CS138 Set #5 Question 1

Chris Beck

May 29, 2009

1

First, a lemma. I claim that, given a fixed $\hat{u} \in S^{n-1}$, and a uniformly randomly distributed $\hat{v}$ on S^{n-1},

$$\Pr_{\hat{v} \in S^{n-1}}[|\hat{u} \cdot \hat{v}| \leq c] = O(c) \ .$$

There are many ways to prove this, but here is one.

Pf: Let R denote the region defined by

$$\{\hat{v} \in S^{n-1} : |\hat{u} \cdot \hat{v}| \leq c\} \ ,$$

and let μ denote the area measure, then we must show $\frac{\mu(R)}{\mu(S^{n-1})} \leq O(c)$. Since the denominator is constant, this is equivalent to $\mu(R) = O(c)$.

Consider a cylinder circumscribing S^{n-1} with axis $\hat{u}$. Consider the projection from the cylinder onto the sphere, which moves points on the cylinder toward the axis until they intersect the sphere. This smooth map only contracts in area, this is well known but we may prove it as follows:

Pf: Our map sends (θ, z) to $(\theta, \cos^{-1}(z))$, both coordinate sets are orthogonal, the $d\theta$ direction contracts by $\sqrt{1-z^2}$, while $\frac{d\cos\phi}{d\phi} = \sin\phi$, which implies $\frac{d\phi}{dz} = \frac{1}{\sin\cos^{-1} z} = \frac{1}{\sqrt{1-z^2}}$, thus our $d\theta dz$ area element maps to a $d\theta d\phi$ infinitesimal rectangle on the sphere, which it is well known has an area element $\sin\phi d\phi d\theta$. Thus the map always contracts area, since $\sin\phi \leq 1$.

So we see in particular

$$\mu(R) \leq \mu(\{\hat{v} \in \text{cylinder} : |\hat{u} \cdot \hat{v}| \leq c\}) = \mu(S^{n-2}) \cdot 2c = O(c)$$

since a cross section of the cylinder perpindicular to $\hat{u}$ is just a copy of S^{n-2}. Thus the lemma follows.

Let σ denote our $O(\log n)$-embedding of the graph G with metric into l_2 space.

Now, we consider an indicator random variable $I_{u,v}$ for the event that u, v project to distance of ≤ 1 on the line l generated by unit vector $\hat{l}$, i.e. $|(\sigma(u) \cdot \hat{l})\hat{l} - (\sigma(v) \cdot \hat{l})\hat{l}| \leq 1$. Note that this is the same as, (since $|\hat{l}| = 1$ and dot product is linear),

$$|(\sigma(u) - \sigma(v)) \cdot \hat{l}| \leq 1 \ ,$$

$$\left| \frac{\sigma(u) - \sigma(v)}{||\sigma(u) - \sigma(v)||_2} \cdot \hat{l} \right| \leq \frac{1}{||\sigma(u) - \sigma(v)||_2} \ .$$

By distortion assumption, $\frac{d_{u,v}}{\log n} \leq ||\sigma(u) - \sigma(v)||_2$, so this event implies the event

$$\left| \frac{\sigma(u) - \sigma(v)}{||\sigma(u) - \sigma(v)||_2} \cdot \hat{l} \right| \leq \frac{\log n}{d_{u,v}} \ .$$

The probability of this new event thus bounds the probability of the event we care about, and so by the lemma applied to new event we obtain the bound, $\Pr[I_{u,v}] = O(\frac{\log n}{d_{u,v}})$. We conclude by applying linearity of expectation to these indicator random variables that

$$E[\ \#\text{ points in distance 1 on } l\ ] = E\left[\sum_{u,v} I_{u,v}\right] = O\left(\log n \sum_{u,v} \frac{1}{d_{u,v}}\right)$$

as desired.

2

The hint follows for these reasons. Since for any $u \sim v$, we have $\text{spread}(u,v) \leq OPT$, we see that if $u_0, u_1, \ldots, u_k$ is a path in G, by triangle inequality we have

$$\text{spread}(u_0, u_k) \leq \text{spread}(u_0, u_1) + \text{spread}(u_1, u_2) + \ldots + \text{spread}(u_{k-1}, u_k) \leq k \cdot OPT\ .$$

Then, since the vertices are given distinct integral values, the number of distinct points with spread $k \cdot OPT$ from a fixed u_0 is at most $2k \cdot OPT$.

Thus we see that for any $v \in V$, $|\{u \in V : d_{u,v} \leq k\}| \leq 2k \cdot OPT$, and so by union bound (and dividing by two because of double counting)

$$n_k = |\{(u,v) \in V \times V : d_{u,v} \leq k\}| \leq n \cdot k \cdot OPT\ .$$

Now, $|d^{-1}(k)| = n_k - n_{k-1}$ by definition. We may now express in terms of n_k,

$$\sum_{u,v} \frac{1}{d_{u,v}} = \sum_k \frac{n_k - n_{k-1}}{k} = \sum_k n_k \frac{1}{k(k+1)}\ ,$$

and applying the bound to each n_k, this is bounded as

$$\sum_{k=1}^{n} \frac{n \cdot OPT}{k+1} \leq n \cdot OPT \cdot O(\log n)$$

since the harmonic series is bounded as $O(\log n)$. Thus the claim follows.

Together with the result from 1, this implies our expected number of pairs within distance 1 on l is $O(n \log^2 n \cdot OPT)$.

(3)

First, it should be clear that the hint implies the result, since this bound implies a lower bound of s^2 on the expectation we computed above;

$$s^2 = O(n \log^2 n \cdot OPT)$$
$$s = O(\sqrt{n \cdot OPT} \log n)$$

To see the hint, we use the other fact about our l_2 embedding, which is that it strictly contracts distances. Projecting orthogonally onto line l only reduces the l_2 distances, thus for any u, v, we see that

$$|(\sigma(u) - \sigma(v)) \cdot \hat{l}| \leq d_{u,v}\ .$$

Meanwhile the spread of u, v is simply the number of vertices on the interval between the two points $\sigma(u) \cdot \hat{l}, \sigma(v) \cdot \hat{l}$, minus one.

Thus for $u \sim v$ a pair with spread $s = OPT$, we have an interval of l of size at most $d_{u,v} = 1$, with $s - 1$ other points on this interval. Looking at these $s + 1$ points alone, we have $(s + 1 \cdot s) - s$ pairs of distinct points, i.e. s^2 pairs of distinct points with a distance of within 1 of eachother on l.