Geometric Network Inference

Dena Marie Asta
Carnegie Mellon University
dasta@andrew.cmu.edu

Abstract

Network analysis has a crucial need for tools to compare networks and assess the significance of differences between networks. Currently, network comparison usually amounts to the ad hoc comparison of common descriptive statistics on graphs (e.g., average shortest path lengths). Such comparisons fail to detect substantial differences in what, at least intuitively, are the large-scale geometric features of networks (e.g. grid-like or tree-like, clusterability). Some of the recent literature, on the other hand, has demonstrated how many real-world networks tend to resemble graphs formed by randomly sampling points on (negatively curved) hyperboloids and connecting them with probabilities decreasing with geodesic distance. For example, the parametric inference of a quasi-uniform density on a hyperboloid from samples of networks effectively detects large-scale geometric features important in assessing routing algorithms for computer networks. We propose a more general, principled statistical approach to network comparison, based on the non-parametric inference and comparison of densities on hyperbolic manifolds from sample networks.

The first phase of our project is to infer the generating densities of nodes from observed graphs. An existing adaptation of multidimensional scaling optimally embeds the nodes of such graphs into hyperbolic manifolds. Non-parametric density estimation on such hyperbolic manifolds then infers the generating densities. This has motivated our development of a generalization of kernel density estimation for a large class of symmetric spaces — Riemannian manifolds with sufficiently many isometries, including the n-dimensional hyperboloid. Using the Helgason-Fourier transform, a generalization of the Fourier transform for such symmetric spaces, we establish the minimax rate for our estimator, matching the classical rate for a given smoothing parameter and local dimension.

The second step of this project is to develop and validate significance tests for differences in network structure by comparing the L2-differences between estimated densities with L2-differences of pooled estimated densities. Power will be assessed empirically and estimated theoretically. Proof-of-concept applications will demonstrate the utility of such tests.

Contents

1 Introduction

The qualitative, large-scale, structure of real-world networks — like collaboration networks, financial networks, and online social networks— undoubtedly contain information of interest – for example in innovation policy, economic behavior, and public health. Such networks are often massive, dynamic in size and connectivity, and noisy. Evidence-based policy requires statistical tools for measuring and interpreting such *statistical networks*. For example, simple and interpretable statistics on collaboration networks could prove useful in the development and evaluation of innovation policies. However, current summary statistics on networks leave big gaps in our toolkit.

Many real-world networks of interest in policy and engineering continuously change over time and conditions. Thus, evidence-based policy requires tools for *comparing* networks. Current methods of summary statistics are insufficient in accessing significance of changes or differences in networks. Moreover, the networks that we observe in the real-world are noisy. The overall theme of this thesis is that hyperbolic geometric models of growing networks give a natural means for network inference (including resampling), network comparison (including significance testing) and subsequent applications in such areas as public health.

2 Motivation

Hyperbolic geometry offers a natural setting for analyzing the global structure of real-world networks. A rigorous and global treatment of networks necessitates a careful choice of generative model §2.1. More specifically, real-world networks suggest for us a continuous latent space model §2.2 when the only relevant differences between nodes are their interconnections to all other nodes. Even more specifically, the precise properties (e.g. tree-like structure) of real-world networks of interest suggest a specific hyperbolic continuous latent space of nodes §2.3.

2.1 Real-world networks

The structure of naturally occurring networks, collections of nodes and relations between them, can illuminate phenomena otherwise hidden in the real-world. One example of interest in research innovation policy is a collaboration network, such as a group of scientists and co-authorship relationships between them. Another example of interest in financial regulation is a financial network, a collection of financial institutions and the various inter-bank lending that relate different pairs of such agents. Yet another example of interest in health policy is an online messaging network, a collection of people and messages between them that can potentially indicate a disease outbreak in the offline world. Thus, we want to capture the salient information from the structure of such networks.

2.1.1 Background

Statistics measuring differences between networks are abundant in network analysis literature. One class of such statistics are differences between numerical statistics, such as average path length or density of triangles. However, such network summarization statistics fail to distinguish between networks exhibiting, say, grid-like versus tree-like structure [11]. Moreover, such tests of significance do not seem to detect heterogeneity in some examples [22]. Choices of such numerical statistics to explain networks tend to be also ad-hoc.

2.1.2 Goal

Instead, we would like methods of network comparison that intuitively distinguish between salient network structure of interest in applications. In particular, we often want to assess when one network significantly differs from another network in a meaningful manner, without regard to node set or node size. A significant difference between collaboration networks for different research domains might suggest some difference in the effectiveness of different grant policies. A critical change in the interdependency of financial institutions

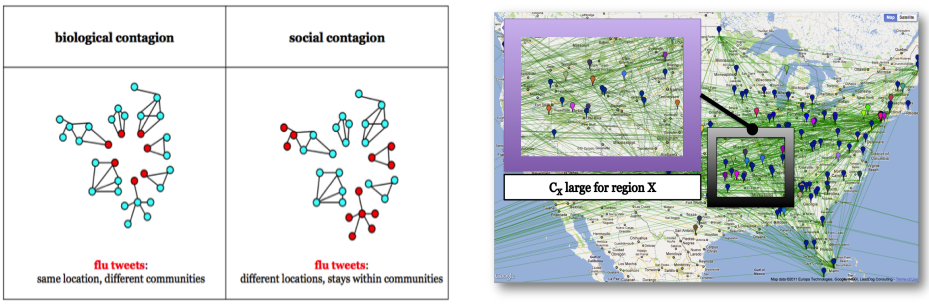

Figure 1: **Contagion: Biological versus Social** Previous work in [3] of the author and C. Shalizi suggests that numbers of communities in social networks of U.S. Twitter users tweeting the word `flu` in different geographical regions closely track corresponding CDC reported flu counts. The left figure illustrates a hypothetical difference between a sample of Twitter users reporting on having flu symptoms (biological contagion) and a sample of Twitter users merely discussing the flu (social contagion). The red circles indicate Twitter users tweeting the word `flu` and blue circles indicate other Twitter users. The right figure represents an actual snapshot of the social graph of `flu`-tweeters for Aug 24, 2011, zooming in a particular geographical region. The different colors represent different communities to which tweeters belong. The number of such colors is shown in [3] to predict flu levels for that region.

might predict a catastrophic collapse in the financial sector. A significant difference between networks of online social messages mentioning the word `flu` from one day to the next may mean all the difference in detecting social versus biological contagion. Methods of network resampling developed along the way should give us a means of assessing the general uncertainty of our network inference.

We want tests of significance that are not overly sensitive to differences in sizes and labelling in node set and noise but can detect differences in qualitative structure. Random graphs having a fixed set of nodes are representable as expected *adjacency matrices*, square matrices whose (i, j)th entry represents the probability that an edge connects nodes i, j. Normed differences $\|M_G - M_H\|$ between expected adjacency matrices M_G, M_H generating graphs G, H, measure structural differences when the graphs *share the same set of nodes*. We often need to assess when two networks share a similar geometric structure *despite having different numbers of nodes and different labels for those nodes*. For example, we might want to compare daily graphs of users tweeting the word `flu` to detect any alarming change in structure; such graphs would have different node sets from day to day. For another example, we want to detect when a growing and dynamic online social network exhibits a critical change in structure; the working assumption is that most network growth is not a critical change. We cannot, for example, take differences of adjacency matrices having different dimension. Moreover, we would have to compare our adjacency matrices up to *exchangeability* [10], even if our node sets have comparable size. We need a *generative model of networks*, unlike expected adjacency matrices, amenable to pairwise comparisons regardless of differences in respective node sets. A challenge is satisfying such criteria in a principled, robust, and computable manner.

2.2 Generative models

Generative models of networks are classes of probability distributions on graphs — prescriptions for generating graphs with certain probabilities. A generative model should explain something about the underlying process generating the actual networks observed. For example, there might exist two different generative models for networks of social online messages — the expression of social contagion and the expression of biological contagion. For another example, different generative models for collaboration networks might explain different research and development policies generating them. Therefore we want to choose generative models that are simple to compare and generate precisely the types of real-world networks of interest to us.

2.2.1 Background

Two common generative models useful for creating and comparing networks of arbitrary sizes are *graphons* and *continuous latent spaces*. Graphons are more general than continuous latent space models. However, continuous latent spaces are still general enough to generate interesting real-world networks but restrictive enough so that simple and elegant tools from geometry can be used for network comparison.

A *graphon* is a measurable function $W : V \times V \to [0, 1]$ for V a probability measure space [20]. A network of size n is generated by a graphon $W : V \times V \to [0, 1]$ by independently sampling n points from V according to the probability measure on V and attaching edges with probability according to W. In other words, V represents a space of vertices and the graphon W represents edge probabilities. This type of generative model can be inconvenient for purposes of network comparison. The *cut distance* $\delta(W_1, W_2)$, the standard notion of distance between two graphons $W_1, W_2 : V \times V \to [0, 1]$, is

$$\delta(W_1, W_2) = \inf_{\varphi : V \cong V} \sup_{A,B} \int_A \int_B |W_1(\varphi(a), \varphi(b)) - W_2(a, b)| \, db \, da, \tag{1}$$

where the infimum is taken over all measure-preserving bijections $\varphi : V \cong V$ [20]. An efficient approximation of the cut metric above presents special challenges, even if the infimum can be taken over a countable set for $V = [0, 1]$ [15, Theorem 6.9 (iv)].

Switching the role of vertices and edges, *continuous latent space models* are another simple generative model for networks of arbitrary size. Consider a metric space V. For each distribution on V, we independently pick nodes from V according to that distribution and then attach edges between nodes $x, y \in V$ with probabilities $W(x, y)$ inversely related to distances in V. A pair of node densities — unlike a pair of adjacency matrices — are comparable regardless of differences in graph size and independent of vertex labelling. A notion of distance analogous to (1) between two densities $f_1, f_2 : V \to \mathbb{R}$ on a metric space V is

$$\delta(f_1, f_2) = \inf_{\varphi : V \cong V} \int_V |f_1(\varphi(x)) - f_2(x)| \, dx, \tag{2}$$

where the infimum is taken over all isometries $\varphi : V \cong V$. For suitable choices of metric space V, the isometries on V may be well-understood enough so that the above infimum is easy to approximate from knowledge of the geometry of V.

While latent spaces in continuous latent space models are typically taken to be Euclidean [12], recent research suggests that other metrics are more suitable for generating the particular sort of networks observed in the real-world [19]. The choice of metric space should be dictated by the abstract geometry of the abstract networks. Social networks and other real-world networks tend to satisfy *power laws*, *small world properties*, and are highly *clusterable*. For example, a network satisfies a *power law* if the probability $P(k)$ that a randomly uniformly selected node has degree k is of the form $P(k) = k^{-\gamma}$ for some real number $\gamma > 0$; real-world networks tend to exhibit approximate power laws. A network satisfies a *small world property* if the average minimum path-length is small relative to the total number of nodes in the network. Recent work has shown that certain natural *quasi-uniform densities* on the hyperboloid, a 2-dimensional surface visualizable in $\mathbb{R}^3$ as a bowl [Figure 3], generate networks with the above properties [19].

2.2.2 Goal

Often, the qualitative structures of networks we wish to compare has something to do with their geometry as metric spaces of nodes under minimum path lengths and nothing to do with their node labellings and size. Networks satisfying power laws, small world properties, and clusterability are *hyperbolic* as abstract metric spaces under the minimum path-length metric [19]. Thus continuous latent space models generating such networks ought to be hyperbolic as metric spaces themselves.

2.3 Hyperbolic spaces

Hyperbolic spaces are metric spaces which are negatively curved — the angles in a triangle of geodesics has sum less than 180 degrees. Examples of hyperbolic spaces are trees, the hyperboloid, and higher dimensional

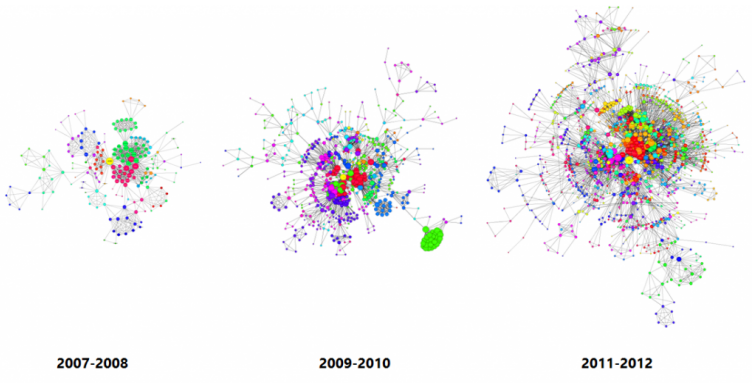

Figure 2: **Growing networks** Illustrated above is Apple's network of inventors (nodes) and joint patents (edges) linking inventors growing over time [1]. Often, the goal of network inference is to determine that growth process, irrespective of a particular snapshot at some point in time.

versions of hyperboloids. We want to extend statistical methods, such as kernel density estimation, from Euclidean space to hyperbolic spaces like the hyperboloid. Just as quasi-uniform densities on the hyperboloid generate certain networks with real-world structure, more general densities on the hyperboloid generate a large class of random networks. Thus the L_2-distance between densities on the hyperboloid, equipped with its standard hyperbolic volume measure, quantifies the degree to which the networks generated differ in their power laws, small world properties, and clusterabilities. The goal will be to analyze and compare networks by inferring and comparing latent node densities on the hyperboloid.

2.3.1 Background

Hyperbolic spaces provide a logical, generative model for real-world networks [19]. An example is the set of nodes of a tree-like network under the minimum path length metric [16]. The space $\mathbb{H}_2$ is the hyperboloid $x_1^2 + x_2^2 - x_3^2 = 1$ equipped with the metric inherited from $\mathbb{R}^3$. This space $\mathbb{H}_2$ is equivalent to the *Poincaré half-plane*

$$\mathbb{H}_2 = \{x + iy \mid x \in \mathbb{R},\ y \in (0, \infty)\}$$

equipped with the metric $ds = (dx^2 + dy^2)/y^2$. The space $\mathbb{H}_2$ is also equivalent to the *Poincaré disk*

$$\mathbb{H}_2 = \{x, y \in \mathbb{R} \mid |x + iy| < 1\}$$

equipped with the metric $ds^2 = 4(dx^2 + dy^2)/(1 - x^2 - y^2)$.

Real-world networks tend to have isometric embeddings into $\mathbb{H}_2$ [7] [16]. Moreover, densities

$$f_\delta(r, \theta) = \frac{\delta \sinh \delta r}{2\pi \cosh(\delta c - 1)}, \quad \delta > 0 \tag{3}$$

on the Poincaré disk generate real-world networks, such as the Internet [19]. The inference of a δ parameter from network samples serves as an effective, if limited, method of network inference and comparison. The graphs generated exhibit typical properties of complex networks, such as small world connections, power law distributions, significant clustering coefficients [19], and exchangeability (by independence of sampling).

2.3.2 Goal

Our goal is to generalize the network inference of densities on $\mathbb{H}_2$ from the limited parametrized setting (3) of [19] to a non-parametric setting and even the higher dimensional setting of *uniform hyperbolic spaces* $\mathbb{H}_{nc}$

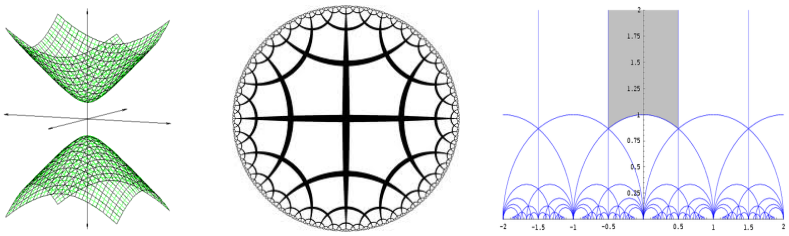

Figure 3: **Different models of** $\mathbb{H}_2$ A connected component of the hyperboloid $x_3^2 = 1 + x_1^2 + x_2^2$ on the left can be represented as the *Poincaré disk* in the middle or the *Poincaré half-plane* on the right. Under suitable choices of metrics, the middle and right spaces are isometric to a connected component of the left space equipped with the minimum path length.

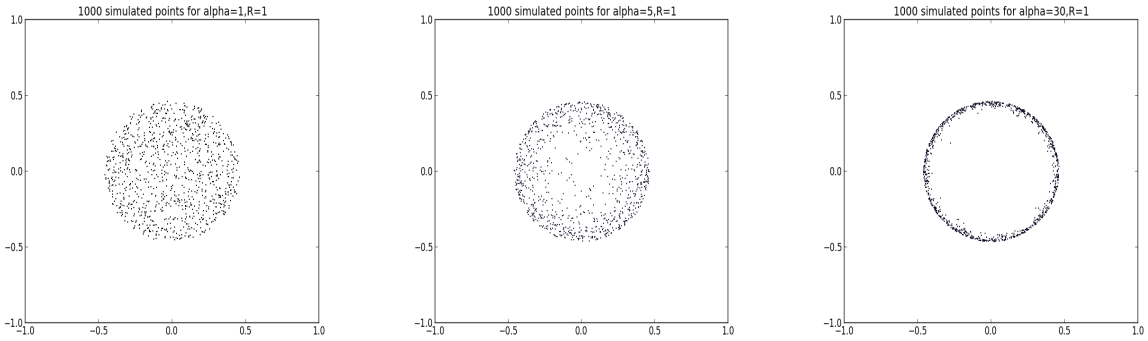

Figure 4: **Simulations of quasi-uniform densities** While quasi-uniform densities (3) are naturally defined in terms of geopolar coordinates, the above plots illustrate sample points in the Poincaré disk for a fixed R value (a compact region of $\mathbb{H}_2$ over which the densities are supported) and varying $\delta = 1, 5, 30$ from left to right.

of dimension n and curvature $-c$. Such generalized inference of densities on hyperbolic space from network samples will require extensions of classical statistics, such as non-parametric density estimation, from the Euclidean setting to the hyperbolic setting.

3 Methodology

Inferring the generative model from observed networks leads to many applications, like a principled method of network comparison. Network inference naturally breaks up into two parts: we first embed observed networks into $\mathbb{H}_2$, and second we estimate densities on $\mathbb{H}_2$ from embedded node coordinates. The former part is either a spectral decomposition problem or an optimization problem of maximizing a likelihood function. The latter part requires a density estimator optimized for the integral of L_2-loss with respect to the natural volume measure $\mathbb{H}_2$. Having obtained an estimator for densities from network samples, we can compare inferred node densities and test if they are statistically significantly different.

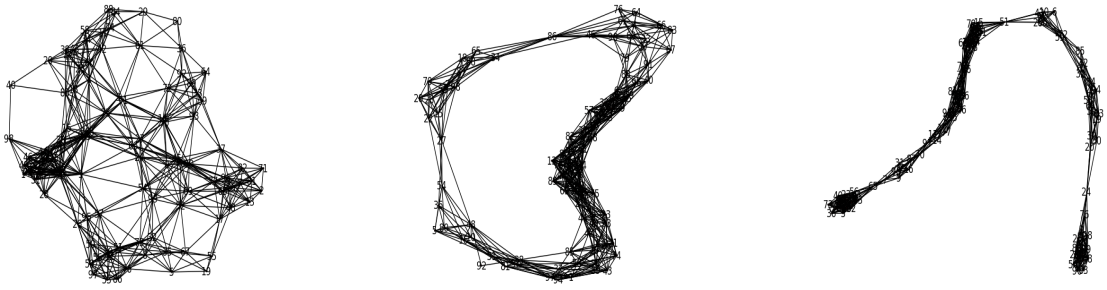

Figure 5: **Networks generated by quasi-uniform densities** The above graphs were generated by sampling 100 points from quasi-uniform densities with fixed R value and varying $\delta = 1, 5, 30$ from left to right. The connection probability function is the Heaviside stepside function, $W(z_1, z_2) = \Theta(R - d_h(z_1, z_2))$ where d_h represents the hyperbolic distance function. Note how the generated networks increase in clusterability as δ, determining the hyperbolic distance away from the center of the Poincaré disk at which the density peaks, increases.

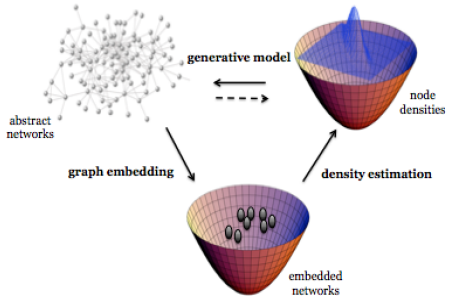

Figure 6: Network inference

4 Theoretical Agenda

Ultimately, we wish to directly describe an estimator for the composite problem [Figure 6] and study its convergence properties. In order to do so, we need to define graph embeddings as certain optimization problems and describe algorithms for their approximate solutions, theoretically understand the non-uniqueness of the solutions, and construct minimax density estimators on $\mathbb{H}_2$.

4.1 Graph embedding

Graph embeddings are embeddings of graphs into metric spaces that preserve the metric, or at least the closest possible approximations to such metric-preserving functions. An embedding of a social network would assign coordinates to nodes in such a way so that the metric differences between coordinates reflects the social distances between nodes. We would like to understand the theoretical properties of various algorithms and formulas for embedding a graph into the hyperboloid.

4.1.1 Background

A first step in estimating a density function f generating a given graph G is to embed the graph G into the metric space X. The ML estimator θ_G of the embedding $V_G \to X$ of graph vertices,

$$\hat{\theta}_G = \arg\max_{\theta: V_G \to X} \prod_{(v,w)\in E_G} W(\hat{\theta}(v), \hat{\theta}(w)) \prod_{(v,w)\notin E_G} (1 - W(\hat{\theta}(v), \hat{\theta}(w))), \tag{4}$$

where W is the connection probability function, is used in the literature [6] to embed real-world networks, such as the Internet for the case $X = \mathbb{H}_2$. Several techniques exist for approximating (4) for $X = \mathbb{H}_2$.

4.1.1.1 Metropolis-Hastings One such technique, thinking of $\mathbb{H}_2$ as the Poincaré disk, applies the *Metropolis-Hastings algorithm* [6]. We approximate the geopolar modulus r_v of a vertex v in G by

$$r_v = 2\log(\frac{8n}{k\pi}) - 2\log(\frac{k_v}{k_{\min}})$$

on moduli as specified in [6], where n denotes the number of vertices in G, k denotes the total sum of all degrees in G, $k_{\min}$ denotes the minimum degree in G, and k_v denotes the degree of vertex v. We then approximate the angle θ_v of a vertex v in G via the Metropolis-Hastings algorithm [6]: over several iterations, we uniformly randomly assign geopolar angles to the points, compare the likelihood p that G arose from such new coordinates with the corresponding likelihood q that G arose from the previous choice of coordinates, and replace the old choice of coordinates with the new choice of coordinates if $p > q$ or with probability greater than the likelihood ratio q/p.

4.1.1.2 Multidimensional Scaling Another technique is a hyperbolic adaptation of multidimensional scaling [4]. First, we treat the nodes of a given graph G as an abstract metric space whose metric is defined by minimum path lengths, D_{ij} for a pair of nodes i and j. Next, we find

$$x_1, ..., x_n \in \mathbb{R} \subset \mathbb{H}_2 \text{ such that } d_H(x_i, x_j) = D_{ij}$$

where d_H denotes the distance on a hyperboloid, $\arccos(x_i J {x_j}^T)$, and J represents the matrix associated with the Minkowski space metric. Finally, spectral decomposition is used to decompose $\cos(D)$ into AJA^T where the rows of A are the points, $x_1, ..., x_n$. In [4], the embeddings are proven to be an exact embedding for a network such that $\cos(D)$ has two negative eigenvalues and one positive eigenvalue.

4.1.2 Proposed work

One goal is to improve the accuracy and speed of one or more of the above embedding techniques. Another goal is to find a Mobius transformation $\varphi : \mathbb{H}_2 \to \mathbb{H}_2$ that minimizes the Hausdorff distance $d_{Haus}(\varphi(A), B)$ between a given pair $A, B \subset \mathbb{H}_2$, perhaps by way of adapting the Procrustes algorithm. Such a technique should prove useful in estimating (2) for the case $V = \mathbb{H}_2$ directly from network samples. Yet another goal is to generalize the setting from $\mathbb{H}_2$ with different choices of negative curvature and use model selection to select the ideal larger space before solving (4). The curvature selected could correspond to combinatorial notions of hyperbolicity (e.g. *δ-hyperbolicity* [16]).

4.2 Density estimation

Given a set of sample points on a space, *density estimation* is the problem of estimating the true density under which the sample points were drawn. Given an optimal choice (4) of hyperboloid coordinates for nodes in a social network, a density estimator on the hyperboloid will then summarize all relationships between nodes in a manner that is independent of the size of the network. Thus we seek a density estimator on the hyperboloid with a fast convergence rate. While the hyperboloid admits a reparametrization as a 2-dimensional surface, ordinary kernel density estimators in $\mathbb{R}^2$ are not optimal with respect to L_2-loss, defined

with respect to the natural volume measure on the hyperboloid. We want a non-parametric density estimator optimal with respect to such loss, because such loss intuitively corresponds to the difference between the networks generated by the estimated density and networks generated by the true density.

4.2.1 Background

The construction and performance of ordinary density estimators in the literature are dependent on specific properties of Euclidean space and its *volume measure*, the Lebesgue measure, with respect to which ordinary risk is defined. Existing generalizations of density estimators from the Euclidean setting require that the given manifold — unlike $\mathbb{H}_2$ — be compact [23], or require that the kernel be chosen for each point of the space and the true densities satisfy a Hölder class condition [17].

4.2.1.1 From $\mathbb{R}^n$ to Riemannian manifolds Data often does not come from Euclidean space, but more general *(Riemannian) manifolds*, which are smooth subspaces of Euclidean space. Besides hierarchical network structure described as 2-dimensional hyperboloids, other examples include directional headings and orientations in air traffic described as the space $\mathbb{SO}_3$, and stress in materials and distortions in spacetime described as the spaces of positive symmetric definite matrices. We can consider the general problem of estimating a density function on a manifold M from some observed sample points $X_1, X_2, \cdots, X_N$. The literature [23] suggests a straightforward generalization for compact manifolds and kernels whose supports are small enough to fit inside normal neighborhoods, patches of the manifold that are Euclidean enough. The normed vector difference $|x - X_i|$ is replaced by the Riemannian distance [23, Equation 9]. The risk of such an estimator, defined in terms of the volume measure of the manifold, shares an upper bound with the risk of a classical kernel density estimator under certain assumptions [23, Theorem 5]. However, the hyperboloid and higher dimensional analogues are not compact manifolds and therefore require different density estimators. Another example is a density estimator for general manifolds, where the domains of kernels are tangent bundles (which means choosing a kernel for each point of the space) and the true densities satisfy a Hölder class condition [17]. A minimax convergence rate in terms of the Hölder class exponent and differentiability of the densities is proven. However, in practice defining a kernel at each point of the manifold poses challenges in the implementation.

4.2.1.2 Convolution The convolution of densities on $\mathbb{R}^n$, necessary both to deconvolve noise and to smooth out empirical observations to obtain density estimators, is defined as follows. For a pair f, g of densities on $\mathbb{R}^n$, define the density $f * g$ on $\mathbb{R}^n$ by the rule

$$(f * g)(t) = \int_{\mathbb{R}^n} f(t - x)g(x)\, d\mu_{\mathbf{X}},$$

where $d\mu_{\mathbf{X}}$ denotes the Lebesgue measure. Convolution, which involves the operation of subtraction, generalizes to spaces $\mathbf{X}$ equipped with actions $\mathbf{G} \times \mathbf{X} \to \mathbf{X}$ of *Lie groups* $\mathbf{G}$. A *Lie group* is a manifold consisting of invertible matrices closed under smooth matrix multiplication and smooth matrix inversion operations. The symmetries of a manifold $\mathbf{X}$ can be described by an *action*

$$\mathbf{G} \times \mathbf{X} \to \mathbf{X}$$

of a Lie group, a function whose restriction to a function $\mathbf{X} \to \mathbf{X}$ for each $g \in \mathbf{G}$ is an isometry (distance-preserving map).

For a density f on $\mathbf{X}$ and a density g on $\mathbf{G}$, we define the density $f * g$ on $\mathbf{X}$ by the rule

$$(f * g)(t) = \int_{\mathbf{G}} f(tx^{-1})g(x)\, d\mu_{\mathbf{G}},$$

where $d\mu_{\mathbf{G}}$ denotes the *Haar measure*, a measure uniquely defined up to a multiplicative constant, on the Lie group $\mathbf{G}$.

Convolutions allow us to define ordinary kernel density estimators on $\mathbb{R}^n$ and more general $\mathbf{G}$-*kernel density estimators* on certain spaces $\mathbf{X}$ with symmetries described by a group $\mathbf{G}$. Just as Fourier analysis is useful for indirectly constructing and analyzing convolutions of densities on $\mathbb{R}^n$, a more general *Helgason-Fourier Analysis* will allow us to indirectly construct and analyze convolution of densities on more general manifolds. We define all relevant spaces, define the transform, and give an explicit formula for the inverse transform.

4.2.1.3 Standard notation We will use the following notation in differential geometry, Fourier Analysis, and kernel density estimation. Fix a semisimple Lie group $\mathbf{G}$ and maximal compact connected Lie subgroup $\mathbf{K}$. Let $\mathbf{X}$ be the symmetric space

$$\mathbf{X} = \mathbf{G}/\mathbf{K}.$$

Let $\mathbf{A}, \mathbf{N}$ be the respectively abelian and nilpotent subgroups of $\mathbf{G}$ such that $\mathbf{G} = \mathbf{KAN}$, the Iwasawa decomposition of $\mathbf{G}$. We define $a_x \in \mathbf{A}$ and $n_x \in \mathbf{N}$ so that x gets sent to the product $a_x n_x$ for each $x \in \mathbf{X}$ under the natural isomorphism $\mathbf{G}/\mathbf{K} \cong \mathbf{AN}$. Let $\mathfrak{g}, \mathfrak{k}, \mathfrak{a}, \mathfrak{n}$ represent the Lie algebras of $\mathbf{G}, \mathbf{K}, \mathbf{A}, \mathbf{N}$, respectively. Let $\langle -, - \rangle$ denote the standard inner product on $\mathfrak{g}$ defined by

$$\langle v, w \rangle = trace([v, -] \circ [w, -])$$

and $|-|$ be the norm on $\mathfrak{g}$ defined by $\langle v, v \rangle = |v|^2$. Let $\langle -, - \rangle$ and $|-|$ also denote the induced inner product and norm on $\mathfrak{g}^*$. Let $\mathfrak{a}^*$ denote the dual vector space of $\mathfrak{a}$. Let $\mathbf{M}$ be the centralizer of $\mathbf{A}$ in $\mathbf{K}$. We write $\mu_{\mathbf{H}}$ to denote a certain measure on $\mathbf{H}$: in the case $\mathbf{H}$ is a Lie group, $\mu_{\mathbf{H}}$ represents the Haar measure; in the case $\mathbf{H}$ is a Riemannian manifold, $\mu_{\mathbf{H}}$ represents the volume measure. We also write c for the *Harish-Chandra function* on $\mathfrak{a}^*$; we refer the reader to [28] for suitable definitions.

4.2.1.4 The transform For $f \in \mathbb{C}_c^\infty(\mathbf{X})$, the *Helgason-Fourier transform*, written $\mathcal{H}$, is a linear map

$$L_2(\mathbf{X}, d\mu_{\mathbf{X}}) \to L_2(\mathfrak{a}^* \times \mathbf{K}/\mathbf{M}, \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}})$$

sending a function f to the function $\mathcal{H}f$ defined by the rule

$$(\mathcal{H}f)(s, k\mathbf{M}) = \int_{\mathbf{G}/\mathbf{K}} f(x) e^{\overline{s(log\ a_{k(x)})}}\, d\mu_{\mathbf{X}}, \quad (s, k\mathbf{M}) \in (\mathfrak{a}^* \otimes \mathbb{C}) \times \mathbf{K}/\mathbf{M},$$

where we take $s = i\lambda + \rho$ for $\lambda \in \mathfrak{a}^*$ and ρ is half of the sum of restricted roots of $\mathbf{G}$.
Like in the classical Fourier case, we have the Plancherel identity

$$\int_{\mathbf{G}/\mathbf{K}} |f(x)|^2 d\mu_{\mathbf{X}} = \int_{\lambda \in \mathfrak{a}^*} \int_{k\mathbf{M} \in \mathbf{K}/\mathbf{M}} |\mathcal{H}f\,(i\lambda + \rho, k\mathbf{M})|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}}, \tag{5}$$

We define the convolution of a density f_1 on $\mathbf{X}$ with a density f_2 on $\mathbf{G}$ by the rule

$$(f_1 * f_2)(x) = \int_{\mathbf{G}} f_1(g) f_2(g^{-1}x)\, d\mu_{\mathbf{G}}.$$

The Helgason-Fourier transform $\mathcal{H}$ sends convolutions to products in the following sense. Call a density f on $\mathbf{G}$ $\mathbf{K}$-*invariant* if $f(k_1 x k_2) = f(x)$ for all $k_1, k_2 \in \mathbf{K}$ and $x \in \mathbf{X}$. A $\mathbf{K}$-invariant density f on $\mathbf{G}$ induces a well-defined density on $\mathbf{G}/\mathbf{K}$ sending each element $g\mathbf{K}$ to $f(g)$. Hence we can define $\mathcal{H}[f]$ of such f as the Helgason-Fourier transforms of the induced densities on $\mathbf{X}$. For each $f_1 \in L_2(\mathbf{X}, d\mu_{\mathbf{X}})$ and $f_2 \in L_2(\mathbf{G}, d\mu_{\mathbf{G}})$ such that f_2 is $\mathbf{K}$-invariant,

$$\mathcal{H}[f_1 * f_2] = \mathcal{H}[f_1]\mathcal{H}[f_2].$$

4.2.1.5 The inverse transform The *inverse Helgason-Fourier transform*, written $\mathcal{H}^{-1}$, is given by

$$f(x) = \int_{\lambda \in \mathfrak{a}^*} \int_{k\mathbf{M} \in \mathbf{K}/\mathbf{M}} \mathcal{H}f(i\lambda + \rho, k\mathbf{M}) e^{(i\lambda+\rho)(log\ a_{k(x)})} \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}}$$

4.2.2 Completed work

We have already introduced a new density estimator on a large class of symmetric spaces, and have proven a minimax rate of convergence identical to the minimax rate of convergence for a n-dimensional Euclidean kernel density estimator [2]. We then specialize our generalized kernel density estimator to $\mathbb{H}_2$ for our hyperbolic network inference methodology.

4.2.2.1 The G-kernel density estimator In Euclidean space, kernel density estimation smooths out the empirical distribution by adding a little bit of noise (distributed according to the kernel) around each observation. Formally, the ordinary kernel density estimator $f^{n,h}$ satisfies

$$f^{n,h} = \mathcal{F}^{-1}\left[\hat{\phi}\mathcal{F}[K_h]\right],$$

where $\mathcal{F}$ denotes the ordinary Fourier transform, $\hat{\phi}$ denotes the empirical characteristic function of the samples, K denotes a kernel, h denotes a bandwidth parameter, and $\mathcal{F}[K_h](s) = \mathcal{F}[K](hs)$. In order to generalize kernel density estimators from $\mathbb{R}^n$ to $\mathbf{X}$, we treat certain densities on $\mathbf{G}$ as kernels and generalize bandwidth in terms of the transform $\mathcal{H}[K]$. For the observed samples $X_1, \ldots, X_n \in \mathbf{X}$ and density K on $\mathbf{G}$ invariant under left and right multiplication by $\mathbf{K}$, we define the $\mathbf{G}$-*Kernel Density Estimator* $f^{(n,T,h)}$ by

$$f^{(n,T,h)} = \mathcal{H}^{-1}\left[\hat{\phi}\mathcal{H}K_h I_{(-T,+T)}\right]$$

where we abuse notation and treat $\mathcal{H}K_h$ as the function sending $(i\lambda+\rho, k\mathbf{M})$ to $\mathcal{H}K(h(i\lambda+\rho), \mathbf{M})$, $I_{(-T,+T)}$ as the function sending $(i\lambda + \rho, k\mathbf{M})$ to 1 if $|\lambda| \leqslant T$ and 0 otherwise, and

$$\hat{\phi}(s, k\mathbf{M}) = \frac{1}{n}\sum_{i=1}^{n} e^{\overline{s(log\ a_{k(X_i)})}} \tag{6}$$

4.2.2.2 Assumptions Let $\mathbf{X}$ denote a symmetric space such that for fixed semisimple Lie group $\mathbf{G}$ $\mathbf{X} = \mathbf{G}/\mathbf{K}$ for a maximal complete subgroup $\mathbf{K}$. Let f_X denote a density on $\mathbf{X}$ with respect to the standard volume measure $d\mu_{\mathbf{X}}$. Let K denote a density on $\mathbf{G}$ with respect to the Haar measure $d\mu_{\mathbf{G}}$.

First of all, we assume our densities are L_2.

(D.1) Assume $f_X \in L^2(\mathbf{X}, d\mu_{\mathbf{X}})$ and $K \in L^2(\mathbf{G}, d\mu_{\mathbf{G}})$.

Second of all, we need to restrict K to guarantee that its Helgason-Fourier transform is well-defined.

(D.2) the kernel K is $\mathbf{K}$-bi-invariant:

$$K(acb) = K(c), \quad c \in \mathbf{G}\, a, b \in \mathbf{K}.$$

Third of all, we make assumptions on the smoothness of the true density f_X. The operator Δ^k defined below in terms of the Helgason-Fourier transform, generalizes the kth derivative operator from integers k to non-negative real numbers k.

(D.3) there exist $\alpha > 1$ and $\mathcal{Q} > 0$ such that

$$f_X \in \mathcal{F}_\alpha(\mathcal{Q}) = \{f_X \in L^2(\mathbf{X}, d\mu_{\mathbf{X}}) : \|\Delta^{\alpha/2} f_X\|^2 \leqslant \mathcal{Q}\},$$

where $\Delta^{\alpha/2} f_X$ denotes the unique function $h \in L^2(\mathbf{X}, d\mu_{\mathbf{X}})$ such that

$$\mathcal{H}h(s, k\mathbf{M}) = \overline{(s(s-2\rho))}^{\alpha/2} \mathcal{H}f_X(s, k\mathbf{M})$$

Last of all, we make assumptions on the smoothness of the kernel K.

(D.4) there exist constants $\beta, \gamma, C_1, C_2 > 0$ such that

$$C_1 e^{-\frac{|s|^\beta}{\gamma}} \leqslant |\mathcal{H}K(s, \mathbf{M})| \leqslant C_2 e^{-\frac{|s|^\beta}{\gamma}}$$

for all $s \in \mathfrak{a}^* \otimes \mathbb{C}$.

(D.5) For some $\alpha > 1$, there exist a constant A such that

$$\underset{s \in \mathfrak{a}^* \otimes \mathbb{C}}{ess\ sup} \frac{|\mathcal{H}K(s, \mathbf{M}) - 1|}{|s|^\alpha} \leqslant A$$

4.2.2.3 Main theorems Proofs of the following main results can be found in the appendix, with more details in the paper [2].

Theorem 4.1. *Assume (D.1)-(D.5). Then for a density f_X on a symmetric space $\mathbf{X}$,*

$$\mathbb{E}\|f_X^{(n,T,h)} - f_X\|^2 \leqslant QA^2h^{2\alpha} + QT^{-2\alpha} + C\frac{T^{\dim \mathbf{X}}}{n} e^{[-(2/\gamma)(h|\rho|)^\beta]}.$$

for some constant $C > 0$ not dependent on $T, \alpha, \mathcal{Q}, n$.

By choosing a smooth enough kernel density K, an optimal cutoff of T and optimal bandwidth h, we obtain the following rate of convergence.

Corollary 4.2. *Assume (D.1)-(D.5). Then*

$$\mathbb{E}\|f_X^{(n,T,h)} - f_X\|^2 \leqslant Cn^{-2\alpha/(2\alpha+\dim \mathbf{X})}$$

for some constant $C > 0$ not dependent on $T, \alpha, \mathcal{Q}, n$ and

$$T = \left[\frac{2\alpha Qn}{\dim \mathbf{X} C} e^{[-(2/\gamma)(h|\rho|)^\beta]}\right]^{1/(2\alpha+\dim \mathbf{X})} \qquad h \in \mathcal{O}(n^{-1/(2\alpha+\dim \mathbf{X})}).$$

The convergence rate for the upper bound is matched by the lower bound, as shown below.

Theorem 4.3. *Assume (D.3). There exists a constant $C > 0$ such that*

$$\inf_{g^{(n)}} \sup_{f_X \in \mathcal{F}_\alpha(\mathcal{Q})} \mathbb{E}\|g^{(n)} - f_X\|^2 \geqslant Cn^{-2\alpha/(2\alpha+\dim \mathbf{X})}$$

where the infimum is taken over all estimators $g^{(n)}$.

By the previous results, we obtain our minimax rate below for our adapted kernel density estimator.

Corollary 4.4. *If f_X and K satisfy (D.1)-(D.5), the minimax rate for $f_X^{(n,T,h)}$ is $n^{-2\alpha/(2\alpha+\dim \mathbf{X})}$.*

4.2.2.4 Specialized case Let $\mathbb{H}_2$ denote the hyperboloid, regarded as the *Poincaré half-plane*

$$\mathbb{H}_2 = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\},$$

where $\mathrm{Im}(z)$ denotes the imaginary part of a complex number z, equipped with metric

$$ds^2 = y^{-2}(dx^2 + dy^2).$$

The space $\mathbb{H}_2$ is isometric as a Riemannian manifold to the quotient space

$$\mathbb{H}_2 = \mathbb{SL}_2/\mathbb{SO}_2.$$

Under this identification, the matrices in $\mathbb{SL}_2$ act on $\mathbb{H}_2$ by *Möbius transformations*:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{az+b}{cz+d}.$$

Therefore our density estimator is defined on $\mathbb{H}_2$ because $\mathbb{SL}_2$ is a semisimple Lie group admitting an Iwasawa decomposition as $\mathbb{SL}_2 = \mathbb{SO}_2\mathbf{AN}$, where $\mathbf{A}$ is the group of diagonal (2×2)-matrices in $\mathbb{SL}_2$ with non-negative entries and $\mathbf{N}$ is the group of upper triangular (2×2)-matrices with 1's along the diagonal. For each $z \in \mathbb{H}_2$, n_z and a_z are characterized by

$$(n_z)_{1,2} = \mathrm{Re}(z), \quad (a_z)_{1,1} = \mathrm{Im}(z)^{1/2}.$$

The Harish-Chandra c-function satisfies the formula

$$c(\lambda)^{-2} = \frac{1}{8\pi^2}\lambda \tanh(\pi\lambda).$$

There exists a unique restricted root of $\mathbb{H}_2 = \mathbb{SL}_2/\mathbb{SO}_2$. Under the natural identification of $\mathbf{A}$ with the Lie group of multiplicative non-negative real numbers and hence an identification of $\mathfrak{a}^*$ with $\mathbb{R}$, we identify the unique restricted root (taking a (2×2)-matrix to the difference in its diagonal elements) with 1 and hence ρ with $1/2$.

4.2.2.5 Choice of kernel We can also choose our kernel K to be the *hypergaussian*, an analogue of a gaussian density on Euclidean space defined as follows in [13]. Just as ordinary gaussians are characterized as solutions to the heat equation, we define K to be the unique ($\mathbb{SO}_2$-invariant solution) to the heat equation on $\mathbb{H}_2$, lifted to a function on $\mathbb{SL}_2$. Concretely,

$$\mathbb{H}[K](s, kM) \propto e^{\overline{s(s-1)}}$$

and hence K satisfies the assumptions (D.1), (D.2), and (D.4) for $\beta = 2$ and $\gamma = 1$.

4.2.2.6 Simplified formula Under these identifications and simplifications, our $\mathbb{SL}_2$-*kernel density estimator* takes the form:

$$f^{(n,T,h)}(z) \propto \frac{1}{n}\sum_{i=1}^{n}\int_{-T}^{+T}\int_{0}^{2\pi} \mathrm{Im}(k_\theta(Z_i))^{\frac{1}{2}-i\lambda} e^{-(\frac{h^2}{4}+h^2\lambda^2)} (\mathrm{Im}(k_\theta(z)))^{\frac{1}{2}+i\lambda} \frac{\lambda\tanh(\pi\lambda)}{8\pi^2}\, d\theta\, d\lambda, \tag{7}$$

where k_θ denotes the rotation matrix associated to the angle θ.

4.2.3 Proposed work

Future work would be to explore other applications of the generalized kernel density estimator. Data often resides in non-Euclidean symmetric spaces. For example, directional headings and orientations of flying objects (e.g. missiles, asteroids) naturally reside on the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ and $\mathbb{SO}_3 \subset \mathbb{R}^9$ [25]. For another example, the data of material stress in material sciences and spacetime distortions in cosmology amount to points in the space of symmetric positive definite matrices [25]. In addition, future work would also include developing an analogue of the Fast Fourier Transform for Helgason-Fourier Analysis in order to speed up the actual computation of the kernel density estimator.

4.3 Network comparison

A suitable comparison of two sets $\mathcal{N}_1, \mathcal{N}_2$ of networks should involve comparing the underlying generative models g_1, g_2 creating the two sets — computing some numerical distance $\rho(g_1, g_2)$ between g_1 and g_2 in a space of possible generative models and assessing the significance of such difference statistics. In our case of interest, $\mathcal{N}_1, \mathcal{N}_2$ might be a pair of snapshots of different public health, financial, or collaboration networks of varying size and g_1, g_2 would describe the overall network geometries as densities on $\mathbb{H}_2$.

4.3.1 Background

A general framework for comparing networks from a generative model has been outlined in [26]. Consider two sets $\mathcal{N}_1, \mathcal{N}_2$ of networks. There exist three inferred generative models g_1, g_2, g_{12} for the three sets $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_1 \cup \mathcal{N}_2$. In order to account for the difference $\rho(g_1, g_2)$ due to statistical noise, we use g_{12} to again generate a pair $\mathcal{N}_1', \mathcal{N}_2'$ of network samples and again infer three generative models g_1', g_2', g_{12}', repeating the process several times to estimate a distribution for differences $\rho(g_1', g_2')$ due to sampling variation alone. Thus we can assess the significance of $\rho(g_1, g_2)$. Moreover, the metric ρ on generative models often takes the form of a difference of distributions (e.g. total variation) and hence assesses the global differences between networks.

4.3.2 Proposed work

As noted in [26], the more constraints we impose on our particular generative models, the greater the power we expect from our hypothesis testing. Our goal is to specialize the above network comparison program for the generative model of densities on $\mathbb{H}_2$ satisfying D.1 and D.3 and difference statistic ρ the L_2-distance, and prove specific properties of hypothesis testing in that setting. We aim to establish the validity of bootstrapping — the determination of a distribution of difference statistics under the null hypothesis by the iterative procedure outlined previously. Under certain growth assumptions for sequences of graphs

$$G = (G_1, G_2, G_3, \ldots), \quad H = (H_1, H_2, H_3, \ldots),$$

we will show that our tests of significance for G and H are consistent. We will obtain formulas and bounds for the power of test statistics as a function of curvature, network size, and network sample size. We also will characterize bounds on the error in our inferred densities from errors in our graph embeddings and errors in density estimation. We will also explore the possibility of using model selection or adaptivity to tune the curvature of the ambient latent hyperboloid to the network data given. All of these goals are best tackled if we can directly and theoretically describe the distribution of $\|g_1 - g_2\|_2$ under the null hypothesis — a plausible goal, given the simplicity of our generative model and the compatibility of our difference statistic $\|g_1 - g_2\|_2$ with such tools as the Helgason-Fourier transform $\mathcal{H}$.

5 Computational Agenda

In addition to the above theoretical investigation, we also plan to implement and possibly refine graph embedding and density estimation to infer the generating densities of abstract sample networks and thus evaluate the feasibility of our proposed methods of network comparison.

5.1 Completed work

5.1.1 Graph embeddings

5.1.1.1 Metropolis-Hastings We have implemented the Metropolis-Hastings approach to embedding graphs in $\mathbb{H}_2$ [6] and illustrate in [Figure 7] the behavior of the algorithm on sample points which came from quasi-uniform densities like in [Figure 4].

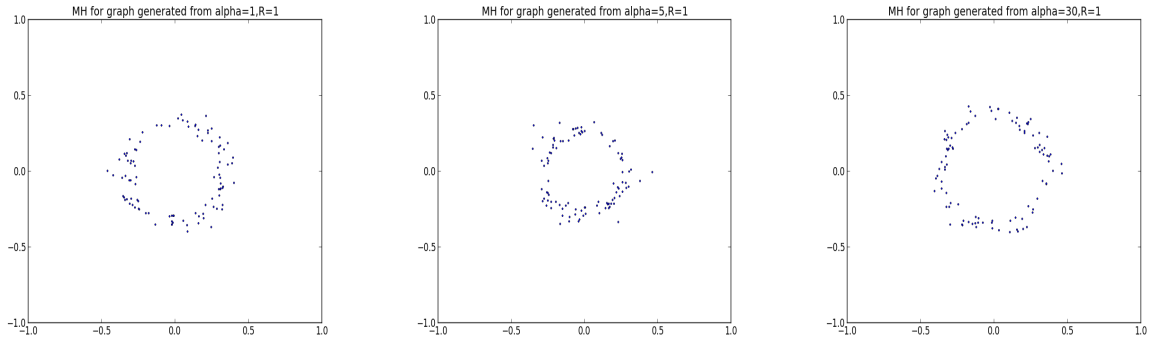

Figure 7: **Metropolis-Hastings** The nodes of abstract graphs [Figure 5] from quasi-uniform densities [Figure 4] are embedded, via the Metropolis-Hastings algorithm, into $\mathbb{H}_2$ to yield the sample points illustrated above. Those sample points in turn can be used to infer a node density describing the global network structure irregardless of node size or node labels.

5.1.1.2 Hyperbolic Multidimensional Scaling

We have also implemented the hyperbolic variant of Multidimensional Scaling to embedding graphs in $\mathbb{H}_2$ [4] and illustrate in [Figure 8] the behavior of the technique on sample points which came from quasi-uniform densities like in [Figure 4].

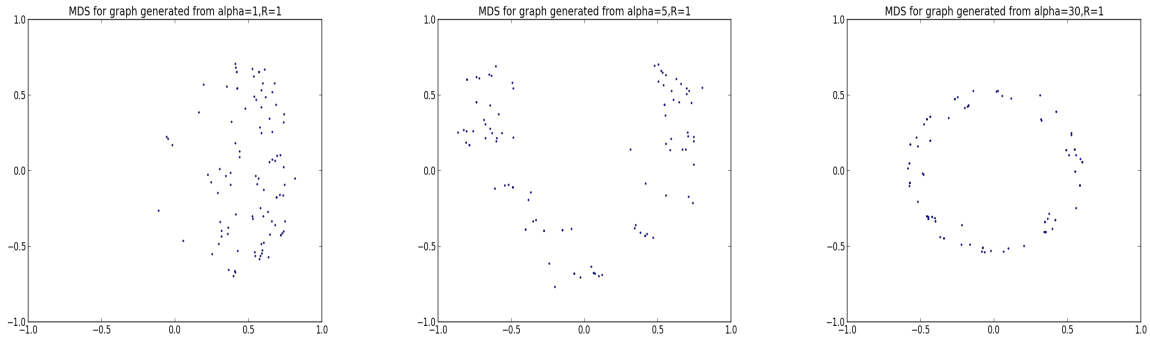

Figure 8: **Multidimensional Scaling** The nodes of abstract graphs [Figure 5] from quasi-uniform densities [Figure 4] are embedded, via a version of Multidimensional Scaling into $\mathbb{H}_2$ to yield the sample points illustrated above. Those sample points in turn can be used to infer a node density describing the global network structure irregardless of node size or node labels.

5.1.2 Density estimation

We have implemented the $\mathbb{SL}_2$-kernel density estimator and compare in [Figure 9] the estimator with the true density, or at least the associated densities of the geopolar magnitude, for different choices of quasi-uniform densities.

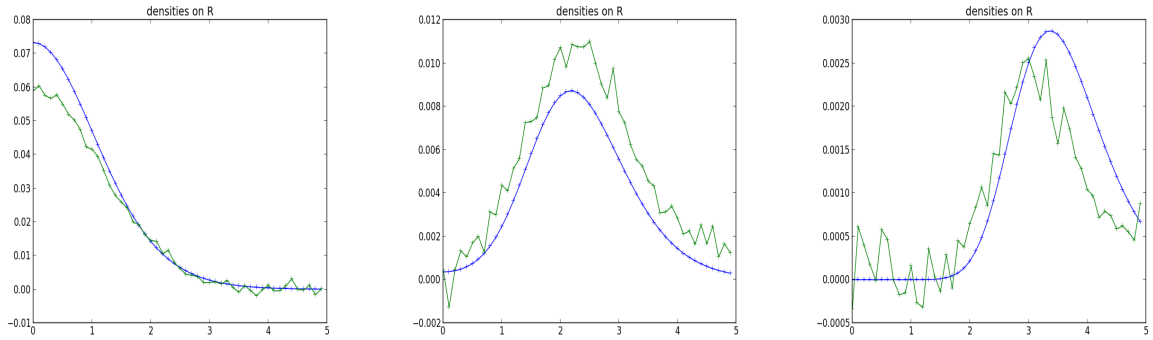

Figure 9: **Density estimators** The true density of the geopolar magnitude (blue) is plotted alongside the estimated density derived from our $\mathbb{SL}_2$-kernel density estimator (green), for quasi-uniform densities with $\delta = 1, 10, 30$ in order from left to right.

5.2 Proposed work

5.2.1 Speed-ups in implementation

Improvements in the algorithms or actual implementations of the above simulations, which rely on repeated iterations, should improve accuracy of the empirical results. Some ideas of such improvements for density estimation involve direct implementations of the estimator, without use of Helgason-Fourier transforms and the associated multiple integrals.

5.2.2 Network comparisons

We will then simulate network comparison tests and empirically estimate power and consistency.

6 Empirical Agenda

In this section, we sketch some methods for validating our models, using our tools to compare real-world examples of networks. In addition, we can use our generative model to do resampling for a couple real-world examples of networks. However, these are just application sketches and subject to change.

6.1 Model validation

One goal is to show that our model is well-suited for studying and comparing the particular sorts of networks we encounter in actual datasets. Specifically, we would like to understand better how hyperbolic distances constrain the types of networks that are generated. Based on combinatorial notions of hyperbolicity, we expect that conditional probabilities of triangles and quadrilaterals given the existence of one or more edges will start to characterize the class of generated networks. We can then start to justify our choice of model, when such assumptions on the networks are warranted by observations on actual datasets.

6.2 Social online networks: health policy

Online social networks can be used as distributed sensors, monitoring offline phenomena [3]. Variation in the network structure of human communications is an under-utilized resource for extracting phenomena of interest in public health. In our case, the networks of interest to us have as nodes a random sampling of U.S. Twitter users tweeting the word `flu` and have as edges reported social links (`friend`/`follower` relationships).

We will show that network hypothesis testing can improve the detection of influenza trends from Twitter, supplementing textual analysis of tweets with change detection in social graph structures that are perhaps more robust to irrelevant variation in language use and choice of language [Figure 1]. We will analyze 1% of all tweets over weeks 34-52 of 2011 and weeks 1-4 of 2012, all restricted to users from identifiable U.S. cities.

6.3 Collaboration networks: innovation policy

Understanding collaboration networks can be useful in crafting innovation policy, including the government regulation of patents and the dispensation of grants. Collaboration networks represent researchers as nodes and co-authorship/co-patents as edges between nodes. The goal would be to explore whether structural differences in innovation networks distinguish between productive and unproductive ecosystems of researchers — where productivity is measured in terms of number of patents, number of research articles, and other (imperfect) measures of creative output.

7 Appendix: Proofs

In the proofs that follow, we only use the following property of the Harish-Chandra c-function as observed in [21].

Lemma 7.1. *For each* $\lambda \in \mathfrak{a}^*$, $|c(\lambda)|^{-2} \leqslant (1+|\lambda|)^{\dim \mathfrak{n}}$.

7.1 Upper bound: Proof of Theorem 4.1

We use the mean integrated squared error to measure the performance of our generalized estimator. We break the mean integrated squared error into two parts, variance and squared bias, and bound each part separately. Proofs here adapt some techniques in [13] for deconvolution on the hyperboloid to convolution on a large class of symmetric spaces.

7.1.1 Variance

We bound the variance from above as follows.

$$\mathbb{E}\left\| f_X^{(n,T,h)} - \mathbb{E}[f_X^{(n,T,h)}] \right\|^2 = \mathbb{E}\int_{\mathbf{G/K}} \left| f_X^{(n,T,h)} - \mathbb{E}[f_X^{(n,T,h)}] \right|^2 d\mu_{\mathbf{X}}$$

The below equality follows by the Plancherel identity (5),

$$\mathbb{E}\int_{\mathbf{G/K}} \left| \mathcal{H}f_X^{(n,T,h)} - \mathbb{E}[\mathcal{H}f_X^{(n,T,h)}] \right|^2 d\mu_{\mathbf{X}} = \mathbb{E}\int_{\mathbf{G/K}} \left| \mathcal{H}K_h \hat{\phi} I_{(-T,+T)} - \mathcal{H}K_h \mathcal{H}f_X \right|^2 d\mu_{\mathbf{X}}.$$

The above expression, by the Fubini-Tonelli theorem, equals

$$\begin{aligned}
&\mathbb{E}\int_{|\lambda|<T}\int_{k\mathbf{M}\in\mathbf{K/M}} |\mathcal{H}K_h|^2 \left| \hat{\phi} - \mathcal{H}f_X \right|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K/M}} \\
&= \int_{|\lambda|<T}\int_{k\mathbf{M}\in\mathbf{K/M}} |\mathcal{H}K_h|^2 \, \mathbb{E}\left[|\hat{\phi}|^2 + |\mathcal{H}f_X|^2 - 2\langle \hat{\phi}, \mathcal{H}f_X \rangle \right] \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K/M}}.
\end{aligned}$$

Deducing from (6) that

$$\mathbb{E}\left| \hat{\phi}(\lambda + i\rho, k\mathbf{M}) \right|^2 = |\mathcal{H}f_X(\lambda + i\rho, k\mathbf{M})|^2 + \frac{\mathbb{E}[|a_k(X)|^{2\rho}] - |\mathcal{H}f_X(\lambda + i\rho, k\mathbf{M})|^2}{n},$$

the previous expression is bounded by

$$\begin{aligned}
&\int_{|\lambda|<T}\int_{k\mathbf{M}\in\mathbf{K/M}} |\mathcal{H}K_h|^2 \left[|\mathcal{H}f_X|^2 + \frac{\mathbb{E}[|a_k(X)|^{2\rho}] - |\mathcal{H}f_X|^2}{n} + |\mathcal{H}f_X|^2 - \mathbb{E}[2\langle \hat{\phi}, \mathcal{H}f_X \rangle] \right] \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K/M}} \\
&= \int_{|\lambda|<T}\int_{k\mathbf{M}\in\mathbf{K/M}} |\mathcal{H}K_h|^2 \left[2|\mathcal{H}f_X|^2 + \frac{\mathbb{E}[|a_k(X)|^{2\rho}] - |\mathcal{H}f_X|^2}{n} - 2|\mathcal{H}f_X|^2 \right] \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K/M}} \\
&= \int_{|\lambda|<T}\int_{k\mathbf{M}\in\mathbf{K/M}} |\mathcal{H}K_h|^2 \left[\frac{\mathbb{E}[|a_k(X)|^{2\rho}] - |\mathcal{H}f_X|^2}{n} \right] \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K/M}} \\
&\leqslant \int_{|\lambda|<T}\int_{k\mathbf{M}\in\mathbf{K/M}} |\mathcal{H}K_h|^2 \left[\frac{\mathbb{E}[|a_k(X)|^{2\rho}]}{n} \right] \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K/M}}.
\end{aligned}$$

In addition, $\mathbb{E}[|a_k(X)|^{2\rho}] = \mathcal{H}f_X(2s, k\mathbf{M})$ implies that for some constant $D_1 > 0$,

$$\int_{k\mathbf{M}\in\mathbf{K/M}} E[|a_k(X)|^{2\rho}] d\mu_{\mathbf{K/M}} = \int_{k\mathbf{M}\in\mathbf{K/M}} |Hf_X(2s, k\mathbf{M})| d\mu_{\mathbf{K/M}} \leqslant D_1$$

Thus the previous expression is in turn bounded by

$$\frac{D_1}{n} \sup_{|\lambda|<T} |\mathcal{H}K_h(i\lambda+\rho)|^2 \int_{|\lambda|<T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}}$$

By Lemma 2.1, $|c(\lambda)|^{-2} \leqslant (1+|\lambda|)^{\dim \mathfrak{n}}$. Therefore for some constant $D_2 > 0$,

$$\int_{|\lambda|<T} \frac{d\mu_\lambda}{|c(\lambda)^2|} \leqslant (2T)^{\dim \mathfrak{a}} \sup_{|\lambda|<T} (1+|\lambda|)^{\dim \mathfrak{n}} = (2T)^{\dim \mathfrak{a}} (1+T)^{\dim \mathfrak{n}} \leqslant D_2 T^{\dim \mathbf{X}}$$

Thus the previous expression is in turn bounded as follows for a constant C by (D.4).

$$\frac{D_1 D_2 T^{\dim \mathbf{X}}}{n} \sup_{|\lambda|<T} |\mathcal{H}K(h(i\lambda+\rho))|^2 \leqslant \frac{CT^{\dim \mathbf{X}}}{n} e^{[-(2/\gamma)(h|\rho|)^\beta]}$$

7.1.2 Squared bias

We now bound the squared bias from above. Note that

$$\left\| \mathbb{E} f_X^{(n,T,h)} - f_X \right\|^2 = \int_{\mathbf{G}/\mathbf{K}} \left| \mathbb{E} f_X^{(n,T,h)} - f_X \right|^2 d\mu_{\mathbf{X}}$$

The below equality follows by the Plancherel identity (5) and the Fubini-Tonelli theorem,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} \left| \mathbb{E}\mathcal{H} f_X^{(n,T,h)} - \mathcal{H} f_X \right|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}} \\
&= \int_{|\lambda|<T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} \left| \mathbb{E}\mathcal{H} f_X^{(n,T,h)} - \mathcal{H} f_X \right|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}} \\
&+ \int_{|\lambda|>T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} \left| \mathbb{E}\mathcal{H} f_X^{(n,T,h)} - \mathcal{H} f_X \right|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}}
\end{aligned}$$

the last equality following from $\mathcal{H} f_X^{(n,T,h)}$ compact on $|\lambda| < T$. Then the above expression equals

$$\begin{aligned}
&\int_{|\lambda|<T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} |\mathcal{H}f_X \mathcal{H}K_h - \mathcal{H}f_X|^2 \, d\frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}} + \int_{|\lambda|>T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} |\mathcal{H}f_X|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}} \\
&= \int_{|\lambda|<T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} |\mathcal{H}K(h(i\lambda+\rho)) - 1|^2 \, |\mathcal{H}f_X(\imath\lambda+\rho)|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}} \\
&+ \int_{|\lambda|>T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} |(i\lambda+\rho)(i\lambda-\rho)|^{-\alpha} \, |(i\lambda+\rho)(i\lambda-\rho)|^{\alpha} \, |\mathcal{H}f_X(i\lambda+\rho)|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}}
\end{aligned}$$

By assumption (D.5), the previous expression equals

$$\begin{aligned}
&\leqslant \int_{|\lambda|<T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} A^2 h^{2\alpha} \, |(i\lambda+\rho)(i\lambda-\rho)|^{\alpha} \, |\mathcal{H}f_X(i\lambda+\rho)|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}} \\
&+ \sup_{|\lambda|>T} |(i\lambda+\rho)(i\lambda-\rho)|^{-\alpha} \int_{|\lambda|>T} \int_{k\mathbf{M}\in\mathbf{K}/\mathbf{M}} |(i\lambda+\rho)(i\lambda-\rho)|^{\alpha} \, |\mathcal{H}f_X(i\lambda+\rho)|^2 \frac{d\mu_\lambda}{|c(\lambda)^2|} d\mu_{\mathbf{K}/\mathbf{M}}
\end{aligned}$$

By assumption (D.3), the previous expression is bounded by

$$QA^2 h^{2\alpha} + Q(T^2 + |\rho|^2)^{-\alpha} \leqslant QA^2 h^{2\alpha} + QT^{-2\alpha}.$$

7.2 Optimal upper bound: Proof of Corollary 4.2

By the previous part, we have obtain an upper bound of

$$\mathbb{E}\|f_X^{(n,T,h)} - f_X\|^2 \leqslant QA^2h^{2\alpha} + QT^{-2\alpha} + C\frac{T^{\dim \mathbf{X}}}{n}e^{[-(2/\gamma)(h|\rho|)^\beta]}.$$

The optimal cutoff of T that minimizes the upper bound is of the form

$$T(n) = \left[\frac{2\alpha Qn}{\dim \mathbf{X} C}e^{[-(2/\gamma)(h|\rho|)^\beta]}\right]^{1/(2\alpha+\dim \mathbf{X})}$$

yielding the following upper bound for positive constants D_1, D_2.

$$\mathbb{E}\|f_X^{(n,T,h)} - f_X\|^2 \leqslant D_1h^{2\alpha} + D_2n^{-2\alpha/(2\alpha+\dim \mathbf{X})}.$$

The upper bound converges at the fastest possible rate when we choose the bandwidth h such that

$$h^{2\alpha} = n^{-2\alpha/(2\alpha+\dim \mathbf{X})},$$

giving us a convergence rate of $n^{-\alpha/(\alpha+1)}$. Thus, the bandwidth h optimizing the upper bound is

$$h(n) = n^{-1/(2\alpha+\dim \mathbf{X})}.$$

Consequently, the above cutoff T and bandwidth h give the inequality

$$\mathbb{E}\|f_X^{(n,T,h)} - f_X\|^2 \leqslant Cn^{-2\alpha/(2\alpha+\dim \mathbf{X})},$$

where C is a positive constant neither dependent on n nor T. The right side is the optimal rate of convergence for the upper bound term.

7.3 Lower bound: Proof of Theorem 4.3

Let U be a normal and convex neighborhood in $\mathbf{X}$, isometric to Euclidean space. Then the Sobolev ball of L_2-functions on U with smoothing parameter α lies in the Sobolev ball of L_2-functions on $\mathbf{X}$ with smoothing parameter α, by extending functions $f : U \to \mathbb{R}$ to $\mathbf{X}$ by setting $f(x) = 0$ for all $x \in \mathbf{X} - U$ by the discussion on Sobolev spaces in [24] on page 7. The minimax rate, $n^{-2\alpha/(2\alpha+\dim \mathbf{X})}$ [29, Theorem 24.4], for Euclidean kernel density estimators on $\mathbb{R}^{\dim \mathbf{X}}$ thus lower bounds the convergence rate for $\mathbf{G}$-kernel density estimators.

References

[1] "Apple's Internal Innovation Network Unraveled - Part1 - Evolving networks." *Kenedict Innovation Analytics*, http://www.kenedict.com/blog/apples-internal-innovation-network-unraveled.

[2] Asta, Dena. "Kernel Density Estimation on Symmetric Spaces." *preprint available upon request.*

[3] Asta, Dena and Cosma Shalizi. "Separating Biological and Social Contagions in Social Media: The Case of Regional Flu Trends in Twitter." *preprint available upon request.*

[4] Begelfor, Evgeni, and Michael Werman. "The world is not always flat or learning curved manifolds." *School of Engineering and Computer Science, Hebrew University of Jerusalem.,Tech. Rep* (2005).

[5] Bhattacharya, Abhishek, Rabi N. Bhattacharya. "Nonparametric statistics on manifolds with applications to shape spaces." *IMS Collections* 3 (2008): 282-301.

[6] Boguna, Marian, Fragkiskos Papadopoulos, and Dimitri Krioukov. "Sustaining the Internet with hyperbolic mapping." *Nature communications* 1 (2010): 62.

[7] Bonk, Mario and Oded Schramm. "Embeddings of Gromov hyperbolic spaces." *In Selected Works of Oded Schramm, pp.243-284.* New York: Springer, 2011.

[8] Cannon, James W., William J. Floyd, Richard Kenyon, and Walter R. Parry. "Hyperbolic geometry." *Flavors of geometry* 31, (1997): 59-115.

[9] Cvetkovski, Andrej, and Mark Crovella. "Multidimensional Scaling in the Poincaré Disk." *arXiv preprint arXiv:* 1105.5332 (2011).

[10] Diaconis, Persi and Svante Janson. "Graph limits and exchangeable random graphs." *Rend. Mat. Appl.(7)* 28, (2008): 33-61.

[11] Henderson, J. A. and P. A. Robinson. "Geometric effects on complex network structure in the cortex." *Physical Review Letters* 107 (2011): 018102.

[12] Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. "Latent space approaches to social network analysis." *Journal of the american Statistical association* 97, no. 460 (2002): 1090-1098.

[13] Huckemann, Stephan F., Peter T. Kim, Ja-Yong Koo, and Axel Munk. "Mobius deconvolution on the hyperbolic plane with application to impedance density estimation." *The Annals of Statistics* 38, no. 4 (2010): 2465-2498.

[14] Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. "Goodness of fit of social network models." *Journal of the american Statistical association* 103, no. 481 (2008).

[15] Janson, Svante. "Graphons, cut norm and distance, couplings and rearrangements." *NYJM Monographs* 4, (2013).

[16] Jonckheere, Edmond, Poonsuk Lohsoonthorn, and Francis Bonahon. "Scaled Gromov hyperbolic graphs." *Journal of Graph Theory* 57, no. 2 (2008): 157-180.

[17] Kim, Yoon Tae, and Hyun Suk Park. "Geometric structures arising from kernel density estimation on Riemannian manifolds." *Journal of Multivariate Analysis* 114 (2013): 112-126.

[18] Krioukov, Dimitri, Fragkiskos Papadopoulos, Marian Boguna, and Amin Vahdat. "Efficient navigation in scale-free networks embedded in hyperblic metric spaces." *CoRR, May* (2008).

[19] Krioukov, Dimitri, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marian Boguna. "Hyperbolic geometry of complex networks." *Physical Review E* 82, no. 3 (2010): 036106.

[20] Lovàsz, L.. "Very Large Graphs." *Current Developments in Mathematics* vol. 2008 (2009): 67-128.

[21] Narayanan, E. K, and S. K. Ray. "The heat kernel and Hardy's theorem on symmetric spaces of noncompact type." *Proceedings of the Indian Academy of Sciences-Mathematical Sciences,* vol. 112, no. 2 (2002): 321-330.

[22] Pao, Henry, Glen A. Coppersmith, and Carey E. Priebe. "Statistical inference on random graphs: Comparative power analyses via Monte Carlo." *Journal of Computational and Graphical Statistics* 20 (2011): 395-416.

[23] Pelletier, Bruno. "Kernel density estimation on Riemannian manifolds." *Statistics & probability letters* 73, no. 3 (2005): 297-304.

[24] Pesenson, Isaac. "A discrete Helgason-Fourier transform for Sobolev and Besov functions on noncompact symmetric spaces." *Contemp. Math* 464 (2008): 231-249.

[25] Rahman, Inam Ur, Iddo Drori, Victoria C. Stodden, David L. Donoho, and Peter Schrder. "Multiscale representations for manifold-valued data." *Multiscale Modeling & Simulation* 4, no. 4 (2005): 1201-1232.

[26] Shalizi, Cosma, Chris Genovese, and Andrew Thomas. "Network Comparison." *Unpublished MS* (2012).

[27] Terras, Audrey. *Harmonic analysis on symmetric spaces and applications I.* New York: Springer-Verlag, 1985.

[28] Terras, Audrey. *Harmonic analysis on symmetric spaces and applications II.* New York: Springer-Verlag, 1985.

[29] Van der Vaart, Aad W.. *Asymptotic Statistics.* Cambridge, UK: Cambridge University Press, 1998.