Theoretical Computer Science 410 (2009) 1081–1092

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Combinatorial and spectral aspects of nearest neighbor graphs in doubling dimensional and nearly-Euclidean spaces

Yingchao Zhao [a,*], Shang-hua Teng [b]

[a] *Department of Computer Science and Technology, Tsinghua University, China*

[b] *Department of Computer Science, Boston University, United States*

ARTICLE INFO

Keywords:
Doubling dimension
Shallow minor
Neighborhood system
Metric embedding
Fiedler value

ABSTRACT

Miller, Teng, Thurston, and Vavasis proved a geometric separator theorem which implies that the k-nearest neighbor graph (k-NNG) of every set of n points in $\mathbb{R}^d$ has a balanced vertex separator of size $O(n^{1-1/d}k^{1/d})$. Spielman and Teng then proved that the Fiedler value — the second smallest eigenvalue of the Laplacian matrix — of the k-NNG of any n points in $\mathbb{R}^d$ is $O((k/n)^{2/d})$. In this paper, we extend these two results to nearest neighbor graphs in a metric space with a finite doubling dimension and in a metric space that is nearly-Euclidean. We prove that for every $l > 0$, if $(X, \mathbf{dist})$ forms a metric space with doubling dimension γ, then the k-NNG of every set P of n points in X has a vertex separator of size $O(k^2 l(64l+8)^{2\gamma} \log^2 \frac{L}{S} \log n + \frac{n}{l})$, where L and S are, respectively, the maximum and minimum distances between any two points in P. We show how to use the singular value decomposition method to approximate a k-NNG in a nearly-Euclidean space by a Euclidean k-NNG. This approximation enables us to obtain an upper bound on the Fiedler value of k-NNGs in a nearly-Euclidean space.

1. Introduction

Graph partitioning is an important combinatorial optimization problem that has been widely used in applications such as parallel processing, VLSI design, and data mining. There are several versions of this problem. The simplest one is to divide a graph into two equal-sized clusters in order to 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 an NP-hard problem if an optimal solution is desired [5]. Because of its importance in practice, various partitioning heuristics and approximation algorithms have been designed and implemented. The spectral method, which uses the eigenvectors of a matrix defined from the graph, is among the most popular ones used in practice [1,13].

In this paper, we study the combinatorial and spectral properties of nearest neighbor graphs defined by points in Euclidean-like metric spaces. We will give an upper bound on the separator size as well as an upper bound on the Fielder value of these graphs. Our study is inspired by the following work on Euclidean nearest neighbor graphs. Miller *et al* [12] proved that the k-nearest neighbor graph (k-NNG) of every set 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. Recall that 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 no component has size more than $f \cdot n$. If f is a constant, independent of n, then we refer to the vertex separator that f-splits a graph as

* Corresponding author.
E-mail addresses: zhaoyingchao@gmail.com (Y. Zhao), steng@cs.bu.edu (S.-h. Teng).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2008.10.032

a *balanced* separator. Building on the technique of Miller *et al*, Spielman and Teng [17] proved that the Fiedler value — the second smallest eigenvalue of the Laplacian matrix — of a k-NNG in $\mathbb{R}^d$ is at $O((k/n)^{2/d})$.

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) was introduced by Karger and Ruhl [8]. The motivation of their work is 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 $l > 0$, if $(X, \mathbf{dist})$ forms a metric space with doubling dimension γ, then the k-NNG of every set P of n points in X has a balanced vertex separator of size $O(k^2 l(64l + 8)^{2\gamma} \log^2 \frac{L}{S} \cdot \log n + \frac{n}{l})$, where L and S are, respectively, the maximum and minimum distances between any two points in X.

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 the k-nearest neighbor graph of every set of n points in a metric space with doubling dimension γ has a 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 also show that the maximum degree of these k-NNGs is at most $O(k \log(L/S))$. Thus, our degree bound also implies that the Fiedler value of the k-nearest neighbor graph of any n points in a metric space with doubling dimension γ is at most $O(k \log (L/S))$.

Key to our proof, we characterize the family of minors excluded by these nearest neighbor graphs: For any given depth parameter t, we show that these graphs cannot contain minor of depth at most t and size $O(kt^\gamma \log(L/S))$. With this graph-theoretic property, we can use the separator theorem of Plotkin, Rao, and Smith [14] to prove our separator bound.

For each k-NNG in a nearly-Euclidean space (see Section 4 for a 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 a doubling dimensional framework.

We organize our paper as follows. 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 (SVD). We will prove the separator theorem for k-NNGs in a finite doubling dimensional space in Section 3. For k-nearest neighbor graphs in nearly-Euclidean space, we bound 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 geometrically defined. We first introduce some notation and definitions that will be used later. Given a graph $G = (V, E)$, we assume V is a set of points from a metric space.

2.1. Metric spaces and doubling dimension

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

- $\forall x, y \in X, \mathbf{dist}(x, y) = 0$ iff $x = y$.
- $\forall x, y \in X, \mathbf{dist}(x, y) = \mathbf{dist}(y, x)$.
- $\forall x, y, z \in X, \mathbf{dist}(x, y) + \mathbf{dist}(y, z) \geq \mathbf{dist}(x, z)$.

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

There are various families of metric spaces, such as the Euclidean spaces and the Hamming spaces, parameterized by their dimensions. Although these two families of metric spaces are simpler and more familiar to us, not all practical problems can be modeled as graphs in a Euclidean space or a Hamming space. The family of doubling dimensional spaces, which has fewer constraints, was introduced by Karger and Ruhl [8]. This family has become useful in several research areas, including graph partitioning and network routing. One objective of this paper is to study the structure properties of graphs in a metric space with finite doubling dimension and design efficient algorithms for them.

Suppose that $v \in X$ is a point and r is a positive real number. Let $B_r(v)$ denote the ball that contains all points in X whose distances to v are at most r. The metric $(X, \mathbf{dist})$ has a doubling dimension γ if any ball of radius r could be covered by 2^γ balls of radius $\frac{r}{2}$. The two-dimensional Euclidean space could be considered as a special doubling dimension space with $\gamma = \log_2 7$. Different from general Euclidean spaces, doubling dimensional spaces have no such definitions as volume and parallelization. We will review some properties of doubling dimensional spaces 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. Then, $\forall p_i, p_j$, if

$p_i \in N_k(p_j)$, then $\|p_i - p_j\| \le R(p_j)$. Suppose that $B_R(p_i)$ is the ball centered at p_i with radius R. We denote the ball centered at p_i with radius αR by $\alpha B_R(p_i)$. With $B_R(p_i)$, we can define k-ply systems, k-nearest neighbor graphs and intersection graphs in general metric spaces.

Definition 1. A k-Nearest Neighbor Graph (k-NNG) of a set of n vertices is an undirected graph with vertex set $P = \{p_1, \ldots, p_n\}$ and edge set $E = \{(p_i, p_j) \colon p_i \in N_k(p_j) \text{ or } p_j \in N_k(p_i)\}$. We denote the k-Nearest Neighbor Graph of P as $N_k(P)$.

Let $B_{N_k}(P) = \{B_{R(p_1)}(p_1), \ldots, B_{R(p_n)}(p_n)\}$ be the corresponding neighborhood system of $N_k(P)$. The ply of B_{N_k} will be discussed in Section 3, as well as the definition of neighborhood system.

2.3. Graph partitioning and vertex separators

A *partition* of a graph $G = (V, E)$ is a division of its vertices. We will focus on two objectives motivated by the application of graph partitioning in parallel processing. The first objective is to minimize the number of the edges cut by the partition. The second objective is to balance the computational load, i.e., to limit the size of each cluster to within a tolerance. We call E_s, a subset of E, an *edge separator* of G, if the removal of E_s from E creates two or more disconnected components of V. We call V_s, a subset of V, a *vertex separator* of G, if the removal of V_s and all incident edges induces 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 adjacency matrix is $A(G) = (a_{ij})_{n\times n}$, 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 G. The Laplacian matrix of G is denoted as $L(G) = D(G) - A(G) = (l_{ij})_{n\times n}$. Hence

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

Because $L(G)$ is real, symmetric and positive semi-definite, its eigenvalues are all non-negative and its smallest eigenvalue is zero, with $(1, \ldots, 1)^{\mathrm{T}}$ being its corresponding eigenvector. Fiedler [6] studied the second smallest eigenvalue of graph Laplacians in the context of connectivity. 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 follows.

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

Given E', a subset of E, the graph $G' = (V, E')$ is called an *edge subgraph* of $G = (V, E)$. And we can get the following property about edge subgraph.

Proposition 2. *For any graph G, the Fiedler value of an edge subgraph of G is no more than the Fiedler value of G.*

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).

Definition 3. A singular value decomposition of an $m \times n$ matrix A with $m \ge n$ is any factorization of the form

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

where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and D is an $m \times n$ diagonal matrix.

In SVD, the quantity σ_i is called a singular value of A, and $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0$. Given a matrix $A = (a_{ij})_{m\times n}$, recall that 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 $\|x\|_2 = (x^{\mathrm{T}}x)^{\frac{1}{2}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

In 1907, Erhard Schmidt [16] 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 of rank s of the original matrix A.

Theorem 4 (*Eckart–Young*). *Let the SVD of A be given by* (1) *with rank* $r = rank(A) \le p = \min\{m, n\}$ *and define*

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

then,

$$\min_{rank(B)\le k}\|A - B\|_F = \|A - A_k\|_F = \sqrt{\sum_{i=k+1}^{p}\sigma_i^2}$$
$$\min_{rank(B)\le k}\|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}.$$

Hence we can use the lower-rank matrix A_k to approximate the original matrix A and Eckart–Young Theorem guarantees that this approximation is the best possible. For more properties of SVD, please refer to [2] and [7].

3. A separator theorem for doubling dimensional spaces

In this section, we prove the following separator theorem.

Theorem 5. *For every* $l > 0$*, if P is a set of n points in a metric space with doubling dimension* γ*, then the k-NNG of P has a vertex separator of size* $O(k^2 l(64l + 8)^{2\gamma}\log^2\frac{L}{S}\cdot\log n + \frac{n}{l})$*, where L and S are, respectively, the maximum and minimum distances between any two points in P.*

We start with the following useful lemma. It can be used to obtain a degree bound for k-NNGs in a finite doubling dimensional metric space.

Lemma 6. *Every ball of radius r in a metric space with doubling dimension* γ *contains at most* 2^{γ} *disjoint balls of radius* $\frac{r}{2}$.

Proof. By the definition of doubling dimension, we see that for any ball B of radius r there exist 2^{γ} balls of radius $\frac{r}{2}$ which can cover B completely, assuming that those balls are $C_1, C_2, \ldots, C_{2^{\gamma}}$. Suppose that $B_1, B_2, \ldots, B_k$ are k disjoint balls of radius $\frac{r}{2}$ in B, and their corresponding centers are $v_1, v_2, \ldots, v_k$. Hence $\forall\, i, j \in \{1, \ldots, k\}$ and $i \ne j$, we have $B_i \cap B_j = \emptyset$, i.e., $\|v_i - v_j\| > r$.

Because the radius of every C_i is $\frac{r}{2}$, according to the triangle inequality, the distance between any two points in ball C_i is no more than r. Thus, each C_i cannot cover more than one point in $\{v_1, v_2, \ldots, v_k\}$. Because $C_1, C_2, \ldots, C_{2^{\gamma}}$ could cover B, they must cover $v_1, v_2, \ldots, v_k$ as well. Hence we know that $k \le 2^{\gamma}$. □

Similarly, we have the following corollary.

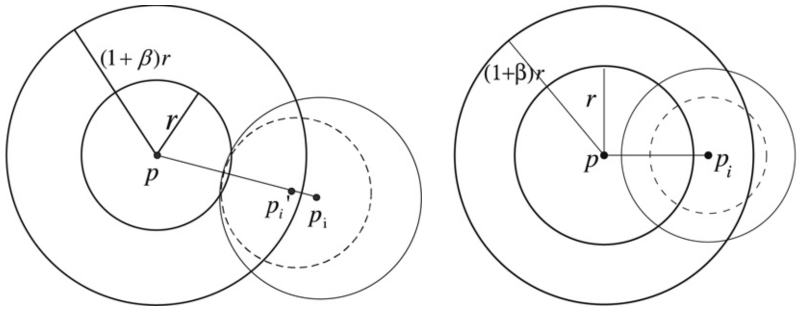
Corollary 7. *Every ball of radius r in a metric space with doubling dimension* γ *contains at most* $2^{\lceil t\rceil\gamma}$ *disjoint balls of radius* $\frac{r}{2^t}$ *for any* $t \ge 1$.

3.1. Shallow minors

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

The key to our analysis is to show that k-NNGs in a finite doubling dimensional metric space exclude certain type of minors.

Definition 9. 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 is centered at p_i and no point p in the metric space is contained in the interior of more than k balls from B.



(a) The first class. (b) The second class.

Fig. 1. The shrinking and moving for two classes.

Definition 10. Given a set Ω, and another set S which is a family of non-empty subsets of Ω, the corresponding intersection graph has a vertex for each subset, and a connecting edge whenever two subsets intersect. Specifically, for a k-ply neighborhood system $\Gamma = \{B_1, B_2, \dots, B_n\}$, the intersection graph of Γ is the undirected graph with vertices $V = \{1, \dots, n\}$ and edges $E = \{(i, j) : B_i \cap B_j \neq \emptyset\}$.

Teng [18] showed that for each ball in a k-ply neighborhood system in a Euclidean space, there cannot be too many balls of large radius intersecting the ball. We can get a similar result for graphs in doubling dimensional spaces. Here we consider the doubling dimensional spaces with the following property.

Definition 11. A metric doubling dimensional space $(X, \mathbf{dist})$ 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 $\mathbf{dist}(\gamma(t), \gamma(s)) = |t - s|$ for all t and s.

More details about the segment property can be found in [15]. Next we will prove that for any ball of radius r, there cannot exist too many balls of radius at least βr that intersect it.

Lemma 12. *Suppose that $\{B_1, \dots, B_n\}$ 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*

$$|\{i : B_i \cap B \neq \phi \text{ and } r_i \geq \beta r\}| \leq (k+1)\left(\frac{4(1+3\beta/2)}{\beta}\right)^{\gamma}$$

where r_i is the radius of B_i.

Proof. Let p be the center of B, and let $B' = (1+\beta)B$. Suppose that t balls with radius at least βr intersect B. Without loss of generality, we assume that those t balls are $B_1, \dots, B_t$, which are divided into two classes. One class contains those balls whose centers are in the exterior of B' and the other class contains those balls whose centers are in B' (see Fig. 1).

For any ball B_i (with center p_i and radius r_i) in the first class, due to Definition 11 (segment property of the doubling dimensional space), we can find a new point p_i' along the curve from p to p_i such that $\|p - p_i'\| + \|p_i' - p_i\| = \|p - p_i\|$, $\|p_i - p_i'\| \leq r_i - \beta r$ and $\|p - p_i'\| \leq (1+\beta)r$ (If $\|p - p_i'\| \geq r_i$, then we can choose p_i' such that $\|p_i - p_i'\| = r_i - \beta r$; otherwise, we can choose p_i' such that $\|p_i - p_i'\| = \|p - p_i\| - \beta r$). Consider any point v such that $\|v - p_i'\| \leq \beta r$, then we have $\|v - p_i\| \leq \|v - p_i'\| + \|p_i' - p_i\| \leq \beta r + (r_i - \beta r) = r_i$. We denote the ball of radius βr with center p_i' as B_i'. Hence B_i' is completely contained in B_i. Additionally, we can see that those B_i''s are all contained in $(1+2\beta)B$. Hence we can shrink each ball in the first class to radius βr and move its center to the corresponding p_i', so that the k-ply condition of the neighborhood system does not change and the new balls are all contained in $(1+2\beta)B$.

We now consider the second class. Similarly, we can shrink each ball B_i (with center p_i and radius r_i) to radius βr and let $p_i' = p_i$. Hence $\|p - p_i'\| = \|p - p_i\| \leq (1+\beta)r$ and the new ball B_i' (with center p_i' and radius βr) is contained in $(1+2\beta)r$. Obviously, each B_i' is contained in B_i and substitution from B_i to B_i' does not increase the ply in the neighborhood system.

Next we shrink each ball B_i' to its corresponding homocentric ball B_i^* with radius $\frac{\beta r}{2}$. We claim that for any ball B_i^*, there are no more than k balls B_j^* having overlap with it.

To prove this claim, we consider the set $S = \{B_1^*, B_2^*, \dots, B_t^*\}$ and will show that the corresponding intersection graph $G(S)$ of S is degree bounded, i.e., the max degree of $G(S)$ is at most k. Suppose that there exists a vertex v_i (whose corresponding ball is B_i^* in S) in $G(S)$ whose degree is greater than k. There are at least $k+1$ balls that intersect B_i^*, and without loss of generality, we assume that those balls are $B_1^*, B_2^*, \dots, B_{k+1}^*$. Because these balls have overlap with B_i^*, we can see that $\|p_i - p_j\| \leq \frac{\beta r}{2} + \frac{\beta r}{2} = \beta r$ for $1 \leq j \leq k+1$. Recall that the radius of each ball B_j' is βr. Thus, the center of ball B_i' is contained in each B_j', i.e., the ply of the original neighborhood system is at least $k+1$, which contradicts the assumption of k-ply condition. Consequently, the claim that for any ball B_i^*, there are no more than k balls B_j^* having overlap with it, is true.

Similar as the method proposed in [11], we color all the vertices (which are the corresponding centers of ball B_i^*) with $k+1$ different colors and make sure that every adjacent two centers have different colors, because the max degree in the intersection graph is no more than k. Then we classify these balls B_i^* according to the colors of their centers and get $k+1$ disjoint sets.

For each color set, there are several disjoint balls B_i^*'s with radius $\frac{\beta r}{2}$ contained in $(1+\frac{3\beta}{2})B$. Applying Corollary 7, we can see that there are at most $(\frac{2(1+3\beta/2)}{\beta/2})^\gamma$ disjoint balls in $(1+3\beta/2)B$ for each color set. Because there are at most $k+1$ different color sets, there are at most $(k+1)(\frac{4}{\beta}+6)^\gamma$ balls with radius at least βr intersecting the ball B. $\square$

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

Definition 13. A depth l minor is a minor of which the length of the longest simple path is l.

Theorem 14. *Suppose that Γ is a k-ply neighborhood system in a metric space with doubling dimension γ and G is the intersection graph of Γ. Then $\forall l > 0$, G excludes K_h as a depth l minor for $h \geq (k+1)(16l+2)^\gamma$.*

Proof. Suppose that 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, assume 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 the 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 $2r$ and 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 ball 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 12, we have $h \leq (k+1)(16l+2)^\gamma$. $\square$

Now we introduce overlap graph, which is defined in a neighborhood system.

Definition 15. 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 $E = \{(i,j) : (B_i \cap \alpha \cdot B_j \neq \emptyset) \textit{ and } (B_j \cap \alpha \cdot B_i \neq \emptyset)\}$.

For those overlap graphs of neighborhood system, we can also get the shallow minor excluded properties as follows.

Theorem 16. *Suppose that 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 $\forall l > 0$, G excludes K_h as a depth l minor for $h \geq (k+1)(16\alpha l+2)^\gamma$.*

Proof. This proof is similar to that of Theorem 14. Suppose that 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 is a ball U_i in Γ_i and a ball U_j in Γ_j such that $U_i \cap \alpha \cdot U_j \neq \emptyset$ and $U_j \cap \alpha \cdot U_i \neq \emptyset$.

Suppose that B_i is the ball of the largest radius in Γ_i. Without loss of generality, assume 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 the ball $(2\alpha l+1)B_1$, because the depth of α-overlap graph of Γ_1 is at most l. According to the second condition, $\forall i > 1$, there exists a ball U_i from Γ_i and a ball U in Γ_1 such that $\alpha U \cap U_i \neq \emptyset$.

We claim that for each $i > 1$, there is a ball in Γ_i of radius at least r that intersects the ball $(4\alpha l-1)B_1$. The proof is similar to that of Theorem 14. Hence, the ball B^* of radius $R = (4\alpha l-1)r$ intersects h balls of radius at least βR where $\beta = 1/(4\alpha l-1)$. Applying Lemma 12, we have $h \leq (k+1)(16\alpha l+2)^\gamma$. $\square$

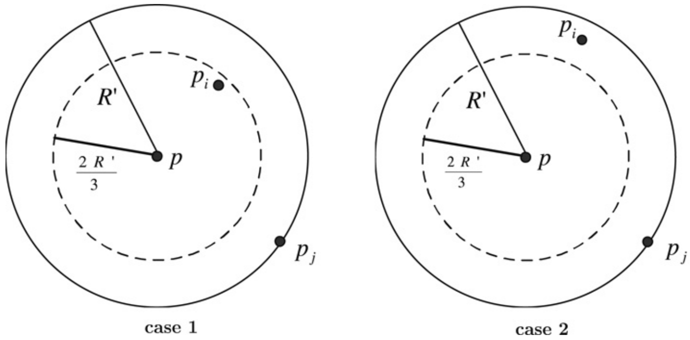


Fig. 2. Two cases in the last iteration.

3.2. Proof of Theorem 5

In this subsection, we give the proof of Theorem 5. First, we bound the maximum degree of nearest neighbor graphs and the ply of neighborhood system in metric spaces with doubling dimension γ.

Lemma 17. *Let $P = \{p_1, \ldots, p_n\}$ be a point set in a metric space with doubling dimension γ. Then the ply of $B_{N_k}(P)$ is bounded by $k \cdot 4^{\gamma} \log_{\frac{3}{2}} \frac{2L}{S}$, where L is the maximum distance between any two points in P, and S is the smallest one.*

Proof. We first consider the case of $k = 1$, i.e., the radius of each ball equals the distance from its center to its nearest neighbor. In such a case, no ball contains the center of other balls in its interior.

Suppose that p is the point which has the largest ply t. Without loss of generality, we assume that these t balls that contain p are $\{B_1, B_2, \ldots, B_t\}$ with corresponding center set $Q = \{p_1, \ldots, p_t\}$. According to the definition of $B_{N_k}(P)$, we can derive that $\forall p_i, p_j \in Q$, $\|p_i - p_j\| \geq \|p - p_i\|$ and $\|p_i - p_j\| \geq \|p - p_j\|$. Then we consider the point set $U = \{p_i : \|p - p_i\| > \frac{2}{3}R\} \subseteq Q$ where $R = \max_{1 \leq i \leq t} \|p - p_i\|$. And for each $p_i \in U$, let C_i be the ball of radius $\frac{1}{3}R$ centered at p_i. We claim that: (1) those C_i's do not have overlap with each other; (2) they are completely contained in a ball centered at p of radius $\frac{4}{3}R$.

As we mentioned above, for any $p_i, p_j \in U$, $\|p_i - p_j\| \geq \|p - p_j\| > 2R/3$, therefore, C_i and C_j do not have overlap with each other. According to the triangle inequality, $\forall C_i$ and for any point $v \in C_i$, $\|p - v\| \leq \|p - p_i\| + \|p_i - v\| \leq R + \frac{1}{3}R = \frac{4}{3}R$. Hence those C_i's where $p_i \in U$ are all contained in a ball centered at p of radius $\frac{4}{3}R$.

Due to Corollary 7, we can see that there are at most $2^{2\gamma} = 4^{\gamma}$ such disjoint C_i in that $\frac{4}{3}R$ radius ball, so there are at most 4^{γ} points in the point set U.

We can remove those points and consider the remaining points with $R' = \max \|p - p_j\| \leq \frac{2}{3}R$. With the similar method, we can estimate the number of points which are at least $\frac{2}{3}R'$ far away from p. We can do this iteration on and on until there are fewer than 2 points left around p within distance $\frac{2}{3}R'$ in the last iteration.

There are two cases shown in Fig. 2. The first case happens when there is only one point left around p within distance $\frac{2}{3}R'$. The second case happens when there is no point left around p with distance less or equal to $\frac{2}{3}R'$. Because there are at least two points around p in the last iteration, without loss of generality, we assume two of these points are p_i and p_j. Hence we can see that $S \leq \|p_i - p_j\| \leq 2R'$, and derive that $R' \geq \frac{1}{2}S$.

From the analysis of the above two cases, we can conclude that $R' \geq \frac{1}{2}S$. Hence the iteration of estimating and removing could repeat for at most $\log_{\frac{3}{2}} \frac{2L}{S}$ times, where $L = \max_{i \neq j} \|p_i - p_j\|$ and $S = \min_{i \neq j} \|p_i - p_j\|$. For each iteration, we can remove at most 4^{γ} points, due to Corollary 7, and there are at most $\log_{\frac{3}{2}} \frac{2L}{S}$ iterations, therefore the ply $m \leq 4^{\gamma} \log_{\frac{3}{2}} \frac{2L}{S}$ for $k = 1$. We have proved the lemma for $k = 1$.

We consider the cases for $k > 1$. Without loss of generality, we assume that $B_1, \ldots, B_t$ contain p, and their corresponding centers are $p_1, \ldots, p_t$. Define a set $Q \subseteq P = \{p_1, \ldots, p_t\}$ by the following procedure. Initially, $Q = \phi$.

while $P \neq \phi$ **do**
  $Q = Q \cup \{q\}$, where q is the point in P with the largest k-nearest neighbor radius;
  $P = P - \text{int}(B_q)$, where B_q stands for the closed ball of $N_k(q)$.
end while

Because each ball covers only k points in its interior, we have $m \geq \lceil t/k \rceil$, where m denotes $|Q|$. Now we will show that for all $q \in Q$, $\text{int}(B_q) \cap Q = \{q\}$.

Suppose that $Q = \{q_1, \ldots, q_m\}$ such that for all $i < j$, q_i is put into Q in the procedure before q_j. Notice that for all $j < i$, $q_j \notin \text{int}(B_{q_i})$ because of the update operation for P in the algorithm. We can find that for all $j > i$, $q_j \notin \text{int}(B_{q_i})$, because the radius of the ball centered at q_i is larger than that of the ball centered at q_j. Therefore, we prove that $\text{int}(B_q) \cap Q = \{q\}$ for all $q \in Q$.

Thus, $m \le 4^\gamma \log_{\frac{3}{2}} \frac{2L}{S}$ which implies $t \le km \le k4^\gamma \log_{\frac{3}{2}} \frac{2L}{S}$. □

Corollary 18. *The max degree of a k-NNG in a metric space with doubling dimension γ is $O(k \log (L/S))$, where L and S are the longest and shortest distances in the graph.*

Proof. Suppose that $B_{N_k}(P)$ is the corresponding neighborhood system and p is a vertex in k-NNG. Then p is contained by at most $k \cdot 4^\gamma \log_{\frac{3}{2}} \frac{2L}{S}$ balls in $B_{N_k}(P)$ and p is connected to those centers. According to the definition of k-NNG, p is also connected to its k nearest neighbors. Hence the degree of vertex p is no more than $k4^\gamma \log_{\frac{3}{2}} \frac{2L}{S} + k$. Generally, the doubling dimension γ is a constant. Because p is chosen arbitrarily , we can see that the max degree in k-NNG is $O(k \log (L/S))$. □

Plotkin, Rao and Smith [14] gave the following theorem and showed that we can find a small size separator for the graph which excludes shallow minors.

Theorem 19. *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.*

We consider the neighborhood system $B_{N_k}(P)$ in a metric space with doubling dimension γ. Suppose that the corresponding k-NNG is G and the intersection graph of $B_{N_k}(P)$ is G', then we can see that G is a subgraph of G'. Applying Lemma 17 gives the ply bound of $B_{N_k}(P)$, therefore, $\forall l > 0$, G' excludes K_h as a depth l minor for $h > (k \cdot 4^\gamma \log_{3/2} \frac{2L}{S} + k + 1)(16l + 2)^\gamma > k(64l + 8)^\gamma \log_{\frac{3}{2}}(\frac{2L}{S})$ according to Theorem 14. Applying Theorem 19 gives the separator bound of G'. Because G is a subgraph of G', a separator of G' is also the separator of G. Therefore, Theorem 5 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-NNGs is $O(k \log(L/S))$, the above degree bound could also give an upper bound of the Fiedler value of a k-NNG of n points in a metric space with doubling dimension γ. Suppose that the sizes of two sides of the separation are x and αx, respectively, where $\alpha \ge 1$. Let Δ be the maximum degree of the k-NNG. Then we can set x_i for each vertex on the side with x vertices to be $\frac{1}{x}$; x_i for each vertex on the other side with αx vertices to be $\frac{-1}{\alpha x}$; and x_i for each vertex in the separator to be zero. Since the separator is a vertex separator, there is no edge between the two sides of the separation. Then we have the following inequality.

$$\begin{aligned}
\lambda_2 &\le \frac{\sum_{cut\ edge(i,j)} (x_i - x_j)^2}{\sum_{\forall i} x_i^2} \\
&\le \frac{(\frac{1}{x})^2 \times \Delta x + (\frac{1}{\alpha x})^2 \times \Delta \alpha x}{1/x + 1/(\alpha x)} \\
&= \Delta \\
&= O(k \log (L/S)).
\end{aligned}$$

The above results hold for spaces with segment property. When there is no segment property in the doubling dimensional space, we have the following similar results.

Lemma 20. *Suppose that $\{B_1, \ldots, B_n\}$ 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*

$$|\{i : B_i \cap B \ne \phi \text{ and } r_i \ge \beta r\}| \le \left\lceil \log_2 \frac{R}{\beta r} \right\rceil (k+1) 2^\gamma \left(5 + \frac{2}{\beta}\right)^\gamma$$

where r_i is the radius of B_i and R is the maximum radius in the system.

Proof. Let p be the center of B. Suppose that t balls with radius at least βr intersect B. Without loss of generality, we assume that those t balls are $B_1, \ldots, B_t$, which are divided into $m + 1$ classes shown in the table below. Here $m = \lceil \log_2 \frac{R}{\beta r} \rceil - 1$ and R is the largest radius in the neighborhood system.

We discuss two typical classes in the table and give the proof for numbers in the last column.

The proof for the first class is similar to that in Lemma 12. Suppose that there are t_1 balls in the first class, and let $B' = (1 + \frac{5\beta}{2})B$. Then we can shrink each ball B_i (with center p_i and radius r_i) to radius βr and let $p_i' = p_i$. Hence $\|p - p_i'\| = \|p - p_i\| \le (1 + 2\beta)r$, and the new ball B_i' is contained in B_i and substitution from B_i to B_i' does not increase the ply in the neighborhood system. Next we shrink each ball B_i' to its corresponding homocentric ball B_i^* with radius $\frac{\beta r}{2}$. We

Range of $\|p - p_i\|$	Radius of B_i	Radius of B_i^*	B'	Number of balls in B'
$[0, (1+2\beta)r]$	$\geq \beta r$	$\frac{\beta r}{2}$	$(1+\frac{5\beta}{2})B$	$(k+1)2^\gamma(5+\frac{2}{\beta})^\gamma$
$((1+2\beta)r, (1+4\beta)r]$	$> 2\beta r$	βr	$(1+5\beta)B$	$(k+1)2^\gamma(5+\frac{1}{\beta})^\gamma$
$((1+4\beta)r, (1+8\beta)r]$	$> 4\beta r$	$2\beta r$	$(1+10\beta)B$	$(k+1)2^\gamma(5+\frac{1}{2\beta})^\gamma$
$((1+8\beta)r, (1+16\beta)r]$	$> 8\beta r$	$4\beta r$	$(1+20\beta)B$	$(k+1)2^\gamma(5+\frac{1}{4\beta})^\gamma$
...	...	...	...	...
$((1+2^j\beta)r, (1+2^{j+1}\beta)r]$	$> 2^j\beta r$	$\frac{2^j\beta r}{2}$	$(1+2^{j-1}5\beta)B$	$(k+1)2^\gamma(5+\frac{1}{2^{j-1}\beta})^\gamma$
...	...	...	...	...
$((1+2^m\beta)r, (1+2^{m+1}\beta)r]$	$> 2^m\beta r$	$\frac{2^m}{2}\beta r$	$(1+2^{m-1}5\beta)B$	$(k+1)2^\gamma(5+\frac{1}{2^{m-1}\beta})^\gamma$

claim that for any ball B_i^*, there are no more than k balls B_j^* having overlap with it. Then we can color all these t_1 centers with $k+1$ colors so that every adjacent two centers have different colors. For each color set, there are some disjoint balls with radius $\frac{\beta r}{2}$ and contained in $B' = (1+\frac{5\beta}{2})B$. Applying Corollary 7, we can see that there at most $2^\gamma(5+\frac{2}{\beta})^\gamma$ disjoint balls in $(1+5\beta/2)B$ for each color set. Because there are at most $k+1$ different color sets, there are at most $(k+1)2^\gamma(5+\frac{2}{\beta})^\gamma$ balls intersecting the ball B in the first class.

Now we consider a general class, for example, the $(j+1)$th class. Suppose that there are t_{j+1} balls in the $(j+1)$th class, and let $B' = (1+2^{j-1}5\beta)B$. Because all these t_{j+1} balls intersect B and their centers are in the range $((1+2^j\beta)r, (1+2^{j+1}\beta)r]$, every ball in this class has radius at least $2^j\beta r$. Then we can shrink each ball B_i in this class (with center p_i and radius r_i) to radius $2^j\beta r$ and let $p_i' = p_i$. Hence $\|p-p_i'\| = \|p-p_i\| \leq (1+2^j\beta)r$, and the new ball B_i' is contained in B_i and substitution from B_i to B_i' does not increase the ply in the neighborhood system. Next we shrink each ball B_i' to its corresponding homocentric ball B_i^* with radius $2^{j-1}\beta r$. We claim that for any ball B_i^*, there are no more than k balls B_j^* having overlap with it. The proof is similar to that in Lemma 12. Then we can color all these t_{j+1} centers with $k+1$ colors so that every adjacent two centers have different colors. For each color set, there are some disjoint balls with radius $2^{j-1}\beta r$ and contained in $B' = (1+2^{j-1}5\beta)B$. Applying Corollary 7, we can see that there are at most $2^\gamma(5+\frac{1}{2^{j-1}\beta})^\gamma$ disjoint balls in $(1+2^{j-1}5\beta)B$ for each color set. Because there are at most $k+1$ different color sets, there are at most $(k+1)2^\gamma(5+\frac{1}{2^{j-1}\beta})^\gamma$ balls intersecting the ball B in the $(j+1)$th class.

Till now, we have proved all the upper bounds in the last column. Hence

$$|\{i : B_i \cap B \neq \phi \text{ and } r_i \geq \beta r\}| \leq t_1 + \cdots + t_{m+1} \leq \left\lceil \log_2 \frac{R}{\beta r} \right\rceil (k+1)2^\gamma \left(5 + \frac{2}{\beta}\right)^\gamma. \quad \square$$

With Lemma 20, we can derive the following theorems similarly as what we have done in Theorem 14, Theorem 16 and Theorem 5.

Theorem 21. *Suppose that Γ is a k-ply neighborhood system in a metric space with doubling dimension γ and G is the intersection graph of Γ. Then $\forall l > 0$, G excludes K_h as a depth l minor for $h \geq (k+1)\lceil \log_2 \frac{R(4l-1)}{r} \rceil (16l+6)^\gamma$, where R is the largest radius and r is the smallest one in the neighborhood system.*

Theorem 22. *Suppose that 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 $\forall l > 0$, G excludes K_h as a depth l minor for $h \geq (k+1)\lceil \log_2 \frac{R(4\alpha l-1)}{r} \rceil (16\alpha l+6)^\gamma$, where R is the largest radius and r is the smallest one in the neighborhood system.*

Theorem 23. *For every $l > 0$, if P is a set of n points in a metric space with doubling dimension γ, then the k-NNG of P has a vertex separator of size $O(k^2 l \log^2 \frac{L}{S}(\log l + \log \frac{L}{S})^2 (16l+6)^{2\gamma} \log n + \frac{n}{l})$, where L and S are, respectively, the maximum and minimum distances between any two points in P.*

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 section, 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 follows.

Theorem 24 (*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 simplified as $O(\frac{1}{n^{2/m}})$.

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 that the column vectors of $\mathbf{Q}$ are $(q_1, \ldots, q_n)$ 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 25. *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$(\mathbf{Q}) = d < m$, 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 $\gamma_d = 2(\pi+1+\frac{\pi}{\alpha})^2(\frac{A_{d+1}}{V_d})^{2/d}$,$A_d$ 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, we 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 25, we need the following preliminary lemmas.

Lemma 26. *Suppose that P is a set of point, and we can use SVD to find an approximate point set Q for P. For each $p_i \in P$, there is a corresponding point q_i in Q. Let R_i be the distance from p_i to its k-th nearest neighbor in P. And let r_i be the distance from q_i to its k-th nearest neighbor in Q. Then $R_i \leq r_i + 2\delta$.*

Proof. Suppose that q_j is the k-th nearest neighbor of q_i in Q. Hence $r_i = \|q_i - q_j\|$. We consider the following three cases.

Case 1. If the k-th nearest neighbor of p_i in P is just p_j, then

$$\begin{aligned} R_i &= \|p_i - p_j\| \\ &\leq \|p_i - q_i\| + \|q_i - q_j\| + \|p_j - q_j\|. \end{aligned}$$

Because $\delta = \max \|p_i - q_i\|$, we can get $R_i \leq r_i + 2\delta$.

Case 2. If the k-th nearest neighbor of p_i in P is p_l and $\|q_i - q_l\| \leq \|q_i - q_j\| = r_i$, then

$$\begin{aligned} R_i &= \|p_i - p_l\| \\ &\leq \|q_i - q_l\| + \|p_i - q_i\| + \|p_l - q_l\| \\ &\leq \|q_i - q_j\| + \|p_i - q_i\| + \|p_l - q_l\|. \end{aligned}$$

Because $\delta = \max \|p_i - q_i\|$, we can also get $R_i \leq r_i + 2\delta$.

Case 3. If the k-th nearest neighbor of p_i in P is p_l and $\|q_i - q_l\| > \|q_i - q_j\| = r_i$, then there must exist some $p_x \in P$, $p_x \notin N_k(p_i)$ and $\|q_i - q_x\| \leq \|q_i - q_j\| = r_i$. Therefore, we can get

$$\begin{aligned} R_i &\leq \|p_i - p_l\| \\ &\leq \|p_i - q_i\| + \|q_i - q_x\| + \|q_x - p_x\| \\ &\leq \|q_i - q_j\| + \|p_i - q_i\| + \|p_j - q_j\|. \end{aligned}$$

Because $\delta = \max \|p_i - q_i\|$, we can get $R_i \leq r_i + 2\delta$ finally. □

Lemma 27. *The 1-overlap graph of $B_{N_k}(P)$ in $\mathbb{R}^m$ is isomorphic to an edge subgraph of the $(1 + 7\delta/s)$-overlap graph of $B_{N_k}(Q)$ in $\mathbb{R}^d$.*

Proof. Suppose that G_p is the k-NNG of P in $\mathbb{R}^m$, G is the 1-overlap graph of $B_{N_k}(P)$, G_Q is the k-NNG of Q in $\mathbb{R}^d$, and G' is the $(1+\delta/s)$-overlap graph of $B_{N_k}(Q)$. If (p_i, p_j) is an arbitrary edge in G, then we have $\|p_i - p_j\| \leq R_i + R_j$.

$$\begin{aligned}\|q_i - q_j\| &\leq \|p_i - p_j\| + 2\delta \\ &\leq R_i + R_j + 2\delta \\ &\leq r_i + 2\delta + r_j + 2\delta + 2\delta \\ &= r_i + r_j + 6\delta.\end{aligned}$$

The first inequality is due to the triangle inequality, and the second one is due to Lemma 26.

Because $\delta = \max \|p_i - q_i\|$, we get $\delta \geq r_i$ and $\delta \geq r_j$. Therefore, we can see that $r_i + r_j + 6\delta \leq r_i + 7\delta = r_i(1 + 7\delta/r_i) \leq r_i(1 + 7\delta/s)$. We can also get $r_i + r_j + 6\delta \leq r_j(1 + 7\delta/s)$ in the similar way. Hence, the edge(q_i, q_j) must be contained in the graph G' as long as the edge (p_i, p_i) is contained in G. Because the choice of edge (p_i, p_j) is arbitrary and $(1 + 7\delta/s)$ is a constant, we prove that G is isomorphic to an edge subgraph of G'. □

Lemma 28. *Let $P = \{p_1, \ldots, p_n\}$ be a set of n points, and the corresponding ball system of $N_k(P)$ is $B_{N_k}(P) = \{B_{R(p_1)}(p_1), \ldots, B_{R(p_n)}(p_n)\}$. Then the k-NNG of P is a subgraph of the 1-overlap graph of $B_{N_k}(P)$.*

Proof. Suppose that G is a k-NNG of point set P, and G' is the 1-overlap graph of $B_{N_k}(P)$. Now we consider an arbitrary edge (p_i, p_j) in G. Since G is a k-NNG, we have $\|p_i - p_j\| \leq r_i$ or $\|p_i - p_j\| \leq r_j$. In addition, $\|p_i - p_j\| \leq r_i + r_j$. 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 of P is a subgraph of the 1-overlap graph of $B_{N_k}(P)$. □

Combining Lemmas 26–28, we can derive the following corollary.

Corollary 29. *The k-NNG of P in $\mathbb{R}^m$ is isomorphic to an edge subgraph of the $(1 + 7\delta/s)$-overlap graph of a ball system $B_{N_k}(Q)$ in $\mathbb{R}^d$, where $\delta = \max \|p_i - q_i\|$ and $s = \min \|q_i - q_j\|$.*

In [12] 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, then we can apply Theorem 24 and get the following corollary directly.

Corollary 30. *If G is a subgraph of the α-overlap graph of $B_{N_k}(P)$ 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 25.

Proof of Theorem 25. Suppose that G' is the $(1 + 7\delta/s)$-overlap graph of a ball system $B_{N_k}(Q)$ in $\mathbb{R}^d$. Applying Corollary 29, we can see that G is isomorphic to a subgraph of G'. Because the isomorphic graphs have the same Laplacian matrices, G and the subgraph of G' have the same Fiedler value.

In addition, the edge subgraph of G' has no larger Fiedler value than G' according to Proposition 2. Hence the Fiedler value of G is no larger than the Fiedler value of G'.

Replacing α with $1 + 7\delta/s$ in Corollary 30, we get an upper bound for the Fiedler value of $L(\mathbf{P})$. Hence the Fiedler value 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 study the combinatorial and spectral aspects of nearest neighbor graphs in doubling dimensional metric spaces and in nearly-Euclidean spaces. We obtain bounds on the largest possible minor and bounds on the size of vertex separators. For those graphs in nearly-Euclidean spaces with high dimension, we prove that 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.

Acknowledgements

We thank the anonymous reviewer(s) for the helpful suggestions and comments, which make the paper more solid and widely applicable.

Part of first author's 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, 2007CB311003), the National Natural Science Foundation of China Grant (60553001, 60621062) and the National Basic Research Program of China Grant (2007CB807900, 2007CB807901).

The second author was partly supported by the NSF grants CCR-0635102 and ITR CCR-0325630. Part of this work was done while visiting Tsinghua University and Microsoft Research Asia Lab.

References

[1] E.R. Barnes, A.J. Hoffman, Partitioning, spectra and linear programming, Progress in Combinatorial Optimization (1984) 13–25.
[2] P. Dewilde, E.F. Deprettere, Singular value decomposition: An introduction, in: Ed.F. Deprettere (Ed.), SVD and Signal Processing: Algorithms, Applications, and Architectures, pages Elsevier Science Publishers, North Holland.
[3] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (1936) 211–218.
[4] C. Eckart, G. Young, A principal axis transformation for non-Hermitian matrices, Bulletin of the American Mathematical Society 45 (1939) 118–121.
[5] T. Feder, P. Hell, S. Klein, R. Motwani, Complexity of graph partition problems, in: STOC, 1999, pp. 464–472.
[6] M. Fiedler, Algebraic connectivity of graphs, Czechoslovak Mathematical Journal 23 (98) (1973) 298–305.
[7] Gene H. Golub, Gharles F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, Maryland, 1992, pp. 16–21, 293.
[8] David R. Karger, Matthias Ruhl, Finding nearest neighbors in growth-restricted metrics, in: STOC 2002, pp. 741–750.
[9] Richard J. Lipton, Robert Endre Tarjan, A separator theorem for planar graphs, SIAM Journal on Applied Mathematics 36 (1979) 177–189.
[10] Richard J. Lipton, Robert Endre Tarjan, Applications of a planar separator theorem, SIAM Journal on Computing 9 (1980) 615–627.
[11] Michael Luby, Removing randomness in parallel computation without a processor penalty, Journal of Computer and System Science 47 (1993) 250–286.
[12] Gary L. Miller, Shang-Hua Teng, William P. Thurston, Stephen A. Vavasis, Separators for sphere-packings and nearest neighbor graphs, Journal of the ACM 44 (1) (1997) 1–29.
[13] A. Pothen, H.D. Simon, K.-P. Liou, Partitioning sparse matrices with eigenvectors of graphs, SIAM Journal on Matrix Analysis and Applications 11 (1990) 430–452.
[14] S. Plotkin, S. Rao, W.D. Smith, Shallow excluded minors and improved graph decomposition, in: Proceedings at the 5th Symposium Discrete Algorithms, SIAM, New York, 1994, pp. 462–470.
[15] Michael Ruzhansky, On uniform properties of doubling measures, Proceedings of American Mathematical Society 129 (2001) 3413–3416.
[16] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener, Mathematische Annalen 63 (1907) 433–476.
[17] Daniel A. Spielman, Shang-Hua Teng, Spectral partitioning works: Planar graphs and finite element meshes, in: FOCS 1996, pp. 96–105.
[18] Shang-Hua Teng, Combinatorial aspects of geometric graphs, Computational Geometry (9) (1998) 277–287.