CS-774 Spectral Graph Theory Aug 14 and 17, 2023

Lecture 3: The Arora-Barak-Steurer Algorithm

Lecturer: Akash Kumar *Scribe: Akash Kumar*

1 A Motivating Discussion: A New Coloring Problem?

In this lecture, we will discuss a coloring problem[1]. So, you know 2-coloring can be solved in polynomial time. On the other hand k-coloring is NP-Complete for $k \geq 3$. Back in 2002, Subhash Khot proposed a generalization of the 2-coloring problem. It is an intriguing generalization because we do not yet know whether it can be solved in polynomial time or it is NP-Complete. Before we state this problem, let us first consider a warmup. Like you know, in 2-coloring problem you are given a graph $G = (V, E)$ and you want to assign colors to its vertices from a palette with $k = 2$ colors such that all edges have their end points colored differently. As a warmup, Khot invented the following problem: You are given an integer $k \in \mathbb{N}$ and a graph $G = (V, E)$. Each edge $e = (u, v)$ announces a permutation $\pi_e \colon [k] \to [k]$ where the set of colorings accepted by this edge is form $(\ell_u, \pi_e(\ell_u))$. Your goal is to find a coloring of all the vertices which satisfies all the constraints or report that no such coloring exists. It is not difficult to see that this problem can be solved in polynomial time using an algorithm almost identical to the algorithm for the 2-coloring problem. Below, I sketch this algorithm, which we call color propagation algorithm.

`ColorPropagation`(G)

1. Apply color c to vertex 1.
2. Pick any $v \in N(u)$ where u is colored. Use the "edge rule" $\pi_{(u,j)}$ to color j with the unique color $\pi_{(u,j)}(\ell_u)$.
3. Continue until all the vertices are colored.
4. If all the edge rules are satisfied the graph is satisfiable. If not go back to step (1) and try another color c'. If all the colors have been tried, then the graph is not satisfiable.

So much for the warmup. It ended up being fairly easy. This is where Khot considers a variant and goes beyond the warmup above. You are given an integer $k \in \mathbb{N}$ and a graph $G = (V, E)$ as in the warmup above. You are promised one of the following situations is true for some sufficiently small $\varepsilon > 0$:

`UniqueLabelingProblem`(G, ε)

1. There exists $\ell \in [k]^n$ so that $\geq (1 - \varepsilon)$ fraction of the edges are satisfied.
2. For all $\ell \in [k]^n$, ℓ satisfies at most ε fraction of the edges.

For a little more intuition, we consider another special case of the problem above where all permutations are cyclic. Indeed, among all things, this new formulation uses linear equations (mod k).

[1] A good part of this lecture borrows from the excellent and extremely educational blog post of [Lipton]

So, in this formulation, you are given

$$\text{lin. eqns of the form } x_i + x_j = c_{ij} \pmod{k}.$$

Denote the given linear system as $\boldsymbol{A}$ and think of k as being some large constant and suppose the system $\boldsymbol{Ax} = \boldsymbol{c}$ given to you is solvable over $\mathbb{F}_k$. Then you can in fact find a satisfying assignment – you just use Gaussian Elimination or if you are feeling particularly adventurous, you can again use the color propagation algorithm from before. We can again ask the question we asked before. Can you distinguish between the following two situations:

`UniqueAssignmentProblem`$(\boldsymbol{A}, \varepsilon)$

1. There exists $\sigma \in [k]^n$ which satisfies $\geq (1-\varepsilon)$ fraction of the equations.
2. There is no $\sigma \in [k]^n$ which satisfies more than ε fraction of the equations.

Khot conjectured that this problem is NP-Complete. This question is still open and pursuing this question has led to some very deep advances in our understanding of approximability (or *inapproximability*). Indeed, if you know about it, what we just considered above is in fact the (in)famous *Unique Games Problem*. For completeness sake, I will state the *Unique Games Conjecture* formally as well[2].

The Unique Games Conjecture:
For any $\varepsilon > 0$, $\exists k = k(\varepsilon)$ s.t. the `UniqueLabelingProblem`(G, ε) problem is NP-hard.

[ABS15] ↓ UGC in 2^{n^2} time.

Alright, now we understand what the *unique games problem* says. But we still have a day to kill. So, for the rest of this lecture, we would see an algorithm due to Arora-Barak-Steurer [ABS15] which solves this problem in subexponential time (like $2^{n^{0.01}}$ish time). This can be a bit unsettling at first blush. Indeed, it appears difficult to square such a running time with the conjecture which boldy asserts that this problem is NP-Complete. The reason for this discomfort is the *Exponential time hypothesis* which asserts that there is no $2^{o(n)}$ time algorithm for the 3-SAT problem. The following remark should help you out of that discomfort.

Remark 1.1. It is conceivable that assuming ETH does not straightaway refute the Unique Games Conjecture. After all, it is still possible that a reduction from a 3-SAT instance of size n runs in polynomial time and returns a Unique Labeling Instance of size $n^{1/\varepsilon}$.

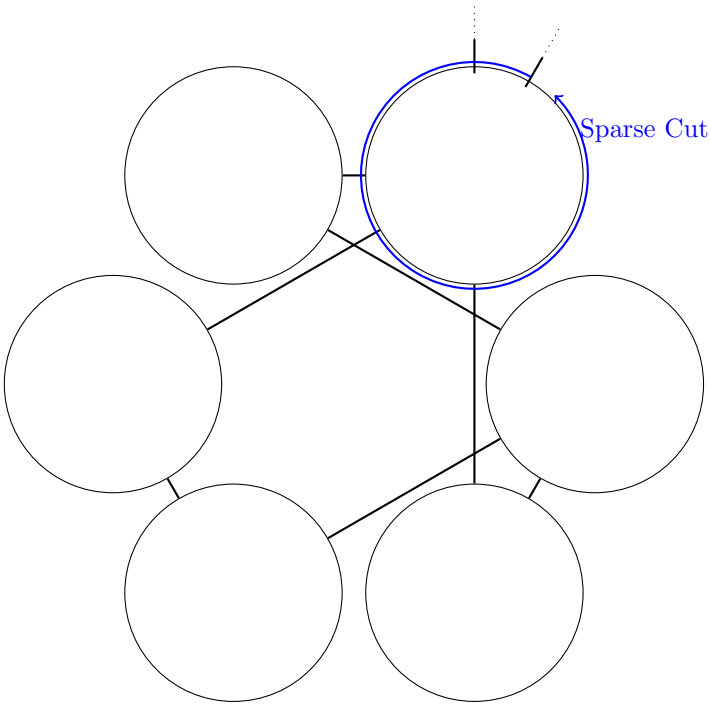
At this stage you might want to see a slightly non-trivial algorithm for Unique Labeling. Of course, there is the trivial brute force solution and the fairly intricate (but beautiful, in my opinion) ABS algorithm which we will cover in today's lecture. Is there something else, kind of a naive approach one might try? Indeed, there is one algorithmic apporach we describe now. As you will see it helps understand the parameter regime for which you might want to approach the conjecture better. So, take a graph with maximum degree at most d. Note that you can color all the edges with at most $d+1$ colors so that no two edges which share a vertex get the same color. Consider the biggest color class which is the biggest set consisting of all edges colored the same way. This contains at least a

[2] You can tease out the statement yourself if you prefer.

$1/(d+1)$ fraction of the total number of edges in G. Note that you can find a coloring restricted to all of these edges which satisfies just these edges! You may extend this to a full coloring of all the vertices and you will, at the end, satisfy at least $1/(d+1)$ fraction of the edges. Well, that does falsify unique games for $1/(d+1) \geq \varepsilon$ and you should think of the conjecture as being asserted for graphs with maximum-degree at least $2/\varepsilon$. Indeed, we know that unique labeling problem on *regular graphs* with $d \geq 2/\varepsilon$ is at least as hard as the general problem and in the rest of this lecture, we will focus on the regular case.

2 The high level ABS approach

The high level approach to the unique labeling problem pioneered by [ABS15] is one of the gems of modern algorithms. In particular, their approach emphasizes the humongous applicability of the expander decomposition toolkit as an enabler of divide-and-conquer algorithms for algorithmic problems on graphs. The key idea is to break the vertex set of the underlying graph into a union of "almost expanders." Then using a postprocessing routine, inspired from [KT07], [ABS15] solve the unique labeling problem on these almost expanders. Their method gives up on the cross edges between the pieces which they ensure are not too many. Here is a pictorial illustration of the high level approach. The circles in this picture denote the expander like pieces [ABS15] decomposition produces which are sparsely connected in the original graph.



3 ABS decomposition

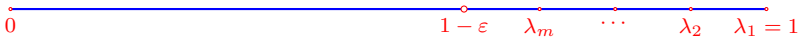
Believe it or not, we are now ready for the main event. It will be useful to list out a few parameters which will aid the rest of the discussion. It is perhaps useful to collect them all at the outset.

- $n \in \mathbb{N}$, the number of vertices in the original graph.
- $\varepsilon > 0$.
- $\beta = O(\varepsilon^{1/3})$.
- $m = n^{O(\beta)}$.

We start by clarifying what almost expanders mean for us.

Definition 3.1. *Fix some sufficiently small $\varepsilon > 0$. Take a d-regular graph $G = (V, E)$ with the lazy random walk matrix being $\boldsymbol{W}$. We say that the $(1-\varepsilon)$ threshold rank of G is at most m if the number of eigenvalues of $\boldsymbol{W}$ which exceed $(1-\varepsilon)$ is at most m. This is written as,*

$$\mathsf{rank}_{1-\varepsilon}(G) = |\{\lambda \colon \lambda \text{ is an eigenvalue of } \boldsymbol{W}, \lambda \geq 1-\varepsilon\}| \leq m.$$



We would like to break an arbitrary d-regular graph into a collection of subsets where each subset induces an almost expander. Before we show such a result, it will be convenient to do a warmup where we show that you can break a graph into a union of expanders provided you do not insist on too much expansion.

Lemma 3.2. *The vertex set of any d regular, n vertex graph can be partitioned into some subsets where each subset induces a $\phi \geq 1/\log^3 n$ expander at the expense of removing at most $o(n)$ edges in time poly(n).*

Proof Sketch for Lemma 3.2. Charging argument. We first present a non-alorithmic version to obtain this partitioning. We will see it can be readily converted to an algorithmic partitioning as well. So, if the graph is already an ϕ expander, we are done without deleting any edges. Otherwise, the graph has a cut with conductance at most ϕ. Let us denote this cut as $(V_S, \overline{V_S})$ where w.log, we let $|V_S| \leq n/2 \leq |\overline{V_S}|$. We remove the edges crossing this cut and we construct a tree T level by level whose root vertex corresponds to $V(G)$, the left child of root corresponds to V_S and the right child corresponds to $\overline{V_S}$. The number of edges removed is at most $\phi d|V_S|$. For each deleted edge $(u, v) \in E(V_S, \overline{V_S})$, add a self-loop each at u and v (this ensures the components remain d-regular). If either among V_S or $\overline{V_S}$ induce ϕ expanders now, we need not refine that component further and we mark such a node as a leaf. On a component which is not yet a ϕ-expander (possibly both), we recursively refine that component by again finding a ϕ sparse cut and extending the tree with a new left and right child of the refined node. At an intermediate stage, this recursive process encounters a set A of vertices. If the graph induced on A (with the self-loops added to maintain d-regularity) is a ϕ-expander, we stop refining A. Otherwise, there exists a cut $(A_S, \overline{A_S})$ with conductance at

most ϕ where w.log we have $|A_S| \leq |A|/2 \leq |\overline{A_S}|$. This time note that the number of edges crossed is at most $\phi d|A_S|$.

We now get into the details of our charging argument. Let $(A_S, \overline{A_S}), |A_S| \leq |\overline{A_S}|$ denote the set of vertices you obtain after deleting a cut with sparsity less than ϕ at some intermediate node in the tree. Let us collect ϕd amount of money from every vertex on the smaller side. In total, the smaller side pays up $\phi d|A_S| \geq |E(A_S, \overline{A_S})|$. That is, the amount of money collected from a vertex on the smaller side upperbounds the number of edges lost in the refinement. Also, whenever a vertex pays up, the size of the connected component it is sent to after refinement, shrinks by 1/2. Thus, in total, a vertex can pay up only $\log n$ times. And the total amount paid by a single vertex therefore, is at most, $\phi d \cdot \log n$. Thus, in total, the number of edges lost is at most $\phi dn \cdot \log n$ which is at most $dn/\log^2 n$. And at the end of this process, all the components are indeed ϕ expanders. You can make this algorithmic by using Cheeger's algorithm to recursively find all the sparse cuts. In all, this would cut $dn/\sqrt{\log n}$ edges. ■

With this warmup out of the way, we are now ready to state the main theorem of this section.

Theorem 3.3. *The vertex set of every d-regular n vertex graph can be partitioned into some subsets $V_1, V_2 \ldots V_b$ such that each subset induces a graph which has $\mathsf{rank}_{1-\varepsilon}(G[V_i]) \leq m = n^{O(\beta)}$ for every $i \leq b$ and the number of edges between these sets is at most 1% fraction of the number of edges in G.*

We show the following lemma which is a big stepping stone towards this result. It asserts that in a graph with $\lambda_m \geq 1 - \varepsilon$, there exists a small set which has small edge boundary to the rest of the graph which you can find efficiently.

Lemma 3.4. *Fix $\varepsilon > 0$. Then, there exists an algorithm which on input a graph H with $n_h \leq n$ vertices and $\mathsf{rank}_{1-\varepsilon}(H) \geq m = n^{O(\beta)}$ – that is, $\lambda_m(\boldsymbol{W}_H) \geq 1-\varepsilon$, runs in time poly$(n)$ and returns a set $S \subseteq V(H)$, which satisfies the following properties:*

- $|S| \leq |V|/m$, *and*
- $\varphi(S) \leq O(\sqrt{\frac{\varepsilon}{\beta}})$.

With this lemma in our hand, we can finish proving Theorem 3.3. Below, we only sketch this proof. You are invited to fill the details in.

Proof Sketch for Theorem 3.3. The idea is to use a argument identical to Lemma 3.2. This time, you just apply Lemma 3.2 with different parameters. Also, you note that the tree is built recursively not by using Cheeger cuts, but rather by using the cuts you get from an application of Lemma 3.4. The size of the smaller side is going down *really fast*. So, a vertex does not pay up $\log n$ times. It only pays at most $1/\beta$ times. You should fill in the details and verify that this gives you the cut bound. ■

It remains to show Lemma 3.4 which is the more challenging part. Your intuition/experience with the proof of the Cheeger bound suggests that it is enough to find *an indicator vector $\mathbf{1}_S$ for a set S* with small Rayleigh Quotient. Intuitively, one natural idea might be to search for a vector like this in the eigenspace spanned by the top m eigenvectors of $\boldsymbol{W}_H$. However, you also want S to have a

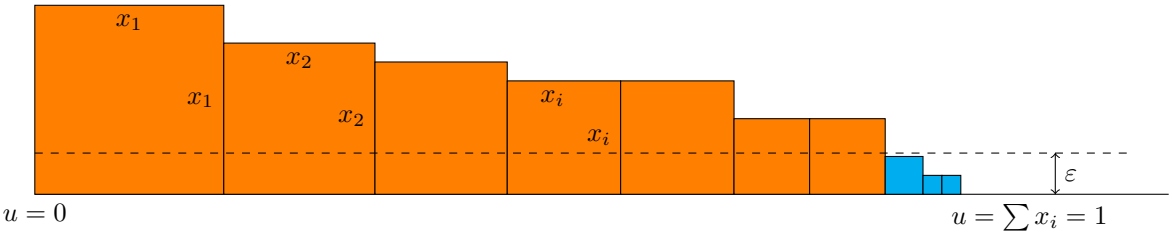
small size viz, $|S| \leq n_H/m$ which is the same thing as insisting supp($\mathbf{1}_S$) to be small. How can be build in an awareness ot support size in the Cheeger argument? Apriori, the space spanned by the top m eigenvectors *does not even have to contain a sparse boolean vector(!)* What else can we try now? [ABS15] offers an insightful way forward. The key is to make the following definition.

Definition 3.5. *A vector $\boldsymbol{x} \in \mathbb{R}^n$ is called δ-analytically sparse if*

$$\|\boldsymbol{x}\|_1 \leq \sqrt{\delta n} \cdot \|\boldsymbol{x}\|_2.$$

Remark 3.6. You call a boolean vector $\boldsymbol{x} \in \{0,1\}^n$ combinatorially sparse if very few coordinates in this vector are 1. Analytic Sparsity attempts to extend a similar definition to real valued vectors $\boldsymbol{x} \in \mathbb{R}^n$. A good way to think about such vectors is to consider what it says for vectors with $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_1 = 1$. For these vectors you get $\|\boldsymbol{x}\|_2^2 \geq \frac{1}{\delta n}$ and these vectors are "morally sparse" because a majority of the entries in such vectors have small absolute value. See picture. If threshold is $\varepsilon = 1/2\delta n$, only $\frac{1}{\varepsilon} = 2\delta n$ entires can be above it (sparsity) and total area below this threshold line is at most $\varepsilon = 1/2\delta n$ (approximate area in the orange bars).

#ques: why this true?



We would like to proceed onward with our plan to Lemma 3.4. It would be good to have the following intermediate claim.

Claim 3.7. *Fix $\varepsilon > 0$ and let H denote a graph on n_H vertices where $\mathsf{rank}_{1-\varepsilon}(H) \geq m = n^{O(\beta)}$ – that is, $\lambda_m(\boldsymbol{W}_H) \geq 1 - \varepsilon$. Let $t = \frac{\beta}{\varepsilon} \cdot \log n$. Then, $\exists u \in V(H)$ such that $\|\boldsymbol{W}_H^t \mathbf{1}_u\|_2^2 \geq n^\beta/n_H$.*

Proof of Claim 3.7. Write $\boldsymbol{W} = \boldsymbol{W}_H$ for notational convenience. Note

$$\begin{aligned}
\mathrm{Tr}(\boldsymbol{W}^{2t}) &= \sum_{u \in V(H)} \mathbf{1}_u^T \boldsymbol{W}^{2t} \mathbf{1}_u \\
&= \sum_{1 \leq i \leq n_H} \lambda_i(\boldsymbol{W}^{2t}) \\
&\geq m \cdot (1-\varepsilon)^{2t} && \text{(large eigs)} \\
&\geq n^{O(\beta)} n^{(-\beta)} && \text{Since } m = n^{O(\beta)} \\
&\geq n^\beta && \text{By choice of constant in big O(.)}
\end{aligned}$$

#ques: huh?

In all, this means

$$\sum_{u \in V(H)} \mathbf{1}_u^T \boldsymbol{W}^{2t} \mathbf{1}_u = \sum_{u \in V} \|\boldsymbol{W}^t \mathbf{1}_u\|_2^2 \geq n^\beta.$$

By averaging, we get a vertex $u \in V(H)$ such that $\|\boldsymbol{W}^t \mathbf{1}_u\|_2^2 \geq n^\beta/n_H$. ■

We now show how Lemma 3.4 follows using Claim 3.7. The high level intuition for this proof is fairly vivid and it is given below. The above claim tells you that there exists a vertex $u \in V(H)$ random walks from which have large 2 norm at the end of t steps. This is another way of saying random walks do not mix well in $V(H)$ after t steps[3]. In this case, you can show that failure to mix after the end of t steps implies a blatant failure to mix (or of the 2-norm to drop) between two succesive steps. But, *this probability vector with large 2-norm is analytically sparse*! We can zoom into in this "bad" step which uses an analytically sparse vector to find a low conductance cut of small size. Details follow.

Proof Sketch for Lemma 3.4. For ease of notation in this proof, let us write $\boldsymbol{W} = \boldsymbol{W}_H$. From Claim 3.7, we know $\exists u \in V(H)$ such that

$$\|\boldsymbol{W}^t \mathbf{1}_u\|_2^2 \geq n^\beta / n_H \geq n^\beta / n.$$

We first show that there is a pair of successive steps, step $1 \leq s, s+1 \leq t$ such that

$$\|\boldsymbol{W}^{s+1} \mathbf{1}_u\|_2^2 \geq \left(1 - \frac{\varepsilon}{\beta}\right) \|\boldsymbol{W}^s \mathbf{1}_u\|_2^2.$$

Indeed, if this fails from one step to the next for all successive pairs of steps, you note that you would have $\|\boldsymbol{W}^t \mathbf{1}_u\|_2^2 \leq (1 - \varepsilon/\beta)^t$. By the choice of $t = \frac{\beta}{\varepsilon} \cdot \log n$, this means $\|\boldsymbol{W}^t \mathbf{1}_u\|_2^2 \ll 1/n$ which is a contradiction. Now, consider a step s such that you have $\|\boldsymbol{W}^{s+1} \mathbf{1}_u\|_2^2 \geq (1 - \varepsilon/\beta) \|\boldsymbol{W}^s \mathbf{1}_u\|_2^2$. Write $\boldsymbol{x} = \boldsymbol{W}^s \mathbf{1}_u$. Thus, $\boldsymbol{W}\boldsymbol{x}$ has a large 2-norm which means $\boldsymbol{x}$ has an even larger 2-norm and it is therefore analytically sparse! Claim 3.8 (when invoked with $\alpha = \varepsilon/\beta$) shows how you can use an analytically sparse vector to find a set $S \subseteq V(H)$ with small conductance where $|S| \leq n_H/m$. Below, we give an intuitive sketch of how this argument goes. You should try and fill in the details yourself before you read up the proof of Claim 3.8.

$$\begin{aligned}
&\|\boldsymbol{W}\boldsymbol{x}\|_2^2 \geq (1 - \varepsilon/\beta) \cdot \|\boldsymbol{x}\|_2^2 \\
\implies &\boldsymbol{W}\boldsymbol{x} \cdot \boldsymbol{W}\boldsymbol{x} \geq (1 - \varepsilon/\beta) \cdot (\boldsymbol{x} \cdot \boldsymbol{x}) \\
\implies &\boldsymbol{W}^2\boldsymbol{x} \cdot \boldsymbol{x} \geq (1 - \varepsilon/\beta) \cdot (\boldsymbol{x} \cdot \boldsymbol{x}) && \text{Because } \boldsymbol{W} \text{ is symmetric}
\end{aligned}$$

We now sketch how to use this *analytically sparse vector* $\boldsymbol{x}$ to finish up by producing the desired set S. Note that the last step above means that $\boldsymbol{x}$ is almost an eigenvector of $\boldsymbol{W}^2$ with eigenvalue $(1 - \varepsilon/\beta)$. Since eigenvalues of $\boldsymbol{W}^2$ are squares of eigenvalues of $\boldsymbol{W}$, it follows that one of the eigenvalues of $\boldsymbol{W}$ is at least $(1 - \frac{\varepsilon}{\beta})$ as well. In turn, this means you can use Cheeger to find a set $S \subseteq V(H)$ with conductance at most $O\left(\frac{\varepsilon}{\beta}\right)$ as desired. Finally, as shown in Claim 3.8, this set S also has size at most n_H/m.

■

Next up, we prove Claim 3.8.

[3] Indeed, since H is a regular graph, the stationary distribution is uniform. Note that uniform distribution over a universe has the smallest 2-norm. So, if the 2-norm is large, walk has not mixed.

Claim 3.8. *Let H be a regular graph on n_H vertices with lazy random walk matrix $\boldsymbol{W} = \boldsymbol{W}_H$ and suppose $\boldsymbol{x}$ is a δ-analytically sparse vector which satisfies $\frac{\|\boldsymbol{W}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \geq (1-\alpha)$ for some sufficiently small $\alpha > 0$. Then, there exists a set $S \subseteq V(H)$ with $|S| \leq \delta n$ such that $\varphi_H(S) \leq 8\sqrt{\alpha}$.*

Recall the proof sketch for Lemma 3.4 cryptically suggests a certain vector is "almost an eigenvector" for the matrix $\boldsymbol{W} = \boldsymbol{W}_H$. This vector is only known to have a Rayleigh Quotient (with respect to $\boldsymbol{W} = \boldsymbol{W}_H$) which is close to 1. The preceding argument then concludes by Cheeger rounding on this "almost eigenvector". Claim 3.8 seeks to establish this rigorously. You want to build size-awareness into the Cheeger bound to prove the above lemma. To this end, you perhaps might hit upon the following plan. Just like we did in Cheeger's proof, you might want to sample a cut from a distribution over cuts you obtain using an analytically sparse vector. Indeed, as [ABS15] showed, this plan indeed works. The proof does some cleanup as a preprocessing step to obtain an analytically sparse $\boldsymbol{y}$ which interfaces better when working with Rayleigh Quotients with respect to $\boldsymbol{W} = \boldsymbol{W}_H$.

Proof of Claim 3.8. As things stand, the vector $\boldsymbol{x}$ satisfies some condition about the 2-norm squared of $\|\boldsymbol{W}\boldsymbol{x}\|$. We want to instead produce a condition that gives quantitative bounds on Rayleigh Quotients of $\boldsymbol{x}$ with respect to $\boldsymbol{W}$. To this end, first note that if $\|\boldsymbol{W}\boldsymbol{x}\|_2^2 = \boldsymbol{W}\boldsymbol{x} \cdot \boldsymbol{W}\boldsymbol{x} = \boldsymbol{x} \cdot \boldsymbol{W}^2\boldsymbol{x}$ is large, then the correlation between $\boldsymbol{x}$ and $\boldsymbol{W}\boldsymbol{x}$ is only larger. In other words, this just says that the next one step distribution is more correlated with the current distribution as compared to the correlation between the current distribution and the distribution you end up with after next two steps. This should be fairly intuitive and it follows directly by noting that $\boldsymbol{W}$ is symmetric and $\boldsymbol{W} - \boldsymbol{W}^2$ is psd (think why this psdness implies that correlation drops as the distribution evolves with more and more random walk steps)[4].

Okay, so the following holds:

$$\frac{\boldsymbol{x} \cdot \boldsymbol{W}\boldsymbol{x}}{\|\boldsymbol{x}\|_2^2} \geq \frac{\|\boldsymbol{W}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \geq 1 - \alpha \tag{3.1}$$

Next, we modify $\boldsymbol{x}$ to obtain a 4δ-analytically sparse vector $\boldsymbol{y}$ which has small support as a real valued vector. By scaling, we may assume that $\|\boldsymbol{x}\|_2^2 = \delta n$ and $\|\boldsymbol{x}\|_1 \leq \delta n$. The key to obtain a vector $\boldsymbol{y}$ is a simple truncation argument. Define

$$\boldsymbol{y}_u = \begin{cases} \boldsymbol{x}_u - 1/4 \text{ if } \boldsymbol{x}_u \geq 1/4 \\ 0 \text{ Otherwise.} \end{cases}$$

It is clear that $|\operatorname{supp}(\boldsymbol{y})| \leq 4\delta n$ because $\|\boldsymbol{x}\|_1 \leq \delta n$ and each non-zero $\boldsymbol{y}_u$ corresponds to $u \in V(H)$ with $\boldsymbol{x}_u \geq 1/4$. Let us now compare the Rayleigh Quotients of unit vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ with respect to $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{\mathcal{L}}$. To do this, it suffices to compare Rayleigh Quotients of these vectors with $\boldsymbol{\mathcal{L}}$. So, consider

$$\mathcal{R}_{\boldsymbol{\mathcal{L}}}(\boldsymbol{x}) = \frac{\sum_{(i,j)\in E(H)} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2}{d \sum \boldsymbol{x}_i^2} \quad ; \quad \mathcal{R}_{\boldsymbol{\mathcal{L}}}(\boldsymbol{y}) = \frac{\sum_{(i,j)\in E(H)} (\boldsymbol{y}_i - \boldsymbol{y}_j)^2}{d \sum \boldsymbol{y}_i^2}.$$

Note that truncations can never make an edge longer and therefore for any $(i,j) \in E(H)$ we have $(\boldsymbol{y}_i - \boldsymbol{y}_j)^2 \leq (\boldsymbol{x}_i - \boldsymbol{x}_j)^2$. So, the numerator of the Rayleigh Quotient of $\boldsymbol{y}$ is no larger than the

[4] An important comment is that this drop also holds even when H is not regular.

numerator of Rayleigh Quotient of $\boldsymbol{x}$ (wrt $\boldsymbol{\mathcal{L}}$). Now, we want to compare the denominators. To this end, note that for all $i \in V(H)$, $\boldsymbol{y}_i^2 \geq \boldsymbol{x}_i^2 - \frac{\boldsymbol{x}_i}{2}$. This means,

$$\sum \boldsymbol{y}_i^2 \geq \sum \boldsymbol{x}_i^2 - \frac{1}{2} \cdot \sum \boldsymbol{x}_i \geq \delta n/2 = \|\boldsymbol{x}\|_2^2/2.$$

Combining, we have $\mathcal{R}_{\boldsymbol{\mathcal{L}}}(\boldsymbol{y}) \leq 2\mathcal{R}_{\boldsymbol{\mathcal{L}}}(\boldsymbol{x})$. By (3.1), this is at most 2α. Now, putting all of this together, we get

huh? #ques

$$\frac{\boldsymbol{y} \cdot \boldsymbol{W}\boldsymbol{y}}{\|\boldsymbol{y}\|_2^2} \geq 1 - 2\alpha \tag{3.2}$$

So, finally we are ready with a vector on which we would like do Cheeger rounding. The proof is almost identical to how Cheeger rounding is done on non-zero vectors on small Rayleigh Quotient vectors from lecture #1 and is therefore omitted. ■

4 Solving Unique Labeling Problem on Almost Expanders

So far in our discussion, we did not use anything specific to the Unique Labeling Problem. Indeed, our discussion so far, in the preceding section, has completely ignored the labels! Indeed the preceding section only partitioned the underlying graph into almost expanders. In this section we present ideas which underlie the second step of the [ABS15] algorithm and we show how you can find a good unique games labeling on an almost expanding instance. Following [KT07, ABS15], we first introduce a helpful auxiliary graph (Definition 4.1) which aids this argument. In particular, in the rest of this section, we will let G denote an almost expander (or a low-rank graph). Going ahead, a convenient reparametrization would be to set $\varepsilon := \gamma^6$. For your convenience, let me recall all the parameters we will use going forward once more.

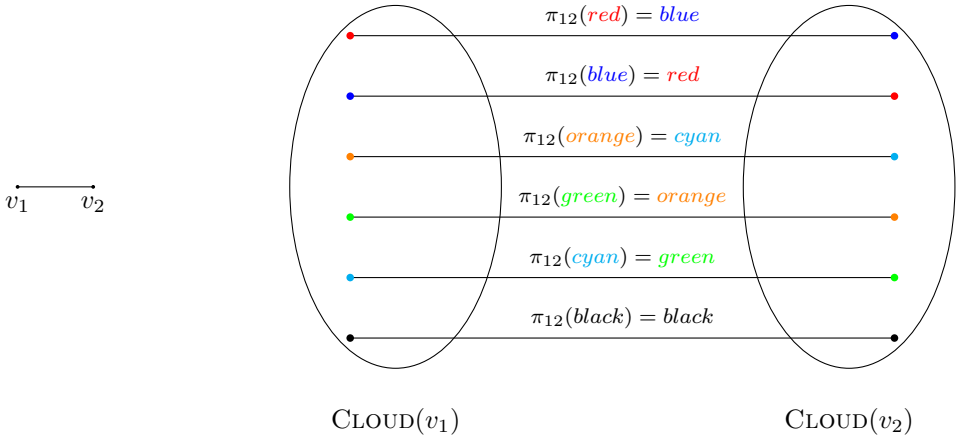
- $n \in \mathbb{N}$, the number of vertices in the original graph.
- $0 < \gamma \ll 0.01$.
- $\varepsilon = \gamma^6$.
- $\beta = O(\varepsilon^{1/3}) = O(\gamma^2)$.
- $m = n^{O(\beta)}$.

Definition 4.1. *Let* $\texttt{UniqueLabeling}(G, \varepsilon, k)$ *denote an instance of the unique labeling problem where* $G = (V, E)$ *is the underlying graph. We define the graph* $\widehat{G} = (\widehat{V}, \widehat{E})$ *where the vertices of* $\widehat{G}$ *have the form* (v, i) *where* $v \in V(G)$ *and* $i \in [k]$*. This is to say, for* $v \in V(G)$*, we let*

$$\textsc{Cloud}(v) = \{(v, 1), (v, 2), \cdots (v, k)\}$$

which is to say $\textsc{Cloud}(v)$ *contains* k *copies of the vertex* v*, one in each color. Note that* $|\widehat{V}| = kn$*. Now, we define* $\widehat{E}$*. An edge in* $\widehat{G}$ *is of the form* $((u, a), (v, b))$ *where* $u, v \in V(G)$ *and the edge rule* π_{uv} *satisfies* $\pi_{uv}(a) = b$*. In other words, given an edge* $(u, v) \in E(G)$*, we include* k *edges in* $\widehat{E}$*.*

These edges form a perfect matching running between vertices in $\textsc{Cloud}(u)$ *and* $\textsc{Cloud}(v)$ *and they satisfy the edge rule* π_{uv} *in the "natural way". This graph is called the* Label-Extended Graph. *The picture below accompanies this definition.*



Edges between cloud(v_1) *and cloud*(v_2) *connect vertices of the form* $(1, a)$ *with* $(2, b)$ *iff* $\pi_{12}(a) = b$.

Suppose the unique labeling instance is completely satisfiable and the satisfying assignment assigns the color $\sigma(v)$ to $v \in V(G)$. Note that this means the set $S \subseteq \widehat{V}$ where $S = \{(v, \sigma(v))\}_{v \in V(G)}$ is disconnected from rest of $\widehat{V}$ (why?[5]). The real utility of the Label Extended Graph stems from Lemma 4.2 which presents a robust version of the foregoing observation. This robust version simply asserts that if you have a good unique labeling, then correspondingly, the label extended graph has a low conductance cut.

Lemma 4.2. *Consider an instance of* $\texttt{UniqueLabeling}(G, \alpha, k)$ *of the unique labeling problem where* $G = (V, E)$ *is a* d*-regular graph with* $d \geq 2/\alpha$*. Suppose this instance admits an assignment that satisfies* $(1-\alpha)$ *fraction of the edges. Let* $\widehat{G}$ *denote the label extended graph (see Definition 4.1) of* G*. Then*

$$\varphi(\widehat{G}) \leq 2\alpha.$$

Proof Sketch for Lemma 4.2. Note that the graph $\widehat{G}$ is also a d-regular graph. Denote by σ the assignment that satisfies maximum fraction of the edges of G. Consider the set

$$S = \{(v, \sigma(v))\}_{v \in V(G)} \subseteq \widehat{V}.$$

We would like to bound the conductance of this set. Next, take an edge $(u, v) \in E(G)$ violated by σ. We can pair this edge with two edges – one incident at $(u, \sigma(u))$ and the other at $(v, \sigma(v))$ –

[5] You just need to verify the following items. If u, v is a non-edge in $E(G)$ then convince yourself that there are no edges between $\textsc{Cloud}(u)$ and $\textsc{Cloud}(v)$. Next, take $(u, v) \in E(G)$. Convince yourself that the only possible edge between the two clouds connects only $(u, \sigma(u))$ with $(v, \sigma(v))$

both of which might cross the cut $(S, \overline{S})$ in $\widehat{G}$. Thus, $|\widehat{E}(S, \overline{S})| \leq 2\alpha d|S|$. This gives the desired bound on the conductance of $\widehat{G}$. ■

How to find good assignment using low conductance cut.

At this stage, your spider-sense might be tingling. You might reflexively ask: could the converse to Lemma 4.2 hold? That is, might a sparse cut in the label extended graph correspond to a good labeling for our `UniqueLabeling` instance? Unfortunately, this is not necessarily true. Indeed, the sparsest cut might correspond to some set $S \subseteq \widehat{V}$ which contains more than one vertex from the same cloud. Sometimes, it is still useful to proceed undeterred. After all, finding sparse cuts is a useful primitive in several other applications. [ABS15, KT07] prove that this is a good time to again pull out our ever-reliable spectral toolkit and think about the span of eigenvectors corresponding to large eigenvalues of the lazy random walk matrix $\widehat{\boldsymbol{W}}$ on the label-extended graph $\widehat{G}$. The key is to show this space contains a real valued vector which can be rounded into a sparse cut which in turn corresponds to a good assignment if G is an almost expander.

Remark 4.3. *There are a few questions you might ask at this stage. Let's list those out.*

1. Okay, so the span of "top" eigenvectors of $\widehat{\boldsymbol{W}}$ only contains a good vector which corresponds to a sparse cut. How do we find it? The brief answer is we kind of brute force it out.
2. How exactly do we get an assignment from this sparse cut? Perhaps this might be evident to you. We will say a few words about it in §4.2.
3. It seems the above discussion skipped a step. We only know G is an almost expander. We want to think about the top eigenspace of $\widehat{\boldsymbol{W}}$ and not $\boldsymbol{W}$. This seems to require a bound on the dimensionality of this top eigenspace, no? How do we get that?

Before we proceed further, let me assign one more exercise out of the way through the following observation. This observation reaffirms that label extended graphs are more than a mere technical artifact and that these graphs are indeed super useful (even spectrally) if you want to attack the unique games problem.

Observation 4.4. *Consider an instance* $\texttt{UniqueLabeling}(G, \alpha, k)$ *of the unique labeling problem where* $G = (V, E)$ *is a* d*-regular graph with* $d \geq 2/\alpha$*. Let* $\widehat{G}$ *denote the label extended graph (see Definition 4.1) of* G*. Then corresponding to every cycle in* G *there are* k*-cycles in* $\widehat{G}$*.*

A bound on number of cycles in $\widehat{G}$ will allow you to bound the trace of some appropriate power of the lazy random walk matrix $\widehat{\boldsymbol{W}}$ on $\widehat{\boldsymbol{G}}$. In turn, trace is the sum of eigenvalues of $\widehat{\boldsymbol{W}}$ raised to that appropriate power which is the spectral connection we will use later.

§4.1 builds ammunition to answer the last question first. Later, in §4.2, I will briefly sketch the answer to the first two questions. A good part of answering these questions will be delegated to the homework.

4.1 Dimensionality of top eigenspace of $\widehat{\boldsymbol{W}}$

Let us start by formally writing out the space we want to control the dimensionality of.

Definition 4.5. *Let $\widehat{\boldsymbol{W}}$ denote the lazy random walk matrix for the label extended graph $\widehat{G}$. The $(1-\gamma^5)$-top eigenspace of $\widehat{\boldsymbol{W}}$ is written as:*

$$\mathcal{U} = \mathcal{U}(\widehat{\boldsymbol{W}}, \gamma^5) = \mathsf{SPAN}\left(\{\boldsymbol{x} : \boldsymbol{x} \text{ is an eigenvector of } \widehat{\boldsymbol{W}} \text{ with eigenvalue } \geq (1-\gamma^5)\}\right).$$

We now bound the size/dimensionality of the space $\mathcal{U}$ defined above. This is done in Lemma 4.6.

Lemma 4.6. *Let* $\texttt{UniqueLabeling}(G, \varepsilon, k)$ *denote an instance of the unique labeling problem which admits an assignment that satisfies at least $(1-\varepsilon)$-fraction of the constraints. Suppose $\lambda_m \geq 1-\alpha$ where*

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$$

denote the eigenvalues of the lazy random walk matrix $\boldsymbol{W}$ of G. Let $\widehat{G}$ denote the label-extended graph of G. Denote the eigenvalues of $\widehat{\boldsymbol{W}}$ as $1 = \widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \widehat{\lambda}_{kn} \geq 0$.

$$\text{Then for any } 0 < \delta < 1, \widehat{\lambda}_{km \cdot n^{O(\beta)}} < 1 - \delta \cdot \alpha.$$

We prove this lemma by exploiting the trace connection alluded to eariler. To cement this connection, let us note that the proof of Theorem 3.3 actually shows a bit more. In particular, it shows that the trace of an almost expander is not too large. Details follow.

Remark 4.7. Recall that the key ingredient in proving Theorem 3.3 was Lemma 3.4 which, in turn, was proved using Claim 3.7. Indeed, this is what you might call a freely upgradable claim. Having fixed $t = \frac{\beta}{\varepsilon} \cdot \log n$, you essentially conclude that if $\mathrm{Tr}(\boldsymbol{W}^{2t}) \geq n^{\beta}$ then the graph contains a small non-expanding set. Indeed, let $G[V_i]$ denote the graph induced on an almost expander produced by Theorem 3.3 and $\boldsymbol{W}$ denote the lazy random walk matrix for this induced expander. Note that you get $\mathrm{Tr}\, \boldsymbol{W}^{2t} < n^{\beta}$. The reason this is an upgrade or a stronger result is because it starts from a weaker premise.

We first show Claim 4.8 which is a convenient stepping stone towards proving the above Lemma. Before we proceed further, it is convenient to introduce the following variant of rank of a lazy random walk matrix. For a lazy random walk matrix $\boldsymbol{W}$, for any $k \in \mathbb{N}$, note that $\mathrm{Tr}(\boldsymbol{W}^{2k}) \geq (1-\alpha)^{2k} \cdot \mathsf{rank}_{1-\alpha}(\boldsymbol{W})$. Define

Why equality eventually? #ques

$$r(\boldsymbol{W}) = r_{1-\alpha}(\boldsymbol{W}) = \inf_{k \in \mathbb{N}} \frac{\mathrm{Tr}(\boldsymbol{W}^{2k})}{(1-\alpha)^{2k}}.$$

In the rest of this document, we pretend that this infimum is in fact the minimum.

Claim 4.8. *Let* $\texttt{UniqueLabeling}(G, \varepsilon, k)$ *denote an instance of the unique labeling problem which admits an assignment that satisfies at least $(1-\varepsilon)$-fraction of the constraints. Suppose $r_{1-\alpha}(\boldsymbol{W}) \leq m$ where $\boldsymbol{W}$ denotes the lazy random walk matrix of G. Let $\widehat{G}$ denote the label-extended graph of G.*

$$\text{Then } r_{1-\alpha}(\widehat{\boldsymbol{W}}) \leq km.$$

Proof. Note that each cycle in G has k copies in $\widehat{G}$ (Observation 4.4). Fix any $t \in \mathbb{N}$. You note

$$r_{1-\alpha}(\widehat{\boldsymbol{W}}) \leq \frac{\mathrm{Tr}(\widehat{\boldsymbol{W}}^{2t})}{(1-\alpha)^{2t}} \leq \frac{k^{2t} \cdot \mathrm{Tr}(\boldsymbol{W}^{2t})}{(1-\alpha)^{2t}}.$$

cycle counting = t length walks to same vertex ⇒ k (t length cycle)

$\mathrm{tr}(A^t) \leq k^t\, \mathrm{tr}(A^t)$

$\mathrm{tr}\left(\left(\frac{I+D}{2}\right)^t\right)$ ⇝ bound, show term by term

#ques: bound = $k^{2t} r_{1-\alpha}(W)$ $\leq k^{2t} m$?

This is true for every t and on choosing t which achieves the infimum, we get $r_{1-\alpha}(\widehat{\boldsymbol{W}}) \leq k \cdot r_{1-\alpha}(\boldsymbol{W}) \leq km$. ■

As a (somewhat) challenging exercise, I invite you to show the following claim. I would like to point out this is the claim which (a little unexpectedly) uses the upgraded version of Theorem 3.3 mentioned in Remark 4.7. In particular, we exploit the fact that almost expanding graphs have small trace.

Claim 4.9. *Suppose $r_{1-\alpha}(\widehat{\boldsymbol{W}}) \leq m$. Then for any $0 < \delta < 1$, we have $\widehat{\lambda}_{m \cdot n^{O(\beta)}} < 1 - \delta\alpha$.* #ques: How to get this

Lemma 4.6 now follows an immediate corollary.

Proof of Lemma 4.6. Follows immediately from Claim 4.8 and Claim 4.9. ■

4.2 Finding a good assignment/vector in $\mathcal{U}$

Finally, we are ready to carry out the first two steps of the plan laid out in Remark 4.3. To this end, recall Lemma 4.2 shows that there exists a set $S \subseteq \widehat{V}$ with conductance at most $2\varepsilon = 2\gamma^6$. I will first argue that you can find a vector in $\mathcal{U}$ which is pretty close to the "normalized indicator" of S (this is the vector $\boldsymbol{\chi}_S = \frac{\mathbf{1}_S}{\sqrt{|S|}}$). To this end, denote the projector on to the space $\mathcal{U}$ as $\boldsymbol{P}_{\mathcal{U}}$. We will show that the squared norm of the projection of $\boldsymbol{\chi}_S$ on the space $\mathcal{U}$ essentially does not loose any ℓ_2 squared-mass.

Claim 4.10. $\|\boldsymbol{P}_{\mathcal{U}}\boldsymbol{\chi}_S\|_2^2 \geq 0.99\|\boldsymbol{\chi}_S\|_2^2$.

Proof. Recall for every unit vector $\boldsymbol{z} \in \mathcal{U}$, we know $\boldsymbol{z}^T\widehat{\boldsymbol{W}}\boldsymbol{z} \geq 1 - \gamma^5$. Also, recall from Lemma 4.2 that $\varphi_{\widehat{G}}(S) \leq 2\varepsilon$. This means

$$\boldsymbol{\chi}_S \cdot \widehat{\boldsymbol{W}}\boldsymbol{\chi}_S \geq (1 - 2\varepsilon)\|\boldsymbol{\chi}_S\|_2^2.$$ #ques: why?

Suppose, $\|\boldsymbol{P}_{\mathcal{U}}\boldsymbol{\chi}_S\|_2^2 = \alpha$. Write $\boldsymbol{\chi}_S = \boldsymbol{x} + \boldsymbol{y}$ where $\boldsymbol{x} \in \mathcal{U}$ and $\boldsymbol{y} \in \mathcal{U}^\perp$. Now, note

$$\begin{aligned}
\boldsymbol{\chi}_S \cdot \widehat{\boldsymbol{W}}\boldsymbol{\chi}_S &= \boldsymbol{x} \cdot \widehat{\boldsymbol{W}}\boldsymbol{x} + \boldsymbol{y} \cdot \widehat{\boldsymbol{W}}\boldsymbol{y} \\
&\leq \alpha + \|\boldsymbol{y}\|_2^2 \cdot (1 - \gamma^5) \\
&= \alpha + \left(\|\boldsymbol{\chi}_S\|_2^2 - \alpha\right) \cdot (1 - \gamma^5) \\
&\leq \alpha + \left(\|\boldsymbol{\chi}_S\|_2^2 - \alpha\right) \cdot (1 - 1000\varepsilon)
\end{aligned}$$

This gives us both an upperbound and a lowerbound on $\boldsymbol{\chi}_S \cdot \widehat{\boldsymbol{W}}\boldsymbol{\chi}_S$. Chaining these bounds together, we get

$$\begin{aligned}
&(1 - 2\varepsilon)\|\boldsymbol{\chi}_S\|_2^2 \leq \alpha + \left(\|\boldsymbol{\chi}_S\|_2^2 - \alpha\right)(1 - 1000\varepsilon) \\
\implies &\alpha \geq 0.99\|\boldsymbol{\chi}_S\|_2^2 \quad \text{As desired.}
\end{aligned}$$

■

This is great. We confirmed that the subspace spanned by the top few eigenvectors of $\widehat{\boldsymbol{W}}$ contains a vector that indicates a sparse cut in the label extended graph. Two things still need to be checked. First off, we need to show this sparse cut in the label extended graph corresponds to a good assignment for the underlying unique labeling instance. Next up, I need to show how you can find (a good approximation to) this set S. Let us take up the second question first. As hinted earlier, the trick is to kind of brute force this set out using the subspace enumeration tool. Below, I present this procedure. It takes as argument two parameters: a graph G and a real $0 < \varepsilon < 1$ where $\varphi(G) = \phi_G(S) \leq \varepsilon := \gamma^6$. The process returns in time $\exp(\mathsf{rank}_{1-\varepsilon^5}(G))$ a good approximation $\tilde{S}$ to S where $\phi_G(\tilde{S}) = O(\varepsilon)$.

`SubspaceEnumeration`(G, ε)

1. Assign $\mathcal{U} = \mathsf{SPAN}(\{\boldsymbol{x} : \boldsymbol{x}$ is an eigenvector of $\boldsymbol{W}$ with eigenvalue $\geq 1 - \gamma^5\})$.
2. Assign $m = dim(\mathcal{U})$ and $\mathcal{C} = \emptyset$.
3. Intialize Σ_m, the $\sqrt{\varepsilon/100\gamma^5} = 0.1\sqrt{\gamma}$-net in the m-dimensional unit ball of $\mathcal{U}$.
4. For $\boldsymbol{v} \in \Sigma_m$,
   (a) For $s \in \{1, 2, \cdots n\}$
      i. Add to $\mathcal{C}$ the set $\{i \colon \boldsymbol{v}(i) \geq 1/2\sqrt{s}\}$.
5. Return $\mathcal{C}$.

Algorithmically, we make a call `SubspaceEnumeration`$(\widehat{G}, \varepsilon)$. This returns a collection of sets, $\mathcal{C}$. I claim this collection contains a set $\tilde{S} \subseteq \widehat{V}$ with conductance $O(\varepsilon)$ such that $\tilde{S}$ has small symmetric difference with S – in particular, $|S \oplus \tilde{S}| \leq O(\gamma|S|)$ (think why `SubspaceEnumeration` gives such a guarantee).

Getting a coloring once you have the set $\tilde{S}$ is pretty automatic. If $\tilde{S}$ misses $\textsc{Cloud}(v)$, then v is given an arbitrary color. Otherwise, if $|\tilde{S} \cap \textsc{Cloud}(v)| > 0$, then v is assigned an arbitrary color from this overlap. You need to argue this assignment gives a good enough solution to `UniqueLabeling` problem. I will leave this as a guided homework problem. One way to approach this question is to exploit the small symmetric difference between S and $\tilde{S}$ I alluded to above.

References

[ABS15] Sanjeev Arora, Boaz Barak and David Steurer, Subexponential Algorithms for Unique Games and Related problems *J.ACM*, 2015

[KT07] Alexandra Kolla and Madhur Tulsiani, Playing unique games using graph spectra *manuscript*, 2007

[Lipton] Richard J. Lipton "Unique Games: A Three Act Play," https://rjlipton.wpcomstaging.com/2010/05/05/unique-games-a-three-act-play/