BRIDGE DETECTION AND ROBUST GEODESICS ESTIMATION VIA RANDOM WALKS

Eugene Brevdo and Peter J. Ramadge

Dept. of Electrical Engineering, Princeton University, Princeton NJ 08544

ABSTRACT

We propose an algorithm for detecting bridges and estimating geodesic distances from a set of noisy samples of an underlying manifold. Finding geodesics on a nearest neighbors graph is known to fail in the presence of bridges. Our method detects bridges using global statistics via a Markov random walk and denoises the nearest neighbors graph using "surrogate" weights. We show experimentally that our method outperforms methods based on local neighborhood statistics.

Index Terms— Unsupervised learning, Diffusion processes, Multidimensional signal processing

1. INTRODUCTION

We are interested in situations where accurate geodesic estimation from data sampled from a manifold is important. For example, estimating feature space distances when the data lies on a manifold can be an important step in unsupervised [1] and semi-supervised [2] learning. In the simplest approach, a nearest neighbors (NN) graph is constructed from either k nearest neighbors, or a δ-ball, at each point. Each graph edge is weighted by the distance between its nodes. A graph shortest path (SP) algorithm, e.g. Dijkstra's, is then used to estimate geodesic distances between pairs of points.

When the manifold samples are noisy or contain outliers, bridges (short circuits between distant parts of the manifold) appear in the NN graph and this has a catastrophic effect [3]. The greedy nature of the SP solution encourages the traversal of bridges, thereby significantly underestimating geodesic distances. Previous work has considered denoising the nearest neighbors graph via rejection of edges based on local distance statistics [4, 5], or via local tangent space estimation [6, 7]. However, unlike the method proposed here, these methods use local rather that global statistics. We have found that using only local statistics can be unreliable. For example, with state of the art robust estimators of the local tangent space ([8]), local rejection of neighborhood edges is not reliable with moderate noise or outliers. Furthermore, edge removal (pruning) based on local edge length statistics is based on questionable assumptions. For example, a thin chain of outliers can form a bridge without unusually long edge lengths.

The Diffusion Maps (DM) approach to estimating feature space distances [9], has experimentally exhibited robustness to noise, outliers, and finite sampling. Diffusion distances, based on random walks on the NN graph, are closely related to "natural" distances (*commute times*) on a manifold [10]. Furthermore, DM coordinates (based on these distances) converge to eigenfunctions of the Laplace Beltrami operator on the underlying manifold. However, a global relationship between geodesic and diffusion distances is unknown.

Our key contribution is the formulation of a global measure of edge reliability based on diffusion distances. We use this measure as the basis of a bridge detector. This detector is then used to inform geodesic estimates. To do so, instead of pruning the NN graph we use surrogate distances when estimating shortest paths. This has the advantage of preserving all sample points and avoiding disconnecting under-sampled sections of the NN graph. We relate the edge reliability measure, as the sample size tends to infinity, to the Laplace Beltrami operator on the underlying manifold. Finally, we verify experimentally the success of the approach against competing methods of bridge detection and geodesic estimation.

The remainder of the paper is organized as follows. §2 introduces notation and describes the basic framework of surrogate distances in geodesics estimation. We also describe methods for bridge detection that will be used as benchmarks. §3 covers our main contribution: a novel method for bridge detection based on random walks. We show its improved performance experimentally in §4.

2. PRELIMINARIES

Let $\mathcal{X} = \{x_i\}_{i=1}^n$ be nonuniformly sampled points from manifold $\mathcal{M}$ embedded in $\mathbb{R}^N$. We observe $\mathcal{Y} = \{y_i = x_i + \nu_i\}_{i=1}^n$, where ν_i is noise. A nearest neighbor (NN) graph $G = (\mathcal{Y}, \mathcal{E}, d)$ is constructed from k-NN or δ-ball neighborhoods of $\mathcal{Y}$ with the scale (k or δ) chosen via cross-validation or prior knowledge. Map $d : \mathcal{E} \to \mathbb{R}$ assigns the cost $d_e = \|x_k - x_l\|_2$ to edge $e = (k, l) \in \mathcal{E}$. Let $\mathcal{D} = \{d_e : e \in \mathcal{E}\}$. The set $\mathcal{E}$ gives initial estimates of neighbors on the manifold. Let $\mathcal{F}_k$ denote the neighbors of x_k in G.

In [1] the geodesic distance between $(i, j) \in \mathcal{Y}^2$ is estimated by $\hat{g}_{ij} = \sum_{e \in \mathcal{P}_{ij}} d_e$ where $\mathcal{P}_{ij}$ is a minimum cost path from i to j in G. When there is no noise, this estimate converges to the true geodesic distance on $\mathcal{M}$ as $n \uparrow \infty$ and neighborhood size $\delta \downarrow 0$. However, in the presence of noise bridges form in the NN graph and this results in significant

estimation error. Forming the shortest path in G is too greedy in the presence of bridges.

If bridges could be detected, their anomalous effect could be removed without disconnecting the graph by substituting a surrogate weight: $\tilde{d}_e = d_e + M, e \in \mathcal{B}$ where $\mathcal{B}$ is the set of detected bridges and $M = n(\max_{e\in\mathcal{E}} d_e)$, larger than the diameter of G, is a penalty. Let $\tilde{G} = (\mathcal{Y}, \mathcal{E}, \tilde{d})$ and $\tilde{\mathcal{P}}_{ij}$ be a minimum cost path between i and j in $\tilde{G}$. The adjusted estimate of geodesic distance is $\tilde{g}_{ij} = \sum_{e\in\tilde{\mathcal{P}}_{ij}} d_e$. With this in mind, we first review some bridge detection methods.

The simplest bridge decision rule (DR) classifies bridges by a threshold on edge length. First, normalize edge lengths for local sampling density by setting $\bar{d}_{kl} = d_{kl}/\sqrt{d_{k\cdot}.d_{l\cdot}}$ where $d_{k\cdot} = \sum_{m\in\mathcal{F}_k} d_{km}$ sums outgoing edge lengths from x_k. Let $\bar{\mathcal{D}} = \{\bar{d}_e : e \in \mathcal{E}\}$. Select a "good edge percentage" $0 < q < 1$ (e.g. 99%) and let $\mathcal{B} = \{e \in \mathcal{E} : \bar{d}_e \geq Q(\bar{\mathcal{D}}, q)\}$, where $Q(\mathcal{D}, q)$ is the q-th quantile of the set $\mathcal{D}$. We call this the length decision rule (LDR); it is similar to the DR of [4].

The Jaccard similarity DR (JDR) classifies bridges as edges between points with dissimilar neighborhoods [5]. Let $j_{lm} = |\mathcal{F}_l \cap \mathcal{F}_m|/|\mathcal{F}_l \cup \mathcal{F}_m|$ be the Jaccard similarity of the neighborhoods of x_l and x_m. Let $\mathcal{J} = \{j_e : e \in \mathcal{E}\}$ be the set of Jaccard similarities. Again select a $q \in (0, 1)$; the Jaccard DR sets $\mathcal{B} = \{e \in \mathcal{E} : j_e < Q(\mathcal{J}, 1 - q)\}$.

Bridges are short cuts for shortest paths. This suggests detecting bridges by counting the traversals of each edge by estimated geodesics. This is the concept of edge centrality (*edge betweenness*) in networks [11]. In a network, the centrality of edge e is $BE(e) = \sum_{(i,j)\in\mathcal{Y}^2} \mathbb{I}(e \in \mathcal{P}_{ij})$. Edge centrality can be calculated in $O(n^2 \log n)$ time and $O(n + |\mathcal{E}|)$ space [11, 12]. However, caution is required in using edge centrality for our purpose. Consider a bridge $(a, b) \in \mathcal{E}$ having high centrality. Suppose there exists a point y_c with $(a, c), (c, b) \in \mathcal{E}$ such that $d_{ac} + d_{cb} < d_{ab} + \delta'$, δ' small. These edges are never preferred over (a, b) in a geodesic path, hence have low centrality. However, once (a, b) is placed into $\mathcal{B}$ and given increased weight, $(a, c), (c, b)$ reveals itself as a secondary bridge in $\tilde{G}$. So detection by centrality should be done in rounds, each adding edges to $\mathcal{B}$. This allows secondary bridges to be detected in subsequent rounds. The Edge Centrality Decision Rule (ECDR) iterates K times on $\tilde{G}$, each time placing $(1-q)n/K$ of the most central edges into $\mathcal{B}$ and updating $\tilde{G}$. Here q and K are parameters. To our knowledge use of centrality as a bridge detector is new. While an improvement over LDR, the deterministic greedy underpinnings of ECDR are a limitation: it initially fails to see secondary bridges, and may also misclassify true high centrality edges as bridges, e.g. the narrow path in the dumbbell manifold [9].

3. NEIGHBOR PROBABILITY DECISION RULE

We now propose a DR based on a Markov random walk that assigns a probability that two points are neighbors. This steps back from immediately looking for a shortest path and instead lets a random walk "explore" the NN graph. To this end, let P be a row-stochastic matrix with $P_{ij} = 0$ if $i \neq j$ and $(i, j) \notin \mathcal{E}$. Let $p \in (0, 1)$ be a parameter and $\bar{p} = 1 - p$. Consider a random walk $s(t)$, $t \geq 0$, on G starting at $s(0) = i$ and governed by P. For each $t \geq 0$, with probability p we stop the walk and declare $s(t)$ a neighbor of i. Let N be matrix of probabilities that the walk starting at node i, stops at node j. The i-th row of N is the neighbor distribution of node i. Since stopping at each time t is mutually exclusive:

$$N = \sum_{t\geq 0} P^t p(1-p)^t = p(I - \bar{p}P)^{-1}$$

A smaller stopping probability p induces greater weighting of long-time diffusion effects, which are more dependent on the topology of $\mathcal{M}$. N is closely related to the *normalized commute time* of [13]. Its computation requires $O(n^3)$ time and $O(n^2)$ space (P is sparse but N may not be).

Lemma 3.1. *N is row-stochastic, shares the left and right eigenvectors of P, and has spectrum $\sigma(N) = \{p(1 - \bar{p}\lambda)^{-1}, \lambda \in \sigma(P)\}$.*

Proof. Clearly, $N^{-1} = (I - \bar{p}P)/p$ is row-stochastic, shares the left and right eigenvectors of P and has spectrum $\{(1 - \bar{p}\lambda)/p, \lambda \in \sigma(P)\}$. □

Following [9], we select P to be the popular Diffusion Maps kernel on G. Choose $\epsilon > 0$ (e.g. $\epsilon = \text{med}_{e\in\mathcal{E}}\, d_e/2$) and let $(\hat{P}_\epsilon)_{lk} = \exp\left(-d_{kl}^2/\epsilon\right)$ if $(l, k) \in \mathcal{E}$, 1 if $l = k$, and 0 otherwise. Then let $\hat{D}_\epsilon$ be diagonal with $(\hat{D}_\epsilon)_{kk} = \sum_l (\hat{P}_\epsilon)_{kl}$, and normalize for sampling density by setting $A_\epsilon = \hat{D}_\epsilon^{-1}\hat{P}_\epsilon\hat{D}_\epsilon^{-1}$. Finally, let D_ϵ be diagonal with $(D_\epsilon)_{kk} = \sum_l (A_\epsilon)_{kl}$ and set $P_\epsilon = {D_\epsilon}^{-1}A_\epsilon$. P_ϵ is row-stochastic and, as is well known, has bounded spectrum:

Lemma 3.2. *The spectrum of P_ϵ is contained in $[0, 1]$.*

Proof. $\hat{P}_\epsilon$, A_ϵ and $S = {D_\epsilon}^{-1/2}A_\epsilon{D_\epsilon}^{-1/2}$, are symmetric PSD. $P_\epsilon = {D_\epsilon}^{-1}A_\epsilon$ is row-stochastic and similar to S. □

Define the Neighbor Probability Decision Rule (NPDR) as follows. Let $N_\epsilon = p(I - \bar{p}P_\epsilon)^{-1}$. Find the restriction $\mathcal{N} = \{(N_\epsilon)_e : e \in \mathcal{E}\}$ of N_ϵ to $\mathcal{E}$. Choose a $q \in (0, 1)$ and let $\mathcal{B} = \{e \in \mathcal{E} : (N_\epsilon)_e < Q(\mathcal{N}, 1 - q)\}$, i.e., edges connecting nodes with a low probability of being neighbors, are bridges.

Calculating N_ϵ can be prohibitive for very large datasets. Fortunately, we can effectively order the elements of $\mathcal{N}$ using a low rank approximation to N_ϵ. By Lemmas 3.1, 3.2, we calculate $N_{lm} \approx \sum_{j=1}^{J} p(1 - \bar{p}\lambda_j)^{-1}(v_j^R)_l(v_j^L)_m$ for $(l, m) \in \mathcal{E}$, where λ_j,v_j^R and v_j^L are the J largest eigenvalues of P_ϵ and the associated right and left eigenvectors. In practice, an effective ordering of $\mathcal{N}$ is obtained with $J \ll n$ and since P_ϵ is sparse, its largest eigenvalues and eigenvectors can be computed efficiently using a standard iterative algorithm.

An important property of P_ϵ is that when $n\uparrow\infty$, as $\epsilon\downarrow 0$, $(I - P_\epsilon)/\epsilon \to c\Delta_{LB}$ both pointwise and in spectrum [9].

Here Δ_{LB} is the Laplace Beltrami operator on $\mathcal{M}$ and c is a constant. Thus, for large n and small ϵ, P_ϵ is a neighborhood averaging operator. This property also holds for N_ϵ.

Theorem 3.3. *As $n\uparrow\infty$ and $\epsilon\downarrow 0$, $(I - N_\epsilon)/\epsilon \to c'\Delta_{LB}$.*

Proof. For $I - S$ invertible, $(I - S)^{-1} = I + S(I - S)^{-1}$. From Lem. 3.2, $I - \bar{p}P_\epsilon$ is invertible. Thus $p(I - \bar{p}P_\epsilon)^{-1} = (I - \frac{\bar{p}}{p}(P_\epsilon - I))^{-1} = I + \frac{\bar{p}}{p}(P_\epsilon - I)(I - \frac{\bar{p}}{p}(P_\epsilon - I))^{-1}$. Therefore

$$\frac{I - N_\epsilon}{\epsilon} = \frac{\bar{p}}{p}\frac{I - P_\epsilon}{\epsilon}(I - \frac{\bar{p}}{p}(P_\epsilon - I))^{-1} = \bar{p}\frac{I - P_\epsilon}{\epsilon}(I - \bar{p}P_\epsilon)^{-1}.$$

The first factor on the RHS converges to $\bar{p}\Delta_{LB}$ and the second to $p^{-1}I$ since $P_\epsilon \to I$. □

By Theorem 3.3, as $n\uparrow\infty$ and $\epsilon\downarrow 0$, N_ϵ acts like P_ϵ. However, experiments indicate that N_ϵ is more informative of neighborhood relationships. For example, using P_ϵ instead of N_ϵ in the NPDR yields performance very similar to LDR.

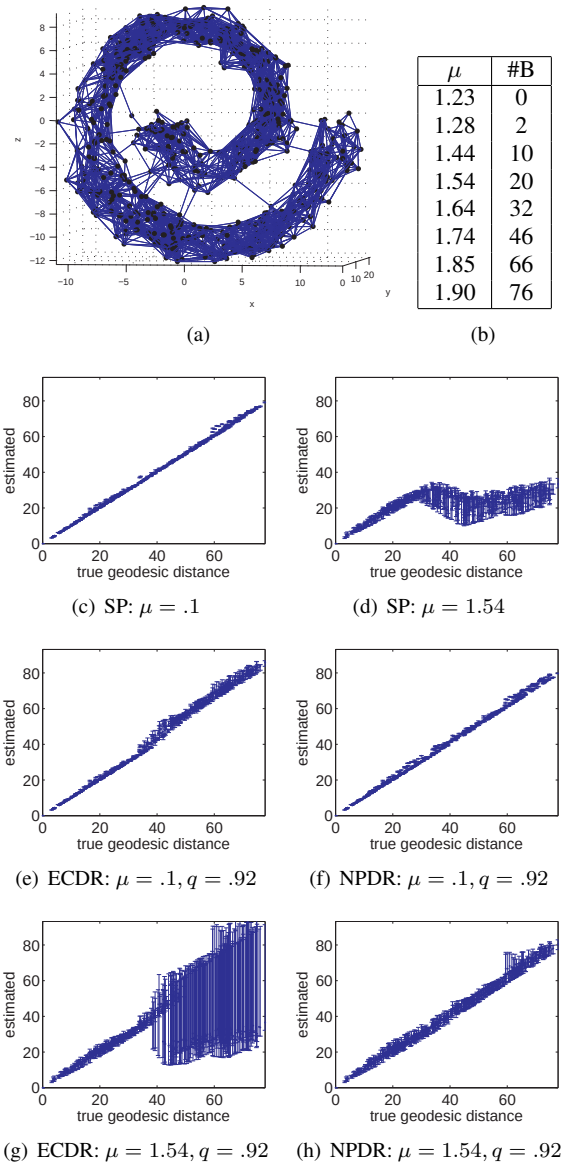
4. EXPERIMENTS

We first test our method on the synthetic swiss roll, parameterized by (a, b) in $U = [\pi, 4\pi]\times[0, 21]$ via the embedding $(x^1, x^2, x^3) = f(a, b) = (a\cos a, b, a\sin a)$. True geodesic distances $\{g_{1j}\}_{j=1}^n$ are computed via:

$$g_{ij} = \int_0^1 \|Df(\boldsymbol{v}(t))\frac{\partial \boldsymbol{v}(t)}{\partial t}\|_2 dt$$

where $\boldsymbol{v}(t) = (1 - t)[a_i \;\; b_i]^T + t[a_j \;\; b_j]^T$, and $Df(\boldsymbol{v})$ is the differential of f evaluated at $(\boldsymbol{v}^1, \boldsymbol{v}^2)$. We sampled $n = 500$ points uniformly in the ambient space $\mathbb{R}^3$ with x_1 fixed ($(a_1, b_1) = (\pi, 0)$). For $t = 1, \ldots, T$ ($T = 100$) we generated n random noise values, $\{u_{ti}\}_{i=1}^n$, uniformly on $[-1, 1]$. Then $y_{ti} = x_i + \mu u_{ti}\boldsymbol{n}_i$, $t = 1, \ldots, T$, where $\boldsymbol{n}_i$ is the normal to $\mathcal{M}$ at x_i. Each experiment was repeated for $\mu \in \{0, .05, \cdots, 1.95, 2\}$. The initial NN graph G was constructed using δ-balls ($\delta = 4$). The median bridge counts over T realizations are shown in Fig. 1(b). Bridges first appear at $\mu \approx 1.2$. Fig. 1(a) shows one realization of $\mathcal{Y}$ and the NN graph G (note bridges). We compare the simple SP with the LDR, ECDR ($K = 15$), and NPDR ($p = 0.01$) -based estimators by plotting the estimates of geodesic distance versus ground truth (sorted by distance from x_1). We plot the median estimate, 33%, and 66% quantiles over the T runs, for $\mu = .1$ and $\mu = 1.54$. The LDR and JDR based estimators' performance is comparable to SP for $q > .9$ (plots not included).

With no noise: SP provides excellent estimates; NPDR estimates are accurate even after removing 8% of the graph edges (Fig. 1(f)); however, ECDR removes important edges (Fig. 1(e)). At $\mu = 1.54$ with approximately 25 bridges: SP has failed (Fig. 1(d)); ECDR is removing bridges but also important edges, resulting in an upward estimation bias (Fig. 1(g)); in contrast, NPDR is successfully discounting



(a)

μ	#B
1.23	0
1.28	2
1.44	10
1.54	20
1.64	32
1.74	46
1.85	66
1.90	76

(b)

(c) SP: $\mu = .1$ (d) SP: $\mu = 1.54$

(e) ECDR: $\mu = .1, q = .92$ (f) NPDR: $\mu = .1, q = .92$

(g) ECDR: $\mu = 1.54, q = .92$ (h) NPDR: $\mu = 1.54, q = .92$

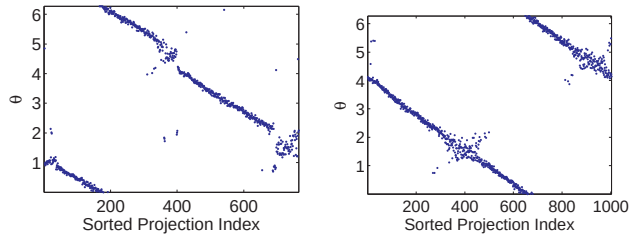
Fig. 1: (a) NN graph of Swiss roll (μ=1.6). (b) Median bridge count vs. μ. (c)-(h) Geodesic estimates vs. ground truth (from x_1).

bridges without any significant upward bias even at $q = .92$ (Fig. 1(h)). This supports our claim that bridges occur between edges with low neighbor probability in the NPDR random walk. Lower values of q remove more edges, including bridges, but removing non-bridges always increases SP estimates, and can lead to an upward bias. The choice of q should be based on prior knowledge of the noise or cross-validation.

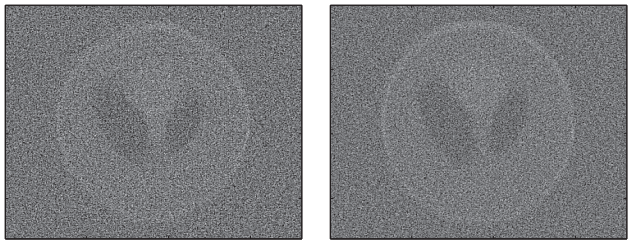
Fig. 2 compares the performance of DRs at moderate noise levels. The error is the mean over the trials of $E = \frac{1}{5n}\sum_{r=1}^5\sum_{i=1}^n|g_{ri} - \tilde{g}_{ri}|$, where the points x_r are well distributed over $\mathcal{M}$. NPDR outperforms the other DRs at moderate noise levels, given the appropriate choice of q.

μ	SP	LDR, q=.92	ECDR, q= .92	.95	.99	NPDR, q= .92	.95	.99
0.10	**0.8**	1.5	4.4	3.8	1.9	2.4	1.9	1.2
1.44	12.0	10.9	10.7	8.1	2.1	3.6	2.5	**1.6**
1.54	13.6	12.8	11.7	7.6	2.5	3.8	2.6	**2.3**
1.64	14.4	13.9	11.6	6.8	5.1	4.0	**3.1**	6.5
1.74	14.9	14.4	11.4	6.5	8.9	**5.0**	5.6	11.1
1.85	15.3	14.9	12.1	8.6	12.0	**8.1**	9.4	13.4

Fig. 2: Comparison of mean error E, varying μ.



(a) Estimated θ, JDR (b) Estimated θ, NPDR



(c) Reconstruction $\hat{I}$, JDR (d) Reconstruction $\hat{I}$, NPDR

Fig. 3: Tomography Reconstructions from Random Projections

We now consider the random projection tomography problem of [5]. For $\theta \in [0, 2\pi]$, $f(\theta) = R_\theta(I)$ where R_θ is the Radon transform at angle θ and I is a phantom image. We observe $n = 1024$ random projections: $y_i = f(\theta_i) + \nu_i$ where $\nu_i \sim N(0, \sigma^2)$. In [5], a NN graph ($k = 50$) is constructed, JDR is used to prune edges, and after pruning, nodes with less than two edges are removed. An eigenvalue problem is then solved to find an angular ordering of the projections and $\hat{I}$ is reconstructed via an inverse Radon transform. We compared JDR to NPDR pruning at a SNR of -2db. Exhaustive search finds the optimal q for JDR at $q = .78$; for NPDR we used $\hat{P}_\infty$ (all edges in G have weight 1), $p = .01$, and $q = .8$. JDR disconnected 277 nodes compared to 21 for NPDR. The estimated sorted angles are shown in Figs. 3(a),3(b), and the rotated reconstructions in Figs. 3(c),3(d). Under the similarity metric $\rho = I^T \hat{I} / (\|I\| \|\hat{I}\|)$, with alignment of $\hat{I}$ with I, the increase in NPDR similarity (0.15) over JDR similarity (0.12) is 25%. Note the clearer boundaries in the NPDR phantom, thanks to 256 additional (unpruned) projections (best viewed on screen). At moderate noise levels, NPDR removes fewer NN graph nodes and yields a more accurate reconstruction.

Preliminary evidence from both experiments indicates that, as §3 suggests, for NPDR one should choose p as small as possible while retaining numerical conditioning of $I - \bar{p} P_\epsilon$.

5. CONCLUSION

Our experiments indicate that NPDR robustly detects bridges in the NN graph without misclassifying edges important for geodesic estimation or tomographic angle estimation. Furthermore, it does so over a wider noise range than competing methods, e.g. LDR and ECDR. It can be calculated efficiently via a sparse eigenvalue decomposition. For very large n and small ϵ, the matrices N_ϵ and P_ϵ are equivalent, but in practical cases N_ϵ yields significantly better performance. More testing of NPDR on non-synthetic datasets is needed.

6. ACKNOWLEDGEMENTS

We thank Amit Singer and Hautieng Wu for discussions and code, and Ingrid Daubechies for advice and discussions. The first author acknowledges support of an NDSEG Fellowship.

7. REFERENCES

[1] J B Tenenbaum, V De Silva, and J C Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–23, 2000.

[2] M Belkin, P Niyogi, and V Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *JMLR*, vol. 7, pp. 2399–2434, 2006.

[3] M Balasubramanian and E L Schwartz, "The isomap algorithm and topological stability," *Science*, vol. 295, pp. 7, 2002.

[4] H Chen, G Jiang, and K Yoshihira, "Robust nonlinear dimensionality reduction for manifold learning," *ICPR*, vol. 1, pp. 447–450, 2006.

[5] A Singer and HT Wu, "2-d tomography from noisy projections taken at unknown random directions," *Submitted*, 2009.

[6] B Li, D-S Huang, and C Wang, "Improving the robustness of isomap by de-noising," *IJCNN*, pp. 266–270, June 2008.

[7] H Chang and D Yeung, "Robust locally linear embedding," *Pattern Recognition*, vol. 39, pp. 1053–1065, June 2006.

[8] R Subbarao and P Meer, "Subspace estimation using projection based m-estimators over grassmann manifolds," *ECCV*, vol. 3951, pp. 301–312, 2006.

[9] R R Coifman, S Lafon, A B Lee, M Maggioni, B Nadler, F Warner, and S W Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps," *PNAS*, vol. 102, pp. 7426–31, 2005.

[10] J Ham, D D Lee, S Mika, and B Schölkopf, "A kernel view of the dimensionality reduction of manifolds," in *ICML*. 2004, ACM New York, NY, USA.

[11] M Newman and M Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, pp. 1–15, 2004.

[12] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, pp. 163177, 2001.

[13] D Zhou and B Schölkopf, "Learning from labeled and unlabeled data using random walks," in *Pattern Recognition: 26th DAGM Symp.* 2004, pp. 237–244, Springer.