

IDENTIFYING THE LATENT SPACE GEOMETRY OF NETWORK MODELS THROUGH ANALYSIS OF CURVATURE

SHANE LUBOLD[§], ARUN G. CHANDRASEKHAR[‡], AND TYLER H. MCCORMICK^{§,*}

ABSTRACT. Statistically modeling networks, across numerous disciplines and contexts, is fundamentally challenging because of (often high-order) dependence between connections. A common approach assigns each person in the graph to a position on a low-dimensional manifold. Distance between individuals in this (latent) space is inversely proportional to the likelihood of forming a connection. The choice of the latent geometry (the manifold class, dimension, and curvature) has consequential impacts on the substantive conclusions of the model. More positive curvature in the manifold, for example, encourages more and tighter communities; negative curvature induces repulsion among nodes. Currently, however, the choice of the latent geometry is an *a priori* modeling assumption and there is limited guidance about how to make these choices in a data-driven way. In this work, we present a method to consistently estimate the manifold type, dimension, and curvature from an empirically relevant class of latent spaces: simply connected, complete Riemannian manifolds of constant curvature. Our core insight comes by representing the graph as a noisy distance matrix based on the ties between cliques. Leveraging results from statistical geometry, we develop hypothesis tests to determine whether the observed distances could plausibly be embedded isometrically in each of the candidate geometries. We explore the accuracy of our approach with simulations and then apply our approach to data-sets from economics and sociology as well as neuroscience.

1. INTRODUCTION

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

Date: January 7, 2021.

We thank Eric Auerbach, Abhijit Banerjee, Emily Breza, Jacob Burchard, Gabriel Carroll, James Evans, Bailey Fosdick, Jeremy Fox, Paul Goldsmith-Pinkham, Ben Golub, Matthew Grant, Fang Han, Rachel Heath, Yunmi Kong, Mengjie Pan, Mallesh Pai, Abel Rodriguez, Anna Smith, Xun Tang, Matt Thirkettle, Aravindan Vijayaraghavan, and Jon Wellner. We thank participants at the 2020 Joint Statistical Meeting, CANSSI-Ontario Data Science Applied Research and Education Seminar, IDEAL (Institute for Data, Econometrics, Algorithms, and Learning), the Joint Econometrics and Statistics Seminar Series at Cornell University, UC-Davis (Statistics), Bocconi University (Statistics), and Rice (Applied Micro and Econometrics).

[‡] Department of Economics, Stanford University; NBER; JPAL.

[§] Department of Statistics, University of Washington.

* Department of Sociology, University of Washington.

settings in which networks play a major role. Modeling network formation is, therefore, essential for both descriptive and counterfactual analyses.

Constructing such models is, however, challenging from a statistical perspective since networks typically feature higher-order dependence between the connections. Phenomena such as transitivity and the tendency for a friend of a friend to be a friend are common in networks and mean that standard statistical approaches, which assume independence across connections, aren't appropriate. One common approach for modeling this dependence structure is the latent space model, introduced by (Hoff, Raftery, and Handcock, 2002). The latent space model estimates a probability distribution over graphs that is consistent with the single instance of the graph observed in practice. It assigns each actor in the network to a position on a low-dimensional manifold. Likelihood of a connection is inversely proportional to distance between actors on a manifold with a pre-specified dimension and geometry. Connections are assumed independent conditional on the latent positions.

In this paper, we address the question of how to choose the type and dimension of the manifold in latent space network models. We present a hypothesis testing framework which connects distances in the latent space to feasible embeddings on simply connected, complete Riemannian manifolds with constant curvature. The intuition is as follows. First, we must characterize each graph in a manner that is agnostic to manifold type (otherwise we assume the result), but is interpretable on each manifold (otherwise we cannot compare). Given a way to characterize the graph, we can construct a test statistic and perform inference. Distance is a natural candidate for such a characterization, particularly given the extensive literature on representing distance matrices on manifolds. Distance is typically defined *given* a particular geometry, however. Our solution is to define distance based on interactions between cliques, or completely connected subgraphs. We define the distance between two cliques as probability of interaction of a node in clique A with a node in clique B , which we can define using the definition of the latent space model plus the observed fraction of realized links between clique members out of the total possible. Given a distance matrix, we rely on three main results from geometry. The first two results give conditions for embedding distance matrices isometrically in a given manifold type and the third provides an estimator of the smallest possible dimension for this embedding. To characterize uncertainty, we develop hypothesis tests for candidate geometries. These tests describe how different we expect contact patterns between cliques to be when seeing single draws from a probability model based on the latent space. We show, theoretically, that our procedure consistently identifies the latent manifold and, empirically, that the method has appealing properties in simulations and consequential implications using two datasets.

The choice of latent manifold type and its curvature is consequential for both the interpretation of the latent space model and its theoretical properties. First, the choice of the geometry 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 isn't 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 spherical space wraps, there is an upper bound to how far apart nodes can be from one another before they start getting closer together. Further, positive curvature 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 spaces—indeed infinite trees cannot be—but fit naturally in hyperbolic space. Recent work in statistics has shown the importance of modeling 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. [Smith et al. \(2019\)](#) provides 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. A question open until recently was on the consistency of estimates of the individual locations on the unobserved manifold. This was addressed by [Shalizi and Asta \(2017\)](#) for a general class of models who demonstrate that latent locations are consistently estimated in the case where the researcher ex-ante assumes a manifold type within a class of rigid manifolds and there are no fixed effects nor covariates. [Breza et al. \(2019\)](#) extend this to the general case including fixed effects and covariates. However, since the distribution of the network formation process depends on the manifold itself, the key question is whether a researcher can consistently estimate the latent space. After all, the network formation process is sensitive to the geometry of the latent space. We study and prove consistent estimation of the latent space geometry in the present paper.

As previously mentioned, it is common practice to assume the latent dimension, the manifold type, or both. We provide a data-driven alternative. It also contrasts with cross-validation based selection procedures that are sometimes used, in particular to estimate the manifold dimension. These approaches subsample connections and then use either model fit diagnostics or out-of-sample prediction metrics. We avoid a critical issue with these approaches, namely that subsampling can fundamentally alter graph properties in unpredictable ways ([Chandrasekhar and Lewis, 2016](#)), calling into question the relevance of the subsampled distribution.

We also 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 limited, but non-trivial, curvature may perform well for a graph generated from Euclidean embeddings, but we can improve upon this. In our approach we isolate uncertainty to only sampling error by using a test for isometric embeddings of distances into the space under consideration. In our context we 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 study 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. In our framework 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 close this section with a formal definition of the latent space model and an overview of the structure of the remainder of the paper. Beginning with the latent space model, consider the graph $G = (V, E)$ where $n = |V|$ are actors (also called individuals or nodes) and E are edges (also called links or connections). 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

$$(1) \quad \mathbb{P}(G_{ij} = 1 \mid \nu, z, \mathcal{M}^p(\kappa)) = \exp(\nu_i + \nu_j - d_{\mathcal{M}^p(\kappa)}(z_i, z_j)),$$

which is the probability model for the graph. In (1) we have $G_{ij} = 1$ if there's a connection between i and j and 0 otherwise. We represent $\nu = (\nu_1, \dots, \nu_n)$ as the vector of individual effects, restricted to be negative to ensure Equation (1) is a probability. These are independent effects (e.g., they can 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 . We assume that the space is a simply connected, complete Riemannian manifold of constant curvature. By the Killing-Hoff theorem (Killing, 1891), there are only three such manifolds: Euclidean space (\mathbb{R}^p equipped with the usual inner product) with $\kappa = 0$, the p -sphere ($\mathbf{S}^p(\kappa)$) with strictly positive curvature, and hyperbolic space ($\mathbf{H}^p(\kappa)$) with strictly negative curvature. We present additional background on these geometries in Section 1.1. The propensity to form ties between pairs of individuals is assumed independent conditional on the vector $z = (z_1, \dots, z_n)$ of these positions on the manifold. Hoff et al. (2002) present this model with a logit link and subsequent authors have extended the mapping from latent space distances to probability using both univariate and multivariate distributions (see Salter-Townshend and McCormick (2017) for example). We choose exponentiation to simplify the decomposition between the ν and z components, though our results are robust to the choice of the link function. We assume the following throughout the paper.

Assumption 1.1 means (by Killing (1891)) that the geometry must be Euclidean, spherical, or hyperbolic with a bounded dimension and possible curvatures in some compact set. This condition implies that, as we accumulate more data, the manifold must meaningfully curve (in that it is not arbitrarily close to Euclidean).

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

Assumption 1.2 ensures that (1) produces probabilities.

ASSUMPTION 1.2. Every node i has a fixed-effect ν_i i.i.d. from a distribution F_ν on $(-\infty, 0]$.

Before we proceed, we use C_ℓ to denote an ℓ clique in G . That is, C_ℓ is a complete subgraph on ℓ nodes. Our results relies on latent positions of nodes in completely connected subgraphs, or cliques. Under any model for F_z , nodes in cliques will be close together in the

latent space, since cliques are completely connected subgraphs and nodes close together in the latent space are likely to connect.

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 of local cliques captures a typical feature of latent space models and empirical data.

ASSUMPTION 1.3. *Every node i resides at a location z_i that is drawn independently and are identically distributed from a distribution F_z on manifold $\mathcal{M}^p(\kappa)$. The latent location distribution must satisfy two properties:*

- (a) Identifiability: *The support of F_z must consist of at least $K > p$ distinct points that uniquely identify the manifold, with $0 \leq p < K \cdot (1 - \sqrt{\frac{1+K}{2K}})$.*
- (b) Local cliques: *The distribution of locations should be such that the following is true. For any collection of ℓ nodes with locations drawn i.i.d. from F_z we have $\forall \delta > 0$,*

$$\lim_{n, \ell \rightarrow \infty} \left(\frac{\mathbb{P}(C_\ell \text{ exists} \mid \max_{i,j} d(z_i, z_j) \leq \delta)}{\mathbb{P}(C_\ell \text{ exists} \mid \max_{i,j} d(z_i, z_j) > \delta)} \times \frac{\mathbb{P}(\max_{i,j} d(z_i, z_j) \leq \delta)}{\mathbb{P}(\max_{i,j} d(z_i, z_j) > \delta)} \right) = \infty.$$

This provides an explicit condition for the rate at which nodes in a clique should near one another as the size of the graph grows. Aside from this condition, we impose no restrictions on the distribution of F_z . This general requirement allows for continuous, discrete, or mixed distributions as well as dependence on n .

To give one expository example of a distribution that has this property and does not depend on n , consider a sphere with K locations. Say these locations are distributed evenly about sphere such that they satisfy part (a) of Assumption 1.3 and all nodes are sampled i.i.d. uniformly distributed over these locations. With this distribution, nodes that are connected in clique will, with high probability be placed at the same location relative to being at several locations. When we consider a sufficiently large clique—in practice even as few as five nodes works well as this yields ten rare events (potential links)—the overwhelming likelihood is that we see nodes at a single location, rather than spread across several locations, which is the requirement for Assumption 1.3. Thus, cross-clique links allow us to estimate cross-location probabilities for these K locations, which play the major role in studying isometric embedding conditions used to determine the geometry.

The remainder of the paper is organized as follows. In Section 1.1 we review key concepts of curvature and embedding from geometry which are crucial for our testing procedure. We present a strategy for creating distance matrices that are manifold-type agnostic in Section 2. Next, we present two uses for these distance matrices, estimating the minimal latent dimension (Section 3) and testing for geometric class (Section 4) We also show that the estimators for minimal dimension and the hypothesis tests are consistent. In Section 5 we present simulation experiments that explore the efficacy of our approach. In Section 6, we apply our results to two empirical examples. The first empirical example considers data from Indian village social networks, comprised of informal finance, information, and social links and also study how the introduction of microfinance impacts geometry. The second example focuses on the neural network of a worm. Section 7 concludes. Code to reproduce our results is available at github.com/slubold/LS_Geometry.

1.1. Preliminaries: A Review of Core Geometric Results.

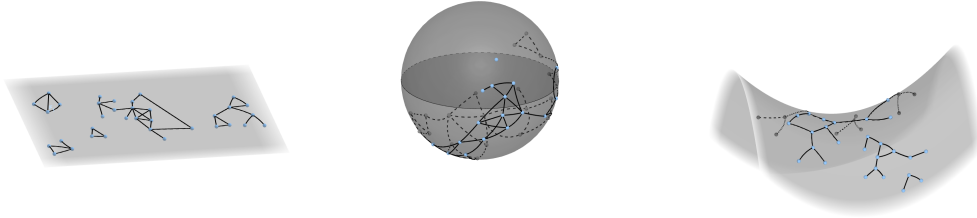


FIGURE 1. How curved geometries affect network embeddings ($n = 36$).

1.1.1. *Sectional Curvature.* We will study geometries given by simply connected, complete Riemannian manifolds of constant sectional curvature, defined formally below. Each of these assumptions are *ex ante* parsimonious. Simple connectedness and completeness are innocuous as well and constant curvature at least provides a place to start and nests all manifolds used in the literature.

The Killing-Hoff theorem in [Killing \(1891\)](#) states that any simply connected complete Riemannian manifold of constant sectional curvature is either Euclidean, spherical, or hyperbolic. This restricts our class to these manifolds. 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 which has constant positive curvature does force such behavior. 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.

1.1.2. *Sectional Curvature Definitions.* 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. 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 $f : \mathcal{M}^p \rightarrow \mathbb{R}$ to $v(f(m)) \in \mathbb{R}$ that is \mathbb{R} -linear and Leibnizian.¹

A Riemannian manifold (\mathcal{M}^p, g) comes with a *metric tensor* g which at every point $m \in \mathcal{M}^p$ takes two vectors in the tangent space 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.

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 $R_m(u, v)w \in T_m(\mathcal{M}^p)$ ²

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

¹An 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.

²Here $[\cdot, \cdot]$ is the usual Lie bracket.

Here is a simple 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

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

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 $\mathcal{M}^p(\kappa)$.

1.1.3. *Candidate Geometries.* In order to study the candidate manifold $\mathcal{M}^p(\kappa)$, we embed them in \mathbb{R}^{p+1} . Clearly the Euclidean case is trivial. 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, given by Q , 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:³

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

This implies a way of calculating distances between points on the manifold with

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

Let us turn to our candidate cases. In the case of the sphere \mathbf{S}^p , we have the usual Euclidean inner product $Q_{\mathbb{R}^{p+1}}(x, y) := \sum_{i=1}^{p+1} x_i y_i$.⁴ The locus of points and distances between two points $x, y \in \mathbb{R}^{p+1}$ for the embedding is

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

Hyperbolic space \mathbf{H}^p is embedded in Minkowski space, $\mathbb{R}^{1,p}$ which is \mathbb{R}^{p+1} equipped with the Minkowski pseudo-metric: $Q_{\mathbb{R}^{1,p}}(x, y) := -x_0 y_0 + \sum_{i=1}^p x_i y_i$.⁵ The important point is that the signature is distinguished from the Euclidean space which will play a key role in

³For the hyperboloid $x_0 > 0$ is an additional restriction.

⁴It can be checked that the metric tensor at a point $x \in \mathbf{S}^p$ is induced by the ambient Euclidean space: $g_x^{\mathbf{S}^p} := Q_{\mathbb{R}^{p+1}}|_{T_x(\mathbf{S}^p)}$ or $g_x^{\mathbf{S}^p}(u, v) = \sum_i u_i v_i$ for $u, v \in T_x(\mathbf{S}^p)$.

⁵One can check that across the tangent bundle at all points this indefinite inner product in the ambient space is positive definite, thereby defining a Riemannian metric: $g_x^{\mathbf{H}^p} := Q_{\mathbb{R}^{p,1}}|_{T_x(\mathbf{H}^p)}$ or $g_x^{\mathbf{H}^p}(u, v) = -u_0 v_0 + \sum_{i=1}^p u_i v_i$ for $u, v \in T_x(\mathbf{H}^p)$.

distinguishing the geometries. As a consequence the locus of points and distances are given by

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

and

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

1.1.4. Embedding Conditions. Let D be a known distance matrix from K points given by $Z = \{z_1, \dots, z_K\}$. 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' , $d_{\mathcal{M}}(\phi(z_l), \phi(z_{l'})) = d_{ll'}$.

In our case, we want to study if Z can be isometrically embedded in some manifold $\mathcal{M}^p(\kappa)$ satisfying Assumption 1.1 and, specifically, given the class of the manifold (Euclidean, spherical, or hyperbolic), the minimal dimension required for the embedding. As such, we review how one determines if it is possible to isometrically embed these points into Euclidean space or any of the curved spaces.

We define the $K \times K$ matrix corresponding to the bilinear form above,

$$W_\kappa(D) = \frac{1}{\kappa} \cos(\sqrt{\kappa}D) \quad \text{if } \kappa \neq 0$$

where we apply the cosine operation element-wise, as before. We use the convention that for $\kappa = 0$, $W_0(D) = -\frac{1}{2}JD \circ DJ$, where $J := I_K - \frac{1}{K}1_K1_K^T$. By using a Taylor series of $W_\kappa(D)$ around $\kappa = 0$, one can see that there is a close relationship between the expression of $W_\kappa(D)$ for $\kappa > 0$ and $W_0(D)$.⁶ The motivation for these expressions is that the signatures of these matrices will tell us whether such an isometric embedding is possible. We write $W_\kappa = W_\kappa(D)$, suppressing the dependency on D unless otherwise noted.

First we review the Euclidean case.

PROPOSITION 1.1 (Schoenberg (1935), Theorem 1). $Z \xrightarrow{\text{isom}} \mathbb{R}^p$ for some p if and only if W_0 is positive semi-definite. In addition, the smallest p such that $Z \xrightarrow{\text{isom}} \mathbb{R}^p$ is $p = \text{rank}(W_0)$.

Next we turn to the case of curved manifolds. Recall that the *signature* of a square matrix A is a triple (a, b, c) , where a, b, c are respectively the number of positive, zero, and negative eigenvalues of A .

Returning to our general formulation above, we can write distance matrix among a vector of points $Z \in \mathcal{M}^p(\kappa)$ as

$$D_{\mathcal{M}^p(\kappa)} = \frac{\arccos(\kappa \phi(Z)' \eta \phi(Z))}{\sqrt{\kappa}}$$

where we have written the general bilinear form as $Q = \phi(Z)' \eta \phi(Z)$. Notice that $\eta = \text{diag}(+1, \dots, +1, 0, \dots, 0)$ in the case of spherical geometry and $\eta = \text{diag}(-1, 0, \dots, 0, +1, \dots, +1)$ in the case of hyperbolic geometry. The logic of the approach is to identify the signature which then tells us which geometry we are in and precisely why W_κ takes its form for $\kappa \neq 0$.

Formally, the following result from Begelfor and Werman (2005) provides necessary and sufficient conditions for points to be isometrically embedded in spherical and hyperbolic space.

⁶We would like to thank Gabriel Carroll for pointing this out to us.

PROPOSITION 1.2 (Begelfor and Werman (2005), Theorem 1). $Z \xrightarrow{\text{isom}} \mathbf{S}^p(\kappa)$ if and only if the signature of W_κ is $(a, n - a, 0)$ for some $a \leq p + 1$ and some $\kappa > 0$. Also, $Z \xrightarrow{\text{isom}} \mathbf{H}^p(\kappa)$ if and only if the signature of W_κ is $(1, n - a - 1, a)$ for some $a \leq p$ and some $\kappa < 0$.

We want the smallest dimension such that Z can be embedded isometrically in $\mathcal{M}^p(\kappa)$. In the Euclidean case, the smallest dimension is $\text{rank}(W_0)$. For the spherical case, any admissible dimension p must satisfy $n_+(W_\kappa) \leq p + 1$, where $n_+(W_\kappa)$ is the number of positive eigenvalues of W_κ . So p must satisfy $p \geq n_+(W_\kappa) - 1$ and the smallest valid value of $p = n_+(W_\kappa) - 1$ for some $\kappa > 0$. Using a similar argument, one can show that the smallest p such that $Z \xrightarrow{\text{isom}} \mathbf{H}^p(\kappa)$ is $p = n_-(W_\kappa) + 1$ for some $\kappa < 0$ where $n_-(W_\kappa)$ is the number of negative eigenvalues of W_κ .

Taken together, these two results allow us to determine when a collection of points can be isometrically embedded in a Euclidean, spherical, or hyperbolic space. In addition, we can determine the smallest possible dimension of the space. To preview how we will use the results from this section, suppose for now that we have access to distances between points on $\mathcal{M}^p(\kappa)$, the latent space in (1). Our approach connects the eigenstructure of the distance matrix to an underlying manifold, leveraging the results from this section which indicate that only distance matrices with certain eigenstructure can be isometrically embedded on a particular manifold. We care about positive definiteness for the Euclidean and spherical cases (for W_0 and W_κ for $\kappa > 0$ respectively). For the hyperbolic case we have a more peculiar requirement: that the spectrum only has one positive eigenvalue and the remainder are zero and negative. Further, in all cases we care about rank of W_κ , which indicates the (smallest possible) dimension of $\mathcal{M}^p(\kappa)$.

In practice, of course, we must use the graph G to estimate a set of distances on $\mathcal{M}^p(\kappa)$. Using this estimator, which we describe in the next section, we can then provide estimators of the curvature of the space, provide a hypothesis testing framework to classify the geometry (Euclidean, spherical, or hyperbolic), and provide an estimator of the dimension of the manifold.

2. MAPPING GRAPHS TO MANIFOLDS VIA DISTANCE MATRICES

We now move from a general discussion of a geometry in the previous section to the specific case of the graph. We first show how we define a distance matrix based on connections between cliques. With an appropriately defined distance matrix, the geometric results from the prior section can be used to generate hypothesis tests for latent geometry type and estimates for the dimension of this geometry. With sufficient data we could estimate the distance matrix with arbitrarily small noise and apply the results from the previous section directly. In practice, however, we see a fixed graph size and face sampling uncertainty about the distance matrix.

With these assumptions in mind, we now move to defining the distance matrix, D and its estimator, \hat{D} . In the typical setup for the latent space model, D is only defined *after* assuming a particular geometry. In this setting, the definition of D changes depending on the presumed geometry (arc-length on a hyper sphere for example, Euclidean distance on a plane). Since we do not want to take the geometry as given, the distance matrix cannot use a metric that is specific to a single geometry. The distance matrix also must be comparable across geometries, however, otherwise we will be performing different tests in each geometry. Our solution is to use probabilities defined by the latent space model in

(1). Rather than using distance on a given manifold as the input to these probabilities, as would be typical in the latent space model, we use overlap between cliques. In particular, we define the fraction of potential interactions between two clique members out of the total possible interactions. A potential consideration with this approach is that it requires the existence of cliques, a requirement not put forth in our assumptions. The existence of cliques is guaranteed by the latent space model under our assumptions as the number of nodes increases, however. In particular, the conditional independence relation that is key to the latent space model requires an assumption of exchangeability. The Aldous-Hoover Theorem implies that exchangeable sequences of nodes correspond to dense graphs in the limit (Aldous, 1981; Orbanz and Roy, 2015), which implies that cliques are present in the limit. We also examine the existence of cliques in practice using our empirical and simulated examples. We find that, in general 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.

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. First, suppose there are no individual effects (so $\nu_i = 0 \forall i$). Second, suppose that the distribution of node locations are degenerate distributions fixed at K locations z_1, \dots, z_K on $\mathcal{M}^p(\kappa)$. That is, all nodes occupy one of K locations in the latent space. Define $V_k = \{j \in \{1, \dots, n\} : z_j = z_k\}$ to be the set of nodes at location z_k . Under this simplification, we can write the distance between points z_k and $z_{k'}$ using the definition of the latent space model in (1) as

$$(2) \quad d_{k,k'} = -\log(p_{k,k'})$$

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

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

In words, this estimator counts the number of observed edges between z_k and $z_{k'}$ and divides by the number of possible edges, given by $|V_k||V_{k'}|$. Since $G_{ij} \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p_{k,k'})$ for $(i, j) \in V_k \times V_{k'}$, this estimator is unbiased for $p_{k,k'}$. In addition, supposing that $|V_i| \rightarrow \infty$ as $n \rightarrow \infty$, the weak law of large numbers implies that $\hat{p}_{k,k'} \xrightarrow{P} p_{k,k'}$. By (2), we can estimate the distance between z_k and $z_{k'}$ by $\hat{d}_{k,k'} = -\log(\hat{p}_{k,k'})$. We can then apply the continuous mapping theorem to show that $\hat{d}_{k,k'} \xrightarrow{P} d_{k,k'}$. To summarize, we have used the edges in G to estimate the connections between locations on the unobserved latent space $\mathcal{M}^p(\kappa)$ and then used the graph model to estimate distances.

We now return to the original model in (1) and relax the two simplifying assumptions made above. The individual effects describe heterogeneity in the propensity for an individual to form connections and are not directly related to which connections they will form, which is what the latent space captures. We would prefer, therefore, 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 ν_i is $(-\infty, 0]$, we integrate to find

that

$$\begin{aligned} \mathbb{P}(G_{ij} = 1|z, \mathcal{M}^p(\kappa)) &= \int_{-\infty}^0 \int_{-\infty}^0 \mathbb{P}(G_{ij} = 1|z, \nu_i, \nu_j, \mathcal{M}^p(\kappa)) dF(\nu_i) dF(\nu_j) \\ &= \int_{-\infty}^0 \int_{-\infty}^0 \exp(\nu_i + \nu_j - d(z_i, z_j)) dF(\nu_i) dF(\nu_j) \\ &= \exp(-d(z_i, z_j)) \int_{-\infty}^0 \int_{-\infty}^0 \exp(\nu_i + \nu_j) dF(\nu_i) dF(\nu_j) \end{aligned}$$

where the last line uses the fact that $\exp(x + y) = \exp(x) \exp(y)$. Since the fixed effects ν_i are i.i.d., we simplify this last double integral so that

$$\int_{-\infty}^0 \int_{-\infty}^0 \exp(\nu_i + \nu_j) dF(\nu_i) dF(\nu_j) = E(\exp(\nu_i))^2 .$$

We therefore conclude that, after marginalizing the individual effects,

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

We now solve for the distance $d(z_i, z_j)$ to conclude that

$$(3) \quad d_{k,k'} = -\log(p_{k,k'}) + 2 \log(E(\exp(\nu_i))) .$$

Note that if $\nu_i = 0$ with probability 1, which we assumed in the simplified model above, then $E(\exp(\nu)) = 1$, so that (3) becomes (2). We must now estimate (i) the term $p_{k,k}$ and (ii) the term $\log(E(\exp(\nu)))$.

We first discuss how to estimate the term $p_{k,k}$. Under (1), individuals have distinct locations in the latent space and we only observe one instance of a connection (or lack thereof) between individuals, so the strategy from the simplified model work with individual position. Instead, we focus on cliques in the graph, G . An ℓ -clique C of a graph $G = (V, E)$ is a set of ℓ nodes such that all possible edges between pairs of nodes in C exist. That is, C is a complete sub-graph of G consisting of ℓ nodes. In the formation model in (1), individuals who are close together in the latent space are likely to form ties. Since cliques are connected we know that these individuals have formed ties, and, thus, will have similar latent locations. More formally, suppose that $\{\tilde{z}_1, \dots, \tilde{z}_\ell\}$ are the latent space locations of ℓ points in an ℓ -clique. Then, it is likely that each $d(\tilde{z}_i, \tilde{z}_j)$ is *small*. If this were not the case (meaning that the locations of the nodes were far apart), then it is unlikely that we would observe such a clique.

Let C_1, \dots, C_K denote a collection of ℓ -cliques. For each $k, k' \in \{1, \dots, K\}$, compute the $K \times K$ matrix $\hat{P} = \{\hat{p}_{k,k'}\}$ with

$$(4) \quad \hat{p}_{k,k'} = \frac{1}{\ell^2} \sum_{i,j=1}^n G_{i,j} \mathbf{1}\{i \in C_k, j \in C_{k'}\} ,$$

In words, this estimator counts the number of edges between cliques C_k and $C_{k'}$ and divides by the number of possible edges ℓ^2 . As previously discussed, the existence of cliques arises through the properties of the formation model in (1) as the number of nodes increases. Ideal cliques would be distinct, meaning that any node in the graph belongs to at most one clique, and complete ℓ -cliques as described above. In practice, however, we require neither of these and demonstrate empirically that the methods we propose maintain desirable properties with cliques that are partially overlapping and are not complete (meaning that most but not all

edges are realized).⁷ Further, for the convenience of notation we assume that all cliques are of size ℓ , though this is also not required.

We now discuss how to estimate the term $\log(E(\exp(\nu)))^2$. To construct such an estimator 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) and thus won't confound estimation of the individual effects. Since we use cliques to estimate $p_{k,k'}$, we may consider using these nodes to also estimate $\log(E(\exp(\nu)))^2$. However, this fails as 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 $I_k(t) := \{j : j \notin C_k, |C_k| > \sum_{i \in C_k} G_{i,j} \geq t\}$ 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 estimate the probability that nodes in $I_k(t)$ connect by

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

To estimate $E(\exp(\nu))$, we average the above term over all cliques, yielding

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

In practice, this approach will suffer from selection bias when the clique size is large. That is, by Assumption 1.2 all individuals have independent and identically distributed v_i 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.

3. ESTIMATING THE CURVATURE AND MINIMAL LATENT DIMENSION

Having defined a consistent estimator of the distance matrix, D , we now explore how to use this estimator in conjunction with the geometry results from Section 1.1. First, we address the question of choosing the dimension of the latent space. We assume here that the manifold type is given and present a test for manifold type in the next section. Propositions 1.1 and 1.2 show that the rank of W_κ tells us the dimension of the underlying space. Since W_κ depends on the curvature of the space, we first present an estimator of the curvature of the latent space and then discuss estimating the minimal dimension. Throughout we focus on estimating the minimal dimension in which points can be embedded isometrically, since points isometrically embedded in dimension p can be trivially embedded in dimension p' , with $p' > p$.

⁷Indeed, our argument extends to the case of subgraphs, where we are interested in cross-link probabilities across structures that are clique-like but have missing links, provided they are more likely to occur locally in the manifold.

We use Proposition 1.2 to motivate our estimate of the curvature of $\mathcal{M}^p(\kappa)$, noting that estimating curvature is required only for the spherical and hyperbolic cases. For illustration, suppose that a set of points Z can be embedded in $\mathbf{S}^p(\kappa_0)$ for some p . Suppose that the value κ_0 is unique. Then, for any $\kappa \neq \kappa_0$, it follows that $|\lambda_1(W_\kappa)| > 0$. This follows from Proposition 1.1, which says that for any $\kappa \neq \kappa_0$, it must be true that $\lambda_1(W_\kappa) \neq 0$, but $\lambda_1(W_{\kappa_0}) = 0$.

The true κ_0 , therefore, satisfies⁸

$$(5) \quad \kappa_0 = \arg \min_{\kappa \geq 0} \left| \lambda_1(\kappa W_\kappa) \right|.$$

In practice, we don't know D but we do have access to $\hat{D} \xrightarrow{p} D$, as described in the previous section. We can therefore use \hat{D} in place of D to estimate κ . This approach leads to the following estimator of the curvature:

$$\hat{\kappa}_S := \arg \min_{\kappa \in [a, b]} \left| \lambda_1(\kappa W_\kappa(\hat{D})) \right|, \quad \hat{\kappa}_H := \arg \min_{\kappa \in [-b, -a]} \left| \lambda_{K-1}(\kappa W_\kappa(\hat{D})) \right|$$

for some $0 < a < b$.

As the graph size approaches infinity the estimates approach the true curvature, as the following proposition shows, which we prove in Appendix A. In Appendix B, we discuss how to pick a and b in practice.

PROPOSITION 3.1. *Suppose that Z contains K points in either $\mathbf{S}^p(\kappa)$ or $\mathbf{H}^p(\kappa)$. Suppose also that \hat{D} is a $K \times K$ matrix containing pairwise distances between points in Z such that $\hat{D} \xrightarrow{p} D$. Finally, suppose that either*

- (1) *there is a unique $\kappa \in [a, b]$ such that $Z \xrightarrow{isom} \mathbf{S}^p(\kappa)$ for some p . Our estimate of the curvature is $\hat{\kappa}_n := \operatorname{argmin}_{\kappa \in [a, b]} \left| \lambda_1(\kappa W_\kappa(\hat{D})) \right|$,*
- (2) *or there is a unique $\kappa \in [-b, -a]$ such that $Z \xrightarrow{isom} \mathbf{H}^p(\kappa)$ for some p . Our estimate of the curvature is $\hat{\kappa}_n := \operatorname{argmin}_{\kappa \in [-b, -a]} \left| \lambda_{K-1}(\kappa W_\kappa(\hat{D})) \right|$.*

In cases (1) and (2), $\hat{\kappa}_n \xrightarrow{p} \kappa$ as $n \rightarrow \infty$.

The estimators we propose for curvature are similar to those proposed in Wilson et al. (2014), though Wilson et al. (2014) does not prove that these estimators are consistent. Wilson et al. (2014) does, however, also note difficulties in finite samples with the hyperbolic curvature estimates.

Given an estimate of the curvature, κ , we are now ready to compute \hat{W}_κ and, in doing so, estimate the minimal latent dimension. We want to estimate the rank of W_κ (or $W_\kappa^T W_\kappa$) since, by Proposition 1.2, this gives the minimal latent dimension. Robin and Smith (2000) provides a method to consistently estimate the dimension of the latent space, p , in our framework. Robin and Smith (2000) shows that, under certain conditions on the population matrix, the eigenvalues of the sampled version of the matrix converge to a weighted sum of independent chi-square random variables. Robin and Smith (2000) uses this fact to propose a consistent estimate of the matrix rank. This approach is appealing because it provides a consistent estimator of the rank; when applied to our problem, we can recover consistent estimates under our assumptions. Recent work, by Luo and Li (2016), has been shown to

⁸It is immaterial whether we rescale by κ .

have more appealing finite sample performance and in Appendix C we provide the algorithm to estimate the rank with this method.

4. TESTING FOR GEOMETRIC CLASS

We now derive a formal test for the geometric class of the latent manifold. We do this by leveraging the formation model and the graph to estimate a distance matrix \hat{D} with estimated curvature $\hat{\kappa}$, which we use to test hypotheses about the geometry of D . We first present two consistent hypothesis tests. One is general to any distance matrix, whether related to a social network or not, but requires an estimator of the distance matrix which goes in probability to the true matrix. For the second, we present a consistent test which is specific to the latent space model for networks. Rather than making any assumptions about the distance matrix, we show the test is consistent based on properties of the graph and latent space model.

To begin, consider the Euclidean case. We are interested in testing the following pair of hypotheses:

$$H_{0,e} = D \text{ is Euclidean, } H_{a,e} = D \text{ is not Euclidean.}$$

By the phrase “ D is Euclidean” we mean that $Z \xrightarrow{\text{isom}} \mathbb{R}^p$ for some p . By Proposition 1.1, we can write the above pair of hypotheses as the equivalent pair

$$(6) \quad H_{0,e} : \lambda_1(W_0) \geq 0, \quad H_{a,e} : \lambda_1(W_0) < 0.$$

Essentially, we use Proposition 1.1 to reframe the geometry requirement of isometric embedding into a testable statement about the eigenstructure of a function of the distance matrix. Using the same reasoning, we can use Proposition 1.2 to write the hypotheses that D is spherical for some $\kappa > 0$ as

$$(7) \quad H_{0,s} : \lambda_1(W_\kappa) \geq 0, \quad H_{a,s} : \lambda_1(W_\kappa) < 0.$$

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

$$(8) \quad H_{0,H} : \lambda_{K-1}(W_\kappa) = 0, \quad H_{a,K} : \lambda_{K-1}(W_\kappa) \neq 0.$$

By Proposition 1.2, concluding that in the hyperbolic case $\lambda_{K-1}(W_\kappa) = 0$ is not enough to conclude that D is hyperbolic, since we must test the first and smallest eigenvalues of W_κ too. In practice, we found that testing only one eigenvalue was typically sufficient and, thus, use this simpler test. It would also be possible to test all three eigenvalues using an intersection test, which we leave to future work.

Given this general setup, we now move to the case where the distance matrix is observed with noise and propose a series of tests for the hypotheses above. To test $H_{0,e}$, first, define a rejection region $\mathcal{R}_n = (-\infty, \delta_n]$ for some real-valued sequence $\delta_n \in (-\infty, 0]$ and we define our test $\phi_n(\hat{W}_0) \in \{0, 1\}$ as

$$(9) \quad \phi_n(\hat{W}_0) = \begin{cases} 0, & \lambda_1(\hat{W}_0) \in \mathcal{R}_n \\ 1, & \lambda_1(\hat{W}_0) \notin \mathcal{R}_n, \end{cases}$$

where 0 indicates that we reject $H_{0,e}$ and 1 indicates that we fail to reject $H_{0,e}$. The motivation for this approach is that if $\lambda_1(\hat{W}_0)$ is sufficiently far enough away from 0, then we can confidently say that $\lambda_1(W_0)$ is not zero.

When testing if D is spherical, let $\hat{\kappa}$ denote an estimate of κ , as in Proposition 3.1. We define our rejection region of $H_{0,s}$ as $\mathcal{R}_n = (-\infty, \delta_n]$ and our test as

$$(10) \quad \phi_n(\hat{W}_{\hat{\kappa}}) = \begin{cases} 0, & \lambda_1(\hat{W}_{\hat{\kappa}}) \in \mathcal{R}_n, \\ 1, & \lambda_1(\hat{W}_{\hat{\kappa}}) \notin \mathcal{R}_n, \end{cases}$$

Finally, when testing if D is hyperbolic, let $\hat{\kappa}$ denote an estimate of $\kappa < 0$, as in Proposition 3.1. Our rejection region of $H_{0,h}$ is $\mathcal{R}_n = [\delta_n, \infty)$ and our test is

$$(11) \quad \phi_n(W_{\hat{\kappa}}) = \begin{cases} 0, & \lambda_1(\hat{W}_{\hat{\kappa}}) \in \mathcal{R}_n, \\ 1, & \lambda_1(\hat{W}_{\hat{\kappa}}) \notin \mathcal{R}_n, \end{cases}$$

We now study what conditions must hold on this sequence δ_n 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 4.1. *Let $\delta_n = o_P(1)$ be a random or deterministic sequence.*

- (1) *If $\delta_n \in (-\infty, 0]$, $\delta_n = o_P(1)$ and $\mathbb{P}(\lambda_1(\hat{W}_0) \leq \delta_n) = 1 - o(1)$, then the test for $H_{0,e}$ in (6) with rejection region $R_n := (-\infty, \delta_n]$ is consistent.*
- (2) *If $\delta_n \in (-\infty, 0]$, $\delta_n = o_P(1)$ and $\mathbb{P}(\lambda_1(\hat{W}_{\hat{\kappa}}) \leq \delta_n) = 1 - o(1)$ with $\hat{\kappa} \in (0, \infty)$, then the test for $H_{0,s}$ in (7) with rejection region $R_n := (-\infty, \delta_n]$ is consistent.*
- (3) *If $\delta_n \in [0, \infty)$, $\delta_n = o_P(1)$ and $\mathbb{P}(\lambda_1(\hat{W}_{\hat{\kappa}}) \geq \delta_n) = 1 - o(1)$, then the test for $H_{0,h}$ in (8) with rejection region $R_n := [\delta_n, \infty)$ is consistent.*

We prove Proposition 4.1 in Appendix A. Intuitively, this Proposition shows that we can use the observed distance matrix \hat{D} to test the hypotheses that the latent space is Euclidean, spherical, or hyperbolic. From these tests we define $\widehat{\mathcal{M}}^{\hat{p}}(\hat{\kappa})$ as the intersection of the three tests. That is, the estimated latent geometry based on \hat{D} is defined by the result of three hypothesis tests in Equations (9), (10), and (11). More specifically, we can select any estimator that preserves the consistency. For example, we can use an ordered test to estimate the geometry type.

Thanks to Proposition 4.1, with sufficiently large n the probability that more than one of these tests will fail to reject the null goes to zero. As the sample size becomes infinite, therefore, we can define our estimated geometry as an indicator for whether each of the tests fails to reject the null. Using this intuition, we can now show consistency for estimating the latent geometry type in two settings. The first, Theorem 4.1, is a general setting which does not rely on our context of graphs, but instead applies in any setting where there is a noisy distance matrix from a manifold of unknown type and dimension. We leverage this more general result to prove consistency for the network case in Theorem 4.2. Before stating this result, we first state some technical conditions needed to verify our results.

ASSUMPTION 4.1. *Suppose that D is a $K \times K$ distance matrix from K points on $\mathcal{M}^p(\kappa)$ satisfying Assumption 1.1 such that $\mathcal{M}^p(\kappa)$ is uniquely identified. Assume that for any κ , there exists \hat{D} and Ω_κ such that $\hat{W}_{\hat{\kappa}} = \hat{W}_{\hat{\kappa}}(\hat{D})$ satisfies $\sqrt{n}(\hat{W}_{\hat{\kappa}} - W_\kappa) \xrightarrow{D} N(0, \Omega_\kappa)$, for some covariance matrix Ω . Suppose there exists some estimator $\hat{\Omega}$ such that $\hat{\Omega} \xrightarrow{P} \Omega$. Finally, suppose that $0 \leq p < K \cdot (1 - \sqrt{\frac{1+K}{2K}})$.*

THEOREM 4.1. *Let Assumption 4.1 hold. Define the tests of manifold type as in (9), (10), and (11). Further, let the real-valued sequence $\delta_n \in (-\infty, 0]$ be consistent with Proposition 4.1. Define \hat{p} as the dimension estimate from Robin and Smith (2000) and let $\hat{\kappa}$ denote the curvature estimate from Proposition 3.1. Then, $\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}} \neq \mathcal{M}^p) = o(1)$ and $\hat{\kappa} - \kappa = o_P(1)$.*

We provide a sketch of the proof here. The full proof is in Appendix A. The key ideas come from Schoenberg (1935) and Begelfor and Werman (2005), which provide necessary and sufficient conditions for a set of points to be embedded in Euclidean, spherical, and hyperbolic spaces. With a sequence δ_n for each of the three spaces, we have consistent tests for each geometry. Taking the intersection, we have the first result. In addition, Proposition 3.1 shows that $\hat{\kappa} - \kappa = o_P(1)$ under the stated conditions and the dimension estimator \hat{p} from Robin and Smith (2000) is consistent for the true dimension.

The following proposition provides upper bounds on the Type 1 error α of a hypothesis test for the geometry. The main idea is that we can control the deviation of the estimated eigenvalue around the population eigenvalue by Weyl's inequality.

PROPOSITION 4.2. *Suppose that D is a $K \times K$ distance matrix from K points on $\mathcal{M}^p(\kappa)$ satisfying Assumption 1.1, with K chosen such that the $\mathcal{M}^p(\kappa)$ is uniquely identified. Assume further that $\hat{D} \xrightarrow{P} D$. Let θ_α be defined as the α th quantile of the distribution of $\|\hat{W}_{\hat{\kappa}} - W_\kappa\|_F$. Then,*

$$(12) \quad \mathbb{P}\left(\lambda_{k^*}\left(W_{\hat{\kappa}}(\hat{D})\right) < \theta_\alpha\right) \leq \alpha.$$

We now move to our second result. Once we have a consistent estimate for the distance matrix, D , we can use the same techniques used in the proof of Theorem 4.1 to prove the following result. To introduce this result, recalling the graph model in Equation (1), the likelihood \mathcal{L} for observing the graph G is given by

$$\begin{aligned} \mathcal{L}(G|z, \nu, \mathcal{M}^p(\kappa)) &= \prod_{i < j}^n \mathbb{P}(G_{ij} = 1 | \nu, z, \mathcal{M}^p(\kappa)) \\ &= \prod_{i < j}^n (\exp(\nu_i + \nu_j - d(z_i, z_j)))^{G_{ij}} \prod_{i < j}^n (1 - \exp(\nu_i + \nu_j - d(z_i, z_j)))^{1-G_{ij}} \end{aligned}$$

and by taking the log we see that the log-likelihood $\ell(G|z, \nu, \mathcal{M}^p(\kappa))$ is

$$\ell(G|z, \nu, \mathcal{M}^p(\kappa)) = \sum_{i < j}^n G_{ij}(\nu_i + \nu_j - d(z_i, z_j)) + \sum_{i < j}^n (1 - G_{ij}) \log(1 - \exp(\nu_i + \nu_j - d(z_i, z_j))).$$

Our goal is to solve for the maximum likelihood estimators $\hat{\nu}_i$ and \hat{z}_i , which satisfy

$$(13) \quad (\hat{\nu}, \hat{z}) \in \arg \max_{(\nu, z) \in (-\infty, 0]^n \times \prod_{i=1}^n \mathcal{M}^p(\kappa)} \ell(G|z, \nu, \mathcal{M}^p(\kappa)).$$

Recall that once we have assumed a geometry $\mathcal{M}^p(\kappa)$, we know the expression for $d(z_i, z_j)$, which allows us to evaluate the log-likelihood and solve for the maximum likelihood estimators. It is challenging, however, to study the accuracy of these maximum likelihood estimators since the number of parameters, which is $2n$ in our case, grows as the graph size grows. Lemma A.1 of Breza et al. (2019) shows that, once we assume a latent space geometry, the maximum likelihood estimators are consistent in the sense given in the following result.

THEOREM 4.2. *Assume that Assumptions 1.1-1.3 hold, and that G is distributed as in (1). Let $\widehat{\mathcal{M}}^{\hat{p}}(\hat{\kappa})$ denote the estimate of the geometry based on the tests from (9), (10), and (11) with $\hat{\kappa}$ and \hat{p} denoted the corresponding estimates of the curvature (from Proposition 3.1) and dimension (from Robin and Smith (2000)). Let \hat{z}_i and $\hat{\nu}_i$ denote the manifold-specific maximum likelihood estimator for z_i and ν_i defined in (13), using the estimated latent space type, dimension, and curvature. Finally, suppose that there exists a consistent estimator of the mean of the individual effects distribution, $E(\exp(\nu_i))$. Then, the following three results hold.*

- (1) *We can consistently estimate the type, dimension and curvature of $\mathcal{M}^p(\kappa)$:*

$$\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}} \neq \mathcal{M}^p) = o(1) \text{ and } \hat{\kappa} - k = o_P(1).$$

- (2) *We can consistently estimate the model parameters (ν, z) in the following sense:*

$$\max_{1 \leq i \leq n} |\nu_i - \hat{\nu}_i| = o_P(1) \text{ and } \inf_{\phi \in \text{isom}(\mathcal{M})} \sum_{i=1}^n d_{\mathcal{M}^p(\kappa)}(z_i, \phi(\hat{z}_i)) = o_P(1)$$

where $\text{isom}(\mathcal{M})$ denotes the set of isometries on $\mathcal{M}^p(\kappa)$.

Recall that an isometry defined on $\mathcal{M}^p(\kappa)$ is a function $\phi : \mathcal{M}^p(\kappa) \rightarrow \mathcal{M}^p(\kappa)$ such that $d_{\mathcal{M}}(\phi(x), \phi(y)) = d_{\mathcal{M}}(x, y)$. In words, an isometry is a function that preserves the distance between its arguments. We denote the set of isometries on $\mathcal{M}^p(\kappa)$ by $\text{isom}(\mathcal{M})$. For example, isometries on \mathbb{R}^p include rotations and translations. The result in Theorem 4.2 says that we can estimate the locations z_1, \dots, z_n in the sense that there exists some isometry ϕ such that $\sum_{i=1}^n d_{\mathcal{M}^p(\kappa)}(z_i, \phi(\hat{z}_i)) = o_P(1)$.

We now outline the proof, with full details given in Appendix A. We first showed how to use the geometry embedding theorems from Schoenberg (1935) and Begelfor and Werman (2005) to conduct hypothesis tests about the geometry of the latent space model. We proposed three tests, one for each geometry type (Euclidean, spherical, or hyperbolic). We showed that these tests are consistent in the sense that, given an appropriate sequence of threshold values and sufficiently large graph, we show that they will always fail to reject in cases where the null is true and reject when the null is false. Finally, we combine our proposal with previous results from Breza et al. (2019) to show that under the model in (1), we can consistently estimate all of the parameters of interest: the type, dimension, and curvature of the latent space, as well as the individual fixed effects and the latent space locations.

While the results from this section describe the behavior of our proposed methods as the graph grows, we are also interested in how to implement these procedures for finite samples. Classical bootstrap approaches fail in this context as we have a parameter-on-the-boundary problem and also it is possible for there to be repeated eigenvalues. We rely instead on a bootstrap that uses sub-sampling based on work by Politis and Romano (1994). We provide full details, including specific algorithms and a discussion of theoretical properties, in Appendix D.

5. SIMULATION EVALUATION

We now examine the performance of our 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 proposed hypothesis tests and then show the performance of our algorithm

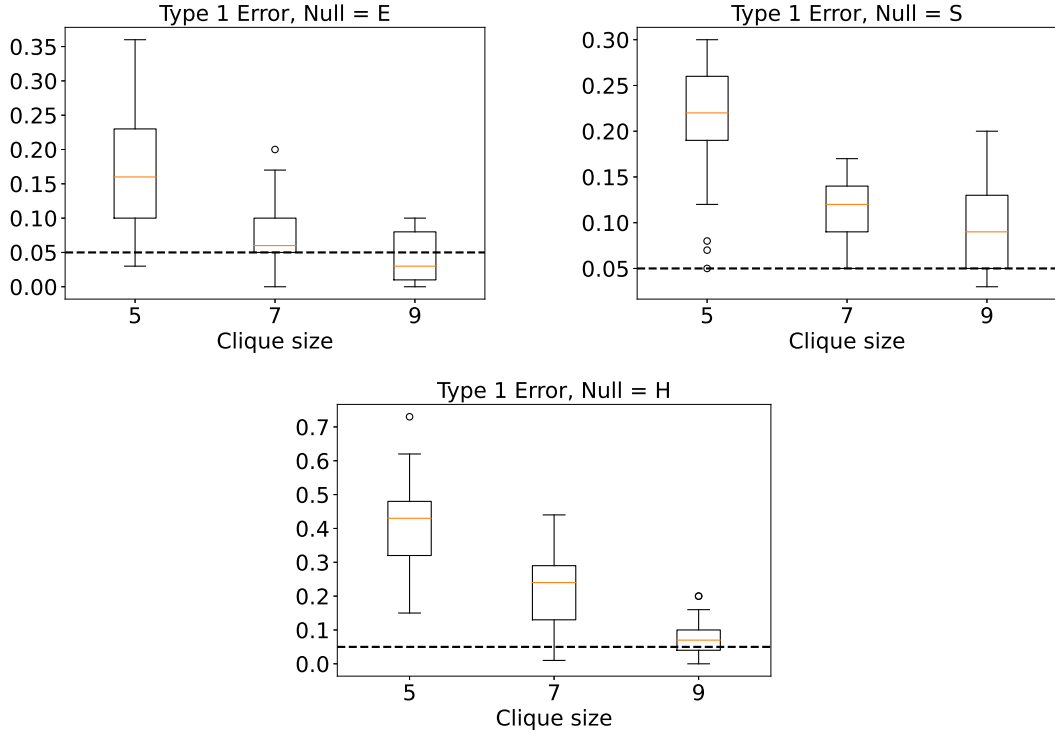


FIGURE 2. Estimated type 1 error using 25 simulated latent space positions. For each set of positions, we perform the test 100 times. Each box contains the 25 points, each representing the fraction of rejections for a set of latent positions.

for estimating the latent dimension. Additional simulation results showing estimates of curvature are in Appendix E.

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 centers randomly in the latent geometry and dimension to be tested. We generated 15 centers and a graph of size $n = 1200$. We then spread points around these latent centers, with an equal number of points at each center. We use this approach rather than (for example) assigning points uniformly across the surface to mimic community structure in empirical networks (Newman et al., 2000; Girvan and Newman, 2002; Jackson, 2008; Leskovec et al., 2008; Newman, 2010). We can, of course, interpolate between a setting with extreme within-group homophily and uniform latent positions by adjusting the spread of points around these centers. We generated 25 sets of latent positions and then, for each set of latent positions, construct 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 F.

Figures 2 and 3 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

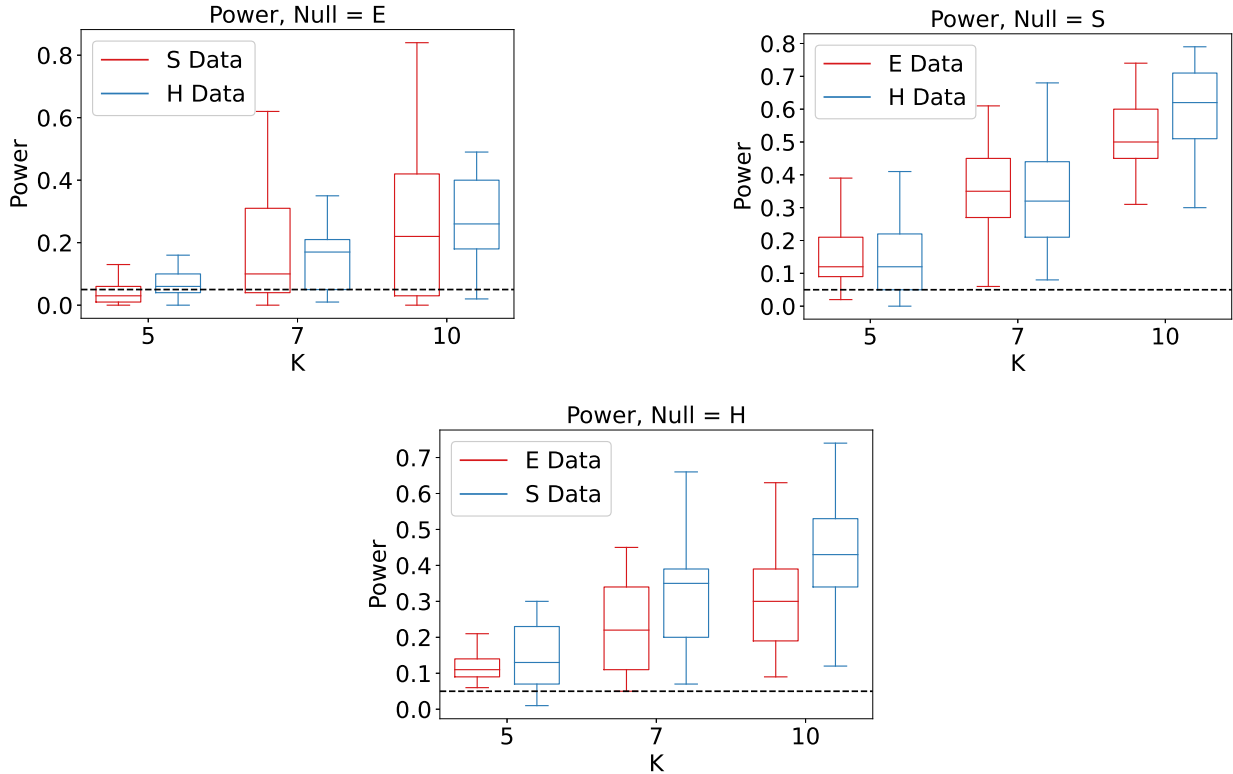


FIGURE 3. Estimated power using 25 simulated latent space positions. For each set of positions, we perform the test 100 times. Each box contains the 25 points, each representing the fraction of rejections for a set of latent positions.

true underlying geometry and the simulation procedure we use. Figure 2 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 demonstrate the impact of increasing the clique size (and, in doing so, providing a more precise estimate of a distance matrix with a fixed dimension). 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 3 we see that the power increases as we increase K for all three geometries. 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.

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 $\mathcal{M}^p(\kappa)$. In Table 1 we give our estimates of the dimension for the three geometries. Table 1 shows that our method was always within one dimension in the simulation experiments we conducted. For the curved space, the simulated graph size was sufficient to achieve an accurate estimate of the minimal dimension in every simulated graph.

TABLE 1. Rank estimator performance with $K = 5$ and $\ell = 9$. using simulated latent spaces.

	\mathbb{R}^3	$\mathbf{S}^2(1)$	$\mathbf{H}^2(-1)$
$P(\hat{p} = 2)$	0.52	1.0	1.0
$P(\hat{p} = 3)$	0.48	0	0
$P(\hat{p} > 3)$	0	0	0

6. EXAMPLES FROM ECONOMICS AND BIOLOGY

We now demonstrate the performance of our method in a setting with the complexity of observed data. We show that in two 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 two contexts: (a) 75 village social networks with fully observed graph data from [Banerjee et al. \(2019\)](#) and (b) a neural network of a single *Caenorhadbitis elegans* worm [Kaiser and Hilgetag \(2006\)](#).

6.1. Choices for Implementation. Since our method relies on distances between cliques, a key decision is how to ascertain cliques for a given graph. Overall, there are two key considerations. First, we would like to take the number (K) and size (ℓ) of cliques to be as large as possible. As ℓ increases, the variance of estimates of \hat{D} decreases. The power of the test increases as K increases, since we have more distances between points on the manifold. Figure 3 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 not overlap (have nodes in common), otherwise there is not a well-defined distance to measure between them. If two cliques are fully disconnected, then the distance between them is not defined.

In practice, our approach is to first select K and ℓ , given the considerations described above. Then, we want to find K cliques, each of size ℓ , such that these cliques have a small overlap. Written formally, our goal is to solve

$$(14) \quad \hat{C}_1, \dots, \hat{C}_K \in \operatorname{argmin}_{C_1, \dots, C_K} \sum_{i,j}^K |C_i \cap C_j|$$

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

As discussed previously, we generally want K and ℓ to be as big as possible. In practice, we set K and ℓ by first looking at the number of cliques of various sizes in the graph and choosing 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 (14). Searching over the set of possible cliques is a well-studied (NP-hard) problem, 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 (14).

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 G. For the *C. Elegans* data, we select $K = 12$ and set $\ell = 5$, which is the size of the largest clique 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) which is the sequel to the Wave I data in Banerjee et al. (2013), both 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 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. The links across these dimensions mostly line up, consistent with a theory of multiplexed incentives to form links, so we study the undirected, unweighted graph following the prior literature using this data (Jackson and Lopez-Pintado, 2013; Banerjee et al., 2013; Breza and Chandrasekhar, 2019; Banerjee et al., 2019).

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. A literature focusing on equilibrium informal financial networks, which facilitates 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 basic 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 good behavior in equilibrium.

From the perspective of a latent space model, 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, as discussed 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

⁹Common modeling assumptions in the socio-economic literature imply constant curvature from the perspective of our model (1), though certainly there are perspectives that would violate this and 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, imagine a model in which agents can take an action to influence the extent to which their neighbors 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 and Seinfeld, 1995). 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

TABLE 2. Descriptive classification (using maximum p -value) for the 75 villages. Villages classified as N/A had p -values less than .05 for all three tests.

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

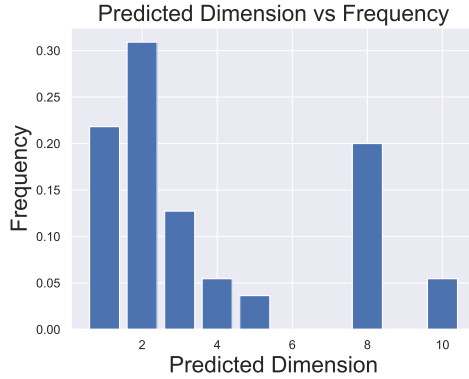


FIGURE 4. Predicted dimensions for the predicted geometries of the Indian village networks. Most predictions are low-dimensional, usually between 1 and 3. Some of the predictions are higher. In these cases, our method is predicting that W_κ is full rank.

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). Table 2 presents the results. We see that we are able to classify 75% of the villages. It is important to note that the vast majority of empirical villages were able to be classified 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. Figure 4 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 the latent geometry and socio-economic phenomena. Note that these are observational, not causal, analyses. First, we look at how the volume of informal financial transactions vary with network geometry. Specifically, we are interested in how the volume of informal loans that a household has with network neighbors (e.g., friends or members of their rotating, savings, and credit associations) varies with geometry. Both the theoretical and empirical economic literatures 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

complex models with heterogeneity would require modeling manifolds with non-constant curvature, which we leave to future work.

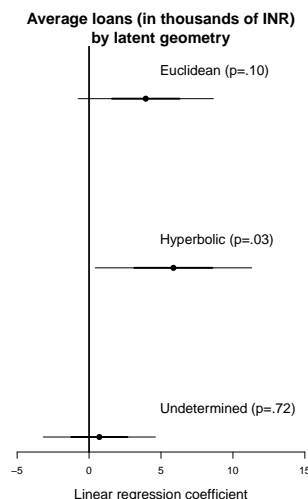


FIGURE 5. Regression coefficients showing the relationship between loans and geometry. We plot the coefficient in a multivariate linear regression where the outcome is the average amount of loans (in thousands of INR) and the predictors are geometry types (with spherical as the reference). Wide bars correspond to one standard error and narrow bars represent two. The reference value for spherical is 16.71 (in thousands of INR).

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, 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:

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

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 \approx INR 73.5). Here, the omitted category (α) corresponds to the loan amount for a sphere.

Figure 5 presents the results. We find that a Euclidean village relative to a spherical one has INR 3940 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 5865 or 35% ($p = 0.034$) more in 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. So, we see greater financial flows precisely in geometries that permit more expansive networks.

Second, we turn to determinants of geometry. Our primary interest is in whether the introduction of a formal credit market (microfinance) to a setting otherwise dominated by informal financial transactions changes the network structure by changing the latent geometry. Microcredit was introduced to only some of the villages, allowing us to compare the impact of access to microfinance on network structure.

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 so 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 endeavors 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., 2013, 2020). 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 ν . 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 and push towards negative curvature. In either case, the answer as to how a large credit intervention affects geometry is of empirical interest.

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

$$\frac{\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}}_i = E)}{\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}}_i = \mathbf{S})} = \exp(\delta_1 + \beta_{\text{MFI}}^E \text{MFI}_i + \beta_{\text{W}}^E \text{Wealth}_i + \beta_{\text{I}}^E \text{Inequality}_i + \beta_{\text{F}}^E \text{Frac}_i)$$

$$\frac{\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}}_i = \mathbf{H})}{\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}}_i = \mathbf{S})} = \exp(\delta_2 + \beta_{\text{MFI}}^H \text{MFI}_i + \beta_{\text{W}}^H \cdot \text{Wealth}_i + \beta_{\text{I}}^H \text{Inequality}_i + \beta_{\text{F}}^H \text{Frac}_i)$$

$$\frac{\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}}_i = \text{N/A})}{\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}}_i = \mathbf{S})} = \exp(\delta_3 + \beta_{\text{MFI}}^N \text{MFI}_i + \beta_{\text{W}}^N \text{Wealth}_i + \beta_{\text{I}}^N \text{Inequality}_i + \beta_{\text{F}}^N \text{Frac}_i) .$$

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 wealth. Finally, $\text{Frac}_i = \alpha_U(1 - \alpha_U)$ where α_U is the share of households that are of upper caste. This is zero if society is homogenous and 1/4 for an even split.

¹⁰The Banerjee et al. (2019) dataset does not have consumption nor expenditure measures. We utilize a score constructed from the first principle component of a number of household features that correlate with wealth: access to private electricity, home ownership, quality of roofing material, and number of rooms in the house. For inequality, we take the score from the first principle component of the within-village standard deviation of each of the constituent wealth measures.

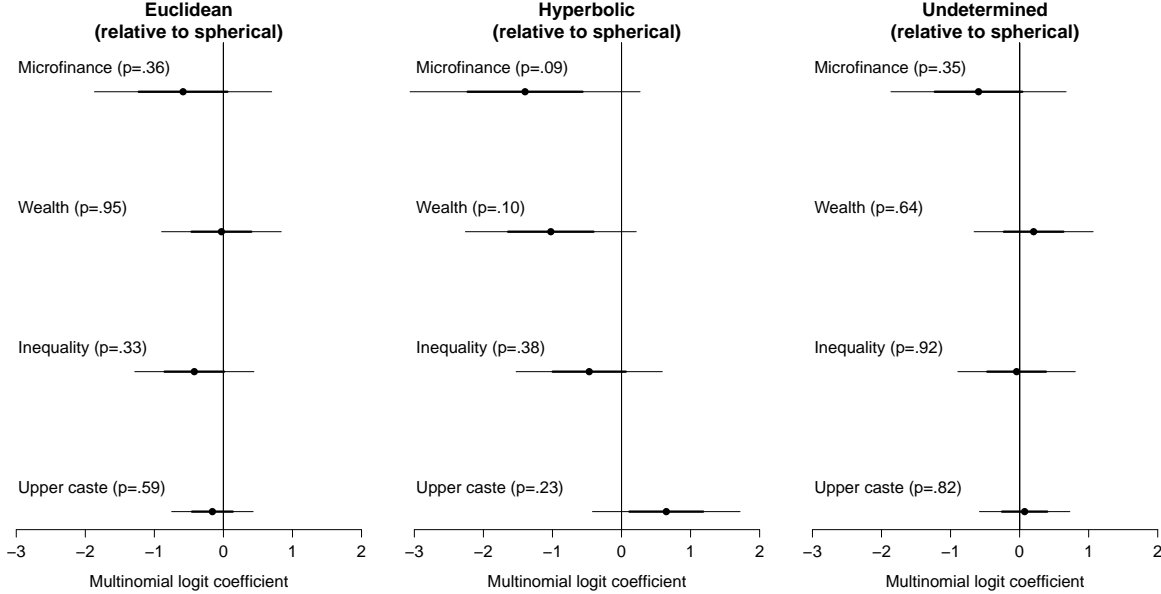


FIGURE 6. Regression coefficients showing the determinants of geometry. Plots show the coefficients from a multinomial logistic regression where the outcome is the predicted geometry type for each village. Each panel shows all coefficients for a particular geometry (with spherical as the reference). Each line in the plot corresponds to an estimated coefficient. Wide bars correspond to one standard error and narrow bars represent two.

Figure 6 presents the results. We begin by looking at microfinance. We estimate $\hat{\beta}_{\text{MFI}}^H = -1.40$ ($p = 0.093$). This means that when a village receives microcredit, there is a 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 (ν_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: $\hat{\beta}_W^H = -1.02$ ($p = 0.098$). 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 inequality nor fractionalization and geometry.

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. Informal financial loans are higher in villages with negative curvature. Finally, and importantly, introducing microcredit nudges villages to have 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 a neuroscience literature that is interested in documenting regularities in network structure as well as modeling 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 has thick tails: certain nodes have numerous connections. Moreover, the network is dynamic: early in age the it exhibits high amounts of homophily whereas as the individual ages this declines.

The second strand attempts to develop a low dimensional statistical representation 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 simply correlated 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 (VanRullen and Reddy, 2019; Karwowski et al., 2019).

Third, and particularly relevant for latent space models, is the concept of the functional graph of neurons rather than the structural graph (Petersen and Sporns, 2015; Abdelnour et al., 2018). 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 *Caenorhabditis elegans*, which are soil-dwelling roundworms. There is a long history of using *C. elegans* as a model organism for studying nervous systems of animals. In fact neurons of *C. elegans* are extremely similar to that of humans Leung, Williams, Benedetto, Au, Helmcke, Aschner, and Meyer (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 modeling neural network structure. For instance, the relative location distribution, how distance affects linking rates, and the geometry all inform how signals could 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 modeled 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 (14) and then take a maximally disjoint clique set which is sufficient for our test. We compute the p values for the geometries and find $p_E = .378$, $p_S = 0.05$, and $p_H = 0.267$, so we reject the spherical hypothesis (noting we can do so despite there being a high level of clustering). However, in this data, we can only say that there is weakly negative curvature, but are not powered to distinguish between hyperbolic nor Euclidean hypotheses.

Thus, we can strongly conclude that the *C. elegans* network of neurons 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 does not 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 space, though there is current discussion about adopting a hyperbolic space in certain contexts. Nonetheless, current methods employed do not provide a way to estimate the geometry. Unfortunately, incorrect embedding spaces can deliver misleading results. While there may be convergence to pseudo-true values, counterfactual analysis will be affected.

The observed network provides information about latent space distances between nodes. A matrix of distances must be consistent with the geometry in which it is embedded; given a collection of distances, this can be checked. A finite sample network corresponds to a noisy set of distances; we develop a procedure to classify and estimate the geometry—meaning manifold type, dimension, and curvature. We show this procedure is consistent as the number of nodes tends to infinity.

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

We also demonstrate the 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 75 villages. 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. These are interpretable economic findings, consistent with theory, borne out of a geometric exercise. When we turn to the structure of a network of neurons, despite the neural network having a high degree of clustering, we are able to strongly reject positive curvature. We do are unable to distinguish between hyperbolic and Euclidean geometries, and can only conclude that in this setting curvature is weakly negative. This is consistent with a setting where there is at least limited value for branching and expansion for rapid signal transmission.

Several next steps come to mind. First, in many contexts aggregate data, such as aggregated relational data (ARD), rather than network data at the edge level is what the researchers have at their disposal (McCormick and Zheng, 2015; Breza et al., Forthcoming). An important avenue for future work is to extend methods to cases where different samplings are involved rather than observing the network. Second, while our assumptions on geometry—that it is a simply connected, complete Riemannian manifold of constant curvature—are natural, they are also limited. Though it nests the current literature—we know of no empirical research that assumes a torus of genus two for instance—it is still lacking. We speculate that there may be strategies to use local structures in the network to patch together 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 these can be stitched together. This is a loose and speculative discussion of course.

ACKNOWLEDGEMENTS

We thank Eric Auerbach, Abhijit Banerjee, Emily Breza, Jacob Burchard, Gabriel Carroll, James Evans, Bailey Fosdick, Jeremy Fox, Paul Goldsmith-Pinkham, Ben Golub, Matthew Grant, Fang Han, Rachel Heath, Yunmi Kong, Mengjie Pan, Mallesh Pai, Abel Rodriguez, Anna Smith, Xun Tang, Matt Thirkettle, Aravindan Vijayaraghavan, and Jon Wellner. We thank participants at the 2020 Joint Statistical Meeting, CANSSI-Ontario Data Science Applied Research and Education Seminar, IDEAL (Institute for Data, Econometrics, Algorithms, and Learning), the Joint Econometrics and Statistics Seminar Series at Cornell University, UC-Davis (Statistics), Bocconi University (Statistics), and Rice (Applied Micro and Econometrics). Chandrasekhar is grateful for support from the Alfred P. Sloan foundation.

REFERENCES

- ABDELNOUR, F., M. DAYAN, O. DEVINSKY, T. THESEN, AND A. RAJ (2018): “Functional brain connectivity is predictable from anatomic network’s Laplacian eigenstructure,” *NeuroImage*, 172, 728–739.
- ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015): “Systemic risk and stability in financial networks,” *American Economic Review*, 105, 564–608.
- ALDOUS, D. J. (1981): “Representations for partially exchangeable arrays of random variables,” *Journal of Multivariate Analysis*, 11, 581–598.
- AMBRUS, A., M. MOBIUS, AND A. SZEIDL (2014): “Consumption Risk-Sharing in Social Networks,” *American Economic Review*, 104, 149–82.
- ANDREWS, D. (2000): “Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space,” *Econometrica*, 68, pp. 399–405.
- ASTA, D. M. AND C. R. SHALIZI (2015): “Geometric network comparisons,” in *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, AUAI Press, 102–110.
- BANERJEE, A., E. BREZA, A. CHANDRASEKHAR, E. DUFLO, C. KINNAN, AND M. JACKSON (2020): “Changes in social network structure in response to exposure to formal credit markets,” *Working Paper*.
- BANERJEE, A., A. CHANDRASEKHAR, E. DUFLO, AND M. JACKSON (2013): “Diffusion of Microfinance,” *Science*, 341, 1–7.
- BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2019): “Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials,” *The Review of Economic Studies*.
- BANSAL, S., J. READ, B. POURBOHLOUL, AND L. A. MEYERS (2010): “The dynamic nature of contact networks in infectious disease epidemiology,” *Journal of biological dynamics*, 4, 478–489.
- 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.
- BEGELFOR, E. AND M. WERMAN (2005): “The world is not always flat or learning curved manifolds,” *School of Engineering and Computer Science, Hebrew University of Jerusalem., Tech. Rep*, 3, 8.
- BREZA, E. AND A. G. CHANDRASEKHAR (2019): “Social networks, reputation, and commitment: evidence from a savings monitors experiment,” *Econometrica*, 87, 175–216.

- BREZA, E., A. G. CHANDRASEKHAR, T. MCCORMICK, AND M. PAN (2019): “Consistently estimating graph statistics using Aggregated Relational Data,” *arXiv preprint arXiv:1908.09881*.
- BREZA, E., A. G. CHANDRASEKHAR, T. H. MCCORMICK, AND M. PAN (Forthcoming): “Using aggregated relational data to feasibly identify network structure without network data,” *American Economic Review*.
- CAI, J. AND A. SZEIDL (2017): “Interfirm relationships and business performance,” *Quarterly Journal of Economics*, 133, 1229–1282.
- CALVO-ARMENGOL, A. (2004): “Job contact networks,” *Journal of Economic Theory*, 115, 191–206.
- CALVÓ-ARMENGOL, A., E. PATACCINI, AND Y. ZENOU (2009): “Peer effects and social networks in education,” *The Review of Economic Studies*, 76, 1239–1267.
- CHANDRASEKHAR, A. AND R. LEWIS (2016): “Econometrics of sampled networks,” Stanford Working Paper.
- CHANEY, T. (2014): “The network structure of international trade,” *American Economic Review*, 104, 3600–3634.
- CHATTERJEE, S., P. DIACONIS, A. SLY, ET AL. (2011): “Random graphs with a given degree sequence,” *The Annals of Applied Probability*, 21, 1400–1435.
- CHO, Y.-S., G. VER STEEG, E. FERRARA, AND A. GALSTYAN (2016): “Latent space model for multi-modal social data,” in *Proceedings of the 25th International Conference on World Wide Web*, 447–458.
- COLEMAN, J. (1988): “Social Capital in the Creation of Human Capital,” *American Journal of Sociology*, 94, S95–S120.
- CURRARINI, S., M. JACKSON, AND P. PIN (2009): “An economic model of friendship: Homophily, minorities, and segregation,” *Econometrica*, 77, 1003–1045.
- DAVID, L. AND J. SEINFELD (1995): “The Pool Guy,” *Seinfeld*.
- DE LANGE, S., M. DE REUS, AND M. VAN DEN HEUVEL (2014): “The Laplacian spectrum of neural networks,” *Frontiers in computational neuroscience*, 7, 189.
- DIPRETE, T. A., A. GELMAN, T. MCCORMICK, J. TEITLER, AND T. ZHENG (2011): “Segregation in social networks based on acquaintanceship and trust,” *American Journal of Sociology*, 116, 1234–83.
- EATON, M. L. AND D. E. TAYLOR (1991): “On Weilandt’s Inequality and Its Application to the Asymptotic Distribution of Eigenvalues of a Random Symmetric Matrix,” *Annals of Statistics*, 19, 260–271.
- ELLIOTT, M., B. GOLUB, AND M. O. JACKSON (2014): “Financial networks and contagion,” *American Economic Review*, 104, 3115–53.
- GAI, P. AND S. KAPADIA (2010): “Contagion in financial networks,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466, 2401–2423.
- GIRVAN, M. AND M. E. NEWMAN (2002): “Community structure in social and biological networks,” *Proceedings of the national academy of sciences*, 99, 7821–7826.
- GRAHAM, B. S. (2017): “An econometric model of network formation with degree heterogeneity,” *Econometrica*, 85, 1033–1063.
- GRANOVETTER, M. S. (1973): “The Strength of Weak Ties,” *The American Journal of Sociology*, 78, 1360–1380.
- HANDCOCK, M. S. AND J. H. JONES (2004): “Likelihood-based inference for stochastic models of sexual network formation,” *Theoretical population biology*, 65, 413–422.

- HEATH, R. (2018): “Why do firms hire using referrals? evidence from bangladeshi garment factories,” *Journal of Political Economy*, 126, 1691–1746.
- HOFF, P., A. RAFTERY, AND M. HANDCOCK (2002): “Latent Space Approaches to Social Network Analysis,” *Journal of the American Statistical Association*, 97:460, 1090–1098.
- JACKSON, M. (2008): *Social and Economic Networks*, Princeton: Princeton University Press.
- (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, DOI: 10.1080/07350015.2013.794095.
- JACKSON, M. AND D. LOPEZ-PINTADO (2013): “Diffusion and Contagion in Networks with Heterogeneous Agents and Homophily,” *Network Science*, 1:1, 49–67.
- KAISER, M. AND C. C. HILGETAG (2006): “Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems,” *PLoS computational biology*, 2.
- KARWOWSKI, W., F. VASHEGHANI FARAHANI, AND N. LIGHTHALL (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, 257–278.
- KINNAN, C. AND R. TOWNSEND (2012): “Kinship and financial networks, formal financial access, and risk reduction,” *The American Economic Review*, 102, 289–293.
- KRIOUKOV, D., F. PAPADOPOULOS, M. KITSAK, A. VAHDAT, AND M. BOGUNÁ (2010): “Hyperbolic geometry of complex networks,” *Physical Review E*, 82, 036106.
- LESKOVEC, J., K. J. LANG, A. DASGUPTA, AND M. W. MAHONEY (2008): “Statistical properties of community structure in large social and information networks,” in *Proceedings of the 17th international conference on World Wide Web*, 695–704.
- LEUNG, M. C., P. L. WILLIAMS, A. BENEDETTO, C. AU, K. J. HELMCKE, M. ASCHNER, AND J. N. MEYER (2008): “Caenorhabditis elegans: an emerging model in biomedical and environmental toxicology,” *Toxicological sciences*, 106, 5–28.
- LUO, W. AND B. LI (2016): “Combining eigenvalues and variation of eigenvectors for order determination,” *Biometrika*, 103, 875–887.
- MCCORMICK, T. H. AND T. ZHENG (2015): “Latent surface models for networks using Aggregated Relational Data,” *Journal of the American Statistical Association*, 110, 1684–1695.
- MYERS, S. A. AND J. LESKOVEC (2014): “The bursty dynamics of the twitter information network,” in *Proceedings of the 23rd international conference on World wide web*, 913–924.
- NEWBY, W. K. AND D. MCFADDEN (1994): “Large sample estimation and hypothesis testing,” *Handbook of Econometrics*, 4, 2111–2245.
- NEWMAN, M. (2010): *Networks: An Introduction*, Oxford University Press.
- NEWMAN, M., C. MOORE, AND D. WATTS (2000): “Mean-field solution of the small-world network model,” *Physical Review Letters*, 84, 3201–3204.
- O’NEILL, B. (1983): *Semi-Riemannian Geometry with Applications to Relativity*, vol. 103, Academic press.
- ORBANZ, P. AND D. M. ROY (2015): “Bayesian models of graphs, arrays and other exchangeable random structures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37, 437–461.

- PETERSEN, S. E. AND O. SPORNS (2015): “Brain networks and cognitive architectures,” *Neuron*, 88, 207–219.
- POLITIS, D. N. AND J. P. ROMANO (1994): “Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions,” *Annals of Statistics*, 22, 2031 – 2050.
- RECANATESI, S., M. FARRELL, G. LAJOIE, S. DENEVE, M. RIGOTTI, AND E. SHEA-BROWN (2019): “Predictive learning extracts latent space representations from sensory observations,” *bioRxiv*, 471987.
- ROBIN, J.-M. AND R. J. SMITH (2000): “Tests of Rank,” *Econometric Theory*, 16, 151–175.
- ROMERO, D. M., B. MEEDER, AND J. KLEINBERG (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*, 695–704.
- SALTER-TOWNSHEND, M. AND T. H. MCCORMICK (2017): “Latent space models for multiview network data,” *The Annals of Applied Statistics*, 11, 1217.
- 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, 724–732.
- SEWELL, D. K. AND Y. CHEN (2015): “Latent space models for dynamic networks,” *Journal of the American Statistical Association*, 110, 1646–1657.
- SHALIZI, C. R. AND D. ASTA (2017): “Consistency of maximum likelihood for continuous-space network models,” *arXiv preprint arXiv:1711.02123*.
- SMITH, A. L., D. M. ASTA, C. A. CALDER, ET AL. (2019): “The geometry of continuous latent space models for network data,” *Statistical Science*, 34, 428–453.
- VANRULLEN, R. AND L. REDDY (2019): “Reconstructing faces from fMRI patterns using deep generative neural networks,” *Communications biology*, 2, 1–10.
- WILSON, R. C., E. R. HANDCOCK, E. PEKALSKA, AND R. P. DUIN (2014): “Spherical and Hyperbolic Embeddings of Data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 2255–2269.

APPENDIX A. PROOFS

A.1. Proof of Proposition 3.1.

Proof of Proposition 3.1. 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 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 $\hat{D} \rightarrow_p D$ as $n \rightarrow \infty$. Since eigenvalues are continuous functions of their matrix arguments, we have by the continuous mapping theorem that $\lambda_1(\kappa W_\kappa(\hat{D})) \xrightarrow{p} \lambda_1(\kappa W_\kappa(D))$ 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 and McFadden \(1994\)](#)): that for any κ_1, κ_2 , $|\lambda_1(\kappa_1 W_{\kappa_1}(\hat{D})) - \lambda_1(\kappa_2 W_{\kappa_2}(\hat{D}))| \leq B_n |\kappa_1 - \kappa_2|$ for some random variable $B_n = O_p(1)$. To do this, fix any $\kappa_1, \kappa_2 \in [a, b]$. By Weyl's inequality,

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

where $\|A\|_F$ is the Frobenius norm of A , that is $\|A\|_F^2 = \sum_{l,l'} a_{ll'}^2$. 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' ,

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

For $\kappa \in [a, b]$,

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

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

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

Putting this all together, we see that

$$\begin{aligned} \left| \lambda_1(\kappa_1 W_{\kappa_1}(\hat{D})) - \lambda_1(\kappa_2 W_{\kappa_2}(\hat{D})) \right| &\leq \sqrt{\sum_{i,j} \left(\kappa_1 W_{\kappa_1}(\hat{D}) - \kappa_2 W_{\kappa_2}(\hat{D}) \right)_{i,j}^2} \\ &\leq \sqrt{\sum_{i,j} \left(\frac{\hat{d}_{i,j}}{2a^{3/2}} |\kappa_1 - \kappa_2| \right)^2} \\ &= \sqrt{\sum_{i,j} \left(\frac{\hat{d}_{i,j}}{2a^{3/2}} \right)^2} |\kappa_1 - \kappa_2|. \end{aligned}$$

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

A.2. Proof of Propositions 4.1 and 4.2.

Proof of Proposition 4.1. We prove the Euclidean case (part a) and note that the proofs of parts b and c (spherical and hyperbolic) are nearly identical. Define $\mathcal{R}_n = (-\infty, \delta_n]$. Let $\mathbb{P}_0(A)$ denote the probability of the event A under the null hypothesis that $\mathcal{M}^p(\kappa)$ is Euclidean. By (9),

$$\mathbb{P}_0(\lambda_1(\hat{W}_0) \in \mathcal{R}_n) = \mathbb{P}_0(\lambda_1(\hat{W}_0) \leq \delta_n) = o(1),$$

by assumption. Under H_1 , $\lambda_1(W_0) < 0$ by Proposition 1.1. Since $\delta_n = o_P(1)$,

$$\mathbb{P}_1(\lambda_1(\hat{W}_0) \in \mathcal{R}_n) = \mathbb{P}(\lambda_1(\hat{W}_0) \leq \delta_n) = 1 - \mathbb{P}(\lambda_1(\hat{W}_0) \geq \delta_n) = 1 - o(1).$$

This proves that the test for (6) is consistent, as claimed. \square

Our second result constructs a sequence of cutoffs for the decision rule that is conservative at an α -level that is known to the researcher.

Now we demonstrate that we can calculate a specific and conservative α -bound for the decision rule. That is, we can specify an upper bound for the Type 1 error rate. The idea is to use Weyl's inequality: the eigenvalue of a perturbed matrix can only vary so much—the extent of this depends on the size of the perturbation, which in our case has a known distribution. This result is given in Proposition 4.2, which we now prove.

Proof of Proposition 4.2. We only prove this claim for the Euclidean case, but the same argument proves the claim for the other two geometries. We have by Weyl's inequality that $|\lambda_1(\hat{W}_0) - \lambda_1(W_0)| \leq \|\hat{W}_0 - W_0\|_F$. Then, we have that $\mathbb{P}(|\lambda_1(\hat{W}_0) - \lambda_1(W_0)| < \theta) \leq \mathbb{P}(\|\hat{W}_0 - W_0\|_F < \theta)$ for all θ . Under $H_{0,e}$, $\lambda_1(W_0) = 0$, so we have that $\mathbb{P}(|\lambda_1(\hat{W}_0)| < \theta) \leq \mathbb{P}(\|\hat{W}_0 - W_0\|_F < \theta)$. By setting θ to be the α quantile of $\|\hat{W}_0 - W_0\|_F$, we conclude (12). This completes the proof. \square

Notice if we take a sequence of cutoffs $\theta_{\alpha(n)}^n$ with $\alpha(n) \rightarrow 0$ then the decision rule has will in the limit never falsely reject the null. To make this not vacuous it needs to be chosen at any rate that $\theta_{\alpha(n)}^n \rightarrow 0$ as $n \rightarrow \infty$ which ensures that the thresholds themselves tend to zero. To see how this can be done, see Section 5 of Robin and Smith (2000).

A.3. Proof of Theorem 4.1.

Proof of Theorem 4.1. By assumption, we know that $\hat{D} \xrightarrow{P} D$, so by Proposition 3.1 we have that $\hat{\kappa} \xrightarrow{P} \kappa$. We will use Proposition 4.1 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 4.1, $\widehat{\mathcal{M}}^{\hat{p}}$ 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 $\mathcal{M}^p(\kappa)$ is Euclidean. If $\phi(\hat{W}_0) = 0$ then we fail to reject the hypothesis that $\mathcal{M}^p(\kappa)$ is Euclidean. Similar definitions hold for the spherical and hyperbolic cases.

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

$$\begin{aligned} \mathbb{P}_S(\widehat{\mathcal{M}}^{\hat{p}} = \mathbf{S}^p(\kappa)) &= \mathbb{P}_S(\phi(\hat{W}_0) = 1, \phi(\hat{W}_{\hat{\kappa}}) = 0) \\ &= \mathbb{P}_S(\phi(\hat{W}_0) = 1)\mathbb{P}_S(\phi(\hat{W}_{\hat{\kappa}}) = 0) \\ &\rightarrow 1, \end{aligned}$$

where the notation \mathbb{P}_S indicates that $\mathcal{M}^p(\kappa) = \mathbf{S}^p(\kappa)$ and the third line follows from Proposition 4.1. A similar argument proves that $\widehat{\mathcal{M}}^{\hat{p}}$ is consistent when $\mathcal{M}^p(\kappa)$ is hyperbolic.

We now prove that the estimate \hat{p} is consistent for p . To do this, we must verify the assumptions of Theorem 5.2 of Robin and Smith (2000). Before continuing, we note that in our discussion of Robin and Smith (2000), we keep the original notation given in that paper for consistency. The goal of the Robin and Smith (2000) paper is to estimate the rank of a matrix $p \times q$ matrix B using an estimate \hat{B} . In practice, Robin and Smith (2000) tries to estimate the rank of the matrix $\Sigma B \Psi B^T$ where B is a $p \times q$ matrix, Σ is a $p \times p$ positive definite matrix and Ψ is a $q \times q$ positive definite matrix. Since Σ and Ψ are positive definite, it follows that $\text{rank}(\Sigma B \Psi B^T) = \text{rank}(B)$. In our work, we set both $\Sigma = I_p$ and $\Psi = I_q$ to be the identity matrix. We now list the assumptions in Robin and Smith (2000).

ASSUMPTION A.1. *The matrix B of interest has finite rank.*

ASSUMPTION A.2. *There is an estimator \hat{B} such that*

$$\sqrt{T} \text{vec}(\hat{B} - B) \xrightarrow{d} \text{N}(0, \Omega)$$

as $T \rightarrow \infty$ for some covariance matrix Ω with $\text{rank}(\Omega) = s \in (0, pq]$.

ASSUMPTION A.3. *There are estimators $\hat{\Sigma}$ and $\hat{\Psi}$ with $\hat{\Sigma} - \Sigma = o_P(1)$ and $\hat{\Psi} - \Psi = o_P(1)$.*

Let C be a $p \times p$ matrix containing the eigenvectors of BB^T . We order the entries of C such that $C = (C_{r^*}, C_{p-r^*})$ such that the first r^* columns correspond to the r^* eigenvalues of BB^T and the remaining $p - r^*$ eigenvectors are in C_{p-r^*} . Similarly, we define D to be a $q \times q$ matrix containing the eigenvectors of $B^T B$. We order the entries of D such that $D = (D_{r^*}, D_{q-r^*})$ such that the first r^* columns correspond to the r^* eigenvalues of $B^T B$ and the remaining $q - r^*$ eigenvectors are in D_{q-r^*} .

ASSUMPTION A.4. *If $r^* < q \leq p$ (meaning that the matrix B is not full rank), then we assume that the $(p - r^*)(q - r^*) \times (p - r^*)(q - r^*)$ matrix*

$$(15) \quad A_r := (D_{q-r^*} \otimes D_{p-r^*})^T \Omega (D_{q-r^*} \otimes C_{p-r^*})$$

is non-zero.

The test-statistic based on the eigenvalues (or characteristic roots, as Robin and Smith (2000) calls them) at an estimate r of the rank is given by

$$CRT_r := T \sum_{i=r+1}^q h(\hat{\lambda}_i),$$

where h is an arbitrary function and $\hat{\lambda}_i$ are the eigenvalues of the matrix $\hat{\Sigma} \hat{B} \hat{\Psi} \hat{B}^T$. The following assumption restricts the set of admissible functions h .

ASSUMPTION A.5. *The function h is non-negative, finite and posses continuous derivatives at least of order 1 with $h(0) = 0$ and $h'(0) = 1$.*

Some examples of admissible h functions include $h(z) = z$.

ASSUMPTION A.6. *There is an estimator $\hat{\Omega}$ with $\hat{\Omega} - \Omega = o_P(1)$.*

We now define the estimator from [Robin and Smith \(2000\)](#). Let Z_1, \dots , be i.i.d. standard normal random variables and let $\{\lambda_i^r\}_{i=1}^{(p-r)(q-r)}$ be the ordered eigenvalues of the matrix $A_r = (D_{q-r} \otimes D_{p-r})^T \hat{\Omega} (D_{q-r} \otimes C_{p-r})$. Then, for $\alpha \in (0, 1)$, define $c_{1-\alpha}^r$ as the solution to

$$(16) \quad \mathbb{P} \left(\sum_{i=1}^r \lambda_i^r Z_i^2 \geq c \right) = 1 - \alpha ,$$

where here the super-script r indicates that these eigenvalues come from the matrix A_r and does not indicate that the eigenvalues are being raised to the power r . This is the notation from [Robin and Smith \(2000\)](#), which we use in this section for consistency. Our estimate of the rank is then

$$(17) \quad \hat{r} = \min_{r \in \{0, 1, \dots, q-1\}} \{r : CRT_i \geq \hat{c}_{1-\alpha iT}^i \text{ for } i = 0, \dots, r-1 \text{ and } CRT_r < \hat{c}_{1-\alpha r T}^r\}$$

In words, the estimate \hat{r} is the first index r such that CRT_r is smaller than the quantile \hat{c}^r defined in (16).

PROPOSITION A.1 (Theorem 5.2 of [Robin and Smith \(2000\)](#)). *Suppose that $r^* < q$ and that Assumptions A1-A6 hold. Then, the estimator \hat{r} defined in (17) is consistent for r^* , provided that $\alpha_T = o(1)$ and $-T \log(\alpha_{rT}) = o(1)$.*

We now prove that these assumptions hold in our problem in order, returning to our notation. First, the matrix B in our case is W_κ . The rank of W_κ is clearly finite, so Assumption A.1 holds. By the assumptions in Theorem 4.1, we can use the delta method to show that Assumption A.2 is satisfied, since in all cases, W_κ is a differentiable transformation of D . Assumption A.3 since we set $\Sigma = \Psi = I$. We discuss Assumption A.4 at the end of this proof. Assumption A.5 is satisfied by taking, say, $h(x) = x$. Assumption A.6 is satisfied by the assumptions of Theorem 4.1, since if we have a consistent estimate of Ω , then we can apply the delta method and obtain a consistent estimate of $W_\kappa(D)$.

Finally, we show that Assumption A.4 is satisfied. To do this, note that the discussion immediately after Assumption 2.4 in [Robin and Smith \(2000\)](#) says that this Assumption is satisfied if

$$p \in [0, K - \frac{1}{2} \sqrt{4K^2 - 4\text{rank}(\Omega_i)}] .$$

One can show that $\text{rank}(\Omega_i) = \binom{K}{2}$. By the assumption given in the theorem, K is chosen such that the above statement is true. Therefore, we have that Assumption A.4 is satisfied. We have shown that all of the assumptions in Theorem 5.2 of [Robin and Smith \(2000\)](#). Therefore, we can conclude that \hat{p} is consistent for the true rank of W_κ . This completes the proof. □

A.4. Proof of Theorem 4.2.

Proof of Theorem 4.2. If we can show that $\hat{D} - D \xrightarrow{p}$, then we can use Proposition 3.1 to conclude that $\hat{\kappa} \xrightarrow{p} \kappa$. By using the same argument as that given in the proof of Theorem 4.1 to argue that \hat{p} is consistent for p . In addition, by Proposition 4.1 and Theorem 4.1 we have that $\widehat{\mathcal{M}}^{\hat{p}}$ is consistent, $\mathbb{P}(\widehat{\mathcal{M}}^{\hat{p}} \neq \mathcal{M}^p(\kappa)) = o(1)$. Further, given a consistent estimator of $\widehat{\mathcal{M}}^{\hat{p}}$,

we can use results from Breza et al. (2019) to show consistency of the model parameters. In constructing this proof, therefore, we first claim that it is possible to estimate a distance matrix such that its deviation from the true distance matrix among a collection of points converges in probability to zero using the properties of the model in (1). This construction, along with Proposition 4.1 and Theorem 4.1, is sufficient for the first two statements in the proof. We then turn to the third, consistency of the model parameters, and demonstrate how it follows from the results in Breza et al. (2019).

We now seek to show that the deviation of a distance matrix estimated using cliques from the true distance matrix converges in probability to zero. Recall from Section 2 that it is possible to construct an estimator that converges in probability under a model with two simplifying assumptions. Relaxing these assumptions, (1) yields, after marginalizing the individual effects,

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

Solving for the distances $d(z_i, z_j)$,

$$(18) \quad d_{k,k'} = -\log(p_{k,k'}) + 2 \log(E(\exp(\nu_i))) .$$

We assume the existence of a consistent estimator of $E(\exp(\nu_i))$. Though it works well in practice, the estimator we present in Section 2 suffers from selection bias due to only using nodes that have many connections (since they are part of near cliques). If we were to make a distributional assumption on the ν_i 's we could correct for this selection bias by adjusting the estimate using the tail probability of assumed distribution. We could also explore a nonparametric estimate that is consistent by using subgraphs of different sizes in an attempt to mitigate the selection bias. Since the approach from Section 2 works well in practice we leave these as areas for future work and proceed here with the assumption that we have used of the two methods described here to drive a consistent estimator.

Moving now to estimating the $p_{k,k'}$ term, recall that in Section 2 we made the simplifying assumption that all individuals had latent positions that were on one of a small number of points. We now instead follow the full generality of (1) and allow each individual to have their own latent position. We argue that, once the graph becomes sufficiently large, the distances between nodes in completely connected cliques becomes negligible and, thus, we can treat each clique as though it were a point in the simplified model presented in Section 2.

LEMMA A.1. *Take D as a $K \times K$ distance matrix of between-clique distances based on the Fréchet mean of the latent space positions of the nodes in each clique. Assume that Assumptions 1.1-1.3 hold, and that G is distributed as in (1). Let $\widehat{\mathcal{M}}^{\hat{p}}$ denote the estimate of the geometry based on the tests from (9), (10), and (11). Using $\widehat{\mathcal{M}}^{\hat{p}}$, let $\hat{\kappa}$ and \hat{p} denote the estimates of the curvature and dimension. Let \hat{z}_i and $\hat{\nu}_i$ denote the manifold-specific maximum likelihood estimator for z_i and ν_i defined in (13), using the estimated latent space type, dimension, and curvature. Finally, suppose that there exists a consistent estimator of the mean of the individual effects distribution, $\hat{\gamma}$, such that $\hat{\gamma}_\ell \xrightarrow{p} E(\exp(\nu_i))$. Let $C_k = (V_{C_k}, E_{C_k})$ be a subgraph clique with $|V_{C_k}| = \ell$ and let K be the number of such cliques. Under these conditions,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\|\hat{D} - D\|_F > \epsilon \right) = 0$$

Proof of Lemma A.1. There are two main points in the proof of Lemma A.1. First, unlike the case in Theorem 4.1 where D is a (fixed) set of positions on the (unknown) manifold, there is an additional layer of randomness in the case of the graph. Specifically, since distances manifest as counts of links, the “true” distance matrix, D based on the Fréchet mean of the latent space positions of the nodes in each clique, and calculate the distances between these points, depends on the graph, which is a single draw from the network formation model in (1). The first task in this proof, therefore, is to show that, for any realization from the graph generating process (1), the observed counts are a sufficient approximation of the underlying (random variable) D . Presented in more detail below, this argument uses the assumption that we have a consistent estimator of the expectation of the individual effects and relies heavily on the idea that, conditional on the latent positions the probabilities of forming ties are independent, which allows us to treat links between cliques as a series of independent Bernoulli trials.

The above argument establishes that the variation in the observed between-clique frequencies vanishes sufficiently quickly as the size of the graph grows. The second issue we must address for this result to hold arises in how we compute these cross-clique frequencies. In contrast to the example we give in Section 2 where all latent positions have the same point in the latent space, in estimating \hat{D} each person has their own latent position. Intuitively people in cliques should be close together in the latent space (since if they weren’t close together they would be unlikely to be connected). To show convergence of the distance matrix, we need to argue that the distances between people in the same clique become sufficiently small as the size of the graph (and thus ℓ) grows that any bias becomes negligible. To establish this result we use the model from (1) to compute the likelihood of observing a clique under the assumption that all members of a clique are within an arbitrarily small distance, δ , compared to the likelihood that one member of the clique is outside the δ -window. A critical point for this logic is that the cliques are observed, so the computations are true likelihood computations rather. That is, we ask, given a model, how likely are the data we have already observed. Though straightforward algebra we show that the ratio of these likelihoods diminishes as the size of graph (and thus cliques) move towards infinity.

We now move to the formal proof. First, consider the Frobenius (vector) norm of the distance between D and \hat{D} :

$$\begin{aligned} \|\hat{D} - D\|_F &= \sum_{k,k'} \left| -\log(\hat{p}_{k,k'}) + 2\log(\hat{\gamma}) + \log(p_{k,k'}) - 2\log(E(\exp(\nu_i))) \right| \\ &\leq \sum_{k,k'} \left| \log(p_{k,k'}) - \log(\hat{p}_{k,k'}) \right| + 2 \sum_{k,k'} \left| \log(\hat{\gamma}) - \log(E(\exp(\nu_i))) \right|. \end{aligned}$$

where the second line follows from the triangle inequality. By the continuous mapping theorem, $\log(\hat{\gamma}) \xrightarrow{p} \log(E(\exp(\nu_i)))$. Thus, since K does not change with ℓ , the second term above goes to zero in probability. We now consider the first term in the second line.

We now recall the definition of $\hat{p}_{k,k'}$,

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

Each G_{ij} is a Bernoulli trial, so $\text{Var}(G_{ij}) \leq 1/4$. In addition, the G_{ij} are all mutually independent, so

$$\text{Var}(\hat{p}_{k,k'}) = \frac{1}{\ell^2} \sum_{i \in C_k} \sum_{j \in C_{k'}} \text{Var}(G_{ij}) \leq \frac{1}{4\ell^2} \rightarrow 0$$

as $\ell \rightarrow \infty$.

Given that the variance of $\hat{p}_{k,k'}$ vanishes as the graph size increase based on the arguments above, we now move to show that the estimate based on cross-clique interactions converges to the correct point.

Recall from Section 2 that if the all points have the same latent position then $\hat{p}_{k,k'}$ is unbiased and the weak law of large numbers gives the desired result. Assumption 1.3 provides the final step to show the result. That is, using Assumption 1.3 we show that indeed all members of the clique are going to be located arbitrarily close together in the limit. To see this, we can calculate the relative likelihood of nodes being within or outside a δ -ball as $n, \ell \rightarrow \infty$. Specifically, observe that for any $\delta > 0$, by Assumption 1.3,

$$\begin{aligned} \frac{\mathbb{P}(\max_{i,j \in C_\ell} d(z_i, z_j) \leq \delta | C_\ell \text{ exists})}{\mathbb{P}(\max_{i,j \in C_\ell} d(z_i, z_j) > \delta | C_\ell \text{ exists})} &= \frac{\mathbb{P}(C_\ell \text{ exists} | \max_{i,j} d(z_i, z_j) \leq \delta)}{\mathbb{P}(C_\ell \text{ exists} | \max_{i,j} d(z_i, z_j) > \delta)} \\ &\times \frac{\mathbb{P}(\max_{i,j} d(z_i, z_j) \leq \delta)}{\mathbb{P}(\max_{i,j} d(z_i, z_j) > \delta)} \\ &\rightarrow \infty \end{aligned}$$

from which the result follows. □

Using Lemma A.1, we can now leverage the same estimator $\hat{p}_{k,k'}$ as in Section 2 and apply a similar logic, appealing to the weak law of large number and continuous mapping theorems, to get the desired result. We move now to the final part of the proof, namely that, given a consistent estimate of geometry type and dimension, we can consistently estimate the remaining parameters using their maximum likelihood estimates. To prove this claim, we apply the following result from Breza et al. (2019). We have slightly rewritten this result in the notation of the current paper.

LEMMA A.2 (Lemma A.1 from Breza et al. (2019)). *Suppose that we observe complete graph data generated from a formation model noted as a Continuous Latent Space (CLS) model in Shalizi and Asta (2017). Specifically, let*

$$\mathbb{P}(g_{ij} = 1 | \nu, z, \beta) \propto \exp\{\nu_i + \nu_j + \beta^T X_{ij} - d_{\mathcal{M}^p}(z_i, z_j)\}$$

where $d_{\mathcal{M}^p}(z_i, z_j)$ is the distance between the latent position of person i and person j on the latent simply connected, complete Riemannian manifold of constant curvature κ , $\mathcal{M}^p(\kappa)$. Our goal is to show that we can consistently estimate ν, z, β given $\mathcal{M}^p(\kappa)$. Let $X_{ij} \in \mathbb{R}^h$ so $\beta \in \mathbb{R}^h$. Let $V^n \subset (-\infty, 0]^n$ a compact subset with $(\nu_1, \dots, \nu_n) \in V^n$. Then, under the same conditions as Shalizi and Asta (2017), we have

$$\max_{1 \leq i \leq n} \left(|\hat{\nu}_i - \nu_i^0| + d(\hat{z}_{1:n}, z_{1:n}) + |\hat{\beta} - \beta| \right) = o_P(1)$$

as $n \rightarrow \infty$, where

$$d(z_{1:n}, z_{1:n}^0) = \inf_{\phi \in \text{isom}(M)} \sum_{i=1}^n d_{\mathcal{M}^p}(z_i, \phi(z_i))$$

where $\text{isom}(\mathcal{M})$ is the set of isometries on $\mathcal{M}^p(\kappa)$ as in [Shalizi and Asta \(2017\)](#) and

$$(\hat{\nu}, \hat{z}, \hat{\beta}) = \underset{\nu, z, \beta \in V^n \times \prod_{i=1}^n \mathcal{M}^p \times \mathbb{R}^h}{\text{argmax}} \ell(\nu, z, \beta),$$

the maximum likelihood estimates.

Since we have that $P(\widehat{\mathcal{M}}^{\hat{p}} = \mathcal{M}) = 1 - o(1)$, the result follows. \square

APPENDIX B. 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

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

By solving for κ , we see that κ satisfies

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

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 7 plots the function $\kappa \mapsto \left| \lambda_1(\kappa W_\kappa) \right|$ when D corresponds to $K = 15$ points drawn randomly on $\mathbf{S}^2(1)$. Figure 7 also plots the function $\kappa \mapsto \left| \lambda_{K-1}(\kappa W_\kappa) \right|$ 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), matching the intuition from (5).

APPENDIX C. RANK ESTIMATOR

In Algorithm 1 we formally describe the algorithm and the estimate of the rank of W_κ . The [Luo and Li \(2016\)](#) estimator uses two pieces of information. The first is the scree function, which plots the sample eigenvalues in order from largest to smallest. In Figure 8 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 1. [Luo and Li \(2016\)](#) argues 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 behavior in Figure 8: 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

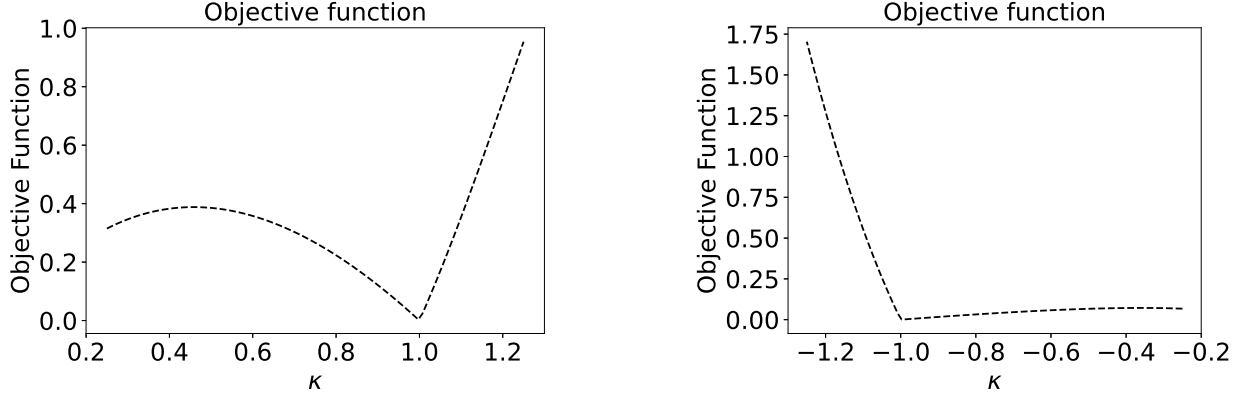


FIGURE 7. Plot of the objective function from (5) when D corresponds to 15 points in $\mathbf{S}^2(1)$ (left) and 15 points in $\mathbf{H}^2(-1)$ (right). On the horizontal axis we plot the curvature κ and on the vertical axis we plot the value of the function $\kappa \mapsto \left| \lambda_1(\cos(\sqrt{\kappa}D)) \right|$. We see that at the true κ , the objective function is minimized.

Algorithm 1: Estimating Rank of W_κ

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

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

- (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

$$\hat{r} = \arg \min_{j \in \{0, 1, \dots, K-2\}} (\phi_n(j) + f_n(j)).$$

on these two pieces of information, [Luo and Li \(2016\)](#) suggests 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 D. BOOTSTRAP PROCEDURE

D.1. A Bootstrap Test for Geometry. We now provide a bootstrapping method to test the three hypotheses in (6), (7), and (8). Before we describe our bootstrapping method, we

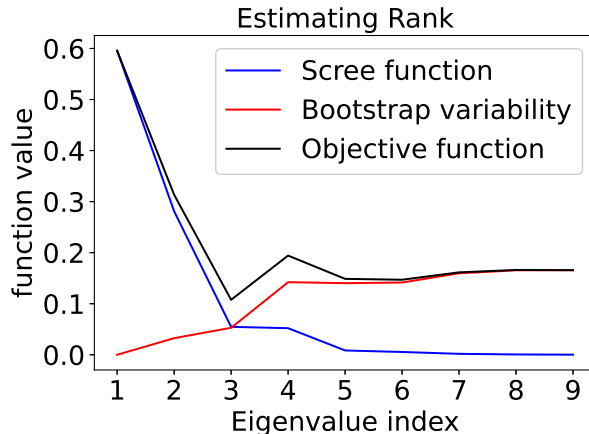


FIGURE 8. We generate a graph using a 3-dimensional Euclidean latent space with $K = 10$ cliques. We plot the scree function ϕ and the bootstrap variability function f_n defined in Algorithm 1. We also plot their sum, defined as the objective function. The horizontal axis represents the possible ranks 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.

describe two problems that make bootstrapping the eigenvalues of W_κ challenging. First, $W_0 \in \mathbb{R}^{K \times K}$ does not have full rank, because

$$\text{rank}(W_0) = \text{rank}(JD \circ DJ) \leq \min(\text{rank}(J), \text{rank}(D)) \leq K - 1$$

where we have used the fact that $\text{rank}(J) = K - 1$. So $\lambda_1(W_0)$ must lie in $(-\infty, 0]$ and under $H_{0,e}$, $\lambda_1(W_0)$ lies on the boundary of the parameter space. Classical bootstrapping in such a case is not valid (Andrews, 2000). The second problem is that W_κ have repeated eigenvalues at zero. Again, classical bootstrapping also does not work in this case (Eaton and Taylor, 1991).

To address these problem and provide a bootstrapping method that delivers strong performance, we will use the sub-sampling method from Politis and Romano (1994), which is valid in a broader set of problems. In Politis and Romano (1994), the data are independent and identically distributed. In our case, they are independent but *not* identically distributed, therefore the coverage guarantees we present in Appendix D.2 are not expected to hold exactly, though in practice these results do hold approximately. We break our method up into two distinct algorithms. The first, Algorithm 2, provides a method to bootstrap the distance matrix D . The second algorithm, Algorithm 3, uses the bootstrapped distance matrix to test the three hypotheses about the underlying geometry. This second algorithm is based on the methods in Politis and Romano (1994). The input to this algorithm is the adjacency matrix G , the number of bootstrap samples B , the geometry under the null hypothesis, and two parameters that appear in the Politis and Romano (1994) method.

D.2. Additional details on the bootstrap procedure. Given n independent and identically distributed data points X_1, \dots, X_n drawn from a distribution H , we want to estimate a parameter $\theta = \theta(H)$ with an estimator $\hat{\theta}_n$. We make the following assumption about $\hat{\theta}_n$, which appears in Politis and Romano (1994).

Algorithm 2: Bootstrapping Distance Matrix. See Appendix D.2 for additional details.

Input: adjacency matrix G , number of bootstrap samples B and sub-sample rate m

- (1) Let $\mathbf{I} := I_1, \dots, I_m$ and $\mathbf{J} := J_1, \dots, J_m$ denote two sets of m -length integers drawn independently and uniformly from $\{1, \dots, \ell\}$ with replacement.
- (2) Compute the bootstrapped between-clique probability matrix P^* by first estimating $p_{k,k'}^*$ where

$$p_{k,k'}^* = \frac{1}{m} \sum_{i,j=1}^n G_{ij} \mathbf{1}\{i \in C_k[\mathbf{I}], j \in C_{k'}[\mathbf{J}]\}$$

where $C_k[\mathbf{I}] := C_k \cap \mathbf{I}$ and $C_{k'}[\mathbf{J}] := C_{k'} \cap \mathbf{J}$ where C_k is the set of nodes in clique k . Then set $P_{k,k'}^* = \max(1/\ell^2, p_{k,k'}^*)$.

- (3) Then set $D_b^* = -\log(P^*/\hat{E}_\nu)$ where the division is component-wise.
- (4) **return** $\{D_b^*\}_{b=1}^B$

Algorithm 3: Hypothesis Testing Geometry via Bootstrapping

Input: adjacency matrix G , number of bootstrap samples B and sub-sample rate m

- (1) Compute the observed eigenvalue $\lambda_{k^*}(\hat{W})$,
- (2) For $b = 1, \dots, B$, do the following:
 - (a) Sample D_b^* from Algorithm 2 and compute W_b^* based on the null hypothesis.
 - (b) Compute the eigenvalue $\lambda_{k^*}(W_b^*)$
- (3) Compute

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

- (4) Compute $c_n(1 - \alpha) = \inf\{x : \hat{L}_n(x) \geq 1 - \alpha\}$ be the $(1 - \alpha)\%$ percentile of $m^2 (\lambda_{k^*}(W_i^*) - \lambda_{k^*}(\hat{W}))$.
- (5) Reject H_0 when

$$\ell^{2\alpha} \lambda_{k^*}(\hat{W}) < c_n(\alpha) \text{ if the null hypothesis is Euclidean or spherical}$$

and we reject H_0 when

$$\ell^{2\alpha} \lambda_{k^*}(\hat{W}) > c_n(1 - \alpha) \text{ if the null hypothesis is hyperbolic.}$$

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, \dots, X_n . To do this, we select a sub-sample rate $m = m(n)$, where $m \leq n$. Then, let $Y_1, \dots, Y_{\binom{n}{b}}$ be all the subsets of X of size b , and let $\hat{\theta}_{n,i}$ be the estimate of θ using the i th 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 $\tau_b(\hat{\theta}_{n,i} - \hat{\theta}_n)$,

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

Intuitively, as n and $b \rightarrow \infty$, 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 $\tau_n(\hat{\theta}_n - \theta)$, 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 and 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 $\tau_b/\tau_n \rightarrow 0$ and $b/n \rightarrow 0$ then*

$$\mathbb{P}\left(\tau_n(\hat{\theta}_n - \theta) \leq c_n(1 - \alpha)\right) \rightarrow 1 - \alpha.$$

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 $\{Y_1, \dots, Y_s\}$ for some integer $s \leq \binom{n}{b}$, and compute

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

According to [Politis and Romano \(1994\)](#), we have the following result:

PROPOSITION D.2 (Theorem 2, (iii) of [Politis and 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 $\tau_b/\tau_n \rightarrow 0$ and $b/n \rightarrow 0$, then*

$$\mathbb{P}\left(\tau_n(\hat{\theta}_n - \theta) \leq \hat{c}_n(1 - \alpha)\right) \rightarrow 1 - \alpha.$$

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 $\lambda_{k^*}(W)$. 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 $\bigcup_{k=1}^K C_k$, the union of all K cliques. We fix some sub-sample rate m . In Section D, we describe how to do this. 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 $C_k \cup C_{k'}$. 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 $\{1, \dots, \ell\}$, without replacement. These indices will be the re-sampled nodes. We then compute

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

Since it is possible that $P_{k,k'}^*$ is zero (meaning that the re-sampled pairs of nodes do not connect), we use $P_{k,k'}^* = \max(1/\ell^2, P_{k,k'}^*)$, 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^* using (1). We provide a step-by-step implementation of the sub-sampling method in Algorithm 2.

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

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

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 $(1 - \alpha)\%$ percentile of $m^{2\alpha} (\lambda_{k^*}(W_i^*) - \lambda_{k^*}(\hat{W}))$. Then, from Proposition 4.2, we know $\mathbb{P}(\ell^{2\alpha} (\lambda_{k^*}(\hat{W}) - \lambda_{k^*}(W)) \leq c_n(1 - \alpha)) \approx 1 - \alpha + o(1)$ for large ℓ . This motivates the bootstrapping method we summarize in Algorithm 3.

APPENDIX E. ADDITIONAL SIMULATION RESULTS

We now give 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).

We now analyze the accuracy of the curvature methods for the spherical and hyperbolic latent space models. The estimator in Proposition 3.1 minimizes $\kappa \mapsto \lambda_1(\kappa \hat{W}_\kappa)$. But from Proposition 1.2, 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

$$(19) \quad \hat{\kappa}(q) = \frac{1}{q} \sum_{i=1}^q \hat{\kappa}_i, \quad \hat{\kappa}_i = \arg \min_{\kappa \in [a, b]} \left| \lambda_i \left(\kappa W(\hat{D})_\kappa \right) \right|$$

Assuming that $q \ll K$, we can reasonably believe that the first through q th eigenvalues of $\cos(\sqrt{\kappa} D)$ are zero. In fact, it is easy to modify the proof of Proposition 3.1 to show that $\hat{\kappa}(q) \xrightarrow{p} \kappa$, provided that $t \ll K$. Taking $q > 1$ does not always reduce the variance of $\hat{\kappa}(t)$, which could be because the $\hat{\kappa}_1, \dots, \hat{\kappa}_t$ are not necessarily independent. In Figure 9 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 $K = 10$. Although Proposition 3.1 says that the estimate is consistent as the sample size grows, we see that in finite samples this estimator performs poorly. In the future we would like to determine an estimator that performs better in finite samples.

APPENDIX F. 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 centers. we then call the first n/K nodes to be in group 1, the second

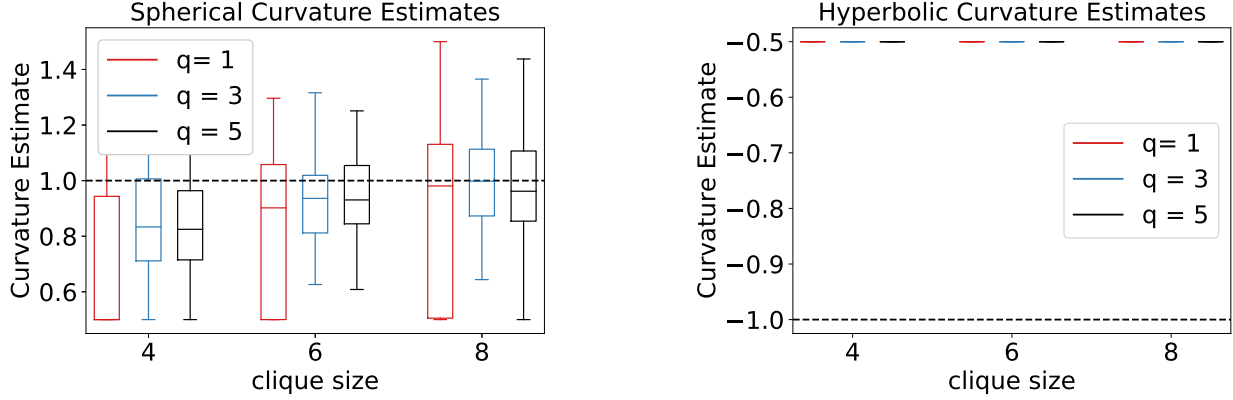


FIGURE 9. Left: Curvature estimates for $\mathbf{S}^2(1)$ using $K = 10$ cliques, with clique size $\ell = 4, 6, 8$ on the horizontal axis. We use $q = 1, 3, 5$ where q is defined in (19). We plot the true curvature $\kappa = 1$ in the black dashed line. Right: Curvature estimates for $\mathbf{H}^2(-1)$ using $K = 10$ cliques, with clique size $\ell = 4, 6, 8$ on the horizontal axis.

n/K nodes to be in group 2, and so on. Let $c_i \in \{1, \dots, K\}$ denote the group membership of node i . Finally, we distribute the node latent space positions centered at their group locations according to some procedure that is unique for each of the three geometries. To generate the LS positions in the Euclidean case, we do the following:

- (1) Generate K group centers $\mu \in \mathbb{R}^p$ distributed according to $\mu \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(\mathbf{0}_p, \sigma^2 I_p)$.
- (2) Then simulate the positions of the nodes as $z_i | c_i \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(\mu_{c_i}, \frac{\sigma^2}{K} I_p)$.

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

- (1) Generate K group centers $\mu \in \mathbf{S}^2(\kappa)$. To do this, we generate two angles: $\theta \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \pi)$ and $\phi \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, 2\pi)$. Then compute

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

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

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

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

- (1) Generate K group centers $\mu \in \mathbf{H}^2(\kappa)$. 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 the table below, 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.

TABLE 3. 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, $\kappa = 1$

APPENDIX G. ADDITIONAL DETAILS FOR THE BANERJEE ET AL. (2019) DATA

In Figure 10 we give violinplots of the number of cliques of size $\ell \in \{4, 5, 6\}$ in the Indian village data set. The median values are 132, 13, and 0, respectively. The variances are 14797, 703, and 12, respectively. We show commutative distribution plots of the number of cliques across the 75 villages in Figure 11.

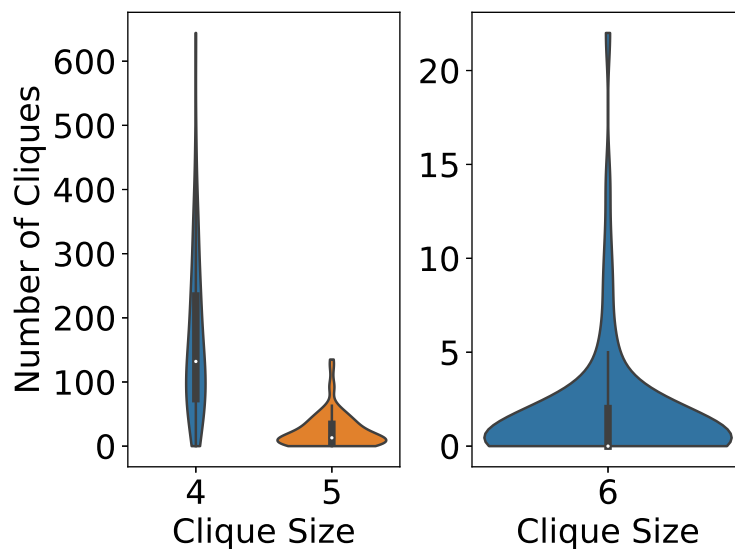


FIGURE 10. Number of cliques of size $\ell \in \{4, 5, 6\}$ for the Indian village data set. The median values for $\ell = 4, 5,$ and 6 are 132, 13, 0, respectively.

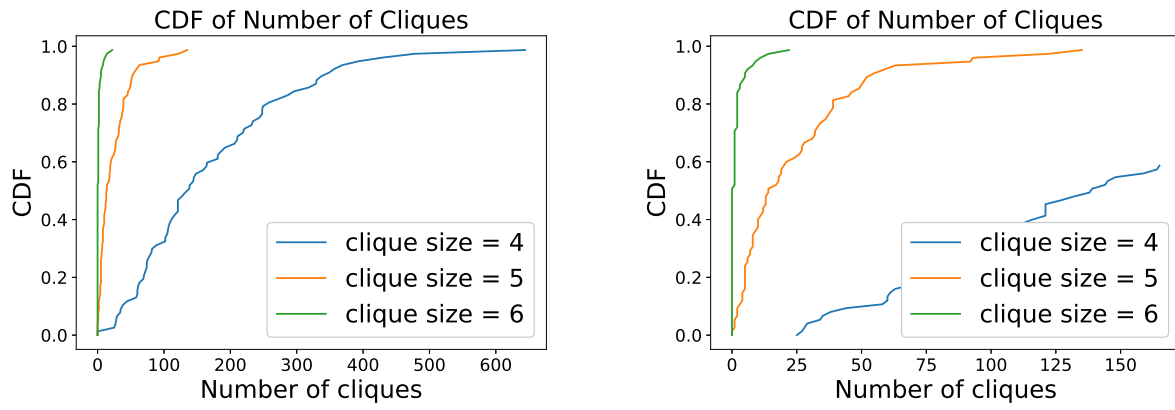


FIGURE 11. CDF of number of cliques for clique sizes $\ell \in \{4, 5, 6\}$ for the 75 Indian villages.