Recall this version of the PCP theorem from last lecture.

Theorem 1. *There exists an alphabet Σ and a constant $\varepsilon > 0$ for which the following task is* NP*-hard: Given a satisfiable 2CSP instance over Σ, find an assignment that satisfies a $1 - \varepsilon$ fraction of constraints.*

In a general 2CSP instance, a variable may be present in an arbitrary number of constraints. What if we restrict our attention to instances where every variable appears in at most d constraints, where d is small compared to the number of variables? When $d = 1$, every variable appears in one constraint and finding a satisfying assignment is easy. When $d = 2$, the task is a bit harder but still solvable in time linear in n. On the other hand, when d is as large as the number of constraints the problem becomes NP-hard. This suggests that the problem may become gradually harder as d gets larger.

It turns out that this intuition is incorrect:

Theorem 2. *There exists an alphabet Σ and constants d and ε such that given a satisfiable 2CSP instance over Σ where every variable appears in at most d constraints, it is* NP*-hard to satisfy a $1 - \varepsilon$ fraction of the constraints.*

We prove this statement by reduction from Theorem 1. Let Φ be the 2CSP instance in question. We want to construct a new instance Φ' which is as hard as Φ, but every variable appears in at most d constraints. Some of the variables in Φ may appear in more constraints. If variable x_i appears in n_i different constraints it is natural to replace it with n_i new variables $x'_{i1}, \ldots, x'_{in_i}$ and impose some additional constraints that force all of $x'_{i1}, \ldots, x'_{in_i}$ to take the same value.

The first thing we may try is to add the constraints $x'_{i1} = x'_{i2}, x'_{i2} = x'_{i3}, \ldots, x'_{i(n_i-1)} = x'_{in_i}$ to Φ'. Then if Φ has a satisfying assignment, the assignment obtained by setting $x'_{i1} = \cdots = x'_{in_i} = x_i$ will be satisfying for Φ'. If Φ' was easy to approximate, we would then be able to obtain an assignment x' that satisfies a $1 - \varepsilon'$ fraction of its constraints. Can we use x' to extract an assignment that satisfies most constraints in Φ?

It is not hard to see that if $\varepsilon' = 0$, the assignment $x_i = x'_{i1} = \cdots = x'_{in_i}$ is satisfying for Φ. However, even if one of the equality constraints is violated, the values of x'_{ij} could split into two equally sized sets. Then it is not clear which value to assign to x_i and it is possible to come up with examples where no matter which value we assign, a large fraction of the constraints of Ψ will be violated.

So we need to make the equality constraints more robust: If there is no clear majority among the values $x'_{i1}, \ldots, x'_{in_i}$, then not one but many of the equality constraints should be violated. One way to do so is to impose the equality constraint $x'_{ij} = x'_{ij'}$ for every pair $j < j'$; but then we have done nothing about reducing the number of constraints a variable appears in.

In general the equality constraints we are looking for can be described by an undirected graph G on the vertices $\{1, \ldots, t\}$. An assignment to $x'_{i1}, \ldots, x'_{in_i}$ can be viewed as a partition of the vertices into sets $A_\sigma = \{j \colon x'_{ij} = \sigma\}$, where σ ranges over Σ. On the one hand, we want the degree of this graph to be constant. On the other hand, we want that a partition $\{A_\sigma\}$ splits many of the edges of G, unless one of the sets A_σ contains most of the vertices.

Both of these properties are achieved by expander graphs. To understand expander graphs and their properties we first need to take a detour into random walks, adjacency matrices, and eigenvalues.

In what follows we will assume the graph G is connected, d-regular, and non-bipartite.

1 Adjacency matrix and eigenvalues

Suppose a particle sits at a vertex s of some graph G. At every step, s moves to a random one of its neighbors. How long will it take s to reach a vertex in G that looks random?

To answer this question, it will be helpful to represent the random walk by a sequence of probability distributions $\mathbf{p}^0, \mathbf{p}^1, \ldots$ on the vertices of G, with the following interpretation: At each step t, $\mathbf{p}^t(u)$ is the probability of the particle ending up at vertex u after t steps of the walk. Initially, we have $\mathbf{p}^0$ assign probability 1 to vertex s, and probability 0 to all the other vertices. The distribution $\mathbf{p}^{t+1}$ can be calculated from $\mathbf{p}^t$ via the formula

$$\mathbf{p}^{t+1}(u) = \sum_{v:(v,u) \text{ is an edge}} \frac{1}{d} \cdot \mathbf{p}^t(v). \tag{1}$$

We are now interested in the following question: When t gets large, how close does the distribution $\mathbf{p}^t$ get to the uniform distribution $\mathbf{u}$ on the set of vertices? To answer this question, we need some way of measuring how "close" two distributions are. In our setting the most convenient measure is the ℓ_2 *norm*. The ℓ_2 norm of a vector $\mathbf{v}$ is the quantity

$$\|\mathbf{v}\| = \Big(\sum_i \mathbf{v}_i^2\Big)^{1/2}$$

and the ℓ_2 distance between two vectors $\mathbf{v}$ and $\mathbf{v}'$ is the ℓ_2 norm of $\mathbf{v} - \mathbf{v}'$. We will think of probability distributions as vectors in $\mathbb{R}^n$ (with one entry for each vertex in the graph), and we will say that two distributions $\mathbf{p}$ and $\mathbf{p}'$ are ϵ-close (in ℓ_2 distance) if $\|\mathbf{p} - \mathbf{p}'\| \leq \epsilon$.

The (normalized) *adjacency matrix* of G is an $n \times n$ matrix A defined as follows:

$$A_{u,v} = \frac{\text{number of edges between } u \text{ and } v \text{ in } G}{d}$$

This matrix is symmetric and the entries in each row add up to one. Using A, we can write equation 1 in matrix form as $\mathbf{p}^{t+1} = \mathbf{p}^t A$ (it is customary to represent $\mathbf{p}^t$ as row vectors) and so we immediately obtain that $\mathbf{p}^t = \mathbf{p}^0 A^t$.

The eigenvalues and eigenvectors of A play a significant role in determining the behavior of random walks on G. Recall that an eigenvalue-eigenvector pair is a complex number λ and a vector $\mathbf{v}$ such that $\mathbf{v}A = \lambda\mathbf{v}$. It is a basic theorem in linear algebra that symmetric matrices have an orthonormal basis of eigenvectors with real eigenvalues. Let's denote these pairs by $(\lambda_1, \mathbf{v}_1), \ldots, (\lambda_n, \mathbf{v}_n)$ where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. (Some of the λ_i may be negative.)

What is the meaning of this? Initially the position of our particle is determined by the distribution $\mathbf{p}^0$. Since the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form an orthonormal basis we can decompose $\mathbf{p}^0$ in the form

$$\mathbf{p}^0 = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$$

where $\alpha_i = \langle \mathbf{p}^0, \mathbf{v}_i \rangle$ and $\alpha_1^2 + \cdots + \alpha_n^2 = 1$.

After one step of the random walk, the distribution becomes

$$\mathbf{p}^1 = \mathbf{p}^0 A = \alpha_1 \mathbf{v}_1 A + \cdots + \alpha_n \mathbf{v}_n A = \alpha_1 \lambda_1 \mathbf{v}_1 + \cdots + \alpha_n \lambda_n \mathbf{v}_n$$

and after t steps

$$\mathbf{p}^t = \mathbf{p}^0 A^t = \alpha_1 \lambda_1^t \mathbf{v}_1 + \cdots + \alpha_n \lambda_n^t \mathbf{v}_n. \tag{2}$$

Let's think of what happens when t becomes large. We will assume the values α_i are nonzero since the initial position of the particle can be arbitrary.[1] Eventually the right hand side of the expression will be dominated by the term in which λ_i has largest absolute value; this is either $|\lambda_1|$ or $|\lambda_n|$. This absolute value cannot exceed 1, because $\mathbf{p}^t$ would then become very large, but its norm is bounded since it is a probability distribution. Similarly, the absolute value cannot be less than 1 because then $\mathbf{p}^t$ would become very small when t gets large.

Therefore, it must be the case that $\lambda_1 = 1$, and

$$\max\{|\lambda_i| : 2 \le i \le n\} = \max(\lambda_2, -\lambda_n) \le 1.$$

The quantity on the left side is denoted by $\lambda = \lambda(G)$ and plays a very important role because of the following. First, note that $\mathbf{u}A = \lambda_1 \mathbf{u}$, so the eigenvector $\mathbf{v}_1$ associated to $\lambda_1 = 1$ equals $\sqrt{n} \cdot \mathbf{u}$. Now from (2) we have that

$$\|\mathbf{p}^t - \alpha_1 \mathbf{v}_1\|^2 = \alpha_2^2 \lambda_2^{2t} + \cdots + \alpha_n^2 \lambda_n^{2t} \le \lambda^{2t}.$$

The left hand side has a natural interpretation. Recall that $\alpha_1 = \langle \mathbf{p}^0, \mathbf{v}_1 \rangle = 1/\sqrt{n}$, so $\alpha_1 \mathbf{v}_1$ equals the uniform distribution $\mathbf{u}$. Thus λ^t measures how close $\mathbf{p}^t$ gets to the uniform distribution after t steps of the walk: $\|\mathbf{p}^t - \mathbf{u}\| \le \lambda^t$. Another way of saying this is that λ determines the *rate* at which $\mathbf{p}^t$ converges to the uniform distribution: The smaller λ is, the faster we will get to a uniformly random vertex.

2 Expander graphs

How can we design a d-regular graph G such that starting from *any* vertex s, we can reach a random vertex as soon as possible? If $d = n$ we can take G to be the complete graph (with a loop around every vertex), but we want d to be small compared to n. It seems a good idea to make as many vertices of G reachable using short walks out of s. This suggests that G should look like a tree rooted at s.

If we start as s, after about $\log_d n$ steps we will find the particle near the leaves of the tree. However, the particle is unlikely to stick at any particular leaf because there is only one path leading to it. A random walk on the tree favors the interior vertices, so the vertex at which the particle ends up won't look random.

In some sense, this is a bit unfair because the leaves have degree one, and the graph is not d-regular. We can "connect up" the leaves in some way so as to make the graph be d-regular. Once we do

[1] This is not quite right: The correct way to say it is that for every index i there exists an initial position for the particle that makes $\alpha_i \neq 0$.

this, it seems plausible that after enough steps the vertex where the particle sits will indeed be uniform (and this is in fact the case), but also that a random vertex is reachable from s rather quickly (because in a tree, paths starting from s "expand out" very quickly).

In the end there is nothing special about s, and what we want in some way is that if we choose any vertex as the root, from the perspective of that vertex the graph looks a lot like a tree.

To be a bit more quantitative, if we start at s, even in the ideal case of a tree, we need just $\Omega(\log n)$ steps out of s to "cover" all the possible vertices in G. So we cannot hope to end up at a random vertex of G before we have completed at least $\Omega(\log n)$ steps. Can we do so in $O(\log n)$ steps no matter at which vertex s we started?

Recall that after t steps of the walk, our distance to the uniform distribution is upper bounded by the value λ^t, where $\lambda = \max(\lambda_2, -\lambda_n)$. This suggests that we want to design a graph whose value λ is as small as possible.

This discussion indicates that λ can in fact never get too small. To get a lower bound on λ, notice that after t steps of the walk, the potential number of vertices that could have been reached from s never exceeds d^{t+1}; there are at most this many vertices at distance $\leq t$ from s. So even when $t = \log_d n - 2$, less than half of the vertices of the random walk are reachable. Therefore the distribution $\mathbf{p}^t$ must assign probability zero to the other half vertices, and

$$\lambda^t \geq \|\mathbf{p}^t - \mathbf{u}\| \geq \left(n/2 \cdot (0 - 1/n)^2\right)^{1/2} = 1/\sqrt{2n}.$$

It follows that $\lambda = \Omega(1/\sqrt{d})$. A more precise analysis shows that for every graph, $\lambda \geq 2\sqrt{d-1}/d - o_n(1)$, where $o_n(1)$ is quantity that converges to zero as n gets large. However, there exist graphs such that $\lambda = 2\sqrt{d-1}/d$ for infinitely many values of n. Such graphs are called *Ramanujan graphs.*[2]

For our purposes, it will be enough to consider graph families for which as n grows, λ stays bounded away from one. If this is the case, then after only $t = \Theta(\log n)$ steps of the random walk, we have that

$$\|\mathbf{p}^t - \mathbf{u}\| \leq \lambda^{\Theta(\log n)} = n^{-\Theta(1)} \tag{3}$$

so $\mathbf{p}^t$ gets very close to the uniform distribution, and in fact *all vertices* of G are reached with probability $\Theta(1/n)$.

Definition 3. A family of graphs $\{G_n\}$, where G_n has n vertices and is d-regular, is called an *expander family* if there is a constant $\epsilon > 0$ such that $\lambda(G_n) \leq 1 - \epsilon$ for every sufficiently large n.

3 Edge expansion

Suppose you start at a random vertex of some set S that is not too large and you take a random edge out of this vertex. How likely are you to get out of S? This is the *edge expansion* of G:

$$h(G) = \min_{S:\, |S| \leq n/2} \Pr_{(u,w)}[w \notin S \mid u \in S]$$

The probability is taken over a random pair of vertices (u, w) that is connected by an edge in G. Clearly it is necessary to put some bound on the size of S to make this quantity meaningful; why we require that $|S| \leq n/2$ is not terribly important for today's lecture.

[2]Ramanujan graphs are known to exist for every n such that $n + 1$ is a power of a prime larger than two.

If a random walk out of any vertex s approaches the uniform distribution quickly, we would expect such a walk to avoid "getting stuck" in any set S. The following claim makes this intuition precise: If λ_2 is bounded away from one, then $h(G)$ cannot be too small:

Theorem 4. $h(G) \geq (1 - \lambda_2)/2$.

To prove this theorem it is useful to describe the eigenvalues of A, the normalized adjacency matrix of G, in an alternative way. To see how this can be done, we diagonalize the matrix A as $S^{\mathrm{T}}\Lambda S$, where S is an orthonormal matrix whose rows are the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and Λ is a diagonal matrix consisting of the entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ in that order. (Recall that when S is orthonormal, $S^{\mathrm{T}} = S^{-1}$.) Then the eigenvectors of A and those of Λ are related by an orthonormal change of basis.

Let's now look at the eigenvalues of Λ, which are its diagonal entries. One way to describe the first eigenvalue λ_1 is to look at all possible vectors $\mathbf{v}$ of norm one, and take the one that maximizes the expression $\mathbf{v}\Lambda\mathbf{v}^{\mathrm{T}}$. Clearly this quantity is maximized by the vector $\mathbf{v} = \mathbf{e}_1 = (1, 0, \ldots, 0)$, which yields the value λ_1. But now notice that

$$\lambda_1 = \max_{\|\mathbf{v}\|=1} \mathbf{v}\Lambda\mathbf{v}^{\mathrm{T}} = \max_{\|\mathbf{v}\|=1} \mathbf{v}(SAS^{\mathrm{T}})\mathbf{v}^{\mathrm{T}} = \max_{\|\mathbf{v}\|=1}(\mathbf{v}S)A(\mathbf{v}S)^{\mathrm{T}} = \max_{\|\mathbf{v}\|=1} \mathbf{v}A\mathbf{v}^{\mathrm{T}}$$

because as $\mathbf{v}$ cycles over all vectors of norm one, so does $\mathbf{v}S$. Now notice that

$$\mathbf{v}A\mathbf{v}^{\mathrm{T}} = \sum_{u,w=1}^{n} A_{uw}\mathbf{v}(u)\mathbf{v}(w) = \sum_{(u,\,w) \text{ is an edge}} \frac{1}{d}\mathbf{v}(u)\mathbf{v}(w) = n\,\mathrm{E}_{(u,w)}[\mathbf{v}(u)\mathbf{v}(w)]$$

where $\mathrm{E}_{(u,w)}[\cdot]$ denotes expectation taken over a random *directed* edge (u, w) in G. This gives the following formula for λ_1:

$$\lambda_1 = n \cdot \max_{\|\mathbf{v}\|=1} \mathrm{E}_{(u,w)}[\mathbf{v}(u)\mathbf{v}(w)].$$

What about λ_2? Again, we look at Λ, but now instead of maximizing over all vectors, we only maximize over those that are orthogonal to the first vector $\mathbf{e}_1$:

$$\lambda_2 = \max_{\|\mathbf{v}\|=1,\mathbf{v}\perp\mathbf{e}_1} \mathbf{v}\Lambda\mathbf{v}^{\mathrm{T}} = \max_{\|\mathbf{v}\|=1,\mathbf{v}\perp\mathbf{e}_1}(\mathbf{v}S)A(\mathbf{v}S)^{\mathrm{T}}.$$

Now notice that as $\mathbf{v}$ cycles over all vectors of norm 1 that are perpendicular to $\mathbf{e}_1$, $\mathbf{v}S$ will cycle over all vectors of norm 1 that are perpendicular to $\mathbf{e}_1 S = \mathbf{v}_1$, which is parallel to $\mathbf{u}$. So we obtain the following expression for λ_2:

$$\lambda_2 = \max_{\|\mathbf{v}\|=1,\mathbf{v}\perp\mathbf{u}} \mathbf{v}A\mathbf{v}^{\mathrm{T}} = n \max_{\|\mathbf{v}\|=1,\mathbf{v}\perp\mathbf{u}} \mathrm{E}_{(u,w)}[\mathbf{v}(u)\mathbf{v}(w)].$$

Since our goal is to bound the value λ_2 away from 1, it will be convenient to look at the expression $1 - \lambda_2$. Using the above formula and simplifying a bit, we obtain

$$1 - \lambda_2 = \frac{n}{2} \min_{\|\mathbf{v}\|=1,\mathbf{v}\perp\mathbf{u}} \mathrm{E}_{(u,w)}[(\mathbf{v}(u) - \mathbf{v}(w))^2] \tag{4}$$

$$= \frac{1}{2} \min_{\mathbf{v}\perp\mathbf{u}} \frac{\mathrm{E}_{(u,w)}[(\mathbf{v}(u) - \mathbf{v}(w))^2]}{\mathrm{E}_u[\mathbf{v}(u)^2]}. \tag{5}$$

Proof of Theorem 4. Let S be any set of vertices of size at most $n/2$ and set

$$\mathbf{v}(u) = \begin{cases} |\overline{S}|/n, & \text{if } u \in S \\ -|S|/n, & \text{if } u \notin S. \end{cases}$$

Notice that $\mathbf{v} \perp \mathbf{u}$, and that $(\mathbf{v}(u) - \mathbf{v}(w))^2$ is 1 when exactly one of u and w is in S and the other is in $\overline{S}$, and 0 otherwise. In the first case we will say (u, w) *crosses* $(S, \overline{S})$. Plugging into (5) we obtain

$$1 - \lambda_2 \leq \frac{1}{2} \frac{\Pr_{(u,w)}[(u,w) \text{ crosses } (S,\overline{S})]}{\mathrm{E}_u[\mathbf{v}(u)^2]}.$$

Now notice that

$$\Pr_{(u,w)}[(u,w) \text{ crosses } (S,\overline{S})] = 2\Pr[w \notin S \text{ and } u \in S] = 2\Pr[w \notin S \mid u \in S] \cdot \frac{|S|}{n}$$

and

$$\mathrm{E}_u[\mathbf{v}(u)]^2 = \frac{1}{n}\left[|S| \cdot \left(\tfrac{|\overline{S}|}{n}\right)^2 + |\overline{S}| \cdot \left(\tfrac{|S|}{n}\right)^2\right] = \frac{|S||\overline{S}|}{n^2} \geq \frac{|S|}{2n}. \qquad \square$$

4 Proof of Theorem 2

We now show how to deduce Theorem 2 from Theorem 1. Let Φ be a 2CSP with no restrictions on the number of occurrences of each variable. We show how to get a new instance Φ' out of Φ where every variable occurs at most d times.

Each variable x_i in Φ gives rise to n_i variables $x'_{i1}, \ldots, x'_{in_i}$ in Φ'. For each constraint $\phi_{ii'}(x_i, x_{i'})$ in Φ we assign unique copies $x'_{ij}, x'_{i'j'}$ in Φ' and add $d/2$ copies of the constraint $\phi_{ii'}(x'_{ij}, x'_{i'j'})$ in Φ'. Finally, for every i we fix a $d/2$-regular graph G_i on n_i vertices with edge expansion $h(G_i) \geq 1/4$ and introduce equality constraints $x'_{ij} = x'_{ij'}$ for every edge (i, i') of G_i. We will call these the equality constraints for i. We will discuss how to construct such an expander in the next lecture.

If Φ has $m/2$ constraints, then Φ' will have m variables and dm constraints. If Φ is satisfiable, then Φ' is clearly satisfiable. Now suppose we could find an assignment x' that satisfies a $1 - \varepsilon$ fraction of the constraints of Φ'. Then the following claim allows us to convert x' into an assignment that satisfies a $1 - 18\varepsilon$ fraction of the constraints of Φ:

Claim 5. *If some assignment x' the violates at most an ε-fraction of contraints in Φ', then the assignment x where*

$$x_i = \textit{plurality (most frequent) value among } x'_{i1}, \ldots, x'_{in_i}$$

violates at most a 34ε fraction of constraints in Φ.

By the definition of edge expansion, within every graph G_i

$$|E(S, \overline{S})| \geq \frac{d|S|}{8}$$

for every subset S of vertices in G_i of size at most $n_i/2$, where $E(S, \overline{S})$ is the number of edges from a vertex in S to a vertex outside S.

Let S_i be the set of variables x'_{ij} that agree with the plurality value x_i. We will argue that because of expansion S_i must contain most of the variables x'_{ij} unless many of the equality constraints for i are violated. To make this quantitative, let ε_i be the fraction of the $dn_i/4$ equality constraints for i violated by the assignment x'. We split the analysis into three cases:

- If $|S_i| > n_i/2$, then $|\overline{S}_i| < n_i/2$ and $|E(S_i, \overline{S}_i)| \geq d|\overline{S}_i|/8$. Since all the equality constraints for i between S_i and $\overline{S}_i$ are violated by x', $|E(S_i, \overline{S}_i)| \leq \varepsilon_i(dn_i/4)$, so $|\overline{S_i}| \leq 2\varepsilon_i n_i$.
- If $n_i/4 \leq |S_i| \leq n_i/2$, then $|E(S_i, \overline{S}_i)| \geq d|S_i|/8 \geq dn_i/32$. Since all the equality constraints for i between S_i and $\overline{S}_i$ are violated by x', it follows that $\varepsilon_i \geq 1/8$, so $|\overline{S_i}| \leq n_i \leq 8\varepsilon_i n_i$.
- If $|S_i| < n_i/4$, then no value in Σ is taken by more than a 1/4-fraction of the x'_{ij}s, so there must exist some subset of values $\Sigma' \subseteq \Sigma$ so the number of x'_{ij} taking values in Σ' is between $n_i/4$ and $n_i/2$. Just like in the previous case, we get $|\overline{S_i}| \leq n_i \leq 8\varepsilon_i n_i$.

We see that no matter what, $|\overline{S_i}| \leq 8\varepsilon_i n_i$ for every i.

Now consider what happens in Φ' when we replace the assignment x' with the plurality assignment $x'_{\text{plur } ij} = x_i$ for every j. Replacing x' by x'_{plur} may cause the violation of at most $(d/2)|\overline{S_i}|$ non-equality constraints for every i. If x' violates εdm constraints, x'_{plur} will then violate at most

$$\varepsilon dm + \sum_{i=1}^{n}(d/2)|\overline{S_i}| \leq \varepsilon dm + \sum_{i=1}^{n}(d/2)(8\varepsilon_i n_i) = \varepsilon dm + 16\sum_{i=1}^{n}\varepsilon_i dn_i/4 \leq 17\varepsilon dm$$

constraints of Φ'. This is a 17ε-fraction of all the constraints in Φ'. Since exactly half the constraints in Φ' are equality constraints, x cannot violate more than a 34ε fraction of constraints in Φ.