Notes and Comments for [1] *

Zhang Qin

March 3, 2007

1 Problems and Results

Given an undirected graph $G = (V, E), |V| = n, |E| = m$ The *Minimum Bandwidth Problem* is to find a one-to-one mapping $f : V(G) \stackrel{1-1}{\rightarrow} [n]$ to minimize

$$bw(f) = max_{(i,j)\in E}|f(i) - f(j)|$$

This is a NP-hard problem. This paper presents a randomized algorithm that runs in nearly linear time and outputs a linear arrangement whose bandwidth is within a $O(polylog(n))$ factor of optimal.

2 Intuition and General Idea

The basic idea to solve the bandwidth problem is: we first randomly map the points in V to real line $\mathbb{R}$, denote the mapping ψ. Such an one to one mapping ψ defines a natural ordering on the points according to their positions in the real line. We then use this ordering to get a mapping $f : V(G) \rightarrow [n]$. To achieve a good ordering, we want to maintain the following two properties. Assume that the real line is already divided into intervals of length l, where l will be specified later.

1. Let $S \subseteq V$ and $|S| = k$. We say S is bad if $\psi(S) \subseteq [tl, (t + 1)l)$ for some integer t, that is, all the points in S are mapped into a narrow interval. We want to make sure that the chance of *"S is bad"* is small. In other words, we hope that the points are well separated after the mapping.

2. We hope that the image $\psi(v_i), \psi(v_j)$ of two endpoints of any edge $e = (v_i, v_j) \in E$ are not far apart.

*Part of this note based on "scribed lecture notes of a course at CMU".

Notice that the final goal of our mapping is trying to minimize the number of points between $\psi(v_i), \psi(v_j)$ where $(v_i, v_j) \in E$. If we can guarantee the two properties above, we can bound the number of points between the image of any pair of endpoints effectively, thus solve the problem.

Feige's method to achieve those two properties relies on a technique called **volume respecting embedding**. The essence behind this technique is trying to find an **(η, k)-well-separated mapping** $\phi : V \to \mathbb{R}^L$. A contracting mapping $\phi : V \to \mathbb{R}^L$ is called (η, k)-well-separated if the following condition holds.

For each set $S \subseteq V, s.t. |S| = k$, there exists an ordering $\{s_0, s_1, \ldots, s_{k-1}\}$ of S such that, for all i, let $L_i = span\{\phi(s_0), \phi(s_1), \ldots, \phi(s_{i-1})\}$, and

$$dist(\phi(s_i), L_i) \geq \frac{q_i}{\eta}$$

where $q_i = d_G(s_i, \{s_0, s_1, \ldots, s_{i-1}\})$ is the minimum length between s_i and some point in $\{s_0, s_1, \ldots, s_{i-1}\}$. Once such a mapping is found, we project the points $(\phi(v) : v \in V)$ in $\mathbb{R}^L$ to a random line using a function φ. Now $\psi = \varphi \circ \phi$.

We can prove that such a transformation is good for our purpose. An intuitive imagination might helps to illustrate the point: the simplex (convex hull of $\phi(s_i)$) obtained by the well-separated mapping must be "fat" (i.e. has a large volume), and there is a big chance to obtain a well separated points configuration after the line projection. Meanwhile, since the mapping is contracting, the length of the vectors corresponding to all edges after the mapping would be less than 1. Therefore the final length of its image after the random projection φ will not be very large.

The remain questions are:

- How to find such a well-separated mapping ϕ? What is the least η such that there exist such a mapping? The value of η is important because it related with the approximation ratio we could get for the problem. And also notice that an (η, k)-well-separated mapping is an η-distortion embedding.

- How to use such a well-separated mapping to prove the first property? Note that the second property is comparatively easier and mainly relies on φ.

3 The Framework of the Method

In this section, we will formalize the idea given in the previous section, and provide the framework of the approximation procedure.

As always, we first need to find a lower bound of the minimum bandwidth. And here comes the notion of *local density*.

Definition 3.1 *The local density D of graph G is defined as*

$$D = D(G) = \max_{v,r} \frac{|B(v, r)| - 1}{2r}$$

where $B(v, r)$ is the ball centered at v containing all the vertices that are within distance r from v, include v itself.

One can easily verify that this is a lower bound of the minimum bandwidth by using the pigeon hole principle.

The two properties described above could be formalized as follows:

Claim 3.1

$$Pr[S \text{ is bad }] \leq \frac{O(\eta l)^{|S|-1}}{Tvol(S)}$$

Claim 3.2

$$Pr[|\overline{\psi(v_i)\psi(v_j)}| \geq l : e = (v_i, v_j) \in E] \leq \frac{1}{2m}$$

The $Tvol(S)$ is an auxiliary concept defined as follows.

Definition 3.2 *Let $S \subseteq V$. Consider the complete graph G_S on S [1], and let the length of an edge d_e for $e = (u, v) \in S \times S$. We define*

$$Tvol = \prod_{e \in T} d_e$$

where T is the minimum spanning tree on G_S.

The reason to introduce $Tvol(S)$ lies in that we can further bound $\sum_{S \subseteq V : |S| = k} \frac{1}{Tvol(S)}$ using an expression containing the lower bound D.

Claim 3.3

$$\sum_{S \subseteq V : |S| = k} \frac{1}{Tvol(S)} \leq n \cdot O(D \log n)^{k-1}$$

Claim 3.1 and 3.3 give us

$$E[\#(\text{bad sets})] \leq n \cdot O(\eta l D \log n)^{k-1}$$

Using Markov's inequality, one can conclude that

$$Pr[\#(\text{bad sets}) \leq c \cdot n \cdot O(\eta l D \log n)^{k-1}] \geq \frac{1}{c}$$

for some constant c. On the other hand, we can conclude from Claim 3.2 that with probability more than $\frac{1}{2}$, all the edges' length are no more than l.

Now the method of *Double Counting* comes into the play. Suppose the bandwidth is B, achieved by edge e, the edge with length no more than l spanning at

[1] Given a set of vertices $S \subseteq V$, let d_S be the induced metric in S. Then (S, d_S) can be viewed as the complete graph with the edge (s_1, s_2) having weight $d_S(s_1, s_2)$.

most 2 intervals. Then at least $B/2$ in-between vertices lies in one interval, and these vertices contribute $\binom{\frac{B}{2}}{k}$ bad sets. Therefore

$$\#\text{bad sets} \geq \binom{\frac{B}{2}}{k} \geq \left(\frac{B}{2k}\right)^k$$

Combine the upper bound and lower bound, we obtain

$$\left(\frac{B}{2k}\right)^k \leq c \cdot n \cdot O(\eta l D \log n)^{k-1} \Rightarrow B \leq (3n)^{1/k} \cdot O(k\eta l D \log n)$$

Set $k = \log n$, we have

$$B \leq O(D\eta l \log^2 n)$$

We can further set $l = \Theta(\sqrt{\log n})$ and $\eta = \Theta(\log^{3/2} n)$ (could be improved to $\eta = \Theta(\log n\sqrt{\log\log n})$), and then obtain an $O(polylog(n))$ approximation ratio.

The proof of Claim 3.2 is comparatively easier, we give it below. We also give an intuitive (and far from rigorous) proof for Claim 3.3. The proof of Claim 3.1 as well as how to find a well-separated mapping (This would be the backbone in the proof of Claim 3.1) will be given at section 5.

Proof of Claim 3.2.

We pick a random vector $\vec{r} = \{r_1, r_2, \ldots, r_L\}$ in $\mathbb{R}^L$ and each $r_i \sim N(0,1)$. Then for a vector $\vec{v} = \overline{\phi(v_i)\phi(v_j)}$ of at most unit length (recall that ϕ is a contracting mapping), we have

$$Pr[|\langle \vec{r}, \vec{v}\rangle| > 2(\sqrt{\log n})] = Pr[\|\vec{v}\|_2 \times |X| > 2(\sqrt{\log n})] \leq Pr[|X| > 2(\sqrt{\log n})]$$

where X is a $N(0,1)$ random variable. Using the fact that $Pr[|X| > t] \leq \frac{2}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$, we have

$$Pr[|\langle \vec{r}, \vec{v}\rangle| > 2(\sqrt{\log n})] \leq \frac{2}{2\sqrt{\log n}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(2\sqrt{\log n})^2} < \frac{1}{n^2} < \frac{1}{2m}$$

An Intuitive Proof of Claim 3.3.

From the definition of *local density* we know that at most n/k vertices at distance at most $n/(k \cdot 2D)$ from v. Hence, when selecting k vertices randomly to form the set S, the expected distance from v to the closest other vertex in S would be $\Omega(n/(kD))$. And then $Tvol(S)$ would be $\Omega((n/kD)^{k-1})$. Finally we have

$$\sum_{S \subseteq V : |S| = k} \frac{1}{Tvol(S)} \leq \binom{n}{k}((n/kD)^{k-1}) < \frac{n}{k}(eD)^{k-1}$$

Almost matches the bound shown above. The more accurate bound needs a formal and a bit complicated proof, which is omitted here.

4 Algorithm and Explanation

The general algorithm is described as follows.

Algorithm Bandwidth(G)

1. Embed G to R^L using an (η, k)-well-separated mapping, where $L = c \log n \log D$, c is some constant and D is the diameter of the original graph.

2. Project all the vertices in the embedding on a random line, obtain for each vertex a point $h(v)$ on the line.

3. Sort $h(v)$, and output the sorted list of vertices as the linear arrangement.

The well-separated mapping could be found by the following algorithms.

Algorithm Well_Separated_Map(G)

1. Let $L = c \log n \log D$

2. For $t = 1$ to $\log D$

   (a) Let $\Delta = 2^t$

   (b) For $j = 1$ to $ck \log n$
   $f_{tj} =$ **Generate_Coordinate**(G, Δ)

3. Let $f = \oplus_{tj} f_{tj}$

4. Let $\phi = f/\sqrt{4L}$ for all t, j. /* *to preserve "contracting"* */

The algorithm Generate_Coordinate(G, Δ) maps all the vertices $v \in V$ into the real line, with coordinate $f_{tj}(v)$.

Algorithm Generate_Coordinate(G, Δ)

1. Let $G' = G$, $S_{tj} = \emptyset$.

2. While $G' \neq \emptyset$ do

   (a) Pick up an arbitrary vertices $v \in G'$, and build a BFS tree rooted at v. Let $l(u)$ denotes the distance (level) from a vertex $u \in G'$ to v.

   (b) Let $r = \Delta/(4 \log n)$.

   (c) Define layers $Lay_k = \{u \mid l(u) \in [(k-1)r, kr)\}$ for all possible k, such that every layer contains r levels.

   (d) Pick one of the layers randomly, with the k^{th} layer chosen with probability $p_k = 2^{-k}$, and pick a level l uniformly at random within that layer.

(e) Add all the vertices at distance l to S_{tj} and delete them as well as the whole component C bounded by those vertices from G'.

3. Choose a parameter $\gamma_{tj}(C) \in [1, 2]$ independently, uniformly at random for all the components resulting from the above decomposition. Assign the value $f_{tj}(v)$ for a vertex $v \in C$ with $f_{tj}(v) = \gamma_{tj} \cdot d_G(S_{tj}, v)$.

A few explanations.

1. The meaning of the two subscripts t, j in f_{tj}.

- t: As shown in the algorithm Well_Separated_Map(G), the value of t denotes the value $\Delta \in \{1, 2, 4, \ldots, D\}$. Obviously, each q_i must fall into a particular $[\Delta, 2\Delta]$. Therefore if we can prove that $dist(\phi(s_i), L_i)$ will be large (say, at least q_i/η for some small η) in many tj^{th} coordinates when j varies, we are done.
- j: the $ck \log n$ copies of js is used to guarantee that with high possibility (by using of Chernoff bound), at least a constant fractional of tj^{th}s will be large.

5 Major Techniques and Key Proofs

5.1 Proof of Claim 3.1

We first assume that the ϕ constructed by algorithm Well_Separated_Map(G) is such a well-separated mapping , and provide a proof for Claim 3.1.

We fix a set $S = \{s_0, s_1, \ldots, s_{k-1}\}$. Recall that $\phi(s_i)$ is the image after the embedding and $L_i = span\{\phi(s_0), \phi(s_1), \ldots, \phi(s_{i-1})\}$. W.l.o.g, we assume that $s_0 = \vec{0}$, and $L_i = span\{\vec{e_1}, \vec{e_2}, \ldots, \vec{e_{i-1}}\}$ is the subspace spanned by the first $i-1$ basis vector. Let $\phi(s_i) = (s_{i1}, s_{i2}, \ldots, s_{ii}, 0, \ldots, 0)$. Here comes the key observation

$$(\eta, k)\text{-well-separatedness} \Rightarrow s_{ii} \geq \frac{q_i}{\eta}$$

Next we define

$$\psi(x) = \langle \phi(x), \vec{r} \rangle$$

where $\vec{r} = (r_1, \ldots, r_k)$ and $r_i \sim N(0, 1)$ Then we have the following expression, let $I = [0, l)$.

$$\begin{aligned} Pr[\psi(S) \subseteq I] &= Pr[\psi(s_0) \in I] \times Pr[\psi(s_1) \in I | \psi(s_0) \in I] \times \\ & \quad \ldots \times Pr[\psi(s_{k-1}) \in I | \wedge_{j=1}^{k-2} \psi(s_j) \in I] \end{aligned}$$

Consider the expression

$$Pr[\psi(s_i) \in I | \wedge_{j=1}^{i-1} \psi(s_j) \in I \wedge (r_1 = \widehat{r_1} \wedge \ldots \wedge r_{i-1} = \widehat{r_{i-1}})]$$

Let $Z = \sum_{j<i} s_{ij}\widehat{r_j}$. If $\psi(s_i)$ fall into the interval I, it must be the case that $s_{ii}r_i \in [-Z, l-Z)$ Since the value r_i is independent of all the conditions, we have $s_{ii}r_i \sim N(0, s_{ii}^2)$ with $s_{ii} \geq q_i/\eta$, and hence[2]

$$Pr[\psi(s_i) \in I| \wedge_{j=1}^{i-1} \psi(s_j) \in I \wedge (r_1 = \widehat{r_1} \wedge \ldots \wedge r_{i-1} = \widehat{r_{i-1}})] \leq \frac{\eta l}{\sqrt{2\pi} q_i}$$

Since the inequality holds for every possible value of $r_j (j < i)$, we have

$$Pr[\psi(s_i) \in I| \wedge_{j=1}^{i-1} \psi(s_j) \in I] \leq \frac{\eta l}{\sqrt{2\pi} q_i}$$

and hence

$$Pr[\psi(S) \in I] \leq \prod_{i=1}^{k-1} \frac{\eta l}{\sqrt{2\pi} q_i} \leq \frac{O(\eta l)^{k-1}}{Tvol(S)}$$

A proof leak!

Notice that here we choose a random vector $\vec{r}$ in space R^k. But we want to get a mapping from R^L to the real line. And intuitively, a vector would be shrunk by a random projection from R^L to R^k. Fortunately, the following proposition comes to rescue.

Proposition 5.1 *Let r be a random unit vector in R^L chosen with spherical symmetry, and let R^k be a subspace of R^L. Let l denote the length of the projection of r on R^k. Then:*

- Small projection: *for every $0 < \epsilon < 1$, $Pr_r[l < \epsilon\sqrt{k}/\sqrt{L}] \leq (\beta\epsilon)^k$, for some universal $\beta > 0$.*
- Large projection: *for $c > 1$ and $k = 1$, $Pr_r[l > \sqrt{c/L}] \leq e^{-c/4}$. When L is large, the exponent tends to $-c/2$.*

5.2 Correctness of the Well-Separated mapping Construction

The proof make use of the following observation.

Claim 5.1 *Algorithm Well_Separated_Map(G) is a randomized construction such that for an arbitrary point $x \in L_i$,*

$$Pr\left[\|\phi(s_i) - x\| > 3\frac{q_i}{\eta}\right] \geq 1 - n^{-3k}$$

where η is chosen to be $\log^{3/2} n$ (A relative weak but simple version)

[2]According to the fact that provided $X \sim N(0, \hat{\sigma}^2)$ with $\hat{\sigma} \geq \sigma$ and I be an interval of length l in the real line, we have $Pr[X \in I] \leq \frac{l}{\sqrt{2\pi}\sigma}$

Why we need such a claim? The intuition is that if we choose an $2/\eta$-net[3] from a ball with radius n that contains the subspace L_i, the number of points of the $2/\eta$-net would be at most $(n\eta/2)^{i-1} = O(n^{2k})$ by a volume argument. And then we can claim that with high probability $(1 - n^{-k})$, x is at least $3q_i/\eta$ away from each point in the $2/\eta$-net. Thus for every point y in the space L_i, suppose x is the closest point to y in the $2/\eta$-net, we have (by the Pythagoras' inequality)

$$\|\phi(s_i) - y\|^2 \geq \|\phi(s_i) - x\|^2 - \|x - y\|^2 \geq \left(\frac{3q_i}{\eta}\right)^2 - \left(\frac{2}{\eta}\right)^2 \geq \left(\frac{q_i}{\eta}\right)^2$$

The rest of our job is to prove Claim 5.1.

Our task is trying to prove that $\phi(s_i)$ is far away from the subspace spanned by $\phi(s_0), \phi(s_1), \ldots, \phi(s_{i-1})$. As each point x in that subspace could be expressed as $x = \sum_{j<i} \lambda_j \phi(s_j)$ with $\sum_{j<i} \lambda_i = 1$, we approach this in 2 steps. We first prove that $f_{tj}(s_i)$ is far away from $f_{tj}(s_j)$ for every s_j $(j < i)$, and then use the variational parameter γ_{tj} to prove that $f_{tj}(s_i)$ is far from $\lambda_j f_{tj}(s_j)$ for every s_j $(j < i)$ at a constant fractional of coordinates.

We say a particular coordinate f_{tj} created by the algorithm is **eligible** if all the components created during its creation have a diameter of at most Δ. For a given set S_{tj} defining a coordinate f_{tj}, we say a node u is **δ-good** if $d(S_{tj}, u) \geq \delta$. We have following observations.

Observation 5.1 *A coordinate is eligible with probability* $1 - \frac{1}{n}$

Observation 5.2 *Each node* $u \in G$ *is* $r = \Delta/(4 \log n)$*-good with constant probability.*

To prove the first observation, we just need to notice that if a coordinate is not eligible, we must choose a level greater than $\Delta/2$, which lies in a layer greater than $(2 \log n)$. But this layer is chosen with probability at most $2^{-2 \log n} = 1/n^2$. Meanwhile, there are at most n components at that time.

To prove the second observation. Suppose that at a particular time, $v \in G$ is chosen to be the root, u lies in the neighborhood of v and a cut happens, see Figure 1 (b). If the cut falls into a level between 0 to $l(u) - r$, it will not effect u, and u will be cut off sometime later. If the cut falls into a level between $l(u) - r$ to $l(u) + r$, we call it *bad* since $d(S_{tj}, u)$ is less than δ. If the cut falls beyond the level $l(u) + r$, we call it *good* since u will be cut off this time and it will be at least δ far away from the boundary S_{tj} forever. Therefore u is δ-good if and only if a good cut happens before a bad one (and the bad one will never happen). And one can easily see that this is a constant, by our choices of p_ks.

Now it is time to prove Claim 5.1, finally!

[3] An η-net is a maximal set of points in the space such that the distance between each pair of points is at least η

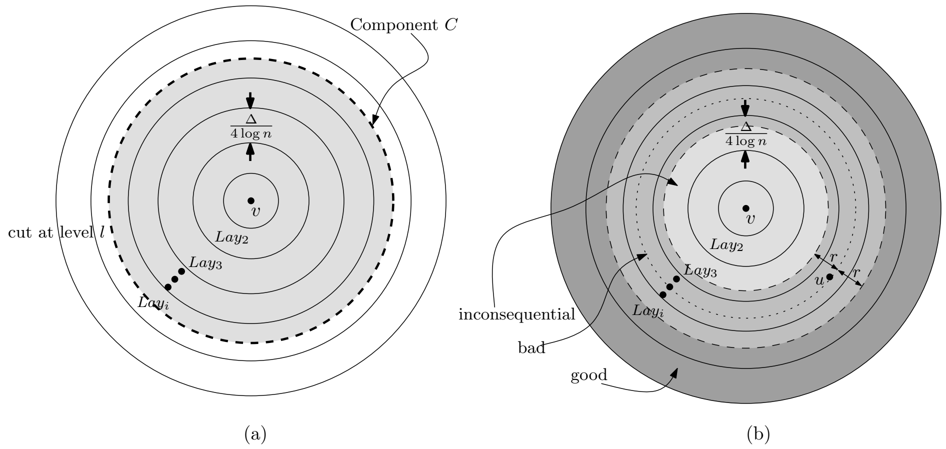

Figure 1: The embedding and coordinate generating.

For point s_i, we choose the coordinate f_{tj}s created when the procedure "Generate_Coordinate(G, Δ)" is called with $\Delta \in [q_i/2, q_i]$. We notice that with some constant possibility β, a coordinate $f_{tj}(s_i)$ is both eligible and $\Delta/(4\log n)$-good. The fact that coordinate is eligible means that s_i lies in a component C different from all the s_j $j < i$ during that decomposition procedure. And the coordinate is also $\Delta/(4\log n)$-good, which means that $f_{tj}(s_i)$ is at least $\gamma_{tj}(C) \cdot \Delta/(4\log n)$ away from all the $f_{tj}(s_j)$ $(j < i)$ at that coordinate. By the fact that $\gamma_{tj}(c)$ varies in the interval $[1, 2]$, we can conclude that $f_{tj}(s_i)$ is at least $\Delta/(12\log n)$ away from $\lambda_j f_{tj}(s_j)$ with possibility $\beta/3$. Since there are $ck\log n$ copies of j, we conclude that there is a constant fraction of coordinates that contribute to $\|\phi(s_i) - x\|$ with high probability ($(1 - n^{-3k})$ if we choose a suitable c and use the Chernoff bound). Therefore, after dividing by $\sqrt{4L} = \sqrt{4ck\log n \log D}$, we conclude that

$$\|\phi(s_i) - x\| \geq O(\sqrt{\left(\frac{\Delta}{12\log n}\right)^2 (ck\log n)} \cdot \frac{1}{\sqrt{4L}}) = O(\frac{q_i}{\log^{3/2} n})$$

6 Conclusion and Comments

Such a long journey to achieve the goal!

References

[1] Uriel Feige. Approximating the bandwidth via volume respecting embeddings (extended abstract). pages 90–99, 1998.