12 Improved Embeddings from cutting schemes (Measured descent)

In this section we show how to improve on the result from the previous lecture by giving an embedding into ℓ_1 with distortion $\sqrt{\log n/\beta}$. We first introduce some terminology.

Given a distance scale i and a partition P_i of the graph (where the diameter of a component is at most 2^i), let $C_i(v)$ denote the component containing a vertex $v \in V$. We say that a pair (x, y) is *δ-separated* by the partition P_i if

- the vertices x and y lie in different components; i.e., $C_i(x) \neq C_i(y)$, and
- both of x and y are "far from the boundary of their components", i.e., $d(x, V \setminus C_i(x)) \geq \delta \cdot d(x, y)$ and $d(y, V \setminus C_i(y)) \geq \delta \cdot d(x, y)$.

A *decomposition suite* Π is a collection $\{P_i\}$ of partitions, one for each distance scale i between 1 and $\lfloor \log \Delta \rfloor$. Given a *separation function* $\alpha(x, y) : V \times V \to [0, 1]$, the decomposition suite Π is said to *$\alpha(x, y)$-separate* (x, y) if for the distance scale i such that $2^i \leq d(x, y) \leq 2^{i+1}$, (x, y) is $\alpha(x, y)$-separated by the corresponding partition $P_i \in \Pi$.

Finally, a *$\alpha(x, y)$-decomposition bundle* is a collection $\{\Pi_j\}$ of decomposition suites such that for each (x, y), at least a constant fraction of the Π_j $\alpha(x, y)$-separate the pair (x, y).

Note that these definitions about decomposition suites and bundles are in some sense an extension of a cutting scheme. By using a cutting scheme for every distance scale we get a decomposition suite where nodes are β-separated with constant probability. In this sense a cutting scheme is a randomized procedure for constructing decomposition suites and can therefore be viewed as a decomposition bundle. Note that one difference between cutting schemes and decomposition bundles is that the separation parameter may depend on the node-pair (x, y).

Theorem 12.1 *Given an $\alpha(x, y)$-decomposition bundle for the metric d, there exists a randomized contracting embedding $\varphi : V \longrightarrow \ell_2$, such that for each pair (x, y),*

$$||\varphi(x) - \varphi(y)||_2 \geq \Omega\left(\sqrt{\frac{\alpha(x, y)}{\log n}}\right) \cdot d(x, y) \quad .$$

We define a measure of "local expansion". Let

$$V(x, y) = \max\left\{\log \frac{|B(x, 2d(x, y))|}{|B(x, d(x, y)/8)|}, \log \frac{|B(y, 2d(x, y))|}{|B(y, d(x, y)/8)|}\right\}$$

where $B(x, r)$ denotes the set of nodes within the ball of radius r around x. We will derive Theorem 12.1 from the following lemma.

Lemma 12.2 *Given an $\gamma(x,y)$-decomposition bundle, there is a randomized contracting embedding $\varphi : V \longrightarrow \ell_2$ such that for every pair (x,y) with constant probability*

$$||\varphi(x)-\varphi(y)||_2 \geq \Omega\left(\sqrt{\frac{V(x,y)}{\log n}} \cdot \gamma(x,y)\right) \cdot d(x,y) \ .$$

By repeatedly applying Lemma 12.2, we obtain the following guarantee:

Corollary 12.3 *Given a $\gamma(x,y)$-decomposition bundle, there is a randomized contracting embedding $\varphi : V \longrightarrow \ell_2$ such that for every pair (x,y),*

$$||\varphi(x)-\varphi(y)||_2 \geq \Omega\left(\sqrt{\frac{V(x,y)}{\log n}} \cdot \gamma(x,y)\right) \cdot d(x,y) \ .$$

Proof. The corollary follows by applying Lemma 12.2 repeatedly and independently for each decomposition suite several times. Then concatenating and rescaling the resulting maps gives with high probability an embedding that fulfills the corollary. ∎

To see that the above corollary implies Theorem 12.1, we observe that the cutting scheme that was used for the FRT-result has the property that with constant probability a pair (x,y) is $\Omega(1/V(x,y))$-separated in the decomposition bundle that results from the cutting scheme. (Recall that in FRT the probability for cutting the ball $B(x,\rho)$ was bounded by $\log(\frac{|B(x,2\delta)|}{|B(x,\delta/8)|}) \cdot \frac{1}{\delta}$ when creating cluster of diameter at most δ. This also gives a decomposition bundle with this parameter).

Applying the corollary to this decomposition bundle, we get an embedding φ_1, such that

$$||\varphi_1(x)-\varphi_1(y)||_2 \geq \Omega\left(\frac{1}{\sqrt{V(x,y)\cdot \log n}}\right) \cdot d(x,y) \ .$$

Applying the corollary to the decomposition bundle assumed by Theorem 12.1, gives an embedding φ_2 with

$$||\varphi_2(x)-\varphi_2(y)||_2 \geq \Omega\left(\sqrt{\frac{V(x,y)}{\log n}} \cdot \alpha(x,y)\right) \cdot d(x,y) \ .$$

Concatenating the two mappings and rescaling, we get a contracting embedding $\varphi = \frac{1}{\sqrt{2}}(\varphi_1 \otimes \varphi_2)$, with

$$\begin{aligned}||\varphi(x)-\varphi(y)||_2 &\geq \Omega\left(\frac{1}{\sqrt{\log n}} \cdot (\frac{1}{V(x,y)^{\frac{1}{2}}} + V(x,y)^{\frac{1}{2}}\ \alpha(x,y))\right) \cdot d(x,y) \\ &\geq \Omega\left(\sqrt{\frac{\alpha(x,y)}{\log n}}\right) \cdot d(x,y)\end{aligned}$$

as desired (note that $\frac{1}{z} + z\alpha \leq \sqrt{\alpha}$ for any value of z). Now it remains to prove Lemma 12.2.

12.1 The embedding

Let $T = \{1, \ldots, \log \Delta\}$ and $Q = \{0, \ldots, m-1\}$, for some suitably chosen constant m. In the following we define an embedding into $|T| \cdot |Q|$ dimensions. For $t \in T$, let $\mathrm{r}_t(x)$ denote the minimum radius r such that the ball $B(x, r)$ contains at least 2^t terminal nodes. We call $\mathrm{r}_t(x)$ the *t-radius of* x. Further, let $\ell_t(x) \in \mathbb{N}$ denote the distance class this radius belongs to (i.e., $2^{\ell_t(x)-1} \leq \mathrm{r}_t(x) \leq 2^{\ell_t(x)}$).

Pick a decomposition suite $\Pi = \{P_s\}$ from the decomposition bundle at random. In the following $\delta(x, y)$ denotes the separation-factor between x and y in this suite, i.e., $\delta(x, y) = \frac{1}{d(x,y)} \min\{d(x, V \setminus C_s(x)), d(y, V \setminus C_s(y))\}$ if $C_s(y) \neq C_s(x)$ and 0, otherwise. Observe that with constant probability we have $\delta(x, y) \geq \alpha(x, y)$. We color all the clusters in the decomposition suite uniformly at random with either "0" or "1".

In order to improve on the results from the previous lecture, the goal is to construct an embedding φ in which the distance between $(\varphi(x), \varphi(y))$ increases as the local expansion $V(x, y)$ increases. This means we not only want that x, y are far apart for a $\Theta(1/\log n)$-fraction of coordinates (this was used in the previous lecture) but for a $\Theta(V(x, y)/\log n)$-fraction.

One (stupid) way to obtain this for a pair (x, y) is to create a coordinate for every $t \in T$ and then embed points in this coordinate according to the partitioning for the distance scale $\ell_t(x)$. If $V(x, y)$ is large this means that for many values of t, $\mathrm{r}_t(x)$ (the length corresponding to distance scale $\ell_t(x)$) will be approximately the distance between x and y (indeed this holds for approximately $V(x, y)$ values of t). Therefore in total this embedding will obtain for a $\Theta(V(x, y)/\log n)$-fraction of coordinates a distance of $\Omega(\alpha(x, y) \cdot d(x, y))$ between x and y. This means that this embedding makes the node x happy since all its distances to other nodes are not distorted by too much; however this is mainly based on the fact that the embedding was specifically designed for x in the sense that *all* nodes choose their color from the partitioning on level $\ell_t(x)$.

We use a different approach now. We want to construct coordinates for every $t \in T$ and then a point x chooses a color according to the partitioning for the distance scale $\ell_t(x)$ (i.e., different points derive their color from partitionings for different distance scales depending on their local expansion).

However, now if we define the set S_t (the subset for the t-th coordinate),as the nodes x that are colored 1 in the partitioning for scale $\ell_t(x)$ we cannot argue that for a pair (x, y) either $d(x, S_t)$ or $d(y, S_t)$ is large, because nodes u very close to x or y may have distance scales $\ell_t(u)$ that are different from $\ell_t(x)$ or $\ell_t(y)$. In order to ensure local consistency such that all nodes close to x obtain their color from the same partitioning, we construct several coordinates in the embedding for every t, such that for each distance scale $\ell_t(x)$ there is a coordinate in which all nodes close to x derive their color from the partitioning for scale $\ell_t(x)$. The details are as follows.

Let $Q = \{0, \cdots, m-1\}$ denote the set of indices of coordinates corresponding to each value of t. For each $q \in Q$, we partition the distance scales into *groups* g_q of size m each, and let the median scale in each group represent that group for the coordinate corresponding to q. In the $(q, t)^{\text{th}}$ coordinate, the color of a node is picked according to the median distance scale in the group g_q to which $\ell_t(x)$ belongs.

In particular, let $g_q(\ell) := \lceil \frac{\ell - q}{m} \rceil$. Note that each distance group contains (at most) m consecutive distance classes which means that distances within a group differ at most by a constant factor – all distances in group g are in $\Theta(2^{m \cdot g})$. We define a mapping π_q between distance classes that maps all classes of a group to the median distance class in this group (the value of π_q for the first and last distance group is rounded off appropriately; we omit a precise definition for the sake of clarity).

$$\pi_q(\ell) := i + m \cdot g_q(\ell) - \lfloor \frac{m}{2} \rfloor$$

Observe that this partitioning satisfies the key property that for each distance class i, there exists a q such that $\pi_q(i) = i$, this means there is a q such that the distance classes close to i (in the range $i - m/2, \ldots, i + m/2$) are mapped to i.

Based on this mapping we define a set S_t^q for each choice of $t \in T$ and $q \in Q$ by $S_t^q = \{x \in V : \mathrm{color}_{\pi_q(\ell_t(x))}(x) = 0\}$, where $\mathrm{color}_i(x)$ denotes the color of the cluster that contains x in partitioning P_i. Note that all nodes whose t-radii fall into the same distance group (w.r.t. parameter q) derive their color (and hence whether they belong to S_t^q) from the same partitioning.

Based on the sets S_t^q we define an embedding $\varphi_{t,q} : V \longrightarrow \mathbb{R}$ for each coordinate (t, q) — $\varphi_{t,q}(x) = d(x, S_t^q)$. The embedding $\varphi : V \longrightarrow \mathbb{R}^{|T||Q|}$ is defined by

$$\varphi(x) := \otimes_{t,q}\, \varphi_{t,q}(x). \tag{12.1}$$

In the next section, we analyse the distortion of the map φ.

12.2 The Analysis

Since each coordinate of φ maps point x to its distance from some subset of points, it follows that each coordinate of this embedding is contracting. Therefore, we have for all $x, y \in V$

$$\begin{aligned} ||\varphi(x) - \varphi(y)||_2 &\leq \sqrt{|T| \cdot |Q| \cdot d(x,y)^2} \\ &\leq O(\sqrt{\log n}) \cdot d(x,y) \ . \end{aligned}$$

Now, we show that for a pair x, y that is $\delta(x,y)$-separated in the partitioning corresponding to its distance scale $\lfloor \log(d(x,y)) \rfloor$, with a constant probability, we get

$$||\varphi(x) - \varphi(y)||_2 \geq \Omega(\delta(x,y) \cdot d(x,y)) \cdot \sqrt{V(x,y)} \ . \tag{12.2}$$

This gives Lemma 12.2 since $\delta(x,y) > \alpha(x,y)$ with constant probability.

Fix a pair (x, y) that is $\delta(x,y)$-separated in the partitioning for distance scale $\lfloor \log(d(x,y)) \rfloor$. Without loss of generality assume that the maximum in the definition of $V(x,y)$ is attained by the first term, i.e. $\frac{|B(x,2d(x,y))|}{|B(x,d(x,y)/8)|} \geq \frac{|B(y,2d(x,y))|}{|B(y,d(x,y)/8)|}$. We show that for any t with $|B(x, d(x,y)/8)| \leq 2^t \leq |B(x, 2d(x,y))|$, there is a $q \in Q$ such that the coordinate (t, q) gives a large contribution, i.e., $|\varphi_{t,q}(x) - \varphi_{t,q}(y)| \geq \Omega(\delta(x,y) \cdot d(x,y))$. Equation 12.2 then follows.

We fix an integer t with $\log(|B(x, d(x,y)/8)|) \leq t \leq \log(|B(x, 2d(x,y))|)$, and we use $i = \lfloor \log d(x,y) \rfloor$ to denote the distance class of $d(x,y)$. Clearly, the distance class $\ell_t(x)$ of the

t-radius of x is in $\{i-4, \ldots, i+2\}$, because $d(x,y)/8 \leq \mathrm{r}_t(x) \leq 2d(x,y)$. The following claim gives a similar bound on the t-radius for nodes that are close to x. This is the key property: it says that nodes close to x use a distance class that is not too far away from the distance class used by x. Because of the partitioning into distance-groups we can then find a coordinate (i.e., a value of q) for which all the nodes around x derive their color from the partitioning corresponding to $\ell_t(x)$.

Claim 12.4 *Let $z \in B(x, \frac{1}{16}d(x,y))$. Then $\ell_t(z) \in \{i-5, i+3\}$.*

Proof. For the t-radius $\mathrm{r}_t(z)$ around z we have $\mathrm{r}_t(x) - d(x,y)/16 \leq \mathrm{r}_t(z) \leq \mathrm{r}_t(x) + d(x,y)/16$. Since $d(x,y)/8 \leq \mathrm{r}_t(x) \leq 2d(x,y)$ we get $\frac{1}{16}d(x,y) \leq \mathrm{r}_t(z) \leq \frac{33}{16}d(x,y)$, which yields the claim. ■

In the following we choose m (the number of distances classes within a group) as 10, and q such that $\pi_q(i) = i$, i.e., i is the median of its distance group. Then the above claim ensures that for all nodes $z \in B(x, \frac{1}{16}d(x,y))$, the distance class $\ell_t(z)$ is in the same distance group as i. Furthermore, these nodes choose their color (that decides whether they belong to S_t^q) according to the partitioning for distance scale i. Recall that x is $\delta(x,y)$-separated in this partitioning. Therefore, we can make the following claim.

Claim 12.5 *If x does not belong to the set S_t^q, then,*

$$d(x, S_t^q) \geq \min\{\tfrac{1}{16}, \delta(x,y)\}\, d(x,y) \geq \frac{1}{16}\,\delta(x,y)\, d(x,y)$$

Now, we consider the following events concerning the distances of x and y from S_t^q, respectively.

- $X_{\text{far}} = \{d(x, S_t^q) > \frac{1}{16}\delta(x,y)d(x,y)\}$
- $Y_0 \;\; = \{d(y, S_t^q) = 0\}$, i.e., $y \in S_t^q$

These events only depend on the random colorings chosen for the partitionings in different distance classes. First we claim that the event X_{far} is independent of Y_0. To see this, note that X_{far} only depends on colors chosen for nodes in $B(x, \frac{1}{16}\delta(x,y)d(x,y))$. Our choice of q ensures that these colors are derived from the partitioning for distance class i, and Claim 12.4 implies that all nodes in $B(x, \frac{1}{16}\delta(x,y)d(x,y))$ get the color assigned to the cluster $C_i(x)$.

The event Y_0, however, depends on the color chosen for y. This color is either derived from a partitioning for a distance class different from i (in this case independence is immediate), or it is equal to the color assigned to the cluster $C_i(y)$. In the latter case the independence follows, since x and y lie in different clusters in this partitioning as they are separated by it.

If $Y_0 \cap X_{\text{far}}$ happens, then the dimension (t,q) gives a contribution of $\Omega(\delta(x,y)d(x,y))$. This happens with probability 1/4 as x needs to get color "0" and y needs to get color "1". This completes the proof of Lemma 12.2.

References

[CGR05] Shuchi Chawla, Anupam Gupta, and Harald Räcke. An improved approximation to sparsest cut. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 102–111, 2005.

[CKR01] Joseph Cheriyan, Howard Karloff, and Yuval Rabani. Approximating directed multicuts. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 320–328, 2001.

[FRT03] Jittat Fakcharoenphol, Satish B. Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th ACM Symposium on Theory of Computing (STOC)*, pages 448–455, 2003.

[KLMN04] Robert Krauthgamer, James Lee, Manor Mendel, and Assaf Naor. Measured descent: A new embedding method for finite metrics. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 434–443, 2004.