Combinatorial and Spectral Aspects of Nearest Neighbor Graphs in Doubling Dimensional and Nearly-Euclidean Spaces

Yingchao Zhao[1,⋆] and Shang-Hua Teng[2,⋆⋆]

[1] Department of Computer Science
Tsinghua University
yczhao@mails.tsinghua.edu.cn
[2] Department of Computer Science
Boston University
steng@cs.bu.edu

Abstract. Miller, Teng, Thurston, and Vavasis proved that every k-nearest neighbor graph (k-NNG) in $\mathbb{R}^d$ has a balanced vertex separator of size $O(n^{1-1/d}k^{1/d})$. Later, Spielman and Teng proved that the Fiedler value — the second smallest eigenvalue of the graph — of the Laplacian matrix of a k-NNG in $\mathbb{R}^d$ is at $O(\frac{1}{n^{2/d}})$. In this paper, we extend these two results to nearest neighbor graphs in a metric space with doubling dimension γ and in nearly-Euclidean spaces. We prove that for every $l > 0$, each k-NNG in a metric space with doubling dimension γ has a vertex separator of size $O(k^2 l(32l+8)^{2\gamma} \log^2 \frac{L}{S} \log n + \frac{n}{l})$, where L and S are respectively the maximum and minimum distances between any two points in P, and P is the point set that constitutes the metric space. We show how to use the singular value decomposition method to approximate a k-NNG in a nearly-Euclidean space by an Euclidean k-NNG. This approximation enables us to obtain an upper bound on the Fiedler value of the k-NNG in a nearly-Euclidean space.

Keywords: Doubling dimension, shallow minor, neighborhood system, metric embedding, Fiedler value.

1 Introduction

Graph partitioning is an important combinatorial optimization problem that is widely used in applications that include parallel processing, VLSI design, and data mining. There are several versions of this problem. Perhaps the simplest

⋆ Part of this work was done while visiting Computer Science Department at Boston University. In part supported by the National Grand Fundamental Research 973 Program of China under Grant (2004CB318108, 2004CB318110, 2003CB317007), the National Natural Science Foundation of China Grant (60553001, 60321002) and the National Basic Research Program of China Grant (2007CB807900, 2007CB807901).
⋆⋆ Part of this work was done while visiting Tsinghua University and Microsoft Research Asia Lab.

J.-Y. Cai, S.B. Cooper, and H. Zhu (Eds.): TAMC 2007, LNCS 4484, pp. 554–565, 2007.

version is to divide a graph into two equal-sized clusters and minimize the number of edges between these two clusters. In general, we may want to divide a graph into multiple clusters and minimize some objective functions such as the total number of inter-cluster edges or the maximum among the ratios defined by the number of edges leaving a cluster to the number of vertices in that cluster [9,10]. Graph partitioning is usually a hard problem if an optimal solution is desired [5]. But, because of its importance in practice, various partitioning heuristics and approximation algorithms are designed and implemented. The spectral method, which uses the eigenvectors of a graph matrix, has been among the most popular ones used in practice [1,12].

In this paper, we study combinatorial and spectral aspects relating with partitioning nearest neighbor graphs defined in Euclidean-like metric spaces. Our study is inspired by two early work on Euclidean nearest neighbor graphs. The first one is by Miller *et al* [11] who shows that every k nearest neighbor graph (k-NNG) of n points in $\mathbb{R}^d$ has a vertex separator of size $O(n^{1-1/d}k^{1/d})$ that $1/(d+2)$ splits the graph. Here, for a parameter $f : 0 < f < 1$, a vertex separator that f-splits a graph is a subset of its vertices whose removal divides the rest of the graph into at least two disconnected components such that the sizes of all components are no more than $f \cdot n$. If f is a constant, independent of n, then we refer to the vertex separator that f-splits as a *balanced* separator. The second one is by Spielman and Teng [17]. It shows that the Fiedler value — the second smallest eigenvalue of the graph — of the Laplacian matrix of a k-NNG in $\mathbb{R}^d$ is at $O(\frac{1}{n^{2/d}})$.

We first consider the k-NNG for points in a metric space of a finite doubling dimension. This family of metric spaces (see Section 2 for a formal definition) is introduced by Karger and Ruhl [8] with the motivation to extend efficient nearest-neighbor-search data structures from Euclidean spaces to other growth-constrained metric spaces arising in internet applications.

As one of the main results of this paper, we prove that for every $l > 0$, each k-NNG in a metric space with doubling dimension γ has a balanced vertex separator of size $O(k^2 l(32l+24)^{2\gamma}\log^2 \frac{L}{S} \cdot \log n + \frac{n}{l})$, where L and S are respectively the maximum and minimum distances between any two points in P. By choosing $l = n^{1/(2\gamma+2)}(k^2\log^2 \frac{L}{S} \cdot \log n)^{-1/(2\gamma+2)}$, we prove that every k-nearest neighbor graph of n points in a metric space with doubling dimension γ has a balanced vertex separator of size

$$O\left(n^{1-1/(2\gamma+2)}k^{1/(\gamma+1)}\log^{1/(\gamma+1)}(L/S) \cdot \log^{1/(2\gamma+2)} n\right)$$

We can also show that the maximum degree of these k-NNG is at most $O(k\log(L/S))$. Thus, this separator bound also implies that the Fiedler value of a k-nearest neighbor graph of n points in a metric space with doubling dimension γ is at most

$$O\left(\frac{n^{\frac{-1}{2\gamma+2}}k^{1+\frac{1}{\gamma+1}}\log^{\frac{1}{2\gamma+2}}(L/S)\log^{\frac{1}{2\gamma+2}} n}{1-2n^{\frac{-1}{2\gamma+2}}k^{\frac{1}{\gamma+1}}\log^{\frac{1}{\gamma+1}}(L/S)\log^{\frac{2}{\gamma+1}} n + n^{\frac{-1}{\gamma+1}}k^{\frac{2}{\gamma+1}}\log^{\frac{2}{\gamma+1}}(L/S)\log^{\frac{1}{\gamma+1}} n}\right)$$

Key to our proof, we characterize the family of minors excluded by these nearest neighbor graphs: For any given depth t, we show that these graphs can not contain a minor of size $O(kt^{\gamma} \log(L/S))$. With this graph-theoretic property, we can use the separator theorem of Plotkin, Rao, and Smith [13] to prove our separator bound.

For each k-NNG in nearly Euclidean spaces (see Section 4 for formal definition), we can apply the singular value decomposition method to find an approximate Euclidean k-NNG. This approximation enables us to obtain a better separator and Fiedler value bound than those that can be derived from doubling-dimensional framework.

We organize our paper as following. In Section 2, we introduce the notation and definitions which will be used in the paper. In particular, we will introduce doubling dimensional spaces, nearest neighbor graphs, the Fiedler value of a graph, and Singular Value Decomposition. We will prove the separator theorem for k-NNG in a finite doubling dimensional space in Section 3. For the k-nearest neighbor graphs in nearly-Euclidean space, we discuss their spectra in section 4. Finally, we conclude our work in Section 5.

2 Graphs and Geometry

In this paper, we consider graphs that are geometric defined. We first introduce some notation and definitions that will be used later. Given a graph $G = (V, E)$, we assume V is the point set from a metric space.

2.1 Metric Spaces and Doubling Dimension

Given a set X of points and a distance function d which is defined as d: $X \times X \longrightarrow [0, \infty)$, we call the pair (X, d) a *metric* space if it satisfies the following axioms.

– $\forall x, y \in X$, $d(x, y) = 0$ iff $x = y$.
– $\forall x, y \in X$, $d(x, y) = d(y, x)$.
– $\forall x, y, z \in X$, $d(x, y) + d(y, z) \geq d(x, z)$.

If (X, d) only satisfies the last two axioms and $d(x, x) = 0$ for all $x \in X$ instead of the first item, we call it a *semimetric* (or *pseudometric*).

There are various metric spaces with different dimensions, for example, the Euclidean space and the Hamming space. Not all the problems in practice can be modeled as graphs in an Euclidean space or a Hamming space. Although these metric spaces are simple and more familiar to us, practical problems may not satisfy all those geometric terms. The doubling dimensional space, which has less constraints, is introduced by Karger and Ruhl [8] and becomes useful in several research areas, such as graph partitioning and network routing. One objective of this paper is to design efficient algorithms for graphs in a metric space with finite doubling dimension.

Denote the space within a distance r to a point $v \in X$ as a ball $B_r(v)$ where r is the radius and v is the center. The metric (X, d) has a doubling dimension γ if any ball of radius r could be covered by 2^{γ} balls of radius $\frac{r}{2}$. Euclidean space could be considered as a special doubling dimension space. Different from general Euclidean spaces, doubling dimensional space has no such definitions as volume and parallelization. However, for those continuous doubling dimensional spaces, they could also have some useful properties, such as segment property as follows.

Definition 1. *A metric doubling dimension space* (X, d) *has segment property if for each pair of points* $x, y \in X$*, there exists a continuous curve* $\gamma = \gamma(t)$ *connecting* x *and* y *such that* $d(\gamma(t), \gamma(s)) = |t - s|$ *for all* t *and* s*.*

Segment property appears naturally in many applications and in this paper, we will mainly focus on those instances where segment property holds. For more details about segment property, please refer to [14]. We will give more properties in Section 3.

2.2 Nearest Neighbor Graphs

Let $P = \{p_1, \ldots, p_n\}$ be a set of n points in a metric space. For each $p_i \in P$, let $N_k(p_i)$ be the set of k points closest to p_i in P (if there are ties, break them arbitrarily). Let $R(p_i)$ be the distance between p_i and its k-th closest neighbor, hence $\forall p_i, p_j$, if $p_i \in N_k(p_j)$ then $\|p_i p_j\| \leq R(p_j)$. Let $B_R(p_i)$ be the ball centered at p_i with radius R, and we denote $\alpha B_R(p_i)$ to be the ball centered at p_i with radius αR. With $B_R(p_i)$, we can define k-ply systems, k-nearest neighbor graphs and intersection graphs in general metric spaces.

Definition 2. *Let* $P = \{p_1, \ldots, p_n\}$ *be points in a metric space, then a* k*-ply neighborhood system for* P *is a set of closed balls,* $B = \{B_1, \ldots, B_n\}$*, such that* B_i *centers at* p_i *and no point* p *in this metric space is contained in the interior of more than* k *balls from* B*.*

Definition 3. *A* k*-Nearest Neighbor Graph (*k*-NNG) of a set of* n *vertices is a graph with vertex set* $P = \{p_1, \ldots, p_n\}$ *and edge set* $E = \{(p_i, p_j)\colon p_i \in N_k(p_j)$ *or* $p_j \in N_k(p_i)\}$*, where* $N_k(p_i)$ *represents the set of* k *points closest to* p_i *in* V*. We denote the* k*-Nearest neighbor Graph of* P *as* $N_k(P)$*.*

Definition 4. *Given a set* S*, and a family of nonempty subsets of* S*, the corresponding intersection graph has a vertex for each subset, and a connecting edge whenever two subsets intersect.*

Definition 5. *Given a* k*-ply neighborhood system* $\Gamma = \{B_1, B_2, \ldots, B_n\}$*. The intersection graph of* Γ *is the undirected graph with vertices* $V = \{1, \ldots, n\}$ *and edges* $E = \{(B_i, B_j) : (B_i \cap B_j \neq \emptyset)\}$*.*

Definition 6. *Given a* k*-ply neighborhood system* $\Gamma = \{B_1, B_2, \ldots, B_n\}$ *and* $\alpha \geq 1$*. The* α*-overlap graph of* Γ *is the undirected graph with vertices* $V = \{1, \ldots, n\}$ *and edges* $E = \{(B_i, B_j) : (B_i \cap \alpha \cdot B_j \neq \emptyset)$ *and* $(B_j \cap \alpha \cdot B_i \neq \emptyset)\}$*.*

In this paper, the subsets are balls in some metric space. Therefore, we can let the vertex for each subset be the corresponding center of the ball. In this way, we can bound the degree of the intersection graph with the help of the ply bound in the original ball system.

2.3 Graph Partitioning and Vertex Separators

A *partition* of a graph $G = (V, E)$ is a division of its vertices into several specified number of subsets. Generally we focus on two objectives when we use graph partitioning : one objective is to minimize the number of the edges cut by the partition; the other one is to balance the computational load, i.e, to limit the size of each subset to within a tolerance. We call E_s, a subset of E, the *edge separator* of G, if removing E_s from E leaves two or more disconnected components of V. We call V_s, a subset of V, the *vertex separator* of G, if removing V_s and all incident edges leaves two or more disconnected components of V.

2.4 Laplacian and the Fiedler Value

Suppose $G = (V, E)$ is an undirected, connected graph, then its adjacent matrix $A(G) = (a_{ij})_{n\times n}$, where

$$a_{ij} = \begin{cases} 1 \text{ if } (i,j) \in E \\ 0 \text{ otherwise} \end{cases}$$

Let $D(G) = (d_{ij})_{n\times n}$ be a diagonal matrix where d_{ii} is the degree of the vertex v_i in the graph G. The Laplacian matrix of G is denoted as $L(G) = D(G) - A(G) = (l_{ij})_{n\times n}$. Hence

$$l_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } (i,j) \in E \\ 0 & \text{if } i \neq j \text{ and } (i,j) \notin E \\ degree(v_i) & \text{if i=j} \end{cases}$$

Because $L(G)$ is real and symmetric, its eigenvalues are all non-negative and its smallest eigenvalue is zero, with $(1, \ldots, 1)^T$ being its corresponding eigenvector. Fiedler [6] associated the second smallest eigenvalue of the Laplacian matrix of the graph with its connectivity. Thus, we call the second smallest eigenvalue of $L(G)$ the *Fiedler value* and call the corresponding eigenvector the *Fiedler vector*. Because G is connected, we know that the Fiedler value is non-zero and can be expressed as following.

$$\lambda_2 = \min_{x \perp (1,\ldots,1)^T} \frac{x^T L(G) x}{x^T x} = \min_{x \perp (1,\ldots,1)^T} \frac{\sum_{(i,j)\in E} (x_i - x_j)^2}{\sum_{i=1}^{n} x_i^2}$$

From the definition, we can get the following property.

Corollary 1. *The Fiedler value of the edge subgraph is no more than the Fiedler value of the original graph.*

2.5 Singular Value Decomposition

To learn more about Laplacian matrix and its Fiedler value, we review a useful technique called *singular value decomposition* (SVD). We give its formal definition below.

Definition 7. *A singular value decomposition of an $m \times n$ matrix A with $m \geq n$ is any factorization of the form*

$$A = UDV^T = \left[u_1, u_2, \ldots, u_n\right] \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{pmatrix} \tag{1}$$

where U is an $m \times n$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and D is an $n \times n$ diagonal matrix with $s_{ij} = 0$ if $i \neq j$ and $s_{ii} = \sigma_i \geq 0$.

In SVD, the quantity σ_i is a singular value of A. Without loss of generality, we assume that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ in this paper. We usually use two norms to describe the matrix. Given a matrix $A = (a_{ij})_{m \times n}$, the *Frobenius* norm ($F$ norm) of A is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} {a_{ij}}^2} = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$$

while the *Euclidean* norm (2-norm) of A is defined as

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 = 1} \|Ax\|_2$$

where x is an n dimensional vector and $\|x\|_2 = (x^T x)^{\frac{1}{2}} = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$.

In 1907, Erhard Schmidt [15] introduced the infinite dimensional analogue of the singular value decomposition. Eckart and Young [3,4] showed that if we replace the smallest $m - s$ singular values with zeros in D, then the new multiplication of UDV^T is the least square approximation in s dimensions of the original matrix A.

Theorem 1 (Eckart-Young). *Let the SVD of A given by (1) with rank $r = rank(A) \leq p = \min\{m, n\}$ and define*

$$A_k = U_k D_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

then A_k is the optimal approximation of A in the view of

$$\min_{rank(B)=k} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sum_{i=k+1}^{p} \sigma_i^2}$$

$$\min_{rank(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

Hence we can find a proper low-rank matrix A_k to approximate the original graph and Eckart-Young Theorem guarantees that this approximation will not cause much difference. For more properties of SVD, please refer to [2] and [7].

3 A Separator Theorem for Doubling Dimensional Spaces

In this section we prove a separator theorem for k-NNG in a metric space with finite doubling dimension.

Theorem 2. *For every $l > 0$, each k-NNG in a metric space with doubling dimension γ, we can find a separator of size $O(k^2 l(32l+8)^{2\gamma} \log^2 \frac{L}{S} \log n + \frac{n}{l})$, where L and S are respectively the maximum and minimum distances between any two points in P.*

We start with the following lemma that will be useful to obtain a degree bound for k-NNG in a doubling dimensional metric space.

Lemma 1. *For any ball of radius r in metric space with doubling dimension γ, it contains at most 2^γ disjoint balls of radius $\frac{r}{2}$.*

Similarly, we can get the following corollary.

Corollary 2. *For any ball of radius r in a metric space with doubling dimension γ, it contains at most $2^{t\gamma}$ disjoint balls of radius $\frac{r}{2^t}$.*

In this section, we will show an extended version of separator theorem in doubling dimensional spaces.

3.1 Shallow Minors

Key to our analysis is to show that k-NNG in finite doubling dimensional metric space excludes certain type of minors.

Definition 8. *A minor of a graph G is a graph obtained from G by a series of edge contractions and edge deletions.*

Teng [18] showed that for those balls in k-ply neighborhood system in Euclidean space, there could not be too many balls of large radius intersecting the same ball. We can get a similar result for graphs in doubling dimensional space, i.e, for any ball of radius r, there could not exist many balls of radius at least βr which intersect it.

Lemma 2. *Suppose $\{B_1, \ldots, B_n\}$ is a k-ply neighborhood system in a metric space with doubling dimension γ. For each ball B with radius r, for all constant $\beta > 0$, we have*

$$|\{i : B_i \cap B \neq \phi \ \ \text{and} \ \ r_i \geq \beta r\}| \leq k \left(\frac{2(1+2\beta)}{\beta}\right)^\gamma$$

where r_i is the radius of B_i.

Although there is no such definition as volume in Euclidean space, the doubling dimensional space does have similar shallow minor properties. We'll show that the intersection graph of a k-ply neighborhood system in doubling dimensional space does exclude shallow minors of a certain size.

Because there are not many intersecting balls, the intersection graph of k-ply neighborhood system does not have a large minor either.

Theorem 3. *Suppose Γ is a k-ply neighborhood system in a metric space with doubling dimension γ and G is the intersection graph of Γ. Then $\forall l$, G excludes K_h as a depth*[1] *l minor for $h \geq k(8l+2)^{\gamma}$.*

Proof. Suppose G has a K_h minor of depth l. We claim that there must exist h sets of balls, $\Gamma_1, \ldots, \Gamma_h \subset \Gamma$, such that:

- The intersection graph of each Γ_i is connected with diameter at most l.
- For each pair $i, j \in \{1, \ldots, h\}$, there's a ball in Γ_i that intersects a ball in Γ_j.

Let B_i be the ball of the largest radius in Γ_i. Without loss of generality, assuming that B_1 is the ball of the smallest radius among $\{B_1, \ldots, B_h\}$ and its radius is r. Hence, all the balls in Γ_1 are contained in a ball $B' = (2l+1)B_1$, because the intersection graph of Γ_1 is connected. According to the second condition, $\forall i > 1$, there is a ball from Γ_i that intersects B'.

We claim that for each $i > 1$, there is a ball in Γ_i of radius at least r that intersects the ball $(4l-1)B_1$.

As we know, the diameter of the intersection graph of Γ_i is at most l and there is a ball from Γ_i that intersects B'. If that intersecting ball has radius at least r, then we are done with Γ_i. If not, we can enlarge the radius of B' by $2r$, at that time, the enlarged B' will completely contain the intersecting ball in Γ_i and meet other balls in Γ_i because of the connectivity of Γ_i. Then we judge whether one of the intersecting balls has radius at least r. If not, we repeat the augment process above. Because B_1 is the ball of the smallest radius among $\{B_1, \ldots, B_h\}$, the process will surely terminate. This process is like a breadth-first-search. The number of iterations is less than $l-1$, since we will surely meet either B_i (the maximum-radius ball in Γ_i, whose radius is at least r) or some other balls in Γ_i that has radius at least r.

Namely, the ball B^* of radius $R = (4l-1)r$ intersects h balls of radius at least βR where $\beta = 1/(4l-1)$. Applying Lemma 2, we have $h \leq k(8l+2)^{\gamma}$. $\square$

Theorem 4. *Suppose Γ is a k-ply neighborhood system in a metric space with doubling dimension γ and G is the α-overlap graph of Γ. Then $\forall l$, G excludes K_h as a depth l minor for $h \geq k(8\alpha l+2)^{\gamma}$.*

3.2 Proof of Theorem 2

In this subsection, we give the proof of Theorem 2. First, let's bound the max degree of nearest neighbor graphs and the ply of neighborhood system in metric spaces with doubling dimension γ.

Lemma 3. *Let $P = \{p_1, \ldots, p_n\}$ be a point set in a metric space with doubling dimension γ. Then the ply of $N_k(P)$ is bounded by $k4^{\gamma}\log_{\frac{3}{2}}\frac{2L}{S}$, where L is the maximum distance between any two points in P, and S is the smallest one.*

[1] The maximum number of edges in each simple path.

The k-nearest neighbor graph has no more ply than $N_k(P)$, therefore we can bound the ply of k-NNG in doubling dimension space.

Corollary 3. *The ply for any k nearest neighbor graph in a metric space with doubling dimension γ is at most $k4^\gamma \log_{3/2} \frac{2L}{S}$, where L is the longest distance in the graph and S is the shortest.*

Plotkin, Rao and Smith [13] gave the following theorem and showed that we can find a small size separator for the graph which excludes shallow minors.

Theorem 5. *For any graph that excludes K_h as a depth l minor, we can find a separator of size $O(lh^2 \log n + n/l)$, where n is the number of vertices of the graph.*

Because every k-NNG in a metric space with doubling dimension γ has ply bound of $k4^\gamma \log_{3/2} \frac{2L}{S}$ and it excludes K_h minor with depth l, where $h > k4^\gamma \log_{3/2} \frac{2L}{S}(8l+2)^\gamma = k(32l+8)^\gamma \log_{\frac{3}{2}}(\frac{2L}{S})$. Applying Theorem 5 gives the separator bound of k-NNG in a metric space with doubling dimension γ. Therefore, Theorem 2 holds.

To minimize the separator size, we choose $l = n^{1/(2\gamma+2)}(k^2 \log^2 \frac{L}{S} \cdot \log n)^{-1/(2\gamma+2)}$ such that the two terms are equal and get that every k-nearest neighbor graph of n points in a metric space with doubling dimension γ has a balanced vertex separator of size

$$O(n^{1-1/(2\gamma+2)} k^{1/(\gamma+1)} \log^{1/(\gamma+1)}(L/S) \cdot \log^{1/(2\gamma+2)} n)$$

Since we have showed that the maximum degree of these k-NNG is at most $O(k \log(L/S))$, the above separator bound could also give an upper bound of the Fiedler value of a k-nearest neighbor graph of n points in a metric space with doubling dimension γ. Assign 1 to each vertex in the vertex separator and $|separator|/(|separator| - n)$ to the remaining vertices, then we have the following inequality.

$$\begin{aligned}
\lambda_2 &\leq \frac{\sum_{cut\ edge(i,j)} (x_i - x_j)^2}{\sum_{\forall i} x_i^2} \\
&\leq \frac{(\frac{n}{|separator|-n})^2 \times |separator| \times |max\ degree|}{n} \\
&= O(\frac{n^{\frac{-1}{2\gamma+2}} k^{1+\frac{1}{\gamma+1}} \log^{\frac{1}{2\gamma+2}}(L/S) \log^{\frac{1}{2\gamma+2}} n}{1 - 2n^{\frac{-1}{2\gamma+2}} k^{\frac{1}{\gamma+1}} \log^{\frac{1}{\gamma+1}}(L/S) \log^{\frac{2}{\gamma+1}} n + n^{\frac{-1}{\gamma+1}} k^{\frac{2}{\gamma+1}} \log^{\frac{2}{\gamma+1}}(L/S) \log^{\frac{1}{\gamma+1}} n})
\end{aligned}$$

4 A Spectral Theorem for Nearly-Euclidean Spaces

Since Fiedler [6] discovered that the second smallest eigenvalue is closely related to the connectivity of the graph, a large amount of work has been done on spectra analysis of graphs. In 1996, Spielman and Teng [17] proved that the Fiedler value of a k-nearest neighbor graph with n vertices in $\mathbb{R}^d$ is bounded by $O(k^{1+2/d}/n^{2/d})$.

In this session, we consider a point set P of n vertices in $\mathbb{R}^m$ space. $P = \{p_1, \ldots, p_n\} \subseteq \mathbb{R}^m$. We can get an $m \times n$ matrix $\mathbf{P}$ with column vectors $(p_1, \ldots, p_n)$. An upper bound of Fiedler value of the Laplacian matrix $L(\mathbf{P})$, given by Spielman and Teng, is as following.

Theorem 6. (Spielman-Teng) *If G is a subgraph of an α-overlap graph of a k-ply neighborhood system in $\mathbb{R}^m$ and the maximum degree of G is Δ, then the Fiedler value of $L(G)$ is bounded by $\gamma_m \Delta \alpha^2 (\frac{k}{n})^{2/m}$, where $\gamma_m = 2(\pi + 1 + \frac{\pi}{\alpha})^2 (\frac{A_{m+1}}{V_m})^{2/m}$.*

A_m is the surface volume of a unit m-dimensional ball, and V_m is the volume of a unit m-dimensional ball. In general case, the numbers k and α are two constants, and the item γ_m can be considered as a constant if the dimension m is fixed. Therefore, the bound can be expressed by $O(\frac{1}{n^{2/m}})$, which is dependant on the dimension of the space.

If we change the base carefully, the dimension could be changed as well. Hence we can consider the Laplacian matrix of a k-NNG and find a low-rank approximation matrix which can be contained in a lower dimension space so that the dimension of the new space is smaller. The changing of basis could make the problem easier, and we call the new space *nearly-Euclidean space*.

As we mentioned in Section 2, SVD could help us get a low-rank approximation matrix $\mathbf{Q}$ whose rank is d with $d < m$. Suppose the column vectors of $\mathbf{Q}$ is $(q_1, \ldots, q_n)$ and these n points form a new point set Q. Suppose that G' is the $(1 + \frac{7\delta}{s})$-overlap graph of the k-NNG of Q, the maximum degree of G' is Δ, s is the length of the shortest edge in G', δ is the maximum distance between each p_i and q_i for any $i \in \{1, \ldots, n\}$, we can prove the following theorem and get a more accurate bound for $L(\mathbf{P})$.

Theorem 7. *If G is the k-NNG of the point set P in $\mathbb{R}^m$ space, then using SVD, we can find an approximate point set Q with $rank(\boldsymbol{Q}) = d < m$, and the Fiedler value of $L(\boldsymbol{P})$ can be bounded by $(1 + \frac{7\delta}{s})^2 \gamma_d \Delta \tau_d k/n)^{\frac{2}{d}}$ where $\gamma_d = 2(\pi + 1 + \frac{\pi}{\alpha})^2 (\frac{A_{d+1}}{V_d})^{2/d}$.*

Here A_d is the surface volume of a unit d-dimensional ball, and V_d is the volume of a unit d-dimensional ball. To make the idea look clearer, let's consider a simple example in $\mathbb{R}^2$ space. $Q = \{q_1, \ldots, q_n\}$ is a set of n points in $\mathbb{R}^2$ space. We perturb these n points in the direction perpendicular to the original plane and get a new set of n points, denoted by $P = \{p_1, \ldots, p_n\}$, in $\mathbb{R}^3$ space. Assuming that the smallest distance between any two points of Q is s, and the perturbation distance is at most δ. If $s \geq \delta$, we can get the following inequalities.

$$\|p_i - p_j\| \leq \sqrt{(2\delta)^2 + \|q_i - q_j\|^2} \leq \sqrt{5}\|q_i - q_j\|$$

If r_i is the k-NNG radius for q_i, and R_i is the k-NNG radius for p_i, then we can see that $R_i \leq \sqrt{5} r_i$ for all $i \in \{1, \ldots, n\}$. Therefore, we can use $\sqrt{5}$-overlap graph G' of Q to approximate the k-NNG G of P. And the Fiedler value of $L(\mathbf{G'})$ can also be bounded by the Fiedler value of $L(\mathbf{G})$. In fact, we can think that all those n points of P in $\mathbb{R}^3$ are perturbed perpendicularly to the same plane and the new point set on the plane is Q.

To prove Theorem 7, let's get some preparations.

Lemma 4. *$\forall p_i \in P$, its k-NNG radius R_i is no more than $r_i + 2\delta$, where r_i is the k-NNG radius of the corresponding point q_i in Q.*

Lemma 5. *The 1-overlap graph of k-NNG of P in $\mathbb{R}^m$ is isomorphic to a subgraph of the $(1 + 7\delta/s)$-overlap graph of k-NNG of Q in $\mathbb{R}^d$.*

Lemma 6. *The k-NNG is a subgraph of its 1-overlap graph.*

Proof. Suppose G is a k-NNG, G' is the 1-overlap graph of G and (p_i, p_j) is an arbitrary edge of G. Then we can see that $\|p_i - p_j\| \le r_i$ or $\|p_i - p_j\| \le r_j$. In addition, $\|p_i - p_j\| \le r_i + r_j$. Hence (p_i, p_j) must exists in the graph G'. From the generality of (p_i, p_j), we can see that the k-NNG graph is a subgraph of its 1-overlap graph. □

Combining Lemma 4, 5 and 6, we can derive the following corollary.

Corollary 4. *The k-NNG of P in $\mathbb{R}^m$ is isomorphic to a subgraph of the $(1 + 7\delta/s)$-overlap graph of a k-NNG of Q in $\mathbb{R}^d$, where $\delta = \max \|p_i - q_i\|$ and $s = \min \|q_i - q_j\|$.*

In [11] it is shown that any k-NNG is a subgraph of a $k\tau_d$-ply neighborhood system where τ_d is the kissing number in dimension d. If G is an α-overlap graph of a k-NNG in $\mathbb{R}^d$ then G is a subgraph of an α-overlap graph of a $k\tau_d$-neighborhood system in $\mathbb{R}^d$. Suppose that the maximum degree of G is Δ, we can apply Theorem 6 and get the following corollary directly.

Corollary 5. *If G is a subgraph of the α-overlap graph of k-NNG in $\mathbb{R}^d$ with maximum degree Δ, then the Fiedler value of $L(G)$ is bounded by $\gamma_d \Delta \alpha^2 (\tau_d k/n)^{2/d}$, where $\gamma_d = 2(\pi + 1 + \pi/\alpha)^2 (A_{d+1}/V_d)^{2/d}$.*

Finally, we give the proof of Theorem 7.

Proof of Theorem 7. The k-NNG of P in $\mathbb{R}^m$ is isomorphic to a subgraph of the $(1+7\delta/s)$-overlap graph of k-NNG of Q in $\mathbb{R}^d$, according to Corollary 6. The isomorphic graph has the same Fiedler value as the original graph because they have the same Laplacian matrices. Hence the k-NNG of P in $\mathbb{R}^m$ has no larger Fiedler value than the $(1 + 7\delta/s)$-overlap graph of k-NNG of Q in $\mathbb{R}^d$ according to Corollary 1.Due to Corollary 5, we know that the Fiedler value of k-NNG in $\mathbb{R}^m$ is bounded by $(1 + \frac{7\delta}{s})^2 \gamma_d \Delta \tau_d k/n)^{\frac{2}{d}}$ where $\gamma_d = 2(\pi + 1 + \frac{\pi}{\alpha})^2 (\frac{A_{d+1}}{V_d})^{2/d}$. □

5 Conclusion

In this paper, we concentrate on the combinatorial and spectral aspects of nearest neighbor graphs in doubling dimensional metric spaces and nearly-Euclidean spaces. For those k-nearest neighbor graphs in metric spaces with doubling dimension γ, we give the ply bound and degree bound, where there are no definitions like volume or parallel. We analyze the shallow minor excluded property and bound the separator size. For those graphs in Euclidean spaces with high dimension, we prove that the k-nearest neighbor graphs could have better spectral properties using SVD. If the number k is a constant, then we can show that its Fiedler value can be bounded by $O(\Delta(1+7\delta/s)^2 n^{-2/d})$ where Δ is the maximum degree of the approximation graph.

References

1. E. R. Barnes and A. J. Hoffman, *Partitioning, spectra and linear programming*, Progress in Combinatorial Optimization, pages 13–25, (1984).
2. P. Dewilde and E. F. Deprettere, *Singular value decomposition: An introduction*, In Ed. F. Deprettere, editor. SVD and Signal Processing: Algorithms, Applications, and Architectures pages Elsevier Science Publishers North Holland.
3. C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1:211-218, (1936)
4. C. Eckart and G. Young, *A principal axis transformation for non-Hermitian matrices*, Bulletin of the American Mathematical Society, 45:118–121 (1939)
5. T. Feder, P. Hell, S. Klein, R. Motwani, *Complexity of Graph Partition Problems*, STOC 1999: 464-472.
6. M. Fiedler, *Algebraic connectivity of graphs.* Czechoslovak Mathematical Journal, 23(98):298-305, (1973)
7. Gene H. Golub and Gharles F. Van Loan,*Matrix Computations*, pages 16-21,293. Johns Hopkins University Press, Baltimore, Maryland (1992).
8. David R. Karger, Matthias Ruhl, *Finding nearest neighbors in growth-restricted metrics.* STOC 2002: 741-750.
9. Richard J. Lipton, Robert Endre Tarjan, *A separator theorem for planar graphs*, SIAM J. Appl. Math. 36:177-189, (1979)
10. Richard J. Lipton, Robert Endre Tarjan, *Applications of a planar separator theorem*, SIAM Journal on Computing 9:615-627, (1980)
11. Gary L. Miller, Shang-Hua Teng, William P. Thurston, Stephen A. Vavasis, *Separators for sphere-packings and nearest neighbor graphs*, J. ACM 44(1): 1-29 (1997)
12. A. Pothen, H. D. Simon, and K.-P. Liou, *Partitioning sparse matrices with eigenvectors of graphs*, SIAM J. Matrix Anal. Appl, 11:430–452, (1990)
13. S. Plotkin, S. Rao, and W.D. Smith, *Shallow excluded minors and improved graph decomposition*, In proceedings at the 5th Symposium Discrete Algorithms. SIAM, New York, pp.462-470 (1994)
14. Michael Ruzhansky, *On uniform properties of doubling measures*, Proc. of Amer. Math. Soc., 129: 3413-3416 (2001)
15. E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener*, Mathematische Annalen, 63:433-476, (1907)
16. Daniel A. Spielman, Shang-Hua Teng, *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, STOC 2004: 81-90.
17. Daniel A. Speilman and Shang-Hua Teng, *Spectral Partitioning Works: Planar Graphs and Finite Element Meshes*, FOCS 1996: 96-105.
18. Shang-Hua Teng, *Combinatorial aspects of geometric graphs*, Computational Geometry no. 9: 277-287 (1998)