Identifying the latent space geometry of network models through analysis of curvature

Shane Lubold[1], Arun G. Chandrasekhar[2,3,4] and Tyler H. McCormick[1,5,6]

[1]Department of Statistics, University of Washington, Seattle, WA, USA
[2]Department of Economics, Stanford University, Stanford, CA, USA
[3]NBER, Cambridge, MA, USA
[4]J-PAL, Cambridge, MA, USA
[5]Department of Sociology, University of Washington, Seattle, WA, USA
[6]Center for Statistics and the Social Sciences, University of Washington Seattle, WA, USA

Address for correspondence: Tyler H. McCormick, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322, USA. Email: tyler.h.mccormick@gmail.com

Abstract

A common approach to modelling networks assigns each node to a position on a low-dimensional manifold where distance is inversely proportional to connection likelihood. More positive manifold curvature encourages more and tighter communities; negative curvature induces repulsion. We consistently estimate manifold type, dimension, and curvature from simply connected, complete Riemannian manifolds of constant curvature. We represent the graph as a noisy distance matrix based on the ties between cliques, then develop hypothesis tests to determine whether the observed distances could plausibly be embedded isometrically in each of the candidate geometries. We apply our approach to datasets from economics and neuroscience.

Keywords: latent embedding, non-Euclidean geometry, social networks, statistical geometry

1 Introduction

Social, economic, biological, and technological networks play a crucial role in a myriad of environments. Job referrals (Beaman, 2012; Calvo-Armengol, 2004; Granovetter, 1973; Heath, 2018), neurological function (Leung et al., 2008), epidemics (Bansal et al., 2010; Hoff et al., 2002; Sewell & Chen, 2015), social media (Cho et al., 2016; Myers & Leskovec, 2014; Romero et al., 2011), informal insurance (Ambrus et al., 2014; Cai & Szeidl, 2017), education decisions (Calvó-Armengol et al., 2009), sexual health (Handcock & Jones, 2004), financial contagion (Acemoglu et al., 2015; Elliott et al., 2014; Gai & Kapadia, 2010), international trade (Chaney, 2014), and politics (DiPrete et al., 2011) are among the many settings in which networks play a major role. Modelling network formation is, therefore, essential for both descriptive and counterfactual analyses.

Constructing such models is challenging from a statistical perspective since networks typically feature higher-order dependence between the connections. Phenomena such as transitivity are common and mean that standard regression approaches, which assume independence across connections, are not appropriate. A common approach for modelling this dependence structure is the latent space model, introduced by Hoff et al. (2002). One estimates a probability distribution over graphs that is consistent with the single, observed graph. The model assigns each node in the network to a position on a low-dimensional manifold. Likelihood of a connection is inversely proportional to the distance between actors on a manifold with a pre-specified dimension and geometry. Connections are assumed independent conditional on the latent positions. Standard practice in this area assumes a manifold class beforehand.

The choice of the latent manifold is extremely consequential for both the interpretation and its theoretical properties. In enumerating these properties, it is first critical to distinguish between intrinsic and extrinsic geometries. Our focus is on the former, which embodies the fundamental properties of the geometric space. Modelling in a geometry manifold and inheriting the distance measure of that geometry (e.g., using a two-dimensional sphere means points exist on the surface of the sphere and distances are measured by arc length).

Moving now to the connection between geometry and networks, we first note that, holding constant the distribution of points in the latent space, the choice of the geometry in particular determines the nature of network structure captured by the latent space. For a simple example, consider a two-dimensional Euclidean space (a plane). Here it is not possible to place four nodes in such a way that they are equidistant from one another, meaning that it is not possible to represent groups of four such that, holding constant node effects, each node has the same likelihood of interacting with any other. Another way to see the impact of geometry is through triangles. Since a sphere has a bounded area, there is an upper bound to how far apart nodes can be from one another before they start getting closer together. Positive curvature also encourages the formation of triangles and communities. Additionally, certain networks, such as a network of neurons or a network exchange built along a supply chain, may have a tree-like structure. Trees are difficult to embed in spherical or Euclidean space but fit more naturally in hyperbolic space. Recent work on statistical modelling has also shown the importance of modelling networks using non-Euclidean latent representations. For instance, McCormick and Zheng (2015) model latent space as a sphere and Krioukov et al. (2010) and Asta and Shalizi (2015) use hyperbolic space. Yu and Rodriguez (2019) explore both spherical and Euclidean representations in a Bayesian model for spatial voting patterns. Weber and Nickel (2018) examine the relationship between the latent space curvature and graph motifs. Wilkins-Reeves and McCormick (2022) propose a test of the assumption that the latent space has constant curvature which uses the clique structure in the graph. Finally, Smith et al. (2019) provide a comprehensive review of the implications and consequences of the choice of geometry.

A second consequence of the choice of geometry arises in the theoretical properties of latent space model estimates. Consistency of the estimates of the individual locations on the unobserved manifold is the subject of recent work by Shalizi and Asta (2017). Since the distribution of the network formation process depends on the manifold itself, the key open question is whether a researcher can consistently estimate the latent space. After all, the network formation process is sensitive to the geometry inclusive of its curvature and dimension.

It is currently common practice to assume the latent dimension and manifold type. Our approach provides a data-driven alternative. It also contrasts with cross-validation-based selection procedures that are sometimes used, in particular, to estimate the dimension. These approaches subsample connections and then use either model fit diagnostics or out-of-sample prediction metrics. Our approach avoids a critical issue with these approaches, namely, subsampling can fundamentally alter graph properties in unpredictable ways (Chandrasekhar & Lewis, 2016), calling into question the relevance of the subsampled distribution. We approach the question from a fundamentally different perspective than currently available alternatives that use the likelihood or penalized likelihood to estimate model fit. First, rather than characterizing fit or predictive accuracy with a particular dataset, our approach takes a more classical hypothesis testing perspective. Comparing (for example) an information metric across a model fit with a spherical or hyperbolic latent space is fundamentally characterizing the congruence between the embedding for a given dataset and the spaces under consideration. Uncertainty in this framework arises from sampling, but also from potential model mis-specification. A likelihood-based metric for a spherical space with small curvature will likely perform quite well for a graph generated from Euclidean embeddings, for example. We isolate uncertainty to only sampling error by using a test for isometric embeddings of distances into the space under consideration. We conceptualize variability in the observed distance matrix as representing expected noise due to sampling realizations of a graph of a given size. Second, we isolate the test to uniquely distance rather than to the model as a whole, as would be the case with a likelihood-based measure. A likelihood ratio test for whether or not the curvature of the space is zero, for example, may seem to be an appealing alternative to our approach. Such a test would, however, confound changes in the latent geometry with changes in the fixed effects. To see this, recall that the surface area of the sphere changes as a function of

the curvature. To preserve the overall density of the graph, therefore, the individual effects must change when the curvature changes. We sidestep this issue by leveraging the structure of the network formation model to isolate the test as specific to the latent geometry. Constructing an appropriate likelihood-based test, in contrast, would require marginalizing over the individual effects, which would be computationally intensive and require specifying distributions for the individual effects (which we do not require).

We address the question of how to choose the manifold $\mathcal{M}^{p^\star}(\kappa^\star)$, meaning the manifold class $(\mathcal{M}^\star)$, curvature $(\kappa^\star)$, and dimension $(p^\star)$ of the latent space. We present a hypothesis testing framework that connects distances in the latent space to feasible embeddings on simply connected, complete Riemannian manifolds with constant curvature. Rather than relying on likelihood or cross-validation techniques, we directly leverage geometric results that provide necessary and sufficient conditions to embed points into particular manifolds, given pairwise distances between the points.

Our main insight is as follows. The manifolds we consider in this work all come equipped with a metric—an inner product which is used in the calculation of distances between points. The metric uniquely identifies the manifold. Given a distance matrix D between K points, even without knowing where the points are located, we can check if the points can be isometrically embedded in the manifold. It can be embedded if and only if the distance matrix is compatible with the candidate manifold's metric. Specifically, given D, we can construct a test matrix $W_{\kappa^\star}(D)$ and check that the eigenvalue spectrum has the same signature as the manifold's metric. The manifold detection problem is, therefore, reduced to testing the spectrum of $W_{\kappa^\star}(D)$.

We make two contributions. Our first contribution is to statistical geometry (Theorem 1.1). We show how to estimate the manifold consistently when observing a noisy estimate $\hat{D}$ of the true distances. For this, we must estimate the signature of the spectrum of $\hat{W}_{\hat{k}}(\hat{D})$, the empirical test matrix. The logic relies on Weyl's inequality, which places bounds on the change in eigenvalues due to perturbations of a matrix. This provides an avenue for consistent estimation of $\mathcal{M}^{p^\star}(\kappa^\star)$.

Our second contribution is to then extend the argument to the latent space network model. The main idea is that we take the observed graph G on n nodes and construct some distance matrix $\hat{D}(G)$ among $K << n$ points and then apply our statistical geometry result. We define distance based on interaction rates between K *groups* of nodes. Using the observed graph, we define the distance between two cliques based on the probability of an edge between a node in clique C_1 and a node in clique C_2, which we can calculate using the definition of the latent space model plus the fraction of realized links between cliques. We estimate the matrix of cross-clique edge probabilities and then use the latent space model to estimate the pairwise distances between cliques. We then leverage our statistical geometric result to estimate the manifold.

In the remainder of this section, we formally define the two problems we address and provide an overview of our approach. Specifically, we address (i) the general problem of estimating geometry from a noisy distance matrix and (ii) estimating network geometry from latent space models using cliques. Using cliques represents the most challenging case for our method since we expect that in many settings cliques may be relatively small. Next, in Section 2, we review key geometric concepts crucial for our testing procedure. Section 3 covers the general geometry problem in detail and develops our general method of estimating geometry from a noisy distance matrix. We turn to study the estimation of the latent space in general using cliques in Section 4. In Section 5, we present simulation experiments that explore the efficacy of our approach. We apply our results to two empirical examples in Section 6. The first empirical example considers data from 75 Indian village social networks, comprised of informal finance, information, and social links. We study the financial flows by geometry and also how the introduction of microfinance impacts geometry. The second example focuses on the neural network of the *C. Elegans* worm. Section 7 concludes. All proofs are in Appendix A unless otherwise noted.

1.1 Statistical geometry problem

The estimation methods we propose are quite general. They apply broadly to a large set of problems (Section 3.1 provides examples) in which the researcher observes a noisy distance matrix $\hat{D}$ and wishes to estimate the properties of the underlying space. We make the following assumption

about the latent space $\mathcal{M}^p(\kappa)$. After that, we provide a broadly applicable classification theorem of $\mathcal{M}^p(\kappa)$ from an estimator $\hat{D}$ of distances computed between points in $\mathcal{M}^p(\kappa)$.

Assumption 1.1 $\mathcal{M}^{p^\star}(\kappa^\star)$ is a simply connected, complete Riemannian manifold of constant sectional curvature $\kappa^\star$, with $p^\star \in \mathbb{Z}$ with known upper bound and $\kappa^\star \in [-b, -a] \cup \{0\} \cup [a, b]$ with $a > 0$, $b > 0$.

The technical geometric definition of simply connected, complete Riemannian manifolds is provided in a self-contained manner in Appendix H. By Killing (1891), Assumption 1.1 means that the manifold must be Euclidean, spherical, or hyperbolic with a bounded dimension and curvature value in some compact set. We emphasize that $a > 0$ means that in the curved cases, the geometry is not arbitrarily close to a Euclidean space ($\kappa = 0$). All such Riemannian manifolds are locally Euclidean, by definition, so this is required to be meaningful. We also emphasize that by dimension $p^\star$ we mean the minimum such dimension, as one can clearly embed $\mathcal{M}^{p^\star}$ in $\mathcal{M}^{p^\star + h}$ for $h > 0$.

Algorithm 1 takes in $\hat{D}$ and returns consistent estimates of the manifold type, curvature, and dimension. We use $\circ$ to denote the Hadamard product and 1_K denotes a vector of ones of length K.

Algorithm 1: Consistent estimation with a noisy distance matrix

---

Input: noisy $K \times K$ distance matrix $\hat{D}$.

1. Estimate the curvature of the manifold for each of the curved cases ($\hat{\kappa}_S$, $\hat{\kappa}_H$), from (7).
2. For each of the three candidate geometries, calculate the test matrix W_κ using the corresponding curvature estimate from step (1), with projection $J = I_K - 1_K 1_K^T / K$:

$$W_\kappa(\hat{D}) := \begin{cases} \frac{1}{\kappa}\cos(\sqrt{\kappa}\hat{D}) & \text{for } \kappa \neq 0 \\ -\frac{1}{2}J\hat{D} \circ \hat{D}J & \text{for } \kappa = 0. \end{cases}$$

3. Construct $\hat{\mathcal{M}}$, the estimate the manifold class, using Proposition 3.4.
4. Given $\hat{\mathcal{M}}$, select $\hat{p}$, the estimate of its dimension using Proposition 3.5.

Theorem 1.1 (Consistent estimation with a noisy distance matrix). Let $\mathcal{M}^{p^\star}(\kappa^\star)$ be a latent space of unknown manifold class, dimension, and curvature that satisfies Assumption 1.1. Fix a set of $K > p^\star$ locations on $\mathcal{M}^{p^\star}(\kappa^\star)$ and suppose they uniquely identify the latent space. Suppose there is a sequence of $K \times K$ matrices $\hat{D}_T$, indexed by $T \in \mathbb{N}$, such that $\hat{D}_T \xrightarrow{p} D$ as $T \to \infty$. Under these assumptions, the estimators produced by Algorithm 1 are consistent as $T \to \infty$. That is, $\mathbb{P}(\hat{\mathcal{M}}^{\hat{p}} \neq \mathcal{M}^{p^\star}) = o(1)$ and $\hat{\kappa} - \kappa^\star = o_P(1)$.

1.2 Latent space model

Having proposed a consistent classification method of the manifold class, dimension, and curvature from noisy distances (see Theorem 1.1), we now turn to the latent space model. Consider a graph $G = (V, E)$ where V are nodes and E are edges (also called links or connections), with $|V| = n$. For simplicity, we assume throughout that the graph is un-directed, all connections are symmetric, and unweighted, all connections are either present or absent. Our methods readily extend to the weighted and directed case, though it increases the complexity in terms of both notation and exposition. We assume that edges in G are drawn independently according to

$$\mathbb{P}\left\{G_{ij} = 1 \mid v^\star, z^\star, X_{ij}^\star, \mathcal{M}^{p^\star}(\kappa^\star)\right\} = \Lambda\left(v_i^\star + v_j^\star - d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i^\star, z_j^\star)\right), \tag{1}$$

for some increasing, invertible link function Λ. We represent $v = (v_1, \ldots, v_n)$ as the vector of individual effects, restricted to lie in some set to ensure (1) is a probability value in [0, 1].[1] These are

---

[1] One way to model a directed graph is to allow each node to have two different fixed effects, v_i and χ_i, one playing a role when i is the sender of the link (G_{ij}) and the other when i is the receiver (G_{ji}).
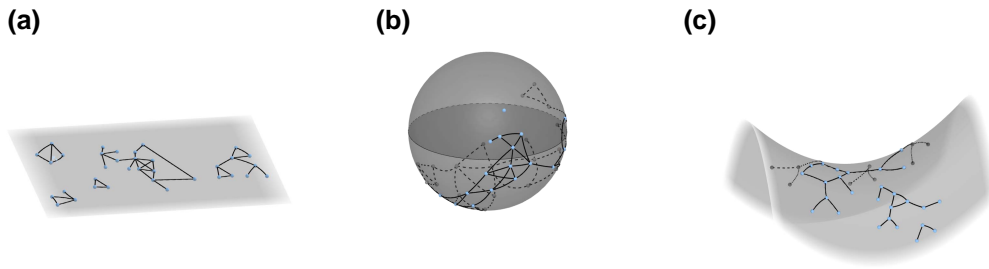
Figure 1. How curved geometries affect network embeddings where each displayed graph has 36 nodes. (a) Flat $\kappa^\star = 0$. (b) Positive $\kappa^\star > 0$. (c) Negative $\kappa^\star < 0$.

independent effects that encode individual gregariousness and are related to the total number of connections (Chatterjee et al., 2011; Graham, 2017). The $d_{\mathcal{M}^p(\kappa)}(z_i, z_j)$ terms represent the distance on the manifold $\mathcal{M}^p(\kappa)$, with dimension p and curvature κ, between locations z_i and z_j. Most of our analysis is done on the model above in (1) using an exponential link function, so $\Lambda(x) = \exp(x)$, but this is mostly out of convenience. In Appendix J, we show how to handle other common link functions (e.g., logistic link, (Hoff et al., 2002)) or how to handle node- and pair-level covariates effects (Graham, 2017).

We now discuss how Assumption 1.1 applies to our network problem. Simple connectedness and completeness are innocuous and constant curvature provides a place to start and nests all manifolds used in the literature, but rules out inhomogeneous latent spaces entirely such as those with 'structural holes' in the manifold, like the torus. Nevertheless, these three types of manifolds span a large and usable set of empirically relevant networks. With zero curvature, we model networks that allow for many paths where following them along nodes takes one increasingly far from nodes in other directions, while preserving local clustering. So while there is clustering, a flat space models a sort of vastness. Meanwhile, a sphere that has constant positive curvature does force such behaviour. Following friends of friends of friends and so forth typically leads to encountering some distant friends in common at a much higher rate. Therefore, there is a sort of cloistering in addition to clustering. Finally, hyperbolic spaces in contrast naturally embed trees or hierarchical networks or any context where expansiveness is a key feature. Intuitively this is because any set of initially parallel lines spread apart. Figure 1 presents intuitions. Smith et al. (2019) provide a comprehensive discussion on the relationship between network properties and the latent space.

Assumption 1.2 is a mild assumption that ensures that (1) produces probabilities.[2]

> **Assumption 1.2** Every node i has a fixed-effect $v_i^\star$ i.i.d. from a distribution F_v. The support of F_v is required to be such that (1) always returns values in [0, 1].

If the link function is exponential, Assumption 1.2 requires that support$(F_v) \subseteq (-\infty, 0]$, but for the logistic link function Assumption 1.2 is satisfied for any distribution. Recall that our goal is to apply Theorem 1.1 to identify the manifold properties from just one network. We need to identify a set of K locations on the manifold and a sequence of estimators $\hat{D}_T$ that satisfy the assumptions in Theorem 1.1.

Our approach, which we study in detail in Section 4, exploits the clique structure in the network. Because of (1), nodes in even modest-sized cliques (e.g., 5 nodes) are very likely to be close in the latent space. In other words, we can imagine that nodes in a clique are at the same location on the manifold. Finding K disjoint cliques in the network, therefore, gives us K distinct points in the latent space. By counting the number of edges between pairs of these K cliques, we can, therefore, estimate the probability that nodes on the latent space connect. Since (1) relates distances in the latent space and edge probabilities, we can, therefore, use the estimated probability of connections to estimate distances between these K points.[3] We use the notation $C(\ell)$ to denote a clique of size ℓ in a graph, and $C_1(\ell), \ldots, C_K(\ell)$ refers to a collection of K cliques each of size ℓ.

---

[2] Equivalently if one defined $\theta_i = \exp(v_i)$, then these fixed effects are simply multiplicative factors on the linking probability due to distances.

Algorithm 2: Estimating geometry of latent space network model using cliques

---

Input: graph G.

1. Construct $\hat{D} = \hat{D}(G)$.
   (a) Identify K disjoint ℓ-cliques $C_1(\ell), \ldots, C_K(\ell)$.
   (b) Estimate the cross-clique linking probability $\hat{P}_{kk'} = (1/\ell^2) \sum_{i \in C_k(\ell)} \sum_{j \in C_{k'}(\ell)} G_{ij}$.
   (c) Calculate $\hat{D}_{kk'} = -\log(\hat{P}_{kk'}) + \log(\hat{\gamma})$, where $\hat{\gamma}$ is an estimate of $E\{\exp(v)\}^2$.
2. Apply Algorithm 1 to $\hat{D}$ to construct estimator $\hat{\mathcal{M}}^{\hat{p}}(\hat{\kappa})$.

We require an assumption to ensure that observed cliques are likely to be comprised of nodes that are near each other in the latent space. Assumption 1.3 sets out a general requirement for F_z which makes explicit the condition that is required for our proofs: proportionally most of the cliques in the graph are comprised of nodes that are proximate in latent space. This assumption captures a typical feature of latent space models and empirical data.

Let G be a graph drawn from (1). For an arbitrary set of nodes $V_0 \subseteq V$, let G_{V_0} denote the subgraph induced by these nodes. If $|V_0| = \ell$, we use $\{G_{V_0} \in C(\ell)\}$ to denote the event that G_{V_0} is an ℓ-clique; that is, G_{V_0} is a complete graph on ℓ nodes.

> **Assumption 1.3** Every node i resides at a location $z_i^\star$ that is drawn independently and are identically distributed from a distribution F_z on manifold $\mathcal{M}^{p^\star}(\kappa^\star)$. The latent location distribution must satisfy two properties:
>
> (a) *Identifiability:* The support of F_z must consist of at least $K > p^\star$ distinct points that uniquely identify the manifold.
> (b) *Local cliques:* For any collection V_0 of ℓ nodes with locations drawn i.i.d. from F_z we have for all $\delta > 0$, $\mathbb{P}\{\max_{ij \in V_0} d(z_i, z_j) \leq \delta | G_{V_0} \in C(\ell)\} \to 1$, as $n, \ell \to \infty$.

Part (a) states that we need K to be larger than the true dimension and that we need there to be only one latent space in which we can embed these points isometrically (Section 2.1 contains definitions of these terms). Part (b) states that given a clique of size ℓ, the probability that nodes in this clique are close to each other goes to 1 as the clique size and graph size grow. Aside from this condition, we impose no restrictions on the distribution of F_z. This allows for continuous, discrete, or mixed distributions as well as dependence on n. In Section 4.4, we provide a discussion on these two conditions in the context of the network model and provide high-level conditions which imply Assumption 1.3. We prove that these high-level conditions hold in common cases, such as when the node locations are drawn from a lattice model, Gaussian mixture model, or uniformly over a bounded but expanding region.

Before continuing, we emphasize a few key points. First, Assumption 1.3(b) is written with $n, \ell \to \infty$. Let $\ell = \ell(n)$ depend on the graph size. In order for there to be cliques of size ℓ as $n \to \infty$, we need $\ell(n)$ to grow slowly, usually $\ell(n) \propto \log(n)$. Appendix K shows that cliques of size $\log(n)$ exist with high probability as n grows for many common location distributions. Second, the existence of cliques is guaranteed by the latent space model under our assumptions as the number of nodes increases. The conditional independence relation that is key to the latent space model requires an assumption of exchangeability.[4] The Aldous-Hoover Theorem implies that exchangeable sequences of nodes correspond to dense graphs in the limit (Aldous, 1981; Orbanz & Roy, 2015), which implies that cliques are present in the limit. We also examine the existence of cliques using our empirical and simulated examples. We find that the number and size of cliques in our empirical examples is sufficient to match settings in simulations where the method controls Type 1 error and has high power.

---

[3]   This makes clear that if the researcher observed weighted graph data, since G_{ij} is now a smooth function of distance and fixed effects, they can dispense with the clique approach altogether, since they directly observe a transformation of distances between specific points. The problem is easier.

[4]   Large graphs with exchangeable nodes are only dense if the graph is a subgraph of an infinitely exchangeable graph. If the distribution of the graph changes with n, the limiting graph need not be dense.

Theorem 1.2   (Estimating geometry via cliques). Let $\mathcal{M}^{p^\star}(\kappa^\star)$ be a latent space of unknown type, dimension, and curvature that satisfies Assumption 1.1. Consider a sequence of graphs on n nodes drawn from the distribution in (1), satisfying Assumptions 1.2–1.3. Under these assumptions, the estimators produced by Algorithm 2 are consistent as $n, \ell \to \infty$. That is, $\mathbb{P}(\hat{\mathcal{M}}^p \neq \mathcal{M}^{p^\star}) = o(1)$ and $\hat{\kappa} - \kappa^\star = o_P(1)$.

2 Overview of geometry and embedding conditions

In this section, we provide a brief overview of the three manifold types we consider in this work: Euclidean, spherical, and hyperbolic space. As noted above, these spaces span a class of empirically relevant manifolds on which much of the latent space network literature focuses (Killing, 1891). We then provide necessary and sufficient conditions on an arbitrary distance matrix that ensures the points from which the distances are computed can be embedded isometrically into one of these three geometries.

2.1 Candidate geometries

To study the candidate manifold $\mathcal{M}^p(\kappa)$, we embed them in $\mathbb{R}^{p+1}$. The Euclidean case is straightforward. In the spherical case, we embed it in Euclidean space ($\mathbb{R}^{p+1}$ with the usual metric) and in the hyperbolic case we use Minkowski space ($\mathbb{R}^{1,p}$). Note that the only difference is that the bilinear form of the space, denoted by Q below, varies in signature described below.

The model for each is constructed by looking at a locus of points in the ambient space in which it is embedded:[5]

$$\mathcal{M}^p(\kappa) := \left\{ x \in \mathbb{R}^{p+1} : Q(x, x) = \kappa^{-1} \right\}.$$

This implies a way of calculating distances between points on the manifold. Specifically,

$$d_{\mathbf{M}^p}(x, y) = \frac{\arccos\{\kappa Q(x, y)\}}{\sqrt{\kappa}}.$$

Let us turn to our candidate cases. The Euclidean space $\mathbb{R}^p$ is the p-dimensional Euclidean space with the usual Euclidean metric. In the case of the sphere $\mathbf{S}^p$, we have the usual Euclidean inner product $Q_{\mathbb{R}^{p+1}}(x, y) := \sum_{i=1}^{p+1} x_i y_i$. The locus of points and distances between two points $x, y \in \mathbb{R}^{p+1}$ for the embedding is

$$\mathbf{S}^p(\kappa) := \left\{ x \in \mathbb{R}^{p+1} : Q_{\mathbb{R}^{p+1}}(x, x) = \kappa^{-1}, \ \kappa > 0 \right\} \quad \text{and} \quad d_{\mathbf{S}^p}(x, y) = \frac{\arccos\{\kappa Q_{\mathbb{R}^{p+1}}(x, y)\}}{\sqrt{\kappa}}.$$

Hyperbolic space $\mathbf{H}^p$ is embedded in Minkowski space, $\mathbb{R}^{1,p}$ which is $\mathbb{R}^{p+1}$ equipped with the Minkowski bilinear form: $Q_{\mathbb{R}^{1,p}}(x, y) := -x_0 y_0 + \sum_{i=1}^p x_i y_i$. The important point is that the signature is distinguished from the Euclidean space, which will play a key role in distinguishing the geometries. The locus of points and distances are given by

$$\mathbf{H}^p(\kappa) := \left\{ x = (x_0, x_{1:p}) \in \mathbb{R}^{1,p} : Q_{\mathbb{R}^{1,p}}(x, x) = \kappa^{-1}, \ x_0 > 0, \ \kappa < 0 \right\}$$

and

$$d_{\mathbf{H}^p}(x, y) = \frac{\arccos\{\kappa Q_{\mathbb{R}^{1,p}}(x, y)\}}{\sqrt{\kappa}}.$$

---

[5]   For the hyperboloid $x_0 > 0$ is an additional restriction.

2.2 Isometric embedding conditions

Equipped with a notion of how distances are calculated between points in our candidate manifolds, we briefly review the conditions to check if a collection of K points can be isometrically embedded in each of the manifolds.

Let D be a known distance matrix from K points given by $Z = \{z_1, \ldots, z_K\}$. We say that Z can be *isometrically embedded* in manifold $\mathcal{M}^p(\kappa)$, written as $Z \xrightarrow{\text{isom}} \mathcal{M}$, if there exists an isometry ϕ such that for all $l, l' \in \{1, \ldots, K\}$, $d_{\mathcal{M}}(\phi(z_l), \phi(z_{l'})) = d_{ll'}$.

Given D, we define the $K \times K$ matrix $W_\kappa(D)$, which will allow us to determine whether an isometric embedding is possible in one of the three candidate geometries. To do this, we choose the matrix to correspond to the bilinear form $Q(\cdot, \cdot)$ above. For the Euclidean case we need a matrix $J := I_K - (1/K)1_K 1'_K$. Our test matrix is given by

$$W_\kappa(D) := \begin{cases} \frac{1}{\kappa}\cos(\sqrt{\kappa}D) & \text{for } \kappa \neq 0 \\ -\frac{1}{2}JD \circ DJ & \text{for } \kappa = 0, \end{cases} \tag{2}$$

where we apply the cosine operation element-wise, as before. We write $W_\kappa = W_\kappa(D)$, suppressing the dependency on D unless otherwise noted. By using a Taylor series of $W_\kappa(D)$ around $\kappa = 0$, one can see the relationship between the expression of $W_\kappa(D)$ for $\kappa > 0$ and $W_0(D)$.[6]

The following lemma characterizes the conditions for isometric embedding and is a concise restatement of classical results: Schoenberg (1935) Theorem 1 (which we include as part (1) of our Lemma 2.1) and Begelfor and Werman (2005) Theorem 1 (which we include as parts (2-3) of our Lemma 2.1). See Belton et al. (2019) for an overview of related topics in distance geometry. The *signature* of a square matrix A is a triple $\text{sig}(A) = (a, b, c)$, where a, b, and c are, respectively, the number of positive, zero, and negative eigenvalues of A. A positive semi-definite matrix has $c = 0$. Throughout the paper, we use the convention that $\lambda_{\max}(A) := \lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_K(A) =: \lambda_{\min}(A)$ are the eigenvalues of the $K \times K$ matrix A sorted in decreasing order.

> **Lemma 2.1** Let $p_{\min}$ be the minimum dimension for which $Z \xrightarrow{\text{isom}} \mathcal{M}^p(\kappa)$ and assume $a > 0$.
>
> 1. $Z \xrightarrow{\text{isom}} \mathbb{R}^p$ for some p if and only if $\text{sig}(W_0) = (a, K - a, 0)$. Further, $p_{\min} = a$.
> 2. $Z \xrightarrow{\text{isom}} S^p(\kappa)$ if and only if $\text{sig}(W_\kappa) = (a, K - a, 0)$. Further, $p_{\min} = a - 1$.
> 3. $Z \xrightarrow{\text{isom}} H^p(\kappa)$ if and only if $\text{sig}(W_\kappa) = (1, K - a - 1, a)$. Further, $p_{\min} = (K - a) - 1$.

The result above tells us, for example, that $\lambda_{\min}(W_0) \geq 0$ when D is computed from points in $\mathbb{R}^p$. This allows us to then phrase the problem of geometry identification, where we do not observe the manifold, into a problem about eigenvalues of W, which we do observe. This re-framing of the statistical geometry problem is the main insight behind our geometry classification procedure in Theorem 1.1. The lemma also allows us to estimate the dimension of the latent space through the rank of W_κ, which provides the basis of our dimension estimation procedure in Section 3.6.

3 Testing geometry from an arbitrary distance matrix estimate $\hat{D}$

This section addresses how to test whether a set of K points can be isometrically embedded into a candidate manifold out of the set we consider, given only a consistent estimator $\hat{D}$ of the pairwise distance matrix D between them. We develop the constituent pieces for our main result, Theorem 1.1.

This section, therefore, is not about networks exclusively but rather about a general statistical geometry problem that we outlined in Theorem 1.1. We show below that the general statistical geometry problem outlined in Section 1.1 is applicable in a wide range of relational data examples, extending the scope of our work beyond the binary adjacency matrix case we consider in Section 4. Each of the following examples will produce an estimator $\hat{D}$ which approximates some unknown matrix D, which contains pair-wise distances between K objects (people, nodes, firms, etc.). For

---

[6] We would like to thank Gabriel Caroll for pointing this out to us.

the asymptotic results presented below, index $\hat{D}$ by some T such that $\hat{D} = \hat{D}_T \xrightarrow{p} D$ as $T \to \infty$. Informally, T maybe be thought of as the sample size. In our main application to networks, T is a function of network size n. However, the general statistical geometry problem may generate an estimator of distances in other ways. For example, longitudinal data where T indexes time in an international trade example or number of samples in a neuroscience example. For convenience, we often drop the notation $\hat{D}_T$ and instead simply write $\hat{D}$.

Because the results of this section are more general than the network model studied in Equation (1), our main application, we provide a few examples to develop an intuition for other potential applications. These examples deal with a general set of problems where we observe weighted relationships between pairs of nodes, and these relationships are formed based on distances or dissimilarities between nodes. The latent space model is a special case where the relationship is binary. The geometric results in this section are not confined to even these applications—in fact, our estimation procedure in Theorem 1.1 only requires a consistent estimate of distances along the latent space.

3.1 Examples of relational data with distances, $\hat{D}$

We provide four motivating examples for $\hat{D}$.

Example 1 (A Single Large Network). The researcher only observes a single large network, G, drawn from the distribution specified in Equation (1). We study it in detail in Section 4.

Example 2 (Relational data, Abdelnour et al., 2018; Ginestet et al., 2017; Petersen & Sporns, 2015; Salter-Townshend & McCormick, 2017; Yanchenko & Hoff, 2020). There are K units, such as individuals, neurons, sensors, or firms. The researcher observes an outcome of an interaction between two units i and j, given by $f_{ij,t}$ at time $t = 1, \ldots, T$ with some disturbance $\epsilon_{ij,t}$. For instance, $f_{ij,t} = \Lambda\{d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i^\star, z_j^\star) + \epsilon_{ij,t}\}$ where Λ is a bijective function. This may be binary or continuous, such as an instance of a signal being transmitted between neurons or sensors, some financial flows between individuals, or some transactions between firms. We can then compute $\hat{D}$ with entries $\hat{d}_{ij} = (1/T) \sum_t \Lambda^{-1}(f_{ij,t})$ and under regularity conditions (such as differentiability of Λ^{-1}) our results will follow.

A particularly relevant case is the following. The researcher observes T networks $G_1, \ldots, G_T$, where T could represent the number of observations of one network with a fixed set of nodes, or it could represent the number of observed networks, each with a potentially distinct set of nodes. Here we write $p_{ij} := \mathbb{P}(g_{ij} = 1 \mid z_i^\star, z_j^\star)$ is the probability that nodes i and j connect, given their latent space locations. This term depends on $d_{\mathcal{M}}(z_i^\star, z_j^\star)$. We write this as $p_{ij} = H_{ij}\{d_{\mathcal{M}}(z_i^\star, z_j^\star)\}$ for some invertible function H_{ij}. Here we suppose that the generative model for the networks is constant across all T networks. We can then estimate $\hat{D}_{ij} = H^{-1}(\hat{P}_{ij})$, where $\hat{P}_{ij} = T^{-1} \sum_{i=1}^{T} G_{ijt}$ is the number of observed edges between nodes i and j, normalized by the number of observations T.

Example 3 (Trait Groups, Killworth et al., 1990; McCarty et al., 2001). The n nodes each have one of K traits and the number with a trait k is given by n_k. Locations in the latent space are determined uniquely by a node's trait, and nodes with the same trait share the same location on the latent space, with $z_{k_i}^\star$ denoting the common location of nodes with trait k_i. An interaction between nodes follows $f_{ij} = \Lambda\{d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_{k_i}^\star, z_{k_j}^\star) + \epsilon_{ij}\}$ where again $\Lambda(\cdot)$ is bijective. In this example, the researcher can construct $\hat{D}$ with entries

$$\hat{D}_{kk'} = \frac{1}{n_k n_{k'}} \sum_{i,j} \Lambda^{-1}\big(f_{ij} \cdot \mathbf{1}\{k_i = k, \, k_j = k'\}\big),$$

where n_k is the number of nodes with trait k, which can be obtained from census or surveys.

3.2 Perturbation

In Section 2.2, we saw that the isometric embedding conditions related to the manifold class, curvature, and dimension to the spectrum of a test matrix $W_\kappa(D)$. In practice, since we observe $\hat{D}$, we must construct $\hat{W}_{\hat{\kappa}}(\hat{D})$ and study its spectrum. The main idea of our approach to estimating the manifold class, curvature, and dimension comes from Weyl's inequality, which says that the eigenvalues of the estimated test matrix, $\hat{W}_{\hat{\kappa}}(\hat{D})$ are very close to those of the target $W_\kappa(D)$ if the estimators of the distance matrix and curvature are consistent. Under our assumptions, we are able to bound how the estimated spectrum may deviate from the true spectrum.

> **Proposition 3.1**   Suppose that D is a $K \times K$ distance matrix from K points on $\mathcal{M}^{p^\star}(\kappa^\star)$ satisfying Assumption 1.1, with K chosen such that the $\mathcal{M}^{p^\star}(\kappa^\star)$ is uniquely identified. Assume further that there are estimators $\hat{D}$ and $\hat{\kappa}$ such that $\hat{D}_T - D \xrightarrow{p} 0$ and $\hat{\kappa} \xrightarrow{p} \kappa^\star$ as $T \to \infty$. Let θ_α be defined as the αth quantile of the distribution of $\|\hat{W}_{\hat{\kappa}} - W_\kappa\|_F$. Then, for every $k \in \{1, \ldots, K\}$,

$$\mathbb{P}\big\{|\lambda_k(\hat{W}_{\hat{\kappa}}) - \lambda_k(W_\kappa)| < \theta_\alpha\big\} \le \alpha. \tag{3}$$

Owing to this result, we can study the estimated spectrum in order to look at the metric signature and therefore estimate the manifold class, curvature, and dimension.

3.3 Hypothesis tests of geometry

We now frame the problem of classifying the geometry of D into three hypothesis tests. We will then combine the results of these three tests into a classifier of the geometry. To determine the geometry of $\mathcal{M}$, we use Lemma 2.1 and develop testable statements about the spectrum of the test matrix W_κ.

For the Euclidean case, we can test positive semi-definiteness of the test matrix as

$$H_{0,e} : \lambda_K(W_0) \ge 0, \quad H_{a,e} : \lambda_K(W_0) < 0. \tag{4}$$

Using the same reasoning, for the spherical case, we can test the hypothesis that the embedding space is spherical for some $\kappa > 0$ as

$$H_{0,s} : \lambda_K(W_\kappa) \ge 0, \quad H_{a,s} : \lambda_K(W_\kappa) < 0. \tag{5}$$

since the test matrix must be positive semi-definite.

Finally, to determine if the embedding space is hyperbolic for some $\kappa < 0$, we want to test

$$H_{0,h} : \lambda_2(W_\kappa) = 0, \quad H_{a,h} : \lambda_2(W_\kappa) \ne 0 \tag{6}$$

since the signature switches sign and that the matrix does not have full rank under the assumption on dimension implies that there are zeros in the spectrum.[7]

---

[7]   By Lemma 2.1, failing to reject in the hyperbolic case $\lambda_2(W_\kappa) = 0$ is not enough to conclude that D is hyperbolic, since we must test the first is positive and smallest eigenvalue is negative as well. In practice, however, we found that testing only one eigenvalue was sufficient and, thus, use this simpler test. Clearly, it would also be possible to test all three eigenvalues using an intersection test, which we leave to future work.

A natural first step in deriving a geometry classifier would be to use Proposition 3.1. While this method produces a type-1 error that is below α, the power of the method may be low. So as a classifier, that procedure can be improved upon. In Section 3.5, we derive more powerful tests by approximating the distribution of the eigenvalues under the assumption that the distances are computed along one of the three geometries (recall from Lemma 2.1 that the eigenvalues of W_κ tell us the underlying geometry type).

To derive tests of the three geometry hypotheses, we first need to estimate the curvature of $\mathcal{M}$, which is computed assuming the geometry of $\mathcal{M}$ is curved (i.e., not Euclidean). We use these estimated curvature values to then identify the geometry type. For this reason, we start our discussion by studying curvature.

3.4 Estimating curvature $\kappa^\star$

We begin assuming that the researcher has a consistent estimator $\hat{D}_T \xrightarrow{p} D$ (which we will develop below). Equipped with $\hat{D}_T$, we construct a consistent estimate of $\kappa^\star$. The core observation comes from Lemma 2.1. Namely, by looking at sig(W_κ), we can use the fact that certain eigenvalues must *exactly* be zero under the various geometries. For instance, both $\lambda_K(W_0(D)) = 0$ and $\lambda_K(W_{\kappa^\star}(D)) = 0$ for the Euclidean and spherical cases, respectively, as both are positive semi-definite, provided $p^\star < K$. A similar phenomenon is true for $\lambda_2(W_{\kappa^\star}(D)) = 0$ for the hyperbolic case.

This observation leads to the following estimators of the curvature:

$$\hat{\kappa}_S := \underset{\kappa \in [a,\, b]}{\arg\min} \left| \lambda_1 \{ \kappa W_\kappa(\hat{D}_T) \} \right|, \quad \hat{\kappa}_H := \underset{\kappa \in [-b,\, -a]}{\arg\min} \left| \lambda_2 \{ \kappa W_\kappa(\hat{D}_T) \} \right| \tag{7}$$

for some $0 < a < b$. In Appendix C, we discuss how to pick a and b in practice. The subscript S indicates that this estimate is used when testing if the manifold is spherical. Similarly, the subscript H indicates that this estimate is used when testing if the manifold is hyperbolic. As $T \to \infty$, the estimates approach the true curvature under the correct geometry.

Proposition 3.2   (Consistency of curvature estimates). Suppose that D is a $K \times K$ matrix containing pairwise distances between points in either $\mathbf{S}^p(\kappa^\star)$ or $\mathbf{H}^p(\kappa^\star)$, where $|\kappa^\star| > 0$. Let Z denote the collection of these K points. Suppose there is an estimate $\hat{D}_T$ such that $\hat{D}_T \xrightarrow{p} D$ as $T \to \infty$. Finally, suppose that there is either

1. a unique $\kappa^\star \in [a, b]$ such that $Z \xrightarrow{\text{isom}} \mathbf{S}^p(\kappa)$ for some p. Set $\hat{\kappa}_T = \hat{\kappa}_S$.

2. or a unique $\kappa^\star \in [-b, -a]$ such that $Z \xrightarrow{\text{isom}} \mathbf{H}^p(\kappa)$ for some p. Set $\hat{\kappa}_T = \hat{\kappa}_H$.

In cases (1) and (2), $\hat{\kappa}_T \xrightarrow{p} \kappa^\star$ as $T \to \infty$.

The estimators we propose for curvature are similar to those proposed in Wilson et al. (2014), though Wilson et al. (2014) do not prove that these estimators are consistent.

Before continuing, we want to discuss the requirements in Proposition 3.2. First, Proposition 3.2 requires that $\kappa^\star$ is bounded away from zero (meaning that the space is not flat and hence has non-zero curvature). In other words, the manifold is either spherical or hyperbolic, and a and b must be chosen to include the true curvature value $\kappa^\star$. We also require that there is a unique curvature in which we can find an isometric embedding to ensure that the statement that $\hat{\kappa}_T \xrightarrow{p} \kappa^\star$ makes sense. We also want to emphasize that $\kappa^\star$ is fixed and again is assumed to be non-zero. The case where $\kappa^\star$ changes with T is an interesting and challenging problem that we leave to future work.

3.5 Estimating geometric class $\mathcal{M}^\star$

Again, we suppose the researcher has access to a noisy distance matrix $\hat{D}_T$ that approximates an unknown distance matrix D of interest. The matrix D consists of pair-wise distances between K objects along the surface of $\mathcal{M}$.

3.5.1 Consistent tests for latent geometry

We begin by showing a consistent testing framework and then give a bootstrap method for implementation in Algorithm 3.

To test $H_{0,e}$, first, define a rejection region $\mathcal{R}_T = (-\infty, \delta_T]$ for some real-valued sequence $\delta_T \in (-\infty, 0]$ and we define our test $\phi_T(\hat{W}_0) \in \{0, 1\}$ as

$$\phi_T(\hat{W}_0) = \begin{cases} 0, & \lambda_K(\hat{W}_0) \in \mathcal{R}_T \\ 1, & \lambda_K(\hat{W}_0) \notin \mathcal{R}_T. \end{cases} \tag{8}$$

If the test is 0, this indicates that we fail to reject the null hypothesis $H_{0,e}$, while if the test is 1, we reject the null hypothesis $H_{0,e}$. We use this notation throughout the paper when discussing the output of a hypothesis test.

When testing if D is spherical, let $\hat{\kappa}$ denote an estimate of κ, defined in Proposition 3.2. We define our rejection region of $H_{0,s}$ as $\mathcal{R}_T = (-\infty, \delta_T]$ and our test as

$$\phi_T(\hat{W}_{\hat{\kappa}}) = \begin{cases} 0, & \lambda_K(\hat{W}_{\hat{\kappa}}) \in \mathcal{R}_T, \\ 1, & \lambda_K(\hat{W}_{\hat{\kappa}}) \notin \mathcal{R}_T, \end{cases} \tag{9}$$

which is a similar test of positive semi-definiteness.

Finally, when testing if D is hyperbolic, let $\hat{\kappa}$ denote an estimate of $\kappa < 0$, defined in Proposition 3.2. We define our rejection region of $H_{0,h}$ as $\mathcal{R}_T = [\delta_T, \infty)$ and define our test as

$$\phi_T(W_{\hat{\kappa}}) = \begin{cases} 0, & \lambda_2(\hat{W}_{\hat{\kappa}}) \in \mathcal{R}_T, \\ 1, & \lambda_2(\hat{W}_{\hat{\kappa}}) \notin \mathcal{R}_T, \end{cases} \tag{10}$$

which looks to reject positivity of the second eigenvalue as per the metric signature.

We now study what conditions must hold on this sequence δ_T in order for the three tests to be consistent, by which we mean that the probability the test rejects the null goes to 1 under the alternative hypothesis and that the probability it fails to reject the null goes to 1 under the null.

> **Proposition 3.3** Let $\delta_T = o_P(1)$ be a random or deterministic sequence and let Assumption 1.1 hold. Let $\hat{D}_T \overset{p}{\to} D$ as $T \to \infty$. Then,
>
> 1. If $\delta_T \in (-\infty, 0]$, $\delta_T = o_P(1)$ and $\mathbb{P}\{\lambda_K(\hat{W}_0) \leq \delta_T\} = 1 - o(1)$, then the test for $H_{0,e}$ in (4) with rejection region $\mathcal{R}_T := (-\infty, \delta_T]$ is consistent.
> 2. If $\delta_T \in (-\infty, 0]$, $\delta_T = o_P(1)$ and $\mathbb{P}\{\lambda_K(\hat{W}_{\hat{\kappa}}) \leq \delta_T\} = 1 - o(1)$ with $\hat{\kappa} \in [a, b]$, then the test for $H_{0,s}$ in (5) with rejection region $\mathcal{R}_T := (-\infty, \delta_T]$ is consistent.
> 3. If $\delta_T \in [0, \infty)$, $\delta_T = o_P(1)$ and $\mathbb{P}\{\lambda_2(\hat{W}_{\hat{\kappa}}) \geq \delta_T\} = 1 - o(1)$ with $\hat{\kappa} \in [-b, -a]$, then the test for $H_{0,h}$ in (6) with rejection region $\mathcal{R}_T := [\delta_T, \infty)$ is consistent.

In order to combine these tests into a single estimate of the latent space geometry, we suggest using an ordered test. There are six possible orderings of such a test (e.g, Euclidean, then spherical, then hyperbolic). The proof is simple and uses the fact that each of the three geometry tests is consistent, under suitable assumptions on the threshold sequence δ_T.

> **Proposition 3.4** (Consistent estimation of geometry type). Under the assumptions on the sequence δ in Proposition 3.3, any of the 6 ordered tests return a consistent estimate of the latent space geometry.

We omit the proof of Proposition 3.4 which follows immediately from Proposition 3.3. We have shown that we can use the observed distance matrix $\hat{D}_T$ to test the hypotheses that the latent space is Euclidean, spherical, or hyperbolic. From these tests, we define $\hat{\mathcal{M}}^{\hat{p}}(\hat{\kappa})$ as the intersection of the three tests. That is, the estimated latent geometry based on $\hat{D}_T$ is defined by the result of three

hypothesis tests in equations (8), (9), and (10). More specifically, we can select any estimator that preserves consistency. As noted, for example, we can use an ordered test to estimate the geometry type. Thanks to Proposition 3.3, with sufficiently large T the probability that more than one of these tests will fail to reject the null goes to zero, leading to a consistent test.

3.6 Estimating dimension $p^\star$

Given $\hat{D}$, $\hat{\kappa}$, and $\hat{\mathcal{M}}$, we develop a consistent estimate of $p^\star$, the minimal dimension of the manifold class in which the points can be embedded. We focus on the minimum dimension since, trivially, if one can embed in p one can embed in $p' > p$ for all p'. As we noted in Lemma 2.1, we see that $p^\star$ relates to the rank. So we proceed by estimating the rank of $W_\kappa(D)$.

We present two approaches. The first continues our use of the logic of Weyl's inequality to propose a consistent estimate as $T \to \infty$. The second uses the laddle plot method of Luo and Li (2016). We did not verify the required assumptions in Luo and Li (2016), so we do not claim their estimator is consistent in our problem, but we find in practice (as do they) that the estimator performs well, so we suggest practitioners actually use this.

3.6.1 Spectral estimate of dimension

We are interested in finding the rank of W_κ using $\hat{W}_{\hat{\kappa}}$. To do this, we let ϵ_T be a (potentially random) sequence such that ϵ_T goes to zero slower than any $\lambda_j(\hat{W}_{\hat{\kappa}})$ for which $\lambda_j(W_\kappa) = 0$. In other words, we select ϵ_T to go zero slower than the slowest zero eigenvalue of the test matrix W_κ.

Since the rank of a matrix is the number of non-zero eigenvalues, we will estimate the number of non-zero eigenvalues of W_κ. To do this, we define a rejection region $\mathcal{R}_T = (-\epsilon_T, \epsilon_T)$. For any index T, define the estimated rank to be

$$\widehat{\text{rank}}(W_\kappa) = \#\{j = 1, \ldots, K : \lambda_j(\hat{W}_{\hat{\kappa}}) \notin \mathcal{R}_T\}. \tag{11}$$

Our estimate of the rank is then the number of observed eigenvalues that are sufficiently far away from zero, as measured by the threshold sequence ϵ_T.

Clearly, the performance of this estimator depends on the choice of the threshold sequence ϵ_T. If ϵ_T does not converge to zero (in probability), then this estimate cannot be consistent, since it will eventually start to count zero eigenvalues of W_κ as being non-zero. Hence convergence to zero is a necessary condition. It must also converge fast enough to zero. For example, if $\lambda_j(W_\kappa) = 0$ and we have access to $\lambda_j(\hat{W}_{\hat{\kappa}}) = 1/T^2$ (i.e., a deterministic sequence), and we use $\epsilon_T = 1/T$, then we will under-count the rank of W_κ because we will classify the eigenvalue $\lambda_j(W_\kappa)$ as non-zero at every T.

We, therefore, must pick an ϵ_T that goes to zero slower than all estimates of eigenvalues for which their true counterpart is zero. From Weyl's inequality, we know what such sequences look like. For any index k for which $\lambda_k(W_\kappa) = 0$,

$$|\lambda_w(\hat{W}_{\hat{\kappa}})| = |\lambda_k(\hat{W}_{\hat{\kappa}}) - \lambda_k(\hat{W}_{\hat{\kappa}})| \leq \|\hat{W}_{\hat{\kappa}} - W_\kappa\|_F,$$

where the inequality is due to Weyl's. By defining $r_T := \|\hat{W}_{\hat{\kappa}} - W_\kappa\|_F$, from the conditions in Theorem 1.1, we know that $r_T = o_P(1)$. We need to select $\epsilon_T \to 0$ with $r_T/\epsilon_T \to 0$, which means that ϵ_T goes to zero slower than r_T does.

> **Proposition 3.5** (Consistency of minimum dimension estimate). Choose $\epsilon_T \to 0$ such that $r_T/\epsilon_T \to 0$. If the assumptions in Theorem 1.1 hold, then $\widehat{\text{rank}}(W_\kappa) \overset{p}{\to} \text{rank}(W_\kappa)$ as $T \to \infty$. Using Lemma 2.1, we can, therefore, consistently estimate $p^\star$ as $T \to \infty$.

While the rank estimator above leads to a consistent estimate for suitable chosen ϵ_T, choosing a sequence that satisfies these conditions in practice is challenging. We, therefore, recommend in practice to estimate the rank with a different estimator.

3.6.2 *Luo and Li (2016): Laddle Plot*

Recent work, by Luo and Li (2016), however, has been shown to have more appealing finite sample performance and in Appendix E, we provide the algorithm to estimate the rank with this method. The intuition for their approach is as follows. Looking at the scree plot (related to our above approach) is consistent. And a bootstrap procedure, leveraging the fact that the eigenvectors corresponding to indices beyond the rank will be uncorrelated in a bootstrap) also performs well. They note that under certain regularity conditions (a sufficiently fast estimate of the matrix W_κ, a self-similar bootstrap estimator), the combination of these two—a scree plot together with a bootstrap evaluation of eigenvector uncorrelatedness—performs better than either.

Luo and Li (2016) prove that in a number of problems, their rank estimate is consistent. We are not able to verify these conditions in practice, since they assume that their data consists of i.i.d. data. However, in our case, our data is independent but *not* identically distributed. Therefore, we do not make a claim about the consistency of this approach when applied to our problem. We do, however, note that in simulations this approach has better finite sample properties.

To summarize, through Propositions 3.2, 3.4, and 3.5, we have established that the procedure in Algorithm 1 is consistent for its estimands, as claimed in Theorem 1.1.

Returning to our four examples, it is easy to see that Examples 2 and 3 immediately satisfy the assumptions in Theorem 1.1, and as a consequence in each of those cases, we can consistently recover the geometry. The situation is more subtle for Example 1, where rather than directly observing K units and noisy distances among them, the researcher observes some graph G is more subtle. From the graph, distances must be constructed and then geometry estimated. This is the subject of Section 4.

4 Identifying the latent space using only graph data

Having developed statistical tests to estimate the geometry behind a collection of points when we observe and arbitrary noisy distance matrix (the statistical geometry problem from Section 1.1), we now turn the original network problem. Specifically, our goal is to use a graph drawn from the latent space model in (1) and estimate the type, curvature, and dimension of the latent space. Theorem 1.2 presents the answer to this problem and demonstrates how the manifold can be consistently estimated. The proof of the theorem is provided in Appendix A. In this section, we lay out the ingredients: how one constructs the noisy distance matrix and adjusts for fixed effects. With these we are able to prove Theorem 1.2, which states that Algorithm 2 returns a consistent estimate of $\mathcal{M}^{p^\star}(\kappa^\star)$.

Importantly, we also provide a number of examples of node location distributions, motivated by empirically-relevant models, to demonstrate that our core Assumption 1.3 for the clique-based method holds (Section 4.4). We conclude by discussing some practical choices for implementation.

4.1 Estimating distances using graph data

In order to apply the geometric test, we use the graph G to estimate a set of distances on $\mathcal{M}^{p^\star}(\kappa^\star)$. Since the probabilities defined by the latent space model in (1), our approach is to use estimated linking frequencies in order to identify a system of implied distances between K points in the manifold.

To motivate our approach to constructing a distance matrix, we consider a simplification of the main model in (1). We make two assumptions for illustration, which are subsequently relaxed in the main analysis. First, suppose there are no individual effects (so $v_i^\star = 0 \; \forall i$). Second, suppose that nodes are assigned to one of K distinct points in the latent space, which we denote by $\zeta_1^\star, \ldots, \zeta_K^\star$. Define $V_k = \{j \in \{1, \ldots, n\} : z_j^\star = \zeta_k^\star\}$ to be the set of nodes at location $z_k^\star$. Under this simplification, we can write the distance between points $z_k^\star$ and $z_{k'}^\star$ using the definition of the latent space model in (1) as

$$d_{k,k'} = -\log(p_{k,k'}) \tag{12}$$

where $p_{k,k'} := \mathbb{P}(G_{k,k'} = 1 | z^\star)$ is the probability that nodes at locations $z_k^\star$ and $z_{k'}^\star$ connect for any $k, k' \in \{1, \ldots, K\}$.[8] Then, we can estimate the probability $p_{k,k'}$ by

$$\hat{p}_{k,k'} := \frac{1}{|V_k||V_{k'}|} \sum_{(i,j) \in V_k \times V_{k'}} G_{ij}. \tag{13}$$

In words, this estimator counts the number of observed edges between $z_k^\star$ and $z_{k'}^\star$ and divides by the number of possible edges, given by $|V_k||V_{k'}|$. Since G_{ij} i.i.d. Bernoulli$(p_{k,k'})$ for $(i, j) \in V_k \times V_{k'}$, this estimator is unbiased for $p_{k,k'}$. In addition, supposing that $|V_i| \to \infty$ as $n \to \infty$, the weak law of large numbers implies that $\hat{p}_{k,k'} - p_{k,k'} \xrightarrow{p} 0$.

By (12), we can estimate a $K \times K$ distance matrix $\hat{D} = \{\hat{d}_{kk'}\}$ comprised of entries with distances between ζ_k and $\zeta_{k'}$ by

$$\hat{d}_{kk'} = -\left[\log(\hat{p}_{k,k'})\right]_{k,k'}.$$

We can then apply the continuous mapping theorem to show that $\hat{d}_{k,k'} - d_{k,k'} \xrightarrow{p} 0$. That is, the assumption in Theorem 1.1 that we have access to a consistent estimate of distances along the latent space holds. Here, the sample size T is the number of edges between the K points on the latent space, given by $|V_k||V_{k'}|$ for every pair of points $\zeta_k^\star$ and $\zeta_{k'}^\star$.

To summarize, in this example, we have used the edges in G to estimate a distance matrix $\hat{D}$ between K points in the unobserved latent space $\mathcal{M}^{p^\star}(\kappa^\star)$. From this, we can apply Theorem 1.1 to consistently estimate the geometry. However, this example made two simplifying assumptions: no fixed effects and every node is located on exactly one of K finite points on the manifold. Our more general result, which we now describe, relaxes these assumptions considerably.

In (1), nodes have individual fixed effects $v_i^\star$ as well as latent positions $z_i^\star$. The individual effects describe heterogeneity in the propensity for an individual to form connections and are not directly related to which connections will form, which is what the latent space captures. We would prefer to estimate the distances used to test hypotheses about latent geometry without potential confounding by individual effects not specifically related to the geometry. We accomplish this by marginalizing over the individual effects in (1). Recalling that the support of $v_i^\star$ is $(-\infty, 0]$, we integrate out the node effects to find that

$$\mathbb{P}\left\{G_{ij} = 1 | z^\star, \mathcal{M}^{p^\star}(\kappa^\star)\right\} = E\{\exp(v_i)\}^2 \exp\{-d(z_i, z_j)\}.$$

Solving for the distance $d(z_i^\star, z_j^\star)$, we have

$$d(z_i^\star, z_j^\star) = -\log(p_{i,j}) + 2\log[E\{\exp(v_i)\}]. \tag{14}$$

Note that if $v_i = 0$ with probability 1, which we assumed in the simplified model above, then $E\{\exp(v)\} = 1$, so that (14) becomes (12). Here we use properties of the exponential function to isolate the node effect term $E\{\exp(v)^2\}$. See Appendix J for a discussion on how to marginalize out the node effects when other link functions are used. We must now estimate (i) the term $p_{k,k}$ and then (ii) the term $\log[E\{\exp(v)\}]$.

Our strategy, therefore, has two components. The first is to identify some $K \times K$ distance matrix D among K points on the manifold that could be used to test the geometry and then develop a consistent estimator of the pair-wise distances between these K points. By showing that the regularity conditions for the geometric result (Theorem 1.1) are met under the network formation model (Assumptions 1.1–1.3), we can identify the geometry. The second (lesser) component is to recognize and adjust for the fact that in going from linking probabilities to distance matrix estimation, we need to adjust for the nuisance of the fixed effects parameters.

---

[8] Clearly in this simplified model, this corresponds to a stochastic block model with K communities with a linking probability law $p_{kk'}$ that satisfies a geometric restriction.

4.2 Estimating probability of edges

We begin by estimating the term $p_{kk'}$. The approach exploits the clique structure in the network. Cliques are useful because under regularity conditions that are useful in applications, a clique of size ℓ tells us that all nodes in that clique are extremely likely to be very close in the latent space. Therefore, we can treat them as if they are all located at the same point in the manifold. Of course, this is not exactly true, but for sufficiently large clique size ℓ, this approximation strategy becomes more and more accurate. If we find K disjoint cliques in the graph, then we have identified K distinct points on the latent space. Then, by counting the number of edges between cliques, we can approximate all pairwise distances between these cliques, as we did above in (13). So by looking at the clique structure, we can compute an estimate $\hat{D}$ which approximates distances along the surface of the latent space. We now describe this approach more formally.

Let $C_1, \ldots, C_K$ denote K disjoint cliques of size ℓ. We can take ℓ to grow like $\log(n)$, and such cliques will exist with probability tending to 1 in the graph (see Appendix K for a discussion). Let ζ_k be the Fréchet mean of the locations of the nodes in clique k:

$$\zeta_k^\star(\ell) = \operatorname*{argmin}_{\zeta \in \mathcal{M}^{p^\star}(\kappa^\star)} \sum_{i \in C_k}^{\ell} d^2_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i^\star, \zeta).$$

Informally, this point represents the average of all node locations in the latent space. We then define a $K \times K$ distance matrix $D = \{d_{kk'}\}$ with entries

$$d_{kk'} := d_{\mathcal{M}^{p^\star}(\kappa^\star)}(\zeta_k^\star(\ell), \zeta_{k'}^\star(\ell)).$$

Note that $\zeta^\star$ and D are indexed by the clique size ℓ and, therefore, n. Conditioned on seeing cliques of size ℓ, we show that the terms $\zeta_k^\star(\ell)$ each converge to some fixed point $\zeta_k^\star$ on the latent space under general and relevant conditions (see Assumption 1.3 and Section 4.4) . Thus, the matrix containing the pairwise distances is the distance matrix we wish to estimate. Recall that this is the distance term on the left hand side of (14).

To estimate the probability on the right hand side of (14), we compute

$$\hat{p}_{kk'} = \frac{1}{\ell^2} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ij} \mathbf{1}\{i \in C_k, j \in C_{k'}\}. \tag{15}$$

4.3 Adjusting for the expected fixed effect

For notational convenience let us denote $\tau := E\{\exp(v_i^\star)\}$. The main observation here is that if we focus on nodes that are close together in the latent space, since the distance term in such cases will be (nearly) zero in (1), it we will be able to estimate and account for the individual effects.

Naively, since we use cliques to estimate $p_{k,k'}$, we may consider using these nodes to also estimate τ. We could, for example, compute the number of edges in the cliques out of the number of possible edges. However, this estimate will be 1, since by definition all edges exist between nodes in the same clique. Therefore, we define a closely related idea, which we call the 'almost-clique'.

Fix an ℓ-clique C_k and a number $t < \ell$. We define an 'almost-clique' $I_k(t)$ by

$$I_k(t) := \left\{ j : j \notin C_k, |C_k| > \sum_{i \in C_k} G_{i,j} \geq t \right\}$$

to be the set of nodes not in C_k that connect to at least t nodes in C_k. The intuition behind this definition is that if $t \approx \ell$, then the distance between nodes in $I_k(t)$ should be close to zero, but since they are not in the clique not all connections will be realized.

We can estimate the probability that nodes in the sub-graph induced by $I_k(t)$ connect,

$$\hat{E}(t, k) = \binom{|I_k(t)|}{2}^{-1} \sum_{(i,j) \in I_k(t) \times I_k(t)} G_{i,j}.$$

To estimate $E(\exp(v))$, we average the above term over all cliques, leading to an estimate

$$\hat{E}(\exp(v)^2) := \frac{1}{K} \sum_{k=1}^{K} \hat{E}_v(t, k).$$

This approach suffers from selection bias when the clique size is large. That is, by Assumption 1.2 all individuals have independent and identically distributed $v_i^\star$ terms. Conditional on being part of a large clique, however, an individual is likely to be on the right tail of the v_i distribution. We could adjust for this bias by, for example, assuming a parametric model for v_i. If we made such an assumption we could compute a correction for the selection bias based on the tail of the assumed distribution. In practice, we found our non-parametric estimator worked sufficiently well without such a correction. We suggest taking t to be large, for example, $t = \ell - 1$ because our simulations suggest that large values of t reduce the selection bias and, therefore, increases the accuracy of our method.

4.4 Examples satisfying Assumption 1.3

Part (a) of Assumption 1.3 is innocuous. It simply requires that K points identify the manifold. Note that the existence of K points out of $n \to \infty$ random points with continuous distribution over any ball on the manifold will identify it outside a measure zero event in the usual measure. To see this, it is easy to observe that with continuous mass on any patch of a sphere, K points will exist that are not simply forming an arc or are arbitrarily local with probability tending to one.

Part (b) of Assumption 1.3 requires that for sufficiently large ℓ, nodes in a clique are at approximately the same location on the latent surface, which allows us to conclude consistency and asymptotic normality of our distance estimates. We provide a general set of assumptions that are sufficient and show several models used in applied work are covered.

> **Assumption 4.1** Let $\Omega_n = [0, A_n^{1/p^\star}]^{p^\star} \subset \mathcal{M}^{p^\star}(\kappa^\star)$ be such that $A_n = o(n)$ is growing. Assume either
>
> 1. *Bounded Support:* Ω_n is the support of the distribution $F_n(z)$ of locations, or
> 2. *Thin Tails:* $F_n(z)$ satisfies, for some constant $b > 0$,
>
> $$\mathbb{P}\left(\max_{i,j} d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j) > A_n^{1/p^\star}\right) \leq \exp\left(-bA_n^{1/p^\star} + \log n\right) \to 0.$$

Obviously (1) implies (2), but (2) allows for sub-Gaussian tails with unbounded support. The next assumption is natural and general, holding as long as distributions are not taking mass only on some sub-manifolds. It says that the odds that all ℓ independent draws are within δ-distance of each other is inversely related to the volume of that ball. We verify this for examples below.

For any set of ℓ points in the latent space, define the event

$$\mathcal{E}_\delta := \left\{\max_{1 \leq i < j \leq \ell} d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j) < \delta\right\}.$$

In words, $\mathcal{E}_\delta$ is the event that the largest distance between these ℓ points is less than δ. We omit the dependence on ℓ when writing E_δ for convenience.

> **Assumption 4.2** The location distribution sequence satisfies, for any δ and ℓ,
>
> $$\mathbb{P}(\mathcal{E}_\delta) = \frac{a(\delta)}{A_n^\ell}(1 + o(1)).$$

for a positive constant $a(\delta)$, which can depend on δ.

Let $\mu_{d,n} := E\{d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)\}$, where z_i is independent of z_j, be the expected distance between two location draws.

Assumption 4.3   For any $\delta > 0$, assume a sequence of clique sizes $\ell_n \to \infty$ satisfying

$$\ell_n \geq 2\frac{\log A_n}{a(\delta)(\mu_{d,n} - \delta)} + 1$$

for all n sufficiently large.

The above assumption is written for any $\delta > 0$, but in fact we are only concerned with $\delta < \mu_{d,n}$, the average distance between two arbitrary points drawn on the latent surface. For $\delta > \mu_{d,n}$, it is unlikely that two nodes that are at least $\mu_{d,n}$ apart would form an edge. Therefore, we are only interested in checking the condition for sufficiently small δ.

We show that Assumptions 1.3 hold under the above assumptions.

Proposition 4.1   Let Assumptions 1.1, 1.2, 4.1, 4.2, and 4.3 hold. Then Assumptions 1.3 holds.

We now consider three common ways of modelling node locations in latent space. For each model, we show that Assumptions 1.1, 1.2, 4.1, and 4.2 hold.

These three examples cover a wide span of location distribution models. The first, a lattice, is a stylized example that simply corresponds to a community block model with a geometric structure, since all nodes exactly live at one of several locations. The second, a uniform distribution, is the other extreme where nodes have no bias towards any specific locations that help organize cliques. This is the most adversarial case. The Gaussian Mixture Model lives in-between and interpolates between these models and is frequently used in practical statistical modelling. It functions much like the uniform if the dispersion of nodes about their type-centres tends to infinity and is similar to the lattice if the dispersion tends to zero.

Example 4   (Lattice). Every node is a member of some community $t_i \in \{1, \ldots, T_n\}$. Node locations on the manifold are determined as follows. Let a lattice $\Lambda_n = \{0, \ldots, T_n^{1/p^\star}\}^{p^\star} \subset \Omega_n$ be a list of coordinates which serve as the support for node placements with a spacing of 1 between points along any axis. Node locations z_i are placed i.i.d. across these T_n points on the manifold, i.e., on the lattice Λ_n, so Assumption 4.1 holds and t_i corresponds to the location drawn. Assumption 4.2 holds for any $\delta < 1$, with $A_n = T_n$, since every location $t \in \{0, \ldots, T_n\}$ is equally likely: $\mathbb{P}(\mathcal{E}_\delta) = (1/T_n)^\ell$.

To understand the growth-rate restrictions on $\ell(n)$, recall that $\ell(n)$ has to grow slowly enough so that there are cliques of size ℓ with probability 1. Consider the lattice above. Let us assume A_n is fixed at $A > 0$. Then, at each location in the lattice, the nodes connect independently with some fixed probability determined by the fixed effects, so these nodes are in an Erdos-Renyi model. So we need $\ell(n) < \log(n)$ in order for there to be an ℓ-clique in the graph with probability approaching one (recall Appendix K). And according to Assumption 4.3, we need $\ell > \log(A_n) = \log(A)$, a constant. So taking $\ell(n) \to \infty$ to be growing slower than $\log(n)$ satisfies Assumption 4.3 for sufficiently large n.

The next example is in some sense the most adversarial model for our method. By placing nodes uniformly over the support, the odds of accumulation, and, therefore, clique formation, at any given point are minimized. Still, the result holds.

Example 5   (Uniform Distribution). Node locations $z_i \sim U(\Omega_n)$ are assumed to be drawn independently, uniformly over Ω_n such that Assumption 4.1 holds. It immediately follows by a calculation that $\mathbb{P}(\mathcal{E}_\delta) = (\delta^{p^\star}/\text{volume}(\Omega_n))^\ell$, so Assumption 4.2 holds as well.

The final example interpolates between the extremes of the lattice and uniform models.

Example 6   (Gaussian Mixture Model). As in the lattice example, every node is a member of some community $t_i \in \{1, \ldots, T_n\}$. Node locations on the manifold are determined as follows. Let $\Lambda_n \subset \Omega_n$ be a lattice of T_n points, which designate community centres, randomly drawn as follows. There is a support $\Omega'_n \subset \Omega_n$, with $\Omega'_n = [0, B_n^{1/p^\star}]^{p^\star}$ and the community centres $\zeta_t \sim U(\Omega'_n)$ are drawn independently, uniformly over this support.

Given the community centres, every node conditional on the community it is assigned to has a location that is dispersed about the centre

$$z_i | t_i = t \sim F_n(z; \zeta_t, \sigma_t^2)$$

independently. Examples include Gaussian in Euclidean space, the von Mises-Fisher on the sphere, and the wrapped-normal distribution on hyperbolic space. Note that community centres reside within $\Omega'_n \subset \Omega_n$ but given the distribution F_n, the random variable z_i may have full support over $\mathcal{M}^{p^\star}(\kappa^\star)$.

The distance between the outer boundary of Ω'_n and Ω, denoted by $\Delta_n :=$ $A_n^{1/p^\star} - B_n^{1/p^\star}$ is assumed to be growing at a sufficiently fast, possibly sublogarithmic rate $\Delta_n = \omega(\sqrt{\log n})$. Then one can calculate that Condition 2 of Assumption 4.1. The function of this is to ensure that even if the z_i have full support, they are going to essentially all live within Ω_n.

In this setup, we can calculate that

$$\mathbb{P}(\mathcal{E}_\delta) = \left(\frac{\delta^{p^\star}}{A_n}\right)^\ell \times (1 + o(1)).$$

so that Assumption 4.2 holds too. Complete calculations are in Appendix A.

4.5 Consistent estimates of location and fixed effects

Suppose that the researcher has access to estimates of the node locations fixed effects $(\hat{z}_i, \hat{v}_i)$. Our result is agnostic to how these estimators are constructed and allows any consistent ones from the literature.

Corollary 1   Let $\hat{\mathcal{M}}^{\hat{p}}(\hat{\kappa})$ denote the estimate of the geometry from Algorithm 2. Let $\hat{z}_i(\mathcal{M}^{p^\star}(\kappa^\star))$ and $v_i(\mathcal{M}^{p^\star}(\kappa^\star))$ be any set of consistent estimators, computed using the assumed geometry $\mathcal{M}^{p^\star}(\kappa^\star)$. Then, $\hat{z}_i(\hat{\mathcal{M}}^{\hat{p}}(\hat{\kappa}))$ and $\hat{v}_i(\hat{\mathcal{M}}^{\hat{p}}(\hat{\kappa}))$ are consistent.

The proof is straightforward consequence of Theorem 1.2, so we omit it. The mode of convergence in Corollary 1 depends on how the estimates of node locations and effects converge. For example, if conditioned on the right geometry, $\max_{1 \le i \le n} |\hat{v}_i - v_i^\star| \overset{p}{\to} 0$, then this same convergence holds in Corollary 1.

4.6 Practical implementation of manifold hypothesis tests via bootstrap

Following Proposition 3.4, any arbitrary implementation of hypothesis testing for (8), (9), and (10) would suffice. For instance, following the logic of Weyl's inequality and Proposition 3.1, since our estimate of W is asymptotically normally distributed, one can readily develop an analytic conservative test. Since Weyl's inequality is often a loose upper bound, we suggest take a different approach and provide a specific method which is extremely fast, easy-to-implement, and effective in simulations and data work via a sub-sample bootstrap.

There are two non-standard features of our problem that make classical bootstrapping challenging. First, W_0 does not have full rank, meaning that $\lambda_K(W_0) \le 0$; under the null $\lambda_K(W_0) = 0$ meaning the parameter is on the boundary of the parameter space. Classical bootstrap is not valid in

such a case (Andrews, 2000). Second, W_κ has repeated eigenvalues at zero under the null for both curved spaces, which again excludes the classical bootstrap (Eaton & Taylor, 1991).

We adapt the sub-sampling method from Politis and Romano (1994) which is valid both with parameters on the boundary and with repeated eigenvalues. Algorithm 3 in the appendix presents the method. It uses sub-sampling to generate a distribution $\{D_b^\star\}_{b=1}^B$ of B bootstraps of the $K \times K$ distance matrices. Then given this distribution, the method constructs the corresponding distribution of the eigenvalue of interest, $\lambda_{\tilde{k}}\{W_\kappa(D_b^\star)\}$, which is then used to test the null hypothesis for the original data. In Appendix K, we investigate another way of testing geometry via the Cayley–Menger determinant. We show numerically that this procedure has lower power than the bootstrap procedure discussed above, so in the following simulations, we use the bootstrap procedure.

5 Simulation evaluation

In this section, we examine the performance of our proposed method on simulated data from each of the three candidate geometries. The goal is to understand how well the methods perform in a setting where we know the (simulated) true geometry. We first examine the Type 1 error and power of the tests to select manifold class and then show the performance of our algorithm for estimating the latent dimension. We provide additional simulation results, including results for estimating curvature, in Appendix G.

We discuss the Type 1 error and power of our proposed tests under various values for the clique size (ℓ) and the number of cliques we select for our estimation (K). In all cases, we simulate graphs in the following way. First, we generate a set of groups centres randomly in the latent geometry and dimension to be tested. For the Type 1 error, we generate graphs on $n = 200$ nodes and use $K = 5$ and clique sizes that are realistic in the data: $\ell \in \{4, 5, 6\}$. When generating the type 1 error figures, the graph statistics are as follows. For the Euclidean graphs, the average degree is 41 and the average clustering coefficient is 0.26. For the spherical graphs, the average degree is 56 and the average clustering coefficient is 0.35. For the hyperbolic graphs, the average degree is 40 and the average clustering coefficient is 0.26.

For the power simulations, we generated 25 sets of latent positions and then, for each set of latent positions, constructed 100 graphs. When comparing across values of K and ℓ, we use the same graphs for all comparisons (e.g., when comparing $K = 10$ vs $K = 5$, the cliques in the $K = 5$ set are randomly selected from among those in $K = 10$). We provide specific values we used for simulations and additional results in Appendix B.

Figures 3 and 4 show results for Type 1 error and power of the tests we propose using the simulation procedure described above. Each point in the boxplot is the fraction of rejections out of 250 graphs for a given set of latent space positions. The variation in the boxplot, therefore, represents heterogeneity across latent space locations that are consistent with the true underlying geometry and the simulation procedure we use. Figure 3 shows boxplots of the Type 1 error for each of the three null hypotheses for three values of ℓ. We focus on variation in the Type 1 error across values of ℓ to see whether the
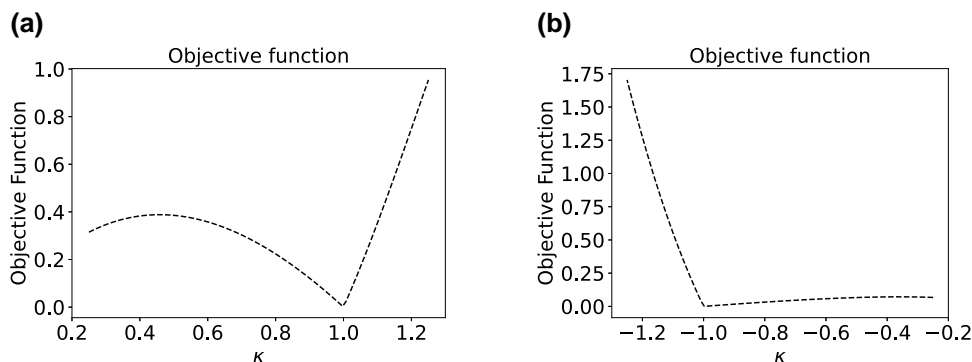

Figure 2. Plot of the objective function from (7) when D corresponds to 15 points in $\mathbf{S}^2(1)$ (left) and $\mathbf{H}^2(-1)$ (right). We plot the curvature κ against the value of the function $\kappa \mapsto |\lambda_1(\cos(\sqrt{\kappa}D))|$ (left) and of the function $\kappa \mapsto |\lambda_2(\cos(\sqrt{\kappa}D))|$ (right). We see that at the true κ, the objective function is minimized. (a) Spherical. (b) Hyperbolic.
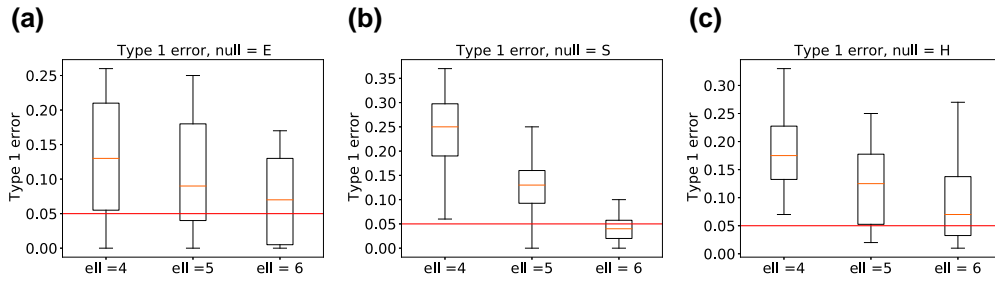
Figure 3. Estimated Type 1 error. For each set of LS positions, we perform the test 100 times and plot the average rejection probability.
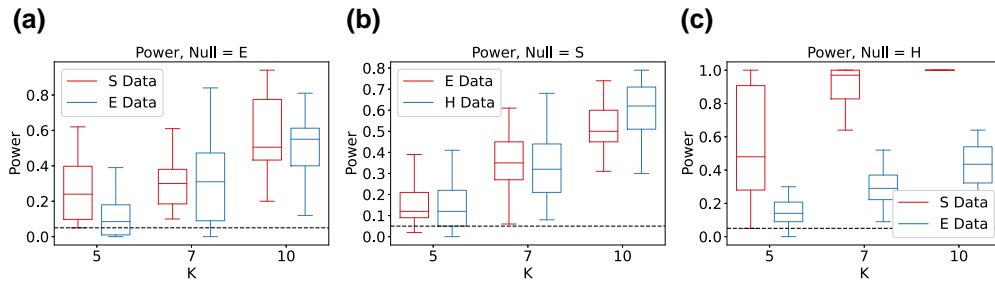

Figure 4. Estimated power using simulated LS positions. For each set of LS positions, we perform the test 100 times and plot the average rejection probability. (a) Euclidean null. (b) Spherical null. (c) Hyperbolic null.

Table 1. Average probability of correctly predicting the dimension of the latent space, averaged across 25 different sets of $n = 200$ latent space positions.

	True Geometry		
	$\mathbb{R}^4$	$\mathbf{S}^3(1)$	$\mathbf{H}^3(-1)$
$\mathbb{P}(\hat{p} < p^\star)$	0.02	0.03	0.06
$\mathbb{P}(\hat{p} \geq p^\star)$	0.98	0.97	0.94

Note. We use $K = 7$ and $\ell = 4$. For each set of latent space positions, we generate 50 networks and predict the dimension.

properties of the Politis and Romano (1994) bootstrap procedure are preserved empirically. We see that, in all three cases, the Type 1 error decreases as the clique size increases. Further, for the Euclidean and hyperbolic cases, the Type 1 error tends to be below the nominal level of five percent, but the spherical type 1 error is higher than five percent. In Figure 4, we see that the power increases as we increase K for all three geometries. In these simulations, we use $\ell = 5$. Recall that all of our manifolds are locally Euclidean—indeed that is part of their definition. So, it is unsurprising, if not expected, that power against Euclidean alternatives rises more slowly than power against alternatives of the opposite curvature. Appendix I contains more simulations.

Moving now to the estimates of the minimal dimension, we consider $p \geq 2$ and take $\max(2, \hat{p})$ as our estimate of the dimension of $\mathcal{M}^{p^\star}(\kappa^\star)$. In Table 1, we give our estimates of the dimension for the three geometries.

6 Examples from economics and biology

In this section, we demonstrate the performance of our method settings with the complexity of observed data. We demonstrate that, in vastly distinct contexts, our approach captures features of the

underlying geometry that provide contextually salient insights. We begin by offering guidance on choices a practitioner would make when implementing the method, then provide examples from three contexts. Access to data from Banerjee et al. (2019) is restricted, however, a similar dataset is freely available here: https://doi.org/10.7910/DVN/U3BIHX. Data from our neural network example is available here: https://www.dynamic-connectome.org/?page·id=25. Replication code is available here: https://zenodo.org/record/7474776.Y6TryS-B2fU.

6.1 Choices for implementation of Algorithm 2

A key decision for implementation is how to identify and then select cliques for a given graph. Overall, there are several considerations.

6.1.1 Choosing clique size

Algorithm 2 requires identifying K cliques of size ℓ (again recalling that we only assume for simplicity that they are of the same size ℓ), with ℓ growing slowly. K is assumed not to grow with n, although more points will raise the power of our tests. This means that the researcher *does not* need to identify all ℓ-cliques in the graph, only K of them which will typically be a small number (e.g., 8–12) and are easily able to be found in our empirical applications, which our simulations show leads to high power. In Appendix K we show that taking $\ell < \log(n)$ is often sufficiently slow to guarantee that cliques of size ℓ exist in the network. Of course, one can take ℓ to grow slower than that, but a higher ℓ results in a lower Type 1 error, since larger cliques gives better estimates of the distances in the latent space. First, we would like to take the number of cliques K and the size of the cliques ℓ of cliques to both be as large as possible. In practice, we use the networkX command `enumerate_all_cliques` in Python, which uses a clique finding algorithm from Langston et al. (2005). While identifying numerous cliques can be challenging, given the modest size of the cliques and that we only need a small, fixed number of cliques, applied researchers can easily implement our technique.

6.1.2 Choosing cliques

As ℓ increases, the variance of estimates of $\hat{D}$ decreases, and the power of the test increases as K increases, since we have more distances between points on the manifold. Figure 4 from our simulations shows that as K increases, the power of our tests increase. Second, we need cliques that are well-separated on the manifold, but connected in the graph. Since we use cliques as 'points' on the manifold to measure distance, the cliques should ideally not have nodes in common, since if two cliques do overlap, its possible these two 'points' on the manifold are close together or even the same point. Third, if two cliques have no edges between them, then our estimate of the distance between the two points is $+\infty$, which contains no information about the geometry.

Motivated by these three considerations, our goal is to solve

$$\hat{C}_1, \ldots, \hat{C}_K \in \underset{C_1, \ldots, C_K}{\operatorname{argmin}} \sum_{i,j}^{K} |C_i \cap C_j| \tag{16}$$

such that $|C_i| = \ell$ for each i and $\hat{P}(C_1, \ldots, C_K)$ does not contains a 0.

In practice, we set K and ℓ by first looking at the number of cliques of various sizes in the graph and choosing and ℓ that is close to the size of the largest cliques in the graph, but where there are still enough cliques of that size to find K and are well-separated. We then take random draws from the (very large) set of possible cliques and evaluate the objective function in (16). Searching over the set of possible cliques is a well-studied (NP-hard) problem in computer science and graph theory, however, we found that our relatively simple approach yielded high-quality cliques after around 10^6 draws from the clique distribution. We evaluate the quality of the cliques we select by running the optimization independently several times. A stable objective function value across the runs indicates high-quality cliques. In the data from Banerjee et al. (2019), we take K as either 7 or 10. The value we choose is based on how easy it is to find appropriate cliques in a given network using the problem formulation in (16). In the Indian village sample, the average number of cliques of size $c_i - 1$ is 80, where c_i is the size of the largest clique in network i, known as the clique number.

It takes on average 0.005 s to find all cliques of size $c_i - 1$ over the 75 networks. It takes 0.004 s in the *C. elegans* sample to find the 29 cliques of size 5 in the network.

To select ℓ, we use the size of the largest clique found in the graph minus one. In most of the villages, choosing ℓ in this way resulted in dozens of possible cliques to choose from. We present more details about cliques in the Banerjee et al. (2019) data in Appendix F. For the *C. elegans* data, we select $K = 12$ and set $\ell = 5$, which is the size of the largest clique in the graph. We reiterate that for our approach we need only K cliques and do not need to enumerate all ℓ cliques in the graph.

6.2 Village risk-sharing networks and the introduction of microfinance

We begin by studying the underlying geometries of Indian village networks. We use the Wave II village network data of Banerjee et al. (2019), in part collected by one of the authors of the present paper. This consists of a collection of graphs for each of 75 villages in Karnataka, India constructed by surveying 89% of all households in each village, thereby generating a 99% edge sample for the resulting undirected graph. There are a total of 16,451 households in the sample. In every village, we have relationship data between households on each of 12 dimensions: 5 social dimensions, 4 financial dimensions, and 3 information sharing dimensions. See Banerjee et al. (2019) for more details including descriptive statistics. The links across these dimensions line up for the most part, consistent with a theory of multiplexed incentives to form links, so we study the undirected, unweighted graph following the prior literature using this data (Banerjee et al., 2013, 2019; Breza & Chandrasekhar, 2019; Jackson & Lopez-Pintado, 2013).

The social networks literature has long been interested in excess closure Coleman (1988). Friends of friends tend to be friends more than one might expect and this is particularly true if network relationships substitute for formal institutions. Literature focusing on equilibrium informal financial networks, which facilities the sharing of risk between households in a village, describes why the equilibrium network shapes exhibit excess closure (e.g., Ambrus et al., 2014; Jackson, 2013). The idea is that in order to maintain cooperation, when individuals can renege on their promises to aid each other in times of need, it is useful to have friends in common to amplify punishment, thereby maintaining a good equilibrium.

From the perspective of a latent space model, this means that we might expect excess closure in the village. There are incentives by households to 'curve' the space, so friends of friends and so on are much more likely to themselves link, discussed in greater depth below. A natural hypothesis, therefore, is that village networks for the most part not be hyperbolic. Rather, they may be more likely to be spherical or, perhaps, Euclidean.[9]

Our proposed method gives hypothesis tests (and corresponding p-values) for each of the three candidate geometries. As a descriptive summary, we 'classify' each of the villages into one of the geometry types using the following procedure. For villages where at least one village has a p-value over .05, we consider, for the purposes of summarizing our results, the manifold type that has the largest p-value. If all three geometries reject the null at the .05 level, then we say that the village cannot be classified. This outcome could mean a number of things, ranging from a false-rejection by chance to a village whose underlying geometry is not captured by one of the three candidates (e.g., curvature may be nonconstant). Figure 5, Panel A, presents classification results for the 75 villages using this descriptive approach. We see that we are able to classify 75% of the villages, despite the fact that N/A was a possibility. Classification was not forced. Further, the results are consistent with the socio-economic hypothesis on villages needing closure. 48% of the classified networks are spherical, 35% are Euclidean, and only 16% are hyperbolic.

---

[9]   We briefly note that common modelling assumptions in the socio-economic literature imply constant curvature from the perspective of our model (1), though certainly there are perspectives that would violate constant curvature which would require future work. To see this, consider two examples. First, imagine a model in which nodes have some random locations. They can choose their efforts to link and the value of their links depends on the number of their friends who are themselves friends in expectation. There is a parameter that governs the value of closure among one's friends which can be positive, zero, or negative, which may depend on the socio-economic context. In such a model, this parameter exactly maps to curvature. Second, one can imagine a model in which agents can take an action to influence the extent to which their neighbours know each other. For instance, the action could be imagined as throwing parties (selecting positive curvature) or the opposite and ensuring 'worlds do not collide' (selecting negative curvature) (David & Seinfeld, 1995). In such a model, if we study the symmetric equilibrium, then the equilibrium choice of the extent of forced socialization or barred socialization among one's friend exactly maps to constant curvature. Both of these examples also illustrate the limitations of such models. While these examples demonstrate how conventional assumptions map to constant curvature, certainly more complex models with heterogeneity would require modelling manifolds with non-constant curvature, which we leave to future work.

(a)

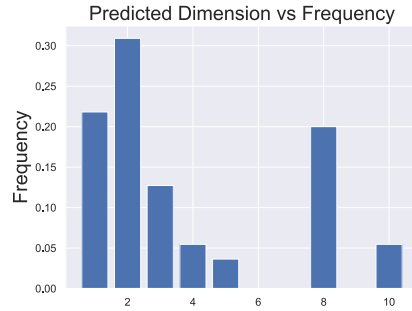| Geometry	$\mathbb{R}^p$	$\mathbf{S}^p$	$\mathbf{H}^p$	N/A
% Classification	26%	36%	12%	25%

(b)

Figure 5. Predicted geometries and dimensions for the Banerjee et al. (2019) village networks. (a) Geometry estimates. (b) Predicted dimensions.

We also examine the estimated dimension of the latent space. Figure 5 presents the estimated dimensions, which irrespective of curvature is important to know the minimal dimension of the space required to model location decisions by agents.

We now explore the relationship between latent geometry and socio-economic phenomena. This is an observational, not causal, analysis. First, we look at how the volume of informal financial transactions varies with network geometry. Specifically, we are interested in how the volume of informal loans that a household has with network neighbours (e.g., friends or members of their rotating, savings, and credit associations) varies with geometry. Both the theoretical and empirical economic literature suggest that it is ex ante ambiguous as to the relationship between the amount of network financial flows and curvature. For example, Kinnan and Townsend (2012) study how informal financial flows efficiently allocate credit to households that experience negative shocks in the network. Theory suggests that such flows are more efficient in more expansive networks, which require negative curvature (Ambrus et al., 2014). At the same time, as discussed above, the ability to facilitate informal financial transactions may increase in the importance of closure, and, therefore, require positive curvature. Which force dominates is an empirical question.

To study this, we estimate the following regression:

$$\text{Network Loan Amount}_i = \alpha + \beta_E \mathbf{1}\{\hat{\mathcal{M}}_i^{\hat{p}} = E\} + \beta_H \mathbf{1}\{\hat{\mathcal{M}}_i^{\hat{p}} = H\} + \beta_N \mathbf{1}\{\text{N/A}_i\} + \epsilon_i$$

where i indexes the village. Network Loan Amount$_i$ is the average volume of loans from either friends or rotating savings and credit association members that a household has in the village. The loan amount is presented in INR (USD 1 ≈ INR 73.5). Here, the omitted category (α) corresponds to the loan amount for a sphere.

The leftmost panel of Figure 6 presents the results. We find that a Euclidean village relative to a spherical one has INR 3,940 or 24% ($p = 0.098$) more informal network loans. Further decreasing curvature, we compare hyperbolic villages to spherical ones and find that hyperbolic villages have INR 5,865 or 35% ($p = 0.034$) more in informal network loans. These increases are extremely large in real economic terms: the difference in credit between the hyperbolic and spherical geometries corresponds to an individual in the hyperbolic geometry receiving additional credit worth 20 days of wages. Taken together, we have seen greater financial flows precisely in geometries that permit more expansive network topologies.

Second, having studied how informal financial transaction patterns are associated with geometry, we now turn to studying determinants of geometry. Our primary interest is in whether the introduction of a formal credit market (microfinance) to a setting otherwise dominated by only informal financial transactions changes the network structure by changing the geometry. In our setting, as described below, microcredit was introduced to only some villages, allowing us to compare the impact of access to microfinance on network structure.
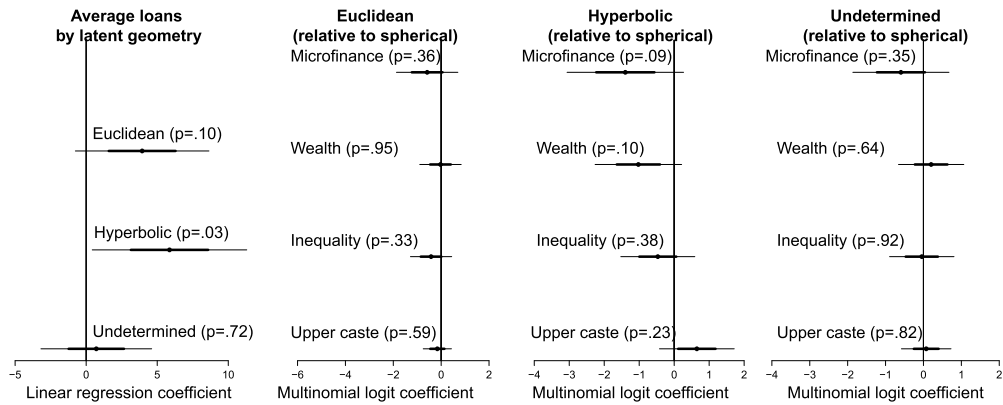
Figure 6. Regression coefficients showing the determinants of geometry. In the left figure, each line in the plot corresponds to the coefficient in a multivariate linear regression where the outcome is the average amount of loans (in thousands of INR) and the predictors are geometry types (with spherical as the reference). The wide bars correspond to one standard error and the narrow bars represent two standard errors. The reference value for spherical is 16.71 (again in thousands of INR). Plots 2–4 show the coefficients from a multinomial logistic regression where the outcome is the predicted geometry type for each village. Each panel shows all coefficients for a particular geometry (with spherical as the reference). Each line in the plot corresponds to an estimated coefficient. The wide bars correspond to one standard error and the narrow bars represent two standard errors. The constant values for the Euclidean, hyperbolic, and undetermined comparisons are 0.005, −0.69, and −0.01, respectively.

In addition to microcredit access, we focus on three other determinants: wealth, inequality, and caste fractionalization. It is ex ante not obvious as to how any of these might correlate with geometry and is, therefore, an important empirical question. For example, wealthier villages may have a reduced need to sustain informal insurance—their worst-case scenario is better off than their poorer counterparts—and as a consequence may require less positive curvature. Or, in contrast, wealthier villages may be able to take on greater entrepreneurial risk as they can sustain losses, and such endeavours require group cooperation and, therefore, closure. Similarly, within-village wealth inequality can change incentives for triadic closure, as can ethnic fractionalization (Currarini et al., 2009). Ultimately, the empirical correlations are of interest. The most important relationship to study is how the introduction of formal credit to villages that otherwise used informal network transactions affects geometry. From 2007, a microfinance institution entered 43 of the 75 villages studied here and the network data we utilize is taken after the intervention (Banerjee et al., 2020, 2013). This allows us to study the effect of the introduction of microcredit on network geometry as a way to understand whether credit access differentially changes the need for ones' friends to maintain relationships with each other. Note that this is different from clustering or other measures of closure per se, which are also affected by the locations z and fixed effects v. So we can specifically address that, all things being equal, whether the demand for one's friends to themselves be linked increases, decreases, or is unchanged when the village now has access to formal financial instruments. It is a priori not obvious. On the one hand, the new credit opportunity may encourage re-lending or joint business ventures among clients of microcredit, increasing the need for closure and generating positive curvature. On the other hand, the new credit opportunity may reduce reliance on informal financial relationships with others in the village and push towards negative curvature. In either case, the answer as to how a large credit intervention may affect geometry is of empirical interest.

To study the determinants of geometry, we estimate a multinomial regression:

$$\frac{\mathbb{P}(\hat{\mathcal{M}}_i^{\hat{p}} = m)}{\mathbb{P}(\hat{\mathcal{M}}_i^{\hat{p}} = \mathbf{S})} = \exp\left(\delta_m + \beta_{\text{MFI}}^m \text{MFI}_i + \beta_{\text{W}}^m \text{Wealth}_i + \beta_{\text{I}}^m \text{Inequality}_i + \beta_{\text{F}}^m \text{Frac}_i\right)$$

where $m \in \{E, \mathbf{H}, \text{N/A}\}$. MFI_i denotes whether the microfinance institution entered village i. Wealth_i denotes a wealth index measure.[10] Inequality_i is within-village standard deviation of

---

[10] The Banerjee et al. (2019) dataset does not have consumption nor expenditure measures. So we utilize the score constructed from the first principle component of a number of household features that correlate with wealth in the village.

wealth.[11] Finally, $\text{Frac}_i = \alpha_U(1 - \alpha_U)$ where α_U is the share of households that are of upper caste. The score is zero if society is perfectly homogenous and 1/4 for an even split. The right three panels of Figure 6 present the results. We begin by looking at microfinance. We estimate $\hat{\beta}^{\text{H}}_{\text{MFI}} = -1.40$ ($p = 0.093$). This means that a village receiving microcredit is associated with an 8.8% decline in the probability of being hyperbolic relative to spherical. In other work, (Banerjee et al., 2020), we have shown that introducing microcredit has decreased density and also the number of triads in the network. Our analysis here demonstrates that the fundamental value of having friends in common itself *increased* suggesting that the effects documented in our prior work came from shifts in node locations (z_i) and efforts of socializing (v_i) in the latent space rather than changes in the relative value of closure which appears to have increased.

We also find that wealthier villages are less likely to be hyperbolic relative to spherical. We estimate $\hat{\beta}^{\text{H}}_{\text{W}} = -1.02$ ($p = 0.098$). This corresponds to a 8.4% decline in the relative probability of being hyperbolic as compared to spherical. We do not find any significant relationship between wealth inequality nor caste fractionalization and geometry.

Taken together, we have shown the empirical content of the estimation of the latent geometry. We can classify the vast majority of villages (despite allowing for N/A) and they are predominantly spherical. We find informal financial loans are higher in villages that exhibit negative curvature. Finally, and importantly, villages where microcredit was introduced tend to have a more spherical structure. This can perhaps be interpreted as showing that access to microcredit generates, *ceteris paribus*, demand for greater triadic closure.

6.3 Network of neurons

Our second setting looks at a network of neurons. There is neuroscience literature that is interested in documenting regularities in network structure as well as modelling network structure through statistical network formation models.

The first strand of the literature looks at how patterns of the graph of neurons relate to neurological mechanisms (Karwowski et al., 2019). For instance, these networks exhibit short path lengths—disparate regions of the human brain are connected by a few steps. Further, the degree distribution reflects thick tails: certain nodes have numerous connections. Moreover, the network is dynamic: early in age the network exhibits high amounts of homophily whereas as the individual ages this declines.

The second strand develops low-dimensional statistical representations of the neural networks since this allows for interpretability, counterfactuals, and deals with the fact that otherwise there is a litany of statistics that can be used to correlate with biological outcomes without any interpretable control (de Lange et al., 2014; Recanatesi et al., 2019). To this end, conditional edge independence models, scale-free models, block models, and latent space models have been explored (Karwowski et al., 2019; VanRullen & Reddy, 2019).

Third, and particularly relevant for latent space models, is the concept of the functional graph of neurons rather than the structural graph of neurons (Abdelnour et al., 2018; Petersen & Sporns, 2015). The idea is that while a graph can be drawn of the physical links between all nodes, predominantly the graph that is able to be activated—the functional network—is a network that is distinct. Much like individuals who reside in geographic space but functionally interact in a network that can be thought of as in a latent space, the functional network perspective presents an opportunity leverage latent space models.

Our specific application is to a network of neurons of *Caenorhadbitis elegans*, which are soil-dwelling roundworms. There is a long history of using *C. elegans* as a model organism for studying the nervous systems of animals. In fact neurons of *C. elegans* are extremely similar to that of humans (Leung et al., 2008). For our example, we use the *C. elegans* neuron data of Kaiser and Hilgetag (2006), which has been used a number of times in order to model neural network structure. There are several goals in modelling neural network structure. The relative location distribution, how distance affects linking rates, and the geometry all inform how signals could

This consists of access to private electricity, home ownership, quality of roofing material, and number of rooms in the household.

[11] Specifically, we take the score from the first principle component of the within-village standard deviation of each of the constituent wealth measures.

be passed across nodes. Moreover, though beyond the scope of our knowledge, there may be interpretations to the distribution of fixed effects—latent heterogeneity in the propensity for certain neurons to systematically link to others.

A priori it is unclear what the right latent geometry ought to be. For instance, if the network of neurons ought to have a high degree of expansiveness, it ought to be embedded in hyperbolic space. In contrast, if it ought to reflect strong, localized redundancies, or a high degree of homophily it may be better modelled as being embedded in a spherical geometry.

The dataset contains a neural network from a single *C. elegans*, consisting of a connected graph of 131 neurons, with 764 edges, and a clustering coefficient of 0.245. The clique number of this graph is 6, but it has only one clique of size 6, but it has 29 cliques of size 5, so we use $\ell = 5$. We find $K = 12$ cliques using the problem formulation in (16) and then take a maximally disjoint clique set which is sufficient for our test. We use Algorithm 2 and compute the p values for the Euclidean, spherical, and hyperbolic geometries and find these values are: $p_E = 0.378$, $p_S = 0.05$, and $p_H = 0.267$, so we reject the spherical hypothesis (noting that we can do so despite there being a high level of clustering). The *C. elegans* network of neurons, therefore, is inconsistent with a latent space with positive curvature, where neurons are excessively likely to exhibit triadic closure relative a flat benchmark. We can only say that there is no or negative curvature, but the data are not sufficiently powered to allow us to distinguish this.

7 Conclusion

Latent space models are widely used in network analysis across numerous disciplines, including, but not limited to sociology, economics, biology, and computer science. The predominant approach is to assume a Euclidean latent space, though there is current discussion about adopting a hyperbolic space in certain contexts. Nonetheless, the current methods employed do not provide a way to estimate the geometry itself. Unfortunately, incorrect embedding spaces can deliver misleading results and while there may be convergence to pseudo-true values, counterfactual analysis will be affected.

We develop methods that the researcher can apply in order to consistently estimate the latent space geometry from network data. Our core observation is that the observed network data encodes information on the distance between nodes in latent space. That is, a finite sample network corresponds to a noisy set of distances. So we transform our network problem to a statistical geometry problem.

In our first result, we study a more general problem: whether an observed estimate of a distance matrix among K points contains enough information to consistently estimate the unobserved manifold in which the K nodes can be isometrically embedded. We answer this in the affirmative: the spectrum of a distance matrix encodes the manifold's metric and, therefore, the manifold class, rank, and curvature. Leveraging results on eigenvalue perturbations, we prove the result. Our second result applies to the network setting. By looking at cross-clique link frequencies, one can construct a noisy distance matrix and therefore estimate the latent manifold consistently.

An important advantage of our approach is that, unlike other strategies, we do not need to estimate the fixed effects or the locations in a candidate manifold (nor integrate them out) in the estimation procedure. Instead, by focusing on a strategy that exploits the fundamentals of geometry, we directly check isometric embeddings, so we can estimate the geometry without ever estimating the numerous other parameters and only move to them after having estimated the geometry.

We also demonstrate the empirical content of estimating the latent geometry which is novel in the literature. Strikingly, even though N/A is a possibility, we were able to classify (75%) of villages, indicating the empirical relevance of our methods. Further, consistent with theory, we show Indian risk-sharing villages are often spherical. Additionally, villages that are more expansive are associated with a greater flow of informal financial loans through the network. Finally, the introduction of microcredit is associated with a shift to positive curvature: the relative value of having triadic closure increases when villages have access to formal credit.

A number of future steps come to mind. While our assumptions on geometry—that it is a simply connected, complete Riemannian manifold—are parsimonious and natural, they are also limited. They nest the current assumptions in the literature (we know of no empirical research that assumes a torus of genus two for instance in the networks literature) but they are still admittedly lacking.

Z"ouboulogloou et al. (2021) have shown a relationship between the torus and the sphere that might allow us to apply our current spherical methodology to the torus. We speculate that there may be strategies to use local structures in the network to patch together some more global structure. That is, for instance, if it can be arranged into a pseudo-block diagonal structure, perhaps in each block, there is room for a different geometry and then these can be stitched together. See Gu et al. (2019), among others, for related work on this topic. Additionally, extending our results to settings where the full graph is not observed, such as Aggregated Relational Data (Breza et al., 2020; McCormick & Zheng, 2015), would allow researchers who do not have resources to collect data on all edges to leverage insights about underlying geometry. Individual node covariates could also be leveraged to form trait groups in settings without complete network data. Finally, an interesting area of future research involves exploring how to optimally combine the results from multiple tests of the three geometries, where each test is computed using a different set of cliques. Of course, the tests based on different cliques would likely be correlated, and so an important question would be to understand how to use the correlation to increase the power of the combined test.

Acknowledgments

Funding

Appendix A. Proofs

A.1 Proofs required for Theorem 1.1 (Section 3)

Proof of Proposition 3.1. We only prove this claim for the Euclidean case, but the same argument proves the claim for the other two geometries. We have by Weyls's inequality that $|\lambda_1(\hat{W}_0) - \lambda_1(W_0)| \le \|\hat{W}_0 - W_0\|_F$, where $\|A\|_F$ is the Frobenius norm of A, that is $\|A\|_F^2 = \sum_{l,l'} a_{ll'}^2$. Then, we have that $\mathbb{P}(|\lambda_1(\hat{W}_0) - \lambda_1(W_0)| < \theta) \le \mathbb{P}(\|\hat{W}_0 - W_0\|_F < \theta)$ for all θ. Under $H_{0,e}$, $\lambda_1(W_0) = 0$, so we have that $\mathbb{P}(|\lambda_1(\hat{W}_0)| < \theta) \le \mathbb{P}(\|\hat{W}_0 - W_0\|_F < \theta)$. By setting θ to be the α quantile of $\|\hat{W}_0 - W_0\|_F$, we conclude (3). This completes the proof. □

Proof of Proposition 3.2. We prove this proposition for the spherical case. The hyperbolic case follows from a similar argument. By Theorem 2.1 in Newey and McFadden (1994), we have consistency if (i) the limit objective function is uniquely maximized at the truth, (ii) the parameter space is compact, (iii) the limit objective function is continuous in the parameter, and (iv) there is uniform convergence of the empirical objective function to its limit. The latter holds if there is point-wise convergence and stochastic

equicontinuity. The parameter space is compact and since under the null $W_\kappa(D)$ is positive semi-definite, the minimum eigenvalue is 0 as long as $K > p$. Identification comes from continuity of eigenvalues in parameters of the matrix. Finally we check uniform convergence. First, note by hypothesis that $\hat{D} \overset{p}{\to} D$ as $T \to \infty$. Since eigenvalues are continuous functions of their matrix arguments, we have by the continuous mapping theorem that $\lambda_1(\kappa W_\kappa(\hat{D})) \overset{p}{\to} \lambda_1(\kappa W_\kappa(D))$ for every $\kappa \in [a, b]$, and so we have pointwise convergence. To complete the proof, we will show stochastic equicontinuity to show uniform convergence. A sufficient condition is a Lipschitz condition (Lemma 2.9, Newey & McFadden, 1994): that for any κ_1, κ_2, $|\lambda_1(\kappa_1 W_{\kappa_1}(\hat{D})) - \lambda_1(\kappa_2 W_{\kappa_2}(\hat{D}))| \le B_T |\kappa_1 - \kappa_2|$ for some random variable $B_T = O_p(1)$. To do this, fix any $\kappa_1, \kappa_2 \in [a, b]$. By Weyl's inequality,

$$\left| \lambda_1(\kappa_1 W_{\kappa_1}(\hat{D})) - \lambda_1(\kappa_2 W_{\kappa_2}(\hat{D})) \right| \le \| \kappa_1 W_{\kappa_1}(\hat{D}) - \kappa_2 W_{\kappa_2}(\hat{D}) \|_F.$$

Since $\kappa W_\kappa(D) = \cos(\sqrt{\kappa}D)$ and $\cos(\cdot)$ is Lipschitz continuous with Lipschitz constant 1, we have for each l, l',

$$\left| \cos(\kappa_1^{1/2}\hat{d}_{l,l'}) - \cos(\kappa_2^{1/2}\hat{d}_{l,l'}) \right| \le \hat{d}_{l,l'} \cdot \left| \kappa_1^{1/2} - \kappa_2^{1/2} \right|.$$

For $\kappa_1, \kappa_2 \in [a, b]$,

$$|\sqrt{\kappa_1} - \sqrt{\kappa_2}| = \left| \frac{\kappa_1 - \kappa_2}{\sqrt{\kappa_1} + \sqrt{\kappa_2}} \right| \le \frac{1}{2\sqrt{a}} |\kappa_1 - \kappa_2|,$$

so for any $\hat{d}_{i,j}$,

$$\left| \cos(\kappa_1^{1/2}\hat{d}_{i,j}) - \cos(\kappa_2^{1/2}\hat{d}_{i,j}) \right| \le \frac{\hat{d}_{i,j}}{2a^{1/2}} |\kappa_1 - \kappa_2|.$$

Putting this all together, we see that

$$\left| \lambda_1(\kappa_1 W_{\kappa_1}(\hat{D})) - \lambda_1(\kappa_2 W_{\kappa_2}(\hat{D})) \right| \le \sqrt{\sum_{i,j} (\kappa_1 W_{\kappa_1}(\hat{D}) - \kappa_2 W_{\kappa_2}(\hat{D}))^2_{i,j}}$$

$$\le \sqrt{\sum_{i,j} \left(\frac{\hat{d}_{i,j}}{2a^{3/2}} |\kappa_1 - \kappa_2| \right)^2}$$

$$= \sqrt{\sum_{i,j} \left(\frac{\hat{d}_{i,j}}{2a^{3/2}} \right)^2} |\kappa_1 - \kappa_2|.$$

Since $\sqrt{\sum_{i,j} (\hat{d}_{i,j}/2a^{3/2})^2} = O_p(1)$, the desired Lipschitz condition holds, which completes the proof. The hyperbolic case is handled in a similar way.   □

Proof of Proposition 3.3. We prove the Euclidean case (part a) and note that the proofs of parts b and c (spherical and hyperbolic) are nearly identical. Define $\mathcal{R}_T = (-\infty, \delta_T]$. Let $\mathbb{P}_0(A)$ denote the probability of event A under the null hypothesis that $\mathcal{M}^p(\kappa)$ is Euclidean. By (8),

$$\mathbb{P}_0(\lambda_1(\hat{W}_0) \in \mathcal{R}_T) = \mathbb{P}_0(\lambda_1(\hat{W}_0) \le \delta_T) = o(1),$$

by assumption. Under H_1, $\lambda_1(W_0) < 0$ by Lemma 2.1. Since $\delta_T = o_P(1)$,

$$\mathbb{P}_1(\lambda_1(\hat{W}_0) \in \mathcal{R}_T) = \mathbb{P}(\lambda_1(\hat{W}_0) \le \delta_T)$$
$$= 1 - \mathbb{P}(\lambda_1(\hat{W}_0) \ge \delta_T) = 1 - o(1).$$

This proves that the test for (4) is consistent, as claimed. $\square$

Proof of Proposition 3.5. For each index $j = 1, \ldots, K$, we consider two different cases. In case 1, $\lambda_j(W_\kappa) \ne 0$. In this case, know that since $\epsilon_T \xrightarrow{p} 0$, $\mathbb{P}(\lambda_j(\hat{W}_{\hat{\kappa}}) \in \mathcal{R}) \to 0$. Here we just use the fact that $\lambda_j(\hat{W}_{\hat{\kappa}}) \to c \ne 0$, so that eventually this eigenvalue is outside the rejection region. Note that this calculation does not require a particular rate on ϵ; we just need ϵ to go to zero in probability. We now handle case 2, in which $\lambda_j(W_\kappa) = 0$. This is the more subtle case. By definition of the rejection region,

$$\mathbb{P}_1(\lambda_j(\hat{W}_{\hat{\kappa}}) \in \mathcal{R}) = \mathbb{P}_1(|\lambda_j(\hat{W}_{\hat{\kappa}})| \le \epsilon_T).$$

where the subscript here indicates that the alternative hypothesis that $\lambda_j \ne 0$ is true. By Weyl's inequality, the above probability then becomes

$$\mathbb{P}_1(|\lambda_j(\hat{W}_{\hat{\kappa}})| \le \epsilon_T) \le \mathbb{P}_1(\|\hat{W}_{\hat{\kappa}} - W_\kappa\|^2 \le \epsilon_T^2) := \mathbb{P}_1(r_T \le \epsilon_T).$$

By assumption, we know that $r_T/\epsilon_T \to 0$ in probability. Therefore, for any eigenvalue that is actually zero, for large enough T we will call this eigenvalue zero and it won't count in the estimated rank. This concludes the proof. $\square$

Proof of Theorem 1.1. By assumption, we know that $\hat{D} \xrightarrow{p} D$, so by Proposition 3.2, we have that $\hat{\kappa} \xrightarrow{p} \kappa$. We will use Proposition 3.3 to argue that $\widehat{\mathcal{M}^{\hat{p}}}$ is consistent for $\mathcal{M}^p(\kappa)$. To do this, note that if $\mathcal{M}^p(\kappa)$ is Euclidean, then by Proposition 3.3, $\widehat{\mathcal{M}^{\hat{p}}}$ is consistent. To prove the claim for the spherical case, recall that we define $\phi(\hat{W}_0) = 1$ to mean that we reject the hypothesis that $\mathcal{M}^{p^\star}(\kappa^\star)$ is Euclidean. If $\phi(\hat{W}_0) = 0$ then we fail to reject the hypothesis that $\mathcal{M}^{p^\star}(\kappa^\star)$ is Euclidean. Similar definitions hold for the spherical and hyperbolic cases.

If $\mathcal{M}^{p^\star}(\kappa^\star)$ is spherical, then we have that

$$\mathbb{P}_S\left(\widehat{\mathcal{M}^{\hat{p}}} = S^p(\kappa)\right) = \mathbb{P}_S(\phi(\hat{W}_0) = 1, \ \phi(\hat{W}_{\hat{\kappa}}) = 0)$$
$$= \mathbb{P}_S(\phi(\hat{W}_0) = 1)\mathbb{P}_S(\phi(\hat{W}_{\hat{\kappa}}) = 0)$$
$$\to 1,$$

where the notation $\mathbb{P}_S$ indicates that $\mathcal{M}^{p^\star}(\kappa^\star) = \mathbf{S}^p(\kappa)$ and the third line follows from Proposition 3.3. A similar argument proves that $\widehat{\mathcal{M}^{\hat{p}}}$ is consistent when $\mathcal{M}^{p^\star}(\kappa^\star)$ is hyperbolic. Therefore, we can conclude that $\hat{p}$ is consistent for the true rank of W_κ. This completes the proof. $\quad\square$

A.2 Proofs required for Theorem 1.2 and Section 4

Proof of Theorem 1.2.   In order to show that estimates of distances computed using cliques are consistent, we recall the form of the estimates from (14),

$$d(z_i^\star, z_j^\star) = -\log(p_{ij}) + \log(\gamma),$$

where $\gamma := E\{\exp(v)\}^2$ and $p_{ij} = \mathbb{P}(G_{ij} = 1 \mid z_i^\star, z_j^\star)$. We estimate $d(z_i^\star, z_j^\star)$ with

$$\hat{d}(z_i, z_j) = -\log(\hat{p}_{ij}) + \log(\hat{\gamma}).$$

Under the assumptions of Theorem 1.2, we have a consistent estimate $\hat{\gamma} \to \gamma$, so we only focus on estimating the term p_{ij}. We estimate this term using

$$\hat{p}_{kk'} = \ell^{-2} \sum_{i \in C_k(\ell)} \sum_{j \in C_{k'}(\ell)} G_{ij},$$

where $C_k(\ell)$ is a clique of size ℓ and $C_{k'}(\ell)$ is another clique of size ℓ. Suppose that each node in a clique is at the same location, say ζ_k for clique k. Then, $\hat{p}_{kk'}$ is a consistent estimate of p_{ij}, the probability that node i at location z_k connects to node j at location $z_{k'}$. In practice, the locations in cliques do not fall at exactly the same location, but as $\ell \to \infty$, under Assumption 1.3, we do know that $\max_{ij} d(z_i^\star, z_j^\star) \xrightarrow{p} 0$ for nodes i and j in any clique. Since the event that all nodes in a clique fall within δ of each other, for any $\delta > 0$, occurs with probability going to 1, we can condition on the event that nodes in a clique are at the same location. By the preceding argument, we can then conclude that $\hat{p}_{kk'} - p_{ij} \xrightarrow{p} 0$. By the continuous mapping theorem, we can then consistently estimate d_{ij}. We can now apply Theorem 1.1. This completes the proof. $\quad\square$

Proof of Proposition 4.1.   Before providing the proof, we provide a brief outline of our strategy. We will suppose that each $v_i^\star = 0$ to simplify notation, but the general result claimed in the Proposition holds. Our goal in this proof is to show that $\mathbb{P}(\mathcal{E}_\delta \mid \text{clique}) \to 1$ as $\ell \to \infty$. To make the proof easier, we will equivalently show that the ratio

$$\frac{\mathbb{P}(\mathcal{E}_\delta \mid clique)}{\mathbb{P}(\mathcal{E}_\delta^c \mid clique)} \to \infty.$$

Showing that this ratio goes to infinity shows the numerator goes to 1 since $x/(1-x) \to \infty$ if and only if $x \to 1$. We now turn to the proof.

We have

$$\frac{\mathbb{P}\big(\mathcal{E}_\delta \mid clique\big)}{\mathbb{P}\big(\mathcal{E}_\delta^c \mid clique\big)} = \frac{\mathbb{P}\big(clique \mid \mathcal{E}_\delta\big)}{\mathbb{P}\big(clique \mid \mathcal{E}_\delta^c\big)} \cdot \frac{\mathbb{P}(\mathcal{E}_\delta)}{1 - \mathbb{P}(\mathcal{E}_\delta)}$$

$$\geq \frac{\mathbb{P}\big(clique \mid \mathcal{E}_\delta\big)}{\mathbb{P}\big(clique \mid \mathcal{E}_\delta^c\big)} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1))$$

for some positive constant $a(\delta)$ by Assumption 4.2.

Next, since we have $L := \binom{\ell}{2}$ possible links and δ is the maximal distance between any two nodes,

$$\mathbb{P}\big(clique \mid \mathcal{E}_\delta\big) \geq \exp(-L\delta),$$

so we have

$$\frac{\mathbb{P}\big(\mathcal{E}_\delta \mid clique\big)}{\mathbb{P}\big(\mathcal{E}_\delta^c \mid clique\big)} \geq \frac{\exp(-L\delta)}{\mathbb{P}\big(clique \mid \mathcal{E}_\delta^c\big)} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1)).$$

For bounded support, $\mu_d = \mu_{d,n} := E\{d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)\}$ is finite. Thus, by Lemma A.1, we have that

$$\mathbb{P}\big(clique \mid \mathcal{E}_\delta^c\big) \leq \exp\big(-L\mu_d\big).$$

Then, substituting this upper bound on $\mathbb{P}(clique \mid \mathcal{E}_\delta^c)$ into the ratio of interest, we have

$$\frac{\mathbb{P}\big(\mathcal{E}_\delta \mid clique\big)}{\mathbb{P}\big(\mathcal{E}_\delta^c \mid clique\big)} \geq \frac{\exp(-L\delta)}{\exp\big(-L\mu_d\big)} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1))$$

$$= \exp\big\{-L\big(\delta - \mu_d\big)\big\} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1)).$$

We can rewrite this as

$$\exp\big\{L\mu_d - L\delta\big\}$$
$$\times \exp\big\{\log a(\delta) + \ell \log A_n^{-1} - \log\big(1 - A_n^{-\ell}\big) + o(1)\big\} \quad \text{(A1)}$$

and now need to show that this lower bound goes to infinity. We do this by appealing to Assumption 4.3. Specifically, by Assumption 4.3, the following condition is true for sufficiently large n

$$L\mu_d - L\delta \geq \ell \log A_n - \log\big(1 - A_n^\ell\big) + \log a(\delta).$$

Substituting the above expression into the lower bound in (A1),

we have

$$\frac{\mathbb{P}\big(\mathcal{E}_\delta \mid clique\big)}{\mathbb{P}\big(\mathcal{E}_\delta^c \mid clique\big)} \geq \exp\big\{L\mu_d - L\delta\big\} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1))$$
$$\to \infty,$$

completing the proof in the case where each Ω_n is bounded.

To handle the case where Ω_n is not bounded, such as when F is a Gaussian distribution on $\mathbb{R}^p$, define an event

$$\mathcal{F}_n := \big\{z_i \in \Omega_n, \text{ for all } i = 1, \ldots, n\big\}.$$

In words, $\mathcal{F}_n$ holds whenever all z_i are in the set Ω_n. Recall that under Assumption 4.1(1), $\mathbb{P}(\mathcal{F}_n) = 1$, and under Assumption 4.1(2) there is an exponentially thin tail. This allows us to conclude that

$$\mathbb{P}(clique \mid \mathcal{E}_\delta^c) = \mathbb{P}(clique \mid \mathcal{E}_\delta^c, \mathcal{F}_n)\mathbb{P}(\mathcal{F}_n) + \mathbb{P}(clique \mid \mathcal{E}_\delta^c, \mathcal{F}_n^c)\mathbb{P}(\mathcal{F}_n^c)$$
$$\leq \mathbb{P}(clique \mid \mathcal{E}_\delta^c, \mathcal{F}_n)\mathbb{P}(\mathcal{F}_n) + \mathbb{P}(\mathcal{F}_n^c)$$

Since the z_i are i.i.d., we also know that $P(F_n^c) = P(z_i \notin \Omega_n)^n := p(n)^n$. Combining this, we have that

$$\mathbb{P}(clique \mid \mathcal{E}_\delta^c) \leq \exp(-\mu_d L) + p(n)^n$$

Supposing that $p(n) = \exp(-bA_n^{1/p^\star})$, where again $p^\star$ is the dimension of the latent space, we see that

$$\frac{\mathbb{P}\big(\mathcal{E}_\delta \mid clique\big)}{\mathbb{P}\big(\mathcal{E}_\delta^c \mid clique\big)} \geq \frac{\exp(-L\delta)}{\exp(-L\mu_d) + p(n)^n} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1))$$
$$= \frac{\exp(-L\delta)}{\exp(-L\mu_d) + \exp(-nbA_n^{1/p^\star})} \times a(\delta) \cdot \frac{1/A_n^\ell}{1 - 1/A_n^\ell}(1 + o(1))$$

Now, if we can show that $\exp(-nbA_n^{1/p^\star}) \to 0$ is negligible in the limit, then the above inequality reduces to the inequality for the bounded case, for which we proved that taking ℓ to grow faster than the term in this Proposition is sufficient. Since $n >> \ell$, the term $\exp(-nbA_n^{1/p^\star})$ is negligible, which allows us to re-use the proof of the bounded case. This completes the proof. $\square$

Lemma A.1 Consider any distribution of locations from which z_i are drawn i.i.d. on $\mathcal{M}^{p^\star}(\kappa^\star)$ with finite expected distance, $\mu_d := E\{d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)\} < \infty$. Then, if again $L = \binom{\ell}{2}$,

$$E\left[\prod_{i<j}\exp\Big\{-d_{\mathcal{M}^{p^\star}(\kappa^\star)}\big(z_i, z_j\big)\Big\}\bigg|\mathcal{E}_\delta^c\right] \leq \exp(-L\mu_d).$$

Proof. We have

$$
\begin{aligned}
E\left[\prod_{i<j} \exp\left\{-d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)\right\} \middle| \mathcal{E}_\delta^c\right] &\le E\left[\prod_{i<j} \exp\left\{-d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)\right\}\right] \\
&\le \prod_{i<j}\left(E\left[\exp\left\{-L \cdot d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)\right\}\right]\right)^{1/L} \\
&\le \prod_{i<j}\left(E\left[\exp\{-L \cdot x_{ij}\}\right]\right)^{1/L} \\
&\le \exp(-L\mu_d) \times \prod_{i<j}\left(E\left[\exp\left\{-L \cdot \eta_{ij}\right\}\right]\right)^{1/L} \\
&\le \exp(-L\mu_d) \times 1
\end{aligned}
$$

where we (a) unconditioned on the event since the probabilities of linking are higher within $\mathcal{E}$ than $\mathcal{E}_\delta^c$, (b) used Holder's generalized inequality, (c) used $\exp(a)^L = \exp(La)$, (d) defined $x_{ij} := d_{\mathcal{M}^{p^\star}(\kappa^\star)}(z_i, z_j)$, (e) decomposed $x_{ij} = \mu_d + \eta_{ij}$ where $\eta_{ij} := x_{ij} - \mu_d$, and (f) used the boundedness of linking probabilities. □

Proof of Example 6. Let $\Omega = [0, B^{1/p}]^p \subset \mathbb{R}^p$ so $\mathrm{vol}(\Omega) = B$ with $B = B_n$. Assume there are $C = C_n$ communities, each with m nodes, distributed uniformly at random in $\Omega' = [0, b^{1/p}]^p \subset \Omega$ with $|B^{1/p} - b^{1/p}| =: t$ and $t = \omega(\sqrt{\log n})$. That is, the community centres—not members necessarily—reside in a subset within the space of interest with a distance between the boundaries of at least $\sqrt{\log n}$. The extra factor controls for tail events.

Given these community centres ζ_c, we have nodes distributed

$$
z_i \sim F_c(\zeta_c, \sigma_c^2)
$$

where F is a Gaussian distribution on $\mathbb{R}^p$ centred at ζ_c with variance σ_c^2.

Note that if $\sigma_c^2 = 0$ then we have an example of an inhomogenous lattice, and with $B = C$ this operates like Example 4 exactly. On the other hand, if $\sigma_c^2 \to \infty$ and we restrict attention only to Ω itself, then we return to the uniform case. In between lies the case of multimodal location distributions with dispersion, governed by community centres. We will identify similar rates, mildly adjusted for tail events of extreme community or individual locations, though the bounds are not tight.

Define an event

$$
\mathcal{F} := \{z_i \in \Omega, \text{ for all } i = 1, \ldots, n\}
$$

and observe that the calculations conditional on $\mathcal{F}$ are identical to the lattice and point process cases. We compute

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_\delta^c | clique) &= \mathbb{P}(\mathcal{E}_\delta^c | clique, \mathcal{F})\mathbb{P}(\mathcal{F}) + \mathbb{P}(\mathcal{E}_\delta^c | clique, \mathcal{F}^c)\mathbb{P}(\mathcal{F}^c) \\
&\le \mathbb{P}(\mathcal{E}_\delta^c | clique, \mathcal{F})\mathbb{P}(\mathcal{F}) + \mathbb{P}(\mathcal{F}^c) \\
&\le \exp(-L\mu_d)(1 + o(1))
\end{aligned}
$$

where μ_d is the expected distance between two points distributed in Ω from the mixture model.

To bound $\mathbb{P}(\mathcal{F}^c)$, observe that

$$\mathbb{P}\left(\max_i \mathrm{dist}(z_i, \zeta_c) > t\right) \le \exp(-\mathrm{const.} \times t^2 + \log(n)) \to 0$$

by the sub-Gaussian distribution of the distance function for normals (the folded normal is sub-Gaussian), the growth assumption on t, that it holds for all communities simultaneously (which is not tight since most nodes will not be within the t-shell of the boundary due to the slow expansion of Euclidean balls). Below we calculate $t > \mathrm{const.} \times \sqrt{LB^{1/p}\sqrt{p}}$, from which the result follows.

By the application of Lemma A.1, and a calculation of the expected distance which is $c_2 B^{1/p}\sqrt{p}$,

$$\mathbb{P}(E_\delta^c | clique, \mathcal{F}) \le \exp(-c_2 LB^{1/p}\sqrt{p}),$$

and so

$$\frac{\mathbb{P}(\mathcal{E}_\delta | clique)}{\mathbb{P}(\mathcal{E}_\delta^c | clique)} = \exp(c_2 LB^{1/p}\sqrt{p} - L\delta) \cdot \frac{\mathbb{P}(E_\delta)}{1 - \mathbb{P}(E_\delta)}.$$

We, therefore, have

$$\ell \ge a(\delta) \frac{\log\left(\dfrac{B}{c_3 \delta^p} - 1\right)}{B^{1/p}\sqrt{p} - \delta}(1 + o(1)) + 1$$

which gives the growth-rate bound on the clique size. Now recall the restriction on the growth rate of B relative to b. If $b = \alpha B$ for some $\alpha < 1$, then

$$t = B^{1/p} - b^{1/p} = B^{1/p}(1 - \alpha^{1/p}) = \Theta(B_n^{1/p})$$

and so the restriction is immediate if for instance $B_n = \omega([\log n]^{p/2})$. This admits many simple rates such as if $C = \log n$, then $m = n/C_n = n/\log n$ or $B_n = \omega([\log(n/\log n)]^{p/2})$ which is a slow poly-logarithmic growth rate. In this case, the number of communities is smaller than the domain. But if $C = \sqrt{n}$, then $m = \sqrt{n}$ and, therefore, $B_n = \omega(2^{-p/2}(\log n)^{p/2})$, the number of communities can grow faster than the domain. $\square$

Appendix B. Generating latent space points

We now describe how we generate our points in the three latent spaces. The basic idea is to generate K group centres. We then call the first n/K nodes to be in group 1, the second n/K nodes to be in group 2, and so on. Let $c_i \in \{1, \ldots, K\}$ denote the group membership of node i. Finally, we distribute the node latent space positions centred at their group locations according to some procedure that is unique for each of the three geometries. To generate the latent space positions in the Euclidean case, we do the following:

Table B1. The parameter values used to make the results in Section 5

	E	S	H
E	$\sigma = 0.5$	$\sigma = 0.8$	$\sigma = 0.8$
S	$\kappa = 0.75$	$\kappa = 1$	$\kappa = 0.75$
H	scale $= 2.5$, $\kappa = 0.75$	scale $= 2.5$, $\kappa = 0.75$	scale $= 2.5$, $\kappa = 1$

Note. The rows correspond to the true data-generating process and the columns correspond to the null hypothesis being tested.

1. Generate K group centres $\mu \in \mathbb{R}^p$ distributed according to μ i.i.d. $N(0_p, \sigma^2 I_p)$.

2. Then simulate the positions of the nodes as $z_i \mid c_i$ i.i.d. $N(\mu_{c_i}, (\sigma^2/K)I_p)$.

To generate the latent space positions in the spherical case, we do the following:

1. Generate K group centres $\mu \in \mathbf{S}^2(\kappa)$. To do this, we generate two angles: θ i.i.d. Unif$(0, \pi)$ and ϕ i.i.d. Unif$(0, 2\pi)$. Then compute

$$\mu_i = \kappa^{-1/2}\left(\sin(\theta_i)\cos(\phi_i),\ \sin(\theta_i)\sin(\phi_i),\ \cos(\phi_i)\right) \in \mathbb{R}^3.$$

2. Then simulate the positions of the nodes. To do this, generate two angles $\theta_i \sim$ Unif$(\theta_{c_i} - \delta, \theta_{c_i} + \delta)$ and $\phi_i \sim$ Unif$(\phi_{c_i} - \delta, \phi_{c_i} + \delta)$ and compute

$$\mu_i = \kappa^{-1/2}\left(\sin(\theta_i)\cos(\phi_i),\ \sin(\theta_i)\sin(\phi_i),\ \cos(\phi_i)\right) \in \mathbb{R}^3.$$

To generate the latent space positions in the Hyperbolic case, we do the following:

1. Generate K group centres $\mu \in \mathbf{H}^2(\kappa)$. To do this, we generate two locations x_i and y_i distributed uniformly on $[-s, s] \times [-s, s]$ and select the third coordinate $z = \sqrt{1/\kappa + x_i^2 + y_i^2}$ so by construction $(x, y, z) \in \mathbf{H}^2(\kappa)$.

2. Then simulate the positions of the nodes. To do this, generate two coordinates x_i and y_i distributed uniformly on $[x_{c_i} - \delta, x_{c_i} + \delta] \times [y_{c_i} - \delta, y_{c_i} + \delta]$ then set $z_i = \sqrt{1/\kappa + x_i^2 + y_i^2}$.

We next present the parameters used for the simulations in Section 5. In Table B1, κ is the curvature used for the Spherical geometry. The σ parameter determines the spread of the points in the Euclidean geometry. For the hyperbolic geometry, the scale refers to the scale of the first two coordinates of the space. In all of these results, we use rate $= 1/3$. We draw the node effects v_i i.i.d. Unif$(\beta, 0)$, where we set $\beta = -0.01$.

Appendix C. Choosing bounds for curvature estimate

We now discuss a way to pick a, the lower bound in the spherical method to pick κ. Note that the maximum distance between any two points is $r\pi = \pi/\sqrt{\kappa}$, which occurs when the points are antipodal. This shows that for a distance matrix $D = \{d_{ij}\}$, which contains distances between K points on $\mathbf{S}^p(\kappa)$, it must be that

$$\max_{1 \le i,j \le K} d_{i,j} \le \frac{\pi}{\sqrt{\kappa}}.$$

By solving for κ, we see that κ satisfies

$$\kappa \leq \left(\frac{\pi}{\max\limits_{1 \leq i, j \leq K} d_{ij}} \right)^2 := b.$$

Based on the discussion in Wilson et al. (2014), we set

$$a := \left(\frac{1}{3 \min_{i,j} d_{i,j}} \right)^2.$$

The suggestion for a comes from Wilson et al. (2014), which says that for curvature values less than a, the space is essentially Euclidean. We use the same bounds for the hyperbolic case, but we flip the signs so that $[a, b] \subseteq (-\infty, 0]$. Future work could more thoroughly investigate how to pick the bounds for the hyperbolic case.

Figure 2a plots the function $\kappa \mapsto |\lambda_1(\kappa W_\kappa)|$ when D corresponds to $K = 15$ points drawn randomly on $\mathbf{S}^2(1)$. Figure 2b plots the function $\kappa \mapsto |\lambda_{K-1}(\kappa W_\kappa)|$ when D corresponds to $K = 15$ points drawn randomly on $\mathbf{H}^2(-1)$. The functions are both minimized at the true curvature ($\kappa_0 = 1$ for the spherical case and $\kappa_0 = -1$ for the hyperbolic case).

Appendix D. Additional details on the bootstrap procedure

Given n independent and identically distributed data points $X_1, \ldots, X_n$ drawn from a distribution we want to estimate a parameter θ with an estimator $\hat{\theta}_n$. We make the following assumption about $\hat{\theta}_n$, which appears in Politis and Romano (1994).

Algorithm 3: Hypothesis Testing via Sub-sample Bootstrap

Input: adjacency matrix G, sub-sample size m, and number of bootstrap samples B, and rate $r \geq 0$.

1. Compute the observed eigenvalue $\lambda_{\tilde{k}}(\hat{W}_{\hat{\kappa}}(\hat{D}))$ for $\tilde{k} = K$ for Euclidean/spherical and $\tilde{k} = 2$ for hyperbolic.
2. Construct a bootstrap distribution of eigenvalues. For $b = 1, \ldots, B$:
   (a) Sample $D_b^\star$
       (i) Let $\mathbf{I} := \{I_1, \ldots, I_m\}$ and $\mathbf{J} := \{J_1, \ldots, J_m\}$ be two sets of integers of length m drawn independently and uniformly from $\{1, \ldots, \ell\}$ with replacement.
       (ii) Calculate $P_b^\star$ with entries

$$p_{b,kk'}^\star = \max\left\{ \frac{1}{m^2} \sum_{i,j} G_{ij} \cdot \mathbf{1}\{i \in C_k \cap \mathbf{I}, \ j \in C_{k'} \cap \mathbf{J}\}, \frac{1}{\ell^2} \right\}.$$

       (iii) Calculate $D_b^\star = -\log(P^\star / \hat{\tau}^2)$ where division is component-wise.
   (b) Calculate $W_b^\star = W_{\hat{\kappa}}(D_b^\star)$ and eigenvalue $\lambda_{\tilde{k}}(W_\kappa(D_b^\star))$.
3. Compute the CDF of the deviation in the bootstrapped and empirical eigenvalue

$$\hat{L}_n(x) := \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{m^{2r} \cdot \left(\lambda_{\tilde{k}}(W_{\hat{\kappa}}(D_b^\star)) - \lambda_{\tilde{k}}(\hat{W}_{\hat{\kappa}}(\hat{D})) \right) \leq x\}, \quad \text{for any } x \in \mathbb{R}.$$

4. Compute critical values $c_{n,1-\alpha} = \inf\{x : \hat{L}_n(x) \geq 1 - \alpha\}$.
5. Test hypotheses:
   (a) Reject $H_{0,e}$ and $H_{0,s}$ when, for each of their respective test matrices,

$$\ell^{2r} \cdot \lambda_K(\hat{W}_{\hat{\kappa}}(\hat{D})) < c_{n,\alpha}.$$

   (b) Reject $H_{0,b}$ when

$$\ell^{2r} \cdot \lambda_2(\hat{W}_{\hat{\kappa}}(\hat{D})) > c_{n,1-\alpha}.$$

Assumption D.1   There exists a deterministic sequence τ_n such that $\tau_n(\hat{\theta}_n - \theta)$ converges in distribution to some random variable L.

Suppose that the goal is to construct confidence intervals for θ using $X_1, \ldots, X_n$. To do this, we select a sub-sample rate $m = m(n)$, where $m \leq n$. Then, let $Y_1, \ldots, Y_{\binom{n}{b}}$ be all the subsets of X of size b, and let $\hat{\theta}_{n,i}$ be the estimate of θ using the ith subset Y_i. Using the rate τ_n from Assumption D.1, with n replaced by the 'sub-sample' size b, we can form the empirical CDF of $\tau_b(\hat{\theta}_{n,i} - \hat{\theta}_n)$,

$$L_n(x) := \frac{1}{\binom{n}{b}} \sum_{i=1}^{\binom{n}{b}} \mathbf{1}\{\tau_b(\hat{\theta}_{n,i} - \hat{\theta}_n) \leq x\}.$$

Intuitively, as n and $b \to \infty$, we expect that L_n converges to the CDF of $\tau_n(\hat{\theta}_n - \theta)$, denoted by L. If this were true, then we could use the quantiles of $\tau_b(\hat{\theta}_{n,i} - \hat{\theta}_n)$ as estimates of the quantiles of $\tau_n(\hat{\theta}_{n,i} - \hat{\theta}_n)$, which would allow us to compute confidence intervals for θ. The following result shows when we can use L_n to construct asymptotically correct confidence intervals for θ.

Proposition D.1   (Theorem 2, (iii) of Politis & Romano, 1994). Let $c_n(1 - \alpha) := \inf\{x : \hat{L}_n(x) \geq 1 - \alpha\}$. Similarly, let $c(1 - \alpha) = \inf\{x : L(x) \geq 1 - \alpha\}$ where L is the CDF of X_1. If the CDF of X_1 is continuous at $c(1 - \alpha)$ and $\tau_b/\tau_n \to 0$ and $b/n \to 0$ then

$$\mathbb{P}\big(\tau_n(\hat{\theta}_n - \theta) \leq c_n(1 - \alpha)\big) \to 1 - \alpha.$$

This proposition allows us to construction asymptotically correct confidence intervals for θ from the sub-sampled data. Note that when n is large, computing all $\binom{n}{b}$ subsets of X is computationally infeasible, so we instead select a collection $\{Y_1, \ldots, Y_s\}$ for some integer $s \leq \binom{n}{b}$, and compute

$$\hat{L}_n(x) := \frac{1}{s} \sum_{i=1}^{s} \mathbf{1}\{\tau_b(\hat{\theta}_{n,i} - \hat{\theta}_n) \leq x\}.$$

According to Politis and Romano (1994), we have the following result:

Proposition D.2   (Theorem 2, (iii) of Politis & Romano, 1994). Let $c_n(1 - \alpha) := \inf\{x : \hat{L}_n(x) \geq 1 - \alpha\}$. Similarly, let $c(1 - \alpha) = \inf\{x : L(x) \geq 1 - \alpha\}$ where L is the CDF of X_1. If the CDF of X_1 is continuous at $c(1 - \alpha)$ and $\tau_b/\tau_n \to 0$ and $b/n \to 0$, then

$$\mathbb{P}\big(\tau_n(\hat{\theta}_n - \theta) \leq \hat{c}_n(1 - \alpha)\big) \to 1 - \alpha.$$

This result allows us to construct confidence intervals for θ. Having described the sub-sampling method from Politis and Romano (1994), we now return to our original problem and show how to apply this method to our problem. The parameter interest θ is the eigenvalue $\lambda_{k^\star}(W)$. To study this, we will show how to use the Politis and Romano (1994) method to sub-sample the distance matrix D. Using this sub-sampled distance matrix, we can then compute sub-sampled matrices W_κ and compute their eigenvalues, since W_κ is just a simple transformation of D.

The data in our problem is the adjacency matrix G. More concretely, it is the adjacency matrix for the subgraph with nodes $\bigcup_{k=1}^{K} C_i(\ell)$, the union of all K cliques. We fix some sub-sample rate m. With the sub-sample rate, we then want to re-sample the entries of D. To do this, we will focus on how to do this for the (k, k') entry of D. This process is repeated for all the entries of D. Let $\tilde{G}_{k,k'}$ denote the adjacency matrix corresponding to the sub-graph induced by the nodes in $C_k(\ell) \cup C_{k'}(\ell)$. For example, if $\ell = 3$, then a potential $\tilde{Y}_{k,k'}$ might take the form

$$\tilde{G}_{k,k'} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

This indicates that the first node in C_k connects to the first node in $C_{k,k'}$ but not to the second or third nodes in $C_{k'}$. We then sample two sets of integers of length m, denoted by I_k and $I_{k'}$, independently and uniformly from $\{1, \ldots, \ell\}$, without replacement. These indices will be the re-sampled nodes. We then compute

$$P_{k,k'}^{\star} = \frac{1}{m^2} \sum [\tilde{G}_{k,k'}]_{ij} \mathbf{1}\{(i, j) \in I_k \times I_{k'}\}.$$

Since it is possible that $P_{k,k'}^{\star}$ is zero (meaning that the re-sampled pairs of nodes do not connect), we use $P_{k,k'}^{\star} = \max(1/\ell^2, P_{k,k'}^{\star})$, since we observe at least one edge in $\tilde{G}_{k,k'}$. We repeat this procedure for all pairs of edges (k, k'). We then compute $D^{\star}$ using (1). We provide a step-by-step implementation of the sub-sampling method in Algorithm 3.

Recalling that our parameter of interest is the eigenvalue $\lambda_{k^\star}(W)$, we use the above procedure to compute $\lambda_{k^\star}(W_b^{\star})$ for $b = 1, \ldots, B$. We then define

$$\hat{L}_n(x) = \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}\{m^{2r}(\lambda_{k^\star}(W_i^{\star}) - \lambda_{k^\star}(\hat{W})) \leq x\}, \quad \text{for any } x \in \mathbb{R}.$$

We then perform hypothesis testing. To do this, we let $c_n(1 - \alpha) = \inf\{x : \hat{L}_n(x) \geq 1 - \alpha\}$ be the $(1 - \alpha)100\%$ percentile of $m^{2r}(\lambda_{k^\star}(W_i^{\star}) - \lambda_{k^\star}(\hat{W}))$. Then, from Proposition 3.1, we know that $P(m^{2r}(\lambda_{k^\star}(\hat{W}) - \lambda_{k^\star}(W)) \leq c_n(1 - \alpha)) \approx 1 - \alpha + o(1)$ for large ℓ. This motivates the bootstrapping method we summarize in Algorithm 3.

D.1 Sensitivity of bootstrapping algorithm

We now analyse the sensitivity of the sub-sampling algorithm in 3 to the parameter B, the number of matrices $D^{\star}$ we generate. To do this, we generate a 2-dimensional lattice in $\mathbb{R}^2$ of length 5 and 7 and randomly select 9 points from these C^2 points. Using these 9 points, we generate 50 networks and compute the p-values for the Euclidean, spherical, and hyperbolic geometries using $B = 1,000$ and $B = 10,000$. In Figures D1 and D2, we plot the 50 p-values for this simulation, and a diagonal
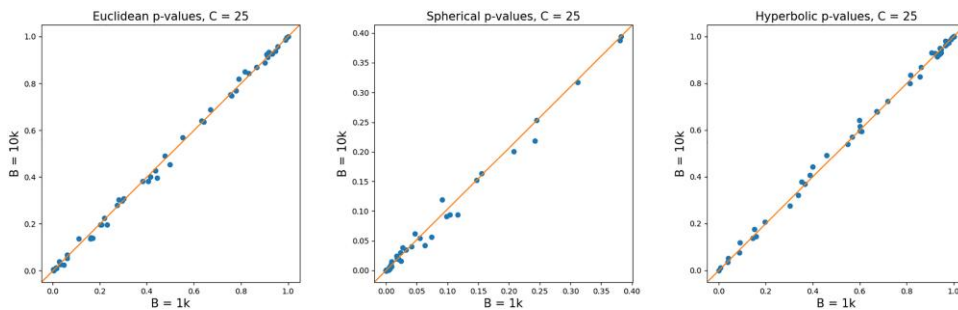

Figure D1. Plot of the p-values for the three geometries (Euclidean, spherical, hyperbolic from left to right). Each point (x, y) corresponds to two p-values. The x coordinate is the p-value computed using $B = 1,000$ and the y coordinate is the p-value computed using $B = 10,000$. The graph is computed using the graph model in (1) and we use 9 latent space positions drawn randomly from a 5×5 lattice in $\mathbb{R}^2$. Since most points fall on or near the diagonal, this is evidence that in this scenario, the sub-sampling algorithm in Algorithm 3 is not very sensitive to the choice of B, the number of distance matrices we sub-sample.
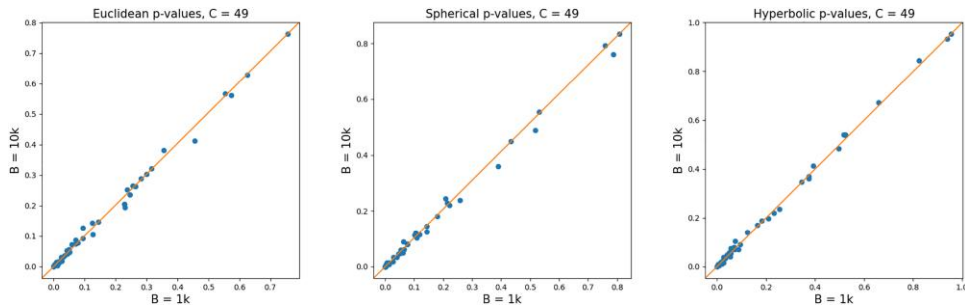
Figure D2. Plot of the *p*-values for the three geometries (Euclidean, spherical, hyperbolic from left to right). Each point (*x*, *y*) corresponds to two *p*-values. The *x* coordinate is the *p*-value computed using $B = 1,000$ and the *y* coordinate is the *p*-value computed using $B = 10,000$. The graph is computed using the graph model in (1) and we use 9 latent space positions drawn randomly from a 7×7 lattice in $\mathbb{R}^2$. Since most points fall on or near the diagonal, this is evidence that in this scenario, the sub-sampling algorithm in Algorithm 3 is not very sensitive to the choice of *B*, the number of distance matrices we sub-sample.

line from (0, 0) to (1, 1). We see that most *p*-values lie on or very near to the diagonal line, which indicates that the sub-sampling algorithm is not very sensitive to the choice of parameter *B*.

Appendix E. Rank estimator

In Algorithm 4, we formally describe the algorithm and the estimate of the rank of W_κ.

Algorithm 4:   Estimating Rank of W_κ

---

1. Compute the scree function $\phi_T(j) := \frac{\hat{\lambda}_{K-j-1}}{\sum_{i=1}^{K} \hat{\lambda}_i}$ for $j \in \{0, 1, \ldots, K-1\}$.
2. Sample *B* bootstrapped $D_1^\star, \ldots, D_B^\star$ matrices from Algorithm 3 and use them to compute $W_1^\star, \ldots, W_B^\star$.
3. For $j \in \{0, 1, \ldots, K-1\}$,
    (a) Define $\hat{A}_j \in \mathbb{R}^{K \times j}$ with $\hat{A}_j = (\hat{v}_{K-j+1}, \ldots, \hat{v}_K)$.
    (b) Let $v_1^\star, \ldots, v_K^\star$ denote the eigenvectors of $W_i^\star$ corresponding to its eigenvalues $\lambda_1^\star \leq \cdots \leq \lambda_K^\star$.
    (c) Set $A_{j,i}^\star \in \mathbb{R}^{K \times j}$ with $A_{j,i}^\star = (v_{K-j+1}^\star, \ldots, v_K^\star)$.
4. Compute

$$f_n^0(j) = 1 - \frac{1}{B} \sum_{i=1}^{B} |\det(\hat{A}_j^T A_{j,i}^\star)|$$

5. Compute

$$f_n(j) = \frac{f_n^0(j)}{\sum_{i=0}^{K-1} f_n^0(i)}.$$

6. The estimate $\hat{r}$ of the rank of W_κ is

$$\hat{r} = \underset{j \in \{0, 1, \ldots, K-2\}}{\arg \min} \left(\phi_T(j) + f_n(j) \right). \tag{E1}$$

Luo and Li (2016) estimator use two pieces of information. The first is the scree function, which plots the sample eigenvalues in order from larges to smallest. In Figure E1, we plot the scree function for a distance matrix computed between $K = 15$ points on a 3-dimensional Euclidean latent space. We see that the scree plot is large but decreasing for the first three eigenvalues but becomes flat after that point. The second piece of information this estimator uses is the variability of the bootstrapped eigenvectors of the matrix W_κ, given in step (4) of Algorithm 4. Luo and Li (2016)
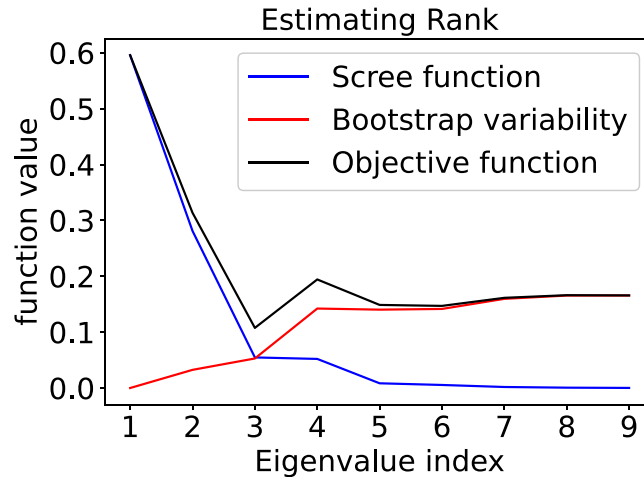
Figure E1. We generate a graph using a 3-dimensional Euclidean latent space with $K = 10$ cliques. We plot the scree function ϕ and the bootstrap variability function f_n defined in Algorithm 4. We also plot their sum, defined as the objective function. The horizontal axis represents the possible ranks of the matrix. We see the objective function has a minimum at 3, so we estimate the rank of the matrix to be 3, which is the true dimension of the latent space.

argue that for $j < r$, the true rank of W_κ, there is little variation in the term $f_n(j)$ in step (4) but for $j \geq r$, this function increases. We see this behaviour in Figure E1: For $j < 3$, the bootstrap variability is lower than when $j \geq 3$. See Luo and Li (2016) for a more thorough explanation of why this phenomenon occurs. Based on these two pieces of information, Luo and Li (2016) suggest adding the two functions together to produce a final objective function. They claim that this new function has a 'ladle shape. The minimum of this new function is our estimate of the rank of W_κ.

Appendix F. Additional details for the Banerjee et al. (2019) data

We show cumulative distribution plots of the number of cliques (sizes 4, 5, and 6) across the 75 villages in Figure F1.

Appendix G. Additional simulation results

We plot curvature estimates for 100 simulated graphs using cliques of size $\ell \in \{5, 7, 9\}$. We see that as ℓ increases, the variance and bias of $\hat{\kappa}_S$ decreases in the spherical case (Figure G1).

We now analyse the accuracy of the curvature methods for the spherical and hyperbolic latent space models. The estimator in Proposition 3.2 minimizes $\kappa \mapsto \lambda_1(\kappa \hat{W}_\kappa)$. But from Lemma 2.1, we in fact know that the first few eigenvalues of $\cos(\sqrt{\kappa}\hat{D})$ are zero, which suggests that we can use the estimator
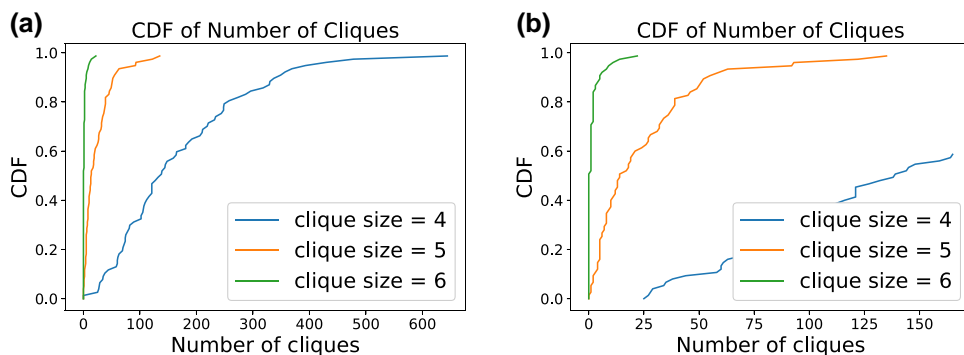

Figure F1. CDF of number of cliques for clique sizes $\ell \in \{4, 5, 6\}$ for the 75 Indian villages.
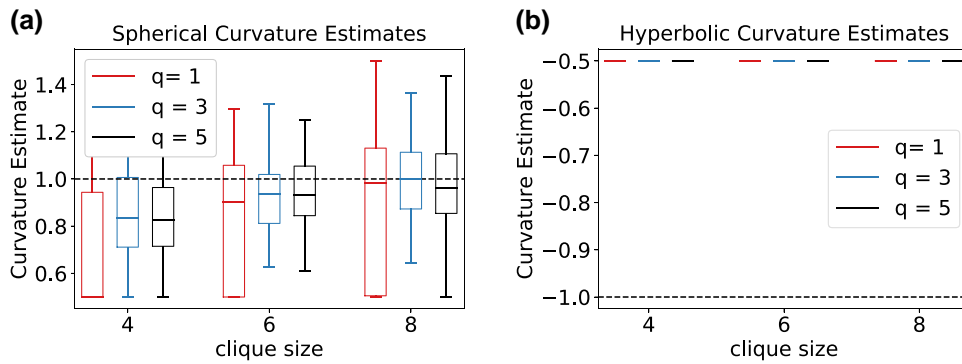
Figure G1. Left: Curvature estimates for $\mathbf{S}^2(1)$ using $K = 10$ cliques, with clique size $\ell = 4, 6, 8$ on the horizontal axis. We use $q = 1, 3, 5$ where q is defined in (G1). We plot the true curvature $\kappa = 1$ in the black dashed line. Right: Curvature estimates for $\mathbf{H}^2(-1)$ using $K = 10$ cliques, with clique size $\ell = 4, 6, 8$ on the horizontal axis. In Figure. G2, we analyse how large the clique size must be for the hyperbolic curvature estimator to perform better. (a) Spherical. (b) Hyperbolic.
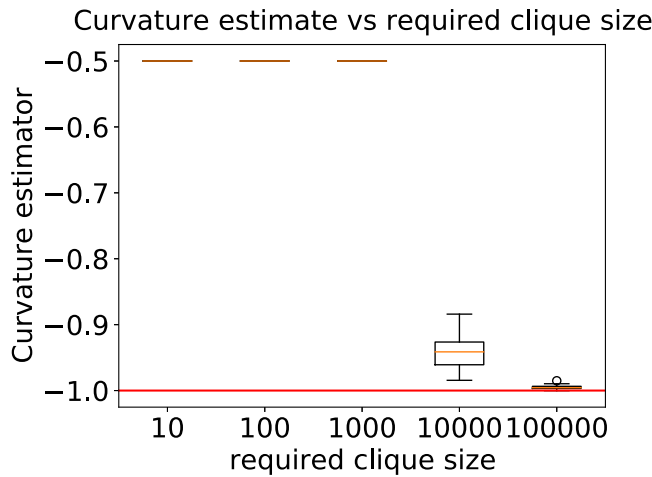
Figure G2. We plot the estimated curvature using distances computed from $K = 10$ points in $\mathbb{H}^2(-1)$ for various clique sizes on the *x*-axis. To simulate this figure, we fix a set of $K = 10$ locations on $H^2(-1)$ and compute pairwise distances d_{ij}. We then simulate 50 independent noisy realizations $\hat{d}_{ij} \sim N(d_{ij}, \sigma^2)$ for $\sigma \in \{0.1, 0.01, 0.001, 0.0001, 0.00001\}$. We then compute the estimate of the curvature from (7) using $\hat{D} = \{\hat{d}_{ij}\}$. Given a certain noise level, we use the fact that when using cliques to estimate distances, the variance of $\hat{d}_{ij}$ is on the order of $1/\ell^2$. So we equate $\sigma^2 = 1/\ell^2$ to compute an approximately required clique sized required. For example, we require clique sizes of approximately 10^4 or higher to obtain an estimator that does not always select the lower bound a in (7).

$$\hat{\kappa}(q) = \frac{1}{q} \sum_{i=1}^{q} \hat{\kappa}_i, \quad \hat{\kappa}_i = \arg\min_{\kappa \in [a,\, b]} \left| \lambda_i \left(\kappa W(\hat{D})_\kappa \right) \right| \tag{G1}$$

Assuming that $q \ll K$, we can reasonably believe that the first through qth eigenvalues of $\cos(\sqrt{\kappa}D)$ are zero. In fact, it is easy to modify the proof of Proposition 3.2 to show that $\hat{\kappa}(q) \xrightarrow{p} \kappa$, provided that $q \ll K$. Taking $q > 1$ does not always reduce the variance of $\hat{\kappa}(t)$, which could be because the $\hat{\kappa}_1, \ldots, \hat{\kappa}_q$ are not necessarily independent. In Figure G1, we plot 250 estimates of κ when $\mathcal{M}^p(\kappa) = \mathbf{S}^2(1)$ and when $\mathcal{M}^p(\kappa) = \mathbf{H}^2(-1)$ using $K = 10$. Specifically, we generate a network and find K cliques of size $4, 5, 6$. We estimate distances between these K groups using the number of cross-clique edges as described in Algorithm 2. Although Proposition 3.2 says that the

estimate is consistent as the sample size grows, we see that the hyperbolic curvature estimate has not reached its asymptotic behaviour for cliques of size 4, 5, and 6. In Figure G2, we determine how big the cliques must be in order for the Hyperbolic curvature estimator to be close to the true curvature. However, as we see in Figure 4, we see that the classification of geometry is still accurate, which is our ultimate goal.

Appendix H. Sectional curvature definitions

To discuss sectional curvature, some preliminary definitions are required. We review these concepts in a self-contained way. The reader may look to O'Neill (1983) for a more in-depth explanation of these concepts. The *tangent space* at $m \in \mathcal{M}^p$ is denoted $T_m(\mathcal{M}^p)$, defined as the set of all tangent vectors to the manifold at m: that is, all real-valued functions v that map any smooth function $f : \mathcal{M}^p \to \mathbb{R}$ to $v(f(m)) \in \mathbb{R}$ that is $\mathbb{R}$-linear and Leibnizian.[12]

A Riemannian manifold $(\mathcal{M}^p, g)$ comes equipped with a *metric tensor g* which at every point $m \in \mathcal{M}^p$ takes two vectors in the tangent space of the manifold at m, $u, v \in T_m(\mathcal{M}^p)$, and maps it to a non-negative number: $g_m(u, v) \mapsto \mathbb{R}_{\geq 0}$ and the map is symmetric, non-degenerate, and bilinear. That is, g defines a scalar product over the manifold; on a smooth manifold the metric tensor smoothly varies over the manifold itself.

To define curvature, we first need to define the *Riemann curvature tensor*, R evaluated at point $m \in \mathcal{M}^p$, which takes three tangent vectors in the tangent space at m—$u, v, w \in T_m(\mathcal{M}^p)$—and returns $R_m(u, v)w \in T_m(\mathcal{M}^p)$[13]

$$R_m(u, v)w := \nabla_{[u,v]}w - [\nabla_u, \nabla_v]w.$$

Here is some intuition. Consider the vector w which is tangent to the manifold at m. Consider the plane defined by u and v which are tangent at m as well. Now take w and parallel transport it, meaning take it along the parallelogram in the u direction and then v direction and compare that to taking the same w along the v direction and then u direction to the same point. The returned vector has entries that describe how much w changes relatively across the two paths. If this is identically zero, this means of course that there was no change in this parallel transportation. Intuitively, if one does this on a flat manifold, for instance $\mathbb{R}^2$ with the usual Euclidean metric, it is clear that the vector w does not change whatsoever. But on a sphere, for instance, the reader can intuit that things change.

Then the *sectional curvature* at m, which we refer to simply as curvature for the remainder of this paper, is given by

$$\kappa_m(u, v) := \frac{g_m(R_m(u, v)v, u)}{g_m(u, u) \cdot g_m(v, v) - g_m(u, v)^2}.$$

It turns out that this is independent of basis u, v whatsoever (see Lemma 39 in O'Neill, 1983 for instance) so we can simply write κ_m. That the manifold has constant sectional curvature means that for all $m \in \mathcal{M}^p$, $\kappa_m = \kappa$ and so we simply write $\mathcal{M}^p(\kappa)$.

Appendix I. Lattice simulations

We now demonstrate the type 1 error and power simulations by drawing points on a lattice in Euclidean, spherical, and hyperbolic space. We present the results in Figures I1–I6.

Appendix J. Other graph models

In (1), we consider an exponential link function that connects distances in the latent space to the probability that nodes form an edge. The exponential function has the desirable property that

[12] An obvious tangent vector is the directional derivative at a point on the manifold: it maps a smooth function to its derivative in that direction evaluated at that point on the manifold.

[13] Here [., .] is the Lie bracket.

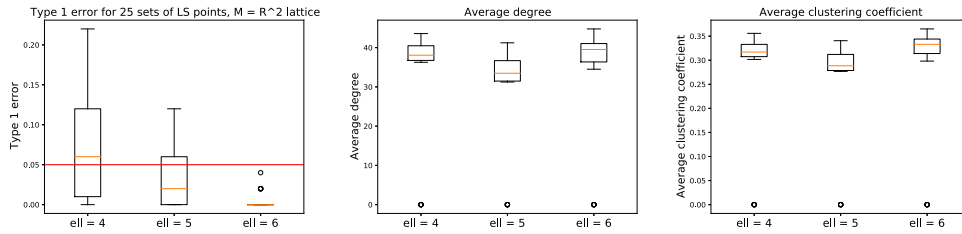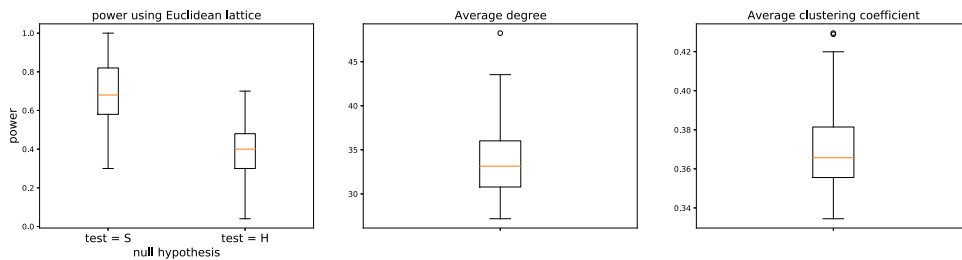**Figure I1.** Simulation results for the two-dimensional lattice. Using a 4×4 lattice in $\mathbb{R}^2$, we randomly select 25 sets of 9 latent space positions. For each set of positions, we generate 50 networks from using the graph model in (1) and calculate how many of these 50 networks we reject the null hypothesis that $\mathcal{M}$ is Euclidean. We repeat this for all 25 sets of latent space positions and plot the resulting probability of type 1 error for $\ell = 4, 5, 6$. We see that the type 1 error is at $\alpha = 0.05$ or below and decreases as ℓ increases. We also report the average degree (middle figure) and average clustering coefficient for the simulated networks. We use $\tau = 0.4$ and $\beta = -0.6$.

Figure I2. Simulation results for the two-dimensional lattice. Using a 4×4 lattice in $\mathbb{R}^2$, we randomly select 25 sets of 9 latent space positions. For each set of positions, we generate 50 networks from using the graph model in (1) and calculate how many of these 50 networks we reject the null hypothesis that $\mathcal{M}$ is spherical or hyperbolic. We repeat this for all 25 sets of latent space positions and plot the resulting power. We also report the average degree (middle figure) and average clustering coefficient for the simulated networks. We use $\tau = 0.4$.
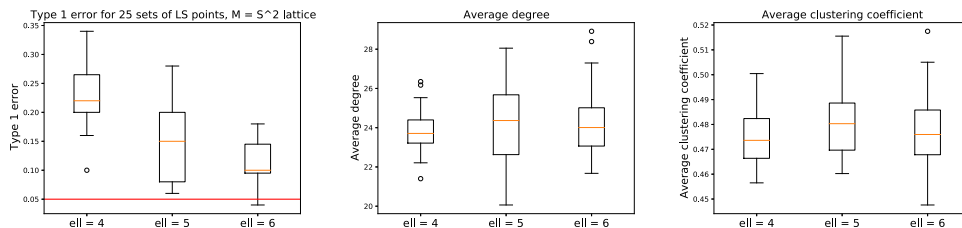

Figure I3. Simulation results for the two-dimensional lattice. Using a 3×3 lattice in $\mathbb{S}^2(1)$, we randomly select 25 sets of 4 latent space positions. For each set of positions, we generate 50 networks from using the graph model in (1) and calculate how many of these 50 networks we reject the null hypothesis that $\mathcal{M}$ is Euclidean. We repeat this for all 25 sets of latent space positions and plot the resulting probability of type 1 error for $\ell = 4, 5, 6$. We see that the type 1 error is above $\alpha = 0.05$ but decreases to about 0.1 as ℓ increases. We also report the average degree (middle figure) and average clustering coefficient for the simulated networks. We use $\beta = -0.2$.

$\exp(a + b) = \exp(a) \exp(b)$ which allows us to isolate the node effects and distances:

$$\mathbb{P}(g_{ij} = 1 \mid v_i^\star, v_j^\star, z_i^\star, z_j^\star) = \exp(v_i^\star + v_j^\star) \exp\{-d(z_i^\star, z_j^\star)\}.$$

So by integrating out the node effects, we see that

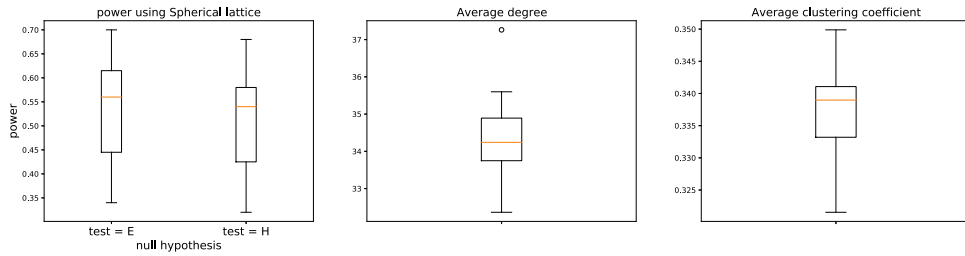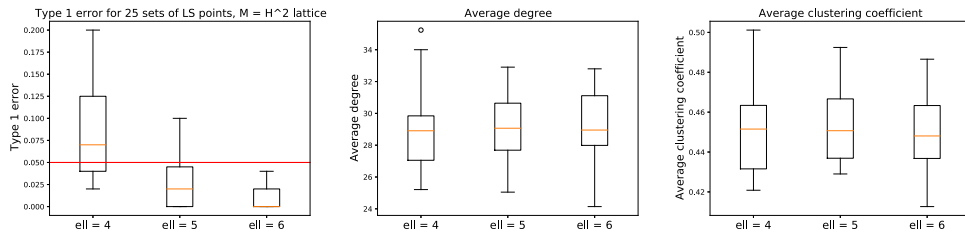$$\mathbb{P}(g_{ij} = 1 \mid z_i^\star, z_j^\star) = E\{\exp(v)\}^2 \exp\{-d(z_i^\star, z_j^\star)\}.$$

Figure I4. Simulation results for the two-dimensional lattice. Using a 5×5 lattice in $\mathbb{S}^2$, we randomly select 25 sets of 9 latent space positions. For each set of positions, we generate 50 networks from using the graph model in (1) and calculate how many of these 50 networks we reject the null hypothesis that $\mathcal{M}$ is Euclidean. We repeat this for all 25 sets of latent space positions and plot the resulting power. We also report the average degree (middle figure) and average clustering coefficient for the simulated networks. We use $\beta = -0.2$.
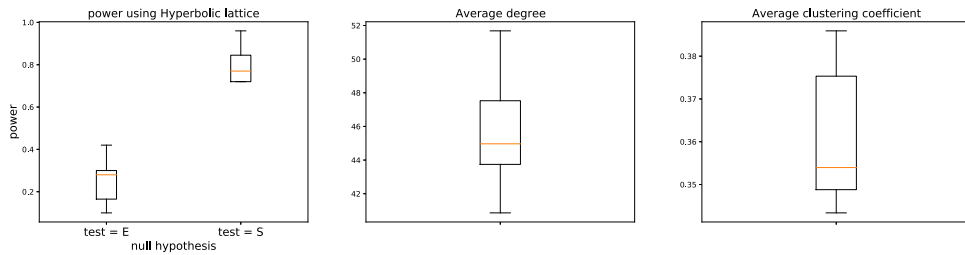

Figure I5. Simulation results for the two-dimensional lattice. Using a 5×5 lattice in $\mathbb{H}^2(-1)$, we randomly select 25 sets of 9 latent space positions. For each set of positions, we generate 50 networks from using the graph model in (1) and calculate how many of these 50 networks we reject the null hypothesis that $\mathcal{M}$ is Euclidean. We repeat this for all 25 sets of latent space positions and plot the resulting power. We also report the average degree (middle figure) and average clustering coefficient for the simulated networks. We use $\beta = -0.2$.

Figure I6. Simulation results for the two-dimensional lattice. Using a 5×5 lattice in $\mathbb{H}^2(-1)$, we randomly select 25 sets of 9 latent space positions. For each set of positions, we generate 50 networks from using the graph model in (1) and calculate how many of these 50 networks we reject the null hypothesis that $\mathcal{M}$ is Euclidean. We repeat this for all 25 sets of latent space positions and plot the resulting power. We also report the average degree (middle figure) and average clustering coefficient for the simulated networks. We use $\beta = -0.2$.

We now illustrate how our approach can be applied to other graph models.

Example 7   (Expit link function). Suppose instead of the exponential link function, we use the expit link function, which was used in Hoff et al. (2002), among many others:

$$\mathbb{P}(g_{ij} = 1 \mid v_i^{\star}, v_j^{\star}, z_i^{\star}, z_j^{\star}) = \text{expit}\{v_i^{\star} + v_j^{\star} - d(z_i^{\star}, z_j^{\star})\},$$

where $\text{expit}(x) = \exp(x)/\{1 + \exp(x)\}$ is the expit function.

The choice of which link function to choose is an important one, which network goodness-of-fit tests can help address (Lubold et al., 2021; Ouadah et al., 2019; Xu & Reinert, 2021).

Now, after integrating out the node effects, we have

$$\mathbb{P}(g_{ij} = 1 \mid v_i^\star, v_j^\star, z_i^\star, z_j^\star) = \iint \text{expit}\{v_i^\star + v_j^\star - d(z_i^\star, z_j^\star)\} dF(v_i^\star) dF(v_j^\star),$$

which again assumes that the node effects are drawn independently of each other. Let us define a function $H : [0, \infty) \to [0, \infty)$ with

$$H(x) := \iint \text{expit}\{v_i^\star + v_j^\star - x\} dF(v_i^\star) dF(v_j^\star).$$

Using Monte Carlo integration, we can approximate H to within any desired certainty. Now, suppose that C_k and $C_{k'}$ are two cliques in the graph G drawn from this graph model. Then,

$$\hat{p}_{kk'} := \frac{1}{\ell^2} \sum_{i \in C_k} \sum_{j \in C_{k'}} G_{ij} \approx H\{d(z_k, z_{k'})\}.$$

We can then solve for the argument of H that solve the above expression. That is, assuming that H is invertible, we can write $\hat{d}_{kk'} = H^{-1}(\hat{p}_{kk'})$. Assuming H^{-1} is continuous, we can estimate distances consistently.

Example 8 (Latent space model with covariates). Consider the model given in Hoff et al. (2002):

$$\mathbb{P}(g_{ij} = 1 \mid \alpha^\star, \beta^\star, z_i^\star, z_j^\star) = \text{expit}\{\alpha^\star + \beta^\star X_{ij} - d(z_i^\star, z_j^\star)\}.$$

where $\alpha^\star$ measures the baseline probability of connecting, X_{ij} is a dyad-level covariate, $\beta^\star$ measures the effect of this covariate on the probability of edges, and d measures distances in the latent space.

To illustrate how to estimate parameters in this model, suppose that X_{ij} is measuring homophily, so that X_{ij} is 1 if nodes i and j share some common trait (like ethnicity, education level, political beliefs, etc.) and is 0 otherwise. Suppose also that covariates are observable.

Suppose we observe a clique of nodes with the same trait. Then,

$$\frac{1}{\binom{\ell}{2}} \sum_{i < j} G_{ij} \approx \text{expit}(\alpha^\star + \beta^\star), \tag{J1}$$

since nodes in the same clique are likely to be close to each other and, therefore, $d(z_i^\star, z_j^\star) = 0$ for i and j in the same clique (Assumption 1.3).

Suppose we also observe an ℓ-clique $C(\ell)$ with nodes that have different traits. Partition these nodes into two groups, $C_0(\ell)$ and $C_1(\ell)$, where the subscript 0 and 1 indicates whether the covariate is 0 or 1, so that $C(\ell) = C_0(\ell) \cup C_1(\ell)$. Then,

$$(|C_0(\ell)||C_1(\ell)|)^{-1} \sum_{i \in C_0(\ell)} \sum_{j \in C_1(\ell)} G_{ij} \approx \text{expit}(\alpha^\star).$$

Thus, we can estimate $\alpha^\star$ by solving the above equation. Using this estimate, we can then plug this value into (J1) and solve for $\hat{\beta}$.

Given estimates of $\alpha^\star$ and $\beta^\star$, we can now estimate distances between cliques. To do this, we define for any two cliques C_k and $C_{k'}$, where all nodes have the same traits,

$$\hat{P}_{kk'} := (|C_k(\ell)||C_{k'}(\ell)|)^{-1} \sum_{i \in C_0(\ell)} \sum_{j \in C_1(\ell)} G_{ij}.$$

where $C_k(\ell)$ and $C_{k'}(\ell)$ are cliques of size ℓ. We can then solve for $\hat{D}_{kk'}$,

$$\hat{D}_{kk'} = \text{logit}(\hat{P}_{kk'}) - \hat{\alpha} - \hat{\beta}.$$

Appendix K. Existence of cliques and locations of nodes in a clique

In Theorem 1.2, we used edges between cliques to estimate distances between points on the unknown latent surface $\mathcal{M}$. Theorem 1.2 then states that as the graph size n and the clique size ℓ both go to infinity, we can consistently estimate all parameters of the latent space graph model. In particular, we have assumed that as ℓ, n both go to infinity, that we observe cliques of size ℓ. To understand the rate at which ℓ, n can grow, we consider a simplifying example. Let G be an undirected random graph drawn from the Erdos-Renyi model with parameters n and p. That is, edges form independently with probability p. The following result is a well-known result which says that in an ER graph, the clique number grows like $\log(n)$ for large n. It can be found in many places, such as in Grimmett and McDiarmid (1975).

> **Proposition K.1** (Grimmett & McDiarmid, 1975). The clique number of an ER model $Z_{n,p}$ satisfies
>
> $$\frac{Z_{n,p}}{\log(n)} \to \frac{2}{\log(1/p)}.$$
>
> almost surely.

In other words, the clique number $Z_{n,p}$ grows like $C\log(n)$ for the constant $C = 2\{\log(1/p)\}^{-1}$. So by taking $\ell = C\log(n)$, we will almost surely see an ℓ clique in the graph as $n, \ell \to \infty$. Now let us return to our problem. We observe a graph drawn from (1), where the node locations and effects are drawn iid from two distributions. Clearly, the probability of observing an $\ell = \ell(n)$ clique depends on the distributions of the points and node effects. To our knowledge, there are no results like Proposition K.1 that hold for arbitrary graph models. However, we would like to investigate the behaviour of the clique number for three common ways of assigning points in the latent space. In particular, we want to determine if there are cliques of size $\log(n)$ in graphs where the node locations are drawn according to a lattice mode, Gaussian mixture model, and uniform model. These three models are the models we study in Section 4.4. We give these results in Figures K1, K2, and K3. We see cliques of size $\log(n)$ with probability going to 1 as $n \to \infty$. The models are listed in increasing 'difficulty', meaning that it should be less likely for there to be clique when points are uniformly drawn (C) than when they are drawn from a GMM (B). And, when nodes are at the same location (A), it should be even more likely that a clique exists. We see
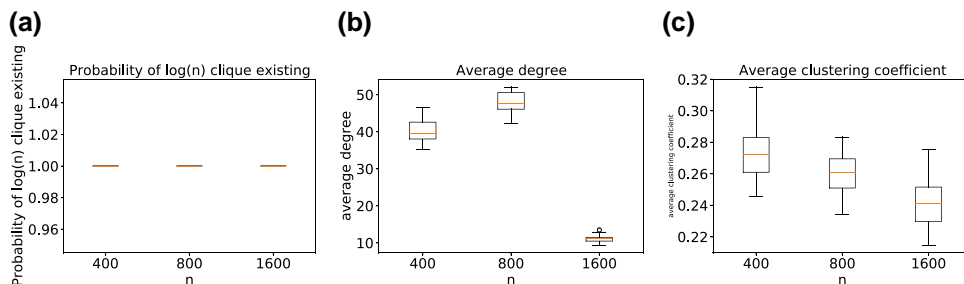

Figure K1. We generate 25 sets of n latent space positions using the lattice model. For each set of LS positions, we generate 50 graphs and count the number of times a clique of size $\log(n)$ exists in the graph. For each set of LS positions, we record the average degree for the 50 graphs and plot the 25 average values in (B). Similarly, in (C), we plot the average clustering coefficient. (a) Estimated probability of $\log(n)$ clique. (b) Average degree. (c) Average clustering coefficient.
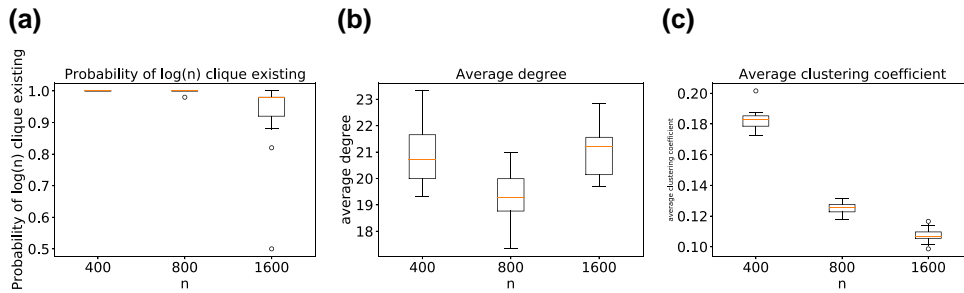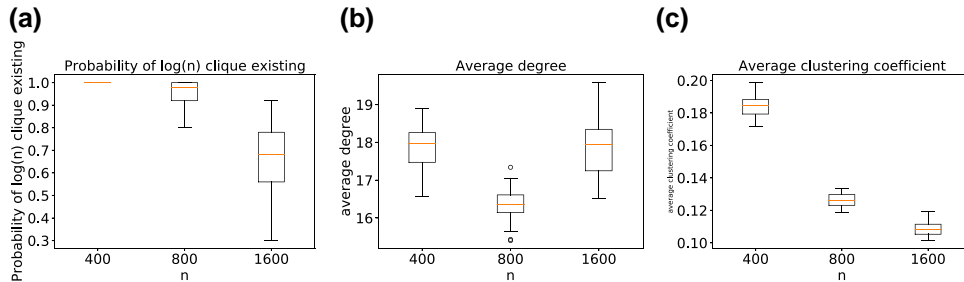
Figure K2. We generate 25 sets of *n* latent space positions using the Gaussian mixture model. For each set of LS positions, we generate 50 graphs and count the number of times a clique of size log (*n*) exists in the graph. For each set of LS positions, we record the average degree for the 50 graphs and plot the 25 average values in (B). Similarly, in (C) we plot the average clustering coefficient. (a) Estimated probability of log (*n*) clique. (b) Average degree. (c) Average clustering coefficient.

Figure K3. We generate 25 sets of *n* latent space positions from a uniform distribution. For each set of LS positions, we generate 50 graphs and count the number of times a clique of size log (*n*) exists in the graph. For each set of LS positions, we record the average degree for the 50 graphs and plot the 25 average values in (B). Similarly, in (C), we plot the average clustering coefficient. (a) Estimated probability of log (*n*) clique. (b) Average degree. (c) Average clustering coefficient.

this pattern in our simulations; for any *n*, the GMM in (B) has a higher probability of containing a log (*n*) clique for every *n*. But, for sufficiently large *n*, all models contain cliques of size at least log (*n*) with relatively high probability.

We also verify in Figure K4 that as the size of a clique goes up, the probability that node locations in the latent space are close to each increases.

Appendix L. Testing geometry using the Cayley–Menger determinant

For $K := 4$ points with pairwise distances $D = \{d_{ij}\}$, define the *Cayley–Menger* determinant to be the determinant of the matrix

$$CM := \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 \\ 1 & d_{21}^2 & 0 & d_{23}^2 & d_{24}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 & d_{34}^2 \\ 1 & d_{41}^2 & d_{42}^2 & d_{43}^2 & 0 \end{pmatrix}.$$

When these points are in $\mathbb{R}^2$, then the determinant of the matrix *CM* is 0. The goal of this section is to derive an estimator of geometry that uses this idea. Similar results hold for points from a spherical space or hyperbolic space, with slightly modified matrices.

We now propose a test of geometry using this method. To ensure that this is a reasonable comparison to our previous results, we generate a stochastic block model with locations $z_1, \ldots, z_K$,
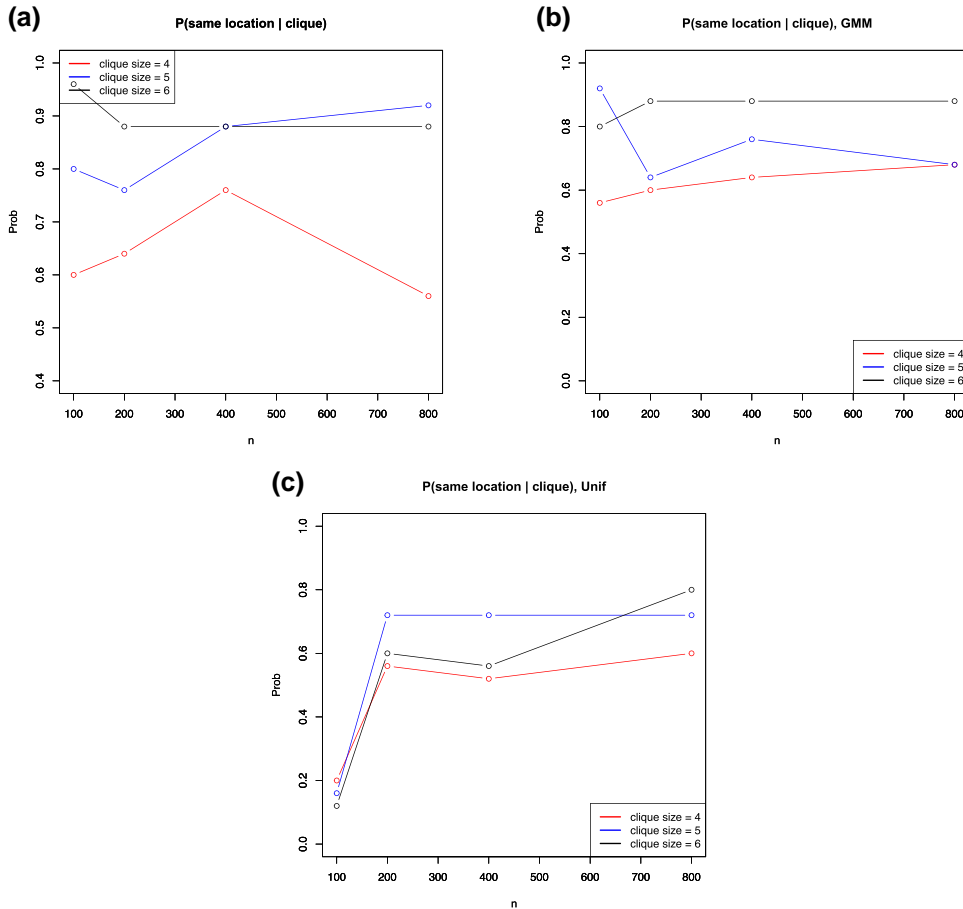
Figure K4. We generate 25 networks on n nodes for $n \in \{100, 200, 400, 800\}$. Using $C = \log(n)$, we generate C communities in a lattice model (A), a GMM (B), or a uniform model (C). We check how many times out of 25 nodes in an arbitrary clique are at the same location. We plot the corresponding probabilities above for clique sizes 4, 5, 6.

and assign edges based on $P(G_{ij} = 1 | z_i, z_j) = \exp(-d(z_i, z_j))$. These locations correspond to K cliques that we might find in a network. We then compute the determinant of $\widehat{CM}$, which is defined as above except that we use $\hat{d}^2_{ij}$ in place of d^2_{ij}. Since the distribution of the random variable under $H_0 : CM = 0$ is hard to derive, let us do a parametric bootstrap. For $b = 1, \ldots, B$, we repeat the following steps:

1. For each $i < j$, draw with replacement ℓ^2 edges from the set of edges that exist between locations z_i and z_j. This is the re-sampling step.
2. Compute the average number of re-sampled edges between i and j for any $i < j$. Call this number $p^\star_{ij}$ and set $d^\star_{ij} = -\log(p^\star_{ij})$.
3. Compute $CM^\star_b = \det(CM^\star)$, where $CM^\star$ is the matrix above with d_{ij} now replaced by $d^\star_{ij}$.

If 0 falls outside of the interval

$$(2\det(\widehat{CM}) - q_{1-\alpha/2}\{\det(CM)^\star\}, 2\det(\widehat{CM}) - q_{\alpha/2}\{\det(CM)^\star\}),$$

then we reject H_0, where $q_{\alpha/2}\{\det(CM)^\star\}$ is the $\alpha/2$ quantile of the empirical distribution $\{\det(CM)^\star\}$.
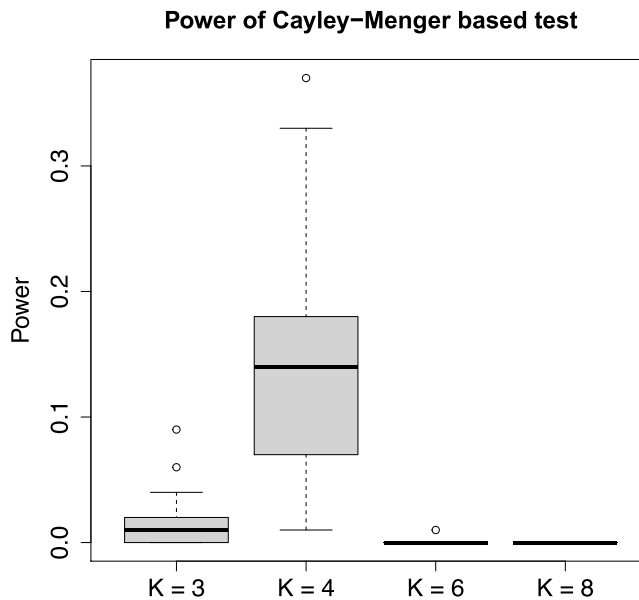
Power of Cayley−Menger based test

Figure L1. Power of the Cayley–Menger-based test of the Euclidean hypothesis. On the *x*-axis, we plot the number of points *K* that were used. On the *y*-axis, we plot the average number of rejections (out of 100) for 25 sets of *K* locations drawn from the sphere $\mathbf{S}^2(1/2)$.

We plot the power of this method using spherical data in Figure L1. We see that when we use $K = p + 2$, then the method works its best, but it still performs relatively poorly. So this seems to suggest that only at this value of K does it work well.

References

Abdelnour, F., Dayan, M., Devinsky, O., Thesen, T., & Raj, A. (2018). Functional brain connectivity is predictable from anatomic network's Laplacian eigen-structure. *NeuroImage*, *172*, 728–739. https://doi.org/10.1016/j.neuroimage.2018.02.016

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, *105*(2), 564–608. https://doi.org/10.1257/aer.20130456

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, *11*(4), 581–598. https://doi.org/10.1016/0047-259X(81)90099-3

Ambrus, A., Mobius, M., & Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review*, *104*(1), 149–182. https://doi.org/10.1257/aer.104.1.149

Andrews, D. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, *68*(2), 399–405. https://doi.org/10.1111/1468-0262.00114

Asta, D. M., & Shalizi, C. R. (2015). Geometric network comparisons. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence* (pp. 102–110). AUAI Press.

Banerjee, A., Breza, E., Chandrasekhar, A., Duflo, E., Kinnan, C., & Jackson, M. (2020). Changes in social network structure in response to exposure to formal credit markets. *Working Paper*.

Banerjee, A., Chandrasekhar, A., Duflo, E., & Jackson, M. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1–7. https://doi.org/10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, *86*(6), 2453–2490. https://doi.org/10.1093/restud/rdz008

Bansal, S., Read, J., Pourbohloul, B., & Meyers, L. A. (2010). The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, *4*(5), 478–489. https://doi.org/10.1080/17513758.2010.503376

Beaman, L. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *Review of Economic Studies*, *79*(1), 128–161. https://doi.org/10.1093/restud/rdr017

Begelfor, E., & Werman, M. (2005). *The world is not always flot or learning curved manifolds* [Tech. Rep.]. School of Engineering and Computer Science, Hebrew University of Jerusalem. 3, 8.

Belton, A., Guillot, D., Khare, A., & Putinar, M. (2019). A panorama of positivity. I: Dimension free. In *Analysis of operators on function spaces* (pp. 117–165). SpringerLink.

Breza, E., & Chandrasekhar, A. G. (2019). Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica*, 87(1), 175–216. https://doi.org/10.3982/ECTA13683

Breza, E., Chandrasekhar, A. G., McCormick, T. H., & Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8), 2454–2484. https://doi.org/10.1257/aer.20170861

Cai, J., & Szeidl, A. (2017). Interfirm relationships and business performance. *Quarterly Journal of Economics*, 133(3), 1229–1282. https://doi.org/10.1093/qje/qjx049

Calvo-Armengol, A. (2004). Job contact networks. *Journal of Economic Theory*, 115(1), 191–206.

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4), 1239–1267.

Chandrasekhar, A., & Lewis, R. (2016). Econometrics of sampled networks. Stanford Working Paper.

Chaney, T. (2014). The network structure of international trade. *American Economic Review*, 104(11), 3600–3634. https://doi.org/10.1257/aer.104.11.3600

Chatterjee, S., Diaconis, P., & Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4), 1400–1435. https://doi.org/10.1214/10-AAP728

Cho, Y.-S., Ver Steeg, G., Ferrara, E., & Galstyan, A. (2016). Latent space model for multi-modal social data. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 447–458). International World Wide Web Conferences Steering Committee.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120. https://doi.org/10.1086/228943

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4), 1003–1045. https://doi.org/10.3982/ECTA7528

David, L. (Writer), Seinfeld, J. (Writer), Mandel, D. (Writer), & Ackerman, A. (Director). (1995). The pool guy (Season 7, Episode 8) [Television series episode]. In L. David & H. West (Executive Producers), *Seinfeld*. West-Shapiro Productions.

de Lange, S. C., de Reus, M. A., & van den Heuvel, M. P. (2014). The Laplacian spectrum of neural networks. *Frontiers in Computational Neuroscience*, 7, 189. https://doi.org/10.3389/fncom.2013.00189

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology*, 116(4), 1234–1283. https://doi.org/10.1086/659100

Eaton, M. L., & Taylor, D. E. (1991). On Weilandt's inequality and its application to the asymptotic distribution of eigenvalues of a random symmetric matrix. *Annals of Statistics*, 19(1), 260–271. https://doi.org/10.1214/aos/1176347980

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10), 3115–3153. https://doi.org/10.1257/aer.104.10.3115

Gai, P., & Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2120), 2401–2423. https://doi.org/10.1098/rspa.2009.0410

Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., & Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, 725–750.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4), 1033–1063. https://doi.org/10.3982/ECTA12679

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6), 1360–1380. https://doi.org/10.1086/225469

Grimmett, G. R., & McDiarmid, C. J. H. (1975). On colouring random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 77(2), 313–324. https://doi.org/10.1017/S0305004100051124

Gu, A., Sala, F., Gunel, B., & Ré, C. (2019). Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*.

Handcock, M. S., & Jones, J. H. (2004). Likelihood-based inference for stochastic models of sexual network formation. *Theoretical Population Biology*, 65(4), 413–422. https://doi.org/10.1016/j.tpb.2003.09.006

Heath, R. (2018). Why do firms hire using referrals? Evidence from Bangladeshi garment factories. *Journal of Political Economy*, 126(4), 1691–1746. https://doi.org/10.1086/697903

Hoff, P., Raftery, A., & Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098. https://doi.org/10.1198/016214502388618906

Jackson, M. (2013). Unraveling peers and peer effects: Comments on goldsmith-pinkham and imbens' "social networks and the identification of peer effects". *Journal of Business and Economic Statistics*, 31(3), 270–273. https://doi.org/10.1080/07350015.2013.794095

Jackson, M., & Lopez-Pintado, D. (2013). Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(1), 49–67. https://doi.org/10.1017/nws.2012.7

Kaiser, M., & Hilgetag, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Computational Biology*, 2(7), e95. https://doi.org/10.1371/journal.pcbi.0020095

Karwowski, W., Vasheghani Farahani, F., & Lighthall, N. (2019). Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review. *Frontiers in Neuroscience*, 13, 585.

Killing, W. (1891). Ueber die Clifford-Klein'schen Raumformen. *Mathematische Annalen*, 39(2), 257–278. https://doi.org/10.1007/BF01206655

Killworth, P. D., Johnsen, E. C., Bernard, H. R., Shelley, G. A., & McCarty, C. (1990). Estimating the size of personal networks. *Social Networks*, 12(4), 289–312. https://doi.org/10.1016/0378-8733(90)90012-X

Kinnan, C., & Townsend, R. (2012). Kinship and financial networks, formal financial access, and risk reduction. *The American Economic Review*, 102(3), 289–293. https://doi.org/10.1257/aer.102.3.289

Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., & Boguná, M. (2010). Hyperbolic geometry of complex networks. *Physical Review E*, 82(3), 036106. https://doi.org/10.1103/PhysRevE.82.036106

Langston, M. A., Zhang, Y., Chesler, E. J., Samatova, N. F., Baldwin, N. E., & Abu-Khzam, F. N. (2005). Genome-scale computational approaches to memory-intensive applications in systems biology. In *SC Conference*, IEEE Computer Society, 12.

Leung, M. C., Williams, P. L., Benedetto, A., Au, C., Helmcke, K. J., Aschner, M., & Meyer, J. N. (2008). Caenorhabditis elegans: An emerging model in biomedical and environmental toxicology. *Toxicological Sciences*, 106(1), 5–28. https://doi.org/10.1093/toxsci/kfn121

Lubold, S., Liu, B., & McCormick, T. H. (2021). Spectral goodness-of-fit tests for complete and partial network data. arXiv preprint arXiv:2106.09702.

Luo, W., & Li, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, 103(4), 875–887. https://doi.org/10.1093/biomet/asw051

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., & Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human Organization*, 60(1), 28–39. https://doi.org/10.17730/humo.60.1.efx5t9gjtgmga73y

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. https://doi.org/10.1080/01621459.2014.991395

Myers, S. A., & Leskovec, J. (2014). The bursty dynamics of the twitter information network. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 913–924). Association for Computing Machinery.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4, 2111–2245.

O'Neill, B. (1983). *Semi-Riemannian geometry with applications to relativity*, vol. 103. Academic Press.

Orbanz, P., & Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2), 437–461. https://doi.org/10.1109/TPAMI.2014.2334607

Ouadah, S., Robin, S., & Latouche, P. (2019). 'Degree-based goodness-of-fit tests for heterogeneous random graph models: Independent and exchangeable cases', arXiv, arXiv: 1507.08140, preprint: not peer reviewed.

Petersen, S. E., & Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron*, 88(1), 207–219. https://doi.org/10.1016/j.neuron.2015.09.027

Politis, D. N., & Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22(4), 2031–2050. https://doi.org/10.1214/aos/1176325770

Recanatesi, S., Farrell, M., Lajoie, G., Deneve, S., Rigotti, M., & Shea-Brown, E. (2019). Predictive learning extracts latent space representations from sensory observations. *bioRxiv*, 471987.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 695–704). Association for Computing Machinery.

Salter-Townshend, M., & McCormick, T. H. (2017). Latent space models for multiview network data. *The Annals of Applied Statistics*, 11(3), 1217. https://doi.org/10.1214/16-AOAS955

Schoenberg, I. J. (1935). Remarks to Maurice Frechet's Article "Sur La Definition Axiomatique D'Une Classe D'Espace Distances Vectoriellement Applicable Sur L'Espace De Hilbert. *Annals of Mathematics*, 36(3), 724–732. https://doi.org/10.2307/1968654

Sewell, D. K., & Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512), 1646–1657. https://doi.org/10.1080/01621459.2014.988214

Shalizi, C. R., & Asta, D. (2017). 'Consistency of maximum likelihood for continuous-space network models', arXiv, arXiv:1711.02123, preprint: not peer reviewed.

Smith, A. L., Asta, D. M., & Calder, C. A. (2019). The geometry of continuous latent space models for network data. *Statistical Science*, 34(3), 428–453. https://doi.org/10.1214/19-STS702

VanRullen, R., & Reddy, L. (2019). Reconstructing faces from fMRI patterns using deep generative neural networks. *Communications Biology*, *2*(1), 1–10. https://doi.org/10.1038/s42003-019-0438-y

Weber, M., & Nickel, M. (2018). Curvature and representation learning: Identifying embedding spaces for relational data. In *NeurIPS Relational Representation Learning*.

Wilkins-Reeves, S., & McCormick, T. (2022). 'Asymptotically normal estimation of local latent network curvature', arXiv, arXiv:2211.11673, preprint: not peer reviewed.

Wilson, R. C., Handcock, E. R., Pekalska, E., & Duin, R. P. (2014). Spherical and hyperbolic embeddings of data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(11), 2255–2269. https://doi.org/10.1109/TPAMI.2014.2316836

Xu, W., & Reinert, G. (2021). A Stein goodness-of-fit test for exponential random graph models. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*. PMLR.

Yanchenko, A. K., & Hoff, P. D. (2020). Hierarchical multidimensional scaling for the comparison of musical performance styles. *The Annals of Applied Statistics*, *14*(4), 1581–1603. https://doi.org/10.1214/20-AOAS1391

Yu, X., & Rodriguez, A. (2019). Spatial voting models in circular spaces: A case study of the U.S. House of Representatives. *Annals of Applied Statistics*, *15*(4), 1897–1922.

Zoubouloglou, P., García-Portugués, E., & Marron, J. S. (2021). 'Scaled torus principal component analysis', arXiv, arXiv:2110.04758, preprint: not peer reviewed.