta})}{\int_{\Theta_{\boldsymbol{\rho}}} \int_{\Theta_{\sigma^2}} \int_{\Theta_{\boldsymbol{\beta}}} p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}) f(y | \boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \mathrm{d}\boldsymbol{\rho}} \tag{8}$$
$$\propto p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}) f(y | \boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}).$$

The denominator of equation (8) is called the *marginal likelihood* and ensures that the posterior integrates to unity. The marginal likelihood does not depend on any model parameters and can be ignored in Bayesian estimation of the model. On the other hand, when testing hypotheses, the marginal likelihood does play a central role, as it quantifies how plausible the data are under a specific hypothesis, which we discuss in the following section.

Next, using the priors in equations (5), (6), and (7), and the likelihood function in equation (2), we can express the posterior $p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta} | y)$ in higher-order network autocorrelation models as

$$p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta} | y) \propto |A_{\boldsymbol{\rho}}| (\sigma^2)^{-\frac{g}{2} - 1}$$
$$\exp\left(-\frac{1}{2} (\boldsymbol{\rho} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\rho} - \boldsymbol{\mu}) - \frac{1}{2\sigma^2} (A_{\boldsymbol{\rho}} y - X\boldsymbol{\beta})^T (A_{\boldsymbol{\rho}} y - X\boldsymbol{\beta}) \right). \tag{9}$$

However, the posterior in equation (9) does not belong to a family of known probability distributions, so we cannot directly infer its posterior mean, its quantiles, or other quantities of interest.[5] In this case, it is common to sample random draws from the posterior and to use these posterior draws to approximate any desired statistic. An efficient method is to sequentially draw from the conditional posteriors, that is, the posterior of one parameter (block) given the remaining parameters and the data (Gelfand and Smith 1990; Geman and Geman 1984).[6] Extending the proposed method for the first-order network autocorrelation model in Dittrich et al. (2017a) to higher-order models, we sample the model parameters according to the following blocks: $(\boldsymbol{\rho}, \beta_1), \sigma^2$, and $\tilde{\boldsymbol{\beta}}$, where β_1 denotes the model's intercept and $\tilde{\boldsymbol{\beta}} := (\beta_2, ..., \beta_k)$ contains the remaining regression coefficients. By simultaneously sampling $\boldsymbol{\rho}$ and β_1, we can better capture potential posterior correlation between the network effects as well as potential correlation between the network effects and the intercept (Dittrich et al. 2017a). The conditional posteriors for the proposed blocks are then given by (see, e.g., LeSage 1997a)

$$
\begin{aligned}
&p(\boldsymbol{\rho}, \beta_1 | \sigma^2, \tilde{\boldsymbol{\beta}}, \boldsymbol{y}) \\
&\propto |A_{\boldsymbol{\rho}}| \exp\left(-\frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\rho} - \boldsymbol{\mu}) - \frac{1}{2\sigma^2}(A_{\boldsymbol{\rho}}y - X\boldsymbol{\beta})^T (A_{\boldsymbol{\rho}}y - X\boldsymbol{\beta}) \right),
\end{aligned}
\tag{10}
$$

$$
p(\sigma^2 | \boldsymbol{\rho}, \beta_1, \tilde{\boldsymbol{\beta}}, \boldsymbol{y}) \sim IG\left(\frac{g}{2}, \frac{(A_{\boldsymbol{\rho}}y - X\boldsymbol{\beta})^T (A_{\boldsymbol{\rho}}y - X\boldsymbol{\beta})}{2} \right),
\tag{11}
$$

$$
p(\tilde{\boldsymbol{\beta}} | \boldsymbol{\rho}, \beta_1, \sigma^2, \boldsymbol{y}) \sim N\left(\boldsymbol{\mu}_{\tilde{\beta}}, \Sigma_{\tilde{\beta}} \right),
\tag{12}
$$

where $IG(\cdot, \cdot)$ denotes the inverse gamma distribution, and $\boldsymbol{\mu}_{\tilde{\beta}}$ and $\Sigma_{\tilde{\beta}}$ are given in Appendix A.

Drawing from the conditional posteriors in equations (11) and (12) can be done using standard statistical software. In contrast, the conditional posterior in equation (10) does not have a well-known form and cannot be directly sampled from. Instead, we use the *Metropolis-Hastings algorithm* (Hastings 1970; Metropolis et al. 1953) to generate draws from the conditional posterior for $(\boldsymbol{\rho}, \beta_1)$. In short, the algorithm generates candidate values for the conditional posterior from a candidate-generating distribution that can be easily sampled from and subsequently accepts, or rejects, the draws with a certain probability.

The algorithm's efficiency mainly depends on the shape of the proposed candidate-generating distribution; if possible, exploiting the form of the conditional posterior and specifying a candidate-generating distribution that closely approximates it results in efficient solutions (Chib and Greenberg 1994, 1995, 1998).

We first approximate $\log(|A_{\boldsymbol{\rho}}|)$ by a quadratic polynomial in $\boldsymbol{\rho}$ by virtue of Jacobi's formula and the Mercator series (see Appendix A). Next, we observe that the logarithm of the exponential in equation (10) can also be written as a quadratic polynomial in $(\boldsymbol{\rho}, \beta_1)$. Hence, the logarithm of the conditional posterior itself can be approximated by a quadratic polynomial in $(\boldsymbol{\rho}, \beta_1)$. Finally, by equating coefficients of this quadratic polynomial with the log-kernel of the probability density function of an $(R+1)$-variate normal distribution, the density in equation (10) can be approximated by an $(R+1)$-variate normal candidate-generating density for $(\boldsymbol{\rho}, \beta_1)$ that is tailored to the conditional posterior for $(\boldsymbol{\rho}, \beta_1)$.[7] All details and the full sampling scheme can be found in Appendix A.

We implemented our proposed approach in R (R Core Team 2017) and compared its performance with a sampling scheme that does not block the network autocorrelation parameters and the intercept but uses one-dimensional random walk algorithms to generate draws for each network effect sequentially, as in Zhang et al. (2013). Figure 2 shows exemplary trace plots of posterior draws for ρ_1 and ρ_2 on the basis of the two sampling schemes and the data in Dall'erba et al. (2009) with model (4). We see that our method results in a more efficient implementation than drawing each network effect separately, as it generates Markov chains that explore the corresponding parameter space of (ρ_1, ρ_2) much faster. Finally, our approach is fully automatic in the sense that there are no parameters to be tuned in the Metropolis-Hastings algorithm, such as the variance of candidate-generating distributions.

To conclude, the presented sampling algorithm allows researchers to automatically and efficiently draw from the posterior on the basis of a general multivariate normal prior for the network autocorrelation parameters, including informative as well as noninformative specifications. Such efficient sampling is essential for performing any Bayesian estimation of the model, which solely relies on the generated posterior draws. An R program that implements our scheme is available at https://github.com/DittrichD/BayesNAM.
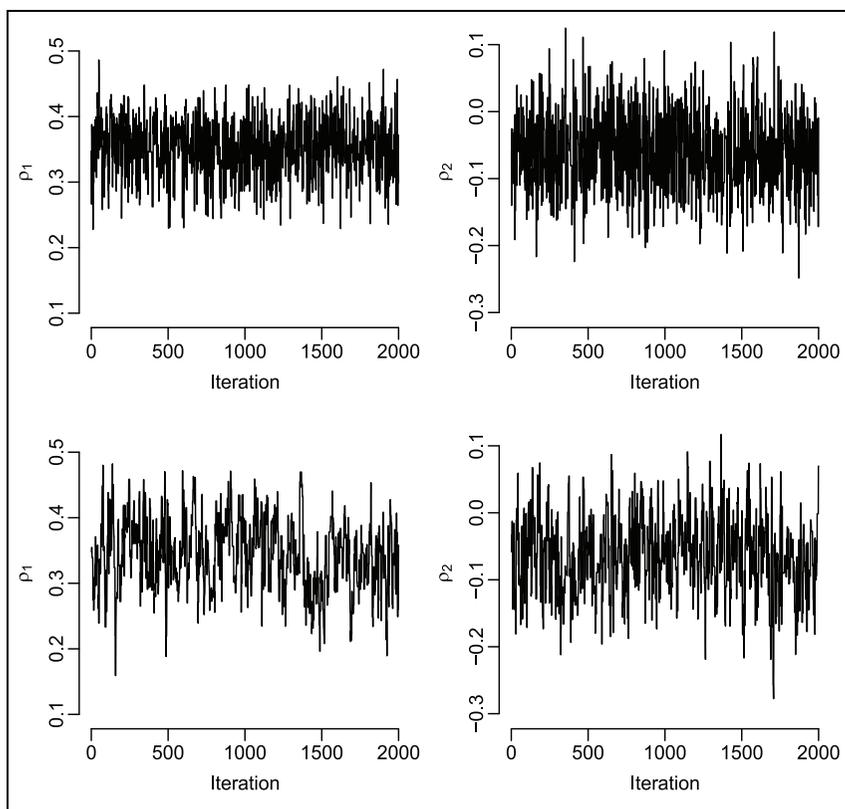
Figure 2. Trace plots of posterior draws for ρ_1 and ρ_2 on the basis of our proposed scheme (top row) and a random walk algorithm (bottom row) for the data in Dall'erba et al. (2009) with model (4).

4. BAYESIAN HYPOTHESIS TESTING IN HIGHER-ORDER NETWORK AUTOCORRELATION MODELS

In many network studies, researchers have competing theories about the specific order of different network effect strengths. These theories can be formulated as hypotheses on the network autocorrelation parameters, for example, as $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, or $H_3 : \rho_1 = \rho_2 > 0$, and can include as many network autocorrelation parameters as relevant to one's theory. The focus of interest then lies on which substantive theory, or hypothesis, is most plausible and most supported by the data and how

strongly. In this section, we consider $T \geq 2$ constrained hypotheses on the network effects, where a hypothesis H_t, $t \in \{0, ..., T-1\}$, contains q_t^E equality and q_t^I inequality constraints on $\boldsymbol{\rho}$, that is,

$$H_t := \begin{cases} R_t^E \boldsymbol{\rho} = r_t^E \\ R_t^I \boldsymbol{\rho} > r_t^I, \end{cases} \tag{13}$$

where R_t^E and r_t^E are a $(q_t^E \times R)$ matrix and a vector of length R, respectively, containing the coefficients of the q_t^E equality constraints under hypothesis H_t. Equivalently, the $(q_t^I \times R)$ matrix R_t^I and the vector r_t^I contain the coefficients of the q_t^I inequality constraints. For example, the constraints induced by the three hypotheses $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, and $H_3 : \rho_1 = \rho_2 > 0$ can be represented by equation (13) as[8]

$$H_1 : \quad R_1^E = (0, 1), r_1^E = 0, \quad R_1^I = (1, 0), r_1^I = 0,$$

$$H_2 : \quad R_2^I = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad r_2^I = (0, 0),$$

$$H_3 : \quad R_3^E = (1, -1), \quad r_3^E = 0, \quad R_3^I = (1, 0), r_3^I = 0.$$

Bayes factors for testing hypotheses of the form in equation (13) have also been proposed for other types of parameters, such as group variances (Böing-Messing et al. 2017), intraclass correlations (Mulder and Fox 2018), and regression effects (Mulder, Hoijtink, and Klugkist 2010).

4.1. *The Bayes Factor*

The Bayes factor is a comparative Bayesian hypothesis testing criterion that directly quantifies the relative evidence in the data in favor of a hypothesis. The Bayes factor of hypothesis H_t against hypothesis $H_{t'}$, $t, t' \in \{0, ..., T-1\}$, is defined as the ratio of the marginal likelihoods under the two hypotheses, that is, in the network autocorrelation model as

$$\begin{aligned} B_{tt'} &= \frac{m_t(\boldsymbol{y})}{m_{t'}(\boldsymbol{y})} \\ &= \frac{\int_{\Theta_{\rho_t}} \int_0^\infty \int_{\mathbb{R}^k} p_t(\boldsymbol{\rho}_t) p(\sigma^2) p(\boldsymbol{\beta}) f(\boldsymbol{y}|\boldsymbol{\rho}_t, \sigma^2, \boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \mathrm{d}\boldsymbol{\rho}_t}{\int_{\Theta_{\rho_{t'}}} \int_0^\infty \int_{\mathbb{R}^k} p_{t'}(\boldsymbol{\rho}_{t'}) p(\sigma^2) p(\boldsymbol{\beta}) f(\boldsymbol{y}|\boldsymbol{\rho}_{t'}, \sigma^2, \boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \mathrm{d}\boldsymbol{\rho}_{t'}}, \end{aligned} \tag{14}$$

Table 1. Evidence Categories for the Bayes Factor $BF_{tt'}$ as Given by Jeffreys (1961)

$BF_{tt'}$	$\log(BF_{tt'})$	Interpretation
>100	>4.61	Decisive evidence for hypothesis H_t
30 to 100	3.40 to 4.61	Very strong evidence for hypothesis H_t
10 to 30	2.30 to 3.40	Strong evidence for hypothesis H_t
3 to 10	1.10 to 2.30	Substantial evidence for hypothesis H_t
1 to 3	0 to 1.10	Not worth more than a bare mention
1/3 to 1	−1.10 to 0	Not worth more than a bare mention
1/10 to 1/3	−2.30 to −1.10	Substantial evidence for hypothesis $H_{t'}$
1/30 to 1/10	−3.40 to −2.30	Strong evidence for hypothesis $H_{t'}$
1/100 to 1/30	−4.61 to −3.40	Very strong evidence for hypothesis $H_{t'}$
<1/100	<−4.61	Decisive evidence for hypothesis $H_{t'}$

where $\boldsymbol{\rho}_t$ are the network autocorrelation parameters under hypothesis H_t, $p_t(\boldsymbol{\rho}_t)$ denotes their prior density, and $\Theta_{\boldsymbol{\rho}_t}$ is the corresponding parameter space (Kass and Raftery 1995). We assume common priors for σ^2 and $\boldsymbol{\beta}$ under both hypothesis H_t and hypothesis $H_{t'}$ as they are seen as nuisance parameters in the presented framework. The exact form of the priors for these nuisance parameters typically does not alter the magnitude of the Bayes factor (Kass and Raftery 1995).

The marginal likelihood under hypothesis H_t, $m_t(\boldsymbol{y})$, is a weighted average likelihood over the parameter space under hypothesis H_t, with the prior $p_t(\boldsymbol{\rho}_t)$ under hypothesis H_t acting as a weight function. Therefore, it can be interpreted as the probability that the data were observed under hypothesis H_t. Hence, the Bayes factor, as the ratio of two marginal likelihoods, quantifies the relative evidence that the data were observed under hypothesis H_t rather than hypothesis $H_{t'}$. For example, when $B_{tt'} = 5$, this indicates that the data are five times more likely to have occurred under hypothesis H_t compared with hypothesis $H_{t'}$. Conversely, when $B_{tt'} = 1/5$, it is five times more likely to have observed the data under hypothesis $H_{t'}$ than under hypothesis H_t.

To facilitate interpretation of the Bayes factor, Jeffreys (1961) proposed a classification scheme that groups Bayes factors into different categories (see Table 1). For example, there is "strong" evidence in the data for hypothesis H_t, relative to hypothesis $H_{t'}$, when $B_{tt'} > 10$ and, equivalently, "strong" relative evidence in the data for hypothesis $H_{t'}$ when $B_{tt'} < 1/10$. This grouping provides verbal descriptions and rules of thumb when speaking of relative evidence in the data in favor of a

hypothesis, but it is still somewhat arbitrary. Ultimately, the interpretation of the magnitude of a Bayes factor should hinge on the context of the research question (Kass and Raftery 1995). For some introductory texts on Bayes factor testing in social science research, we refer the interested reader to Braeken et al. (2015), Raftery (1995), van de Schoot et al. (2011), or Wagenmakers (2007).

4.2. *Bayes Factor Computation*

In this section, we present efficient methods to compute marginal likelihoods and Bayes factors in higher-order network autocorrelation models. Using a multivariate normal prior for $\boldsymbol{\rho}_t$ under hypothesis H_t, $p_t(\boldsymbol{\rho}_t) = \phi_{\boldsymbol{\mu}_t, \Sigma_t}(\boldsymbol{\rho}_t) \mathbb{1}_{\Theta_{\rho_t}}(\boldsymbol{\rho}_t) c_t^{-1}$, $c_t := \int_{\Theta_{\rho_t}} \phi_{\boldsymbol{\mu}_t, \Sigma_t}(\boldsymbol{\rho}_t) d\boldsymbol{\rho}_t$, the noninformative prior $p(\sigma^2, \boldsymbol{\beta}) \propto 1/\sigma^2$ for the nuisance parameters σ^2 and $\boldsymbol{\beta}$, and after analytically integrating out σ^2 and $\boldsymbol{\beta}$, the Bayes factor of hypothesis H_t against hypothesis $H_{t'}$ in equation (14) reduces to

$$
\begin{aligned}
B_{tt'} &= \frac{m_t(\boldsymbol{y})}{m_{t'}(\boldsymbol{y})} \\
&= \frac{c_{t'} |2\pi\Sigma_t|^{-\frac{1}{2}}}{c_t |2\pi\Sigma_{t'}|^{-\frac{1}{2}}} \\
&\quad \frac{\int_{\Theta_{\rho_t}} |A_{\boldsymbol{\rho}_t}| \exp\left(-\frac{1}{2}(\boldsymbol{\rho}_t - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{\rho}_t - \boldsymbol{\mu}_t)\right) \boldsymbol{y}^T A_{\boldsymbol{\rho}_t}^T M A_{\boldsymbol{\rho}_t} \boldsymbol{y}^{-\frac{g-k}{2}} d\boldsymbol{\rho}_t}{\int_{\Theta_{\rho_{t'}}} |A_{\boldsymbol{\rho}_{t'}}| \exp\left(-\frac{1}{2}(\boldsymbol{\rho}_{t'} - \boldsymbol{\mu}_{t'})^T \Sigma_{t'}^{-1}(\boldsymbol{\rho}_{t'} - \boldsymbol{\mu}_{t'})\right) \boldsymbol{y}^T A_{\boldsymbol{\rho}_{t'}}^T M A_{\boldsymbol{\rho}_{t'}} \boldsymbol{y}^{-\frac{g-k}{2}} d\boldsymbol{\rho}_{t'}},
\end{aligned} \tag{15}
$$

where $M := I_g - X(X^T X)^{-1} X^T$.[9]

The normalizing constants c_t and $c_{t'}$ in equation (15) correspond to the prior probabilities that the unconstrained priors for $\boldsymbol{\rho}_t$ under hypothesis H_t and for $\boldsymbol{\rho}_{t'}$ under hypothesis $H_{t'}$, $N(\boldsymbol{\mu}_t, \Sigma_t)$ and $N(\boldsymbol{\mu}_{t'}, \Sigma_{t'})$, are in agreement with the constraints imposed under the two hypotheses. They can be approximated by simple rejection sampling, that is, by sampling draws from the unconstrained priors and recording the proportions of draws that are in agreement with the constraints. The remaining integrals in the numerator and denominator of equation (15) do not have closed-form solutions and have to be evaluated numerically. For this purpose, we rely on an *importance sampling* procedure (Owen and Zhou 2000) that is explained next.

Let $h_t(\boldsymbol{\rho}_t) := |A_{\boldsymbol{\rho}_t}| \exp\left(-\frac{1}{2}(\boldsymbol{\rho}_t - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{\rho}_t - \boldsymbol{\mu}_t)\right) \boldsymbol{y}^T A_{\boldsymbol{\rho}_t}^T M A_{\boldsymbol{\rho}_t} \boldsymbol{y}^{-\frac{g-k}{2}}$ denote the integrand in the numerator of equation (15) (all steps apply

equivalently to $h_{t'}(\boldsymbol{\rho}_{t'})$). Then, we can write for the numerator of equation (15)

$$
\begin{aligned}
I_t &:= \int_{\Theta_{\boldsymbol{\rho}_t}} h_t(\boldsymbol{\rho}_t)\mathrm{d}\boldsymbol{\rho}_t = \int_{\Theta_{\boldsymbol{\rho}_t}} q_t(\boldsymbol{\rho}_t)\frac{h_t(\boldsymbol{\rho}_t)}{q_t(\boldsymbol{\rho}_t)}\mathrm{d}\boldsymbol{\rho}_t = \mathbb{E}\left[\frac{h_t(\boldsymbol{P})}{q_t(\boldsymbol{P})}\right] \\
&\approx N^{-1}\sum_{i=1}^{N}\frac{h_t(\boldsymbol{\rho}_i)}{q_t(\boldsymbol{\rho}_i)} := \widehat{I}_t,
\end{aligned}
\tag{16}
$$

where $\boldsymbol{P}$ is a random variable with probability density function $q_t(\cdot)$ known as the *importance density*, $\mathbb{E}[h_t(\boldsymbol{P})/q_t(\boldsymbol{P})]$ denotes the expected value for $h_t(\boldsymbol{P})/q_t(\boldsymbol{P})$, and $\boldsymbol{\rho}_i$ are draws from $q_t(\cdot)$, forming realizations of $\boldsymbol{P}$. The specification of the importance density is crucial for the algorithm's efficiency, where we aim to construct a density that closely follows the actual integrand but has heavier tails than the latter and is easy to sample from (Owen and Zhou 2000).

As in Section 3.2, we approximate $\log(|A_{\boldsymbol{\rho}_t}|)$ by a second-order polynomial in $\boldsymbol{\rho}_t$ at its maximum, the origin. This results in a normal approximation of $|A_{\boldsymbol{\rho}_t}|$. We apply the same rationale to the third term in $h_t(\boldsymbol{\rho}_t)$, $\boldsymbol{y}^T A_{\boldsymbol{\rho}_t}^T M A_{\boldsymbol{\rho}_t} \boldsymbol{y}^{-(g-k)/2}$. Hence, $h_t(\boldsymbol{\rho}_t)$ can be approximated by the product of three multivariate normal densities that itself is a multivariate normal density, which we use as importance density in equation (16).[10] Finally, as $\boldsymbol{\rho}_t$ approaches the boundary of $\Theta_{\boldsymbol{\rho}_t}$, the proposed normal importance density has heavier tails than $h_t(\boldsymbol{\rho}_t)$, because in this case $|A_{\boldsymbol{\rho}_t}|$ decreases toward zero, but the normal importance density does not. This ensures a finite variance of the importance sampling estimate $\widehat{I}_t$ and reliable estimation of the associated Bayes factors. All details can be found in Appendix B.

4.3. *A Default Prior for* ρ

When testing multiple hypotheses against one another, a prior for the tested model parameters has to be specified under each hypothesis. Arguably, eliciting a prior under each hypothesis directly can become difficult and cumbersome, especially with a large number of hypotheses at hand. As an alternative, we propose an automatic *empirical Bayes* procedure (Carlin and Louis 2000) for constructing a default prior $p_t(\boldsymbol{\rho}_t)$ under each hypothesis H_t such that the marginal likelihood under every hypothesis H_t is maximized.

First, we center the multivariate normal default prior $p_t(\boldsymbol{\rho}_t)$ under hypothesis H_t around the origin. The motivation for this choice is that the origin is located at the boundary of typical (in)equality constrained hypotheses in the network autocorrelation model, such as $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, or $H_3 : \rho_1 = \rho_2 > 0$, and previous literature on order constrained hypothesis testing suggests "there is a gain of evidence for the inequality constrained hypothesis that is supported by the data when the unconstrained prior is located on the boundary" (Mulder 2014:452). Second, in contrast to Bayesian estimation, assigning very large values to the diagonal elements of the prior's covariance matrix Σ_t is not feasible in hypothesis testing. In hypothesis testing, we need to explicitly calculate the normalizing constant $c_t = \int_{\Theta_{\rho_t}} \phi_{\boldsymbol{\mu}_t, \Sigma_t}(\boldsymbol{\rho}_t) d\boldsymbol{\rho}_t$, and a vague formulation of Σ_t makes this computation either unstable or tremendously time consuming because of the fairly small parameter space $\Theta_{\boldsymbol{\rho}_t}$.[11] Instead, we set the prior covariance matrix Σ_t of the free network autocorrelation parameter(s) under a hypothesis, for example, ρ_1 under hypothesis $H_1 : \rho_1 > \rho_2 = 0$, to the product of the corresponding asymptotic variance-covariance matrix of the maximum likelihood estimate of $\boldsymbol{\rho}_t$ and a hypothesis-specific scaling factor σ_t^2, similar to Zellner's g-prior (Zellner 1986). In mathematical notation, $\Sigma_t = \sigma_t^2 I(\boldsymbol{\rho}_t)^{-1}$, where $I(\boldsymbol{\rho}_t)$ denotes the submatrix of the network autocorrelation model's Fisher information matrix $I(\boldsymbol{\rho}_t, \sigma^2, \boldsymbol{\beta})$. Hence, there is only one free parameter in the prior specification of Σ_t left, σ_t^2. Following Hansen and Yu (2001) and Liang et al. (2008), we use a local empirical Bayes approach and choose σ_t^2 such that the associated marginal likelihood $m_t(\boldsymbol{y})$ is maximized, avoiding arbitrary prior specification. Because there is no analytic solution to this maximization problem, one way to approximate the maximum of $m_t(\boldsymbol{y})$ is to compute the marginal likelihood on a grid of increasing values for σ_t^2 until a stopping rule is reached, for example, until the marginal likelihood is not increasing anymore, or until it is not increasing by more than some tolerance factor.[12]

Figure 3 shows the marginal likelihoods under the three constrained hypotheses $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, and $H_3 : \rho_1 = \rho_2 > 0$; the marginal likelihood under an unconstrained hypothesis $H_u : (\rho_1, \rho_2) \in \Theta_{(\rho_1, \rho_2)}$; and the logarithm of the Bayes factors of the three constrained hypotheses against the unconstrained hypothesis H_u as a function of σ_t^2, $t \in \{1, 2, 3, u\}$, for the data in Dall'erba et al. (2009) with model (4). All of the marginal likelihoods sharply increase for
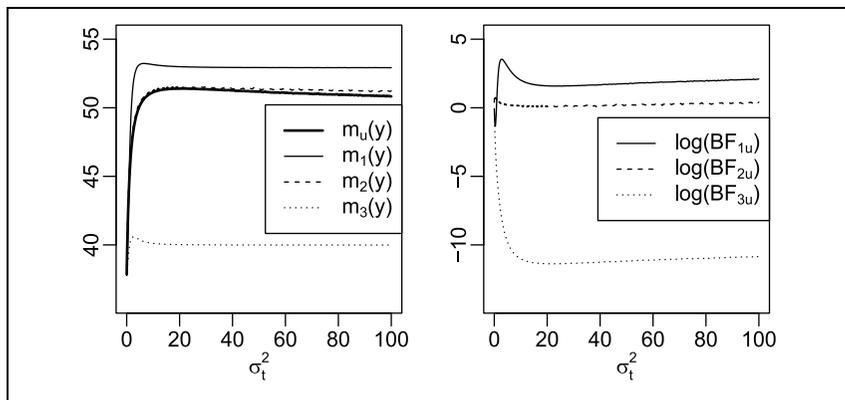
Figure 3. Marginal likelihoods $m_t(\mathbf{y})$, $t \in \{1, 2, 3, u\}$, under the hypotheses $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, $H_3 : \rho_1 = \rho_2 > 0$, and $H_u : (\rho_1, \rho_2) \in \Theta_{(\rho_1, \rho_2)}$ as a function of σ_t^2 (left), and the logarithm of the Bayes factors $\log(BF_{1u})$, $\log(BF_{2u})$, and $\log(BF_{3u})$ as a function of σ_t^2 (right) for the data in Dall'erba et al. (2009) with model (4).

smaller values for σ_t^2 before they gradually decrease after having reached their respective maxima. The associated Bayes factors, in which we are ultimately interested, appear fairly robust to the choice of σ_t^2, except for extremely small values for σ_t^2. For the vast majority of data sets we looked at, we observed essentially the same pattern, with almost all the optimal values for σ_t^2 lying between 2 and 10.

In summary, we showed how Bayes factors can be used to quantify the evidence in the data for hypotheses with order constraints on the network autocorrelation parameters. In addition, we provided methodology to efficiently compute such Bayes factors without any need to subjectively elicit priors for the network effects. This ultimately allows network scholars to test and verify any kind of expectations they have about the strength of different network effects. An R program that implements our methodology will be available at https://github.com/DittrichD/BayesNAM.

5. SIMULATION STUDY

We performed a simulation study to investigate the performance of the proposed Bayesian estimator and Bayes factors in a second-order network autocorrelation model. First, we compared the Bayesian estimator

from Section 3.2 with the maximum likelihood estimator in terms of bias of the network effects and frequentist coverage of the corresponding credible and confidence intervals. Here, we use the term *coverage* to indicate the proportion of times in which the true, that is, data-generating, network effects were contained in the credible and confidence intervals. Second, because researchers are generally interested in testing whether (some) network effects are zero or whether one network effect is larger than another, we considered a multiple hypothesis test with the following five hypotheses: $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, $H_3 : \rho_1 = \rho_2 > 0$, $H_4 : 0 < \rho_1 < \rho_2$, and $H_5 : 0 = \rho_1 < \rho_2$. We investigated if and how fast the different Bayes factors converge to a true data-generating hypothesis and how robust these findings are to various degrees of overlap between two connectivity matrices.

5.1. *Study Design*

In our simulation study, we generated data $\boldsymbol{y}$ via $\boldsymbol{y} = A_{(\rho_1, \rho_2)}^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}_g, I_g)$, for four network sizes g ($g \in \{50, 100, \ 200, 400\}$), three levels of overlap between W_1 and W_2 (0 percent, 20 percent, and 40 percent), and both W_1 and W_2 having an average degree of four. We simulated random nonsymmetric binary connectivity matrices using the rgraph() function from the sna package in R (Butts 2008), randomly rearranged ties when accounting for overlap, and subsequently row-standardized the raw connectivity matrices. Furthermore, we drew independent values from a standard normal distribution for the elements of $X \in \mathbb{R}^{g \times 4}$ (excluding the first column which is a vector of ones), $\boldsymbol{\beta} \in \mathbb{R}^4$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^g$.

In our first experiment, we set the two network effects to $(\rho_1, \rho_2) = (0.2, 0.2)$ and simulated 1,000 data sets for each of the 12 scenarios (four network sizes × three levels of overlap × one network effects size).[13] For the Bayesian estimator, we used the standard improper prior $p(\sigma^2, \boldsymbol{\beta}) \propto 1/\sigma^2$ for the nuisance parameters and a noninformative bivariate normal prior for (ρ_1, ρ_2), $p(\rho_1, \rho_2) \propto N(\boldsymbol{0}_2, 100 \times I_2)$, which essentially corresponds to a uniform prior for $(\rho_1, \rho_2) \in \Theta_{(\rho_1, \rho_2)}$. We drew 1,000 realizations from the resulting posteriors relying on the methods described in Section 3.2, taking the maximum likelihood estimate of $((\rho_1, \rho_2), \sigma^2, \boldsymbol{\beta})$ as the starting value in the sampling algorithm (see Appendix A). We used the marginal posterior median as point
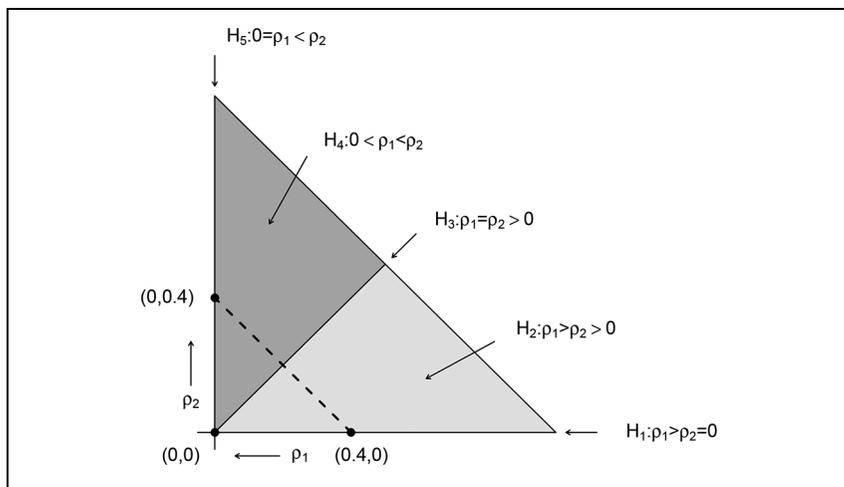
Figure 4. Admissible subspaces of (ρ_1, ρ_2) under the five constrained hypotheses and the trajectory of the data-generating network effects (dashed line).

estimator and the 95 percent equal-tailed credible interval for coverage analysis. We obtained the maximum likelihood estimates as well as their standard errors and associated asymptotic confidence intervals applying the lnam() function from the sna package in R.

In our second experiment, we considered 41 network effects sizes (ρ_1, ρ_2) $((\rho_1, \rho_2) \in \{(0.40, 0), (0.39, 0.01), \ldots, (0, 0.40)\})$ and simulated 100 data sets for each of the 492 scenarios (four network sizes $\times$ three levels of overlap $\times$ 41 network effects sizes). Figure 4 shows the trajectory of the network effects and depicts the five tested hypotheses, $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, $H_3 : \rho_1 = \rho_2 > 0$, $H_4 : 0 < \rho_1 < \rho_2$, and $H_5 : 0 = \rho_1 < \rho_2$. We specified the prior under each of the five hypotheses on the basis of the proposed empirical Bayes procedure in Section 4.3.[14] To compute the normalizing constants c_2 and c_4, we generated draws from the unconstrained bivariate normal prior for (ρ_1, ρ_2) until we obtained 1,000 draws in agreement with the constraints imposed under hypothesis H_2 and hypothesis H_4, respectively. Then, we approximated the normalizing constants by the reciprocals of the proportion of the total number of draws in agreement with the constraints. For the hypotheses with only one free network autocorrelation parameter, that is, $H_1 : \rho_1 > \rho_2 = 0$, $H_3 : \rho_1 = \rho_2 > 0$, and $H_5 : 0 = \rho_1 < \rho_2$, we directly obtained the corresponding normalizing constants by using the pnorm()

function in R, as the bounds of the feasible range of a single free network autocorrelation parameter are known exactly (see Section 2.1). Finally, for all hypotheses we drew 1,000 realizations from their (unconstrained) importance densities and computed the logarithm of the Bayes factor of each constrained hypothesis against an unconstrained reference hypothesis $H_u : (\rho_1, \rho_2) \in \Theta_{(\rho_1, \rho_2)}$.[15]

5.2. *Simulation Results*

Table 2 shows the average estimates and root mean squared errors of ρ_1 and ρ_2 for the Bayesian as well as the maximum likelihood estimator. Overall, the two estimators yield nearly identical results for all considered scenarios. As expected, the (negative) bias in the estimation of the network effects and the associated root mean squared errors are decreasing with the network size, the bias being virtually nonexistent for $g = 400$. Introducing 20 percent and 40 percent overlap between two connectivity matrices does not appear to affect the estimation results, even if there is mild negative correlation between the estimated network effects in these cases (see Figure 5).

Table 3 reports the empirical frequentist coverage of Bayesian equal-tailed 95 percent credible intervals and asymptotic 95 percent maximum likelihood-based confidence intervals for ρ_1 and ρ_2. The coverage of Bayesian credible intervals is very close to the nominal 0.95 for all considered scenarios, and the coverage of confidence intervals is below nominal for network sizes of 50. These observations are in line with the subpar coverage of maximum likelihood-based confidence intervals for small samples in the first-order network autocorrelation model reported by Dittrich et al. (2017a).

On the basis of results from our first simulation experiment, we draw two main conclusions. First, we recommend using the Bayesian estimator over the maximum likelihood estimator, as both estimators yield nearly identical network effect estimates, but the coverage of Bayesian credible intervals appears accurate, whereas for smaller network sizes the coverage of maximum likelihood-based confidence intervals is much less accurate. Second, estimating second-order network autocorrelation models with moderately overlapping connectivity matrices, that is, with up to 40 percent shared ties, does not affect estimation of the network effects. This second finding is of particular importance to social network

Table 2. Average Posterior Median Estimates and MLEs of $(\rho_1, \rho_2) = (0.2, 0.2)$ and Corresponding Average RMSEs for 1,000 Simulated Data Sets

	0 Percent Overlap				20 Percent Overlap				40 Percent Overlap			
	Estimate		RMSE		Estimate		RMSE		Estimate		RMSE	
	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2
$g = 50$												
Bayes	0.149	0.152	0.155	0.157	0.160	0.148	0.163	0.151	0.159	0.166	0.168	0.168
MLE	0.157	0.160	0.156	0.157	0.167	0.154	0.164	0.151	0.164	0.172	0.168	0.169
$g = 100$												
Bayes	0.180	0.182	0.107	0.104	0.178	0.184	0.111	0.107	0.188	0.183	0.107	0.107
MLE	0.182	0.184	0.108	0.104	0.179	0.186	0.111	0.107	0.189	0.185	0.108	0.108
$g = 200$												
Bayes	0.189	0.187	0.074	0.074	0.198	0.190	0.076	0.075	0.194	0.195	0.077	0.077
MLE	0.190	0.188	0.074	0.074	0.198	0.190	0.076	0.075	0.194	0.195	0.078	0.077
$g = 400$												
Bayes	0.196	0.196	0.052	0.051	0.197	0.197	0.050	0.051	0.198	0.196	0.054	0.053
MLE	0.196	0.196	0.052	0.051	0.197	0.197	0.050	0.051	0.198	0.196	0.053	0.053

Note: MLE = maximum likelihood estimate; RMSE = root mean squared error.
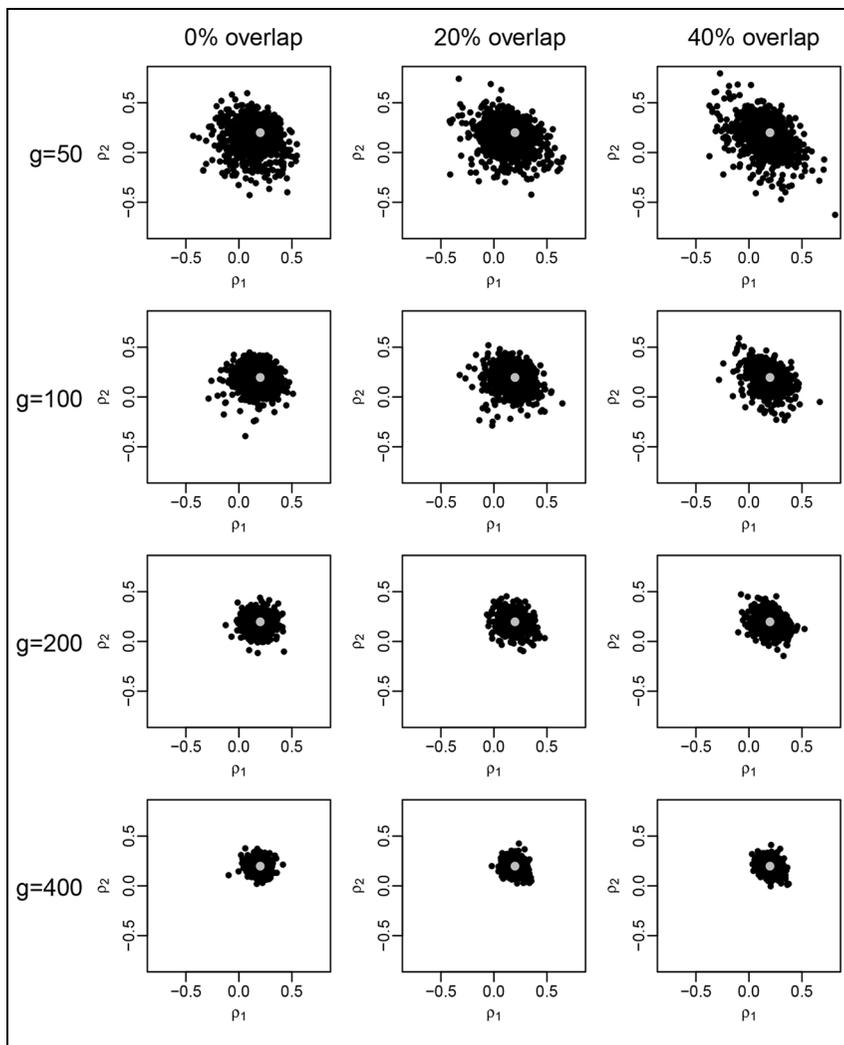
Figure 5. Posterior median estimates (black) of $(\rho_1, \rho_2) = (0.2, 0.2)$ (gray) for 1,000 simulated data sets.

researchers who often encounter distinct but partially overlapping networks in empirical practice.

Figure 6 displays the average logarithm of the Bayes factor of hypothesis H_1 (thick solid line), H_2 (thick dashed line), H_3 (dotted line), H_4 (dashed line), and H_5 (solid line) against an unconstrained reference

Table 3. Empirical Frequentist Coverage of 95 Percent Credible and Confidence Intervals for ρ_1 and ρ_2 for 1,000 Simulated Data Sets

	0 Percent Overlap		20 Percent Overlap		40 Percent Overlap	
	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2
$g = 50$						
Bayes	0.953	0.949	0.936	0.946	0.948	0.943
MLE	0.928	0.928	0.912	0.937	0.923	0.930
$g = 100$						
Bayes	0.949	0.958	0.949	0.946	0.957	0.961
MLE	0.936	0.950	0.942	0.936	0.951	0.955
$g = 200$						
Bayes	0.943	0.948	0.947	0.953	0.954	0.937
MLE	0.934	0.943	0.950	0.951	0.951	0.932
$g = 400$						
Bayes	0.954	0.948	0.964	0.940	0.953	0.943
MLE	0.955	0.949	0.964	0.946	0.953	0.943

Note: MLE = maximum likelihood estimate.

hypothesis, H_u, respectively, as a function of network effects (ρ_1, ρ_2) from $(0.40, 0)$ to $(0, 0.40)$. Overall, the results indicate that the Bayes factor shows consistent behavior, that is, there is the most evidence for the data-generating hypothesis if the network size is large enough. This evidence monotonically increases with network size. In particular, there is little discrimination between the five hypotheses for $g = 50$, whereas there is clear support for the data-generating hypothesis for network sizes of 200 and 400. Two lines in Figure 6 are discontinued for numerical reasons: when computing the Bayes factor, we need to calculate the probability mass of the unconstrained importance density contained in the parameter space imposed by the constraints under a hypothesis (see Appendix B). For the Bayes factors involving the purely inequality constrained hypotheses $H_2 : \rho_1 > \rho_2 > 0$ and $H_4 : 0 < \rho_1 < \rho_2$, we approximated these probabilities numerically by the proportion of 1,000 draws from the unconstrained importance densities that were in agreement with hypothesis H_2 and hypothesis H_4, respectively. For some data sets, however, none of the draws were in agreement with hypothesis H_2, or hypothesis H_4, in which case we set the corresponding marginal likelihood to $-\infty$. If this happened for at least one of the 100 simulated data sets, then the average logarithm of the Bayes factor was $-\infty$ as well. Finally, as in our first simulation experiment, these findings are robust to moderate degrees of overlap between two connectivity matrices.
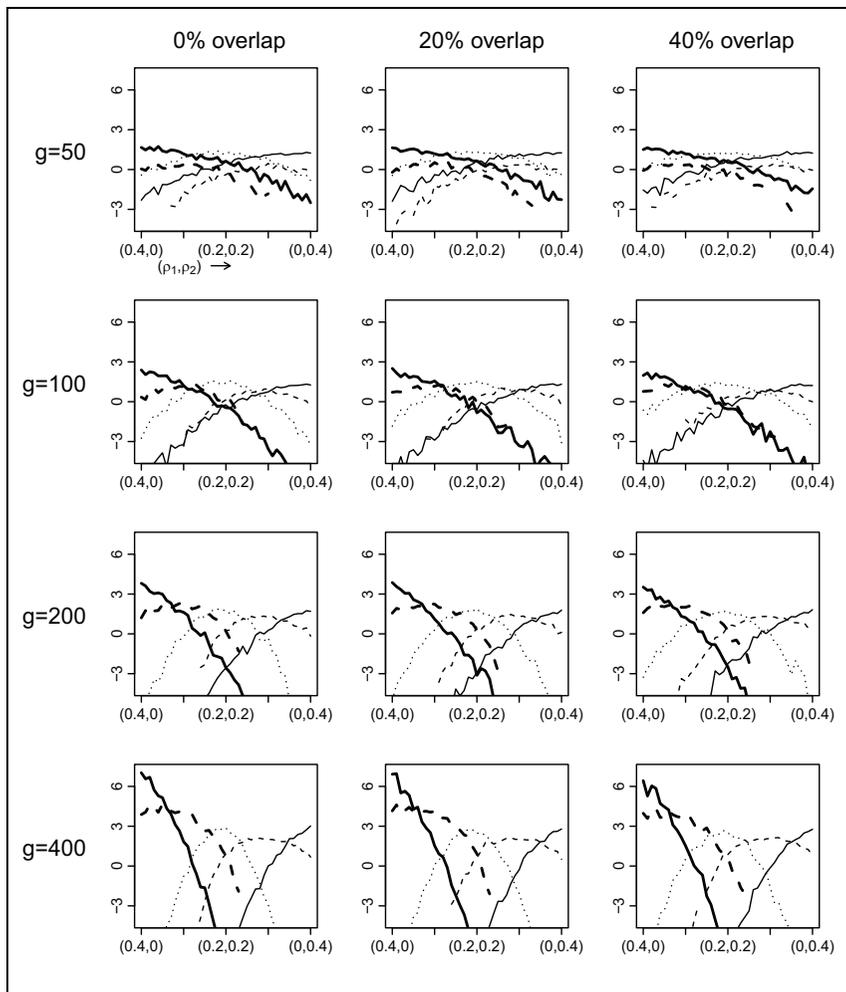
Figure 6. Average logarithm of the Bayes factors $\log(B_{tu})$, $t \in \{1, 2, 3, 4, 5\}$, of the hypotheses $H_1 : \rho_1 > \rho_2 = 0$ (thick solid line), $H_2 : \rho_1 > \rho_2 > 0$ (thick dashed line), $H_3 : \rho_1 = \rho_2 > 0$ (dotted line), $H_4 : 0 < \rho_1 < \rho_2$ (dashed line), and $H_5 : 0 = \rho_1 < \rho_2$ (solid line) against $H_u : (\rho_1, \rho_2) \in \Theta_{(\rho_1, \rho_2)}$ as a function of network effects (ρ_1, ρ_2) from $(0.40, 0)$ to $(0, 0.40)$ for 100 simulated data sets.

6. APPLICATION REVISITED

In this section, we reanalyze a data set from the economic growth litera-ture initially studied by Dall'erba et al. (2009) and address the questions

Table 4. Posterior Median Estimates and MLEs and Associated 95 Percent Bayesian Credible and Confidence Intervals (in Parentheses) for the Data in Dall'erba et al. (2009) with Model (4)

Parameter	Bayes	MLE
ρ_1	0.350	0.348
	$(0.238, 0.464)$	$(0.235, 0.460)$
ρ_2	-0.058	-0.058
	$(-0.169, 0.052)$	$(-0.168, 0.052)$
Intercept	-0.682	-0.696
	$(-0.887, -0.451)$	$(-0.924, -0.469)$
Market service growth	0.484	0.483
	$(0.384, 0.585)$	$(0.385, 0.580)$
Productivity gap	0.212	0.218
	$(0.122, 0.296)$	$(0.127, 0.309)$
Urbanization	3.094×10^{-5}	3.119×10^{-5}
	$(4.728 \times 10^{-6},$	$(5.894 \times 10^{-6},$
	$5.703 \times 10^{-5})$	$5.648 \times 10^{-5})$
Accessibility	1.052×10^{-5}	1.164×10^{-5}
	$(-3.937 \times 10^{-6},$	$(-3.132 \times 10^{-6},$
	$2.516 \times 10^{-5})$	$2.642 \times 10^{-5})$

Note: MLE = maximum likelihood estimate.

raised in Section 2.3. First, we reestimated the second-order network autocorrelation model in equation (4) on the basis of noninformative priors for all model parameters and compared the results with those coming from maximum likelihood estimation. Second, we used Bayes factors to quantify the relative evidence in the data for different competing hypotheses of interest with respect to this data set. Finally, we considered a network autocorrelation model with two subgroups, assuming only one dominant common influence mechanism within and between the two subgroups.

6.1. *Bayesian Estimation of a Second-Order Network Autocorrelation Model*

Table 4 displays the results of a Bayesian estimation of the second-order model in equation (4), along with the corresponding maximum likelihood estimates.[16] The Bayesian and the maximum likelihood estimates of all parameters are similar to each other, in line with results from our simulation study in Section 5.2. In particular, the (Bayesian) estimate of

Table 5. Bayes Factors $B_{tt'}$, $t, t' \in \{0, 1, 2, 3, c\}$, for Hypotheses $H_0 : \rho_1 = \rho_2 = 0$, $H_1 : \rho_1 > \rho_2 = 0$, $H_2 : \rho_1 > \rho_2 > 0$, $H_3 : \rho_1 = \rho_2 > 0$, and $H_c : \neg(H_0 \vee H_1 \vee H_2 \vee H_3)$ for the Data in Dall'erba et al. (2009) with Model (4)

Hypothesis	H_0	H_1	H_2	H_3	H_c
H_0	—	1.893×10^{-7}	3.043×10^{-5}	0.059	2.966×10^{-6}
H_1	5.283×10^6	—	160.746	3.117×10^5	14.085
H_2	3.286×10^4	6.220×10^{-3}	—	1.939×10^3	0.089
H_3	16.945	3.208×10^{-6}	5.157×10^{-4}	—	4.539×10^{-5}
H_c	3.732×10^5	0.071	11.359	2.203×10^4	—

ρ_1, reflecting interactions within the same country, is of large positive magnitude (0.350), and the (Bayesian) estimate of ρ_2, reflecting spillovers from regions in neighboring countries, is much smaller and close to zero (–0.058).[17] Dall'erba et al. (2009) concluded by saying that "the results obtained also confirm the hypothesis that economic interactions decrease very substantially when a national border is passed (indeed, the coefficient reflecting external spillovers is not statistically significant)" (p. 342).

6.2. *Bayesian Hypothesis Testing in a Second-Order Network Autocorrelation Model*

Using Bayes factors, we quantified the evidence in the data for two hypotheses representing the notion of decreasing economic interactions once a national border is passed, $H_1 : \rho_1 > \rho_2 = 0$ and $H_2 : \rho_1 > \rho_2 > 0$, and tested them against two competing hypotheses, $H_0 : \rho_1 = \rho_2 = 0$ and $H_3 : \rho_1 = \rho_2 > 0$.[18] We also included a hypothesis $H_c : \neg(H_0 \vee H_1 \vee H_2 \vee H_3)$ that represents the complement of all other possible hypotheses on (ρ_1, ρ_2) except hypotheses $H_0, H_1, H_2,$ and H_3; that is, hypothesis H_c contains all the orders of network effects we did not hypothesize.

Table 5 provides the Bayes factors for every pair out of the set of the five considered hypotheses above using the prior specifications from Sections 4.2 and 4.3. Notably, $H_1 : \rho_1 > \rho_2 = 0$ is the hypothesis most supported by the data; it is approximately 5.3×10^6, 160.7, 3.2×10^5, and 14.1 times more supported than hypothesis H_0, H_2, H_3, and H_c, respectively. Moreover, we see the least evidence in the data in favor of

the null. Consequently, regardless of the specification of alternative expectations about ρ_1 and ρ_2, the hypothesis that both network effects are zero has to be strongly rejected. Although these implications seem in line with the authors' claim that network effects decrease after a national border is passed, using Bayes factors provides us with much more extensive conclusions about the characteristic evidence in the data. Hence, we can now quantify how much more likely these conclusions are than competing conclusions (hypotheses) and how (un)likely it is that an entirely different mechanism H_c generated the data. Ultimately, this data set contains the most and very strong evidence for a positive within-country network effect only.

6.3. *Bayesian Hypothesis Testing in a Fourth-Order Network Autocorrelation Model*

Dall'erba et al. (2009) pointed to potentially asymmetric growth rates across the regions, depending on a region's initial productivity level. Thus, the authors proceeded by dividing the sample into two clusters: 111 initially more productive regions and 77 initially less productive regions, implying a core-periphery pattern (see Figure 7).[19] Next, they separately estimated two second-order network autocorrelation models for the two clusters. Here, for illustrative purposes, we allowed for varying levels of network autocorrelation within and between the two clusters and consider a model with two subgroups instead. For example, we could expect network effects within regions of the same subgroup to be larger than network effects between regions of different subgroups, or we could expect initially more productive regions to influence initially less productive ones more strongly than the other way around.

Our analyses in Section 6.2 suggest that there is very strong evidence in the data for a positive within-country network effect only; that is, $\rho_1 > 0$ and $\rho_2 = 0$. Thus, we merely considered spillover effects within the same country, in other words, we assume only W_1 plays a role, not W_2. We denote by $\rho_{hh}, \rho_{hl}, \rho_{lh}$, and ρ_{ll} the network effect within regions with initially higher productivity levels, the network effect of initially less productive regions on initially more productive regions, the network effect of initially more productive regions on initially less productive ones, and the network effect within regions with initially lower productivity levels, respectively. Accordingly, $\boldsymbol{y}_h \in \mathbb{R}^{111}$ and $\boldsymbol{y}_l \in \mathbb{R}^{77}$ contain the growth rates of labor productivity of the initially more and less
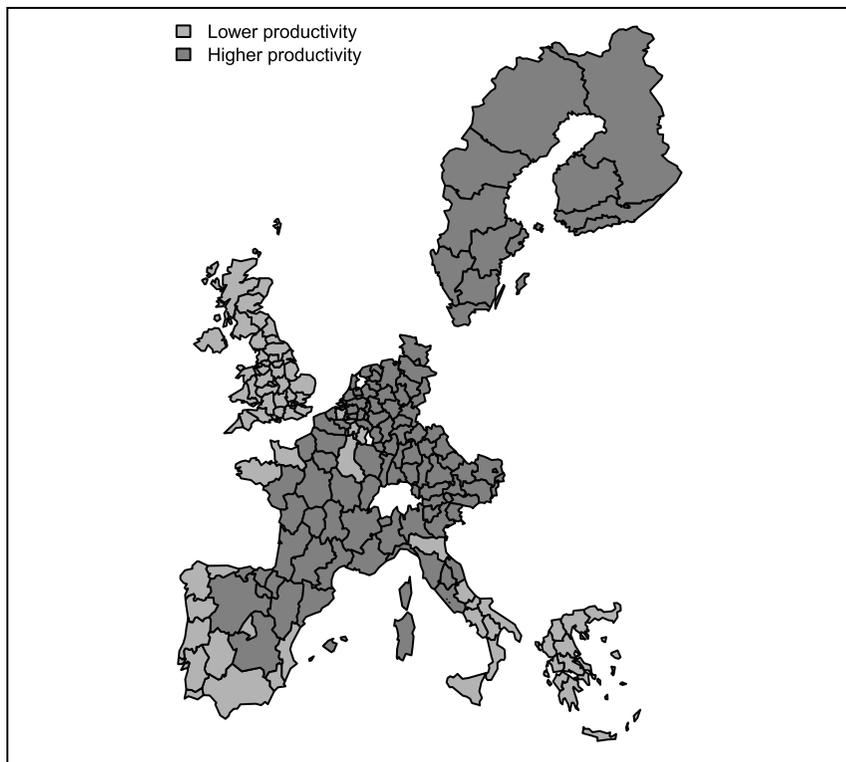
Figure 7. Spatial distribution of productivity levels in 1980 across the 188 regions.

productive regions, respectively, and we partitioned W_1, the unstandardized connectivity matrix using the region's three nearest neighbors within the same country, into the four submatrices $W_{hh} \in \mathbb{R}^{111 \times 111}$, $W_{hl} \in \mathbb{R}^{111 \times 77}$, $W_{lh} \in \mathbb{R}^{77 \times 111}$, and $W_{ll} \in \mathbb{R}^{77 \times 77}$, representing ties within and between the two subgroups.[20] This results in the following fourth-order network autocorrelation model

$$
\begin{aligned}
y = \begin{bmatrix} y_h \\ y_l \end{bmatrix} &= \begin{bmatrix} \rho_{hh} W_{hh} & \rho_{hl} W_{hl} \\ \rho_{lh} W_{lh} & \rho_{ll} W_{ll} \end{bmatrix} \begin{bmatrix} y_h \\ y_l \end{bmatrix} + \beta_1 X_{\cdot 1} + \beta_2 X_{\cdot 2} + \beta_3 X_{\cdot 3} + \beta_4 X_{\cdot 4} + \beta_5 X_{\cdot 5} + \varepsilon \\
&= \left(\rho_{hh} \begin{bmatrix} W_{hh} & 0 \\ 0 & 0 \end{bmatrix} + \rho_{hl} \begin{bmatrix} 0 & W_{hl} \\ 0 & 0 \end{bmatrix} + \rho_{lh} \begin{bmatrix} 0 & 0 \\ W_{lh} & 0 \end{bmatrix} + \rho_{ll} \begin{bmatrix} 0 & 0 \\ 0 & W_{ll} \end{bmatrix} \right) \begin{bmatrix} y_h \\ y_l \end{bmatrix} \\
&\quad + \beta_1 X_{\cdot 1} + \beta_2 X_{\cdot 2} + \beta_3 X_{\cdot 3} + \beta_4 X_{\cdot 4} + \beta_5 X_{\cdot 5} + \varepsilon.
\end{aligned}
$$

$$(17)$$

Table 6. Bayes Factors $B_{tt'}$, $t, t' \in \{0, 1, 2, 3, c\}$, for Hypotheses $H_0 : \rho_{hh} = \rho_{hl} = \rho_{lh} = \rho_{ll} = 0$, $H_1 : \{\rho_{hh} > \rho_{ll} > 0\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\}$, $H_2 : \{\rho_{hh} = \rho_{ll} > 0\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\}$, $H_3 : \{0 < \rho_{hh} < \rho_{ll}\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\}$, and $H_c : \neg(H_0 \vee H_1 \vee H_2 \vee H_3)$ for the Data in Dall'erba et al. (2009) with Model (17)

Hypothesis	H_0	H_1	H_2	H_3	H_c
H_0	—	1.023×10^{-5}	3.176×10^{-6}	1.177×10^{-5}	1.800×10^{-4}
H_1	9.773×10^4	—	0.310	1.151	17.241
H_2	3.149×10^5	3.222	—	3.704	55.556
H_3	8.497×10^4	0.869	0.270	—	14.925
H_c	5.563×10^3	0.058	0.018	0.067	—

We generally expect the network effects within the two subgroups to be larger than the network effects between subgroups, that is, $\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}$, where the ">" sign holds pairwise for any two elements of the first and second set, respectively. Furthermore, hypotheses of substantial interest might be based on expectations of positive network effects within both subgroups but with potentially differing magnitudes. We translated these expectations to three hypotheses, H_1, H_2, and H_3, and supplemented them with the hypothesis of no network effects and the complement of all the orders of network effects we did not have hypotheses for.[21] Formally,

$$H_0 : \rho_{hh} = \rho_{hl} = \rho_{lh} = \rho_{ll} = 0,$$
$$H_1 : \{\rho_{hh} > \rho_{ll} > 0\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\},$$
$$H_2 : \{\rho_{hh} = \rho_{ll} > 0\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\},$$
$$H_3 : \{0 < \rho_{hh} < \rho_{ll}\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\},$$
$$H_c : \neg(H_0 \vee H_1 \vee H_2 \vee H_3).$$

Table 6 shows the Bayes factors for every pair out of the set of the five considered hypotheses. We see that $H_2 : \{\rho_{hh} = \rho_{ll} > 0\} \wedge \{\{\rho_{hh}, \rho_{ll}\} > \{\rho_{hl}, \rho_{lh}\}\}$ is the hypothesis most supported by the data; it receives approximately 3.1×10^5, 3.2, 3.7, and 55.6 times more support than do hypothesis H_0, H_1, H_3, and H_c, respectively. Hence, there is no evidence in the data for differing network effects within the initially more and less productive regions, but there is very strong evidence that network effects within the two subgroups are larger than network effects between subgroups.

7. CONCLUSIONS

In this article, we developed Bayesian techniques for estimating and, primarily, testing higher-order network autocorrelation models with multiple network autocorrelations. In particular, we provided default Bayes factors that enable researchers to test hypotheses with order constraints on the network effects in a direct manner. The proposed methods allow researchers to simultaneously test any number of competing hypotheses on the relative strength of network effects against one another and to quantify the amount of evidence in the data for each hypothesis. This has not yet been possible using currently available statistical techniques for network autocorrelation models. Our proposed methods can straightforwardly be extended to test hypotheses on network autocorrelation parameters in heteroskedastic network autocorrelation models (LeSage 1997a) and network disturbances models (Leenders 2002).

We ran a simulation study to evaluate the numerical behavior of the presented Bayesian procedures for a number of different network specifications, including varying network sizes and network overlap. Our simulation study showed that, first, the Bayesian estimator and the maximum likelihood estimator yield similar parameter estimates of the network autocorrelation parameters for all scenarios. This was expected because we relied on noninformative priors for the network autocorrelation parameters. As a next step, it would be interesting to explore the use of (weakly) informative priors for multiple network autocorrelation parameters. Such priors can either be derived from published estimates in previous literature (similar to what Dittrich et al. [2017a] did for the first-order network autocorrelation model) or by eliciting experts on anticipated network effects in a given case study. Given previous findings in the literature when using a weakly informative prior for estimating a single network autocorrelation parameter (Dittrich et al. 2017a), we expect a carefully specified weakly informative prior for multiple network autocorrelation parameters will also decrease the negative bias of network autocorrelation parameters associated with maximum likelihood estimation of the model. Second, the Bayesian estimator exhibits nominal coverage of credible intervals and is more accurate than the maximum likelihood estimator, which is a strong argument in favor of the Bayesian approach, even when noninformative priors are used. Third, we found that the proposed Bayes factors always result in the

largest evidence for the true data-generating hypothesis, with this evidence increasing further with network size. In other words, the proposed Bayes factors tend to point researchers to the correct (or best) hypothesis out of a set of competing hypotheses; these hypotheses can represent highly complex relations between the autocorrelation parameters, and the set of hypotheses tested against each other simultaneously can, in principle, be arbitrarily large.

The practical tools needed to perform the methods developed in this article are, or will be, freely accessible in an R package. Given the many, often implicit, expectations researchers have about the relative importance of different network effects, we hope that by enabling researchers to test these expectations directly and explicitly, higher-order network autocorrelation models will bring forth a more thorough understanding of social contagion processes that goes beyond the current state of the art.

APPENDIX A: POSTERIOR SAMPLING

We outlined the procedure for sampling from the full posterior $p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}|\boldsymbol{y})$ in higher-order network autocorrelation models in Section 3.2. However, the exact form of the candidate-generating distribution for the conditional posterior $p(\boldsymbol{\rho}, \beta_1|\sigma^2, \tilde{\boldsymbol{\beta}}, \boldsymbol{y})$ and the expressions $\boldsymbol{\mu}_{\tilde{\beta}}$ and $\Sigma_{\tilde{\beta}}$ in equation (12) remain to be specified.

A.1. *Approximating the Conditional Posterior for* $(\boldsymbol{\rho}, \beta_1)$

In the following, we show how to approximate $p(\boldsymbol{\rho}, \beta_1|\sigma^2, \tilde{\beta}, \boldsymbol{y})$ by an $(R+1)$-variate normal distribution. First, by Jacobi's formula (see, e.g., Hall 2003, Theorem 2.11), for any complex matrix X it holds that $\det(\exp(X)) = |\exp(X)| = \exp(\mathrm{tr}(X))$. Because we set the R-dimensional parameter space of $\boldsymbol{\rho}$ as the space containing the origin for which $A_{\boldsymbol{\rho}}$ is nonsingular, we know that $|A_{\boldsymbol{\rho}}| > 0$, $A_{\boldsymbol{\rho}}$ is invertible, and $\log(A_{\boldsymbol{\rho}})$ exists (see, e.g., Higham 2008, Theorem 1.27). Thus, we can write for $X := A_{\boldsymbol{\rho}} = I_g - \sum_{r=1}^{R} \rho_r W_r$,

$$
\begin{aligned}
&|\exp(A_{\boldsymbol{\rho}})| = \exp(\mathrm{tr}(A_{\boldsymbol{\rho}})) \\
\Leftrightarrow\ &|\exp(\log(A_{\boldsymbol{\rho}}))| = \exp(\mathrm{tr}(\log(A_{\boldsymbol{\rho}}))) \\
\Leftrightarrow\ &\log(|A_{\boldsymbol{\rho}}|) = \mathrm{tr}(\log(A_{\boldsymbol{\rho}})).
\end{aligned}
\tag{A1}
$$

Using the Mercator series for the matrix logarithm (see, e.g., Hall 2003, Theorem 2.7), we can rewrite equation (A1) as

$$
\begin{aligned}
\log\left(|A_{\boldsymbol{\rho}}|\right) = \operatorname{tr}\left(\log(A_{\boldsymbol{\rho}})\right) &= \operatorname{tr}\left(\sum_{m=1}^{\infty} (-1)^{m+1} \frac{(A_{\boldsymbol{\rho}} - I_g)^m}{m}\right) \\
&= \sum_{m=1}^{\infty} (-1)^{m+1} \frac{1}{m} \operatorname{tr}\left(\left(-\sum_{r=1}^{R} \rho_r W_r\right)^m\right).
\end{aligned}
\tag{A2}
$$

The first two sum terms in equation (A2) are given by

$$
m=1: \quad (-1)^2 \operatorname{tr}\left(-\sum_{r=1}^{R} \rho_r W_r\right) = -\sum_{r=1}^{R} \rho_r \operatorname{tr}(W_r) = 0,
$$

$$
m=2: \quad (-1)^3 \frac{1}{2} \operatorname{tr}\left(\left(-\sum_{r=1}^{R} \rho_r W_r\right)^2\right) = -\frac{1}{2} \sum_{r,r'=1}^{R} \rho_r \rho_{r'} \operatorname{tr}(W_r W_{r'}).
$$

Hence, $\log\left(|A_{\boldsymbol{\rho}}|\right)$ can be approximated by a quadratic polynomial in $\boldsymbol{\rho}$ as

$$
\begin{aligned}
\log\left(|A_{\boldsymbol{\rho}}|\right) &\approx -\frac{1}{2} \sum_{r,r'=1}^{R} \rho_r \rho_{r'} \operatorname{tr}(W_r W_{r'}), \\
\Rightarrow |A_{\boldsymbol{\rho}}| &\approx \exp\left(-\frac{1}{2} \sum_{r,r'=1}^{R} \rho_r \rho_{r'} \operatorname{tr}(W_r W_{r'})\right).
\end{aligned}
\tag{A3}
$$

Second, after some algebraic manipulation,

$$
\exp\left(-\frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\rho} - \boldsymbol{\mu})\right) \propto \exp\left(-\frac{1}{2}\left(\sum_{r,r'=1}^{R} \rho_r \rho_{r'} \Sigma_{rr'}^{-1} - 2\sum_{r,r'=1}^{R} \rho_r \Sigma_{rr'}^{-1} \mu_{r'}\right)\right), \tag{A4}
$$

$$
\exp\left(-\frac{1}{2\sigma^2}(A_{\boldsymbol{\rho}} \boldsymbol{y} - X\boldsymbol{\beta})^T (A_{\boldsymbol{\rho}} \boldsymbol{y} - X\boldsymbol{\beta})\right)
$$

$$
\propto \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{r,r'=1}^{R} \rho_r \rho_{r'} \boldsymbol{y}^T W_r^T W_{r'} \boldsymbol{y} - 2\sum_{r=1}^{R} \rho_r \boldsymbol{y}^T W_r^T (\boldsymbol{y} - \tilde{X}\tilde{\boldsymbol{\beta}})\right.\right.
$$

$$
\left.\left. + 2\sum_{r=1}^{R} \rho_r \beta_1 \boldsymbol{y}^T W_r^T \mathbf{1}_g - 2\beta_1 \mathbf{1}_g^T (\boldsymbol{y} - \tilde{X}\tilde{\boldsymbol{\beta}}) + \beta_1^2 g\right)\right), \tag{A5}
$$

where the proportionality in equation (A4) holds with respect to $\boldsymbol{\rho}$, in equation (A5) with respect to $(\boldsymbol{\rho}, \beta_1)$, $\tilde{X}$ denotes the matrix X with its first column removed, $\tilde{\boldsymbol{\beta}} = (\beta_2, ..., \beta_k)$, and $\mathbf{1}_g$ is the vector of ones of length g. By equating the coefficients of the product of equations (A3), (A4), and (A5), that is, the

approximated conditional posterior for $(\boldsymbol{\rho}, \beta_1)$, to the kernel of the probability density function of a multivariate normal distribution $N(\boldsymbol{\mu}_{MH}, \Sigma_{MH})$, we obtain for the coefficients of Σ_{MH}^{-1}

$$\Sigma_{MH_{rr'}}^{-1} = \text{tr}(W_r W_{r'}) + \Sigma_{rr'}^{-1} + \boldsymbol{y}^T W_r^T W_{r'} \boldsymbol{y} / \sigma^2, \quad r, r' \in \{1, ..., R\},$$

$$\Sigma_{MH(R+1)r'}^{-1} = \Sigma_{MHr'(R+1)}^{-1} = \boldsymbol{y}^T W_{r'}^T \mathbf{1}_g / \sigma^2,$$

$$\Sigma_{MH(R+1)(R+1)}^{-1} = g / \sigma^2,$$

and $\boldsymbol{\mu}_{MH} = \Sigma_{MH} z_{MH}$, where z_{MH} is a vector of length $R+1$ with

$$z_{MH_r} = \boldsymbol{y}^T W_r^T \left(\boldsymbol{y} - \tilde{X}\tilde{\boldsymbol{\beta}} \right) / \sigma^2 + \sum_{r'=1}^{R} \Sigma_{rr'}^{-1} \boldsymbol{\mu},$$

$$z_{MH_{R+1}} = \mathbf{1}_g^T \left(\boldsymbol{y} - \tilde{X}\tilde{\boldsymbol{\beta}} \right) / \sigma^2.$$

If the $(R \times R)$ matrix T, $T_{rr'} := tr(W_r W_{r'})$, is not positive definite, Σ_{MH}^{-1}, and consequently Σ_{MH}, may not be positive definite either. In this case, we consider the nearest positive definite matrix to Σ_{MH}^{-1} as Σ_{MH}^{-1} instead. Subsequently, we use $q_{MH}(\boldsymbol{\rho}, \beta_1) = \phi_{\boldsymbol{\mu}_{MH}, \Sigma_{MH}}(\boldsymbol{\rho}, \beta_1)$ as candidate-generating density in the Metropolis-Hastings algorithm.

A.1.1. Sampling from the Conditional Posterior for $\tilde{\boldsymbol{\beta}}$. The $(k-1)$-variate normal conditional posterior for $\tilde{\boldsymbol{\beta}}$ in equation (12) can be directly sampled from, with its mean vector and covariance matrix given by $\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}} = \boldsymbol{\mu}_{\boldsymbol{\beta}_2} + \Sigma_{\boldsymbol{\beta}_{21}} \Sigma_{\boldsymbol{\beta}_{11}}^{-1} \left(\beta_1 - \mu_{\beta_1} \right)$ and $\Sigma_{\tilde{\boldsymbol{\beta}}} = \Sigma_{\boldsymbol{\beta}_{22}} - \Sigma_{\boldsymbol{\beta}_{21}} \Sigma_{\boldsymbol{\beta}_{11}}^{-1} \Sigma_{\boldsymbol{\beta}_{12}}$, where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = (X^T X)^{-1} X^T A_{\boldsymbol{\rho}} \boldsymbol{y} = \begin{pmatrix} \mu_{\beta_1} \\ \boldsymbol{\mu}_{\boldsymbol{\beta}_2} \end{pmatrix} \quad \text{sized} \begin{pmatrix} 1 \times 1 \\ (k-1) \times 1 \end{pmatrix}, \tag{A6}$$

$$\Sigma_{\boldsymbol{\beta}} = \sigma^2 (X^T X)^{-1} = \begin{pmatrix} \Sigma_{\boldsymbol{\beta}_{11}} & \Sigma_{\boldsymbol{\beta}_{12}} \\ \Sigma_{\boldsymbol{\beta}_{21}} & \Sigma_{\boldsymbol{\beta}_{22}} \end{pmatrix} \text{sized} \begin{pmatrix} 1 \times 1 & 1 \times (k-1) \\ (k-1) \times 1 & (k-1) \times (k-1) \end{pmatrix}. \tag{A7}$$

A.1.2. Sampling Algorithm. The full sampling algorithm for drawing from the posterior $p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta} | \boldsymbol{y})$ in higher-order network autocorrelation models can be written as follows:

1. Set starting values $\left(\boldsymbol{\rho}^0, \beta_1^0 \right), (\sigma^2)^0$, and $\tilde{\boldsymbol{\beta}}^0$, for example, to their maximum likelihood estimates, and the number of draws N.
2. Repeat steps (3) to (5) for $i = 1 : N$.

3. Perform a Metropolis-Hastings step for $(\boldsymbol{\rho}, \beta_1)$ with the target density $p(\boldsymbol{\rho}, \beta_1 | \sigma^2, \tilde{\boldsymbol{\beta}}, \boldsymbol{y})$ and the candidate-generating density $q_{MH}(\boldsymbol{\rho}, \beta_1)$, that is,

- Draw from $q_{MH}(\boldsymbol{\rho}, \beta_1)$ until a draw $(\hat{\boldsymbol{\rho}}, \hat{\beta}_1)$ satisfies $(\hat{\boldsymbol{\rho}}, \hat{\beta}_1) \in \Theta_{\boldsymbol{\rho}} \times \mathbb{R}$. Draw u from the uniform distribution $U(0,1)$.
- Calculate the acceptance probability $\alpha\big[(\boldsymbol{\rho}^{i-1}, \beta_1^{i-1}), (\hat{\boldsymbol{\rho}}, \hat{\beta}_1)\big]$, defined as

$$
\alpha\big[(\boldsymbol{\rho}^{i-1}, \beta_1^{i-1}), (\hat{\boldsymbol{\rho}}, \hat{\beta}_1)\big] :=
$$
$$
\min\left(\frac{p\big(\hat{\boldsymbol{\rho}}, \hat{\beta}_1 | (\sigma^2)^{i-1}, \tilde{\boldsymbol{\beta}}^{i-1}, \boldsymbol{y}\big) q_{MH}(\boldsymbol{\rho}^{i-1}, \beta_1^{i-1})}{p\big(\boldsymbol{\rho}^{i-1}, \beta_1^{i-1} | (\sigma^2)^{i-1}, \tilde{\boldsymbol{\beta}}^{i-1}, \boldsymbol{y}\big) q_{MH}(\hat{\boldsymbol{\rho}}, \hat{\beta}_1)}, 1\right).
$$

- If $u \leq \alpha\big[(\boldsymbol{\rho}^{i-1}, \beta_1^{i-1}), (\hat{\boldsymbol{\rho}}, \hat{\beta}_1)\big]$, set $(\boldsymbol{\rho}^i, \beta_1^i) := (\hat{\boldsymbol{\rho}}, \hat{\beta}_1)$.
- Else, set $(\boldsymbol{\rho}^i, \beta_1^i) := (\boldsymbol{\rho}^{i-1}, \beta_1^{i-1})$.

4. Draw $(\sigma^2)^i$, given $(\boldsymbol{\rho}^i, \beta_1^i)$ and $\tilde{\boldsymbol{\beta}}^{i-1}$, from the inverse gamma distribution in equation (11).
5. Draw $\tilde{\boldsymbol{\beta}}^i$, given $(\boldsymbol{\rho}^i, \beta_1^i)$ and $(\sigma^2)^i$, from a $(k-1)$-variate normal distribution with mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}$ and covariance matrix $\Sigma_{\tilde{\boldsymbol{\beta}}}$ as in equations (A6) and (A7).

The implementation of the above algorithm is straightforward, but its performance might severely worsen when dealing with very large networks, that is, consisting of thousands of actors, as it requires evaluation of the determinants in the conditional posteriors in each Metropolis-Hastings step. One potential remedy in such cases is to use the exchange algorithm (Murray et al. 2006) instead, as in Han et al. (2017), which adds auxiliary draws to the initial algorithm but avoids calculation of the determinants in the Metropolis-Hastings steps and thereby reduces the overall computational burden.

APPENDIX B: BAYES FACTOR COMPUTATION

In the following, we show how the integral $I_t = \int_{\Theta_{\rho_t}} h_t(\boldsymbol{\rho}_t) \mathrm{d}\boldsymbol{\rho}_t$ in equation (16) can be effectively approximated by its importance sampling estimate $\widehat{I}_t$,

$$\widehat{I}_t = N^{-1} \sum_{i=1}^{N} \frac{h_t(\boldsymbol{\rho}_i)}{q_t(\boldsymbol{\rho}_i)}$$

$$= N^{-1} \sum_{i=1}^{N} \frac{\left|A_{\boldsymbol{\rho}_i}\right| \exp\left(-\frac{1}{2}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)\right) \boldsymbol{y}^T A_{\boldsymbol{\rho}_i}^T M A_{\boldsymbol{\rho}_i} \boldsymbol{y}^{-\frac{g-k}{2}}}{q_t(\boldsymbol{\rho}_i)}, \qquad (B1)$$

where $\boldsymbol{\rho}_i$ are draws from a suitable importance density $q_t(\cdot)$. We specify $q_t(\cdot)$ such that it closely follows the integrand $h_t(\boldsymbol{\rho}_t)$ but has heavier tails than the latter, which ensures a reliable estimation of I_t.

As in Appendix A, we approximate $\log(|A_{\boldsymbol{\rho}_t}|)$ by a quadratic polynomial in $\boldsymbol{\rho}_t$ at its maximum value, the origin. This results in a normal approximation of $|A_{\boldsymbol{\rho}_t}|$, that is, $|A_{\boldsymbol{\rho}_t}| \approx N(\mathbf{0}_R, T^{-1})$, where $T_{rr'} := \mathrm{tr}(W_r W_{r'})$, $r, r' \in \{1, ..., R\}$. In the case that T is not positive definite, we use the nearest positive definite matrix to T instead. The second term in the denominator of equation (B1) already equals the kernel of the probability density function of the normal distribution $N(\boldsymbol{\mu}_t, \Sigma_t)$. Finally, we also approximate the logarithm of the third term in $h_t(\boldsymbol{\rho}_t)$ by a second-order Taylor polynomial at its maximum. It follows that $\boldsymbol{y}^T A_{\boldsymbol{\rho}_t}^T M A_{\boldsymbol{\rho}_t} \boldsymbol{y}^{-(g-k)/2} \approx N(\boldsymbol{\mu}_3, \Sigma_3)$, where $\boldsymbol{\mu}_3 = (\boldsymbol{y}^T W^T M W \boldsymbol{y})^{-1} \boldsymbol{y}^T M W \boldsymbol{y}$, $\Sigma_3 = (\boldsymbol{y}^T W^T M W \boldsymbol{y})^{-1} (\boldsymbol{y}^T A_{\boldsymbol{\rho}_t}^T M A_{\boldsymbol{\rho}} \boldsymbol{y})/(g-k)$ and $(\boldsymbol{y}^T W^T M W \boldsymbol{y})_{rr'} := \boldsymbol{y}^T W_r^T M W_{r'} \boldsymbol{y}$, $(\boldsymbol{y}^T M W \boldsymbol{y})_r := \boldsymbol{y}^T M W_r \boldsymbol{y}$. Thus, $h_t(\boldsymbol{\rho}_t)$ can be approximated by the product of three multivariate normal densities that is multivariate normal itself, so $q_t(\boldsymbol{\rho}_t) := \phi_{\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t}}(\boldsymbol{\rho}_t) \mathbb{1}_{\Theta_{\boldsymbol{\rho}_t}}(\boldsymbol{\rho}_t) c_{IS_t}^{-1}, c_{IS_t} := \int_{\Theta_{\boldsymbol{\rho}_t}} \phi_{\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t}}(\boldsymbol{\rho}_t) d\boldsymbol{\rho}_t$, with $\Sigma_{IS_t} = \left(T + \Sigma_t^{-1} + \Sigma_3^{-1}\right)^{-1}$, $\boldsymbol{\mu}_{IS_t} = \Sigma_{IS_t}\left(\Sigma_t^{-1} \boldsymbol{\mu}_t + \Sigma_3^{-1} \boldsymbol{\mu}_3\right)$.

Calculating $\widehat{I}_t$ directly might result in underflow in R, which is why we next show how to compute its logarithm only. We can write

$$\log\left(\widehat{I}_t\right) = \log\left(N^{-1} \sum_{i=1}^{N} \frac{h_t(\boldsymbol{\rho}_i)}{q_t(\boldsymbol{\rho}_i)}\right) = -\log(N) + \log\left(\sum_{i=1}^{N} \frac{h_t(\boldsymbol{\rho}_i)}{q_t(\boldsymbol{\rho}_i)}\right)$$

$$= -\log(N) + \log\left(\sum_{i=1}^{N} \exp\left(\log\left(\frac{h_t(\boldsymbol{\rho}_i)}{q_t(\boldsymbol{\rho}_i)}\right) + d - d\right)\right)$$

$$= -\log(N) - d + \log\left(\sum_{i=1}^{N} \exp\left(\log\left(\frac{h_t(\boldsymbol{\rho}_i)}{q_t(\boldsymbol{\rho}_i)}\right) + d\right)\right)$$

$$= -\log(N) - d + \log\left(\sum_{i=1}^{N} \exp\left(\log(|A_{\boldsymbol{\rho}_i}|) - \frac{g-k}{2}\log\left(\boldsymbol{y}^T A_{\boldsymbol{\rho}_i}^T M A_{\boldsymbol{\rho}_i} \boldsymbol{y}\right)\right.\right.$$

$$\left.\left. -\frac{1}{2}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)\right.\right.$$

$$- \log\left(\left(\frac{\phi_{\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t}}(\boldsymbol{\rho}_i)\mathbb{1}_{\Theta_{\boldsymbol{\rho}_i}}(\boldsymbol{\rho}_i)}{c_{IS_t}}\right) + d\right)\right)$$

$$\approx - \log(N) - d + \log\left(\sum_{i=1}^{N} \exp\left(\log(|A_{\boldsymbol{\rho}_i}|) - \frac{g-k}{2}\log\left(\boldsymbol{y}^T A_{\boldsymbol{\rho}_i}^T MA_{\boldsymbol{\rho}_i}\boldsymbol{y}\right)\right.\right.$$

$$-\frac{1}{2}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)$$

$$\left.\left. - \log\left(\frac{\phi_{\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t}}(\boldsymbol{\rho}_i)\mathbb{1}_{\Theta_{\boldsymbol{\rho}_t}}(\boldsymbol{\rho}_i)}{\frac{\sum_{i=1}^{N}\mathbb{1}_{\Theta_{\boldsymbol{\rho}_t}}(\boldsymbol{\rho}_i)}{N}}\right) + d\right)\right)$$

$$= - 2\log(N) + \log\left(\sum_{i=1}^{N}\mathbb{1}_{\Theta_{\boldsymbol{\rho}_t}}(\boldsymbol{\rho}_i)\right) - d$$

$$+ \log\left(\sum_{i=1}^{N}\exp\left(\log(|A_{\boldsymbol{\rho}_i}|) - \frac{g-k}{2}\log\left(\boldsymbol{y}^T A_{\boldsymbol{\rho}_i}^T MA_{\boldsymbol{\rho}_i}\boldsymbol{y}\right)\right.\right.$$

$$-\frac{1}{2}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{\rho}_i - \boldsymbol{\mu}_t)$$

$$\left.\left. - \log\left(\phi_{\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t}}(\boldsymbol{\rho}_i)\mathbb{1}_{\Theta_{\boldsymbol{\rho}_t}}(\boldsymbol{\rho}_i)\right) + d\right)\right),$$

where $\boldsymbol{\rho}_i$ are draws from the unconstrained importance density $N(\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t})$ and d is an auxiliary constant, for example, $d = -\frac{g-k}{2}\min\limits_{i \in \{1, ..., N\}}\left(\boldsymbol{y}^T A_{\boldsymbol{\rho}_i}^T MA_{\boldsymbol{\rho}_i}\boldsymbol{y}\right)$, which is added to prevent the marginal likelihood to become too small to be distinguished from zero in R. The auxiliary constant d is set in advance after generating the N draws from the unconstrained importance density $N(\boldsymbol{\mu}_{IS_t}, \Sigma_{IS_t})$ first.

ORCID iD

Dino Dittrich 🄳 https://orcid.org/0000-0002-1672-1684

Notes

1. Except for Leenders (1995), the literature on the network autocorrelation model ignores the occurrence of potentially complex eigenvalues that can arise when considering nonsymmetric connectivity matrices. If W is row-standardized, it follows that $Re(\lambda_1[W]) = 1$. Finally, as $\det(A_\rho) > 0$ for all $\rho \in \Theta_\rho$, we simply write $|A_\rho|$ for $|\det(A_\rho)|$ in the following.

2. Within this framework, it is also possible to model network influence attributable to different types of ties, for example, by defining connectivity matrices that contain only mutual ties or asymmetric ties. We thank an anonymous reviewer for this suggestion.

3. In Elhorst et al. (2012:213), the term $\tan(\alpha)/r_{\max}[W^*]$ needs to be replaced by $\tan(\alpha)/r_{\min}[W^*]$ in equation (15).

4. If a researcher relies on an informative prior for the network autocorrelation parameters, we advise the researcher to also perform a prior sensitivity analysis, in which the estimates are compared with those based on using a vague prior instead. If the network effects estimates are close to one another, this implies the data contain sufficient information and the estimates are most likely highly reliable; otherwise, this strongly points at (negative) bias in the estimation of the network effects. We thank an anonymous reviewer for this suggestion.

5. The posterior $p(\boldsymbol{\rho}, \sigma^2, \boldsymbol{\beta}|y)$ in equation (9) is proper given very mild regularity conditions. The proof for higher-order models is quasi-identical to and an adaptation of the one for the first-order model in Dittrich et al. (2017a).

6. This is iteratively repeated N times for a large number N. Geman and Geman (1984) showed that as $N \to \infty$, the draws based on the sequence of conditional posteriors can be seen as samples from the actual marginal posteriors, that is, the posteriors for a parameter (block) given the data, for example, $p(\boldsymbol{\rho}|y)$.

7. It can (rarely) happen that, after equating coefficients, the obtained covariance matrix of the normal candidate-generating distribution is not positive definite, as the Hessian in the second-order approximation of $\log(|A_{\boldsymbol{\rho}}|)$ itself is not always positive definite, for example, for $W_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $W_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Thus, if the initially obtained matrix is not positive definite, we instead use its nearest positive definite one as the normal candidate-generating distribution's covariance matrix. This can be done using the nearPD() function from the Matrix package in R (Bates and Maechler 2017).

8. For hypothesis H_1, R_1^E and r_1^E imply $0 \times \rho_1 + 1 \times \rho_2 = 0$, and R_1^I and r_1^I lead to $1 \times \rho_1 + 0 \times \rho_2 > 0$. Together, this constitutes hypothesis H_1. For hypothesis H_2, we have $1 \times \rho_1 - 1 \times \rho_2 > 0$ and $0 \times \rho_1 + 1 \times \rho_2 > 0$. Analogously, for hypothesis H_3, $1 \times \rho_1 - 1 \times \rho_2 = 0$ and $1 \times \rho_1 + 0 \times \rho_2 > 0$.

9. We can use an improper prior for the nuisance parameters σ^2 and $\boldsymbol{\beta}$, $p(\sigma^2, \boldsymbol{\beta}) \propto 1/\sigma^2$, as they appear in both hypothesis H_t and hypothesis $H_{t'}$ and the

corresponding normalizing constants cancel out after integrating out σ^2 and $\boldsymbol{\beta}$ (Hepple 1995a). Note that depending on the specification of hypothesis H_t and hypothesis $H_{t'}$, the dimensions of Σ_t and $\Sigma_{t'}$ may differ, for example, for $H_1 : \rho_1 > \rho_2 = 0$ and $H_2 : \rho_1 > \rho_2 > 0$.

10. As in Section 3.2, if the resulting covariance matrix of the normal distribution approximating $|A_{\boldsymbol{\rho}_t}|$ is not positive definite, we use its nearest positive definite matrix as covariance matrix instead.

11. Because $\Theta_{\boldsymbol{\rho}_t}$ is bounded and small, Bartlett's paradox (Bartlett 1957) is not an issue in the considered tests on the network autocorrelations.

12. The marginal likelihood $m_t(\boldsymbol{y})$ can be strictly increasing with σ_t^2, which is why we cannot use more efficient optimization techniques, such as Newton's method or the BFGS algorithm (Nocedal and Wright 2006).

13. Simulation results for different combinations of (ρ_1, ρ_2) and nonrandom specifications of the connectivity matrices are available from the authors upon request. We do not present them here as they do not provide any additional insights.

14. We used the lnam() function from the sna package in R to obtain the asymptotic variance-covariance matrices of the maximum likelihood estimates of the free network autocorrelation parameters. We started with $\sigma_t^2 = 1$, $t \in \{1, 2, 3, 4, 5\}$, and kept increasing σ_t^2 by 1 until either the increment in the marginal likelihood $m_t(\boldsymbol{y})$ was less than 0.01 or σ_t^2 reached the cutoff value 20.

15. For the hypotheses with only one free network autocorrelation parameter, that is, $H_1 : \rho_1 > \rho_2 = 0$, $H_3 : \rho_1 = \rho_2 > 0$, and $H_5 : 0 = \rho_1 < \rho_2$, we directly generated 1,000 draws in agreement with the respective constraints by using the rtruncnorm() function from the truncnorm package in R (Trautmann et al. 2015). As the unconstrained hypothesis H_u only serves as the reference hypothesis with which all other hypotheses are compared, we did not maximize over σ_u^2 when computing $m_u(\boldsymbol{y})$ but fixed σ_u^2 to 5.

16. As in our simulation study in Section 5, we used the prior $p(\rho_1, \rho_2, \sigma^2, \boldsymbol{\beta}) \propto N(\mathbf{0}_2, 100 \times I_2) \times 1/\sigma^2$ and relied on the lnam() function from the sna package in R to compute the maximum likelihood estimates and corresponding confidence intervals.

17. To truly be able to speak of a large positive magnitude of ρ_1, we advise researchers to consider the posterior predictive distribution of the data and compare it with the one based on setting ρ_1 to zero, *ceteris paribus*. Alternatively, researchers can use indirect effects as another measure for the magnitude of a network effect (Elhorst 2010). We thank an anonymous reviewer for this suggestion.

18. We considered the hypothesis $H_2 : \rho_1 > \rho_2 > 0$ rather than a hypothesis $H_{2'} : \rho_1 > \rho_2$ here, as the hypotheses H_1 and $H_{2'}$ would not be unambiguous. This does not mean hypothesis $H_{2'}$ in itself cannot be tested against other hypotheses, only the combination of hypotheses in a given test has to be considered carefully.

19. We created the map of the European NUTS-2 regions using Eurostat data from http://ec.europa.eu/eurostat/cache/GISCO/geodatafiles/NUTS_2010_60M_SH.zip, file NUTS_RG_60M_2010.shp, and the readOGR() function from the rgdal package in R (Bivand, Keitt, and Rowlingson 2017). The map shows 189 instead of the original 188 regions, as the Trentino-Alto Adige region in Italy has been divided into two NUTS-2 regions.

20. We row-standardized W_{hh}, W_{hl}, W_{lh}, and W_{ll} separately.
21. Another hypothesis of interest could be that (with respect to the growth of labor productivity) the initially more productive regions influence the initially less productive ones stronger than the other way around, that is, $\rho_{lh} > \rho_{hl}$. We did not include this hypothesis, as it would overlap with the other considered hypotheses and because of the undesirable behavior of such overlapping-hypotheses Bayes factors (Morey and Rouder 2011). Separate analyses show, however, that there is no evidence in the data for a hypothesis involving $\rho_{lh} > \rho_{hl}$.

References

Anselin, Luc. 2001. "Rao's Score Test in Spatial Econometrics." *Journal of Statistical Planning and Inference* 97(1):113–39.

Badinger, Harald, and Peter Egger. 2013. "Estimation and Testing of Higher-Order Spatial Autoregressive Panel Data Error Component Models." *Journal of Geographical Systems* 15(4):453–89.

Bartlett, M. S. 1957. "A Comment on D. V. Lindley's Statistical Paradox." *Biometrika* 44(3–4):533–34.

Bates, Douglas, and Martin Maechler. 2017. "Matrix: Sparse and Dense Matrix Classes and Methods." Retrieved March 9, 2020. http://CRAN.R-project.org/package= Matrix.

Beck, Nathaniel, Kristian S. Gleditsch, and Kyle Beardsley. 2006. "Space Is More Than Geography: Using Spatial Econometrics in the Study of Political Economy." *International Studies Quarterly* 50(1):27–44.

Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2017. "rgdal: Bindings for the 'Geospatial' Data Abstraction Library." Retrieved March 9, 2020. http://CRAN.R-project.org/package=rgdal.

Böing-Messing, Florian, Marcel A. van Assen, Abe D. Hofman, Herbert Hoijtink, and Joris Mulder. 2017. "Bayesian Evaluation of Constrained Hypotheses on Variances of Multiple Independent Groups." *Psychological Methods* 22(2):262–87.

Braeken, Johan, Joris Mulder, and Stephen Wood. 2015. "Relative Effects at Work: Bayes Factors for Order Hypotheses." *Journal of Management* 41(2):544–73.

Burt, Ronald S., and Patrick Doreian. 1982. "Testing a Structural Model of Perception: Conformity and Deviance with Respect to Journal Norms in Elite Sociological Methodology." *Quality and Quantity* 16(2):109–50.

Butts, Carter T. 2008. "Social Network Analysis with sna." *Journal of Statistical Software* 24(6):1–51.

Carlin, Bradley C., and Thomas A. Louis. 2000. "Empirical Bayes: Past, Present and Future." *Journal of the American Statistical Association* 95(452):1286–89.

Chib, Siddharta, and Edward Greenberg. 1994. "Bayes Inference in Regression Models with ARMA (p, q) Errors." *Journal of Econometrics* 64(1–2):183–206.

Chib, Siddharta, and Edward Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *American Statistician* 49(4):327–35.

Chib, Siddharta, and Edward Greenberg. 1998. "Analysis of Multivariate Probit Models." *Biometrika* 85(2):347–61.

Dall'erba, Sandy, Marco Percoco, and Gianfranco Piras. 2009. "Service Industry and Cumulative Growth in the Regions of Europe." *Entrepreneurship and Regional Development* 21(4):333–49.

Dittrich, Dino, Roger Th. Leenders, and Joris Mulder. 2017a. "Bayesian Estimation of the Network Autocorrelation Model." *Social Networks* 48:213–36.

Dittrich, Dino, Roger Th. Leenders, and Joris Mulder. 2017b. "Network Autocorrelation Modeling: A Bayes Factor Approach for Testing (Multiple) Precise and Interval Hypotheses." *Sociological Methods and Research* 48(3):642–76.

Doreian, Patrick. 1981. "Estimating Linear Models with Spatially Distributed Data." *Sociological Methodology* 12:359–88.

Elhorst, J. Paul. 2010. "Applied Spatial Econometrics: Raising the Bar." *Spatial Economic Analysis* 5(1):9–28.

Elhorst, J. Paul, Donald J. Lacombe, and Gianfranco Piras. 2012. "On Model Specification and Parameter Space Definitions in Higher Order Spatial Econometric Models." *Regional Science and Urban Economics* 42(1–2):211–20.

Fujimoto, Kayo, Chih-Ping Chou, and Thomas W. Valente. 2011. "The Network Autocorrelation Model Using Two-Mode Data: Affiliation Exposure and Potential Bias in the Autocorrelation Parameter." *Social Networks* 33(3):231–43.

Gelfand, Alan E., and Adrian F. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85(410):398–409.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2003. *Bayesian Data Analysis*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC Press.

Geman, Stuart, and Donald Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6):721–41.

Gimpel, James G., and Jason E. Schuknecht. 2003. "Political Participation and the Accessibility of the Ballot Box." *Political Geography* 22(5):471–88.

Gupta, Abhimanyu, and Peter M. Robinson. 2015. "Inference on Higher-Order Spatial Autoregressive Models with Increasingly Many Parameters." *Journal of Econometrics* 186(1):19–31.

Hall, Brian C. 2003. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. New York: Springer.

Han, Xiaoyi, Chih-Sheng Hsieh, and Lung-Fei Lee. 2017. "Estimation and Model Selection of Higher-Order Spatial Autoregressive Model: An Efficient Bayesian Approach." *Regional Science and Urban Economics* 63:97–120.

Hansen, Mark H., and Bin Yu. 2001. "Model Selection and the Principle of Minimum Description Length." *Journal of the American Statistical Association* 96(454):746–74.

Hastings, W. K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57(1):97–109.

Hepple, Leslie W. 1995a. "Bayesian Techniques in Spatial and Network Econometrics: 1. Model Comparison and Posterior Odds." *Environment and Planning A: Economy and Space* 27(3):447–69.

Hepple, Leslie W. 1995b. "Bayesian Techniques in Spatial and Network Econometrics: 2. Computational Methods and Algorithms." *Environment and Planning A: Economy and Space* 27(4):615–44.

Higham, Nicholas J. 2008. *Functions of Matrices: Theory and Computation*. Philadelphia: Society for Industrial and Applied Mathematics.

Holloway, Garth, Bhavani Shankar, and Sanzidur Rahman. 2002. "Bayesian Spatial Probit Estimation: A Primer and an Application to HYV Rice Adoption." *Agricultural Economics* 27(3):383–402.

Jeffreys, Harold. 1961. *Theory of Probability*. 3rd ed. Oxford, UK: Oxford University Press.

Kalenkoski, Charlene M., and Donald J. Lacombe. 2008. "Effects of Minimum Wages on Youth Employment: The Importance of Accounting for Spatial Correlation." *Journal of Labor Research* 29(4):303–17.

Kass, Robert E., and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90(430):773–95.

Klugkist, Irene, Olav Laudy, and Herbert Hoijtink. 2005. "Inequality Constrained Analysis of Variance: A Bayesian Approach." *Psychological Methods* 10(4):477–93.

Lacombe, Donald J. 2004. "Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques." *Geographical Analysis* 36(2):105–18.

Lee, Lung-Fei, and Xiaodong Liu. 2010. "Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances." *Econometric Theory* 26(1):187–230.

Leenders, Roger Th. 1995. *Structure and Influence: Statistical Models for the Dynamics of Actor Attributes, Network Structure and Their Interdependence*. Amsterdam, the Netherlands: Thela Thesis.

Leenders, Roger Th. 2002. "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix." *Social Networks* 24(1):21–47.

LeSage, James P. 1997a. "Bayesian Estimation of Spatial Autoregressive Models." *International Regional Science Review* 20(1–2):113–29.

LeSage, James P. 1997b. "Regression Analysis of Spatial Data." *Journal of Regional Analysis and Policy* 27(2):83–94.

LeSage, James P., and R. Kelley Pace. 2007. "A Matrix Exponential Spatial Specification." *Journal of Econometrics* 140(1):190–214.

LeSage, James P., and R. Kelley Pace. 2008. "Spatial Econometric Modeling of Origin-Destination Flows." *Journal of Regional Science* 48(5):941–67.

LeSage, James P., and R. Kelley Pace. 2011. "Pitfalls in Higher Order Model Extensions of Basic Spatial Regression Methodology." *Review of Regional Studies* 41(1):13–26.

LeSage, James P., and Olivier Parent. 2007. "Bayesian Model Averaging for Spatial Econometric Models." *Geographical Analysis* 39(3):241–67.

Liang, Feng, Rui Paulo, German Molina, Merlise A. Clyde, and Jim O. Berger. 2008. "Mixtures of g Priors for Bayesian Variable Selection." *Journal of the American Statistical Association* 103(481):410–23.

Lin, Tse-Min, Chin-En Wu, and Feng-Yu Lee. 2006. "'Neighborhood' Influence on the Formation of National Identity in Taiwan: Spatial Regression with Disjoint Neighborhoods." *Political Research Quarterly* 59(1):35–46.

McMillen, Daniel P., Larry D. Singell Jr., and Glen R. Waddell. 2007. "Spatial Competition and the Price of College." *Economic Inquiry* 45(4):817–33.

McPherson, Michael A., and Michael L. Nieswiadomy. 2005. "Environmental Kuznets Curve: Threatened Species and Spatial Effects." *Ecological Economics* 55(3): 395–407.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. "Equations of State Calculations by Fast Computing Machines." *Journal of Chemical Physics* 21(6):1087–92.

Mizruchi, Mark S., and Linda B. Stearns. 2006. "The Conditional Nature of Embeddedness: A Study of Borrowing by Large U.S. Firms, 1973–1994." *American Sociological Review* 71(2):310–33.

Morey, Richard D., and Jeffrey N. Rouder. 2011. "Bayes Factor Approaches for Testing Interval Null Hypotheses." *Psychological Methods* 16(4):406–19.

Mulder, Joris. 2014. "Prior Adjusted Default Bayes Factors for Testing (In)equality Constrained Hypotheses." *Computational Statistics and Data Analysis* 71:448–63.

Mulder, Joris. 2016. "Bayes Factors for Testing Order-Constrained Hypotheses on Correlations." *Journal of Mathematical Psychology* 72:104–15.

Mulder, Joris, and Jean-Paul Fox. 2018. "Bayes Factor Testing of Multiple Intraclass Correlations." *Bayesian Analysis* 14(2):521–52.

Mulder, Joris, Herbert Hoijtink, and Irene Klugkist. 2010. "Equality and Inequality Constrained Multivariate Linear Models: Objective Model Selection Using Constrained Posterior Priors." *Journal of Statistical Planning and Inference* 140(4): 887–906.

Mulder, Joris, and Eric-Jan Wagenmakers. 2016. "Editors' Introduction to the Special Issue 'Bayes Factors for Testing Hypotheses in Psychological Research: Practical Relevance and New Developments.'" *Journal of Mathematical Psychology* 72:1–5.

Mur, Jesús, Fernando López, and Ana Angulo. 2008. "Symptoms of Instability in Models of Spatial Dependence." *Geographical Analysis* 40(2):189–211.

Murray, Iain, Zoubin Ghahramani, and David J. MacKay. 2006. "MCMC for Doubly-Intractable Distributions." Pp. 359–66 in *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*. Arlington, VA: AUAI Press.

Neuman, Eric J., and Mark S. Mizruchi. 2010. "Structure and Bias in the Network Autocorrelation Model." *Social Networks* 32(4):290–300.

Nocedal, Jorge, and Stephen J. Wright. 2006. *Numerical Optimization*. 2nd ed. New York: Springer.

Ord, Keith. 1975. "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70(349):120–26.

Owen, Art, and Yi Zhou. 2000. "Safe and Effective Importance Sampling." *Journal of the American Statistical Association* 95(449):135–43.

R Core Team. 2017. "R: A Language and Environment for Statistical Computing." Retrieved March 9, 2020. http://www.R-project.org/.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–63.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* 92(437):179–91.

Smith, Tony E. 2009. "Estimation Bias in Spatial Models with Strongly Connected Weight Matrices." *Geographical Analysis* 41(3):307–32.

Tita, George E., and Steven M. Radil. 2011. "Spatializing the Social Networks of Gangs to Explore Patterns of Violence." *Journal of Quantitative Criminology* 27(4):521–45.

Trautmann, Heike, Detlef Steuer, Olaf Mersmann, and Björn Bornkamp. 2015. "truncnorm: Truncated Normal Distribution." Retrieved March 9, 2020. http://CRAN.R-project.org/package=truncnorm.

van de Schoot, Rens, Joris Mulder, Herbert Hoijtink, Marcel A. van Aken, Judith S. Dubas, Bram Orobio de Castro, Wim Meeus, and Jan-Willem Romeijn. 2011. "An Introduction to Bayesian Model Selection for Evaluating Informative Hypotheses." *European Journal of Developmental Psychology* 8(6):713–29.

Wagenmakers, Eric-Jan. 2007. "A Practical Solution to the Pervasive Problems of *p* Values." *Psychonomic Bulletin and Review* 14(5):779–804.

Zellner, Arnold. 1986. "On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions." Pp. 233–43 in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, edited by P. K. Goel, and A. Zellner. Amsterdam, the Netherlands: North-Holland.

Zhang, Bin, Andrew C. Thomas, Patrick Doreian, David Krackhardt, and Ramayya Krishnan. 2013. "Contrasting Multiple Social Network Autocorrelations for Binary Outcomes, with Applications to Technology Adoption." *ACM Transactions on Management Information Systems* 3(4):18:1–21.

Author Biographies

Dino Dittrich holds a PhD in statistics from Tilburg University. His work centers on Bayesian estimation and hypothesis-testing techniques for social network models. Currently, he works as a data scientist at Health Care Systems GmbH.

Roger Th. A. J. Leenders is a professor at the Jheronimus Academy of Data Science and in the Department of Organization Studies at Tilburg University. He holds a PhD in sociology from the University of Groningen. He has published broadly on social network analysis, teams, innovation, and organization behavior in leading journals such as *Organization Science*, the *Journal of Applied Psychology*, the *Journal of Product Innovation Management, Social Networks*, and the *Academy of Management Journal*.

Joris Mulder is an associate professor in the Department of Methodology and Statistics at Tilburg University. He holds a PhD in applied Bayesian statistics from Utrecht University. His research focuses on Bayesian model selection and social network modeling.