Suppose you have an undirected graph G on n vertices and a particle sits at vertex s. The particle takes a *random walk* on the graph: At each step, it picks one of its neighbors at random and moves to that vertex. How long will it take for the particle to reach a random vertex in the graph?

The answer to this question depends a lot on the shape of G, and we will see what kinds of graphs are "the best" in this respect. These graphs are called expanders.

We will assume that the graph G is *d-regular*, that is every vertex has the same degree d. We will think of d as being much smaller than the number of vertices.

1 Some intuition

Let's think of finding a good G as an engineering problem: We want to design a graph G so that starting from any vertex s, we can reach a random vertex as soon as possible. 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, very quickly (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 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).

However, 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?

Here is another way to look at the problem. Let S be an arbitrary set containing about half the vertices in the graph and also containing s. If we aim at reaching a random vertex quickly, then the particle better have a good chance of "escaping" the set S fairly soon.

2 Random walks

Throughout this lecture, we will use G to denote the graph in question, n for the number of vertices of G, and we will assume that G is d-regular, with $d \geq 3$. In addition, we will assume that there is a loop at every vertex of G, that is (u, u) is an edge of G for every vertex u.

A *random walk* on G is 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 ending up at vertex u after i 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

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

If we let t grow very large, what does the distribution $\mathbf{p}^t$ look like? That is, after we have walked for a long time, where do we expect to land? It seems reasonable that we should be at a random vertex of the graph — that is, be at every vertex v with probability $1/n$, and this is indeed the case. We sketch an informal argument. First, it is easy to see that if at any point in time we reach a uniformly random vertex, then after one step of the walk we are again at a uniformly random vertex; that is, if $\mathbf{p}^t$ equals the uniform distribution $\mathbf{u}$, then so does $\mathbf{p}^{t+1}$.

But is it not possible that starting from some vertex s, after many steps we reach not a uniform vertex but a vertex following some other distribution $\mathbf{u}'$? Consider the following two scenarios: In scenario 1, our particle (let's call it particle 1) starts at s, and in scenario 2 our particle (let's call it particle 2) starts at a uniformly random vertex. Now we claim that if we observe the position of the particle at some very large time t, the two scenarios look exactly alike! This is because at some point in time before t, particle 1 and particle 2 must surely have landed at the same vertex — as t tends to infinity, the probability of this event tends to one[1] — after which they become indistinguishable. Since the position of particle 2 is uniformly random at every step, the position of particle 1 must also eventually become uniformly random.

There is a way to make this argument precise, but instead of pursuing this let's immediately turn to the question of interest: How soon does it take until the random walk starting from s becomes close to uniform? To answer this we must first introduce a "uniformity measure" for probability distributions. We already saw such a measure, namely statistical distance, but here we will look at another, more convenient quantity: 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 $\|\mathbf{p} - \mathbf{p}'\| \leq \varepsilon$.

[1] This is the Borel-Cantelli lemma in probability: Informally, if an event *can* ever happen, it *will* eventually happen.

3 Adjacency matrix and eigenvalues

We now introduce some algebraic tools that happen to be extremely useful in understanding this problem.

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

$$A_{u,v} = \begin{cases} 1/d, & \text{if } (u,v) \text{ is an edge of } G \\ 0, & \text{otherwise.} \end{cases}$$

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 $\mathbf{p}^{t+1} = A\mathbf{p}^t$ (we think of $\mathbf{p}^t$ as column vectors) and so we immediately obtain that $\mathbf{p}^t = A^t\mathbf{p}_0$.

It turns out that 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 $A\mathbf{v} = \lambda\mathbf{v}$. 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

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

and after t steps

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

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.[2] 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. Also, we cannot have $\lambda_n = -1$ because then $\mathbf{p}^t$ would oscillate when t is large, and we know it converges to the uniform distribution $\mathbf{u}$.

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

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

[2]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 $\alpha_i \neq 0$.

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 $A\mathbf{u} = \lambda_1 \mathbf{u}$, so the eigenvector $\mathbf{v}_1$ associated to $\lambda_1 = 1$ equals $\sqrt{n} \cdot \mathbf{u}$. Now from equation 2 we have that

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

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: $\|\mathbf{p}^t - \mathbf{u}\| \leq \lambda^t$. 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.

4 Expanders

So how small can λ get? Let us first see that no matter what the graph looks like, λ can never get too small. For this, let's picture a random walk on an arbitrary graph starting at an arbitrary vertex s. 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 $\lambda = \Omega(1/\sqrt{d})$. Thus when the degree is constant and the number of vertices grows, λ is bounded away from zero by some constant. Graphs that attain the smallest possible value of λ are called *Ramanujan graphs*.

For our purposes, it will be enough to consider graphs 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

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

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

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 $\varepsilon > 0$ such that $\lambda(G_n) \leq 1 - \varepsilon$. There are several examples of expander families, some of which are very simple to describe, but proving that a family of graphs is expanding is fairly difficult and we won't do so here.

For example, the following family of 4-regular graphs is expanding: For every prime number p, think of the vertices of G_p as elements of the field $\mathbb{F}_p$. For every $x \in \mathbb{F}_p$, the edges going out of x are $(x, x-1)$, (x, x), $(x, x+1)$, (x, x^{-1}) where all operations are over $\mathbb{F}_p$.

5 The expander mixing lemma

Now let's revisit one of the initial questions: Given a set of vertices S, how long will it take for the particle to escape S in the random walk? There are several ways to formalize this question, and the following one turns out to be particularly nice. Suppose that the particle starts at a uniformly random vertex in S. Then what is the probability it ends up outside S after t steps of the walk?

Let's start with $t = 1$. Then we ask: Choose a random vertex in S, and a random neighbor. What is the probability that the neighbor is in $\overline{S}$? Up to normalization, this is the same as asking the following: What is the probability that a random edge of G crosses the cut $(S, \overline{S})$?

Using vectors and matrices, this probability can be expressed conveniently as follows. Let $\mathbf{s}$ denote the indicator vector of S, namely the vector that assigns value 1 to every vertex in S and 0 to the other vertices. Similarly we define $\overline{\mathbf{s}}$ for $\overline{S}$. Then the expression $d \cdot \langle A\mathbf{s}, \overline{\mathbf{s}} \rangle$ counts the number of edges in the cut $(S, \overline{S})$, so a probability that a random edge crosses this cut from S to $\overline{S}$ is exactly $\langle A\mathbf{s}, \overline{\mathbf{s}} \rangle / n$.

To bound this expression, we reason as follows. The fact that random walks on expanders converge quickly to the uniform distribution says that, in some sense, an expander is a good approximation of a complete graph: In a complete graph (with a loop around every vertex), the uniform distribution is reached after one step of the random walk, no matter where we start from. In general, an expander cannot do this because it has bounded degree, but in certain situations we can think of an expander graph as a complete graph "plus" some error.

This viewpoint works well for bounding the expression $\langle A\mathbf{s}, \overline{\mathbf{s}} \rangle$. We want to think of the adjacency matrix A of G as representing the adjacency martix J of the complete graph (all the entries of J have value $1/n$) plus some "error matrix" C. We write $A = J + C$, and

$$\langle A\mathbf{s}, \overline{\mathbf{s}} \rangle = \langle J\mathbf{s}, \overline{\mathbf{s}} \rangle + \langle C\mathbf{s}, \overline{\mathbf{s}} \rangle$$

so that

$$|\langle A\mathbf{s}, \overline{\mathbf{s}} \rangle - \langle J\mathbf{s}, \overline{\mathbf{s}} \rangle| = |\langle C\mathbf{s}, \overline{\mathbf{s}} \rangle| \leq \|C\mathbf{s}\| \cdot \|\overline{\mathbf{s}}\|. \tag{3}$$

Two of the entries in this expression can be calculated directly: $\langle J\mathbf{s}, \overline{\mathbf{s}} \rangle = |S| \cdot |\overline{S}|/n$ and $\|\overline{\mathbf{s}}\|^2 = |\overline{S}|$. It remains to bound the term $\|C\mathbf{s}\|$.

Claim 1. *For every vector* $\mathbf{v}$, $\|C\mathbf{v}\| \leq \lambda \|\mathbf{v}\|$.

Proof. We can assume that the entries of $\mathbf{v}$ add up to one because both sides can be scaled accordingly. We write $\mathbf{v}$ in the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ as

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

Then $\alpha_1 \mathbf{v}_1 = \langle \mathbf{v}, \mathbf{v}_1 \rangle \mathbf{v}_1 = \mathbf{u}$ — the uniform distribution, and

$$C\mathbf{u} = A\mathbf{u} - J\mathbf{u} = \mathbf{u} - \mathbf{u} = 0.$$

Also,

$$C(\mathbf{v} - \mathbf{u}) = A(\mathbf{v} - \mathbf{u}) - J(\mathbf{v} - \mathbf{u}).$$

The second term vanishes because

$$J(\mathbf{v}-\mathbf{u}) = J\mathbf{v} - J\mathbf{u} = \mathbf{u} - \mathbf{u} = 0.$$

So, we have that

$$C\mathbf{v} = C\mathbf{u} + C(\mathbf{v}-\mathbf{u}) = A(\mathbf{v}-\alpha_1\mathbf{v}_1) = \lambda_2\alpha_2\mathbf{v}_2 + \cdots + \lambda_n\alpha_n\mathbf{v}_n$$

and therefore

$$\|C\mathbf{v}\|^2 = \lambda_2^2\alpha_2^2 + \cdots + \lambda_n^2\alpha_n^2 \leq \lambda^2\|\mathbf{v}\|^2. \qquad \square$$

Applying the claim to equation 3, we obtain the following important theorem.

Theorem 2 (Expander mixing lemma). *For every graph G and every cut $(S, \overline{S})$,*

$$\left|\Pr_{\text{edge } (u,v)}[u \in S \text{ and } v \in \overline{S}] - \tfrac{|S|}{n}\tfrac{|\overline{S}|}{n}\right| \leq \lambda(G)\sqrt{\tfrac{|S|}{n}\tfrac{|\overline{S}|}{n}}.$$

When $\lambda(G)$ is small, this expression says that the probability that a random edge crosses the cut $(S, \overline{S})$ can be approximated with the probability that two independently chosen vertices u and v fall on opposing sides of the cut. The error term will be relatively small as long as the sets S and $\overline{S}$ are not too small.

The expander mixing lemma answers the question of how likely the particle is to cross a cut in one step of the random walk. How about t steps? For this, let's look at the graph G^t, which has the same vertices as v and has an edge for every path of length t in G. This may result in multiple edges, so in general G^t will be a multigraph of degree d^t; but everything we have done in this lecture applies to multigraphs as well. In particular, $\lambda(G^t) = \lambda(G)^t$, so as t becomes larger, $\lambda(G^t)$ becomes very small. Thus if we start with an expander, after $t = O(\log 1/\delta)$ steps $\lambda(G^t) \leq \delta$, and then the expander mixing lemma tells us that the probability that a random t-step path in G crosses the cut $S, \overline{S}$ is at least

$$\Pr_{\text{edge } (u,v) \text{ in } G^t}[u \in S \text{ and } v \in \overline{S}] \geq \tfrac{|S|}{n}\tfrac{|\overline{S}|}{n} - \delta\sqrt{\tfrac{|S|}{n}\tfrac{|\overline{S}|}{n}}.$$

If $|S| = |\overline{S}| = n/2$, this implies

$$\Pr[v \in \overline{S} \mid u \in S] \geq 1/2 - \delta$$

which is about as large as it can be — even in a complete graph the probability of crossing from S to $\overline{S}$ is $1/2$.

What happens when the set S is small? Then the expander mixing lemma does not give a useful bound on the probability of crossing the cut $(S, \overline{S})$. However using a different proof one can show the following.

Theorem 3. *For every graph G and every cut $(S, \overline{S})$ where $|S| \leq n/2$,*

$$\Pr_{\text{edge } (u,v)}[v \in \overline{S} \mid u \in S] \geq \frac{1-\lambda}{2}.$$