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 $x'_{i1}, \dots, x'_{in_i}$ and impose some additional constraints that force all of $x'_{i1}, \dots, x'_{in_i}$ to take the same value.

The first thing we may try is to add the constraints $x'_{i1} = x'_{i2}, x'_{i2} = x'_{i3}, \dots, x'_{i(n_i-1)} = x'_{in_i}$ 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'. Suppose that we could then find an assignment x' that satisfies a $1-\varepsilon'$ fraction of its constraints of Φ'. Can we use x' to obtain 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 Φ will be violated.

So we need to make the equality constraints more robust: If there is no clear majority among the values $x'_{i1}, \dots, x'_{in_i}$, 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 $\{1, \dots, t\}$. An assignment to $x'_{i1}, \dots, x'_{in_i}$ 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 undirected, connected, and d-regular.

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 and independent of s?

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. Finally, the largest λ_i in absolute value cannot be -1, because $\mathbf{p}^t$ would then eventually be shifting signs; since it is a vector of probabilities, its entries must always be nonnegative.

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 $\mathbf{u}A = \lambda_1 \mathbf{u}$, 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:

$$\|\mathbf{p}^t - \mathbf{u}\| \le \lambda^t. \tag{3}$$

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

To get some intuition about equation (3), notice that in t steps the particle can reach at most $1 + (d-1) + \cdots + (d-1)^t \le (d-1)^{t+1}$ vertices of the graph. This value is attained when the t-neighborhood of s is a d-regular tree. Let t be the largest value for which $(d-1)^{t+1}$ is at most $n/2$. Then at least half the entries of $\mathbf{p}^t$ are zero and

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

from where $\lambda \ge (1/\sqrt{d-1}) \cdot (2(d-1)^2)^{-1/2t}$. As n gets larger, the second term approaches 1 and λ must be at least as large as $1/\sqrt{d-1}$.

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

[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 $\alpha_i \neq 0$.

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

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

$$\|\mathbf{p}^t - \mathbf{u}\| \leq \lambda^{\Theta(\log n)} = n^{-\Theta(1)} \tag{4}$$

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

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? 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. The probability is taken over a pair of endpoints (u, w) of a random *directed* edge of G.

Theorem 4. *For every set S of vertices,*

$$\Pr_{(u,w)}[u \in S \text{ and } w \notin S] \geq (1 - \lambda_2) \Pr_u[u \in S] \Pr_w[w \notin S].$$

To prove this theorem it is useful to describe the eigenvalues of A, the normalized adjacency matrix of G, in an alternative way. We look at the value of the expression $\mathbf{v}A\mathbf{v}^{\mathrm{T}}$ as $\mathbf{v}$ ranges over all vectors of norm 1. We expand $\mathbf{v}$ in the basis of eigenvectors

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

where $\alpha_1^2 + \cdots + \alpha_n^2 = 1$. Then

$$\mathbf{v}A\mathbf{v}^{\mathrm{T}} = \Big(\sum_{i=1}^n \alpha_i \mathbf{v}_i\Big) A \Big(\sum_{j=1}^n \alpha_j \mathbf{v}_j\Big) = \sum_{i,j=1}^n \alpha_i \alpha_j \cdot \mathbf{v}_i A \mathbf{v}_j^{\mathrm{T}}.$$

Since $\mathbf{v}_i A \mathbf{v}_j^{\mathrm{T}} = \lambda_i \mathbf{v}_i \mathbf{v}_j^{\mathrm{T}} = \lambda_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ takes value λ_i when $i = j$ and zero otherwise, we get

$$\mathbf{v}A\mathbf{v}^{\mathrm{T}} = \alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_2 + \cdots + \alpha_n^2 \lambda_n.$$

It follows that $\mathbf{v}A\mathbf{v}^{\mathrm{T}}$ can be at most λ_1 and this value is attained when $\alpha_1 = 1$, namely when $\mathbf{v} = \mathbf{v}_1$. So we can describe λ_1 as

$$\lambda_1 = \max_{\|\mathbf{v}\|=1} \mathbf{v}A\mathbf{v}^{\mathrm{T}}.$$

Similarly, we can describe λ_2 as the maximum of $\mathbf{v}A\mathbf{v}^{\mathrm{T}}$ but taken only over those $\mathbf{v}$ for which $\alpha_1 = 0$, namely those $\mathbf{v}$ that are perpendicular to $\mathbf{v}_1$. In our case $\mathbf{v}_1$ is parallel to $\mathbf{u}$ so we can write

$$\lambda_2 = \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{u}} \mathbf{v}A\mathbf{v}^{\mathrm{T}}.$$

We will now give a probabilistic interpretation to the quantity $1 - \lambda_2$. Fix $\mathbf{v}$ such that $\|\mathbf{v}\| = 1$ and notice that

$$\sum_{u,w=1}^n A_{uw}(\mathbf{v}(u) - \mathbf{v}(w))^2 = \sum_{u,w=1}^n A_{uw}\mathbf{v}(u)^2 + \sum_{u,w=1}^n A_{uw}\mathbf{v}(w)^2 - 2\sum_{u,w=1}^n A_{uw}\mathbf{v}(u)\mathbf{v}(w).$$

Since each row and each column of A adds up to one, each of the first two sums equals the sum of squares of the entries of $\mathbf{v}$, which is 1. The third sum equals $\mathbf{v}A\mathbf{v}^{\mathrm{T}}$. Therefore we can write

$$1-\lambda_2 = \frac{1}{2}\min_{\|\mathbf{v}\|=1,\mathbf{v}\perp\mathbf{u}} \sum_{u,v=1}^{n} A_{uw}(\mathbf{v}(u)-\mathbf{v}(w))^2 = \frac{1}{2}\min_{\mathbf{v}\perp\mathbf{u}} \frac{\sum_{u,v=1}^{n} A_{uw}(\mathbf{v}(u)-\mathbf{v}(w))^2}{\sum_{u=1}^{n}\mathbf{v}(u)^2}.$$

Since there are dn directed edges of G, we get that $\sum A_{uw}(\mathbf{v}(u)-\mathbf{v}(w))^2 = n\,\mathrm{E}_{(u,w)}[(\mathbf{v}(u)-\mathbf{v}(w))^2]$. We also have $\sum \mathbf{v}(u)^2 = n\,\mathrm{E}_u[\mathbf{v}(u)^2]$. Therefore

$$1-\lambda_2 = \frac{1}{2}\min_{\mathbf{v}\perp\mathbf{u}} \frac{\mathrm{E}_{(u,w)}[(\mathbf{v}(u)-\mathbf{v}(w))^2]}{\mathrm{E}_u[\mathbf{v}(u)^2]}. \tag{5}$$

Proof of Theorem 4. Let S be any set of vertices, $\alpha = \Pr[u \in S] = |S|/n$ and set

$$\mathbf{v}(u) = \begin{cases} 1-\alpha, & \text{if } u \in S \\ -\alpha, & \text{if } u \notin S. \end{cases}$$

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

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

where

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

and

$$\mathrm{E}_u[\mathbf{v}(u)]^2 = \alpha(1-\alpha)^2 + (1-\alpha)\alpha^2 = \alpha(1-\alpha) = \Pr[u \in S]\Pr[w \notin S]. \qquad \square$$

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 $\lambda(G_i) \ge 1/2$ 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 talk about how to construct such an expander in the next two lectures.

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 Theorem 4, within every graph G_i

$$|E(S, \overline{S})| \geq \frac{d|S||\overline{S}|}{4n_i}$$

for every subset S of vertices in G_i, 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. Let ε_i be the fraction of the $dn_i/4$ equality constraints for i violated by the assignment x'. We will argue that $|\overline{S_i}| \leq 8\varepsilon_i n_i$:

- If $|S_i| > n_i/2$, then $|E(S_i, \overline{S}_i)| \geq d|\overline{S}_i|/8$. Since all the equality constraints for i between S_i and $\overline{S}_i$ are violated by x', $\varepsilon_i(dn_i/4) \geq |E(S_i, \overline{S}_i)|$, 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 $|\overline{S_i}| \leq n_i \leq 8\varepsilon_i n_i$.

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.