Networks in Space

Spillovers Effects in a Hierarchical Model

Nikolas Kuschnig*

Preliminary Draft

February 19, 2024

Networks and spillover effects between agents are ubiquitous in theory and practice, but network data remains scarce and models suffer from the curse of dimensionality. As a result, empirical analyses rely on proxies and strong assumptions. In this paper, I propose a hierarchical approach to jointly model spillovers and the latent networks behind them. The approach flexibly leverages information where available, and imposes structure and shrinkage where suitable, allowing for more nuanced, less restrictive assumptions. It is widely applicable to spillovers in social and spatial contexts, even with no, limited, or uncertain network information. I provide nuanced priors, novel identification results, and new and efficient sampling procedures to allow for effective estimation and the quantification of uncertainty. I demonstrate the approach with practical applications to deforestation in Brazil.

Keywords: spatial spillovers, peer effects, network formation, latent space

*Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria; correspondence to nikolas.kuschnig@wu.ac.at. Part of the research presented here was conducted at Johns Hopkins University and the University of California, Berkeley.

1 Introduction

Economic theory frequently suggests connections between individual units of observation. Connectivity and the resulting spillover effects between units lie at the heart of pressing research questions in microeconomics (Ambrus et al., 2014; Chyn and Katz, 2021; Ioannides and Datcher Loury, 2004; Weidmann and Deming, 2021) and macroeconomics (Acemoglu, Akcigit, et al., 2016; Crespo Cuaresma et al., 2019; Gofman, 2017; Rose, 2004). Good answers to these questions and deeper insights can be important for education policy (Board and Meyer-ter-Vehn, 2021; Lin, 2010; Mele, 2020), labor market outcomes (Beaman, 2012; Hensvik and Skans, 2016; Munshi, 2003), supply chain management (Acemoglu, Carvalho, et al., 2012; Atalay, Hortaçsu, and Syverson, 2014; Atalay, Hortaçsu, Roberts, et al., 2011; Kranton and Minehart, 2001), innovation (Bloom et al., 2013; Ductor et al., 2014; König et al., 2019; Newman, 2001). However, empirical studies of spillover effects abstract from the nature of connections and modeling approaches are limited. Current econometric methods for analyzing spillover effects generally rely on holding the structure of connectivity fixed, thus obscuring uncertainties and potentially distorting results.

This paper introduces an integrated model to jointly analyze both the structure and consequences of connectivity between units. I propose a Bayesian hierarchical approach to comprehensively address both issues. Suitable shrinkage priors allow for more nuanced assumptions; prior information is flexibly imposed where available and needed, while important aspects of the model are freed up and learned from the data. The resulting model facilitates the explicit treatment of connectivity structures and spillovers in a layered framework. As I demonstrate with an application to deforestation spillovers from croplands in the Brazilian Amazon, this model avoids bias from misspecified connectivities, better reflecting reality and the surrounding uncertainties.

The main contribution of this paper is a framework for jointly estimating spillover/peer effects and connectivity structures, i.e., spatial weight or adjacency matrices. A number of earlier studies pursue similar goals in the spatial econometric (Debarsy and LeSage, 2020; Lam and Souza, 2020; Qu and L. Lee, 2015; Zhang and Yu, 2018) and the network econometric (Goldsmith-Pinkham and Imbens, 2013; Hsieh and L. F. Lee, 2016; Johnsson and Moon, 2021) literature. These studies generally adopt model averaging approaches, and are limited to few discrete candidates for the connectivity structure. The hierarchical setup that I propose in this paper eliminates this constraint. It provides a flexible and nuanced way of modeling and estimating more general forms of connectivity, conveying a more comprehensive picture of spillover effects. For this approach to work in practice, there are three main obstacles to overcome.

The first challenge lies in the forms of the connectivity structure themselves. Here, I provide a unified framework for common forms of connectivity used in both the spatial and network econometric literatures. I characterize connectivity as a weighted digraph, and develop a functional approximation that nests many common connectivity structures. This approximation places individual units in a metric space, augmented with a notion of potential, and highlights the strong simplifying assumptions made in the literature. Lastly, I derive conclusive bounds for the autoregressive parameter of overall connectivity strength, ensuring non-singularity and stationarity of the model. This allows for more flexible models that don't impose a row-stochastic form a priori, and, as I demonstrate, avoid bias from this potential misspecification of connectivity.

The second challenge is related to the curse of dimensionality. I address these using suitable prior distributions that can induce regularization. With few exceptions (e.g. Lam and Souza, 2020; de Paula et al., 2023), the notion of regularization is neglected in the literature, and even Bayesian approaches generally use flat priors. I show how established (expressed and implicit) priors are limited in their flexibility, impose strong information in unsuspected dimensions, and often cannot accommodate the nuanced prior information that is available. I introduce priors that can reflect this prior information, e.g., location-based priors for the structural parameters in the connectivity function. For the autoregressive parameter of overall connectivity strength, I propose a Beta-Gamma mixture prior that can accommodate flexible shapes and provides sensible regularization without distorting estimates. These priors facilitate the efficient estimation of connectivity parameters that have previously been fixed.

The third challenge is of a technical nature and concerns estimation of the model, and the computations involved. Posterior inference relies on Markov chain Monte Carlo (MCMC) sampling or variational methods, and the interdependence that results from connectivity can make even simple models computationally prohibitive. Models with an autoregressive term, in particular, rely on the evaluation of a costly Jacobian determinant. In this paper, I focus on full posterior inference, and develop efficient sampling schemes for the hierarchical and structural parameters, which facilitate the straightforward estimation of extensible models. For such models, the Jacobian determinant features at least one additional dimension, making established procedures obsolete. In order to still allow for rapid and accurate estimation, I develop an adaptive Gaussian process approximation. These optimizations allow for full Bayesian inference in spatial econometric models that can be extended with little overhead.

The remainder of this paper is structured as follows.

2 Methods

Let $\mathscr{A} = \{1, \dots, N\}$ be a finite set of agents, for whom we observe some *response* $Y \in \mathbb{R}$ and a vector of *characteristics* $X \in \mathbb{R}^P$ a total of T times. For clarity and ease of notation, we will assume that $T = 1$ unless mentioned otherwise. We want to learn about the relationship between the response and the characteristics, $Y = f(X) + \varepsilon$, but suspect that agents are not independent of each other. Instead, they are connected through a set of *links* and form a *network*. To conduct inference, we need to impose structure on the functional form of the relationship and the network.

2.1 Framework

To effectively model the network, we formalize it by assuming that it can be represented as a graph. Specifically, by the *weighted digraph* $\mathscr{G} = \left(\mathscr{A}, \mathscr{E}, g\right)$, where the links (or *edges*) $\mathscr{E}$ are induced by the *link function*

$$g : \mathscr{A} \times \mathscr{A} \mapsto \mathbb{R}, \tag{1}$$

meaning that $g(i, j) = 0$ if and only if $\left(i, j\right) \notin \mathscr{E}$. This function is unknown, and will be the target of our modeling efforts. We do not allow for self-links, i.e., $g(i, i) = 0$ for all $i \in \mathscr{A}$, and assume that there is one unique link between individual agents in this formulation. This assumption can be prohibitive, so we informally introduce the generalization g_s, which is specific to some situation s.

Adjacency matrix. The graph $\mathscr{G}$ corresponds to an *adjacency matrix* $\mathbf{G}$, which is defined in terms of the link function, i.e., $g_{ij} := g(i, j)$. This allows us to express the graph in a familiar form. The *characteristic polynomial* of the graph is that of the adjacency matrix, and provides information on the structure of the graph. In practice, we work with a *normalized* adjacency matrix, which we denote as $\mathbf{W}$ and will define later on. We use a subscript, such as $\mathbf{W}_s$, to indicate extensions of these definitions to the situation-specific variant.

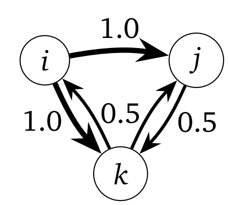

Figure 1: Example of a weighted digraph with three agents $\{i, j, k\}$.

2.1.1 Linear network model

We consider

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{W} \mathbf{X} \boldsymbol{\theta} + \mathbf{X} \boldsymbol{\beta} + \mathbf{e}, \tag{2}$$

for the functional form of f. This model extends the classical linear model with two network regressors — $\mathbf{Wy}, \mathbf{WX}$. These relate to the responses and characteristics of linked agents, and are constructed by pre-multiplication with the normalized adjacency matrix. They give rise to two different types of spillover effects, i.e. effects of an agent i on another agent $j \neq i$. These types become apparent an alternative formulation,

$$\mathbf{y} = (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{z}, \tag{3}$$

$$\mathbf{z} = \mathbf{WX}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \tag{4}$$

Here, we introduce a latent variable Z that allows us to decompose the model in a *network filter* and a *nested linear model*. The filter in Equation 3 captures the coordination behind the equilibrium response of linked agents, and results in *global* spillovers between all connected agents. The latent variable itself arises from the linear model Equation 4, which captures *local* spillovers from the characteristics of directly connected agents. Global spillovers are also referred to as endogenous peer effects, while local spillovers are also known as exogenous or contextual (peer) effects.

The linear network model, as introduced here, reflects our earlier assumption of a single, unique link between agents. To highlight this assumption, consider

$$\mathbf{y}_t = \lambda_t \mathbf{W}_{t,y} \mathbf{y}_t + \mathbf{W}_{t,x_1} \mathbf{x}_{t,1} \theta_{t,1} + \cdots + \mathbf{W}_{t,x_p} \mathbf{x}_{t,x_p} \theta_{t,x_p} + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{e}_t,$$

where we allow for links to be time- and regressor-specific.

2.2 Hierarchical network model

Models of our link function suffer the *curse of dimensionality*, as the number of unknown links increases with the square of the number of observed agents. This means that we require additional information, either from observed data or in the form of structural assumptions, to model the $\mathcal{O}(N^2)$ links in Equation 1.

Latent space. Let $\mathcal{S} = (\mathcal{P}, w)$ be a *latent space*, where the set $\mathcal{P}$ is equipped with a notion of *similarity*, which is measured by the *similarity function* w. One way to view this function is in relation to some *metric*, in which it is non-increasing. Assume that, for some suitable similarity function, our agents can be embedded in such a latent space. Then it is natural to model the network in terms of this function, i.e.

$$g_{ij} = w\left(\Omega, P_i, P_j\right), \tag{5}$$

where Ω collects parameters of the function, and $P_i \in \mathbb{R}^D$ denotes the latent position of an agent i. This model effectively reduces the dimensionality of our target to the order $\mathcal{O}(N)$, allowing us to flexibly work with limited or no actual network data.

2.2.1 Similarity

We will base our approach on the exponential similarity function

$$w := \exp\left\{\delta \times -d(P_i, P_j)\right\}, \tag{6}$$

where the strength of links decays exponentially in the metric d. An alternate form for settings without a clear notion of a metric space to orient on, will be discussed later. For now, just note that d can simply be replaced with an absolute inner product. The sole parameter δ governs the speed of decay, and serves as a point of departure for more flexible extensions.

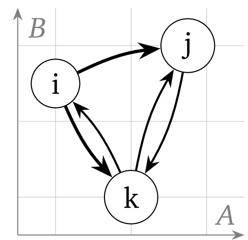

FIGURE 2: An example of an embedding in a two-dimensional space.

A natural extension of this setup addresses the issue of situation-specific links. Consider, e.g., a network of trade relationships over time; over long periods, the structure of such a network is unlikely to be static. By adapting δ_t to be time-specific, our model can, e.g., differentiate between, e.g., a decreasing importance of physical distance (perhaps due to increasing trade in services) and a mere increase in spillover effects. Another extension addresses the possibility of asymmetric links, which are not accommodated in our baseline model.[1] Consider two latent characteristics that govern an agent's potential to link to and influence others, and their susceptibility to being linked to and influenced. The extended similarity function

$$\exp\left\{\varphi_i^{-1} \times -d(P_i, P_j) \times \psi_j\right\}, \tag{7}$$

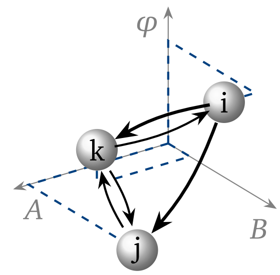

FIGURE 3: An example of an embedding in a three-dimensional space, where the coordinates A, B are augmented by a measure of potential φ.

explicitly reflects an agent's potential or *popularity* in φ, and susceptibility, or *gravity*, in ψ. These parameters allow for asymmetries by only affecting one direction of a link, illustrated in Figure 3.

Specifications such as Equation 6, where similarity decays in the distance provide an extensible baseline, and the composition of such functions offers even greater flexibility. Consider, e.g., a composition of distance- and nearest-neighbor decay. By extending Equation 6 with an indicator $1\left(d(P_i, P_j) \leqslant k_i(\kappa)\right)$, where κ indicates the number of neighbors, and k_i returns the distance to the κ'th nearest neighbor, the specification can readily

[1] For examples of such links consider, e.g., the links behind potential policy spillovers following the 2024 elections in the United States and Austria, or the link embodied by this citation of Strogatz (2001).

accommodate discontinuities. The positivity of links is another constraint to potentially be loosened. In the context of our model, positivity implies that spillovers between all linked agents are either positive or negative, i.e., they either cooperate or compete. One way to address this within the baseline framework is via latent group structures, where agents cooperate within their group, and do not link (or compete) with agents outside of their group. Another, more direct option makes way with the underlying metric, and uses an inner product instead. Consider the similarity function

$$\mathrm{sign}(P_1 \cdot P_2) \exp \left\{ \delta - |P_1 \cdot P_2| \right\},$$

where the inner product natively allows for negative links, and is able capture group-structures more directly.[2]

Positions. The latent positions are to be understood in the context of the similarity function, and could be understood as coordinates in some Euclidean space, or as certain homophilous characteristics. We will generally assume that some (noisy) measurement of or approximation to these latent positions is available, e.g., from the centroid of a region. Then it is useful to think of the latent positions in terms of

$$\mathbf{P} = \mathbf{C}\boldsymbol{\vartheta} + \mathbf{V}, \tag{8}$$

where the matrix $\mathbf{C} \in \mathbb{R}^{N \times D}$ holds approximations to the latent positions. We will make three simplifying assumptions to simplify our approach. First and second, we will assume that $\vartheta_d = 1$ for $d \leqslant \hat{D} \leqslant D$ and zero otherwise. This implies that latent positions are centered around the approximations, and emphasizes the fact that we may be dealing with a low-rank approximation. In practice, the dimensionality will be determined by the availability of good and conveniently measurable approximations. Third, we will assume that $\mathbf{V} \sim \mathrm{N}\left(\mathbf{0}, \varsigma^2 \mathbf{I}\right)$, meaning that our latent positions follow a Gaussian distribution. This interpretation of positions generalizes a large portion of approaches in the literature, relations to which (and possible extensions) will be discussed later.

2.2.2 Related approaches

The hierarchical latent space approach proposed here has interesting parallels and relations to other approaches in the literature. Most prominently, it provides a graph-

[2] To see this, consider these alternative expressions of the Euclidean metric, $||x-y||^2 = ||x||^2 + ||y||^2 - 2x'y$, and the inner product, $||x|| \times ||y|| \times x'y$. In slight abuse of language, the former considers length and adjusts for the angle (i.e., similarity in direction), while the latter relies on all three.

theoretic foundation and explicit model for many of the the networks considered in the spatial econometric literature. Spatial weights matrices are often constructed based on distances between geographic location (see, e.g., Halleck Vega and Elhorst, 2015), and, on a grid, even contiguity-based approaches (such as the one by Harari and Ferrara, 2018) can be viewed as binary decay between the centroids of cells.

Latent space models also have long been used to help model social networks (going back to Hoff et al., 2002), and have been coupled with 'aggregated relational data' to investigate spillover effects. There is also a number of other notable approaches that avoid imposing a functional form on the network. Lewbel et al. (2023) approach Equation 2 as the simultaneous equation model that it is, and show that links can be estimated directly if the network is constrained to small sizes. de Paula et al. (2023) show that long panel data and the imposition of sparsity can be used to recover links in small to moderately sized networks.

2.3 Identification

When analyzing spillover effects in the context of Equation 2, our main concern are the parameters λ and $\boldsymbol{\theta}$, measuring the strength of global and local spillover effects. It is clear that the interpretation of these parameters depends on the structure of the network; with it being treated endogenously as part of the model, identification rests on the normalization applied to the matrix $\mathbf{W}$. Before diving into the normalization procedure, it is useful to recall desirable goals of the procedure. These are

1. to facilitate *interpretation* of $\lambda, \boldsymbol{\theta}$,
2. to guarantee the *stationarity* of the network filter,
3. to preserve the *structure of the network* $\mathcal{G}$.

First goal is straightforward — pinning parameters as reflecting partial effects is convenient The second goal arises from Equation 3, and is more peculiar and specific to the model. Similar constraints on time series models.[3] In practical terms, the third goal is specific to our approach — identification of the network model's parameters improves estimation considerably. Injective normalization is what allows us to discover network parameters from Equation 2. However, I will also argue that this goal has been glossed over in terms of theory, and closer (or more flexible) ties between the underlying network and its representation in the model are warranted in general.

[3]In terms of the motivating game theoretical model, this means that a unique and stable equilibrium exists.

Row normalization. A popular choice of normalization is *row-normalization*, which scales $\mathbf{W}$ to be row-stochastic. As we will see shortly, this procedure allows us to identify both λ and $\boldsymbol{\theta}$. However, in our general setup, row-normalization entails the implicit assumption of equal out-degree centrality, which alters the structure of the network. To illustrate, consider the row-normalization visualized in Figure 4; the out-degrees of agents are trivially equalized, affecting the overall network structure. For instance, eigenvector centrality is equalized as well, and the influential position of agent i is forfeited.[4]

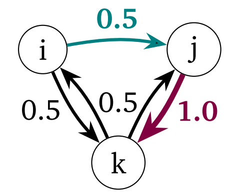

FIGURE 4: The row-stochastic equivalent of the graph in Figure 1; note the strength of highlighted links.

2.3.1 Scalar normalization

An alternative that avoids distorting the network structure is *scalar-normalization* of the adjacency matrix.[5] This normalization has been neglected in the literature, and no there exist no conclusive results on the choice of normalizing factors. Next, we will remedy this situation by showing that two *different* normalizing scalars need to be applied to identify the two types of network effects.

When normalizing the endogenous network lag, we are looking for a range of λ that guarantees the existence and stationarity of the filter in Equation 3. As we will see, this range is determined by the spectral radius of the adjacency matrix, $\rho(\mathbf{G})$, making its inverse a suitable normalizing factor.

Theorem 1. *Let* $\mathbf{I}$ *denote the identity matrix, and* α *be a real scalar. Then* $(\mathbf{I} - \alpha\mathbf{G})$ *is* invertible *for* $\alpha \in \left(\omega_{\min}^{-1}, \omega_{\max}^{-1}\right)$*, where* $\omega_{\min}$ *and* $\omega_{\max}$ *are the minimum and maximum real eigenvalues of* $\mathbf{G}$.

Proof. This statement is true if $\alpha \times \omega_i \neq 1$ for all i, which we will show directly. For $\omega_i = 0$, this is trivially the case; we need to show it for all $\omega_i \neq 0$. Notice that $\text{trace}(\mathbf{G}) = 0$, which (combined with $\omega_i \neq 0$) implies that $\omega_{\min} < 0$ and $\omega_{\max} > 0$. In order to show our result, we have two requirements. For positive eigenvalues we need to show that $\alpha < \omega_i^{-1}$, and for negative ones that $\alpha > \omega_i^{-1}$. The result follows from knowing that $\omega_{\min}^{-1} < \alpha < \omega_{\max}^{-1}$. □

Non-singularity ensures the existence of the filter and the continuity of its determinant. Next, we are concerned with its stationarity,[6] which we can guarantee if the Neumann

[4]The eigenvector centrality is transformed from $0.5, 0.1\dot{6}, 0.\dot{3}$ for agents i, j, k to $0.\dot{3}$ for all agents.

[5]This normalization will also allow us to extend the domain of Equation 1 from the positive to all reals.

[6]In the network context, stationarity implies that the (absolute) connectivity should be decreasing with the order of neighbors, i.e., units are more connected to their direct neighbors than to their neighbors'

series of the filter is convergent, or equivalently, if the spectral radius of $\mathbf{S}$ is less than unity. The following corollary follows from the definition of the spectral radius.

Corollary 1. *The spectral radius of* $(\mathbf{I} - \alpha\mathbf{G})$ *is less than unity for* $\alpha \in \left(-\rho(\mathbf{G})^{-1}, \rho(\mathbf{G})^{-1}\right)$.

As can be seen, conclusive bounds for λ are determined by the spectral radius of $\mathbf{G}$. Under relatively weak additional constraints, the spectral radius coincides with the maximum real eigenvalue. This result is good news for the literature, where the standard domain, $\lambda \in (-1, 1)$, is supported by the standard choice of row-stochastic normalization for $\mathbf{W}$, which guarantees a unit spectral radius. At the same time, the inverse spectral radius, $\rho(\mathbf{G})$, emerges as an alternative, structure-preserving normalization factor.[7]

3 Estimation

We consider the model in Equations 2 and 5, where we constrain the connectivity to $f(\lambda, \delta)$ to illustrate (i.e., we consider the network model in Equation 6, and will assume that positions are known). In this section, we describe a Markov chain Monte Carlo (MCMC) approach to obtain full posteriors of this setup.

3.1 Posteriors

First, note that we can readily obtain posterior draws of $\left(\boldsymbol{\beta}, \sigma^2\right)$ conditional on (λ, δ, τ) using the approach by Makalic and Schmidt (2015). Next, we draw from the conditional posterior of λ, and then τ. Finally, we draw from the conditional posterior δ, and repeat — giving us the procedure in Figure 5.

For the first step, we can rely on standard techniques, such as Gibbs sampling, by conditioning on the connectivity parameters. In the second and third steps, the conditional posteriors of λ, τ, and δ have no well-known form, and we must use another approach.

For λ, we can use a Metropolis-Hastings step to draw from its conditional posterior,

$$p(\lambda \mid \mathbf{y}, \tau, \cdot) \propto |\mathbf{S}(\lambda, \delta)| \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right\} p(\lambda \mid \tau).$$

neighbors.

[7] For large connectivity matrices, the prospect of computing its spectral radius may appear daunting, especially when that matrix is mutable. While direct methods for determining eigenvalues can be computationally prohibitive at a complexity of $\mathcal{O}(N^3)$, iterative methods, such as the Lanczos method or Arnoldi iteration, can provide a remedy. These methods allow us to only compute the required largest eigenvalue, quickly converge to an exact result, and particularly benefit from sparsity, which is a common feature of larger networks and a desirable property of approximations.

0. Set starting values for $\lambda, \delta, \tau, \sigma^2$.
1. Draw from the conditional posteriors of the nested linear model,
   a) $p\left(\boldsymbol{\beta} \,|\, \mathbf{y}, \lambda, \delta, \tau, \sigma^2\right)$,
   b) $p\left(\sigma^2 \,|\, \mathbf{y}, \lambda, \delta, \tau, \boldsymbol{\beta}\right)$.
2. a) Draw from $p\left(\lambda \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \delta, \tau\right)$, and
   b) $p\left(\tau \,|\, \mathbf{y}, \lambda, \cdot\right)$.
3. Draw from $p\left(\delta \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \delta\right)$.
4. Go to the first step until enough draws are obtained.

FIGURE 5: Stylized algorithm for sampling from the model.

The conditional posterior of τ can be expressed as

$$\begin{aligned} p(\tau \,|\, \mathbf{y}, \lambda, \cdot) &\propto \lambda^{\tau} \, (1-\lambda)^{\tau} \, \tau^{a-1} \exp^{-\tau b}, \\ &\propto \tau^{a-1} \exp^{-\tau\left[b - \log\left(\lambda - \lambda^2\right)\right]}, \end{aligned}$$

which is the kernel of a Gamma density, which we can directly draw from using a Gibbs step. This means that our hierarchical prior setup for λ imposes essentially no overhead over conventional specifications.

Lastly, another Metropolis-Hastings step allows us to draw from

$$p(\delta \,|\, \mathbf{y}, \cdot) \propto |\mathbf{S}(\lambda, \delta)| \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right\} p(\delta).$$

With the exception of τ, these sampling steps are well-known, and all of them are conceptually straightforward. However, they pose one major computational challenge — that is, the determinant of the $N \times N$ Jacobian matrix $\mathbf{S}(\lambda, \delta)$.

3.1.1 Evaluating the Jacobian determinant

The likelihood, and hence the posterior, of our model involves a Jacobian determinant, which poses a central computational constraint for estimation (Bivand et al., 2013). In standard models, we can use a spectral decomposition of the fixed connectivity matrix $\mathbf{W}$

to compute the determinant with the eigenvalue method, using

$$\ln |\mathbf{I} - \lambda \mathbf{W}| = \sum_{i=1}^{n} \ln \left(1 - \lambda \omega_i\right).$$

There are other approaches for large matrices, e.g. based on the lower-upper decomposition, spline approximations, or algebraic results that make use of special connectivity structures (see Bivand et al., 2013). However, all of these approaches rely on the connectivity structure in $\mathbf{W}$ being fixed, and would thus present a potentially insurmountable computational challenge for more flexible models.

In order to still allow for rapid estimation using MCMC, we introduce the following Gaussian process approximation

$$|\mathbf{S}(\lambda, \delta, \dots)| \approx \mathrm{GP}\left(\mu(\lambda, \delta, \dots), \mathbf{\Sigma}(\lambda, \delta, \dots)\right),$$

This allows us to approximate the Jacobian determinant with high accuracy (cf. the Supplementary Material). Essentially, we compute the eigenvalues for a grid of values (of δ, etc.), use those to determine $|\mathbf{S}(\lambda, \delta, \dots)|$ using the eigenvalue method, and fit these training samples using Gaussian process regression. This approach provides a quantification of uncertainty, and allows for retraining if the sampler moves to values far from the grid. For our approach, we rely on a constant mean, μ, and a Gaussian kernel for Σ, but other options are available. Notably, Gaussian processes are widespread in the field of spatial statistics, and have a parallel in spline regression, which can be used to approximate one-dimensional Jacobian determinants.

3.2 Priors

In the wider econometric literature, the overall connectivity strength λ is the only parameter considered to be unknown, and plays a central role when analyzing connectivity. With a standardized connectivity matrix, the domain $(-1, 1)$ guarantees an invertible and stationary network filter. Bayesian approaches generally assign the parameter a Uniform prior; a useful generalization (first proposed by LeSage and Parent, 2007) is the Beta prior

$$p(\bar{\lambda}) \sim \mathrm{Be}(1 + \tau, 1 + \tau),$$

where $\tau \geq 0$, and we use $\bar{\lambda} = (\lambda + 1)/2$, scaled to live on $(0, 1)$, for simplicity. Here, $\text{Be}(a, b)$ denotes the density of a Beta distribution with shapes a and b, i.e.

$$\text{Be}(x \mid a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\text{Beta}(a, b)}.$$

For $\tau = 0$, the prior is uniform over all values. With this in mind, the prior parameter τ can be understood as excess support for the origin — but, as we will see, this interpretation is misleadingly narrow.

Beyond the Uniform prior, we sometimes encounter $\tau = 0.01$ in the literature, which results in a fairly flat prior that places slightly higher weight at the origin. On one hand, this reflects a clear orientation on flat priors that are uninformative with respect to certain values. On the other hand, however, this is driven by necessity due to the Beta distribution's undesirable properties. These undesirable properties are moderate peaks in density, and excessive drop-offs towards the tails, as illustrated in Figure 6. There, we can see three Beta densities, placing increasing mass at the origin. Even for $\tau = 100$, the peak remains moderate, while the density at the tails (e.g. at 0.9) becomes miniscule, and the credible support of the prior incredibly narrow.

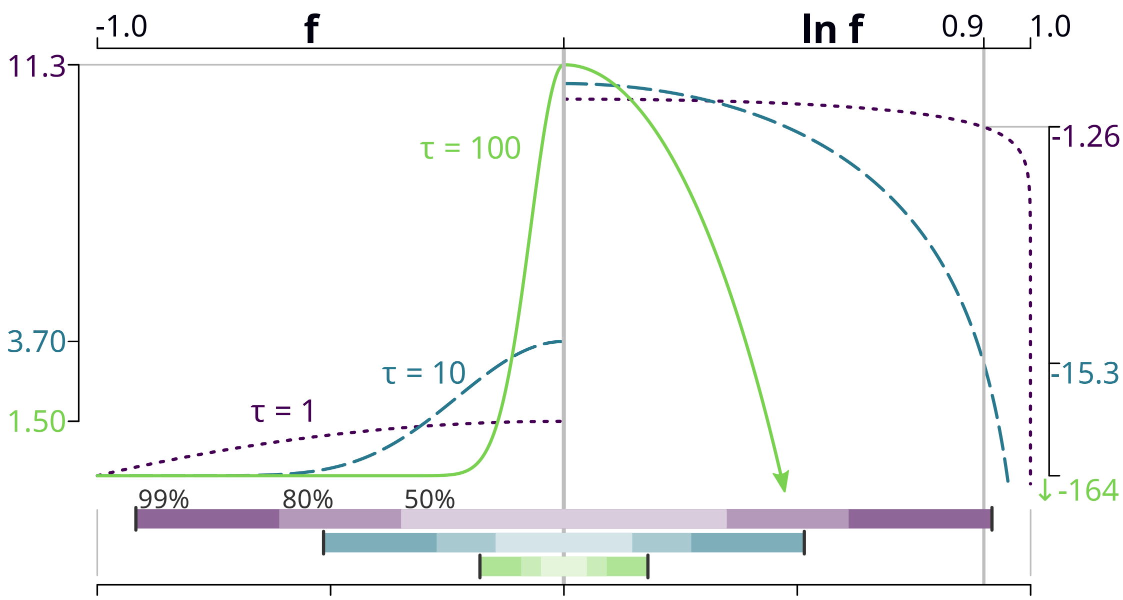

FIGURE 6: Density (left), log-density (right), and 99%, 80%, and 50% credible intervals (bottom) for λ with a Beta prior, i.e. $p(\bar{\lambda}) \sim \text{Be}(1+\tau, 1+\tau)$, with $\tau \in \{1, 10, 100\}$. More informative priors (with increasing τ) lead to narrow credible intervals, with values in the tail (e.g. 0.9, compare the right panel) receiving infinitesimal prior support.

With any prior distribution, we want to express the prior information that is available to us without distorting insights that we can obtain from the data. For our prior for λ, we thus want to incorporate information and avoid potential distortions. First, consider the often implicit, and sometimes explicit preference for parsimony, i.e. $\lambda = 0$. Without decent support for non-zero values, it appears reasonable to skip the superfluous complexity of connectivity. This potentially special role of zero highlights the distorting information induced by the Uniform prior. While this flat prior indicates no preference for any values, it implicitly prefers large values, essentially imposing connectivity on the model. For instance, we have $p(|\lambda| > 0.1) = 9 \cdot p(|\lambda| \leq 0.1)$, i.e., .

This motivates our departing point, which is the the following mixture prior

$$p(\bar{\lambda}) \sim \begin{cases} \mathrm{Be}(1+\tau_0, 1+\tau_0), & \text{if } \gamma = 1, \\ \mathrm{Be}(1+\tau_1, 1+\tau_1), & \text{if } \gamma = 0, \end{cases}$$

where $\tau_0 \ll \tau_1$ are shape parameters, and γ is an indicator. This prior essentially represents a variable selection procedure for λ. For $\gamma = 0$, the sharp spike at zero that is induced by τ_1 leads to a collapse to the linear model. For $\gamma = 1$, we have a comparatively flat prior for λ, mirroring standard setups. This small adaptation to a spike-and-slab prior takes us in the right direction conceptually, but arguably remains too rigid (except, perhaps, for panels with time-specific connectivity). Next, we consider further sources of prior information, and develop a practical prior that can accommodate our prior convictions.

The second source of prior information we want to reflect concerns the model specification — in particular, the boundaries of λ. Parameters that lie at their boundary commonly cause issues for statistical models (Chernoff, 1954; Self and Liang, 1987; Chen and Liang, 2010), and λ is no exception. Numerical instability from the filter approaching singularity is an obvious example, but not the most damning. Instead, a major issue are pathological solutions[8] that arise, not from the phenomenon under investigation, but the peculiar structure of the model (see, for example, Angrist, 2014; Halleck Vega and Elhorst, 2015). In the face of these technical and theoretical caveats, it is sensible to regularize λ and limit the support at the boundaries by imposing a suitable prior.

I propose the following continuous mixture of Beta distributions as a prior, which can coalesce the above sources of prior information in a straightforward way.

$$p(\bar{\lambda}) \sim \mathrm{Be}(1+\tau, 1+\tau), \qquad p(\tau) \sim \mathrm{Ga}(a, b), \tag{9}$$

where the mixing density is a Gamma distribution with shape a and rate b, which has

[8]One example for a model considered here, is the perfect fit from $\lambda \to -1$, $\delta \to 0$, and $\alpha = \sum_{i=1}^{n} y_i$.

proven to be useful for this purpose in similar settings (see Park and Casella, 2008; Griffin and Brown, 2010). This mixture affords us considerable flexibility for prior elicitation, allowing us to act upon the prior information discussed above.

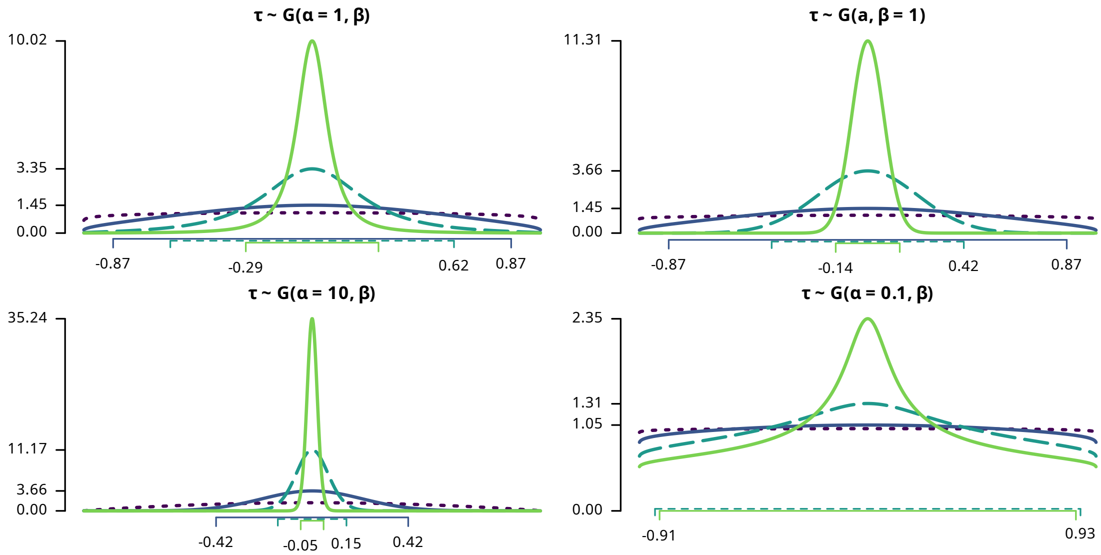

FIGURE 7: Density (left), log-density (right), and 99%, 80%, and 50% credible intervals (bottom) with a Beta-Gamma prior.

The Beta-Gamma mixture can accommodate a wide range of shapes, as exemplified in Figure 7. We can place considerable mass at the origin, while also accommodating values in the tails. If we constrain the prior to the special case of an Exponential ($a = 1$, top-left) and orient ourselves on Figure 6, we can clearly see that the mixture prior yields more pronounced peaks with wider credible support and without excessive drop-offs. By varying both parameters, we are able to flexibly induce fine-tuned priors, essentially without overhead.

The proposed Beta-Gamma shrinkage prior changes the prior specification from an issue of choosing specific values, to one concerned with parsimony and regularization. Increased weight at the origin means that the prior does not induce spillover effects per se, while the tail behavior allows us to provide regularization without limiting support to narrow regions a priori. The result is a flexible prior that can better express many prior convictions. Nonetheless, there may be reservations to even weakly informative priors.[9] While they may not always be a necessity, it is important to keep in mind that we introduce this structure to free up the previously implicit and infinitely informative priors that are fixed structural parameters.

[9]Ignore, for this example, that the standard flat priors are heavily informative in terms of size.

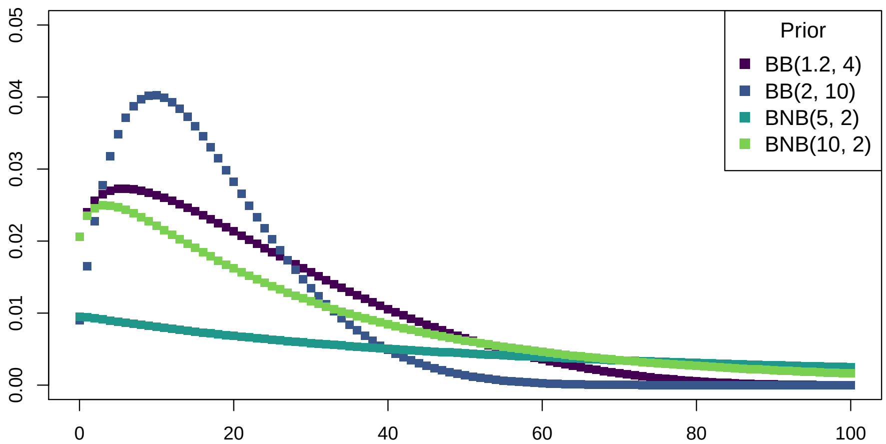

FIGURE 8: Visualization of different prior setups for k, with $n = 100$.

Priors for the structure. Parameters of the network structure include (1) the locations, (2) the speed of distance-decay, and (3) the number of neighbors. For each, I present suitable prior distributions that allow for more flexible and credible models.

The first candidate for more explicit treatment is the number of neighbors in a k-nearest neighbor specification. Limited variations to the parameter are commonly considered for robustness checks, and have previously been addressed within a model averaging framework (see Lesage and Fischer, 2008; Debarsy and LeSage, 2020; Zhang and Yu, 2018, for instance). While the discrete parameter lends itself to model averaging, we will take it one step further and treat k itself in a fully Bayesian way. For this, we can consider it as the result of N trials that determine whether any two units are neighbors or not. Such a trial can be modeled with a Beta-binomial distribution, i.e.

$$p(k) \sim \mathrm{BB}(N, a, b),$$

where N is the number of trials (i.e. potential neighbors), a and b describe the probability of success, and $\mathrm{BB}(N, of\, a, b)$ denotes the density of a Beta-binomial distribution, i.e.

$$\mathrm{BB}(x \,|\, N, a, b) = \binom{N}{x} \frac{\mathrm{Beta}(a + x, b + N - x)}{\mathrm{Beta}(a, b)}.$$

With this prior, there is no need to constrain our model to a selection of values. We can open the full model space by setting $N = n - 1$, retain sparsity in connections by choosing $a < b$ appropriately, and allow for efficient estimation using Markov chain Monte Carlo methods.

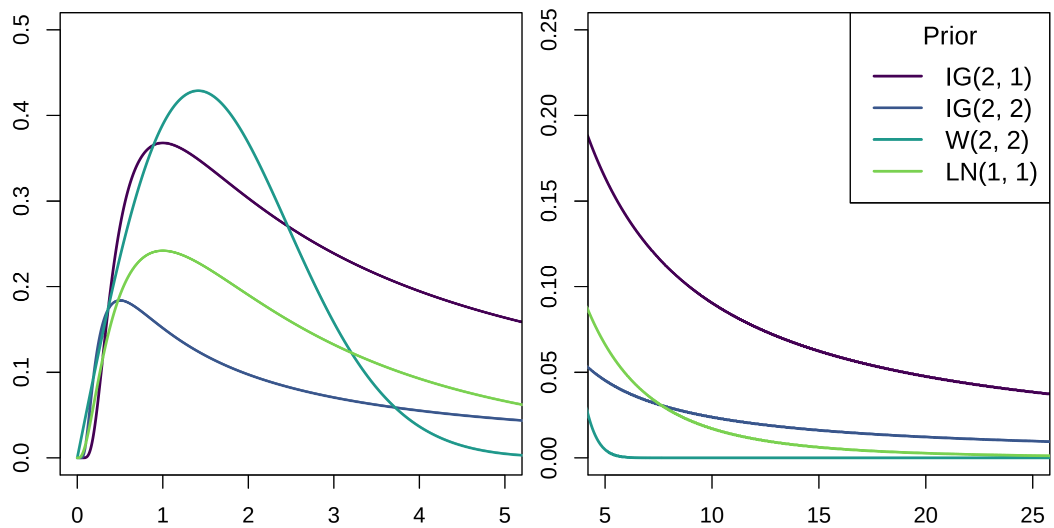

FIGURE 9: Visualization of different prior setups for δ. The left panel highlights behavior near the focal value of $\delta = 1$; the right panel shows the tail behavior.

In the literature, there is a strong preference for few neighbors, expressed in empirical work and motivated by practical and theoretical reasons. In Figure 8, we can see that the Beta-Binomial prior allows for nuanced priors that reflect this preference. However, it mirrors the underlying Beta distribution, in that it experiences excessive drop-off in the tails. If this aspect is considered restrictive, a Beta-negative-Binomial prior (visualized) can present an even less informative alternative.

Next, we consider the distance-decay parameter δ. Earlier works that are limited to local lags show that there is impactful uncertainty around this parameter (Halleck Vega and Elhorst, 2015; Kuschnig, 2022). Standard model averaging approaches provide no remedy due to the continuous nature of δ. With our fully Bayesian approach, however, we merely need a sensible prior for the parameter. A useful option is

$$\delta \sim \mathrm{Ga}^{-1}(a, b).$$

The inverse-Gamma distribution is flexible enough to accommodate general prior conceptions. For our parameter, an important benefit is that it avoids placing weight at and near zero values, where every unit is equally connected to every other unit. If these small values at the left tail are a problem, a log-Normal prior can be a useful alternative. When strong and explicit prior information is available, the Weibull distribution can be another useful alternative. For a visualization of selected priors, see Figure 9.

Regarding specific values for δ, the literature is not particularly informative. While suitable values depend on the distances involved, $\delta = 1$ can serve as an anchor for prior elicitation. Below it, i.e. for $\delta \ll 1$, connections between neighbors are weighted more equally, with less regard to the distance. For $\delta \gg 1$, by contrast, only the closest neighbors retain relevance as connectivity levels off faster than the distance.

4 Application

Deforestation of the Brazilian Amazon continues to be a pressing issue, threatening the forest's roles in, e.g., stabilizing the global climate and harboring vast biodiversity. As a public good, the Amazon sustainably provides vast benefits on global and local levels, but is susceptible to over-exploitation by actors that can appropriate a larger share of returns (in the short term) at the cost of the total benefit. In the case of the Brazilian Amazon these returns stem from the value of land, which is largely unowned or only protected by relatively weak property rights.

To combat the rampant deforestation rates of the early 2000s (cf. Figure 10), the Brazilian government and private actors implemented a number of interventions to protect the Brazilian Amazon. These include better facilities for monitoring deforestation (such as the PRODES and DETER programs), a central property registry (CAR), and strengthened property rights for local and indigenous peoples. These interventions can mostly be characterized as isolating forested land from markets. Particularly salient examples include the 'Soy Moratorium', which bans soy from previously forested areas in the Amazon from large parts of supply chains, or the 'Cattle Agreements' for an attempted parallel for beef supply chains. Other, arguably more impactful interventions of the kind include various protected and conservation areas, and the Brazilian Forest Code, which establishes, inter alia, minimum areas of private land that must be maintained as forest. Notably, most legislation (and repercussions to transgressions) only apply to private land, and the transition from public to private land via land grabbing is comparatively unregulated.

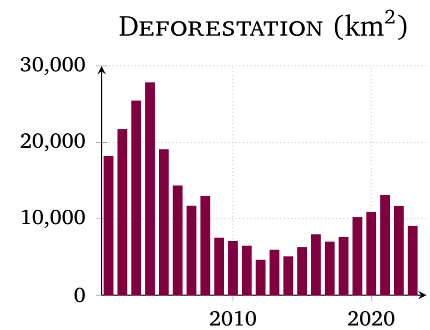

FIGURE 10: Annual deforestation rates in the Brazilian Amazon in square kilometers (Source: INPE / PRODES).

This policies have had some success in the past, but have proven irresilient more recently (Kuschnig, Vashold, et al., 2023). The underlying issue behind deforestation remains largely unaddressed, and deforestation (for the purpose of land grabbing, agriculture, etc.) remains a rational choice for many individual actors. This policy environment is especially conducive to a number of spillover effects. Examples include the role of high-

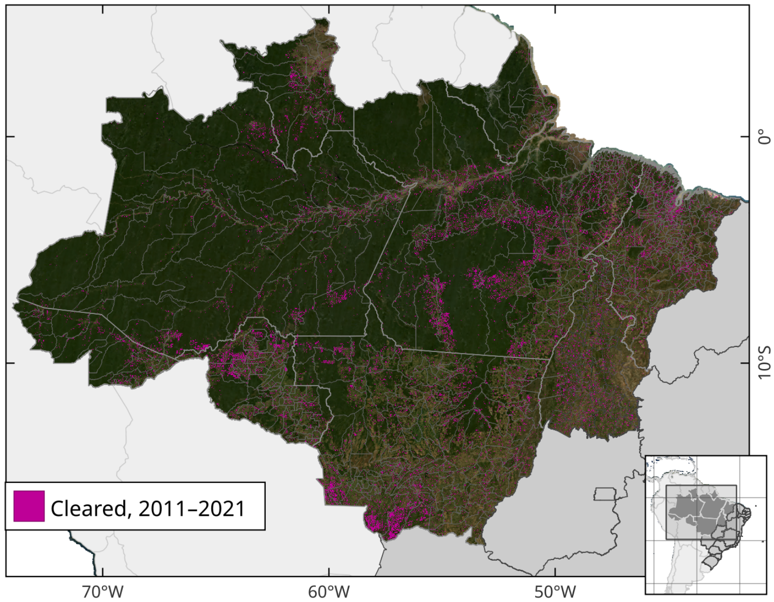

Figure 11: The extent of the Brazilian 'Legal Amazon' and forest areas that were cleared between 2011–2021. Pixels indicating cleared areas are resampled from a 30 × 30 meter grid by MapBiomas; the base image is a 2016 (06–11) cloud-free compound by Planet/NICFI.

ways, such as the BR-163, which can be traced via the deforestation aisle in the center of Figure 11 (around 55°W, 12°S). These open up vast areas of forest to commercial exploitation, driving motives for land grabbing and deforestation. Another example concerns the leakage of deforestation (and other land degradation) from the (more stringently protected) Amazon biome into the neighboring Cerrado biome, or the displacement of cattle from pastures with soy, driving cattle herds into newly deforested areas. Due to their nature, these shortcomings of the current interventions are hard to quantify — not least due to various spillover effects — and hard to address in a forward-looking, sustainable way.[10]

Deforestation itself is an inherently spatial process, as is agriculture and many other potential uses for cleared land, as are many of the policies addressing it. The network dynamics of land use decisions are arguably even more pronounced when considering, inter alia, the risk of receiving a fine, the probability of being sentenced, and the ease of registering appropriated land as coordination games.[11] Moreover, spatial spillover ef-

[10]Some issues could be addressed by plugging holes in their implementation, e.g., by improving monitoring of beef supply chains. However, if one assumes that there is (near) a continuum of goods (in the long-term) this is like plugging holes on a sinking ship.

[11]Empirically, we know that, e.g, capacity limits and interference by the Bolsonaro government severely reduced the fine intensity per deforestation, but also undermined their efficacy by instituting mandatory 'reconciliation hearings' for environmental offenses. This resulted in massive backlogs at court. Together with limitation periods after which fines are written off, this can be seen as inducing a coordination game

fects occur at local, regional, and national levels (Sá et al., 2013; Gollnow et al., 2018; Kuschnig, Crespo Cuaresma, et al., 2021). Most empirical models of deforestation processes, however, generally aggregate independent land use decisions to some local level, abstracting from spatial and network dynamics. Exceptions include studies with a local focus that sidestep the issue to an extent, and studies that explicitly allow for spatial spillovers in their approach. These studies, which will serve as our baseline for the state of the art, rely on strong and potentially distortive assumptions regarding the structure of the spatial network, which is generally assumed known.

With this in mind, we will entertain a simple hierarchical extension of such approaches. We will consider an autoregressive model, with regressors that include the initial levels of population, GDP per capita, soy prices, cattle herd sizes, as well as controls for forested area, dry weather, protected areas, and fine intensities. The model is given by

$$\begin{aligned} \mathrm{def}_t &= \lambda \mathbf{W} \mathrm{def}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \\ \mathbf{W} &= f(\delta, P_1, P_2 \mid \kappa, d), \end{aligned}$$

where def_t is log-deforestation at time t, $\boldsymbol{\varepsilon}_t$ also holds state- and time-fixed effects, f is an exponential distance-decay function, κ is a distance-threshold, d is the Euclidean distance, and P_1, P_2 are the latent positions of municipalities. The prior for these positions is given by

$$\left(P_1, P_2\right) \sim \mathrm{N}\left((\mathrm{lon}, \mathrm{lat}), \varsigma \mathbf{I}\right),$$

where the mean of the multivariate Normal is given by the longitude and latitude of the respective municipality. Other priors are standard (as described in section 3), and κ is fixed such that links beyond the 95$^{\text{th}}$ quantile of *a priori* distances are suppressed.[12] To reduce estimation error and guarantee convergence of the sampler, we base our estimates on 10,000 draws across 20 independent chains (after burn-in periods of 5,000 draws).

Results are reported in Table 1. First, note that estimates of the $\boldsymbol{\beta}$ coefficients are relatively stable across the models considered; minor reductions seem to occur for the posterior means of population and soy price coefficients in more flexible models that include endogenous network effects. These minor changes are rather unsurprising, considering that we suppress local spillovers of these variables in all models. In terms of endogenous spillovers, we see that both contiguity and naively fixed distance-decay specifications underestimate the magnitude of the spatial filter effect. When parameterizing the decay

for the risk of having to bear fines. A similar issue exists for registrations in CAR — the more properties are registered, the fewer can be reviewed due to capacity constraints.

[12] Robustness is checked, inter alia, to the prior location and scale of the latent positions, and the exact value of κ.

	LM	**R**	**D**	$g(\delta)$	$g(\delta, P)$	90% HPDI
forest	0.587	0.555	0.579	0.548	0.550	[0.54, 0.57]
population	0.211	0.178	0.209	0.178	0.178	[0.16, 0.20]
GDP/capita	-0.374	-0.382	-0.376	-0.398	-0.401	[-0.43, -0.37]
soy-price	-0.462	-0.366	-0.441	-0.337	-0.340	[-0.42, -0.26]
cattle-heads	0.308	0.301	0.310	0.309	0.308	[0.30, 0.32]
λ		0.146	0.035	0.198	0.187	[0.14, 0.22]
δ			1.00	0.076	0.092	[0.04, 0.19]

TABLE 1: Estimation results obtained using a classical linear model ('LM'), a fixed and row-stochastic contiguity matrix ('**R**'), a fixed exponential distance-decay matrix ('**D**'), a hierarchical model of the decay parameter ('$g(\delta)$'), and of the decay parameter and positions ('$g(\delta, P)$'). The last column reports the highest posterior density interval (HPDI) covering 90% of the posterior of the most flexible hierarchical specification.

parameter δ and considering uncertainty around latent positions, we find that $\lambda \approx 0.19$. Due to the sampling-based approach, we can work with the full posteriors of free parameters, allowing us, e.g., to summarize uncertainty using highest posterior density intervals. Notably, this allows a degree of non-parametric identification.

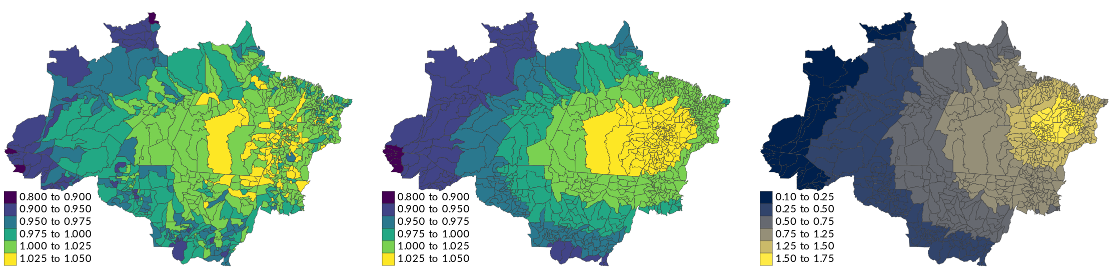

FIGURE 12: Relative eigenvector centrality of municipalities in the network recovered by estimating (δ, P) (left), just δ (center), and implied by fixing $\delta = 1$ (right). When the network is transformed to be row-stochastic all eigenvector (and out-degree) centralities are equalized.

In Figure 12, we see that the hierarchical network model can non-parametrically recover structure of the underlying network. Current approaches generally equalize eigenvector centrality by design (via row-scaling), or impose a certain structure via the geography. In the right-most map of Figure 12, we can see how the deterministic decay-specification leads to centralities that are dominated by the small municipalities in the North-East due to their small size and contiguous nature. It should be clear from Figure 11 that these regions are not particularly 'central' in terms of forest or deforestation, and *a posteriori* it shouldn't be surprising that a more flexible model yields much slower speeds of distance-decay, in which larger municipalities can achieve higher centrality (center map, cf. the largest municipality in the center, which is the location of the BR-163 highway). Lastly, we see that a specification that allows for uncertain, latent positions further shifts the focus of the network towards the Amazon.

5 Conclusion

In this paper, I presented a hierarchical approach to jointly model spillovers and the latent networks behind them. I used a hierarchical setup to estimate a latent space network model behind the linear network model. I provided a theoretical foundation for the network in graph theory, and showed how to identify the strength of spillover effects. I discussed interpretation of spillovers, and the possibility of conducting inference with respect to the recovered network. I proposed a Bayesian sampling approach to estimate the model, facilitating full posterior inference. I discussed the role of sensible priors, and proposed alternatives to the ones currently established in the literature. Lastly, I demonstrated the approach in an investigation of spillover effects in Brazilian deforestation.

References

Acemoglu, Daron, Ufuk Akcigit, and William R. Kerr (2016). “Innovation network”. In: *Proceedings of the National Academy of Sciences* 113.41, pp. 11483–11488. ISSN: 0027-8424. DOI: 10.1073/pnas.1613559113.

Acemoglu, Daron, Vasco M. Carvalho, et al. (2012). “The network origins of aggregate fluctuations”. In: *Econometrica* 80.5, pp. 1977–2016. ISSN: 0012-9682. DOI: 10.3982/ECTA9623.

Ambrus, Attila, Markus Mobius, and Adam Szeidl (2014). “Consumption risk-sharing in social networks”. In: *American Economic Review* 104.1, pp. 149–82. ISSN: 0002-8282. DOI: 10.1257/aer.104.1.149.

Angrist, Joshua D. (2014). “The perils of peer effects”. In: *Labour Economics* 30, pp. 98–108. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2014.05.008.

Atalay, Enghin, Ali Hortaçsu, James Roberts, et al. (2011). “Network structure of production”. In: *Proceedings of the National Academy of Sciences* 108.13, pp. 5199–5202. ISSN: 0027-8424. DOI: 10.1073/pnas.1015564108.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson (2014). “Vertical integration and input flows”. In: *American Economic Review* 104.4, pp. 1120–48. ISSN: 0002-8282. DOI: 10.1257/aer.104.4.1120.

Beaman, Lori A. (2012). “Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S.” In: *Review of Economic Studies* 79.1, pp. 128–161. ISSN: 0034-6527. DOI: 10.1093/restud/rdr017.

Bivand, Roger, Jan Hauke, and Tomasz Kossowski (2013). "Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods". In: *Geographical Analysis* 45.2, pp. 150–179. DOI: 10.1111/gean.12008.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013). "Identifying technology spillovers and product market rivalry". In: *Econometrica* 81.4, pp. 1347–1393. ISSN: 0012-9682. DOI: 10.3982/ECTA9466.

Board, Simon and Moritz Meyer-ter-Vehn (2021). "Learning dynamics in social networks". In: *Econometrica* 89.6, pp. 2601–2635. ISSN: 1468-0262. DOI: 10.3982/ECTA18659.

Chen, Yong and Kung-Yee Liang (2010). "On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems". In: *Biometrika* 97.3, pp. 603–620. ISSN: 0006-3444. DOI: 10.1093/biomet/asq031.

Chernoff, Herman (1954). "On the distribution of the likelihood ratio". In: *The Annals of Mathematical Statistics*, pp. 573–578. DOI: 10.1214/aoms/1177728725.

Chyn, Eric and Lawrence F. Katz (2021). "Neighborhoods matter: Assessing the evidence for place effects". In: *Journal of Economic Perspectives* 35.4, pp. 197–222. ISSN: 0895-3309. DOI: 10.1257/jep.35.4.197.

Crespo Cuaresma, Jesús et al. (2019). "Spillovers from US monetary policy: Evidence from a time varying parameter global vector autoregressive model". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182.3, pp. 831–861. DOI: 10.1111/rssa.12439.

de Paula, Áureo, Imran Rasul, and Pedro Souza (2023). *Identifying network ties from panel data: Theory and an application to tax competition*. DOI: 10.48550/arXiv.1910.07452.

Debarsy, Nicolas and James P. LeSage (2020). "Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices". In: *Journal of Business & Economic Statistics*, pp. 1–12. DOI: 10.1080/07350015.2020.1840993.

Ductor, Lorenzo et al. (2014). "Social networks and research output". In: *Review of Economics and Statistics* 96.5, pp. 936–948. ISSN: 0034-6535. DOI: 10.1162/REST_a_00430.

Gofman, Michael (2017). "Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions". In: *Journal of Financial Economics* 124.1, pp. 113–146. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2016.12.009.

Goldsmith-Pinkham, Paul and Guido W. Imbens (2013). "Social networks and the identification of peer effects". In: *Journal of Business & Economic Statistics* 31.3, pp. 253–264. DOI: 10.1080/07350015.2013.801251.

Gollnow, Florian et al. (2018). "Property-level direct and indirect deforestation for soybean production in the Amazon region of Mato Grosso, Brazil". In: *Land Use Policy* 78, pp. 377–385. DOI: 10.1016/j.landusepol.2018.07.010.

Griffin, Jim E. and Philip J. Brown (2010). "Inference with Normal-Gamma prior distributions in regression problems". In: *Bayesian Analysis* 5.1, pp. 171–188. DOI: 10.1214/10-BA507.

Halleck Vega, Solmaria and J. Paul Elhorst (2015). "The SLX model". In: *Journal of Regional Science* 55.3, pp. 339–363. DOI: 10.1111/jors.12188.

Harari, Mariaflavia and Eliana La Ferrara (2018). “Conflict, climate, and cells: A disaggregated analysis”. In: *Review of Economics and Statistics* 100.4, pp. 594–608. DOI: 10.1162/rest_a_00730.

Hensvik, Lena and Oskar Nordström Skans (2016). “Social networks, employee selection, and labor market outcomes”. In: *Journal of Labor Economics* 34.4, pp. 825–867. DOI: 10.1086/686253.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock (2002). “Latent space approaches to social network analysis”. In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098. ISSN: 0162-1459. DOI: 10.1198/016214502388618906.

Hsieh, Chih-Sheng and Lung Fei Lee (2016). “A social interactions model with endogenous friendship formation and selectivity”. In: *Journal of Applied Econometrics* 31.2, pp. 301–319. DOI: 10.1002/jae.2426.

Ioannides, Yannis M. and Linda Datcher Loury (2004). “Job information networks, neighborhood effects, and inequality”. In: *Journal of Economic Literature* 42.4, pp. 1056–1093. ISSN: 0022-0515. DOI: 10.1257/0022051043004595.

Johnsson, Ida and Hyungsik Roger Moon (2021). “Estimation of peer effects in endogenous social networks: Control function approach”. In: *Review of Economics and Statistics* 103.2, pp. 328–345. ISSN: 0034-6535. DOI: 10.1162/rest_a_00870.

König, Michael D., Xiaodong Liu, and Yves Zenou (2019). “R&D networks: Theory, empirics, and policy implications”. In: *Review of Economics and Statistics* 101.3, pp. 476–491. ISSN: 0034-6535. DOI: 10.1162/rest_a_00762.

Kranton, Rachel E. and Deborah F. Minehart (2001). “A theory of buyer-seller networks”. In: *American Economic Review* 91.3, pp. 485–508. ISSN: 0002-8282. DOI: 10.1257/aer.91.3.485.

Kuschnig, Nikolas (2022). “Bayesian spatial econometrics: A software architecture”. In: *Journal of Spatial Econometrics* 3.1, pp. 6–25. ISSN: 2662-298X. DOI: 10.1007/s43071-022-00023-w.

Kuschnig, Nikolas, Jesús Crespo Cuaresma, et al. (2021). “Spillover effects in agriculture drive deforestation in Mato Grosso, Brazil”. In: *Scientific Reports* 11.1, pp. 1–9. DOI: 10.1038/s41598-021-00861-y.

Kuschnig, Nikolas, Lukas Vashold, et al. (2023). “Eroding resilience of deforestation interventions — evidence from Brazil’s lost decade”. In: *Environmental Research Letters* 18.7, p. 074039. ISSN: 1748-9326. DOI: 10.1088/1748-9326/acdfe7.

Lam, Clifford and Pedro Souza (2020). “Estimation and selection of spatial weight matrix in a spatial lag model”. In: *Journal of Business & Economic Statistics* 38.3, pp. 693–710. DOI: 10.1080/07350015.2019.1569526.

Lesage, James P. and Manfred M. Fischer (2008). “Spatial growth regressions: Model specification, estimation and interpretation”. In: *Spatial Economic Analysis* 3.3, pp. 275–304. DOI: 10.1080/17421770802353758.

LeSage, James P. and Olivier Parent (2007). “Bayesian model averaging for spatial econometric models”. In: *Geographical Analysis* 39.3, pp. 241–267. DOI: 10.1111/j.1538-4632.2007.00703.x.

Lewbel, Arthur, Xi Qu, and Xun Tang (2023). "Social networks with unobserved links". In: *Journal of Political Economy*. DOI: 10.1086/722090.

Lin, Xu (2010). "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables". In: *Journal of Labor Economics* 28.4, pp. 825–860. DOI: 10.1086/653506.

Makalic, Enes and Daniel F. Schmidt (2015). "A simple sampler for the horseshoe estimator". In: *IEEE Signal Processing Letters* 23.1, pp. 179–182. DOI: 10.1109/LSP.2015.2503725.

Mele, Angelo (2020). "Does school desegregation promote diverse interactions? an equilibrium model of segregation within schools". In: *American Economic Journal: Economic Policy* 12.2, pp. 228–57. ISSN: 1945-7731. DOI: 10.1257/pol.20170604.

Munshi, Kaivan (2003). "Networks in the modern economy: Mexican migrants in the U.S. labor market". In: *Quarterly Journal of Economics* 118.2, pp. 549–599. ISSN: 0033-5533. DOI: 10.1162/003355303321675455.

Newman, Mark E. J. (2001). "The structure of scientific collaboration networks". In: *Proceedings of the National Academy of Sciences* 98.2, pp. 404–409. ISSN: 0027-8424. DOI: 10.1073/pnas.98.2.404.

Park, Trevor and George Casella (2008). "The Bayesian lasso". In: *Journal of the American Statistical Association* 103.482, pp. 681–686. DOI: 10.1198/016214508000000337.

Qu, Xi and Lung-fei Lee (2015). "Estimating a spatial autoregressive model with an endogenous spatial weight matrix". In: *Journal of Econometrics* 184.2, pp. 209–232. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2014.08.008.

Rose, Andrew K. (2004). "Do we really know that the WTO increases trade?" In: *American Economic Review* 94.1, pp. 98–114. ISSN: 0002-8282. DOI: 10.1257/000282804322970724.

Sá, Saraly Andrade de, Charles Palmer, and Salvatore Di Falco (2013). "Dynamics of indirect land-use change: Empirical evidence from Brazil". In: *Journal of Environmental Economics and Management* 65.3, pp. 377–393. DOI: 10.2139/ssrn.2201634.

Self, Steven G. and Kung-Yee Liang (1987). "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions". In: *Journal of the American Statistical Association* 82.398, pp. 605–610. ISSN: 0162-1459. DOI: 10.1080/01621459.1987.10478472.

Strogatz, Steven H. (2001). "Exploring complex networks". In: *Nature* 410.6825, pp. 268–276. ISSN: 1476-4687. DOI: 10.1038/35065725.

Weidmann, Ben and David J. Deming (2021). "Team players: How social skills improve team performance". In: *Econometrica* 89.6, pp. 2637–2657. ISSN: 1468-0262. DOI: 10.3982/ECTA18461.

Zhang, Xinyu and Jihai Yu (2018). "Spatial weights matrix selection and model averaging for spatial autoregressive models". In: *Journal of Econometrics* 203.1, pp. 1–18. DOI: 10.1016/j.jeconom.2017.05.021.