

RESEARCH ARTICLE

Spherical latent space models for social networks: Geometry-aware inference and comparison across latent geometries

Carlos Nosa and Juan Sosa 

Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Corresponding author: Juan Sosa; Email: jcsosam@unal.edu.co

Abstract

This article studies latent space models for social network data in which actors are embedded on a hyper-sphere and link probabilities depend on angular similarity. In contrast to Euclidean embeddings, the spherical formulation provides a compact parameter space, stabilizes the linear predictor through bounded inner products, and offers a natural representation of directional and cyclic structure. For inference, we combine maximum likelihood estimation, used to obtain initial values for latent positions and model parameters, with geometry-aware Bayesian methods based on Metropolis–Hastings and Hamiltonian Monte Carlo algorithms, including a geodesic Hamiltonian scheme for manifold-constrained parameters. We conduct a systematic empirical comparison between Euclidean and spherical latent space models on a benchmark social network dataset, evaluating model fit, predictive performance, and interpretability. The results show that spherical representations provide competitive performance while offering a more constrained and geometrically interpretable structure. Overall, the paper clarifies the role of latent space geometry in network modeling and highlights the importance of geometry-aware inference in statistical analysis of relational data.

Keywords: Latent representation; spherical embedding; statistical network models; MCMC-based estimation

1. Introduction

The study of complex systems through their interconnected elements has become central to numerous scientific disciplines. Such systems are commonly represented as networks or graphs, comprising autonomous actors (nodes) and the relational ties (edges) between them. These structures reveal patterns underlying phenomena ranging from disease transmission to social collaboration. The breadth of applications underscores the versatility of network models, which can represent entities as diverse as individuals, organizations, or nations, linked by relationships such as friendships, collaborations, or information exchange, respectively.

Modern statistical network analysis has three primary objectives: (1) identifying and summarizing structural patterns, (2) developing stochastic models to explain network formation processes, and (3) predicting unobserved or future connections based on network properties and actor attributes (see Pósfai and Barabási, 2016, Al-Taie and Kadry, 2017, Newman, 2018, Brugere *et al.*, 2018, and Drobysheskiy and Turdakov, 2019 for comprehensive reviews of network definitions, properties, methods, models, and processes across multiple fields). A central challenge lies in formulating statistical models (rather than deterministic mechanisms) that capture the inherent dependencies among ties, particularly reciprocity and transitivity. Traditional approaches, such as the exponential random graph models of Robins *et al.* (2007), address these dependencies but are often susceptible to degeneracy and limited interpretability.

In statistical network modeling, latent space models offer a compelling alternative (see Goldenberg *et al.*, 2010, Salter-Townshend *et al.*, 2012, Matias and Robin, 2014, Sosa and Buitrago, 2021, and Kaur *et al.*, 2023 for comprehensive reviews of latent space modeling and related statistical approaches). In particular, the latent space model, introduced in the foundational work of Hoff *et al.* (2002), embeds actors in an unobserved, typically low-dimensional, latent space (often referred to as the social space) where the probability of a tie is modeled as a function of the distance or similarity measure between actors in that space. This formulation naturally captures transitivity, as actors positioned closer together in latent space are more likely to form connections. The geometric structure of these models renders them both computationally tractable and interpretable. Nonetheless, they may be inadequate for networks generated by highly nonlinear processes or exhibiting complex geometric features beyond the representational capacity of Euclidean spaces.

Latent space models have been extensively extended to a wide range of settings. Beyond simple distance-based formulations, alternative latent structures have been proposed, including projection models (Hoff *et al.*, 2002), bilinear models (Hoff, 2005, Hoff, 2021), and eigenmodels (Hoff, 2008, Hoff, 2009, Minhas *et al.*, 2019). These frameworks have been generalized to accommodate multiple structural features, such as multilayer networks (Hoff, 2011, Kivelä *et al.*, 2014, Gollini and Murphy, 2016, Salter-Townshend and McCormick, 2017, D'Angelo *et al.*, 2019, Zhang *et al.*, 2020, Sosa and Rodríguez, 2021, Sosa and Betancourt, 2022, D'Angelo *et al.*, 2023) and dynamic network settings (Durante and Dunson, 2014, Sewell and Chen, 2015, Hoff, 2015, Sewell and Chen, 2016, Sewell and Chen, 2017, Kim *et al.*, 2018, He and Hoff, 2019, Guhaniyogi and Rodríguez, 2020). They have also been adapted to perform diverse statistical tasks, including covariate testing (Austin *et al.*, 2013, Fosdick and Hoff, 2015, Wang *et al.*, 2023), clustering (Nowicki and Snijders, 2001, Krivitsky *et al.*, 2009, Ryan *et al.*, 2017, Rastelli and Friel, 2018, Ng and Murphy, 2022), modeling bipartite relational data (Friel *et al.*, 2016), deduplication and record linkage (Sosa and Rodríguez, 2022, Sosa and Rodríguez, 2023), and the study of influence and diffusion processes (Xu, 2018, Sánchez *et al.*, 2026). Moreover, significant efforts have focused on developing faster computational alternatives to improve scalability and efficiency (Hunter *et al.*, 2012, Raftery *et al.*, 2012, Salter-Townshend and Murphy, 2013, Caimo and Gollini, 2016, Ma *et al.*, 2020, Aliverti and Russo, 2022, Spencer *et al.*, 2022, Rastelli *et al.*, 2024).

Most of these extensions adopt the standard choice of a Euclidean latent space, i.e. embedding actors in $\mathbb{R}^d$ for some positive integer d , and model tie probabilities as decreasing functions of Euclidean distance. This specification is attractive for its simplicity and interpretability, but it has structural limitations that are important in some applications. First, the latent space is unbounded, so latent positions can move very far apart and this can produce extreme values in the linear predictor and almost deterministic tie probabilities, especially in very sparse or highly clustered networks. Second, the Euclidean formulation is invariant under global translation and rescaling of the configuration, so the overall scale of the latent positions is not identified separately from the intercept and other global parameters, which complicates prior specification and interpretation. Third, while Euclidean geometry is flexible and widely applicable, it can be less naturally suited to representing directional or cyclic organization. For example, when actors are arranged around ideological orientations or functional roles, a Euclidean embedding may require additional dimensions or ad hoc constraints to approximate the underlying structure, potentially reducing parsimony or interpretability. These considerations motivate the exploration of latent spaces with bounded geometry and a natural notion of angular similarity.

One way to address these limitations is to embed actors on a hypersphere instead of in a flat geometry. The use of circular or spherical latent spaces in network modeling is well established in the literature. For example, Hoff *et al.* (2002) introduced a projection model in which latent positions lie on the unit sphere and similarity is captured through angular structure. More recently, spherical and other non-Euclidean geometries have been studied in broader geometric frameworks for network models (e.g., McCormick and Zheng, 2015; Smith *et al.*, 2019;

Papamichalis *et al.*, 2022; Lubold *et al.*, 2023), where curvature and manifold structure are treated as central modeling components. Our contribution builds on this line of work but differs in emphasis and scope. We focus on a latent distance model for general individual-level social network data in which the hypersphere is adopted as the primary latent space and inference is conducted within a Bayesian framework, implementing Markov chain Monte Carlo (MCMC) algorithms (e.g., Gamerman and Lopes, 2006) that explicitly respect the manifold constraints of spherical latent positions.

From a modeling perspective, the spherical geometry responds directly to the problems described above. By constraining all latent positions to lie on the unit sphere we fix the origin and the overall scale of the configuration, and the contribution of the latent space enters only through inner products in the interval $[-1, 1]$, which keeps the linear predictor in a moderate range and prevents tie probabilities from becoming numerically zero or one, stabilizing posterior inference and prediction. Similarity is captured through angular separation, so the geometry naturally represents directional clustering and cyclic structures and provides additional modeling flexibility in applications where actors are organized around latent directions. Angular separation also defines latent directions that correspond to interpretable groups or factions. In this paper we develop a general Bayesian inference framework for this spherical latent space model using MCMC and we demonstrate its performance on benchmark social network data sets, including those in Hoff *et al.* (2002), where it demonstrates competitive model fit and enhanced geometric interpretability relative to the traditional Euclidean formulation. We do not claim that spherical geometry universally dominates Euclidean latent spaces. Rather, our objective is to demonstrate when a bounded and angular latent representation can provide computational stability, interpretability, and competitive predictive performance relative to standard Euclidean formulations using a benchmark example.

In this spirit, the paper pursues three main goals. First, we formulate a spherical latent space model for social networks and develop simulation-based inference procedures that respect the geometry of the hypersphere using Metropolis and Hamiltonian based updates. Second, we carry out a detailed comparative analysis of Euclidean and spherical latent spaces on the Florentine marriage network as in Hoff *et al.* (2002), evaluating model fit, recovery of centrality, community detection, and predictive performance. Finally, we provide free public code available at https://github.com/cnosa/Latentspacemodels_Networks, implementing all the sampling algorithms in plain Python, allowing direct modification and extension without reliance on automatic differentiation frameworks that assume unconstrained Euclidean parameter spaces. These contributions show that spherical latent space models provide an interpretable directional representation of network structure and achieve competitive predictive performance under a more constrained formulation.

The remainder of this article is structured as follows. Section 2 describes the fundamentals of network structures and their characterization, reviews related work on latent space modeling, and presents the proposed Euclidean and spherical latent space framework, along with its statistical properties. Section 3 details the Bayesian inference procedure and computational strategies. Section 4 presents the experimental results, and Section 5 concludes with an analysis of the implications and avenues for future research.

2. Fundamentals

Network data consist of actors, nodal attributes measured on individuals, and dyadic attributes measured on actor pairs. Dyadic variables may be binary, indicating the presence or absence of a tie (binary networks), or weighted, quantifying relational aspects (valued networks). Relations can be undirected, with a single value per pair, or directed, with two values capturing each actor's perspective. Networks are typically represented as graphs or adjacency matrices $\mathbf{Y} = [y_{ij}]$, where y_{ij} denotes the presence, absence, or weight of a tie. Self-loops generate structural zeros on the

diagonal, and the symmetry or asymmetry of $\mathbf{Y}$ indicates whether the network is undirected or directed.

Statistical network analysis methods fall into three categories: descriptive, which visualize and quantify structure; modeling and inference, which explain network formation; and process, which study how interactions influence attributes. This work focuses on modeling and inference, emphasizing structural features such as cohesion, connectivity, and assortativity that reflect dependencies like reciprocity and clustering. Accounting for these dependencies is crucial for constructing realistic statistical models. For comprehensive reviews of network properties and measures, see Menczer *et al.* (2020) and Kolaczyk and Csárdi (2020).

In this context, a statistical network model is a probability distribution on a sociomatrix $\mathbf{Y}$ indexed by an unknown parameter $\theta \in \Theta$, $p(\mathbf{Y} | \theta)$. Rather than merely visualizing and describing the topological characteristics of a network, statistical models seek to capture the essential aspects of the stochastic mechanism by which the network may have arisen. They enable testing the significance of predefined structural features, assessing associations between node or edge attributes and the overall network structure, and imputing missing observations. Unlike deterministic or purely algorithmic models, statistical approaches also quantify the uncertainty associated with unknown quantities. Importantly, the very nature of network data induces dependencies both between actors and between ties. Accounting for these dependencies is indispensable for formulating statistically sound and substantively meaningful network models.

2.1 Latent space models

Latent space models are a widely used approach to network modeling, originally introduced by Hoff *et al.* (2002) and further developed by Handcock *et al.* (2007) and Krivitsky *et al.* (2009). In this framework, each node is assigned a latent position in a Euclidean space, and the probability of an edge between two nodes depends on their proximity in that space. These models can be viewed as generalized linear models with random effects, allowing them to capture complex network structures. For conditionally independent $y_{i,j}$ in the undirected binary case, the interaction probabilities are

$$\Pr(y_{i,j} = 1 | \mathbf{x}_{i,j}, \boldsymbol{\beta}, \zeta_{i,j}) = \text{logit}^{-1}(\mathbf{x}_{i,j}^\top \boldsymbol{\beta} + \zeta_{i,j}), \quad 1 \leq i < j \leq n,$$

where $\boldsymbol{\beta}$ are fixed effects, $\mathbf{x}_{i,j}^\top \boldsymbol{\beta}$ represents the contribution of covariates $\mathbf{x}_{i,j}$, and $\zeta_{i,j}$ captures unobserved effects, and $\text{logit}(x) = \log \frac{x}{1-x}$ is the link function. As noted in Hoff (2008), following foundational results from Hoover (1982) and Aldous (1985), a jointly exchangeable random effects matrix $[\zeta_{i,j}]$ can be expressed as $\zeta_{i,j} = h(\mathbf{z}_i, \mathbf{z}_j)$, where $h(\cdot, \cdot)$ is a symmetric function and $\mathbf{z}_1, \dots, \mathbf{z}_n$ are independent latent variables (vectors). The function $h(\cdot, \cdot)$ is central to modeling relational data, and various formulations have been proposed. In this paper, we focus on the distance formulation for undirected binary networks. For a comprehensive review, see Sosa and Buitrago (2021).

A common specification is the latent distance model (Hoff *et al.*, 2002), in which each actor i occupies an unobserved position $\mathbf{z}_i \in \mathbb{R}^d$, often referred to as the latent or social space. Conditional on these positions, ties are modeled as

$$y_{i,j} | \eta_{i,j} \stackrel{\text{ind}}{\sim} \text{Ber}(\text{logit}^{-1}(\eta_{i,j})), \quad \eta_{i,j} = \alpha - d(\mathbf{z}_i, \mathbf{z}_j),$$

where $\mathbf{z}_i$ and $\mathbf{z}_j$ are latent position vectors capturing unobserved social characteristics, $d(\cdot, \cdot)$ is a distance function, typically $d(\mathbf{z}_i, \mathbf{z}_j) = \|\mathbf{z}_i - \mathbf{z}_j\|$ with $\|\cdot\|$ denoting the Euclidean norm in $\mathbb{R}^d$, and α is an intercept representing the baseline log odds of a tie when two actors occupy the same latent position. The subtraction of $d(\mathbf{z}_i, \mathbf{z}_j)$ from α reflects the assumption that the probability of a tie decreases as the distance between latent positions increases, encouraging connections

between actors located close together in latent space. This property naturally induces transitivity and clustering, as actors close to a common neighbor are also likely to be close to each other.

Alternatively, for modeling in a spherical latent space, the probability of a tie can be parameterized using the inner product between latent vectors, capturing proximity through angular similarity:

$$\eta_{i,j} = \text{logit } P(y_{i,j} = 1 \mid \alpha, \beta, \mathbf{z}_i, \mathbf{z}_j) = \alpha + \beta \langle \mathbf{z}_i, \mathbf{z}_j \rangle,$$

where $\mathbf{z}_i, \mathbf{z}_j \in \mathbb{S}^{d-1}$ are unit vectors on the $(d-1)$ -dimensional sphere, $\langle \mathbf{z}_i, \mathbf{z}_j \rangle = \mathbf{z}_i^\top \mathbf{z}_j$ denotes the standard inner product in $\mathbb{S}^{d-1}$, α is an intercept representing the baseline log odds of a tie when the latent vectors are orthogonal, and β is a scaling parameter controlling how strongly angular similarity influences tie formation. Geometrically, β determines how sharply the log odds change as the angle between $\mathbf{z}_i$ and $\mathbf{z}_j$ decreases: larger values of β make the probability of a tie increase more rapidly as the two latent positions align. Lastly, we emphasize that, analogous to the Euclidean latent distance model, the latent positions $\mathbf{z}_1, \dots, \mathbf{z}_n$ serve as embeddings of actors that capture unobserved affinities, providing a foundation for visualization, clustering, and structural analysis of the network.

We constrain the latent positions to the $(d-1)$ -dimensional unit sphere $\mathbb{S}^{d-1}$, which fixes the origin and scale of the configuration, removing translation and scaling invariances, reducing each position's degrees of freedom from d to $d-1$, and thus ensuring identifiability up to a global rotation (see Section 2.2 for details). On $\mathbb{S}^{d-1}$, the inner product $\langle \mathbf{z}_i, \mathbf{z}_j \rangle = \cos(\gamma_{i,j})$ relates directly to the geodesic distance $\gamma_{i,j}$, defined here as the intrinsic (angular) distance on the sphere. We use the cosine similarity $\langle \mathbf{z}_i, \mathbf{z}_j \rangle$ rather than $\gamma_{i,j} = \arccos \langle \mathbf{z}_i, \mathbf{z}_j \rangle$ to avoid the numerical instability of the arccos gradient. The spherical formulation yields a compact latent space, preventing positions from drifting to arbitrarily large norms as in Euclidean spaces, thereby avoiding degeneracy and improving numerical stability.

In Hoff *et al.*, (2002, Sec. 2.2), the projection model, which is related but distinct from our approach, specifies

$$\eta_{i,j} = \alpha + \frac{\mathbf{z}_i^\top \mathbf{z}_j}{\|\mathbf{z}_j\|},$$

so the likelihood depends on the configuration $\{\mathbf{z}_i\}$ through inner products and norms in an unconstrained Euclidean space. As stated by Hoff *et al.* (2002), this model is invariant under common rotations and reflections of the latent positions, but not under translations and not under global rescaling. In our spherical specification the latent positions are constrained to lie on the unit sphere, $\mathbf{z}_i \in \mathbb{S}^{d-1}$, and the linear predictor depends on them only through the inner product $\langle \mathbf{z}_i, \mathbf{z}_j \rangle = \cos(\gamma_{i,j})$, where $\gamma_{i,j}$ is the geodesic distance. The spherical constraint thus fixes both the location and the overall scale of the configuration and leaves only global rotations and reflections as symmetries, which in turn simplifies identifiability and the interpretation of the latent positions, reduces each position to $d-1$ free parameters, and prevents norms from diverging.

Given the conditional independence of connections given the intercept term and latent positions, the likelihood is

$$\mathcal{L}(\alpha, \beta, \mathbf{Z}) = \prod_{i < j} p_{i,j}^{y_{i,j}} (1 - p_{i,j})^{1-y_{i,j}},$$

where $p_{i,j} = \text{logit}^{-1}(\eta_{i,j}) = \text{logit}^{-1}(\alpha + \beta \langle \mathbf{z}_i, \mathbf{z}_j \rangle)$ is the interaction probability between nodes i and j , and $\mathbf{Z} \in \mathbb{S}^{(d-1) \times n}$ is the matrix of latent positions with $\|\mathbf{z}_i\| = 1$. Recall that the inner product $\langle \mathbf{z}_i, \mathbf{z}_j \rangle$ equals $\cos(\gamma_{i,j})$, where $\gamma_{i,j}$ is the geodesic distance on the sphere, so similarity is modeled

directly via cosine values rather than angular distances. Thus, the log-likelihood takes the form

$$\ell(\alpha, \beta, \mathbf{Z}) = \sum_{i < j} [y_{ij} \eta_{ij} - \log(1 + e^{\eta_{ij}})],$$

with $\eta_{ij} = \alpha + \beta \langle \mathbf{z}_i, \mathbf{z}_j \rangle$, for which the corresponding gradients are given by

$$\begin{aligned} \frac{\partial}{\partial \mathbf{z}_k} \ell(\alpha, \beta, \mathbf{Z}) &= \sum_{i \neq k} [y_{ik} - \text{expit}(\eta_{ik})] \frac{\partial \eta_{ik}}{\partial \mathbf{z}_k}, \\ \frac{\partial}{\partial \alpha} \ell(\alpha, \beta, \mathbf{Z}) &= \sum_{i < j} [y_{ij} - \text{expit}(\eta_{ij})], \\ \frac{\partial}{\partial \beta} \ell(\alpha, \beta, \mathbf{Z}) &= \sum_{i < j} [y_{ij} - \text{expit}(\eta_{ij})] \langle \mathbf{z}_i, \mathbf{z}_j \rangle. \end{aligned}$$

Analogous expressions can be derived for the Euclidean latent distance model.

2.2 Model properties

Latent space models are non-identifiable because the linear predictor is invariant under specific transformations of the latent space. In particular, the likelihood function $\mathcal{L}$ remains unchanged under certain transformations applied to the $d \times n$ matrix $\mathbf{Z}$ containing the latent positions. Such transformations generate equivalence classes of configurations of $\mathbf{Z}$ that produce identical likelihood values.

In the Euclidean formulation, the model is invariant under isometries in $\mathbb{R}^d$ of the form $\mathbf{Z} \mapsto \mathbf{Z}' = \mathbf{Q}\mathbf{Z} + \mathbf{b}\mathbf{1}_n^\top$, where $\mathbf{Q} \in \mathbb{R}^{d \times d}$ is orthogonal, representing a rotation or reflection, $\mathbf{b} \in \mathbb{R}^d$ is a translation vector applied uniformly to all points, and $\mathbf{1}_n \in \mathbb{R}^n$ is a column vector of ones. Indeed, for transformed latent positions $\mathbf{z}'_i = \mathbf{Q}\mathbf{z}_i + \mathbf{b}$ and $\mathbf{z}'_j = \mathbf{Q}\mathbf{z}_j + \mathbf{b}$, we have

$$\|\mathbf{z}'_i - \mathbf{z}'_j\| = \|(\mathbf{Q}\mathbf{z}_i + \mathbf{b}) - (\mathbf{Q}\mathbf{z}_j + \mathbf{b})\| = \|\mathbf{Q}(\mathbf{z}_i - \mathbf{z}_j)\| = \|\mathbf{z}_i - \mathbf{z}_j\|,$$

since $\|\mathbf{Q}\mathbf{z}\| = \|\mathbf{z}\|$, for any $\mathbf{z} \in \mathbb{R}^d$ when $\mathbf{Q}$ is orthogonal. Hence, all pairwise distances are preserved, and therefore, $\mathcal{L}(\alpha, \mathbf{Z}) = \mathcal{L}(\alpha, \mathbf{Z}')$, inducing equivalence classes of $\mathbf{Z}$ that yield identical likelihoods and making the latent positions non-identifiable without further constraints.

In the spherical formulation, the latent positions are constrained to the $(d-1)$ -dimensional unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$, fixing both the origin and the scale of the configuration, removing translation and scaling invariances, and reducing each position's degrees of freedom from d to $d-1$. However, the model remains invariant under global orthogonal transformations $\mathbf{Z} \mapsto \mathbf{Z}' = \mathbf{Q}\mathbf{Z}$, where $\mathbf{Q} \in \mathbb{R}^{d \times d}$ is orthogonal, representing rotations or reflections that preserve both geodesic distances and inner products on the sphere. Indeed, for transformed positions $\mathbf{z}'_i = \mathbf{Q}\mathbf{z}_i$ and $\mathbf{z}'_j = \mathbf{Q}\mathbf{z}_j$, we have $\|\mathbf{z}'_i\| = \|\mathbf{Q}\mathbf{z}_i\| = \|\mathbf{z}_i\| = 1$, so the unit norm is preserved. Moreover,

$$\mathbf{z}'_i{}^\top \mathbf{z}'_j = (\mathbf{Q}\mathbf{z}_i)^\top (\mathbf{Q}\mathbf{z}_j) = \mathbf{z}_i^\top \mathbf{Q}^\top \mathbf{Q} \mathbf{z}_j = \mathbf{z}_i^\top \mathbf{z}_j,$$

so all pairwise inner products, and therefore all geodesic distances, are preserved. Consequently, $\mathcal{L}(\alpha, \beta, \mathbf{Z}) = \mathcal{L}(\alpha, \beta, \mathbf{Q}\mathbf{Z})$, and the configuration is identifiable only up to a global rotation or reflection. This does not affect inference about relative positions but requires post-processing (e.g., Procrustes alignment, see Borg and Groenen, 2005 for an in-depth discussion) for visualization and interpretation.

2.3 Model extensions

In addition to the baseline spherical latent distance model, the framework admits several natural extensions. One direction is to include node specific random effects or covariate terms in the linear predictor in order to capture degree heterogeneity and assortative mixing beyond what is encoded by the latent positions. Another is to adapt the model to multilayer or multiplex networks by sharing or correlating latent positions across layers while allowing layer specific parameters. A third direction is to consider dynamic networks in which latent positions evolve over time according to a stochastic process on the sphere. These extensions are outlined in more detail in Appendix B, where we provide the corresponding formulations and discuss their potential applications.

3. Inference and computation

The Euclidean and spherical formulations of latent space models introduced above involve the intercept parameter α , the scaling parameter β , and the matrix of latent positions $\mathbf{Z}$, which together determine the structure of the observed network. Once the model formulation is complete, the next step is to estimate these parameters given the observed network data $\mathbf{Y}$. This section provides a detailed treatment of both classical and Bayesian inference approaches, covering optimization techniques for point estimation as well as sampling algorithms for full posterior inference. All inference procedures described in this section are implemented in an open-source Python repository available at https://github.com/cnosa/Latentspacemodels_Networks.

3.1 Classical inference

Here, we conduct classical inference via maximum likelihood estimation (e.g., Casella and Berger, 2024), which we use primarily to obtain initial values for the latent positions and model parameters before applying MCMC-based algorithm. In latent space models, the likelihood function is generally not globally concave. This non-concavity complicates classical inference, as multiple local maxima necessitate iterative optimization methods and carefully selected initialization strategies to mitigate the risk of suboptimal convergence. To obtain the maximizer of the log-likelihood, we use gradient descent with a backtracking line search based on the Armijo condition (see Boyd and Vandenberghe, 2004 and Nocedal and Wright, 2006), ensuring sufficient decrease of the objective at each iteration even in non-concave settings.

Let $\ell(\theta, \mathbf{Z})$ denote the log-likelihood, viewed as a function of the latent positions $\mathbf{Z}$ and the other model parameters collected in θ . Starting from initial values $(\theta^{(0)}, \mathbf{Z}^{(0)})$, the parameters are updated jointly as

$$(\theta^{(b+1)}, \mathbf{Z}^{(b+1)}) = (\theta^{(b)}, \mathbf{Z}^{(b)}) + r^{(b)} \mathbf{D}^{(b)},$$

where $\mathbf{D}^{(b)} = \nabla \ell(\theta^{(b)}, \mathbf{Z}^{(b)})$ is the gradient of the log-likelihood with respect to all model parameters, and $r^{(b)}$ is a step size chosen to satisfy the Armijo condition

$$\ell(\theta^{(b+1)}, \mathbf{Z}^{(b+1)}) \geq \ell(\theta^{(b)}, \mathbf{Z}^{(b)}) + c r^{(b)} \nabla \ell(\theta^{(b)}, \mathbf{Z}^{(b)})^\top \mathbf{D}^{(b)},$$

for some constant $c \in (0, 1)$. This ensures a sufficient ascent in the log-likelihood at each step while preserving convergence guarantees.

Because the log-likelihood landscape is highly sensitive to initialization, we examine multiple choices for the starting point $(\theta^{(0)}, \mathbf{Z}^{(0)})$, including random configurations. In the spherical model, optimization is additionally constrained by the manifold geometry, requiring each latent position to remain on the unit sphere. This is enforced by projecting the updated positions back onto the sphere after each gradient step. The procedure is summarized in Algorithm 1. Given these challenges and the model's structure, classical inference may not always produce

Algorithm 1 Gradient ascent with Armijo backtracking line search for maximum likelihood estimation in latent space models

Require: Initial parameters $(\theta^{(0)}, \mathbf{Z}^{(0)})$, step size parameters $\eta > 0$, $\rho \in (0, 1)$, Armijo constant $c \in (0, 1)$

```

1: Set  $k \leftarrow 0$ 
2: while not converged do
3:   Compute gradient:  $\mathbf{D}^{(b)} \leftarrow \nabla \ell(\theta^{(b)}, \mathbf{Z}^{(b)})$ 
4:   Initialize step size:  $r^{(b)} \leftarrow \eta$ 
5:   while Armijo condition not satisfied do
6:     Check if  $\ell((\theta^{(b)}, \mathbf{Z}^{(b)}) + r^{(b)} \mathbf{D}^{(b)}) \geq \ell(\theta^{(b)}, \mathbf{Z}^{(b)}) + c r^{(b)} \nabla \ell(\theta^{(b)}, \mathbf{Z}^{(b)})^\top \mathbf{D}^{(b)}$ 
7:     if condition not satisfied then
8:        $r^{(b)} \leftarrow \rho r^{(b)}$ 
9:     end if
10:  end while
11:  Update parameters:  $(\theta^{(b+1)}, \mathbf{Z}^{(b+1)}) \leftarrow (\theta^{(b)}, \mathbf{Z}^{(b)}) + r^{(b)} \mathbf{D}^{(b)}$ 
12:  if spherical model then
13:    Project each row of  $\mathbf{Z}^{(b+1)}$  onto the unit sphere
14:  end if
15:   $k \leftarrow k + 1$ 
16: end while
17: Output: Estimates  $(\theta^{\text{ML}}, \mathbf{Z}^{\text{ML}})$ 

```

reliable or interpretable estimates. This motivates the use of Bayesian inference, which not only accounts for parameter uncertainty but also incorporates regularization naturally through the prior distribution on the model parameters.

3.2 Bayesian inference

Bayesian inference offers a flexible alternative to traditional statistical learning methods (e.g., Gelman *et al.*, 2013a), providing both parameter estimation and uncertainty quantification through the full posterior distribution $p(\theta, \mathbf{Z} | \mathbf{Y})$. To perform full Bayesian inference, it is first necessary to specify a sensible prior distribution $p(\theta, \mathbf{Z})$. To do so, we adopt independent priors for the latent positions $\mathbf{z}_1, \dots, \mathbf{z}_n$, chosen to encode plausible assumptions about their spatial configuration before observing the data, and separately assign a prior distribution to the remaining model parameters collected in θ . These priors regularize while respecting the model's invariances, preventing degeneracy and stabilizing computation.

In the Euclidean case, we assign an isotropic Gaussian prior to the latent positions, $\mathbf{z}_i \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I}_d)$, for $i = 1, \dots, n$, and independently set $\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$. This prior is rotationally

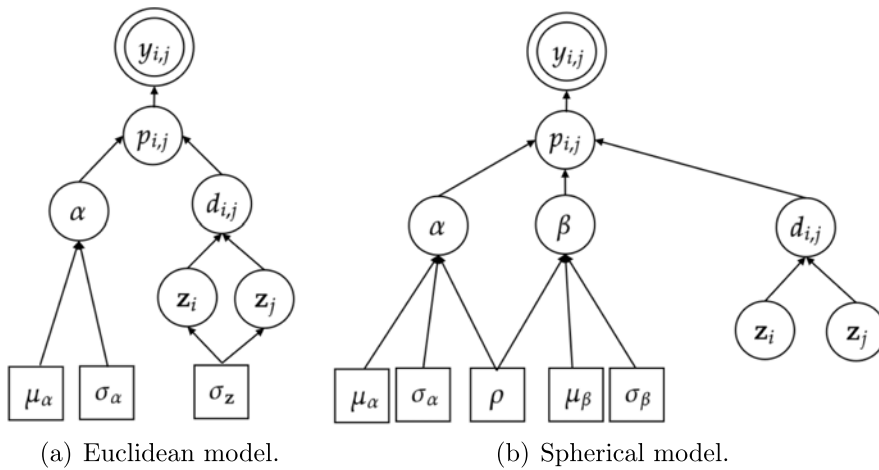


Figure 1. Directed acyclic graphs (DAGs) for the latent space models, where circles denote random variables or random vectors, squares indicate fixed parameters, and edges represent dependencies.

and translationally symmetric, aligning with the likelihood’s invariance under orthogonal transformations. It avoids directional bias, preserves the geometry of the latent space, and softly shrinks positions toward the origin, thereby controlling the unbounded nature of Euclidean embeddings and preventing norm divergence.

In the spherical case, we place a uniform prior on the hypersphere, $\mathbf{z}_i \stackrel{\text{iid}}{\sim} \mathcal{U}(\mathbb{S}^{d-1})$, for $i = 1, \dots, n$, and a correlated Gaussian prior on (α, β) with $\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$, $\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$, and $\text{Cov}(\alpha, \beta) = \rho \sigma_\alpha \sigma_\beta$. The joint prior on (α, β) allows for prior dependence between baseline tie propensity and the strength of the latent space term, since both influence overall network density in logistic latent space models. This motivates a bivariate normal prior with unrestricted covariance rather than independent priors. In practice, however, empirical results are essentially unchanged under independent normal priors, so readers who prefer a simpler specification can safely use independent priors for α and β . The uniform prior corresponds to the uniform measure on $\mathbb{S}^{d-1}$, ensuring invariance under all orthogonal transformations in $\mathbb{R}^d$. This symmetry makes all orientations of the latent configuration equally likely, fully respecting the rotational and reflectional invariance of the spherical latent space.

To clarify the probabilistic structure of the models, we present in Figure 1 the directed acyclic graphs (DAGs) corresponding to both the Euclidean and spherical formulations. These graphical representations explicitly illustrate the conditional dependencies among observed ties, latent positions, and global parameters, making clear the conditional independence assumptions underlying the model specification.

Under the prior specifications given above, let the interaction probabilities be defined as $p_{i,j} = \text{logit}^{-1}(\eta_{i,j})$, where $\eta_{i,j} = \eta_{i,j}(\theta, \mathbf{z}_i, \mathbf{z}_j)$ is a linear predictor specified as in Section 2.1, depending on the case. The posterior distribution for each model is proportional to:

- Euclidean latent distance model:

$$p(\mathbf{Z}, \alpha \mid Y) \propto p(Y \mid \alpha, \mathbf{Z}) \cdot \prod_i p(\mathbf{z}_i) \cdot p(\alpha) \\ \propto \prod_{i < j} p_{i,j}^{y_{i,j}} (1 - p_{i,j})^{1-y_{i,j}} \cdot \exp\left(-\frac{1}{2\sigma_z^2} \sum_i \|\mathbf{z}_i\|^2\right) \cdot \exp\left(-\frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2\right),$$

where μ_α , σ_α^2 , and σ_z^2 are known hyperparameters.

- Spherical latent distance model:

$$p(\alpha, \beta, \mathbf{Z} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \alpha, \beta, \mathbf{Z}) \cdot \prod_i p(\mathbf{z}_i) \cdot p(\alpha, \beta) \\ = \prod_{i < j} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}} \cdot \exp \left(-\frac{1}{2(1-\rho^2)} Q(\alpha, \beta) \right),$$

where $\mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2, \rho$, and σ_z^2 are known hyperparameters, and

$$Q(\alpha, \beta) = \frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2} - \frac{2\rho \alpha(\beta - \mu_\beta)}{\sigma_\alpha \sigma_\beta} + \frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2}.$$

Posterior summaries such as the posterior mean ($\theta^{\text{PM}}, \mathbf{Z}^{\text{PM}} = \mathbb{E}(\theta, \mathbf{Z} \mid \mathbf{Y})$) and the maximum a posteriori ($\theta^{\text{MAP}}, \mathbf{Z}^{\text{MAP}} = \arg \max p(\theta, \mathbf{Z} \mid \mathbf{Y})$) provide interpretable point estimates of the model parameters, capturing the central tendency of the latent structure, while credible regions derived from the posterior distribution quantify the uncertainty associated with these estimates.

3.2.1 Sampling algorithms

The posterior distributions of the latent space models described here are high-dimensional and analytically intractable. To address this challenge, we approximate the posterior using MCMC methods (e.g., Gamerman and Lopes, 2006), specifically the Metropolis–Hastings (MH) algorithm (e.g., Albert *et al.*, 2009), the Hamiltonian Monte Carlo (HMC) algorithm (e.g., Gelman *et al.*, 2013a), and the Spherical HMC algorithm (e.g., Yu, 2020), which are developed throughout this section. These approaches provide the flexibility to design geometry-aware proposal distributions tailored to both Euclidean and spherical latent spaces.

Metropolis–Hastings algorithm

Let $\mathbf{Z}$ denote the matrix of latent positions, and let θ represent the global model parameters (α in the Euclidean case, and both α and β in the spherical formulation). The unnormalized log-posterior density is

$$\psi(\theta, \mathbf{Z}) = \log p(\mathbf{Y} \mid \theta, \mathbf{Z}) + \log p(\mathbf{Z}) + \log p(\theta),$$

which is the sum of the log-likelihood and the log-prior densities for the latent positions and the model parameters. The Markov chain is initialized by sampling θ and $\mathbf{Z}$ from their prior distributions, although maximum likelihood estimates may also serve as effective starting values. We employ a Metropolis algorithm with symmetric proposal distributions, adjusting their variances during the burn-in phase to achieve target acceptance rates, typically between 30% and 50%. The joint posterior is explored by alternately updating θ and $\mathbf{Z}$ via Metropolis steps, with minor procedural differences between the Euclidean and spherical formulations.

Latent positions are updated sequentially by proposing a candidate $\mathbf{z}_i^*$, for each $i = 1, \dots, n$, one at a time. In the Euclidean case, proposals are drawn from a multivariate Normal distribution centered at the current position, $\mathbf{z}_i^* \sim \mathcal{N}(\mathbf{z}_i^{(b)}, \tau_z \mathbf{I}_d)$, where τ_z is a tuning parameter. In the spherical case, following Wood (1994) and Mardia and Jupp (2000), proposals are drawn from a von Mises–Fisher distribution with mean direction equal to the current position, $\mathbf{z}_i^* \sim \text{vMF}(\mathbf{z}_i^{(b)}, \tau_z)$. The proposed value is accepted with probability $\min \{1, \exp[\psi(\theta^{(b)}, \mathbf{z}_i^*) - \psi(\theta^{(b)}, \mathbf{z}_i^{(b)})]\}$, where the log-posterior is evaluated holding all latent positions fixed except for node i . If accepted, the proposed value becomes the new state $\mathbf{z}_i^{(b+1)} = \mathbf{z}_i^*$; otherwise, the position remains unchanged.

The global parameters are updated sequentially. For each parameter, a candidate θ^* is proposed from a Normal distribution centered at the current value, $\theta^* \sim \mathcal{N}(\theta^{(b)}, \tau_\theta)$, where τ_θ is once again a tuning parameter. The acceptance probability is computed as

$\min\{1, \exp[\psi(\theta^*, \mathbf{Z}^{(b)}) - \psi(\theta^{(b)}, \mathbf{Z}^{(b)})]\}$, analogous to the latent position updates, and the proposed value is accepted or rejected accordingly.

Finally, to account for isometric invariance in the latent space, each sampled configuration $\mathbf{Z}^{(b+1)}$ is aligned to the maximum likelihood estimate $\mathbf{Z}^{\text{ML}}$ using a Procrustes transformation (e.g., Borg and Groenen, 2005). In the Euclidean case, the transformation $\mathbf{T}$ minimizes the trace of the squared distances:

$$\tilde{\mathbf{Z}}^{(b+1)} = \arg \min_{\mathbf{T}} \text{tr} \left[(\mathbf{Z}^{\text{ML}} - \mathbf{T} \mathbf{Z}^{(b+1)})^\top (\mathbf{Z}^{\text{ML}} - \mathbf{T} \mathbf{Z}^{(b+1)}) \right].$$

In the spherical case, the alignment is performed via an orthogonal transformation $\mathbf{T}$ that minimizes the Frobenius norm (e.g., Golub and Van Loan, 2013):

$$\tilde{\mathbf{Z}}^{(b+1)} = \arg \min_{\mathbf{T}} \|\mathbf{Z}^{\text{ML}} - \mathbf{T} \mathbf{Z}^{(b+1)}\|_F,$$

with $\mathbf{T}^\top \mathbf{T} = \mathbf{I}_n$.

After discarding an initial burn-in period and applying thinning when necessary, the retained posterior samples can be used to compute summary statistics of interest, including point estimates and marginal credible intervals for all model parameters, including latent positions. Convergence and mixing are evaluated using standard MCMC diagnostics across multiple chains, such as the effective sample size (ESS) and the potential scale reduction factor $\hat{R}$. The complete estimation procedure is outlined in Algorithm 2.

Hamiltonian Monte Carlo algorithm

Alongside the Metropolis–Hastings algorithm we consider two advanced sampling methods, HMC and GHMC. These approaches exploit gradient information and the underlying geometric structure to explore the posterior distribution more efficiently, particularly in high dimensional or curved latent spaces. In our latent space models we use HMC to update unconstrained global parameters and Euclidean latent coordinates, and we use GHMC for latent positions constrained to lie on the hypersphere. Each Hamiltonian update targets the full conditional distribution of either the parameter vector θ or a single latent position $\mathbf{z}_i$, with all remaining components held fixed, which leads to an MCMC scheme where Hamiltonian steps are applied one at a time to the components of $(\theta, \mathbf{Z})$. While these updates can achieve higher sampling efficiency by accounting for the curvature of the full conditional distributions, they are computationally more demanding because they require gradient evaluations and auxiliary momentum variables. For clarity we first recall the generic HMC and GHMC schemes in abstract form before describing their implementation for the latent space models.

The HMC algorithm expands the parameter space by introducing an auxiliary momentum variable $\mathbf{p}$, which enables the use of Hamiltonian dynamics to generate proposals (Neal, 2011). For a target unnormalized density $f(\mathbf{x})$, the Hamiltonian is given by

$$H(\mathbf{x}, \mathbf{p}) = U(\mathbf{x}) + \frac{1}{2} \mathbf{p}^\top \mathbf{M}^{-1} \mathbf{p},$$

where $U(\mathbf{x}) = -\log f(\mathbf{x})$ represents the potential energy associated with the unnormalized log-density $\log f(\mathbf{x})$, and $\mathbf{M}$ is a symmetric positive-definite mass matrix, often chosen as the identity for simplicity.

The dynamics are determined by Hamilton's equations and are numerically integrated via the leapfrog method using a fixed step size ϵ and a predetermined number of steps L (Sanz-Serna and Calvo, 2018). These algorithmic parameters are tuned during execution to balance computational efficiency and acceptance probability, which typically lies between 60% and 70%. A proposal state is obtained by simulating the Hamiltonian dynamics forward in time, after which it is accepted or rejected according to a Metropolis rule that preserves detailed balance. This approach enables the

Algorithm 2 Metropolis–Hastings algorithm for latent space models**Require:** Observed data $\mathbf{Y}$, priors for θ and $\mathbf{Z}$, number of iterations B

- 1: Compute $(\theta^{\text{ML}}, \mathbf{Z}^{\text{ML}})$ via optimization using Algorithm 1
- 2: Initialize $(\mathbf{Z}^{(0)}, \theta^{(0)})$
- 3: **for all** chains **do**
- 4: **for** $b = 1$ **to** B **do**
- 5: **for each** node $i = 1, \dots, n$ **do**
- 6: Propose $\mathbf{z}_i^*$:
 Euclidean: $\mathbf{z}_i^* \sim \mathcal{N}(\mathbf{z}_i^{(b)}, \tau_z \mathbf{I}_d)$
 Spherical: $\mathbf{z}_i^* \sim \text{VMF}(\mathbf{z}_i^{(b)}, \tau_z)$
- 7: Accept $\mathbf{z}_i^*$ with probability $\min \{1, \exp[\psi(\theta^{(b)}, \mathbf{z}_i^*) - \psi(\theta^{(b)}, \mathbf{z}_i^{(b)})]\}$
- 8: Update $\mathbf{z}_i^{(b+1)} \leftarrow \mathbf{z}_i^*$ if accepted, else keep $\mathbf{z}_i^{(b)}$
- 9: **end for**
- 10: **for each** global parameter θ_j **do**
- 11: Propose $\theta_j^* \sim \mathcal{N}(\theta_j^{(b)}, \tau_\theta)$
- 12: Accept with probability $\min \{1, \exp[\psi(\theta_j^*, \mathbf{Z}^{(b)}) - \psi(\theta_j^{(b)}, \mathbf{Z}^{(b)})]\}$
- 13: Update $\theta_j^{(b+1)} \leftarrow \theta_j^*$ if accepted, else keep $\theta_j^{(b)}$
- 14: **end for**
- 15: Align $\mathbf{Z}^{(b+1)}$ to $\mathbf{Z}^{\text{ML}}$ using the corresponding transformation
- 16: **end for**
- 17: **end for**
- 18: Discard burn-in and apply thinning
- 19: **Output:** Chains of posterior samples $(\theta^{(1)}, \mathbf{Z}^{(1)}), \dots, (\theta^{(B)}, \mathbf{Z}^{(B)})$

sampler to perform large, gradient-informed moves across the posterior space, offering substantial efficiency gains over regular random-walk proposals (Neal, 2011). Algorithm 3 outlines the HMC procedure for sampling from a general target distribution. For the Euclidean latent space model we apply the generic HMC step to the full conditional of each latent coordinate. In the notation of Algorithm 3 we take $x = \mathbf{z}_i$ and $f(x) \propto \exp\{\psi(\theta, \mathbf{Z})\}$ with all components except $\mathbf{z}_i$ fixed, and we use the same scheme for blocks of unconstrained global parameters when convenient.

Spherical Hamiltonian Monte Carlo algorithm

Standard Hamiltonian Monte Carlo (HMC) performs effectively in flat Euclidean spaces, but it can be restrictive when the posterior distribution is supported on a curved manifold. To handle parameters constrained to lie on the hypersphere, in particular the latent positions $\mathbf{z}_i \in \mathbb{S}^{d-1}$ in the spherical model, we adopt Geodesic Hamiltonian Monte Carlo (GHMC), an extension of HMC designed for embedded Riemannian manifolds such as spheres or Stiefel manifolds and inspired

Algorithm 3 Hamiltonian Monte Carlo algorithm

Require: Target density $f(\mathbf{x})$, step size ϵ , number of leapfrog steps L , mass matrix $\mathbf{M}$

- 1: Initialize $\mathbf{x}^{(0)}$
- 2: **for** $b = 1$ **to** B **do**
- 3: Sample $\mathbf{p}^{(b)} \sim \mathcal{N}(\mathbf{0}, \mathbf{M})$
- 4: Set $\mathbf{x}^* \leftarrow \mathbf{x}^{(b)}$ and $\mathbf{p}^* \leftarrow \mathbf{p}^{(b)}$
- 5: **for** each leapfrog step $\ell = 1, \dots, L$ **do**
- 6: $\mathbf{p}^* \leftarrow \mathbf{p}^* + \frac{\epsilon}{2} \nabla \log f(\mathbf{x}^*)$
- 7: $\mathbf{x}^* \leftarrow \mathbf{x}^* + \epsilon \mathbf{M}^{-1} \mathbf{p}^*$
- 8: $\mathbf{p}^* \leftarrow \mathbf{p}^* + \frac{\epsilon}{2} \nabla \log f(\mathbf{x}^*)$
- 9: **end for**
- 10: Compute $H_{\text{current}} = -\log f(\mathbf{x}^{(b)}) + \frac{1}{2} \mathbf{p}^{(b)\top} \mathbf{M}^{-1} \mathbf{p}^{(b)}$
- 11: Compute $H_{\text{proposed}} = -\log f(\mathbf{x}^*) + \frac{1}{2} \mathbf{p}^{*\top} \mathbf{M}^{-1} \mathbf{p}^*$
- 12: Accept $\mathbf{x}^*$ with probability $\min \{1, \exp[H_{\text{current}} - H_{\text{proposed}}]\}$
- 13: Update $\mathbf{x}^{(b+1)} \leftarrow \mathbf{x}^*$ if accepted, else $\mathbf{x}^{(b+1)} \leftarrow \mathbf{x}^{(b)}$
- 14: **end for**
- 15: Discard burn-in and apply thinning
- 16: **Output:** Chains of posterior samples $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(B)}$

by the Riemannian Manifold HMC framework of Girolami and Calderhead (2011). In the spherical latent space model, however, the latent positions are constrained to lie on the hypersphere $\mathbb{S}^{d-1}$. Standard HMC algorithms are formulated for unconstrained Euclidean parameter spaces, and a direct application of a generic Euclidean HMC update would generate proposals outside the manifold. Enforcing the spherical constraint would then require projection steps that disrupt the symplectic structure of the Hamiltonian dynamics and can degrade mixing. In particular, Stan's default No-U-Turn Sampler (NUTS) implementation assumes unconstrained Euclidean parameters and would require nontrivial reparameterization together with Jacobian adjustments to accommodate spherical constraints. By contrast, the GHMC algorithm operates directly on the embedded sphere, preserving the manifold constraint intrinsically and evolving proposals along exact geodesic flows without the need for reparameterization.

Unlike RMHMC, which requires computing local metric tensors and numerically integrating the geodesic equations, GHMC exploits the manifold's known geodesic structure to evolve trajectories exactly along curvature aware paths. In this setting, the target density is expressed with respect to the Hausdorff measure (Federer, 2014), leading to the modified Hamiltonian

$$H_{\mathcal{H}}(\mathbf{x}, \mathbf{p}) = -\log f_{\mathcal{H}}(\mathbf{x}) + \frac{1}{2} \mathbf{p}^{\top} \mathbf{G}(\mathbf{x})^{-1} \mathbf{p},$$

where $\mathbf{G}(\mathbf{x})$ is the intrinsic metric tensor of the manifold and $f_{\mathcal{H}}(\mathbf{x})$ is the target density adjusted by the Riemannian volume element $\sqrt{|\mathbf{G}(\mathbf{x})|}$ (Byrne and Girolami, 2013). In our application each

GHMC update targets the full conditional distribution of a single latent position $\mathbf{z}_i$ on $\mathbb{S}^{d-1}$ and the proposal follows a geodesic flow on the sphere.

GHMC alternates between exact geodesic updates of position and velocity (the kinetic flow) and momentum adjustments driven by the gradient of the log-density (the potential flow). On the hypersphere, these geodesic updates admit closed-form solutions involving trigonometric functions of the angular velocity, enabling efficient and numerically stable integration. After a predetermined number of integration steps, a Metropolis acceptance step ensures that the resulting Markov chain targets the correct posterior distribution. To preserve the validity of the dynamics, the momentum is projected onto the tangent space of the manifold after each update. By explicitly exploiting the manifold's geometric structure, GHMC retains the principal advantages of HMC, such as long-distance proposals and reduced random-walk behavior, while remaining well-defined on curved latent spaces. In contrast to RMHMC, which requires computing and differentiating a position-dependent metric tensor at each iteration (Girolami and Calderhead, 2011), GHMC avoids these computational burdens by relying on analytically tractable geodesic flows, making it a scalable alternative for manifold-constrained models (Yu, 2020). Algorithm 4 presents the GHMC procedure for sampling from an arbitrary density supported on a spherical manifold. This algorithm summarizes one update for a latent position in the spherical model. In this case we set $x = \mathbf{z}_i$ and $f(x) \propto \exp\{\psi(\boldsymbol{\theta}, \mathbf{Z})\}$ restricted to $\mathbb{S}^{d-1}$, and the projection and geodesic moves act directly on the latent coordinate $\mathbf{z}_i$.

Hamiltonian-based algorithms must be adapted to latent space network models by tuning hyperparameters, updating parameters sequentially to improve acceptance rates, and respecting the geometry of the latent space. Convergence diagnostics ensure adequate posterior exploration and reliability of inferences, while multimodality and identifiability issues require running multiple chains from diverse initializations and checking the consistency of marginal posteriors. Geometric constraints add further complexity, necessitating careful algorithmic design to balance exploration efficiency, numerical stability, and convergence guarantees.

For readers interested in a concrete illustration of these algorithms in a controlled setting, Appendix A presents a detailed comparison of Metropolis–Hastings, HMC, and GHMC for sampling from the bivariate Rosenbrock density. This target has strong curvature and non quadratic contours, and is a standard benchmark for assessing the ability of MCMC schemes to explore narrow, curved posterior regions. The example reports ESSs, acceptance rates, and trace plots, which help to visualize the efficiency gains of Hamiltonian based methods when needed.

4. Illustration: Florentine families

We illustrate the latent space models using two benchmark social science networks originally analyzed by Hoff *et al.* (2002). The main text presents results for the Florentine families data, while the supplementary material reports findings for the monks data. These datasets differ in scale, domain, and type of interaction, offering a diverse testbed for assessing model behavior and interpretability. For each, we fit the proposed model to the observed relational data and analyze the latent structure, inference results, and model fit via posterior summaries and predictive checks. The examples include a historical marriage network among Florentine families and interpersonal ties among monks in a monastery.

In historical social network analysis, relationships between influential families reveal power dynamics, alliances, and social structures. A notable example is the study by Padgett and Ansell (1993), which examines marriage and business ties among 16 prominent Florentine families in the 15th century, drawing on historical accounts that document the socio-political landscape of Renaissance Florence.

This model focuses on marriage relations between these families, where a tie indicates at least one marriage between them. The network is undirected. One of the 16 families, having no marriage ties, was excluded to avoid infinite distances in maximum likelihood estimation and large

Algorithm 4 Spherical Hamiltonian Monte Carlo algorithm

Require: Target density $f(\mathbf{x})$, step size ϵ , number of leapfrog steps L

- 1: Initialize $\mathbf{x}^{(0)}$
- 2: **for** $b = 1$ **to** B **do**
- 3: Sample momentum $\mathbf{p}^{(b)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$
- 4: Project onto the tangent space of the manifold $\mathbf{p}^{(b)} \leftarrow (\mathbf{I}_d - \mathbf{x}^{(b)}\mathbf{x}^{(b)\top})\mathbf{p}^{(b)}$
- 5: Set $\mathbf{x}^* \leftarrow \mathbf{x}^{(b)}$ and $\mathbf{p}^* \leftarrow \mathbf{p}^{(b)}$
- 6: **for** each leapfrog step $\ell = 1, \dots, L$ **do**
- 7: $\mathbf{p}^* \leftarrow \mathbf{p}^* + \frac{\epsilon}{2} \nabla \log f(\mathbf{x}^*)$
- 8: Project $\mathbf{p}^* \leftarrow (\mathbf{I}_d - \mathbf{x}^*\mathbf{x}^{*\top})\mathbf{p}^*$
- 9: Let $v \leftarrow \|\mathbf{p}^*\|$
- 10: Geodesic flow update:

$$\mathbf{x}^* \leftarrow \mathbf{x}^* \cos(v\epsilon) + \frac{1}{v} \mathbf{p}^* \sin(v\epsilon)$$

$$\mathbf{p}^* \leftarrow \mathbf{p}^* \cos(v\epsilon) - v \mathbf{x}^* \sin(v\epsilon)$$
- 11: $\mathbf{p}^* \leftarrow \mathbf{p}^* + \frac{\epsilon}{2} \nabla \log f(\mathbf{x}^*)$
- 12: Project $\mathbf{p}^* \leftarrow (\mathbf{I}_d - \mathbf{x}^*\mathbf{x}^{*\top})\mathbf{p}^*$
- 13: **end for**
- 14: Compute $H_{\mathcal{H} \text{ current}} = -\log f(\mathbf{x}^{(b)}) + \frac{1}{2}\mathbf{p}^{(b)\top}\mathbf{p}^{(b)}$
- 15: Compute $H_{\mathcal{H} \text{ proposed}} = -\log f(\mathbf{x}^*) + \frac{1}{2}\mathbf{p}^{*\top}\mathbf{p}^*$
- 16: Accept $\mathbf{x}^*$ with probability $\min\{1, \exp[H_{\mathcal{H} \text{ current}} - H_{\mathcal{H} \text{ proposed}}]\}$
- 17: Update $\mathbf{x}^{(b+1)} \leftarrow \mathbf{x}^*$ if accepted, else $\mathbf{x}^{(b+1)} \leftarrow \mathbf{x}^{(b)}$
- 18: **end for**
- 19: Discard burn-in and apply thinning
- 20: **Output:** Chains of posterior samples $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(B)}$

finite distances in a Bayesian framework. Modeling these relations with network analysis techniques reveals the structural properties of the Florentine elite and their strategic matrimonial alliances, offering insights into how marriage shaped political and economic power during the Renaissance. Figure 2 shows the network, its relabeling, and the sociomatrix.

We fit the models in $\mathbb{R}^1$, $\mathbb{R}^2$, $\mathbb{R}^3$, $\mathbb{S}^1$, and $\mathbb{S}^2$ using Algorithms 2, 3, and 4, with hyperparameters $\sigma_{\mathbf{z}} = 5.0$ and $\sigma_{\alpha} = 5.0$ for the Euclidean model, and $\rho = -0.5$, $\mu_{\alpha} = 0.0$, $\sigma_{\alpha} = 1.0$, $\mu_{\beta} = 10.0$, and $\sigma_{\beta} = 5.0$ for the spherical model. Hyperparameters were selected via sensitivity analysis to optimize information and predictive criteria (supplementary material). While Hamiltonian dynamics methods (Algorithms 3 and 4) yield higher ESS per iteration, they require gradient evaluations that increase computation time and may introduce numerical instabilities.

To examine this tradeoff directly within the Florentine network model, we estimated both the Euclidean model in $\mathbb{R}^3$ and the spherical model in $\mathbb{S}^2$ using Metropolis–Hastings (MH),

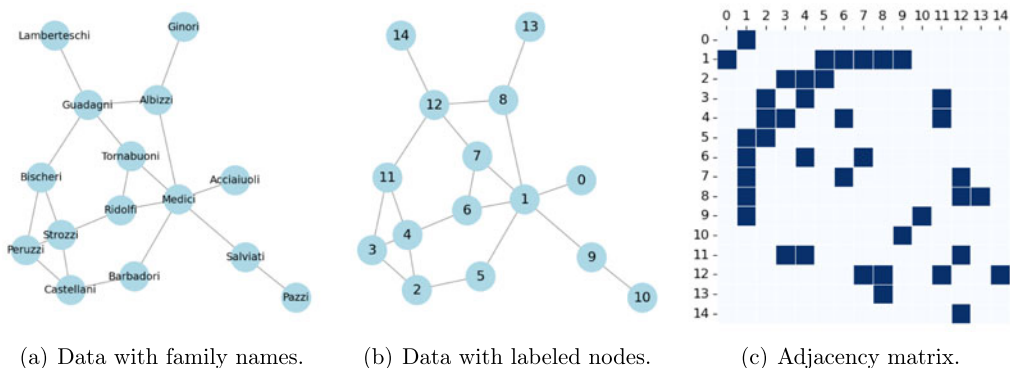


Figure 2. Visualization of the marriage network among Florentine families.

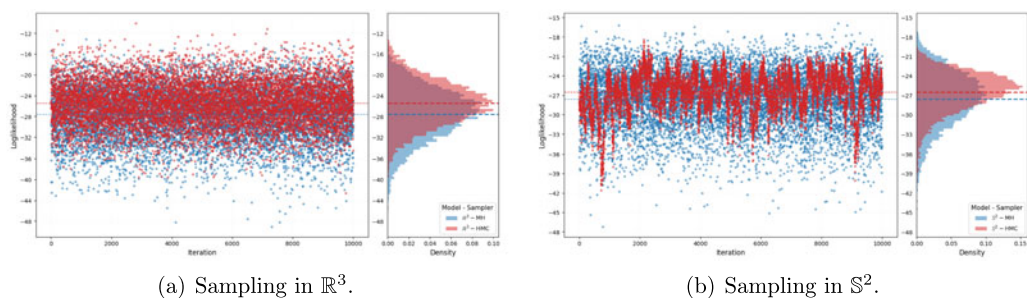


Figure 3. Likelihood traces for latent position model parameters in $\mathbb{R}^3$ and $\mathbb{S}^2$ for the Florentine families example, obtained using Metropolis and Hamiltonian based sampling algorithms.

Hamiltonian Monte Carlo (HMC), and Geodesic Hamiltonian Monte Carlo (GHMC). For the $\mathbb{R}^3$ model, latent positions were initialized from anisotropic Gaussian distributions with $\alpha^{(0)} = 0$. We ran two chains of 10,000 posterior samples after a burn-in of 10,000 iterations with thinning of 50. Proposal scales were $\sigma_{q,Z} = \sigma_{q,\alpha} = 1.0$, and HMC step sizes were $\epsilon_Z = \epsilon_\alpha = 0.1$ with $L = \lfloor \epsilon^{-1} \rfloor$. For the $\mathbb{S}^2$ model, latent positions were initialized from a von Mises–Fisher distribution centered at $(0, 0, 1)$ with concentration $\kappa = 0$, ensuring uniform initialization on the sphere. We used analogous chain lengths and tuning strategies, with Geodesic HMC step sizes $\epsilon_Z = \epsilon_\alpha = \epsilon_\beta = 0.1$ and trajectory lengths defined similarly. During burn-in, proposal scales were tuned to achieve acceptance rates between 30% and 50% for MH and between 60% and 80% for HMC and GHMC.

Figure 3 shows the likelihood traces for each sampler, and Table 1 summarizes convergence diagnostics, ESS, and computational cost. Across geometries, Hamiltonian-based samplers generally achieve higher minimum relative ESS than MH, indicating more efficient posterior exploration per iteration, at the expense of increased computational cost due to gradient evaluations. However, the behavior of GHMC differs from standard HMC. In the spherical setting, GHMC exhibits stronger autocorrelation and slower mixing in the trace plots, along with a tendency to explore a narrower region of the posterior distribution. This pattern is consistent with the interaction between manifold curvature and the numerical integration of the geodesic flow, which can increase sensitivity to tuning parameters. Despite these differences in mixing behavior, posterior summaries remain qualitatively consistent across samplers, indicating that the choice of algorithm primarily affects computational efficiency and sampling dynamics rather than substantive inference.

Table 1. Comparison of MH, HMC, and GHMC for the Florentine families latent space model in $\mathbb{R}^3$ and $\mathbb{S}^2$

Model	$\mathbb{R}^3$		$\mathbb{S}^2$	
	MH	HMC	MH	GHMC
Execution time (seconds)	1446	11154	2112	6623
Maximum $\hat{R}$	1.0007	1.0010	1.0010	1.0994
Minimum relative ESS	0.4786	0.5867	0.6021	0.7050
Acceptance rate	0.3500	0.6600	0.4500	0.6800

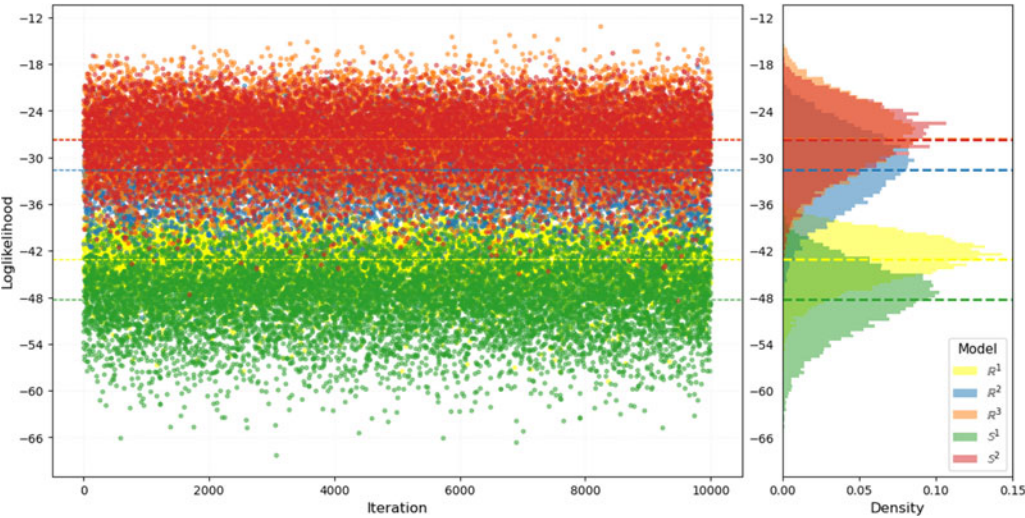


Figure 4. Log-likelihood trajectories for all models obtained during MCMC sampling.

Given this qualitative agreement across samplers, and in order to maintain computational comparability across all geometric specifications and dimensions, we use the MH algorithm for the reported results. Convergence was confirmed through standard diagnostics (Gelman *et al.*, 2013a). We then assess model fit and performance by inspecting log-likelihood trajectories from MH sampling (Figure 4). All models exhibit good convergence, with the Euclidean model in $\mathbb{R}^3$ achieving the highest likelihood, followed by the spherical model in $\mathbb{S}^2$.

Table 2 reports the Watanabe–Akaike information criteria (e.g., Watanabe, 2013, Gelman *et al.*, 2014) for each model. Spherical models achieve log-likelihood values close to their Euclidean counterparts, indicating comparable ability to capture network structure despite geometric constraints that reduce the effective degrees of freedom: for n nodes and dimension d , a model in $\mathbb{R}^d$ has $nd + 1$ degrees of freedom, whereas a model in $\mathbb{S}^{d-1}$ has $n(d - 1) + 2$ due to constraints on latent positions. WAIC favors the Euclidean model in $\mathbb{R}^3$, with the one in $\mathbb{R}^2$ close behind, while the model in $\mathbb{S}^2$ attains values close to the best Euclidean specifications and the model in $\mathbb{S}^1$ performs noticeably worse, reflecting the limited flexibility of a circle. Overall, the spherical models remain competitive in terms of fit, with $\mathbb{S}^2$ in particular combining parsimony, geometric interpretability, and good likelihood based performance.

Figure 5 shows the estimated node positions for the models in $\mathbb{R}^1$ and $\mathbb{S}^2$. Both Euclidean and spherical embeddings preserve much of the original graph structure, but the circular layout provides a more compact configuration with clearer group separation, potentially reflecting structural features such as hierarchical clustering or cyclical subgroups. Average posterior variance was

Table 2. Model comparison metrics for each latent space specification. The mean log-likelihood is averaged over posterior samples, while ML, MAP, and CM denote the log-likelihood evaluated at the maximum likelihood estimate, the maximum a posteriori estimate, and the posterior mean, respectively

	$\mathbb{R}^1$	$\mathbb{R}^2$	$\mathbb{R}^3$	$\mathbb{S}^1$	$\mathbb{S}^2$
WAIC	207.860	170.072	161.099	384.106	199.367
Mean log-lik.	−43.105	−31.639	−27.642	−48.244	−27.692
ML log-lik.	−47.632	−29.819	−26.517	−40.270	−31.852
MAP log-lik.	−48.392	−37.978	−33.138	−46.553	−16.570
CM log-lik.	−51.085	−36.771	−26.935	−45.115	−13.469

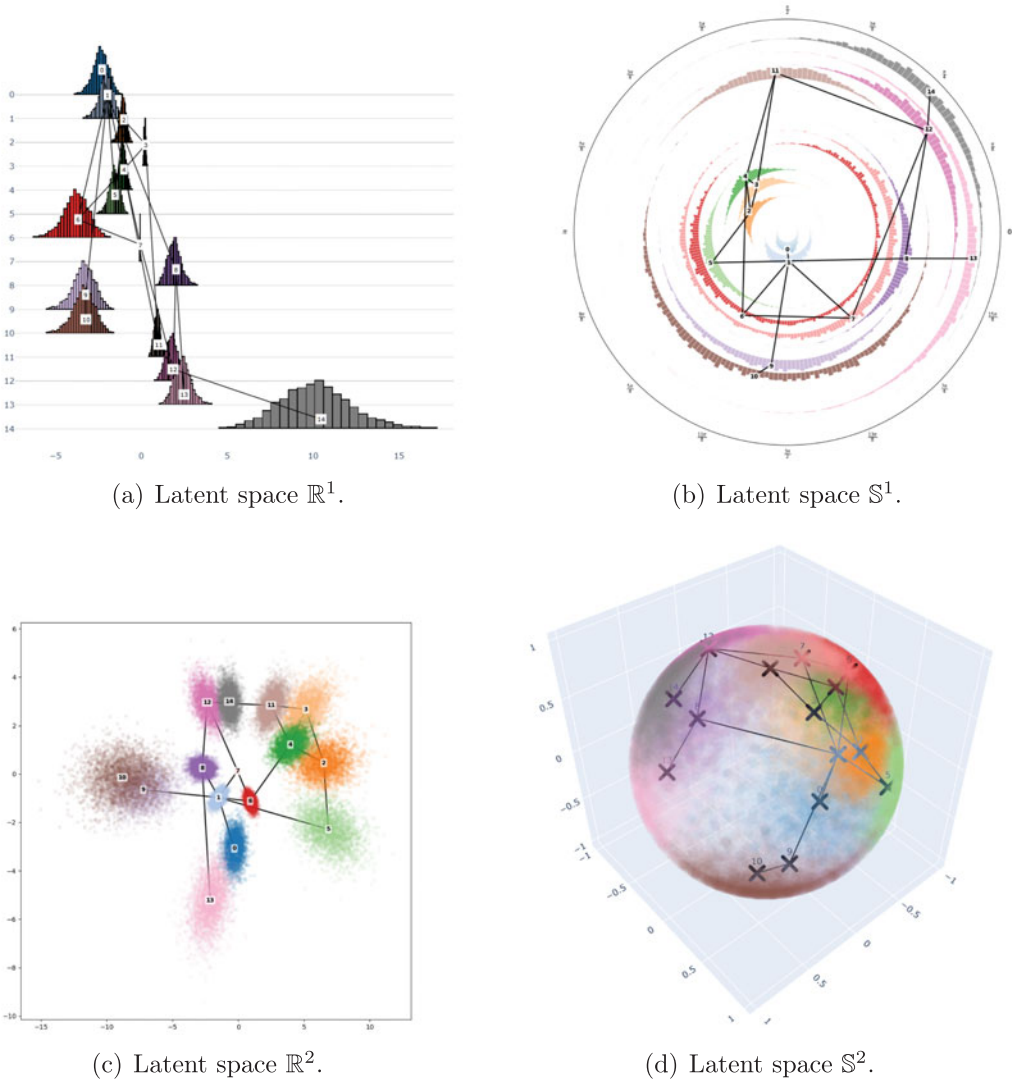


Figure 5. Latent space representations for $\mathbb{R}^1$ and $\mathbb{S}^1$ models, showing each node’s posterior distribution and mean position, with edges indicating ties from the original network.

Table 3. Correlation between latent space centrality measures and node degree

Model	$\mathbb{R}^1$	$\mathbb{R}^2$	$\mathbb{R}^3$	$\mathbb{S}^1$	$\mathbb{S}^2$
Probability connection	0.434	0.539	0.751	0.318	0.754
Mean distance	−0.400	−0.483	−0.698	−0.258	−0.658
Center distance	−0.432	−0.472	−0.692	−0.031	−0.817

highest for $\mathbb{R}^3$ (3.157), followed by $\mathbb{R}^2$ (0.987) and $\mathbb{R}^1$ (0.429), with spherical models yielding the tightest configurations $\mathbb{S}^1$ (0.393) and $\mathbb{S}^2$ (0.154).

In the inferred latent spaces, certain families consistently occupy central positions, most notably the Medici, who appear close to many others, indicating high latent connectivity. This agrees with historical accounts (Padgett and Ansell, 1993) highlighting their strategic marriage alliances and political integration (node 1). In contrast, families such as the Strozzi (node 4) and Albizzi (node 8) appear more peripheral, reflecting fewer matrimonial ties or deliberate exclusionary practices. In latent space models, proximity directly influences tie probability, so central positions correspond to latent social capital, revealing each family's embeddedness in the marriage economy of Renaissance Florence. These patterns motivate examining node centrality and assessing the model's ability to recover historically documented alliances and clustering structures.

To quantify network centrality, we use degree and closeness from classical network theory (Kolaczyk and Csárdi, 2020), along with three latent-space-based measures (Sosa and Betancourt, 2022). The probability connection of node i is the expected average probability of connection to other nodes, $E[\sum_{j \neq i} p_{ij} | \mathbf{Y}]$, where higher values indicate denser latent regions. The mean distance of node i is the expected average Euclidean or geodesic distance to other nodes, $E[\sum_{j \neq i} h(\mathbf{z}_i, \mathbf{z}_j) | \mathbf{Y}]$, with $h(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$ in $\mathbb{R}^d$ and $h(\mathbf{u}, \mathbf{v}) = \arccos(\mathbf{u}^\top \mathbf{v})$ in $\mathbb{S}^{d-1}$. The center distance of node i measures the expected distance to the latent space centroid $\bar{\mathbf{z}}$, $E[h(\mathbf{z}_i, \bar{\mathbf{z}}) | \mathbf{Y}]$, where $\bar{\mathbf{z}} = \arg \min_{\mathbf{z}} \sum_j h(\mathbf{z}_j, \mathbf{z})^2$. These latent space-based measures offer a probabilistic and geometric view of centrality that complements traditional metrics.

Table 3 shows that Euclidean models yield higher correlations with node degree, indicating stronger alignment between spatial proximity and structural centrality. Across geometries, nodes with high degree generally exhibit higher connection probability and lower center distance in Euclidean spaces, reinforcing their spatial coherence. Spherical models display greater variability and weaker alignment, likely due to geometric constraints. Certain nodes (e.g., 1 and 7) remain central across all metrics, underscoring their consistent prominence. These patterns are evident in the latent space visualizations (Figure 5), where central nodes cluster near core regions.

Overall, WAIC favors the three dimensional Euclidean model $\mathbb{R}^3$, with $\mathbb{R}^2$ also performing well, while $\mathbb{S}^2$ attains intermediate values that are close to those of the best Euclidean specifications. This indicates that increasing the Euclidean dimension improves in sample fit, but that the spherical model in $\mathbb{S}^2$ achieves a comparable average log likelihood under a more constrained geometry. In contrast, the one dimensional spherical model $\mathbb{S}^1$ shows poorer WAIC, reflecting the limited flexibility of a circle for this network. Taken together, these results suggest that the spherical models are competitive in terms of overall fit, with $\mathbb{S}^2$ in particular striking a balance between parsimony, geometric interpretability, and likelihood based performance. These conclusions are based on benchmark social science networks and should be interpreted as illustrative of the potential advantages of bounded latent geometries rather than as universal dominance over Euclidean formulations.

Once central nodes are identified, we detect communities by clustering each latent space. Point estimates of latent positions are obtained and used in spectral clustering (Ng *et al.*, 2002) with geometry-consistent affinity matrices. The number of clusters is chosen via the silhouette score,

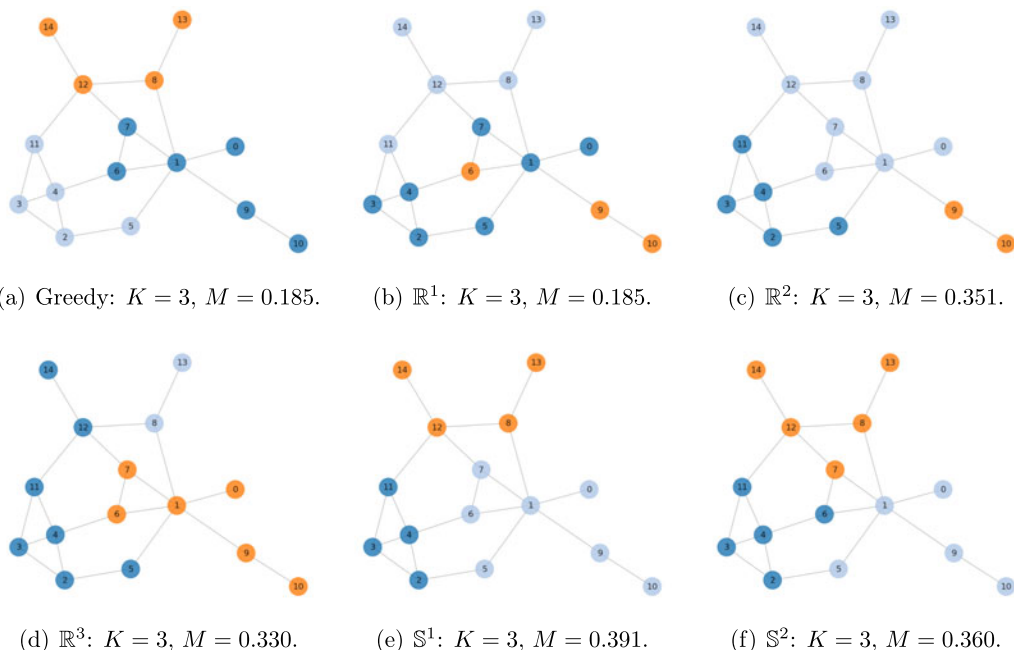


Figure 6. Clustering across latent space representations, showing number of communities K and modularity M .

and the partition with maximum modularity is taken as the representative structure. Figure 6 shows the detected communities, their counts, and modularity scores. The $\mathbb{S}^1$ model achieves the highest modularity, indicating the best capture of the network's community structure. This is consistent with the directional organization induced by the spherical geometry on $\mathbb{S}^1$, which tends to place densely connected families in compact angular sectors while maintaining separation between factions.

Taking together the results in Table 3 and the modularity scores in Figure 6, we see that the model in $\mathbb{S}^2$ essentially matches the model in $\mathbb{R}^3$ in terms of the correlation between degree and connection based centrality, and it provides the strongest alignment between radial position and degree through the center distance metric. By contrast, the model in $\mathbb{S}^1$ exhibits weaker correlations with degree, which is consistent with its more rigid one dimensional structure. Combining these results with the modularity scores, we see that $\mathbb{S}^1$ offers a very sharp community partition along the circle, while $\mathbb{S}^2$ yields slightly lower but still competitive modularity together with a latent geometry that supports both clustering and prediction.

Model adequacy was evaluated through posterior predictive checks using key network-level statistics: degree distribution, geodesic distances, shared edges, and triad census (Figure 7). In this network, latent space models primarily capture degree and geodesic distance patterns, as evidenced by the posterior predictive p -values and the inclusion of observed statistics within the posterior predictive intervals (Gelman *et al.*, 2014). For statistics requiring connectedness, computations use the largest connected component of simulated networks drawn from the posterior. Modularity is computed for both observed and simulated networks using the Clauset–Newman–Moore greedy maximization algorithm (Clauset *et al.*, 2004).

Finally, predictive performance was assessed using out of sample AUC, accuracy and F1 score (Figure 8 and Table 4). The model in $\mathbb{S}^2$ attains the highest values across all three metrics, with clear gains over the Euclidean models in $\mathbb{R}^2$ and $\mathbb{R}^3$, and even the simpler model in $\mathbb{S}^1$ improves on the lower dimensional Euclidean specifications. Higher AUC and F1 score indicate better

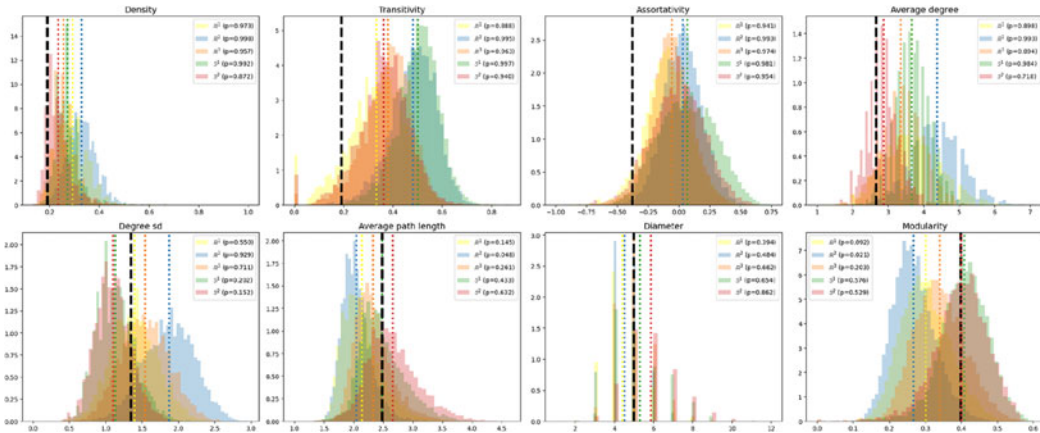


Figure 7. Posterior predictive checks for all models using network-level statistics in the Florentine families example. Each plot shows model-based histograms with means (dashed line) and the observed value (black line), with the legend reporting the posterior predictive p -value for each model.

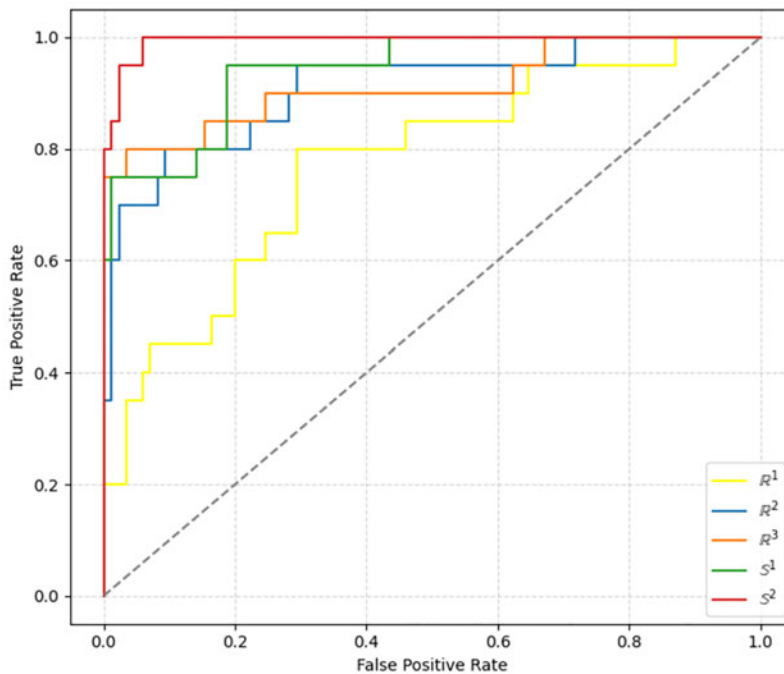


Figure 8. Receiver operating characteristic (ROC) curves for each model.

discrimination between ties and non ties and a more balanced recovery of present and absent edges. These patterns show that spherical latent spaces, particularly $\mathbb{S}^2$, can exploit curvature and angular separation, together with bounded inner products that stabilize tie probabilities, to represent transitivity, clustering and overlapping directional communities more faithfully than Euclidean alternatives, and that these geometric advantages translate into tangible gains in predictive performance.

Table 4. Predictive performance metrics for each model

Model	$\mathbb{R}^1$	$\mathbb{R}^2$	$\mathbb{R}^2$	$\mathbb{S}^1$	$\mathbb{S}^2$
AUC	0.773	0.910	0.913	0.941	0.994
Accuracy	0.780	0.819	0.933	0.923	0.952
F1-score	0.439	0.627	0.820	0.789	0.857

5. Discussion

This paper introduces a spherical latent space model for social network analysis, extending Hoff *et al.* (2002) by embedding nodes on a hypersphere instead of in Euclidean space. This geometry naturally captures transitivity, directional clustering, and cyclic structures while avoiding degeneracies such as unbounded distances. The approach reflects a broader trend in statistics and machine learning toward manifold-based modeling, that provide intrinsic structures for representing hierarchies, cycles, symmetries, and other topological constraints difficult to model in Euclidean settings. Within this framework, latent space models on manifolds offer principled representations of network geometry, enabling faithful structural characterization and interpretable visualizations. Our empirical results show that these geometric properties translate into higher modularity for $\mathbb{S}^1$ and improved AUC, accuracy, and F1 scores for $\mathbb{S}^2$ when compared with Euclidean counterparts, indicating that spherical latent spaces can provide competitive predictive performance and community detection results relative to Euclidean specifications in this empirical setting. These methodological developments and empirical findings constitute the central contributions of this work, namely the formulation and Bayesian implementation of spherical latent space models and their detailed comparison with Euclidean specifications on the Florentine marriage network.

Several recent contributions have incorporated circular or spherical geometry in network related models, including one dimensional circular visualizations in Hoff *et al.* (2002), spherical latent structure for aggregated relational data in McCormick and Zheng (2015), generative models on the sphere in Smith *et al.* (2019) and Papamichalis *et al.* (2022), and methods for detecting the curvature of the underlying latent geometry in Lubold *et al.* (2023). Our approach is complementary. We assume spherical geometry from the outset and study its implications in a latent distance model for individual level social network data, with an emphasis on Bayesian inference via Hamiltonian based MCMC and on an empirical comparison with Euclidean latent spaces in the Florentine marriage network.

Spherical latent spaces naturally encode relationships with topological constraints such as cyclicity or periodicity, relevant to settings from social networks to recurring spatial processes. By mapping unobserved relationships to positions on the manifold, they support probabilistic inference and interpretable visualizations of complex structures. However, manifold embeddings constrain the parameter space, may impose unnecessary geometric restrictions, and increase computational cost, especially in Bayesian settings with Hamiltonian or geodesic dynamics. In these contexts, directional statistics are essential, redefining concepts like mean, variance, and independence in terms of geodesic distances and angular metrics. Tools such as the von Mises–Fisher distribution, geodesic means, and angular concentration measures enable coherent inference in these nonlinear domains. Empirical results suggest that spherical models can achieve competitive performance relative to Euclidean alternatives in prediction and community detection for the networks studied here.

Future work includes adopting variational inference as a scalable alternative to Monte Carlo methods, lowering computational cost while preserving the model’s geometric structure. Another direction is refining MCMC estimation through better tuning of step sizes, proposals, and acceptance criteria, along with adaptive algorithms and parallel chain diagnostics, enhancements that

could improve convergence, posterior exploration, and parameter reliability, particularly in low-dimensional or geometrically constrained settings. Methodological extensions include exploring latent structures beyond the symmetric inner product, such as bilinear forms, asymmetric kernels, or other geometric operators, enabling richer relational patterns and the modeling of directed networks by relaxing the symmetry assumption. The framework can also be extended to multilayer and dynamic networks by modeling latent trajectories that evolve over time or span multiple interaction layers, with the geometry adapting to reflect structural changes and contextual information.

From a broader perspective, extending the geometric stock to hyperbolic spaces or other embedded manifolds would allow model geometry to be tailored to specific domains. Such developments would enable systematic study of how geometric assumptions affect fit, predictive accuracy, and interpretability, advancing manifold-based latent space modeling for complex networks.

Acknowledgments. None.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/nws.2026.10032>.

Data availability statement. All data are publicly available online.

Funding statement. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Statements and declarations. During the preparation of this work, the authors used ChatGPT-5.2 in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- Al-Taie, M. Z., & Kadry, S. (2017). *Python for graph and network analysis*. Springer.
- Albert, J., (2009). *Bayesian computation with R*. Springer.
- Aldous, D. J. (1985). Exchangeability and related topics. In Hennequin, P. L., ed. *École d'Été de Probabilités de Saint-Flour XIII-1983* (pp. 1–198). Springer Berlin Heidelberg.
- Aliverti, E., & Russo, M. (2022). Stratified stochastic variational inference for high-dimensional network factor model. *Journal of Computational and Graphical Statistics*, 31(2), 502–511.
- Austin, A., Linkletter, C., & Wu, Z. (2013). Covariate-defined latent space random effects model. *Social Networks*, 35(3), 338–346.
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877.
- Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer.
- Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Brugere, I., Gallagher, B., & Berger-Wolf, T. Y. (2018). Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys (CSUR)*, 51(2), 1–39.
- Byrne, S., & Girolami, M. (2013). Geodesic Monte Carlo on embedded manifolds. In *Scandinavian Journal of Statistics*, 40(4), 825–845.
- Caimo, A., & Gollini, I. (2016). Bayesian computational algorithms for social network analysis. *Computational Network Analysis with R: Applications in Biology, Medicine, and Chemistry*, 7, 63–82.
- Casella, G., & Berger, R. (2024). *Statistical inference*. Chapman and Hall/CRC.
- Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.
- Drobyshevskiy, M., & Turdakov, D. (2019). Random graph modeling: A survey of the concepts. *ACM Computing Surveys (CSUR)*, 52(6), 1–36.

- Durante, D., & Dunson, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika*, 101(4), 883–898.
- D'Angelo, S., Alfö, M., & Fop, M. (2023). Model-based clustering for multidimensional social networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(3), 481–507.
- D'Angelo, S., Murphy, T. B., & Alfö, M. (2019). Latent space modelling of multidimensional networks with application to the exchange of votes in Eurovision song contest. *The Annals of Applied Statistics*, 13(2), 900–930.
- Federer, H. (2014). *Geometric measure theory*. Springer.
- Fosdick, B. K., & Hoff, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association*, 110(511), 1047–1056.
- Friel, N., Rastelli, R., Wyse, J., & Raftery, A. E. (2016). Interlocking directorates in Irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, 113(24), 6629–6634.
- Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A., & Rubin, D. B. (2013a). *Bayesian Data Analysis* (3rd ed.). Chapman & Hall/CRC.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016.
- Girolami, M., & Calderhead, B. (2011). Riemann Manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 73(2), 123–214.
- Goldenberg, A., Zheng, A. X., Fienberg, S. E., Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2), 129–233.
- Gollini, I., & Murphy, T. B. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1), 246–265.
- Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. JHU press.
- Guhaniyogi, R., & Rodriguez, A. (2020). Joint modeling of longitudinal relational data and exogenous variables. *Bayesian Analysis*, 15(2), 477–503.
- Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 170(2), 301–354.
- He, Y., & Hoff, P. D. (2019). Multiplicative coevolution regression models for longitudinal networks and nodal attributes. *Social Networks*, 57, 54–62.
- Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469), 286–295.
- Hoff, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems*, 20, 657–664.
- Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory*, 15(4), 261–272.
- Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1), 530–543.
- Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics*, 9(3), 1169.
- Hoff, P. D. (2021). Additive and multiplicative effects network models. *Statistical Science*, 36(1), 34–50.
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098.
- Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1), 1303–1347.
- Hoover, D. N. (1982). *Some comments on exchangeability*. Princeton, NJ: Institute for Advanced Study. Unpublished manuscript.
- Hunter, D. R., Krivitsky, P. N., & Schweinberger, M. (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21(4), 856–882.
- Kaur, H., Rastelli, R., Friel, N., & Raftery, A. (2023). Latent position network models. In *Sage Handbook of Social Network Analysis* (pp. 526–541). SAGE Publications.
- Kim, B., Lee, K. H., Xue, L., & Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12, 105.
- Kivelä, M., Arenas, A., Barthélemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203–271.
- Kolaczyk, E. D., & Csárdi, G. (2020). *Statistical analysis of network data with R*. Use R! Cham: Springer International Publishing.
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(3), 204–213.

- Lubold, S., Chandrasekhar, A. G., & McCormick, T. H. (2023). Identifying the latent space geometry of network models through analysis of curvature. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2), 240–292.
- Ma, Z., Ma, Z., & Yuan, H. (2020). Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, 21(4), 1–67.
- Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons.
- Matias, C., Robin, S., Abergel, F., Aiguier, M., Challet, D., Cournède, P.-H., Faÿ, G., & Lafitte, P. (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. *ESAIM: Proceedings and Surveys*, 47, 55–74.
- McCormick, T. H., & Zheng, T. (2015). Latent surface models for networks using aggregated relational data. *Journal of the American Statistical Association*, 110(512), 1684–1695.
- Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A first course in network science*. Cambridge University Press.
- Minhas, S., Hoff, P. D., & Ward, M. D. (2019). Inferential approaches for network analysis: AMEN for latent factor models. *Political Analysis*, 27(2), 208–222.
- Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In: Brooks, S., Gelman, A., Jones, G., & Meng, X.-L., ed. *Handbook of Markov Chain Monte Carlo* (pp. 113–162). Chapman and Hall/CRC.
- Newman, M. (2018). *Networks*. Oxford University Press.
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In: *Advances in Neural Information Processing Systems*, Vol. 14 (pp. 849–856). MIT Press.
- Ng, T. L. J., & Murphy, T. B. (2022). Model-based clustering for random hypergraphs. *Advances in Data Analysis and Classification*, 16(3), 691–723.
- Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). Springer.
- Nowicki, K., & Snijders, T. A. (2001). A latent space model for random graphs with latent block structure. *Journal of the American Statistical Association*, 96(455), 1077–1087.
- Padgett, J. F., & Ansell, C. K. (1993). Robust action and the rise of the Medici, 1400–1434. *American Journal of Sociology*, 98(6), 1259–1319.
- Papamichalis, M., Turnbull, K., Lunagomez, S., & Airolidi, E. (2022). Latent space network modelling with hyperbolic and spherical geometries. arXiv preprint arXiv: [2109.03343](https://arxiv.org/abs/2109.03343).
- Pósfai, M., & Barabási, A.-L. (2016). *Network science*, Vol. 3. Cambridge, UK: Cambridge University Press.
- Raftery, A. E., Niu, X., Hoff, P. D., & Yeung, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of computational and graphical statistics*, 21(4), 901–919.
- Rastelli, R., & Friel, N. (2018). Optimal Bayesian estimators for latent variable cluster models. *Statistics and Computing*, 28(6), 1169–1186.
- Rastelli, R., Maire, F., & Friel, N. (2024). Computationally efficient inference for latent position network models. *Electronic Journal of Statistics*, 18(1), 2531–2570.
- Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (P*) models for social networks. *Social Networks*, 29(2), 173–191.
- Rosenbrock, H. H. (1960). An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3), 175–184.
- Ryan, C., Wyse, J., & Friel, N. (2017). Bayesian model selection for the latent position cluster model for social networks. *Network Science*, 5(1), 70–91.
- Salter-Townshend, M., & McCormick, T. H. (2017). Latent space models for multiview network data. *The Annals of Applied Statistics*, 11(3), 1217.
- Salter-Townshend, M., & Murphy, T. B. (2013). Variational Bayesian inference for the Latent Position Cluster Model for network data. *Computational Statistics & Data Analysis*, 57(1), 661–671.
- Salter-Townshend, M., White, A., Gollini, I., & Murphy, T. B. (2012). Review of statistical network analysis: models, algorithms, and software. *Statistical Analysis and Data Mining*, 5(4), 243–264.
- Sánchez, S., Sosa, J., & Luque, C. (2026). Influence networks: Bayesian modeling and diffusion. *Journal of Applied Statistics*, 53(3), 462–483.
- Sanz-Serna, J.-M., & Calvo, M.-P. (2018). *Numerical hamiltonian problems*, Vol. 7. Courier Dover Publications.
- Sewell, D. K., & Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512), 1646–1657.
- Sewell, D. K., & Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44, 105–116.
- Sewell, D. K., & Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, 12(2), 351–377.
- Smith, A. L., Asta, D. M., & Calder, C. A. (2019). The geometry of continuous latent space models for network data. *Statistical Science: A review journal of the Institute of Mathematical Statistics*, 34(3), 428.
- Sosa, J., & Buitrago, L. (2021). A review of latent space models for social networks. *Revista Colombiana de Estadística*, 44(1), 171–200.

- Sosa, J., & Rodríguez, A. (2022). A Bayesian approach for de-duplication in the presence of relational data. *Journal of Applied Statistics*, 51(2), 197–215.
- Sosa, J., & Rodríguez, A. (2021). A latent space model for cognitive social structures data. *Social Networks*, 65, 85–97.
- Sosa, J., & Rodríguez, A. (2023). A Bayesian record linkage model incorporating relational data. *Applied Stochastic Models in Business and Industry*, 39(6), 755–771.
- Sosa, J. C., & Betancourt, B. (2022). A latent space model for multilayer network data. *Computational Statistics & Data Analysis*, 169, 107432.
- Spencer, N. A., Junker, B. W., & Sweet, T. M. (2022). Faster MCMC for Gaussian latent position network models. *Network Science*, 10(1), 20–45.
- Wang, S., Paul, S., & De Boeck, P. (2023). Joint latent space model for social networks with multivariate attributes. *Psychometrika*, 88(4), 1197–1227.
- Watanabe, S. (2013). A widely applicable Bayesian information criterion. In: *Mathematical Monographs on the Statistical and Probabilistic Sciences*. Springer.
- Wood, A. T. (1994). Simulation of the von mises fisher distribution. *Communications in Statistics - Simulation and Computation*, 23(1), 157–164.
- Xu, R. (2018). Alternative estimation methods for identifying contagion effects in dynamic social networks: a latent-space adjusted approach. *Social Networks*, 54, 101–117.
- Yu, X. (2020). Spherical latent factor model for binary and ordinal data. PhD thesis, University of California, Santa Cruz.
- Zhang, X., Xue, S., & Zhu, J. (2020). A flexible latent space model for multilayer networks. In: *International Conference on Machine Learning* (pp. 11288–11297). PMLR.

Appendix A. Illustration: Rosenbrock function

In order to illustrate the sampling algorithms given above, we consider sampling from a target distribution of the form $f(\mathbf{x}) \propto \exp(-R(\mathbf{x}))$, where $R(\mathbf{x})$ is the Rosenbrock function (Rosenbrock, 1960) given by:

$$R(x_1, x_2) = (a - x_1)^2 + b(x_2 - x_1^2)^2, \quad \mathbf{x} \in \mathbb{R}^2 \text{ or } \mathbb{S}^1,$$

with parameters $a \in \mathbb{R}$ and $b \in \mathbb{R}$, with $b > 0$. The purpose of this example is not to replicate the latent space model itself, but to provide a controlled benchmark that isolates the geometric challenges faced by the sampling algorithms. In latent space models, and particularly in the spherical formulation considered in this paper, the posterior distribution of latent positions can exhibit strong curvature, anisotropy, and narrow high-density regions due to dependence induced by the network likelihood and geometric constraints. The Rosenbrock function is a standard test case with these characteristics, featuring a curved, narrow valley that is difficult to explore with random-walk proposals. As such, it provides a useful proxy for understanding how Metropolis–Hastings, Hamiltonian Monte Carlo, and GHMC behave in posterior landscapes with similar geometric features to those arising in latent space models.

Thus, we apply the Metropolis–Hastings algorithm (Algorithm 2), Hamiltonian Monte Carlo (Algorithm 3), and Geodesic Hamiltonian Monte Carlo (Algorithm 4). In the experiments, the Rosenbrock parameters are set to $a = 1.0$ and $b = 5.0$. For the Euclidean case $\mathbb{R}^2$, we use initial state $\mathbf{x}^{(0)} = (0.0, 0.0)$, 5000 samples, step size 0.05, burn-in 50000, thinning 100, and 2 chains. For the circular manifold $\mathbb{S}^1$, we use initial state $\mathbf{x}^{(0)} = (1.0, 0.0)$ with the same sampling settings. These configurations ensured adequate exploration of the target distributions while keeping computational cost reasonable (full implementation details are available at https://github.com/cnosa/Latentspacemodels_Networks).

The sampling results for the Rosenbrock distribution are presented in Figure 9 and Table 5. Panels (a) and (b) show the contours of the Rosenbrock function and the sample distributions in $\mathbb{R}^2$ obtained with the MH and HMC algorithms, respectively. HMC more effectively explores the narrow curved valley of the Rosenbrock function, yielding samples with lower variance than MH, which exhibits broader dispersion. Panels (c) and (d) display the Rosenbrock function evaluated on the circle and the corresponding samples under the constraint $\mathbb{S}^1$. Here, MH produces a more

Table 5. Comparison of Metropolis–Hastings (MH) and Hamiltonian Monte Carlo (HMC) algorithms for sampling from the Rosenbrock distribution on $\mathbb{R}^2$ and $\mathbb{S}^1$

Model Sampling algorithm	$\mathbb{R}^2$		$\mathbb{S}^1$	
	MH	HMC	MH	GHMC
Execution time (seconds)	40.2300	903.9240	98.2749	27363.6318
Maximum $\hat{R}$	1.0001	1.0003	1.0001	1.0090
Minimum relative ESS	0.1931	0.9999	0.9425	0.9482
Acceptance rate	0.3634	0.9340	0.1250	0.6819
Step size after tuning	0.4292	0.1567	0.0001	0.0004

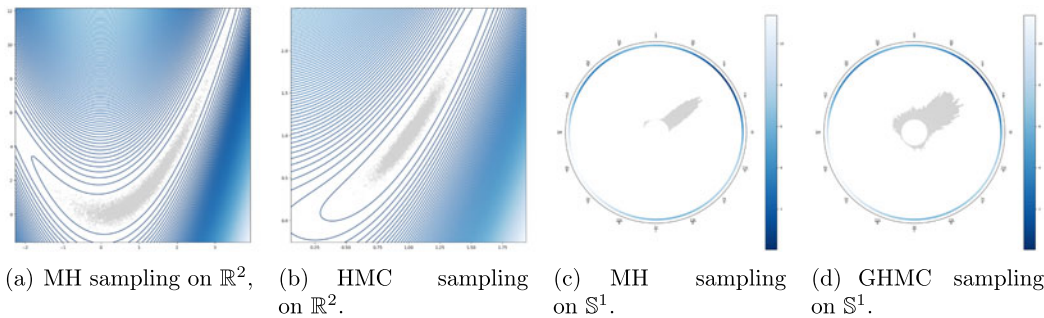


Figure 9. Comparison of sampling results for the rosenbrock distribution across different spaces and algorithms.

limited exploration than Geodesic HMC, reflected in lower sample variance; in both cases, the estimates reveal a bimodal distribution along the circumference.

Table 5 reports quantitative performance metrics, including average execution time, maximum $\hat{R}$, minimum relative ESS, average acceptance rate, mean value of the Rosenbrock function, and the tuned step size. HMC consistently achieves higher acceptance rates and larger ESS, indicating better convergence and more efficient exploration of the target distribution. In contrast, MH shows lower acceptance rates and reduced ESS, reflecting slower mixing, but with substantially lower computational cost, as evidenced by shorter execution times.

Differences between the $\mathbb{R}^2$ and $\mathbb{S}^1$ columns arise because these correspond to distinct target distributions, one defined on the full Euclidean plane and one restricted to the circle. Within each space, discrepancies across algorithms reflect differences in how effectively they explore the high density region. For example, in $\mathbb{R}^2$ the HMC sampler is able to traverse the narrow curved valley of the Rosenbrock function and spends more time near its minimum, producing a substantially lower average value of $R(\cdot, \cdot)$ than random walk Metropolis, which mixes poorly and remains in higher valued regions. This example is included as a controlled benchmark to illustrate the relative efficiency and robustness of MH, HMC, and GHMC in exploring curved and anisotropic target distributions. These geometric features are representative of posterior distributions arising in latent space models, particularly under spherical constraints.

Appendix B. Model extensions

Under the Bayesian paradigm, the spherical latent space formulation admits several natural extensions that enhance modeling flexibility, allow richer structural assumptions, and facilitate scalability in more complex settings. Below, we outline representative extensions together with their mathematical specifications.

Mixed effects model. Let $\mathbf{x}_{i,j} \in \mathbb{R}^p$ denote covariates associated with the actor pair (i, j) . The linear predictor can be extended to $\eta_{i,j} = \alpha + \boldsymbol{\gamma}^\top \mathbf{x}_{i,j} + \beta \langle \mathbf{z}_i, \mathbf{z}_j \rangle$, where $\boldsymbol{\gamma} \in \mathbb{R}^p$ are regression coefficients. Priors for $\boldsymbol{\gamma}$ include Gaussian for regularization, Laplace for shrinkage, or horseshoe for sparse, high-dimensional settings. This specification enables joint inference on covariate effects and latent geometry.

Weighted inner product model. Analogous to the eigenmodel, the linear predictor can be generalized to $\eta_{i,j} = \alpha + \beta \mathbf{z}_i^\top \boldsymbol{\Lambda} \mathbf{z}_j$, with $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{d-1})$, assigns dimension-specific weights. This allows certain hypersphere dimensions to influence tie probabilities more strongly, with Normal or hierarchical shrinkage priors on the diagonal elements of $\boldsymbol{\Lambda}$ controlling effective rank and mitigating overparameterization.

Finite mixture prior. To capture latent community structure, we assign a finite mixture prior to the latent positions $\mathbf{z}_i \mid \xi_i, \boldsymbol{\mu}_{\xi_i}, \kappa_{\xi_i} \stackrel{\text{ind}}{\sim} \text{VMF}(\boldsymbol{\mu}_{\xi_i}, \kappa_{\xi_i})$, with $\xi_i \stackrel{\text{iid}}{\sim} \text{Cat}(\boldsymbol{\pi})$, and $\boldsymbol{\pi} \sim \text{Dir}(\boldsymbol{\alpha})$ controlling the cluster weights. This specification induces clustering in the spherical latent space, producing interpretable block structures while preserving the manifold's geometric constraints.

Dirichlet process prior. A nonparametric alternative replaces the finite mixture with a Dirichlet process (DP) prior, $\mathbf{z}_i \mid G \stackrel{\text{iid}}{\sim} G$, with $G \sim \text{DP}(\alpha_0, G_0)$, where α_0 is a concentration parameter and G_0 is a suitable base measure on $\mathbb{S}^{d-1}$, such as a $\text{VMF}(\boldsymbol{\mu}_0, \kappa_0)$ distribution. The DP formulation enables the number of latent clusters to be inferred directly from the data, preserving the geometric constraints of the spherical latent space.

Multilayer networks. For multiplex or multilayer relational data $\mathbf{Y}_1, \dots, \mathbf{Y}_L$, the linear predictor in layer ℓ can be expressed as $\eta_{i,j,\ell} = \alpha_\ell + \beta_\ell \langle \mathbf{z}_i, \mathbf{z}_j \rangle$, with layer-specific intercepts α_ℓ and similarity coefficients β_ℓ assigned hierarchical Normal priors to enable information sharing across layers, thus capturing both layer-specific tie formation mechanisms and cross-layer dependencies.

Dynamic networks. For temporal networks $\mathbf{Y}_1, \dots, \mathbf{Y}_T$, latent positions $\mathbf{z}_{i,t}$ can be assigned a first-order Markov process on the sphere via $\mathbf{z}_{i,t} \mid \mathbf{z}_{i,t-1} \sim \text{VMF}(\mathbf{z}_{i,t-1}, \kappa_z)$, while layer-specific intercepts and similarity coefficients evolve as $\alpha_t \mid \alpha_{t-1} \sim \text{N}(\alpha_{t-1}, \sigma_\alpha^2)$ and $\beta_t \mid \beta_{t-1} \sim \text{N}(\beta_{t-1}, \sigma_\beta^2)$, introducing Markovian dependence that allows the model to capture smooth temporal variation in both latent geometry and tie formation tendencies.

Alternative computational strategies. While MCMC is the standard inference approach for latent space models, alternative methods can improve scalability. Variational inference (e.g., Blei *et al.*, 2017) and stochastic variational inference (e.g., Hoffman *et al.*, 2013) provide deterministic approximations to the posterior, with stochastic gradient optimization enabling application to large-scale networks.