## Stat260/CS294: Spectral Graph Methods

Lecture 9 - 02/19/2015

Lecture: Expanders, in theory and in practice (2 of 2)

Lecturer: Michael Mahoney

Scribe: Michael Mahoney

Warning: these notes are still very rough. They provide more details on what we discussed in class, but there may still be some errors, incomplete/imprecise statements, etc. in them.

# 9 Expanders, continued

Here, we will describe how expanders are the metric spaces that are least like low-dimensional Euclidean spaces (or, for that matter, any-dimensional Euclidean spaces). Someone asked at the end of the previous class about what would an expander "look like" if we were to draw it. The point of these characterizations of expanders—that they don't have good partitions, that they embed poorly in low dimensional spaces, etc.—is that you can't draw them to see what they look like, or at least you can't draw them in any particularly meaningful way. The reason is that if you could draw them on the board or a two-dimensional piece of paper, then you would have an embedding into two dimensions. Relatedly, you would have partitioned the expander into two parts, i.e., those nodes on the left half of the page, and those nodes on the right half of the page. Any such picture would have roughly as many edges crossing between the two halves as it had on either half, meaning that it would be a non-interpretable mess. This is the reason that we are going through this seemingly-circuitous characterizations of the properties of expanders—they are important, but since they can't be visualized, we can only characterize their properties and gain intuition about their behavior via these indirect means.

## 9.1 Introduction to Metric Space Perspective on Expanders

To understand expanders from a metric space perspective, and in particular to understand how they are the metric spaces that are least like low-dimensional Euclidean spaces, let's back up a bit to the seemingly-exotic subject of *metric spaces* (although in retrospect it will not seem so exotic or be so surprising that it is relevant).

- Finite-dimensional Euclidean space, i.e.,  $\mathbb{R}^n$ , with  $n < \infty$ , is an example of a metric space that is very nice but that is also quite nice/structured or limited.
- When you go to infinite-dimensional Hilbert spaces, things get much more complex; but ∞dimensional RKHS, as used in ML, are ∞-dimensional Hilbert spaces that are sufficiently regularized that they inherit most of the nice properties of ℝ<sup>n</sup>.
- If we measure distances in  $\mathbb{R}^n$  w.r.t. other norms, *e.g.*,  $\ell_1$  or  $\ell_{\infty}$ , then we step outside the domain of Hilbert spaces to the domain of Banach spaces or normed vector spaces.
- A graph G = (V, E) is completely characterized by its shortest path or geodesic metric; so the metric space is the nodes, with the distance being the geodesic distance between the nodes.

Of course, you can modify this metric by adding nonnegative weights to edges like with some nonlinear dimensionality reduction methods. Also, you can assign a vector to vertices and thus view a graph geometrically. (We will get back to the question of whether there are other distances that one can associate with a graph, e.g., resistance of diffusion based distances; and we will ask what is the relationship between this and geodesic distance.)

• The data may not be obviously a matrix or a graph. Maybe you just have similarity/dissimilarity information, *e.g.*, between DNA sequences, protein sequences, or microarray expression levels. Of course, you might want to relate these things to matrices or graphs in some way, as with RKHS, but let's deal with metrics first.

So, let's talk about metric spaces more generally. The goal will be to understand how good/bad things can be when we consider metric information about the data.

So, we start with a definition:

**Definition 1.** (X, d) is a metric space if

- $d: X \times X \to \mathbb{R}^+$  (nonnegativity)
- d(x,y) = 0 iff x = y
- d(x,y) = d(y,x) (symmetric)
- $d(x,y) \le d(x,z) + d(z,y)$  (triangle inequality)

The idea is that there is a function over the set X that takes as input pairs of variables that satisfies a generalization of what our intuition from Euclidean distances is: namely, nonnegativity, the second condition above, symmetry, and the triangle inequality. Importantly, this metric does not need to come from a dot product, and so although the intuition about distances from Euclidean spaces is the motivation, it is significantly different and more general. Also, we should note that if various conditions are satisfied, then various metric-like things are obtained:

- If the second condition above is relaxed, but the other conditions are satisfied, then we have a *psuedometric*.
- If symmetry is relaxed, but the other conditions are satisfied, then we have a *quasimetric*.
- If the triangle inequality is relaxed , but the other conditions are satisfied, then we have a *semimetric*.

We should note that those names are not completely standard, and to confuse matters further sometimes the relaxed quantities are called metrics—for example, we will encounter the so-called *cut metric* describing distances with respect to cuts in a graph, which is not really a metric since the second condition above is not satisfied.

More generally, the distances can be from a Gram matrix, a kernel, or even allowing algorithms in an infinite-dimensional space.

Some of these metrics can be a little counterintuitive, and so for a range of reasons it is useful to ask how similar or different two metrics are, *e.g.*, can we think of a metric as a tweak of a low-dimensional space, in which case we might hope that some of our previous machinery might apply. So, we have the following question:

**Question 1.** How well can a given metric space (X, d) be approximated by  $\ell_2$ , where  $\ell_2$  is the metric space  $(\mathbb{R}^n, || \cdot ||)$ , where  $\forall x, y \in \mathbb{R}^n$ , we have  $||x - y||^2 = \sum_{i=1}^n (x_i - y_i)^2$ .

The idea here is that we want to replace the metric d with something d' that is "nicer," while still preserving distances—in that case, since a lot of algorithms use only distances, we can work with d' in the nicer place, and get results that are algorithmically and/or statistically better without introducing too much error. That is, maybe it's faster without too much loss, as we formulated it before; or maybe it is better, in that the nicer place introduced some sort of smoothing. Of course, we could ask this about metrics other than  $\ell_2$ ; we just start with that since we have been talking about it.

There are a number of ways to compare metric spaces. Here we will start by defining a measure of distortion between two metrics.

**Definition 2.** Given a metric space (X, d) and our old friend the metric space  $(\mathbb{R}^n, \ell_2)$ , and a mapping f:  $X \to \mathbb{R}^n$ :

- $expansion(f) = max_{x_1, x_2 \in X} \frac{||f(x_1) f(x_2)||_2}{d(x_1, x_2)}$
- contraction(f) =  $max \frac{d(x_1, x_2)}{||f(x_1) f(x_2)||}$
- $distortion(f) = expansion(f) \cdot contraction(f)$

As usual, there are several things we can note:

- An embedding with distortion 1 is an *isometry*. This is very limiting for most applications of interset, which is OK since it is also unnecessarily strong notion of similarity for most applications of interest, so we will instead look for low-distortion embeddings.
- There is also interest in embedding into  $\ell_1$ , which we will return to below when talking about graph partitioning.
- There is also interest in embedding in other "nice" places, like trees, but we will not be talking about that in this class.
- As a side comment, a Theorem of Dvoretzky says that any embedding into normed spaces,  $\ell_2$  is the hardest. So, aside from being something we have already seen, this partially justifies the use of  $\ell_2$  and the central role of  $\ell_2$  in embedding theory more generally.

Here, we should note that we have already seen one example (actually, several related examples) of a low-distortion embedding. Here we will phrase the JL lemma that we saw before in our new nomenclature.

**Theorem 1** (JL Lemma). Let X be an n-point set in Euclidean space, i.e.,  $X \subset \ell_2^n$ , and fix  $\epsilon \in (0,1]$ . Then  $\exists a (1+\epsilon)$ -embedding of X into  $\ell_2^k$ , where  $k = O\left(\frac{\log n}{\epsilon^2}\right)$ .

That is, Johnson-Lindenstrauss says that we can map  $x_i \to f(x)$  such that distance is within  $1 \pm \epsilon$  of the original.

A word of notation and some technical comments: For  $x \in \mathbb{R}^d$  and  $p \in [1, \infty)$ , the  $\ell_p$  norm of x is defined as  $||x||_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$ . Let  $\ell_p^d$  denote the space  $\mathbb{R}^d$  equipped with the  $\ell_p$  norm. Sometimes we are interested in embeddings into some space  $\ell_p^d$ , with p given but the dimension d unrestricted, *e.g.*, in *some* Euclidean space s.t. X embeds well. Talk about:  $\ell_p$  = the space of all sequences  $(x_1, x_2, \ldots)$ , with  $||x||_p < \infty$ , with  $||x||_p$  defined as  $||x||_p = (\sum_{i=1}^\infty |x_i|^p)^{1/p}$ . In this case, embedding into  $\ell_p$  is shorthand for embedding into  $\ell_p^d$  for some d.

Here is an important theorem related to this and that we will return to later.

**Theorem 2** (Bourgain). Every n-point metric space (X, d) can be embedded into Euclidean space  $\ell_2$  with distortion  $\leq O(\log n)$ .

Proof Idea. (The proof idea is nifty and used in other contexts, but we won't use it much later, except to point out how flow-based methods do something similar.) The basic idea is given (X, d), map each point  $x \to \phi(x)$  in  $O(\log^2 n)$ -dimensional space with coordinates equal to the distance to  $S \subseteq X$  where S is chosen randomly. That is, given (X, d), map every point  $x \in X$  to  $\phi(x)$ , an  $O(\log^2 n)$ -dimensional vector, where coordinates in  $\phi(\cdot)$  correspond to subsets  $S \subseteq X$ , and s.t. the s-th in  $\phi(x)$  is  $d(x, S) = \min_{s \in S} d(x, s)$ . Then, to define the map, specify a collection of subsets we use selected carefully but randomly—select  $O(\log n)$  subsets of size 1,  $O(\log n)$  subsets of size 2, of size 4, 8, ...,  $\frac{n}{2}$ . Using that, it works, *i.e.*, that is the embedding.

Note that the dimension of the Euclidean space was originally  $O(\log^2 n)$ , but it has been improved to  $O(\log n)$ , which I think is tight. Note also that the proof is algorithmic in that it gives an efficient randomized algorithm.

Several questions arise:

- Q: Is this bound tight? A: YES, on expanders.
- Q: Let  $c_2(X, d)$  be the distortion of the embedding of X into  $\ell_2$ ; can we compute  $c_2(X, d)$  for a given metric? A: YES, with an SDP.
- Q: Are there metrics such that  $c_2(X, d) \ll \log n$ ? A: YES, we saw it with JL, *i.e.*, high-dimensional Euclidean spaces, which might be trivial since we allow the dimension to float in the embedding, but there are others we won't get to.

### 9.1.1 Primal

The problem of whether a given metric space is  $\gamma$ -embeddable into  $\ell_2$  is polynomial time solvable. Note: this does not specify the dimension, just whether there is some dimension; asking the same question with dimension constraints or a fixed dimension is in general much harder. Here, the condition that the distortion  $\leq \gamma$  can be expressed as a system of linear inequalities in Gram matrix correspond to vectors  $\phi(x)$ . So, the computation of  $c_2(x)$  is an SDP—which is easy or hard, depending on how you view SDPs—actually, given an input metric space (X, d) and an  $\epsilon > 0$ , we can determine  $c_2(X, d)$  to relative error  $\leq \epsilon$  in  $poly(n, 1/\epsilon)$  time.

Here is a basic theorem in the area:

**Theorem 3** (LLR).  $\exists$  a poly-time algorithm that, given as input a metric space (X, d), computes  $c_2(X, d)$ , where  $c_2(X, d)$  is the least possible distortion of any embedding of (X, d) into  $(\mathbb{R}^n, \ell_2)$ .

*Proof.* The proof is from HLW, and it is based on semidefinite programming. Let (X, d) be the metric space, let |X| = n, and let  $f : X \to \ell_2$ . WLOG, scale f s.t. contraction(f) = 1. Then,  $distortion(f) \leq \gamma$  iff

$$d(x_i, x_j)^2 \le ||f(x_i) - f(x_j)||^2 \le \gamma^2 d(x_i, x_j)^2.$$
(1)

Then, let  $u_i = f(x_i)$  be the *i*-th row of the embedding matrix U, and let  $Z = UU^T$ . Note that  $Z \in PSD$ , and conversely, if  $Z \in PSD$ , then  $Z = UU^T$ , for some matrix U. Note also:

$$||f(x_i) - f(x_j)||^2 = ||u_i - u_j||^2$$
  
=  $(u_i - u_j)^T (u_i - u_j)$   
=  $u_i^T u_i + u_j^T u_j - 2u_i^T u_j$   
=  $Z_{ii} + Z_{jj} - 2Z_{ij}.$ 

So, instead of finding a  $u_i = f(x_i)$  s.t. (1) holds, we can find a  $Z \in PSD$  s.t.

$$d(x_i, x_j)^2 \le Z_{ii} + Z_{jj} - 2Z_{ij} \le \gamma^2 d(x_i, x_j)^2.$$
(2)

Thus,  $c_2 \leq \gamma$  iff  $\exists Z \in SPSD$  s.t. (2) holds  $\forall ij$ . So, this is an optimization problem, and we can solve this with simplex, interior point, ellipsoid, or whatever; and all the usual issues apply.  $\Box$ 

#### 9.1.2 Dual

The above is a Primal version of the optimization problem. If we look at the corresponding Dual problem, then this gives a characterization of  $c_2(X, d)$  that is useful in proving lower bounds. (This idea will also come up later in graph partitioning, and elsewhere.) To go from the Primal to the Dual, we must take a nonnegative linear combination of constraints. So we must write  $Z \in PSD$  in such a way, since that is the constraint causing a problem; the following lemma will do that.

**Lemma 1.**  $Z \in PSD$  iff  $\sum_{ij} q_{ij} z_{ij} \ge 0, \forall Q \in PSD$ .

*Proof.* First, we will consider rank 1 matrices; the general result will follow since general PSD matrices are a linear combination of rank-1 PSD matrices of the form  $qq^T$ , *i.e.*,  $Q = qq^T$ .

First, start with the  $\Leftarrow$  direction: for  $q \in \mathbb{R}^n$ , let Q be PSD matrix s.t.  $Q_{ij} = q_i q_j$ ; then

$$q^t Z q = \sum_{ij} q_i Z_{ij} q_j = \sum_{ij} Q_{ij} z_{ij} \ge 0,$$

where the inequality follows since Q is PSD. Thus,  $Z \in PSD$ .

For the  $\Rightarrow$  direction: let Q be rank-1 PSD matrix; thus, it has the form  $Q = qq^T$  or  $Q_{ij} = q_iq_j$ , for  $q \in \mathbb{R}^n$ . Then,

$$\sum_{ij} Q_{ij} z_{ij} = \sum_{ij} q_i Z_{ij} q_j \ge 0,$$

where the inequality follows since A is PSD.

Thus, since  $Q \in PSD$  implies that  $Q = \sum_{i} q_i q_i^T = \sum_{i} \Omega_i$ , with  $\Omega_i$  being a rank-*i* PSD matrix, the lemma follows by working through things.

Now that we have this characterization of  $Z \in PSD$  in terms of a set of (nonnegative) linear combination of constraints, we are ready to get out Dual problem which will give us the nice characterization of  $c_2(X, d)$ .

Recall finding an embedding  $f(x_i) = u_i$  iff finding a matrix Z iff  $\sum_{ij} q_{ij} z_{ij} \ge 0, \forall Q \in SPSD$ . So, the Primal constraints are:

- I.  $\sum q_{ij} z_{ij} \ge 0$  for all  $Q \in PSD$
- II.  $z_{ii} + z_{jj} 2z_{ij} \ge d(x_i, x_j)^2$
- III.  $\gamma^2 d(x_i, x_j)^2 \ge z_{ii} + z_{jj} 2z_{ij}$ ,

which hold  $\forall ij$ . Thus, we can get the following theorem.

#### Theorem 4 (LLR).

$$C_2(X,d) = \max_{(P \in PSD, P.1=0)} \sqrt{\frac{\sum_{P_{ij} > 0} P_{ij} d(x_i, x_j)^2}{-\sum_{(P_{ij} < 0)} P_{ij} d(x_i, x_j)^2}}$$

*Proof.* The dual program is the statement that for  $\gamma < C_2(X, d)$ , thre must exist a non-negative combination of the constraints of the primal problem that yields a contradiction.

So, we will assume  $\gamma < C_2(x, d)$  and look for a contradiction, *i.e.*, look for a linear combination of constraints such that the primal gives a contradiction. Thus, the goal is to construct a nonnegative linear combination of primal constraints to give a contradiction s.t.  $Q.Z = \sum q_{ij} z_{ij} \ge 0$ .

Recall that the cone of PSD matrices is convex.

The goal is to zero out the  $z_{ij}$ .

(I) above says that  $Q \cdot Z = \sum_{ij} q_{ij} z_{ij} \ge 0$ . Note that since PSD cone is convex, a nonnegative linear combination of the form  $\sum_k \alpha_k \langle Q, Z \rangle = P \cdot z$ , for some  $P \in PSD$ . So, modifying first constraint from the primal, you get

- I'.  $\sum_{ij} P_{ij} Z_{ij} = P \cdot Z \ge 0$ , for some  $P \epsilon PSD$ To construct P, choose the elements such that you zero out  $z_{ij}$  in the following manner.
  - If  $P_{ij} > 0$ , multiply second constraint from primal by  $P_{ij}/2$ , (*i.e.*, the constraint  $d(x_i, x_j)^2 \le z_{ii} + z_{jj} 2z_{ij}$ )
  - If  $P_{ij} < 0$ , multiply third constraint from primal by  $-P_{ij}/2$ , (*i.e.*, the constraint  $z_{ii} + z_{jj} 2z_{ij} \le \gamma^2 d(x_i, x_j)^2$ )
  - If  $P_{ij} = 0$ , multiply by 0 constraints involving  $z_{ij}$ .

This gives

$$\frac{P_{ij}}{2} (z_{ii} + z_{jj} - 2z_{ij}) \geq \frac{P_{ij}}{2} d(x_i, x_j)^2 - \frac{P_{ij}}{2} \gamma^2 d(x_i, x_j)^2 \geq -\frac{P_{ij}}{2} (z_{ii} + z_{jj} - 2z_{ij})^2$$

)

from which it follows that you can modify the other constraints from primal to be:

II'. 
$$\sum_{ij,P_{ij}>0} \frac{P_{ij}}{2} (z_{ii} + z_{jj} - 2z_{ij}) \ge \sum_{ij,P_{ij}>0} \frac{P_{ij}}{2} d(x_i, x_j)^2$$
  
III'. 
$$\sum_{ij,P_{ij}<0} \frac{P_{ij}}{2} (z_{ii} + z_{jj} - 2z_{ij}) \ge \sum_{ij,P_{ij}<0} \frac{P_{ij}}{2} \gamma^2 (d(x_i, x_j)^2)$$

If we add those two things, then we get,

$$\sum_{i} P_{ii} z_{ii} + \sum_{ij: P_{ij} > 0} \frac{P_{ij}}{2} (z_{ii} + z_{jj}) + \sum_{ij} \frac{P_{ij}}{2} (z_{ii} + z_{jj}) \ge \text{RHS Sum},$$

and so

$$\sum_{i} P_{ii} z_{ii} + \sum_{ij: P_{ij} \neq 0} \frac{P_{ij}}{2} (z_{ii} + z_{jj}) \ge \text{RHS Sum},$$

and so, since we choose P s.t.  $P \cdot \vec{1} = \vec{0}$ , (i.e.  $\sum_{j} P_{ij} = 0$  for all i, and  $\sum_{i} P_{ij} = 0$  for all j by symmetry) we have that

$$0 = \sum_{i} \left( P_{ii} + \sum_{j:P_{ij} \neq 0} P_{ij} \right) z_{ij} \ge RHS = \sum_{ij:P_{ij} > 0} \frac{P_{ij}}{2} d(x_i, x_j)^2 + \sum_{ij:P_{ij} < 0} \frac{P_{ij}}{2} \gamma^2 d(x_i, x_j)^2$$

So, it follows that

$$0 \ge \sum_{ij:P_{ij}>0} P_{ij}d(x_i, x_j)^2 + \sum_{ij:P_{ij}<0} \gamma^2 d(x_i, x_j)^2.$$

This last observation is FALSE if

$$\gamma^2 < \frac{\sum_{ij:P_{ij}>0} P_{ij} d(x_i, x_j)^2}{\sum_{ij:P_{ij}<0} (-P_{ij}) d(x_i, x_j)^2}$$

and so the theorem follows.

(In brief, adding the second and third constraints above,

$$0 \ge \sum_{ij, P_{ij} > 0} P_{ij} d(x_i, x_j)^2 + \sum_{ij, P_{ij} < 0} P_{ij} \gamma^2 (d(x_i, x_j)^2)$$

This will be false if you choose  $\gamma$  to be small—in particular, it will be false if  $\gamma^2 \leq top/bottom$ , from which the theorem will follow.

#### 9.2 Metric Embedding into $\ell_2$

We will show that expanders embed poorly in  $\ell_2$ —this is the basis for the claim that they are the metric spaces that are least like low-dimensional spaces in a very strong sense.

It is easy to see that an expander can be embedded into  $\ell_2$  with distortion  $O(\log n)$  (just note that any graph can be embedded with distortion equal to its diameter)—in fact, any metric space can be embedded with that distortion. We will show that this result is tight, and thus that expanders are the worst.

The basic idea for showing that expanders embed poorly in  $\ell_2$  is: If G is a d-regular,  $\epsilon$ -expander, then  $\lambda_2$  of  $A_G$  is  $\langle d - \delta$  for  $\delta = \delta(d, \epsilon) \neq \delta(n)$ . The vertices of a bounded degree graph can be paired up s.t. every pair of vertices are a distance  $\Omega(\log n)$ . We can then let B be a permutation matrix for the pairing, and use the matrix  $P = dI - A + \frac{\delta}{2}(B - I)$ .

Note: We can have a simpler proof, using the theorem of Bourgain that expanders don't embed well in  $\ell_2$ , since we can embed in  $\mathbb{R}^{diameter}$ , and  $diameter(expander) = \log n$ . But we go through this here to avoid (too much) magic.

Start with the following definitions:

**Definition 3.** A Hamilton cycle in a graph G = (V, E) is a cycle that visits every vertex exactly once (except for the start and end vertices).

**Definition 4.** A matching is a set of pairwise non-adjacent edges, i.e., no two edges share a common vertex. A vertex is matched if it is incident to an edge. A perfect matching is a matching that matched all the vertices of a graph.

The following theorem is the only piece of magic we will use here:

**Theorem 5.** A simple graph with  $n \ge 3$  edges is Hamiltonian if every vertex has degree  $\ge \frac{n}{2}$ .

Note that if every vertex has degree  $\geq \frac{n}{2}$ , then the graph is actually quite dense, and so from Szemeredi-type results relating dense graphs to random graphs it might not be so surprising that there is a lot of wiggle room.

Note: A cycle immediately gives a matching.

Thus, we have the following lemma:

**Lemma 2.** Let G = (V, E) be a d-regular graph, with |V| = n If H = (V, E') is a graph with the same vertex set as G, in which two vertices u and v are adjacent iff  $d_G(u, v) \ge \lfloor \log_k n \rfloor$ . Then, H has a matching with n/2 edges.

Proof. Since G is a d-regular graph, hence for any vertex  $x \in V$  and any value r, it has  $\leq k^r$  vertices  $y \in V$  can have  $d_G(x, y) \leq r$ , *i.e.*, only that many vertices are within a distance r. If  $r = \lfloor \log_k n \rfloor - 1$ , then  $\exists \leq \frac{n}{2}$  vertices within distance r; that is, at least half of the nodes of G are further than  $\log_k n - 2$  from x; this means every node in H has at least degree n/2. So H has a Hamilton cycle and thus a perfect matching, and by the above theorem the lemma follows.  $\Box$ 

Finally we get to the main theorem that says that expanders embed poorly in  $\ell_2$ —note that this is a particularly strong statement or notion of nonembedding, as by Bourgain we know any graph (with the graph distance metric) can be embedded into  $\ell_2$  with distortion  $O(\log n)$ , so expander is the worst case in this sense.

**Theorem 6.** Let  $d \ge 3$ , and let  $\epsilon > 0$ . If G = (V, E) is a (n, d)-regular graph with  $\lambda_2(A_G) \le d - \epsilon$ and |V| = n, then

 $C_2(G) = \Omega(\log n)$ 

where the constant inside the  $\Omega$  depends on  $d, \epsilon$ .

*Proof.* To prove the lower bound, we use the characterization from the last section that for the minimum distortion in embedding a metric space (X, d) into  $l_2$ , denoted by  $C_2(X, d)$ , is:

$$C_2(X,d) = \max_{P \in PSD, P. \vec{1} = \vec{0}} \sqrt{\frac{\sum_{p_{ij} > 0} p_{ij} d(x_i, x_j)^2}{-\sum_{p_{ij} < 0} p_{ij} d(x_i, x_j)^2}}$$
(3)

and we will find some P that is feasible that gives the LB.

Assume B is the adjacency matrix of the matching in H, whose existence is proved in the previous lemma. Then, define

$$P = (dI - A_G) + \frac{\epsilon}{2}(B - I).$$

Then, we claim that  $P\overrightarrow{1} = \overrightarrow{0}$ . To see this, notice both  $(dI - A_G)$  and I - B are Laplacians (not normalized), as B is the adjacency matrix of a perfect matching (i.e. 1-regular graph). Next, we claim that  $P \in PSD$ . This proof of this second claim is because, for any  $x \perp \overrightarrow{1}$ , we have

$$x^{T}(dI - A_{G})x \ge dx^{T}x - x^{T}Ax \ge (d - \lambda_{2})||x||^{2} \ge \epsilon ||x||^{2}$$

(by the assumption on  $\lambda_2$ ); and

$$\begin{aligned} x^{T}(B-I)x &= \sum_{(i,j)\in B} 2x_{i}x_{j} - \sum_{i} x_{i}^{2} \\ &= \sum_{(i,j)\in B} (2x_{i}x_{j} - x_{i}^{2} - x_{j}^{2}) \quad \text{since } ||x||^{2} = \sum_{(i,j)\in B} x_{i}^{2} + x_{j}^{2} \\ &\geq -2\sum_{(i,j)\in B} x_{i}^{2} + x_{j}^{2} \\ &= -2||x||^{2} \end{aligned}$$

The last line is since  $||x||^2 = \sum_{(i,j)\in B} x_i^2 + x_j^2$  and since B is a matching so each i shows up in the sum exactly once.

So, we have that

$$x^{T}Px = x^{T}(dI - A_{G})x + x^{T}\frac{\epsilon}{2}(B - I)x$$
$$\geq \epsilon ||x||^{2} - \frac{2||x||^{2}\epsilon}{2}$$
$$= 0.$$

Next evaluate the numerator and the denominator.

$$-\sum_{P_{ij}<0} d(i,j)^2 P_{ij} = dn$$
$$\sum_{P_{ij}>0} d(i,j)^2 P_{ij} \ge \frac{\epsilon}{2} n \lfloor \log_d n \rfloor^2$$

where the latter follows since the distances of edges in B are at least  $\lfloor \log_d n \rfloor$ . Thus, for this P, we have that:

$$\sqrt{\frac{\sum_{p_{ij}>0} p_{ij} d(x_i, x_j)^2}{-\sum_{p_{ij}<0} p_{ij} d(x_i, x_j)^2}} \ge \sqrt{\frac{\frac{\epsilon}{2} n \lfloor \log_k n \rfloor^2}{dn}} \sim \Theta(\log n)$$

and thus, from (3), that  $C_2$  is at least this big, *i.e.*, that:

$$C_2(G) \ge \Omega(\log n)$$

# References

[1] N. Linial and A. Wigderson, "Expander graphs and their applications"