



HAL
open science

On the Estimation of Latent Distances Using Graph Distances

Ery Arias-Castro, Antoine Channarond, Bruno Pelletier, Nicolas Verzelen

► **To cite this version:**

Ery Arias-Castro, Antoine Channarond, Bruno Pelletier, Nicolas Verzelen. On the Estimation of Latent Distances Using Graph Distances. *Electronic Journal of Statistics*, 2021, 15 (1), pp.722-747. 10.1214/21-EJS1801 . hal-02376073

HAL Id: hal-02376073

<https://hal.science/hal-02376073>

Submitted on 22 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Estimation of Latent Distances Using Graph Distances

Ery Arias-Castro* Antoine Channarond† Bruno Pelletier‡ Nicolas Verzelen§

Abstract

We are given the adjacency matrix of a geometric graph and the task of recovering the latent positions. We study one of the most popular approaches which consists in using the graph distances and derive error bounds under various assumptions on the link function. In the simplest case where the link function is an indicator function, the bound is (nearly) optimal as it (nearly) matches an information lower bound.

1 Introduction

Suppose that we observe a undirected graph with adjacency matrix $W = (W_{ij} : i, j \in [n])$ (where $[n] := \{1, \dots, n\}$ and $n \geq 3$) with $W_{ij} \in \{0, 1\}$ and $W_{ii} = 0$. We assume the existence of points, $x_1, \dots, x_n \in \mathbb{R}^v$, such that

$$\mathbb{P}(W_{ij} = 1 \mid x_1, \dots, x_n) = \phi(\|x_i - x_j\|), \quad (1)$$

for some non-increasing link function $\phi : [0, \infty) \mapsto [0, 1]$. The $(W_{ij}, i < j)$ are assumed to be independent. We place ourselves in a setting where the adjacency matrix W is observed, but the underlying points are unknown. We will be mostly interested in settings where ϕ is unknown (and no parametric form is known).

Our most immediate interest is in the pairwise distances

$$d_{ij} := \|x_i - x_j\|. \quad (2)$$

Since we assume the link function is unknown, all we can hope for is to rank these distances. Indeed, the most information we can aspire to extract from W is the probability matrix $P := (p_{ij})$, where

$$p_{ij} := \mathbb{P}(W_{ij} = 1 \mid x_1, \dots, x_n), \quad (3)$$

and even with perfect knowledge of P , the distances can only be known up to a monotone transformation, since $p_{ij} = \phi(d_{ij})$ and ϕ is an arbitrary monotone (here non-increasing) function.

Once a possibly incomplete, and likely only approximate, ranking of the distances has been produced, the problem of recovering the latent positions amounts to a problem of ordinal embedding (aka, non-metric multidimensional scaling), which has a long history [39] dating back to pioneering work by Shepard [29, 30] and Kruskal [15]. Because of this, we focus on the pairwise distances rather than the latent positions themselves.

* Dept. of Mathematics, University of California, San Diego (USA)

† LMRS – UMR 6085 CNRS – Université de Rouen (France)

‡ IRMAR – UMR CNRS 6625 – Université Rennes II (France)

§ INRA – UMR 729 MISTEA – Montpellier (France)

This work was partially supported by grants from the US National Science Foundation (DMS 1513465).

1.1 Related work

This is an example of a latent graph model and the points are often called latent positions. In its full generality, the model includes the planted partition model popular in the area of graph partitioning. To see this, take $r = 1$ and let v denote the number of blocks and, with e_k denoting the k -th canonical basis vector, set $x_i = e_k$ if i belongs to block k . The planted partition model is a special case of the stochastic blockmodel of Holland et al. [13]. This is also a special case of our model, as can be seen by changing e_k to z_s , chosen so that $\phi(\|z_s - z_\ell\|) = p_{k\ell}$, the connection probability between blocks k and ℓ . Mixed-membership stochastic blockmodels as in [1, 2, 38] are also special cases of latent graph models, but of a slightly different kind. The literature on the stochastic blockmodel is now substantial and includes results on the recovery of the underlying communities; see, e.g., [6, 11, 17, 21, 24, 31] and references therein.

Our contribution here is of a different nature as we focus on the situation where the latent positions are well spread out in space, forming no obvious clusters. This relates more closely to the work of Hoff et al. [12]. Although their setting is more general in that additional information may be available at each position, their approach in our setting reduces to the following logistic regression model:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = -d_{ij}, \quad (4)$$

which is clearly a special case of (1), with $\phi(d) = 1/(1 + e^d)$. Sarkar et al. [25] consider this same model motivated by a link prediction problem where the nodes are assumed to be embedded in space with their Euclidean distances being the dissimilarity of interest. In fact, they assume that the points are uniformly distributed in some region. They study a method based on the number of neighbors that a pair of nodes have in common, which is one of the main methods for link prediction [18, 19]. Parthasarathy et al. [23] consider a more general setting where a noisy neighborhood graph is observed: if (x_i) are points in a metric space with pairwise distances (d_{ij}) , then an adjacency matrix, $W = (W_{ij})$, is observed, where $W_{ij} = 1$ with probability $1 - p$ if $d_{ij} \leq r$ and with probability q if $d_{ij} > r$. Under fairly general conditions on the metric space and the sampling distributions, and additional conditions on (n, r, p) , they show that the graph distances computed based on W provide, with high probability, a 2-approximation to the underlying distances in the case where $q = 0$. In the case where $q > 0$, the same is true, under some conditions on (n, r, p, q) , if W is replaced by $\tilde{W} = (\tilde{W}_{ij})$ where $\tilde{W}_{ij} = 1$ exactly when $\frac{N_{ij}}{N_i + N_j - N_{ij}} \geq \tau$, where τ is a (carefully chosen) parameter of the method, where $N_i := \#\{j : W_{ij} = 1\}$ (number of neighbors) and $N_{ij} := \#\{k : W_{ik} = W_{jk} = 1\}$ (number of common neighbors).

Scheinerman and Tucker [26] and Young and Scheinerman [40] consider what they call a dot-product random graph model where $p_{ij} = \langle x_i, x_j \rangle$, where it is implicitly assumed that $\langle x_i, x_j \rangle \in [0, 1]$ for all $i \neq j$. This model is a special case of (1), with $\phi(d) = 1 - \frac{1}{2}d^2$. Sussman et al. [32] consider recovering the latent positions in this model (with full knowledge of the link function). They devise a spectral method which consists in embedding the items $\{1, \dots, n\}$ as points in \mathbb{R}^v (where v is assumed known) as the row vectors of $U_{(v)} \Theta_{(v)}^{1/2}$, where $W = U \Theta V^\top$ is the SVD of W , and for a matrix $A = (A_{ij})$ and an integer $s \geq 1$, $A_{(s)} = (A_{ij} : i \vee j \leq s)$. They analyze their method in a context where the latent positions are in fact a sample from a (possibly unknown) distribution. The same authors extended their work in [33] to an arbitrary link function, which may be unknown, although the focus is on a binary classification task in a setting where for each $i \in [n]$ a binary label y_i is available.

von Luxburg and Alamgir [36] consider the closely related problem of recovering the latent positions in a setting where a nearest-neighbor graph is available. They proposed a method based on estimating the underlying density (denoted f). If \hat{f}_i denotes the density estimate at x_i , a graph

is defined on $[n]$ with weights $w_{ij} = (\hat{f}_i^{-1/v} + \hat{f}_j^{-1/v})/2$, and d_{ij} is estimated by the graph distance between nodes i and j . Terada and von Luxburg [35] derive some theory for this method based on results conjectured in [36].

1.2 Our contribution

Graph distances are well-known estimates for the Euclidean distances in the context of graph drawing [16, 28], where the goal is to embed items in space based on an incomplete distance matrix. They also appear in the literature on link prediction [18, 19] and are part of the method proposed in [36]. We examine the use of graph distances for the estimation of the Euclidean distances (2). As we shall see, the graph distances are directly useful when the link function ϕ is compactly supported, which is for example the case in the context of a neighborhood graph where $\phi(d) = \mathbb{I}\{d \leq r\}$ for some connectivity radius $r > 0$. In fact, the method is shown to achieve a minimax lower bound in this setting (under a convexity assumption). This setting is discussed in Section 2. In Section 3, we extend the analysis to other (compactly supported) link functions. In Section 4, we briefly discuss a regularization known as Maximum Variance Unfolding (MVU), due to Weinberger et al. [37]. In Section 5, we discuss the problem of embedding a nearest-neighbor graph in a setting where the positions represent a sample from a uniform distribution. In this context, the graph distance method can be seen as a simple variant of the method proposed by von Luxburg and Alamgir [36] and subsequently analyzed by Terada and von Luxburg [35]. We show that, in dimension $v \geq 2$, the method is biased due to a boundary effect that persists even as the number of positions increases without bound. We end with Section 6, where we discuss some important limitations of the method based on graph distances and consider some extensions, including localization (to avoid the convexity assumption) and the use of the number of common neighbors (to accommodate non-compact link functions). Proofs are gathered in Section 7.

1.3 Preliminaries

Given the adjacency matrix W , the graph distance between nodes i and j is defined as

$$\delta_{ij} := \inf \left\{ \ell : \exists k_0, \dots, k_\ell \in [n] \text{ s.t. } k_0 = i, k_\ell = j, \text{ and } W(k_{s-1}, k_s) = 1, \forall s \in [\ell] \right\}, \quad (5)$$

where $\inf \emptyset = \infty$ by convention. (Here and elsewhere, we will sometimes use the notation $W(i, j)$ for W_{ij} , $d(i, j)$ for d_{ij} , etc.) We propose estimating, up to a scale factor, the Euclidean distances (2) with the graph distances (5). Note that since ϕ is unknown the scale factor cannot be recovered from the data.

The method is the analog of the MDS-D method of Kruskal and Seery [16] for graph drawing, which is a setting where some of the distances (2) are known and the goal is to recover the missing distances. Let \mathcal{E} denote the set of pairs $i < j$ for which d_{ij} is known. MDS-D estimates the missing distances with the distances in the graph with node set $[n]$ and edge set \mathcal{E} , and with edge $(i, j) \in \mathcal{E}$ weighed by d_{ij} . This method was later rediscovered by Shang et al. [28], who named it MDS-MAP(P), and coincides with the IsoMap procedure of Tenenbaum et al. [34] for isometric manifold embedding. (For more on the parallel between graph drawing and manifold embedding, see the work of Chen and Buja [5].)

As we shall see, the graph distance method is most relevant when the positions are sufficiently dense in their convex hull, which is a limitation it shares with MDS-D. For $\Omega \subset \mathbb{R}^v$ and $x_1, \dots, x_n \in \mathbb{R}^v$, define

$$\Lambda_\Omega(x_1, \dots, x_n) = \sup_{x \in \Omega} \min_{i \in [n]} \|x - x_i\|, \quad (6)$$

which measures how dense the latent points are in Ω . We also let $\Lambda(x_1, \dots, x_n)$ denote (6) when Ω is the convex hull of $\{x_1, \dots, x_n\}$.

2 Simple setting

In this section we focus on the simple, yet emblematic case of a neighborhood (ball) graph, that is, a setting where the link function is given by $\phi(d) = \mathbb{I}\{d \leq r\}$ for some $r > 0$. We start with a performance bound for the graph distance method and then establish minimax lower bound. Similar results are available in [3, 4, 23], among other places, and we only provide a proof for completeness, and also to pave the way to the more sophisticated Theorem 3.

Theorem 1. *Assume that $\phi(d) = \mathbb{I}\{d \leq r\}$ for some $r > 0$, and define $\hat{d}_{ij} = r\delta_{ij}$. Then*

$$0 \leq \hat{d}_{ij} - d_{ij} \leq 4(\varepsilon/r)d_{ij} + r, \quad \forall i, j \in [n], \quad (7)$$

for any set of points x_1, \dots, x_n that satisfy $\Lambda(x_1, \dots, x_n) \leq \varepsilon$ with $\varepsilon \leq r/4$.

For a numerical example, see Figure 1. In Figure 2 we confirm numerically that the method is biased when the underlying domain from which the positions are sampled is not convex. That said, the method is robust to mild violations of the convex constraint, as shown in Figure 3, where the positions correspond to $n = 3000$ US cities.¹

Remark 1. Computations were done in R, with the graph distances computed using the `igraph` package, to which classical scaling was applied, followed by a procrustes alignment and scaling using the package `vegan`.

Note that \hat{d} is not a true estimator in general as it relies on knowledge of r , which is a feature of the unknown link function. Nevertheless, the result says that, up to that scale parameter, the graph distances achieve a nontrivial level of accuracy.

It turns out that the graph distance method comes close to achieving the best possible performance (understood in a minimax sense) in this particularly simple setting. Indeed, we are able to establish the following general lower bound that applies to any method.

Theorem 2. *Assume that $\phi(d) = \mathbb{I}\{d \leq r\}$ with $r \leq 1/2$ (without loss of generality). Then there is a numeric constant $c > 0$ with the property that, for any $\varepsilon > 0$ and any estimator² \hat{d} , there is $x_1, \dots, x_n \in \Omega := [0, 1]$ (so that $v = 1$ here) such that $\Lambda_\Omega(x_1, \dots, x_n) \leq \varepsilon$ and, for at least half of the pairs $i \neq j$,*

$$|\hat{d}_{ij} - d_{ij}| \geq \frac{c\varepsilon}{r \vee \varepsilon} d_{ij}. \quad (8)$$

Note that ϕ is known in Theorem 2 and that we are talking about bonafide estimators.

3 General setting

Beyond the setting of a neighborhood graph considered in Section 2, the graph distance method, in fact, performs similarly in more generality. We consider in this section the case of a link function that is compactly supported and establish a performance bound that is comparable to that of Theorem 1.

¹ These were sampled at random from the dataset available at simplemaps.com/data/us-cities

² An estimator here is a function on the set of n -by- n symmetric binary matrices with values in $\mathbb{R}_+^{n(n-1)/2}$.

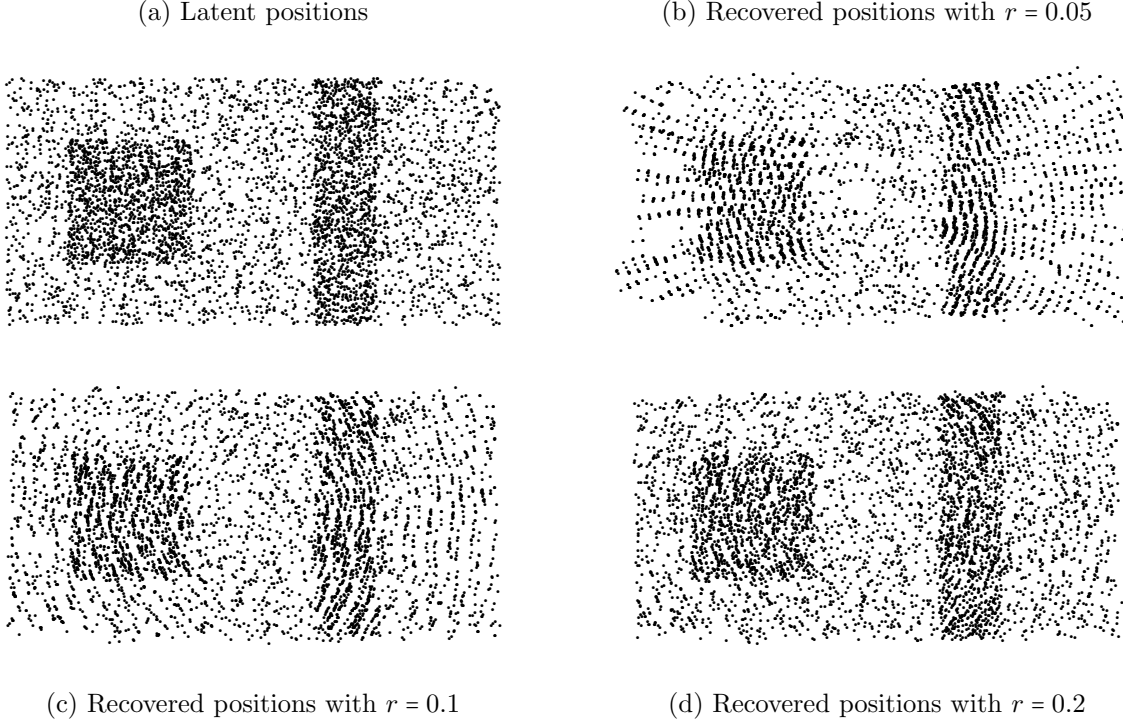


Figure 1: A numerical example illustrating the setting of Theorem 1. Here $n_0 = 3000$ positions were sampled uniformly at random from $\Omega_0 := [0, 2] \times [0, 1]$, $n_1 = 1000$ from $\Omega_1 := [0.25, 0.75] \times [0.25, 0.75]$, and $n_2 = 1000$ from $\Omega_2 := [1.25, 1.5] \times [0, 1]$, for a total of $n = 5000$ positions.

Theorem 3. Assume that ϕ has support $[0, r]$, for some $r > 0$, and define $\hat{d}_{ij} = r\delta_{ij}$. Assume that, for some $c_0 > 0$ and $\alpha \geq 0$, $\phi(d) \geq c_0(1 - d/r)^\alpha$ for all $d \in [0, r]$. Then there are $C_1, C_2 > 0$ depending only on (α, c_0) such that, whenever $r/\varepsilon \geq C_1(\log n)^{1+\alpha}$, with probability at least $1 - 1/n$,

$$0 \leq \hat{d}_{ij} - d_{ij} \leq C_2 \left[(\varepsilon/r)^{\frac{1}{1+\alpha}} d_{ij} + r \right], \quad \forall i \neq j, \quad (9)$$

for any points x_1, \dots, x_n that satisfy $\Lambda(x_1, \dots, x_n) \leq \varepsilon$ and $\text{diam}(x_1, \dots, x_n) \leq 1$.

See Figure 4 for a numerical example. In the bound (9), the behavior of $|\hat{d}_{ij} - d_{ij}|$ depends mainly on how $\phi(d)$ behaves near the edge of its support, namely as $d \nearrow r$. If $\alpha = 0$, then ϕ is discontinuous at r , and we recover a bound similar to that of Theorem 1, proved there for the case where $\phi(d) = \mathbb{I}\{d \leq r\}$. The bound obtained in this case appears near-optimal in view of Theorem 2. For $\alpha > 0$, we do not know whether the bound (9) is similarly near-optimal. In general, we expect the graph distance procedure to be less accurate as α increases, and in particular, even though the result does not cover the case of a function ϕ converging quicker to 0, such as $\phi(d) = c_0 \exp(-(r - d)_+^{-1})$, we speculate that the graph distance does not perform well in this case.

4 Regularization by maximum variance unfolding

With the graph distances (5) computed, it might be tempting to regularize them by an approximation with Euclidean distances. We say that a metric γ on $[n]$ is Euclidean if there are points in some Euclidean space, y_1, \dots, y_n , such that $\gamma_{ij} = \|y_i - y_j\|$. Let \mathcal{E}_n denote the set of such metrics on

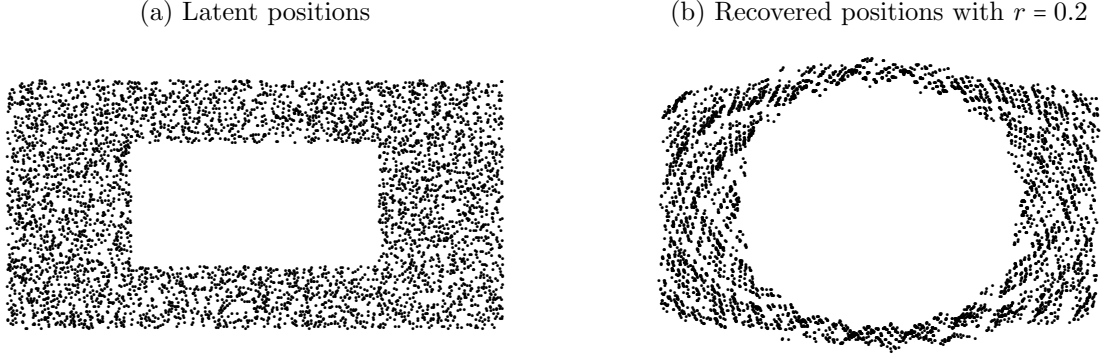


Figure 2: A numerical example illustrating the setting of Theorem 1 showing that the convexity constraint is indeed required for the graph distance method to be unbiased. Here $n = 5000$ positions were sampled uniformly at random from $\Omega := [0, 2] \times [0, 1] \setminus [0.5, 1.5] \times [0.25, 0.75]$.

$[n]$ and consider the following optimization problem

$$\min_{\gamma \in \mathcal{E}_n} \sum_{i < j} |\gamma_{ij}^2 - \delta_{ij}^2|. \quad (10)$$

Let γ^* denote a solution and let y_1^*, \dots, y_n^* be points in a Euclidean space such that $\gamma_{ij}^* = \|y_i^* - y_j^*\|$.

Other measures of discrepancy between γ and δ are of course possible. The reason we consider this particular form is that it corresponds to Maximum Variance Unfolding (MVU), a well-known method for graph drawing proposed by Weinberger, Sha, Zhu, and Saul [37], although it is perhaps best known as a method for manifold embedding. Given a weighed graph on $[n]$ with weight matrix W , MVU consists in solving the following optimization problem:

$$\text{maximize } \sum_{i \neq j} \|y_i - y_j\|^2 \quad \text{over } y_1, \dots, y_n \in \mathbb{R}^d, \quad (11)$$

$$\text{such that } \|y_i - y_j\| \leq W_{ij}, \quad \forall i, j \in [n]. \quad (12)$$

Paprotny and Garcke [22] have shown that the two formulations are equivalent when (δ_{ij}) denotes the corresponding graph distances.

Proposition 1. *We have $\gamma_{ij}^* \leq \delta_{ij}$ for all $i, j \in [n]$.*

Proof. We only need to consider a pair of nodes $i \neq j$ such that $\delta_{ij} < \infty$, for otherwise the bound holds by convention. Let $k_0 = i, k_1, \dots, k_\ell = j$ denote a shortest path in the graph connecting i and j , so that $\ell = \delta_{ij}$. We then derive

$$\gamma_{ij}^* = \|y_i^* - y_j^*\| \leq \sum_{s=0}^{\ell-1} \|y_{k_{s-1}}^* - y_{k_s}^*\| \leq \ell = \delta_{ij}, \quad (13)$$

using the triangle inequality and then the fact that $\|y_{k_{s-1}}^* - y_{k_s}^*\| \leq 1$ for all s , since y_1^*, \dots, y_n^* satisfies (12) and $W(k_{s-1}, k_s) = 1$ for all s . \square

We also have the following, which results from a straightforward adaptation of [22, Th 3].

Proposition 2. *Assume there is $r > 0$ and $\eta \in (0, 1)$ such that $\hat{d}_{ij} := r\delta_{ij}$ satisfies $(1 - \eta)d_{ij} \leq \hat{d}_{ij} \leq (1 + \eta)d_{ij}$ for all $i, j \in [n]$. Letting $\tilde{d}_{ij} := r\gamma_{ij}^*$, we have that there is a universal constant $c > 0$ such that*

$$\sum_{i < j} |\tilde{d}_{ij}^2 - d_{ij}^2| \leq c\eta \sum_{i < j} d_{ij}^2. \quad (14)$$

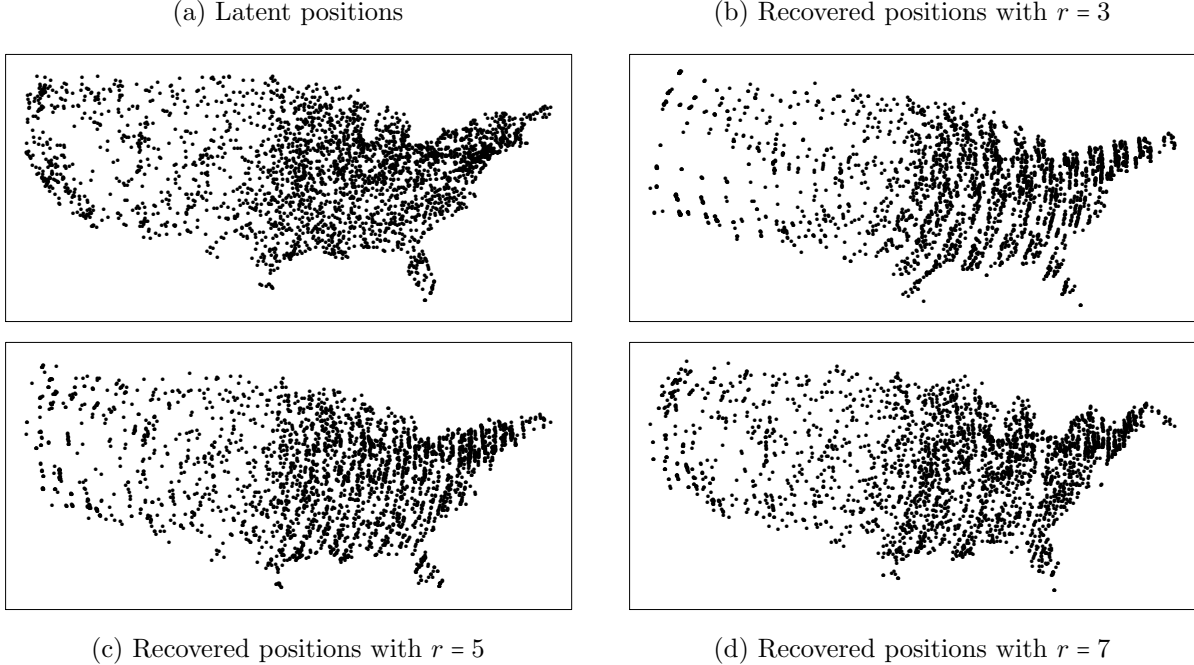


Figure 3: A numerical example illustrating the setting of Theorem 1. The latent positions are located at the coordinates of $n = 3000$ US cities and the connectivity radius varies (in degrees).

This proposition, together with Theorem 3, allows one to bound the performance of the MVU regularization of the graph distances.

5 Embedding a nearest-neighbor graph

In this section we turn our attention to a different, yet closely related setting, that of embedding a nearest-neighbor graph. We observe the κ -nearest-neighbor graph given by the adjacency matrix $A = (A_{ij})$, where $A_{ij} = 1$ if x_j is among the κ -nearest-neighbors of x_i , and $A_{ij} = 0$ otherwise. Note that A is not symmetric in general. The goal remains the same, that is, to recover the Euclidean distances (2).

For a set $\Omega \subset \mathbb{R}^v$, define its radius, $\rho(\Omega)$, as half its diameter; its inradius, $\rho_o(\Omega)$, as the radius of the largest ball included in Ω ; and for $u > 0$, $\Omega^{\ominus u} = \{x \in \Omega : \text{dist}(x, \partial\Omega) > u\}$, which is possibly empty (and called an erosion in mathematical morphology). If Ω is measurable, then $|\Omega|$ denotes its Lebesgue measure. Let β denote the volume of the unit ball in \mathbb{R}^v .

Theorem 4. *Let $v \geq 2$ and assume that x_1, \dots, x_n were generated iid from the uniform distribution on a compact, convex set $\Omega \subset \mathbb{R}^v$ with non-empty interior, and define $\omega = |\Omega|/\beta$. There are $C_1, C_2 > 0$ depending only on $(v, \rho(\Omega), \rho_o(\Omega))$ such that the following holds. Define $\hat{d}_{ij} = r\delta_{ij}$, where $r := r_o + \varepsilon$, where $r_o := (\omega\kappa/n)^{1/v}$ and $\varepsilon := C_1(\log(n)/n)^{1/v}$. If $\kappa \geq C_2 \log n$ and $n \geq C_2\kappa$, with probability at least $1 - 1/n$,*

$$\hat{d}_{ij} - d_{ij} \leq 8(\varepsilon/r)d_{ij} + r, \quad \forall i, j \in [n], \quad (15)$$

as well as

$$\hat{d}_{ij} \geq d_{ij}, \quad \forall i, j \in [n] \text{ such that } d_{ij} \geq 2r \text{ and } x_i, x_j \in \Omega^{\ominus(d_{ij}/2)}. \quad (16)$$

In fact, when $v = 1$, $\hat{d}_{ij} \geq d_{ij}$ for all $i, j \in [n]$.

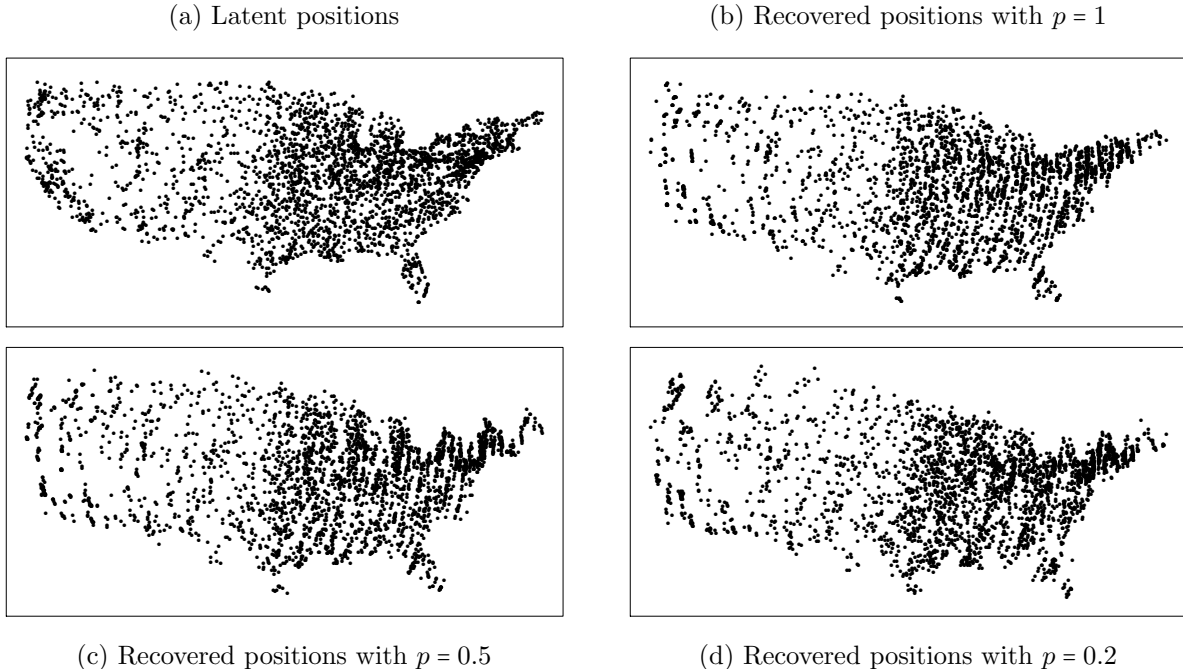


Figure 4: Same setting as in Figure 3. Here we set $r = 5$ and vary p . (In fact, to ease the comparison, we coupled the different adjacency matrices in the sense that the $(p = 0.2)$ -matrix was built by erasing edges from the $(p = 0.5)$ -matrix independently with probability $0.2/0.5 = 0.4$.

Remark 2. A similar, yet slightly different result may be obtained for the setting where $v = 1$. We leave the details up to the reader’s curiosity.

Thus r , as defined in Theorem 4 is an appropriate scaling for the graph distances for the pairs of items satisfying (16). However there are pairs for which this scaling is not accurate, and the two statement combined show that there is no scaling that works in the sense of making the graph distances close to the Euclidean distances. That boundary effect does not go away in the large-sample limit. See Figure 5 for a numerical example. What happens is that the boundary serves as a ‘freeway’, because for a point near the boundary, its k nearest neighbors are farther away. See Figure 6 for an illustration.

Rather than argue the existence of pairs for which r is not an accurate scaling for a general Ω (satisfying the conditions of Theorem 4), which we claim is true as stated, we focus on an example. Arguably, this is the most extreme situation, but it is not hard to see that it generalizes to other set Ω by zooming in on the boundary. The example is in dimension $v = 2$, but only for concreteness, as the same phenomenon is easily seen to generalize to any dimension $v \geq 3$.

Proposition 3. *Let $\Omega = [0, b] \times [0, 1] \subset \mathbb{R}^2$, with $b \geq 3$, and consider the statement of Theorem 4. When C_2 is sufficiently large, there is $q > 1$ depending only on b such that, with probability at least $1 - 1/n$,*

$$\hat{d}_{ij} \leq (1 - q)d_{ij}, \quad \text{for all } i, j \in [n] \text{ such that } d_{ij} \geq 2. \quad (17)$$

Remark 3. Above we considered the approach put forth by von Luxburg and Alamgir [36] naturally adapted to the setting where the underlying distribution is a uniform distribution. The actual method proposed by von Luxburg and Alamgir [36], however, may actually be consistent for estimating the distances (again, up to scaling). Indeed, it is quite possible that the density estimate along the boundary is biased in just the proper way.

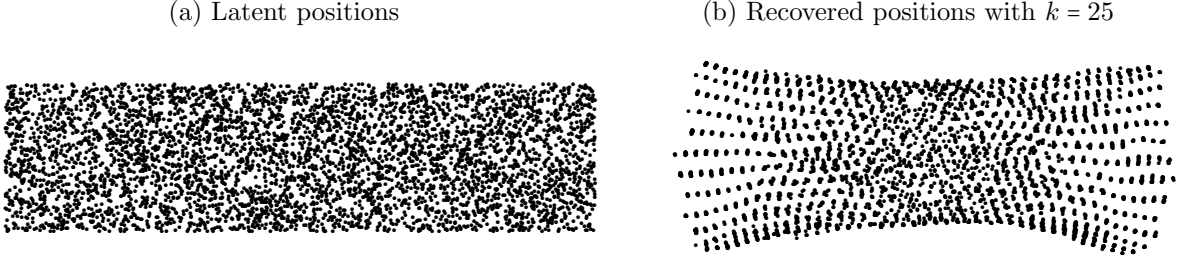


Figure 5: A numerical example illustrating the setting of Proposition 3. Here $\Omega = [0, 4] \times [0, 1]$ (so that $\omega = 4$) and $n = 5000$.

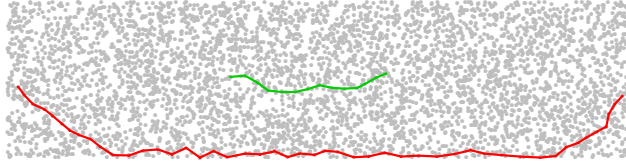


Figure 6: This continues with the setting of Figure 5. Each colored polygonal line is the shortest path in the κ -nearest-neighbor graph between its endpoints. While the distance estimate for the pair of positions that are nearby (joined by the green path) is accurate, the distance estimate the pair of positions that are farther apart (joined by the red path) is not. This is congruent with our statements in Theorem 4 and Proposition 3.

Remark 4. We mention that the nearest-neighbor setting is also considered in [8], where a synchronization approach is proposed. More generally, we are looking at an ordinal embedding problem, since triplet distance comparisons can be obtained as follows

$$A_{ij} = 1 \text{ and } A_{ik} = 0 \Rightarrow d_{ij} < d_{ik}. \quad (18)$$

However, none of the methods we know of come with theoretical guarantees — although [14] comes close.

6 Discussion

The method based on graph distances studied in the previous section suffers from a number of serious limitations:

1. The positions need to span a convex set. (In practice, the method is robust to mild violations of this constraint as exemplified in Figure 3.)
2. Even in the most favorable setting of Section 2, the relative error is still of order r , as established (9), and this does not seem improvable. (To see this, contrast the situations where d_{ij} is just below r with the case where it is just above r .)
3. The link function needs to be compactly supported. Indeed, the method can be grossly inaccurate in the presence of long edges, as in the interesting case where the link function is of the form

$$\phi(d) = p\mathbb{I}\{d \leq r\} + q\mathbb{I}\{d > r\}, \quad (19)$$

where $0 < q < p \leq 1$.

We address each of these three issues in what follows.

6.1 Localization

A possible approach to addressing Issue 1 is to operate locally. This is well-understood and is what lead Shang and Ruml [27] to suggest MDS-MAP(P), which effectively localizes MDS-MAP [28]. (As we discussed earlier, the latter is essentially a graph-distance method and thus bound by the convexity constraint.) More recent methods for graph drawing based on ‘synchronization’ also operate locally [7, 9].

Experimentally, this strategy works well. See Figure 7 for a numerical example, which takes place in the context of the rectangle with a hole of Figure 2. We adopted a simple approach: we kept the graph distances that were below a threshold, leaving the other ones unspecified, and then applied a method for multidimensional scaling with missing values, specifically SMACOF [10] (initialized with the output of the graph distance method).

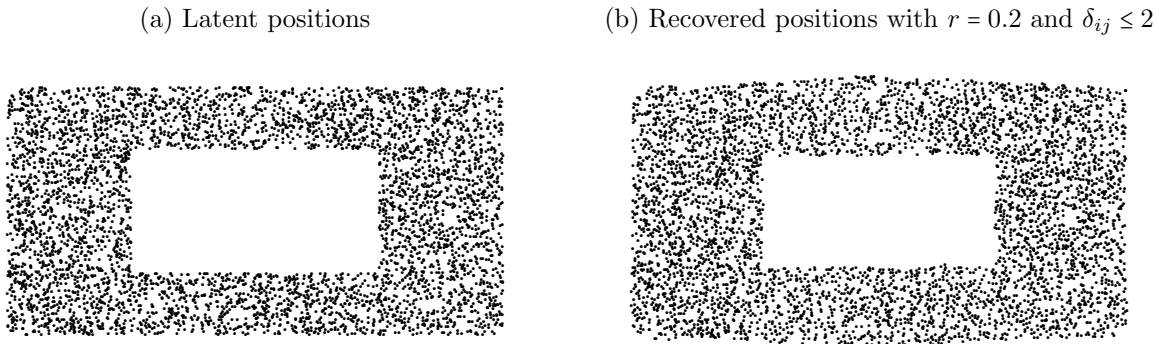


Figure 7: Same setting as in Figure 2.

6.2 Regularization by multidimensional scaling

Regarding Issue 2, in numerical experiments we have found that the graph distances, although grossly inaccurate, are nevertheless useful for embedding the points using (classical) multidimensional scaling. This remains surprising to us and we do not have a good understanding of the situation. (Note if one is truly interested in estimating the Euclidean distances, one may use graph distances, apply multidimensional scaling, and then compute the distances between the embedded points.) For a numerical illustration, see Figure 8.

6.3 Number of common neighbors

A possible approach to addressing Issue 3 (and also Issue 2) is to work with the number of common neighbors, which provides an avenue to ‘super-resolution’ in a way. By this we mean that, say in the simple setting of Section 2, although the adjacency matrix only tells whether two positions are within distance r , it is possible to gather all this information to refine this assessment. Similarly, in the setting where (19) is the link function, it is possible to tell whether two positions are nearby or not. This sort of concentration is well-known to the expert and seems to be at the foundation of spectral methods (see, e.g., [32, Prop 4.2]). We refer the reader to [23, 25], where such an approach is considered in great detail.

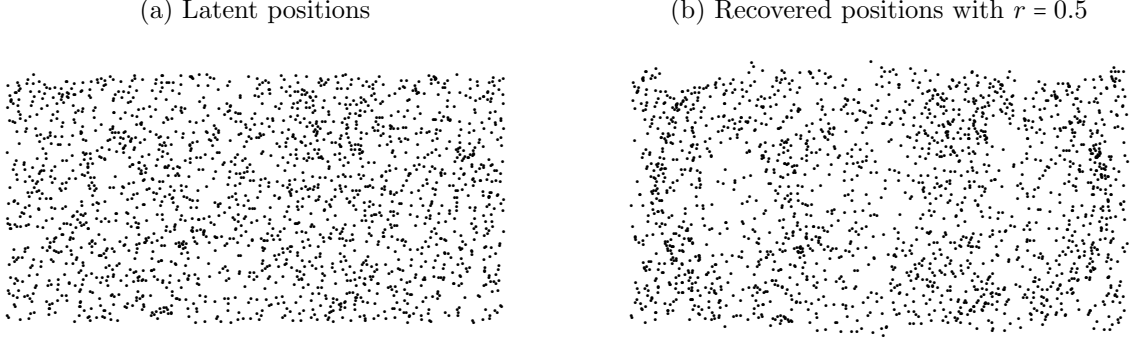


Figure 8: Here $n = 2000$ positions were sampled uniformly at random from $\Omega := [0, 2] \times [0, 1]$. For all $i \neq j$, $\delta_{ij} \in \{1, 2, 3, 4, 5\}$, so that the graph distances are rather discrete, yet the embedding computed by classical multidimensional scaling is surprisingly accurate.

7 Proofs

7.1 Proof of Theorem 1

Fix $i, j \in [n]$ distinct.

Let $m := \lfloor d_{ij}/(r - 2\varepsilon) \rfloor$ and note that $m(r - 2\varepsilon) \leq d_{ij} \leq (m + 1)(r - 2\varepsilon)$. For $s \in \{0, \dots, m + 1\}$, let $z_s = x_i + \frac{s}{m+1}(x_j - x_i)$. We have $z_0 = x_i$ and $z_{m+1} = x_j$, and z_0, z_1, \dots, z_{m+1} are on the line joining x_i and x_j and satisfy $\|z_s - z_{s+1}\| \leq r - 2\varepsilon$ for all s . Let x_{k_s} be such that $\|z_s - x_{k_s}\| \leq \varepsilon$, with $x_{k_0} = x_i$ and $x_{k_{m+1}} = x_j$. Note that x_{k_s} is well-defined since z_s belongs to the convex hull of $\{x_1, \dots, x_n\}$ and we have assumed that $\Lambda(x_1, \dots, x_n) \leq \varepsilon$. By the triangle inequality, for all $s \in \{0, \dots, m\}$,

$$\|x_{k_s} - x_{k_{s+1}}\| \leq \|x_{k_s} - z_s\| + \|z_s - z_{s+1}\| + \|z_{s+1} - x_{k_{s+1}}\| \leq \varepsilon + (r - 2\varepsilon) + \varepsilon \leq r. \quad (20)$$

Hence, $(x_{k_0}, x_{k_1}, \dots, x_{k_{m+1}})$ forms a path in the graph, and as a consequence, $\delta_{ij} \leq m + 1$. In turn, this implies that

$$\hat{d}_{ij} = r\delta_{ij} \leq rm + r \leq r \frac{d_{ij}}{r - 2\varepsilon} + r \leq d_{ij} + 4(\varepsilon/r)d_{ij} + r, \quad (21)$$

using the fact that $\varepsilon \leq r/4$.

Resetting the notation, let $k_0 = i, k_1, \dots, k_\ell = j$ denote a shortest path joining i and j , so that $\ell = \delta_{ij}$. By the triangle inequality,

$$d_{ij} = \|x_{k_0} - x_{k_\ell}\| \leq \sum_{s=0}^{\ell-1} \|x_{k_s} - x_{k_{s+1}}\| \leq \ell r = r\delta_{ij} = \hat{d}_{ij}, \quad (22)$$

using the fact that $\|x_{k_s} - x_{k_{s+1}}\| \leq r$ for all s .

7.2 Proof of Theorem 2

We construct two point configurations that yield the same adjacency matrix and then measure the largest difference between the corresponding sets of pairwise distances.

Assume that $m := r(n - 1)$ is an integer for convenience.

- *Configuration 1.* In this configuration,

$$x_i = \frac{i - 1}{n - 1}, \quad i \in [n]. \quad (23)$$

Note that $\Lambda_\Omega(x_1, \dots, x_n) = 1/(2n - 2)$.

- *Configuration 2.* In this configuration,

$$x_i = \frac{(i-1)(1-\eta(i-1))}{(n-1)(1-\eta(n-1))}, \quad i \in [n]. \quad (24)$$

for some $\eta > 0$ chosen small later on. When $\eta \leq 1/(2n-3)$, which we assume, x_i is increasing with $x_1 = 0$ and $x_n = 1$. Note that

$$\Lambda_\Omega(x_1, \dots, x_n) = \frac{1-\eta}{(2n-2)(1-\eta(n-1))}.$$

The two configurations coincide when $\eta = 0$, but we will choose $\eta > 0$ in what follows. Under Configuration 1, the adjacency matrix W is given by $W_{ij} = \mathbb{I}\{|i-j| \leq m\}$. For the design matrix to be the same under Configuration 2, it suffices that x_1 have (exactly) m neighbors (to the right) and that x_n have (exactly) m neighbors (to the left); this is because $i \mapsto x_i - x_{i-1}$ is decreasing in this configuration. These two conditions correspond to four equations, given by

$$x_{m+1} - x_1 \leq r, \quad x_{m+2} - x_1 > r, \quad |x_n - x_{n-m}| \leq r, \quad x_n - x_{n-m-1} > r. \quad (25)$$

We need only consider the first and fourth as they imply the other two. After some simplifications, we see that the first one holds when $r \leq 1$, while the fourth holds when $r \leq 1 - 2/(n-1)$ and $\eta \leq 1/(2n-3+m(n-m-3))$. Since $r \leq 1/2$, $r \leq 1 - 2/(n-1)$ when $n \geq 5$, and we choose $\eta = 1/(2n+m(n-m))$ for example. Then $\Lambda_\Omega(x_1, \dots, x_n) \sim 1/2n$ in Configuration 2 (same as in Configuration 1). We choose $n = n_\varepsilon$ just large enough that $\Lambda_\Omega(x_1, \dots, x_n) \leq \varepsilon$ in both configurations. In particular, $\varepsilon \sim 1/n_\varepsilon$ as $\varepsilon \rightarrow 0$. Since the result only needs to be proved for ε small, we may take n as large as we need.

Now that the two designs have the same adjacency matrix, we cannot distinguish them with the available information. It therefore suffices to look at the difference between the pairwise distances. Indeed, if $d_{ij}^{(k)}$ denotes the distance between x_i and x_j in Configuration k , then for any estimator \hat{d}_{ij} ,

$$\max_{k \in \{1,2\}} |\hat{d}_{ij} - d_{ij}^{(k)}| \geq \frac{1}{2} |d_{ij}^{(1)} - d_{ij}^{(2)}|, \quad \forall i < j. \quad (26)$$

For $i < j$, we have

$$d_{ij}^{(1)} = \frac{j-i}{n-1}, \quad d_{ij}^{(2)} = \frac{j-i}{n-1} \frac{1-\eta(j+i-2)}{1-\eta(n-1)}, \quad (27)$$

so that

$$d_{ij}^{(1)} - d_{ij}^{(2)} = \frac{j-i}{n-1} \frac{\eta(i+j-n-1)}{1-\eta(n-1)}. \quad (28)$$

Let

$$\mathcal{P} = \{(i, j) : 1 \leq i < j \leq n \text{ and } i+j \leq \lceil n/\sqrt{2} \rceil + 1\}, \quad (29)$$

and note that \mathcal{P} include at least half of all pairs in $\{(i, j) : 1 \leq i < j \leq n\}$. For $(i, j) \in \mathcal{P}$, $d_{ij}^{(1)} < d_{ij}^{(2)}$ so that $d_{ij}^{(1)} = d_{ij}^{(1)} \wedge d_{ij}^{(2)}$, and given the condition on r and our choice for η ,

$$|d_{ij}^{(1)} - d_{ij}^{(2)}| \geq d_{ij}^{(1)} \frac{\eta(n+1-n/\sqrt{2}-2)}{1-\eta(n-1)} \geq (d_{ij}^{(1)} \wedge d_{ij}^{(2)}) \frac{c_0}{1 \vee rn}, \quad n \geq n_0, \quad (30)$$

for some universal constants $n_0 \geq 1$ and $c_0 > 0$. We then conclude with the fact that $1/(1 \vee rn) = (1/n)/(r \vee (1/n)) \geq c_1 \varepsilon / (r \vee \varepsilon)$ for some universal constant $c_1 > 0$, due to the fact that $\varepsilon \sim 1/n$.

7.3 Proof of Theorem 3

As before in (22), we have $\hat{d}_{ij} \geq d_{ij}$ for all $i \neq j$. Recall the definition of $p_{ij} \equiv p(i, j)$ in (3). Let $p_0 = \phi(r/2) > 0$ and note that $p_0 \geq c_0(1/2)^\alpha$. Below, c_1, c_2 , etc, will denote positive constants that only depend on (α, c_0) . Since the result only needs to be proved for large r/ε , we will take this quantity as large as needed. In what follows, we connect each node in the graph to itself. This is only for convenience and has no impact on the validity of the resulting arguments.

Special case. Suppose that $d_{kl} \leq r/2$ for all $k \neq l$. In that case, for all $i \neq j$, $p_{ij} = \phi(d_{ij}) \geq \phi(r/2) = p_0$. For (i, j, k) distinct, (x_i, x_k, x_j) forms a path in the graph if and only if $W_{ik}W_{kj} = 1$, which happens with probability $p_{ik}p_{kj} \geq p_0^2$. Therefore, by independence,

$$\mathbb{P}(\delta_{ij} > 2) \leq \mathbb{P}(W_{ik}W_{kj} = 0, \forall k \notin \{i, j\}) \leq (p_0^2)^{n-2}. \quad (31)$$

Therefore, by the union bound, with probability at least $1 - n^2 p_0^{2n-4} \geq 1 - n^2 \exp(-c_1 n)$, we have $\delta_{ij} \leq 2$, implying $\hat{d}_{ij} \leq 2r$, for all $i \neq j$.

Henceforth, we assume that

$$\max_{i \neq j} d_{ij} > r/2. \quad (32)$$

Claim 1. By choosing C_1 large enough, the following event happens with probability at least $1 - 1/n^2$,

$$\mathcal{A}_1 := \{\hat{d}_{ij} \leq 9d_{ij} + 2r, \text{ for all } i, j \in [n]\}. \quad (33)$$

Take i, j such that $d_{ij} \leq r/4$. We first note that there is j_* such that $d(i, j_*) > r/4$, for otherwise, for all $k, l \in [n]$, $d_{kl} \leq d_{ki} + d_{il} \leq r/4 + r/4 = r/2$, which would contradict our assumption (32).

Define

$$z_s = x_i + s \frac{\varepsilon(x_{j_*} - x_i)}{d(j_*, i)}, \quad s \in [m], \quad (34)$$

where $m := \lfloor (r/4 - \varepsilon)/\varepsilon \rfloor$. By construction each z_s is on the line segment joining x_i and x_{j_*} , and so belongs to the convex hull of x_1, \dots, x_n ; hence, by the fact that $\Lambda(x_1, \dots, x_n) \leq \varepsilon$, there is $i_s \in [n]$ be such that $\|x_{i_s} - z_s\| \leq \varepsilon$. By the triangle inequality,

$$d(i, i_s) = \|x_i - x_{i_s}\| \leq \|x_i - z_s\| + \|z_s - x_{i_s}\| \leq s\varepsilon + \varepsilon \leq m\varepsilon + \varepsilon \leq (r/4 - \varepsilon) + \varepsilon = r/4, \quad (35)$$

and

$$d(i_s, j) \leq d(i_s, i) + d(i, j) \leq r/4 + r/4 = r/2. \quad (36)$$

Therefore, for each $s \in [m]$, (x_i, x_{i_s}, x_j) forms a path with probability at least p_0^2 . By independence, therefore, there is such an $s \in [m]$ with probability at least $1 - (1 - p_0^2)^m$.

With the union bound and the fact that $m \asymp r/\varepsilon$, we may conclude that, if C_1 is chosen large enough, the event

$$\mathcal{A}_2 := \{\hat{d}_{ij} \leq 2r \text{ for all } i \neq j \text{ such that } d_{ij} \leq r/4\}, \quad (37)$$

has probability at least $1 - 1/n^2$, since

$$\mathbb{P}(\mathcal{A}_2^c) \leq n^2(1 - p_0^2)^m \leq n^2(1 - c_0^2/2^{2\alpha})^m \leq n^2 \exp(-c_2(r/\varepsilon)) \leq 1/n^2, \quad (38)$$

eventually, by our assumption that $r/\varepsilon \geq C_1(\log n)^{1/(1+\alpha)}$, where C_1 will be chosen large.

Next, we prove that \mathcal{A}_2 implies \mathcal{A}_1 , which will suffice to establish the claim. For this, it remains to consider the other cases. Therefore, take i, j such that $d_{ij} > r/4$, and define $z_0 = x_i$ and

$$z_s = x_i + s \frac{\varepsilon(x_j - x_i)}{d(j, i)}, \quad \text{for } s \in [m]. \quad (39)$$

As before, for each $s \in [m]$, there is $i_s \in [n]$ such that $\|x_{i_s} - z_s\| \leq \varepsilon$. We let $i_0 = i$ and $i_m = j$. The latter is possible since $\|z_m - x_j\| \leq \varepsilon$.

We have

$$d(i_s, i_{s'}) = \|x_{i_s} - x_{i_{s'}}\| \leq \|z_s - z_{s'}\| + 2\varepsilon = \varepsilon|s - s'| + 2\varepsilon, \quad (40)$$

so that, under \mathcal{A}_2 ,

$$d(i_s, i_{s'}) \leq r/4 \quad \text{when } |s - s'| \leq h := \lfloor r/4\varepsilon - 2 \rfloor, \quad (41)$$

implying that $\hat{d}(i_s, i_{s'}) \leq 2r$ when $|s - s'| \leq h$. Thus, by the triangle inequality, under \mathcal{A}_2 ,

$$\hat{d}(i, j) = \sum_{k=0}^{\lfloor m/h \rfloor - 1} \hat{d}(i_{kh}, i_{(k+1)h}) + \hat{d}(i_{\lfloor m/h \rfloor h}, j) \leq \lfloor m/h \rfloor 2r + \hat{d}(i_{\lfloor m/h \rfloor h}, j). \quad (42)$$

By the triangle inequality,

$$d(i_{\lfloor m/h \rfloor h}, j) \leq \|z_{\lfloor m/h \rfloor h} - x_j\| + \varepsilon = |d_{ij} - \lfloor m/h \rfloor h \varepsilon| + \varepsilon, \quad (43)$$

and it is not hard to verify that $0 \leq d_{ij} - \lfloor m/h \rfloor h \varepsilon \leq (h+1)\varepsilon$, so that $d(i_{\lfloor m/h \rfloor h}, j) \leq (h+2)\varepsilon \leq r/4$, implying under \mathcal{A}_2 that $\hat{d}(i_{\lfloor m/h \rfloor h}, j) \leq 2r$. Hence, under \mathcal{A}_2 ,

$$\hat{d}(i, j) \leq 2r(\lfloor m/h \rfloor + 1) \leq 2r(d_{ij}/(r/4 - 3\varepsilon) + 1) \leq 9d_{ij} + 2r, \quad (44)$$

when ε/r is small enough, as we needed to prove.

The claim, of course, falls quite short of proving the theorem, but we use it in the remainder of the proof. However, it does prove that (9) holds for all $i, j \in [n]$ such that $d_{ij} \leq 2r$. Thus, for the remainder of the proof, we only need focus on i, j such that $d_{ij} > 2r$. Define m and z_0, \dots, z_m as before, and also the corresponding $i_s \in [n]$.

As before, (40) implies that

$$d(i_s, i_{s'}) \leq r \quad \text{when } |s - s'| \leq h := \lfloor r/\varepsilon - 2 \rfloor, \quad (45)$$

and in particular

$$d(i_s, j) \leq r \quad \text{when } s \geq m - h. \quad (46)$$

(We changed the definition of h .) Similarly, we have

$$d(i_s, j) = \|x_{i_s} - x_j\| \geq \|z_s - z_m\| - \varepsilon = \varepsilon|s - s'| - \varepsilon, \quad (47)$$

so that

$$d(i_s, j) > r \quad \text{when } s \leq m - h + 2. \quad (48)$$

Starting at $S_0 = 0$, if $S_{t-1} = s \in \mathbb{N}$, define

$$S_t = \max \{s' : s \leq s' \leq s + h : W(i_s, i_{s'}) = 1\}. \quad (49)$$

This is a maximum since we have set $W(k, k) = 1$ for all $k \in [n]$. Define the following stopping time

$$T = \inf \{t : S_t \geq s^* := m - h + 3\}, \quad (50)$$

with the convention that $\inf \emptyset = \infty$.

Define $U_t = h - (S_t - S_{t-1})$ and then $V_t = U_1 + \dots + U_t$, for $t \geq 1$, and $U_0 = 0$, and note that $S_t = th - V_t$. Importantly, if $S_{t-1} < s^*$ and $U_t = h$, then $U_{t'} = h$ for all $t' \geq t - 1$, or equivalently, $S_{t'} = S_{t-1}$ for all $t' \geq t$, so that $T = \infty$.

Take $s < s^*$. Conditional on $S_{t-1} = s$,

$$U_t = \inf \{0 \leq u \leq h : W(i_s, i_{s+h-u}) = 1\}. \quad (51)$$

If $U_{t-1} < h$, $W(i_s, i_{s+h-u}) \sim \text{Bern}(p(i_s, i_{s+h-u}))$ are jointly independent, and also independent of S_{t-1} . Hence,

$$\mathbb{P}(U_t > u \mid S_{t-1} = s, U_{t-1} < h) = \prod_{k=0}^u (1 - p(i_s, i_{s+h-k})), \quad 0 \leq u \leq h - 1.$$

By the fact that ϕ is non-increasing and (40),

$$p(i_s, i_{s'}) = \phi(d(i_s, i_{s'})) \geq \phi(\varepsilon|s - s'| + 2\varepsilon) =: g(|s - s'|). \quad (52)$$

Hence,

$$\mathbb{P}(U_t > u \mid S_{t-1} = s, U_{t-1} < h) \leq \prod_{k=0}^u (1 - g(h - k)), \quad 0 \leq u \leq h. \quad (53)$$

Let $\{\bar{U}_t : t \geq 1\}$ be iid with distribution that of \bar{U} , supported on $\{0, \dots, h\}$ and with survival function

$$\mathbb{P}(\bar{U} > u) = \prod_{k=0}^u (1 - g(h - k)), \quad 0 \leq u \leq h, \quad (54)$$

and define $\bar{V}_t = \bar{U}_1 + \dots + \bar{U}_t$. Then U_t , conditional on $U_{t-1} < h$, is stochastically dominated by \bar{U}_t . Moreover, V_t , conditional on $U_{t-1} < h$, is stochastically dominated by \bar{V}_t , which follows from [20, Th 5.4] and a simple recursion.

For $0 \leq u \leq h$, we have

$$\mathbb{P}(\bar{U} > u) = \prod_{k=0}^u (1 - g(h - k)) \leq \exp\left(-\sum_{k=0}^u g(h - k)\right), \quad (55)$$

and by the fact that g is non-increasing on $[0, \infty)$ with support in $[0, h + 1]$,

$$\begin{aligned} \sum_{k=0}^u g(h - k) &\geq \sum_{k=0}^u \int_{h-k}^{h-k+1} g(y) dy = \int_{h-u}^{h+1} g(y) dy \geq \int_{h-u}^{h+1} c_0 (1 - (\varepsilon y + 2\varepsilon)/r)_+^\alpha dy \\ &= c_0 (\varepsilon/r)^\alpha \int_{h-u}^{h'} (h' - y)^\alpha dy \\ &= c_0 (\varepsilon/r)^\alpha (u - h + h')^{1+\alpha} / (1 + \alpha) \\ &\geq c_3 (\varepsilon/r)^\alpha (u - 1)_+^{1+\alpha}, \end{aligned}$$

where $h' := (r - 2\varepsilon)/\varepsilon$. Therefore,

$$\mathbb{P}(\bar{U} > u) \leq 2 \exp\left(-c_3 (\varepsilon/r)^\alpha (u - 1)_+^{1+\alpha}\right), \quad u \geq 0. \quad (56)$$

Hence, \bar{U} is stochastically bounded by $1 + \nu Y$, where $\nu := (r/\varepsilon)^{\alpha/1+\alpha}$ and Y is defined by its distribution

$$\mathbb{P}(Y > y) = \exp(-c_3 y^{1+\alpha}), \quad y \geq 0. \quad (57)$$

Let $\{Y_t : t \geq 1\}$ be iid with distribution that of Y .

Using the fact that $S_t = th - V_t$, we have

$$T > t \Leftrightarrow S_t < s^* \Leftrightarrow V_t > th - s^*, \quad (58)$$

and in particular

$$\begin{aligned} \mathbb{P}(T > t) &= \mathbb{P}(V_t > th - s^*) \leq \mathbb{P}(V_t > th - s^* \mid U_t < h) + \mathbb{P}(U_t = h) \\ &\leq \mathbb{P}(\bar{V}_t > th - s^*) + \mathbb{P}(U_t = h). \end{aligned}$$

On the one hand,

$$\mathbb{P}(U_t = h) \leq \mathbb{P}(U_t = h \mid U_{t-1} < h) + \mathbb{P}(U_{t-1} = h) \leq \dots \leq \sum_{k=1}^t \mathbb{P}(U_k = h \mid U_{k-1} < h) \leq t\mathbb{P}(\bar{U} = h),$$

with

$$\mathbb{P}(\bar{U} = h) = \mathbb{P}(\bar{U} > h - 1) \leq 2 \exp(-c_3(\varepsilon/r)^\alpha(h-1)^{1+\alpha}) \leq \exp(-c_4(r/\varepsilon)), \quad (59)$$

using the fact that $h \asymp r/\varepsilon$. On the other hand, \bar{V}_t is stochastically dominated by $t + \nu(Y_1 + \dots + Y_t)$, so that by Chernoff's bound,

$$\mathbb{P}(\bar{V}_t > t(1 + \nu a)) \leq \mathbb{P}(Y_1 + \dots + Y_t > ta) \leq \exp(-t\zeta(a)), \quad \forall a \geq 0, \quad (60)$$

where ζ is the rate function of Y ; in particular, there is $c_5 > 0$ such that $\zeta(a) \geq c_5(a - a_0)$ for all $a \geq a_0 := \mathbb{E}(Y)$. (Note that a_0 is just another constant.) In particular,

$$\mathbb{P}(\bar{V}_t > th - s^*) \leq \exp(-tc_5[(h - s^*/t - 1)\nu^{-1} - a_0]_+). \quad (61)$$

Define $t \geq t^* := 2d_{ij}/r$, and note that, if $D := \text{diam}(x_1, \dots, x_n)$, then $t^* \leq 2D/r$, with $r \geq C_1\varepsilon(\log n)^{1+\alpha}$ and $\varepsilon \geq D/2n$ so that $t^* \leq \frac{2}{C_1}n/(\log n)^{1+\alpha}$, so we may assume that $t^* \leq n$. Now, for r/ε sufficiently large,

$$\begin{aligned} h - s^*/t &\geq h - s^*/t^* = h - (m - h + 3)/(2d_{ij}/r) \\ &\geq r/\varepsilon - 3 - (d_{ij}/\varepsilon - r/\varepsilon + 5)/(2d_{ij}/r) \\ &\geq r/\varepsilon - 3 - (d_{ij}/\varepsilon + 5)/(2d_{ij}/r) \geq r/2\varepsilon - 4, \end{aligned}$$

using the fact that $d_{ij} > 2r$, and this yields

$$\mathbb{P}(\bar{V}_t > th - s^*) \leq \exp(-tc_6(r/\varepsilon)^{\frac{1}{1+\alpha}}). \quad (62)$$

Thus, we have

$$\mathbb{P}(T > t) \leq \exp(-tc_6(r/\varepsilon)^{\frac{1}{1+\alpha}}) + t \exp(-c_4r/\varepsilon), \quad \forall t \geq t^*, \quad (63)$$

and since T cannot be finite and exceed m , and $m \leq n$, this implies that

$$\mathbb{P}(T = \infty) \leq n \exp(-c_4r/\varepsilon). \quad (64)$$

Furthermore, again for any $t \geq t^*$,

$$\mathbb{P}(V_T > t(1 + 2\nu a_0) \cup T > t) \leq \mathbb{P}(V_T > t(1 + 2\nu a_0) \cap T \leq t) + \mathbb{P}(T > t) \quad (65)$$

$$\leq \mathbb{P}(V_t > t(1 + 2\nu a_0)) + \mathbb{P}(T > t) \quad (66)$$

$$\leq \exp(-t\zeta(2a_0)) + 2n \exp(-c_4 r/\varepsilon) + \exp(-tc_6(r/\varepsilon)^{\frac{1}{1+\alpha}}) \quad (67)$$

$$\leq 2 \exp(-tc_7) + 2n \exp(-c_4 r/\varepsilon), \quad (68)$$

when r/ε is large enough. For $t = t^\circ := t^* + (4/c_7) \log n$, the right-hand side is bounded by $2/n^4$ eventually, thus the following event happens with probability at least $1 - 1/n^4$,

$$\mathcal{C} := \left\{ V_T \leq t^\circ(1 + 2\nu a_0) \text{ and } T \leq t^\circ \right\}. \quad (69)$$

A joint control on V_T and T is useful because of the following. Under $\{T < \infty\}$, $(i, i_{S_1}, \dots, i_{S_T})$ forms a path in the graph, so that $\hat{d}(i, i_{S_T}) \leq Tr$. We also have

$$d_{ij} \geq \sum_{t=1}^T \|z_{S_t} - z_{S_{t-1}}\| = \sum_{t=1}^T \varepsilon(S_t - S_{t-1}) = \varepsilon(S_T - S_0) = \varepsilon S_T. \quad (70)$$

Thus, under \mathcal{C} , we have $S_T = hT - V_T \geq hT - (1 + 2\nu a_0)t^\circ$, implying via (70) that

$$d_{ij} \geq \varepsilon S_T \geq \varepsilon hT - \varepsilon(1 + 2\nu a_0)t^\circ \geq rT - c_8 \nu \varepsilon t^\circ, \quad (71)$$

using the fact that $h \geq r/\varepsilon - 3$ and $\nu \geq 1$ when r/ε is large enough. Furthermore,

$$\begin{aligned} \nu \varepsilon t^\circ &= (r/\varepsilon)^{\alpha/(1+\alpha)} \varepsilon (2d_{ij}/r + (3/c_7) \log n) \\ &= (\varepsilon/r)^{1/(1+\alpha)} d_{ij} + [(3/c_7)(\varepsilon/r)^{1/(1+\alpha)} (\log n)] r \\ &\leq (\varepsilon/r)^{1/(1+\alpha)} d_{ij} + (3/c_7) C_1^{-1/(1+\alpha)} r \\ &\leq (\varepsilon/r)^{1/(1+\alpha)} d_{ij} + r, \end{aligned}$$

when C_1 is large enough. Thus, under \mathcal{C} ,

$$\hat{d}(i, i_{S_T}) \leq Tr \leq d_{ij} + c_8 \nu \varepsilon t^\circ \leq d_{ij} + c_8 ((\varepsilon/r)^{1/(1+\alpha)} d_{ij} + r). \quad (72)$$

And since, as in (40),

$$d(i_{S_T}, j) \leq \varepsilon(m - S_T) + \varepsilon \leq \varepsilon(h - 3) + \varepsilon \leq \varepsilon(r/\varepsilon - 5) + \varepsilon \leq r. \quad (73)$$

Using the union bound, we have established the following.

Claim 2. *By choosing C_1 large enough, with probability at least $1 - 1/n^2$, for all $i, j \in [n]$ such that $d_{ij} > 2r$, there is $k \in [n]$ such that $\hat{d}_{ik} \leq d_{ij} + c_8 ((\varepsilon/r)^{1/(1+\alpha)} d_{ij} + r)$ and $d_{jk} \leq r$. (We denote this event by \mathcal{A}_3 .)*

Assume that $\mathcal{A}_1 \cap \mathcal{A}_3$ holds, which happens with probability at least $1 - 2/n^2$ based on the two claims that we have established above. In that case, let $i, j, k \in [n]$ be as in the last claim. Because \mathcal{A}_1 holds, $d_{jk} \leq r$ implies that $\hat{d}_{jk} \leq 11r$. Thus, with the union bound,

$$\hat{d}_{ij} \leq \hat{d}_{ik} + \hat{d}_{jk} \leq d_{ij} + c_8 ((\varepsilon/r)^{1/(1+\alpha)} d_{ij} + r) + 11r, \quad (74)$$

so that

$$\hat{d}_{ij} - d_{ij} \leq c_9 ((\varepsilon/r)^{1/(1+\alpha)} d_{ij} + r). \quad (75)$$

This proves that (9) holds for all $i, j \in [n]$ such that $d_{ij} > 2r$.

7.4 Proof of Theorem 4

We will need the following lemmas. Let $\mathbf{B}(x, r)$ denote the open ball centered at x with radius r .

Lemma 1. *Let $\Omega \subset \mathbb{R}^v$ be compact, convex, and with non-empty interior. Then $0 < \rho_o(\Omega) \leq \rho(\Omega) < \infty$, and there is a constant C that depends only on $(v, \rho(\Omega), \rho_o(\Omega))$ such that*

$$|\Omega \cap \mathbf{B}(x, r)| \geq C(r \wedge 1)^v, \quad \forall x \in \Omega, \quad \forall r > 0. \quad (76)$$

Proof. Let ρ and ρ_o be short for $\rho(\Omega)$ and $\rho_o(\Omega)$. It is enough to establish such a lower bound for $r \leq \rho_o$, which is what we assume below. Let $x_0 \in \Omega$ be such that $\mathbf{B}(x_0, \rho_o) \subset \Omega$. Fix $x \in \Omega$ and let $h = \|x - x_0\|$, and also let u denote an arbitrary normed vector.

Case 1: $h \leq \rho_o$. Then

$$|\Omega \cap \mathbf{B}(x, r)| \geq |\mathbf{B}(x_0, \rho_o) \cap \mathbf{B}(x, r)| \geq |\mathbf{B}(0, \rho_o) \cap \mathbf{B}(\rho_o u, r)|. \quad (77)$$

Given that $r \leq \rho_o$, the last terms is easily seen to be $\geq C_1 r^v$, for a constant $C_1 > 0$ that only depends on (v, ρ_o) .

Case 2: $h > \rho_o$. Let $\mathbf{C}(x, x', b)$ denote the right circular cone with apex x and base with center x' and radius b . By convexity of Ω , and because x and $\mathbf{B}(x_0, \rho_o)$ are both in Ω , so is the convex hull of $\{x\} \cup \mathbf{B}(x_0, \rho_o)$. Since that convex hull contains $\mathbf{C}(x, x_0, \rho_o)$, Ω also contains that cone, which implies that

$$|\Omega \cap \mathbf{B}(x, r)| \geq |\mathbf{C}(x, x_0, \rho_o) \cap \mathbf{B}(x, r)| = |\mathbf{C}(0, hu, \rho_o) \cap \mathbf{B}(0, r)|. \quad (78)$$

Using the fact that $t \mapsto |\mathbf{C}(0, tu, \rho_o) \cap \mathbf{B}(0, r)|$ is decreasing on $[r, \infty)$, together with $r \leq \rho_o < h \leq 2\rho_o$, we obtain

$$|\mathbf{C}(0, hu, \rho_o) \cap \mathbf{B}(0, r)| \geq |\mathbf{C}(0, 2\rho_o u, \rho_o) \cap \mathbf{B}(0, r)|, \quad (79)$$

and the last term is easily seen to be $\geq C_2 r^v$ for a constant $C_2 > 0$ that only depends on (v, ρ, ρ_o) . \square

Lemma 2. *For integer $n \geq 1$, any $p \in (0, 1)$, and any $\eta \geq 0$, we have*

$$\mathbb{P}(|\text{Bin}(n, p) - np| \geq n\eta) \leq 2 \exp\left(-\frac{n\eta^2}{2p(1-p) + 2\eta/3}\right). \quad (80)$$

Proof. The result follows from an application of Bernstein's inequality. \square

Lemma 3. *Consider points on the real line, $x_1 < \dots < x_n$, and the κ -nearest-neighbor graph they form. Then all the shortest paths are increasing in the sense that, if $k_0 = i, k_1, \dots, k_\ell = j$ is a shortest path joining i and j , with $i < j$, then necessarily $k_0 < k_1 < \dots < k_\ell$.*

Proof. Suppose that $k_0 = i, k_1, \dots, k_\ell = j$ is a path joining i and j , with $i < j$ and all nodes being distinct. If the path is not increasing, there are $s < s' < s''$ such that $k_{s'} < k_s < k_{s''}$ with $x_{k_{s''}}$ being a κ -nearest neighbor of $x_{k_{s'}}$. But this necessarily implies that $x_{k_{s''}}$ is also a κ -nearest neighbor of x_{k_s} , so that the chain $k_0, \dots, k_s, k_{s''}, \dots, k_\ell$ is also a path joining i and j , and has strictly smaller length. \square

Proof of Theorem 4. We divide the proof into several steps. Below, c_0, c_1, c_2 , etc, denote are positive constants that only depend on $(v, \rho(\Omega), \rho_o(\Omega))$. Let $\lambda = \Lambda_\Omega(x_1, \dots, x_n)$, which is random in the current situation. For $x \in \Omega$ and $t > 0$, let $N(x, t) = \#\{i : x_i \in \mathbf{B}(x, t)\}$, so that $N(x, t) \sim \text{Bin}(n, p(x, t))$

where $p(x, t) := |\Omega \cap \mathbf{B}(x, t)|/|\Omega|$, although note that $N(x_i, t) | x_i \sim \text{Bin}(n-1, p(x_i, t))$. We have $p(x, t) = t^v/\omega$ when $x \in \Omega^{\ominus t}$, and in any case, by Lemma 1,

$$p(x, t) \geq c_0(t \wedge 1)^v, \quad \forall x \in \Omega, \quad \forall t > 0. \quad (81)$$

For $i \in [n]$, define r_i to be the distance from x_i to the κ -th nearest point among $x_j, j \neq i$, which are random in the present setting. The statements below are when the constants in the statement, C_1, C_2 , are chosen large enough, which in particular implies that n is large enough as well. (Note also that C_1 is chosen large enough and then, C_1 being fixed, C_2 is chosen large enough.)

Step 1: $\mathcal{A}_1 := \{\lambda \leq \varepsilon\}$ happens with probability at least $1 - 1/n^2$.

For $t > 0$, let $m(t)$ be smallest³ such that there is $y_1, \dots, y_m \in \Omega$ with the property that $\Omega \subset \cup_{s=1}^m \mathbf{B}(y_s, t)$. Note that $m(t) \leq c_{1,1}(t \wedge 1)^{-v}$. If there is an x_i in each ball in that collection, then $\lambda \leq 2t$. Therefore, assuming $t \leq 1$,

$$\begin{aligned} \mathbb{P}(\lambda \geq 2t) &\leq \mathbb{P}(\exists s \in [m(t)] : N(y_s, t) = 0) \\ &\leq \sum_{s=1}^{m(t)} (1 - p(y_s, t))^n \\ &\leq m(t)(1 - c_0 t^v)^n \\ &\leq \exp(-nc_0 t^v - v \log t + c_{1,2}), \end{aligned}$$

using the union bound and then (81). Hence, for n large enough that $\varepsilon \leq 1$,

$$\begin{aligned} \mathbb{P}(\lambda \geq \varepsilon) &\leq \exp(-nc_0(\varepsilon/2)^v - \log(\varepsilon/2)^v + c_{1,2}) \\ &\leq \exp(-(c_0(C_1/2)^v - 1) \log n - \log \log n - \log C_1 + c_{1,3}) \\ &\leq 1/n^2, \quad \text{eventually.} \end{aligned}$$

Step 2.1: $\mathcal{A}_2^+ := \{r_i \leq r, \forall i \in \mathcal{J}_r\}$, where $\mathcal{J}_r := \{i : x_i \in \Omega^{\ominus r}\}$, happens with probability at least $1 - 1/n^2$.

For $i \in \mathcal{J}_r$, we have $p(x_i, r) = p_1 := r^v/\omega$, so that

$$\begin{aligned} \mathbb{P}(r_i \geq r) &= \mathbb{P}(N(x_i, p(x_i, r)) < \kappa) \\ &= \mathbb{P}(\text{Bin}(n-1, p_1) \leq \kappa - 1) \\ &\leq \mathbb{P}(\text{Bin}(n, p_1) \leq \kappa). \end{aligned}$$

Letting $p_\circ := r_\circ^v/\omega = \kappa/n$, and noting that $p_\circ \leq p_1$, we may apply Lemma 2 to obtain

$$\begin{aligned} \mathbb{P}(\text{Bin}(n, p_1) \leq \kappa) &\leq 2 \exp\left(-\frac{n(p_1 - p_\circ)^2}{2p_1(1 - p_1) + 2(p_1 - p_\circ)/3}\right) \\ &\leq 2 \exp\left(-\frac{n(p_1 - p_\circ)^2}{3p_1}\right). \end{aligned}$$

We have

$$p_1 - p_\circ = (r^v - r_\circ^v)/\omega = ((r_\circ + \varepsilon)^v - r_\circ^v)/\omega \geq v r_\circ^{v-1} \varepsilon / \omega. \quad (82)$$

We also have

$$r_\circ/\varepsilon = (\omega \kappa/n)^{1/v} / C_1 (\log(n)/n)^{1/v} \geq \omega^{1/v} C_2^{1/v} / C_1, \quad (83)$$

³ m may be called the covering number of Ω , although the term has several definitions in use.

and in particular $p_1 \leq 2p_o$ when C_2 is large enough. In that case,

$$\begin{aligned} n(p_1 - p_o)^2/p_1 &\geq c_{2,1}n(r_o^{v-1}\varepsilon)^2/r_o^v \\ &= c_{2,2}C_1^2\kappa^{1-2/v}(\log n)^{2/v} \\ &\geq c_{2,3}C_1^2C_2^{1-2/v}\log n \\ &\geq c_{2,3}C_1^2\log n, \end{aligned}$$

when $C_2 \geq 1$. Hence, if C_1 is large enough,

$$\mathbb{P}(\text{Bin}(n, p_1) \leq \kappa) \leq 1/n^3.$$

We conclude with the union bound that $\mathbb{P}(\mathcal{A}_2^+) \geq 1 - n/n^3 = 1 - 1/n^2$.

Step 2.2: $\mathcal{A}_2^- := \{r_i \geq r - 2\varepsilon, \forall i \in [n]\}$ happens with probability at least $1 - 1/n^2$.

For $i \in [n]$, we have $p(x_i, r - 2\varepsilon) \leq p_2 := (r - 2\varepsilon)^v/\omega$, so that

$$\begin{aligned} \mathbb{P}(r_i < r - 2\varepsilon) &= \mathbb{P}(N(x_i, p(x_i, r - 2\varepsilon)) \geq \kappa) \\ &\leq \mathbb{P}(\text{Bin}(n, p_2) \geq \kappa). \end{aligned}$$

Since $p_o = \kappa/n$ and $p_o \geq p_2$, we may apply Lemma 2 to obtain

$$\begin{aligned} \mathbb{P}(\text{Bin}(n, p_2) \geq \kappa) &\leq 2 \exp\left(-\frac{n(p_o - p_2)^2}{2p_2(1 - p_2) + 2(p_o - p_2)/3}\right) \\ &\leq 2 \exp\left(-\frac{n(p_o - p_2)^2}{3p_o}\right). \end{aligned}$$

The rest of the calculations are otherwise the same as in Step 2.1 above, in particular since

$$p_o - p_2 = (r_o^v - (r_o - \varepsilon)^v)/\omega \geq c_{2,3}r_o^{v-1}\varepsilon, \quad (84)$$

when r_o/ε is sufficiently large, which happens when C_2 is large enough.

Step 3: (15) holds under $\mathcal{A}_1 \cap \mathcal{A}_2^-$.

The arguments backing this claim are exactly the same as those leading to (21). To be sure, we reproduce them here. We assume that C_2 is large enough that $r/\varepsilon \geq 8$. Thus fix $i, j \in [n]$ distinct. Let $m := \lfloor d_{ij}/(r - 4\varepsilon) \rfloor$, and for $s \in \{0, \dots, m+1\}$, let $z_s = x_i + \frac{s}{m+1}(x_j - x_i)$. We have $z_0 = x_i$ and $z_{m+1} = x_j$, and z_0, z_1, \dots, z_{m+1} are on the line joining x_i and x_j and satisfy $\|z_s - z_{s+1}\| \leq r - 4\varepsilon$ for all s . Let x_{i_s} be such that $\|z_s - x_{i_s}\| \leq \varepsilon$, with $x_{i_0} = x_i$ and $x_{i_{m+1}} = x_j$.

By the triangle inequality, for all $s \in \{0, \dots, m\}$,

$$\begin{aligned} \|x_{i_s} - x_{i_{s+1}}\| &\leq \|x_{i_s} - z_s\| + \|z_s - z_{s+1}\| + \|z_{s+1} - x_{i_{s+1}}\| \\ &\leq \varepsilon + (r - 4\varepsilon) + \varepsilon = r - 2\varepsilon \leq r_{i_s}. \end{aligned}$$

Hence, $(x_{i_0}, x_{i_1}, \dots, x_{i_{m+1}})$ forms a path in the graph, and as a consequence, $\delta_{ij} \leq m + 1$. In turn, this implies that

$$\hat{d}_{ij} = r\delta_{ij} \leq rm + r \leq r\frac{d_{ij}}{r - 4\varepsilon} + r \leq d_{ij} + 8(\varepsilon/r)d_{ij} + r. \quad (85)$$

Step 4.1: (16) holds under \mathcal{A}_2^+ .

Fix distinct $i, j \in \mathcal{J}_d$, where $d \geq r$ is left unspecified for now. Let $k_0 = i, \dots, k_\ell = j$ denote a shortest path in the graph joining i and j , so that $\ell = \delta_{ij}$. If $k_s \in \mathcal{J}_d$ for all s , then because $\mathcal{J}_d \subset \mathcal{J}_r$ and \mathcal{A}_2^+ holds, we have $\|x_{k_{s-1}} - x_{k_s}\| \leq r_{k_{s-1}} \leq r$, and this implies that

$$d_{ij} = \|x_{k_0} - x_{k_\ell}\| \leq \sum_{s=0}^{\ell-1} \|x_{k_s} - x_{k_{s+1}}\| \leq \ell r = r \delta_{ij} = \hat{d}_{ij}. \quad (86)$$

Otherwise, let $s_- = \min\{s : k_s \notin \mathcal{J}_d\}$ and $s^+ = \max\{s : k_s \notin \mathcal{J}_d\}$. For $s \leq s_-$, $k_{s-1} \in \mathcal{J}_d$ and therefore $\|x_{k_{s-1}} - x_{k_s}\| \leq r_{k_{s-1}} \leq r$, and because $\|x_i - x_{k_{s_-}}\| \geq d$, necessarily $s_- \geq d/r$. Similarly, for $s \geq s^+ + 1$, $k_s \in \mathcal{J}_d$ and therefore $\|x_{k_{s-1}} - x_{k_s}\| \leq r_{k_s} \leq r$, and because $\|x_j - x_{k_{s^+}}\| \geq d$, necessarily $\ell - s^+ \geq d/r$. Hence, $\delta_{ij} \geq s_- + \ell - s^+ \geq 2d/r$. Hence, if we choose $d = d_{ij}/2$, we obtain the desired result.

Step 4.2: when $v = 1$, $\hat{d}_{ij} \geq d_{ij}$ for all $i, j \in [n]$ under $\mathcal{A}_1 \cap \mathcal{A}_2^+$.

Rescaling if needed, we may assume without loss of generality that $\Omega = [0, 1]$. Relabel the points in such a way that $x_1 < \dots < x_n$. Fix $i < j$ and let $k_0 = i, k_1, \dots, k_\ell = j$ be a shortest path joining i and j , so that $k_0 < k_1 < \dots < k_\ell$ by Lemma 3. If $x_j \leq r$ or $x_i \geq 1 - r$, then $d_{ij} \leq r$, and we conclude with the fact that $\hat{d}_{ij} = r \delta_{ij} \geq r$. If $x_i \geq r$ and $x_j \leq 1 - r$, then $x_{k_s} \in [r, 1 - r]$ for all s , implying that $k_s \in \mathcal{J}_r$ for all s , implying that $\|x_{k_s} - x_{k_{s+1}}\| \leq r$ by \mathcal{A}_2^+ , and thus, by the triangle inequality, we have (86). Otherwise, let $s_- = \min\{s : x_{k_s} \geq r\}$ and $s^+ = \max\{s : x_{k_s} \leq r\}$, and let $i' = k_{s_-}$ and $j' = k_{s^+}$. Note that $\delta_{i'j'} \geq d_{i'j'}/r$, by the fact that $x_{i'} \geq r$ and $x_{j'} \leq 1 - r$. If $x_i < r$ and $r < x_j \leq 1 - r$, so that $i' > i$ and $j' = j$, then

$$\delta_{ij} \geq 1 + \delta_{i'j} \geq 1 + d_{i'j}/r \geq 1 + (d_{ij} - r)/r = d_{ij}/r, \quad (87)$$

using the triangle inequality at the end. The case where $r \leq x_i < 1 - r$ and $x_j > 1 - r$ can be dealt with similarly. For the remaining case where $x_i < r$ and $x_j > 1 - r$, we have

$$\delta_{ij} \geq 2 + \delta_{i'j'} \geq 2 + d_{i'j'}/r \geq 2 + (d_{ij} - 2r)/r = d_{ij}/r, \quad (88)$$

using the triangle inequality at the end. \square

7.5 Proof of Proposition 3

We use the notation of Theorem 4 and its proof, except that new constants introduced here are denoted by q_1, q_2 , etc. The constant C_1 from Theorem 4 remains the same (specialized to the present choice of Ω), while C_2 may be chosen even larger (as large as needed).

Step I: there is $q_1 > 1$ such that $\mathcal{A}_3 := \{r_i \geq q_1 r, \forall i \in \mathcal{J}_\varepsilon^c\}$ happens with probability at least $1 - 1/n^2$.

For any $x \in \Omega$ such that $\text{dist}(x, \partial\Omega) \leq t/2$, for some $t \leq 1/2$, we have $p(x, t) \leq p_t := q_{1,1} t^2 / \omega$, for some $q_{1,1} < 1$, stemming from the fact that for such an x , $\mathbb{B}(x, t) \not\subset \Omega$. Take $i \in \mathcal{J}_\varepsilon^c$, so that $\text{dist}(x_i, \partial\Omega) \leq \varepsilon$. For $2\varepsilon \leq t \leq 1/2$,

$$\begin{aligned} \mathbb{P}(r_i < t) &= \mathbb{P}(N(x_i, p(x_i, t)) \geq \kappa) \\ &\leq \mathbb{P}(\text{Bin}(n, p_t) \geq \kappa). \end{aligned}$$

Fix $q_{1,2} \in (q_{1,1}, 1)$ and define $q_{1,3} = (q_{1,2}/q_{1,1})^{1/2}$, so that $q_{1,3} > 1$, and if $t = q_{1,3} r_\circ$, then $p_t = q_{1,2} p_\circ$. As before, $p_\circ = \kappa/n$, and because $p_t \leq p_\circ$, we may apply Lemma 2 to get

$$\begin{aligned} \mathbb{P}(\text{Bin}(n, p_t) \geq \kappa) &\leq 2 \exp\left(-\frac{n(p_\circ - p_t)^2}{2p_t(1 - p_t) + (p_\circ - p_t)/3n}\right) \\ &\leq 2 \exp\left(-\frac{n(p_\circ - p_t)^2}{3p_\circ}\right) \\ &\leq 2 \exp(-q_{1,4} n p_\circ). \end{aligned}$$

We then conclude with the union bound and the fact that $np_o = \kappa \geq C_2 \log n$, where C_2 may be chosen arbitrarily large, together with the fact that $r = r_o + \varepsilon \leq r_o$ when $\varepsilon/r_o \leq 1$, which is also guaranteed by choosing C_2 large enough.

Step II: (17) holds under $\mathcal{A}_1 \cap \mathcal{A}_2^- \cap \mathcal{A}_3$ hold.

Consider $i, i' \in [n]$ such that $d_{ii'} \geq 2$. Let z and z' be the orthogonal projections of x_i and $x_{i'}$ onto the horizontal axis. Under \mathcal{A}_1 , there are $j, j' \in [n]$ such that

$$\|x_j - z\| \leq \varepsilon, \quad \|x_{j'} - z'\| \leq \varepsilon. \quad (89)$$

By the triangle inequality,

$$\hat{d}_{ii'} \leq \hat{d}_{ij} + \hat{d}_{jj'} + \hat{d}_{j'i'}. \quad (90)$$

We saw in (85) that, under $\mathcal{A}_1 \cap \mathcal{A}_2^-$, we have

$$\begin{aligned} \hat{d}_{ij} &\leq d_{ij} + 8(\varepsilon/r)d_{ij} + r, \\ \hat{d}_{i'j'} &\leq d_{i'j'} + 8(\varepsilon/r)d_{i'j'} + r. \end{aligned}$$

Remains to bound $\hat{d}_{jj'}$ from above. Let $m := \lfloor \|z' - z\|/(q_1 r - 2\varepsilon) \rfloor$, and for $s \in \{0, \dots, m+1\}$, let $z_s = z + \frac{s}{m+1}(z' - z)$. We have $z_0 = z$ and $z_{m+1} = z'$, and $\|z_s - z_{s+1}\| \leq q_1 r - 2\varepsilon$ for all s . Under \mathcal{A}_1 , for each $s \in [m]$, there is x_{i_s} be such that $\|z_s - x_{i_s}\| \leq \varepsilon$. We take $x_{i_0} = x_j$ and $x_{i_{m+1}} = x_{j'}$. Note that $i_s \in \mathcal{J}_\varepsilon^c$ for all s , due to the fact that $z_s \in \partial\Omega$ for all s . By the triangle inequality, for all $s \in \{0, \dots, m\}$,

$$\begin{aligned} \|x_{i_s} - x_{i_{s+1}}\| &\leq \|x_{i_s} - z_s\| + \|z_s - z_{s+1}\| + \|z_{s+1} - x_{i_{s+1}}\| \\ &\leq \varepsilon + (q_1 r - 2\varepsilon) + \varepsilon \leq q_1 r \leq r_{i_s}, \end{aligned}$$

where the last inequality is due to \mathcal{A}_3 holding. Hence, $(x_{i_0}, x_{i_1}, \dots, x_{i_{m+1}})$ forms a path in the graph, and as a consequence, $\delta_{jj'} \leq m+1$. In turn, this implies that

$$\hat{d}_{jj'} = r\delta_{jj'} \leq rm + r \leq r \frac{d_{jj'}}{q_1 r - 2\varepsilon} + r \leq d_{jj'}/q_1 + q_{2,1}(\varepsilon/r)d_{jj'} + r. \quad (91)$$

Therefore,

$$\begin{aligned} \hat{d}_{ii'} &\leq d_{ij} + 8(\varepsilon/r)d_{ij} + r \\ &\quad + d_{jj'}/q_1 + q_{2,1}(\varepsilon/r)d_{jj'} + r \\ &\quad + d_{i'j'} + 8(\varepsilon/r)d_{i'j'} + r \\ &\leq d_{ii'} - q_{2,3}d_{jj'} + q_{2,2}(\varepsilon/r)d_{ii'} + 3r, \end{aligned}$$

with $q_{2,3} := 1 - 1/q_1 > 0$. Let $t = \|z - z'\|$, and note that $d_{ii'} \leq t+1$ and $d_{jj'} \geq t - 2\varepsilon$, by the triangle inequality, so that $t \geq d_{ii'}/2$ when $d_{ii'} \geq 2$. We thus have

$$\hat{d}_{ii'} \leq d_{ii'} - q_{2,3}d_{ii'}/2 + q_{2,4}[(\varepsilon/r)d_{ii'} + r + \varepsilon] \leq (1 - q_{2,5})d_{ii'}, \quad (92)$$

for some $q_{2,5} > 0$.

References

- [1] E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, 9:1981–2014, 2008.
- [2] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade. A tensor spectral approach to learning mixed membership community models. *arXiv preprint arXiv:1302.2684*, 2013.
- [3] M. Bernstein, V. De Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Technical report, Department of Psychology, Stanford University, 2000.
- [4] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM*, 60(6):41, 2013.
- [5] L. Chen and A. Buja. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association*, 104(485):209–219, 2009.
- [6] Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.
- [7] M. Cucuringu. Asap: An eigenvector synchronization algorithm for the graph realization problem. In *Distance Geometry*, pages 177–195. Springer, 2013.
- [8] M. Cucuringu and J. Woodworth. Ordinal embedding of unweighted knn graphs via synchronization. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*, pages 1–6, 2015.
- [9] M. Cucuringu, Y. Lipman, and A. Singer. Sensor network localization by eigenvector synchronization over the euclidean group. *ACM Transactions on Sensor Networks*, 8(3):19, 2012.
- [10] J. de Leeuw and P. Mair. Multidimensional scaling using majorization: Smacof in r. *Journal of Statistical Software*, 31(i03), 2009.
- [11] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.
- [12] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical association*, 97(460):1090–1098, 2002.
- [13] P. W. Holland, K. Blackmond-Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.
- [14] L. Jain, K. G. Jamieson, and R. Nowak. Finite sample prediction and recovery bounds for ordinal embedding. In *Advances in Neural Information Processing Systems*, pages 2711–2719, 2016.
- [15] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27, 1964.
- [16] J. B. Kruskal and J. B. Seery. Designing network diagrams. In *Proceedings of the 1st General Conference on Social Graphics*, pages 22–50, 1980.
- [17] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.
- [18] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. In *Conference on Information and Knowledge Management (CIKM’03)*, pages 556–559, 2003.
- [19] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the Association for Information Science and Technology*, 58(7):1019–1031, 2007.
- [20] K. Mosler and M. Scarsini. Some theory of stochastic dominance. *Lecture Notes-Monograph Series*, pages 261–284, 1991.
- [21] E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. In

- Proceedings of the 47th Annual ACM Symposium on the Theory of Computing*, pages 69–75, 2015.
- [22] A. Paprotny and J. Garcke. On a connection between maximum variance unfolding, shortest path problems and isomap. In *Artificial Intelligence and Statistics*, pages 859–867, 2012.
- [23] S. Parthasarathy, D. Sivakoff, M. Tian, and Y. Wang. A quest to unravel the metric structure behind perturbed networks. *arXiv preprint arXiv:1703.05475*, 2017.
- [24] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.
- [25] P. Sarkar, D. Chakrabarti, and A. W. Moore. Theoretical justification of popular link prediction heuristics. In *International Conference on Learning Theory*, 2010.
- [26] E. R. Scheinerman and K. Tucker. Modeling graphs using dot product representations. *Computational Statistics*, 25(1):1–16, 2010.
- [27] Y. Shang and W. Ruml. Improved mds-based localization. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 4, pages 2640–2651, 2004.
- [28] Y. Shang, W. Ruml, Y. Zhang, and M. P. Fromherz. Localization from mere connectivity. In *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 201–212, 2003.
- [29] R. N. Shepard. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27:125–140, 1962.
- [30] R. N. Shepard. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27:219–246, 1962.
- [31] D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.
- [32] D. L. Sussman, M. Tang, and C. E. Priebe. Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):48–57, 2014.
- [33] M. Tang, D. L. Sussman, and C. E. Priebe. Universally consistent vertex classification for latent positions graphs. *The Annals of Statistics*, 41(3):1406–1430, 2013.
- [34] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [35] Y. Terada and U. von Luxburg. Local ordinal embedding. In *Proceedings of the 31st International Conference on Machine Learning*, pages 847–855, 2014.
- [36] U. von Luxburg and M. Alamgir. Density estimation from unweighted k-nearest neighbor graphs: a roadmap. In *Advances in Neural Information Processing Systems*, pages 225–233, 2013.
- [37] K. Q. Weinberger, F. Sha, Q. Zhu, and L. K. Saul. Graph laplacian regularization for large-scale semidefinite programming. In *Advances in Neural Information Processing Systems*, pages 1489–1496, 2006.
- [38] J. Yang and J. Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, pages 587–596, 2013.
- [39] F. W. Young and R. M. E. Hamer. *Multidimensional scaling: history, theory, and applications*. Lawrence Erlbaum Associates, Inc, 1987.
- [40] S. J. Young and E. R. Scheinerman. Random dot product graph models for social networks. In *Algorithms and models for the web-graph*, pages 138–149. Springer, 2007.