Lecture 14

Jonathan Katz

1 Randomized Space Complexity

1.1 Undirected Connectivity and Random Walks

A classic problem in $\mathcal{RL}$ is *undirected connectivity* (UConn). Here, we are given an *undirected* graph and two vertices s, t and are asked to determine whether there is a path from s to t. An $\mathcal{RL}$ algorithm for this problem is simply to take a "random walk" (of sufficient length) in the graph, starting from s. If vertex t is ever reached, then output 1; otherwise, output 0. (We remark that this approach does *not* work for *directed* graphs.) We analyze this algorithm (and, specifically, the length of the random walk needed) in two ways; each illustrates a method that is independently useful in other contexts. The first method looks at random walks on *regular* graphs, and proves a stronger result showing that after sufficiently many steps of a random walk the location is close to uniform over the vertices of the graph. The second method is more general, in that it applies to any (non-bipartite) graph; it also gives a tighter bound.

1.1.1 Random Walks on Regular Graphs

Fix an undirected graph G on n vertices where we allow self-loops and parallel edges (i.e., integer weights on the edges). We will assume the graph is d-regular and has at least one self-loop at every vertex; any graph can be changed to satisfy these conditions (without changing its connectivity) by adding sufficiently many self-loops. Let G also denote the (scaled) adjacency matrix corresponding to this graph: the (i, j)th entry is k/d if there are k edges between vertices i and j. Note that G is *symmetric* ($G_{i,j} = G_{j,i}$ for all i, j) and *doubly stochastic* (all entries are non-negative, and all rows and columns sum to 1). A *probability vector* $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n$ is a vector each of whose entries is non-negative and such that $\sum_i p_i = 1$. If we begin by choosing a vertex v of G with probability determined by $\mathbf{p}$, and then take a "random step" by choosing (uniformly) an edge of v and moving to the vertex v' adjacent to that edge, the resulting distribution on v' is given by $\mathbf{p}' = G \cdot \mathbf{p}$. Inductively, the distribution after t steps is given by $G^t \cdot \mathbf{p}$. Note that if we set $\mathbf{p} = \mathbf{e}_i$ (i.e., the vector with a 1 in the ith position and 0s everywhere else), then $G^t \cdot \mathbf{p}$ gives the distribution on the location of a t-step random walk starting at vertex i.

An *eigenvector* of a matrix G is a vector $\mathbf{v}$ such that $G \cdot \mathbf{v} = \lambda \mathbf{v}$ for some $\lambda \in \mathbb{R}$; in this case we call λ the associated *eigenvalue*. Since G is a symmetric matrix, standard results from linear algebra show that there is an orthonormal basis of eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with (real) eigenvalues $\lambda_1, \ldots, \lambda_n$, sorted so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. If we let $\mathbf{1}$ denote the vector with $1/n$ in each entry — i.e., it represents the uniform distribution over the vertices of G — then $G \cdot \mathbf{1} = \mathbf{1}$ and so G has eigenvalue 1. Moreover, since G is a (doubly) stochastic matrix, it has no eigenvalues of absolute value greater than 1. Indeed, let $\mathbf{v} = (v_1, \ldots, v_n)$ be an eigenvector of G with eigenvalue λ, and let j be such that $|v_j|$ is maximized. Then $\lambda \mathbf{v} = G \cdot \mathbf{v}$ and so

$$|\lambda v_j| = \left|\sum_{i=1}^n G_{j,i} \cdot v_i\right|$$

$$\leq \ |v_j| \cdot \sum_{i=1}^{n} |G_{j,i}| \ = \ |v_j| \, ;$$

we conclude that $|\lambda| \leq 1$. If G is connected, then it has no other eigenvector with eigenvalue 1. Since G is non-bipartite (because of the self-loops), -1 is not an eigenvalue either.

To summarize, if G is connected and not bipartite then it has (real) eigenvectors $\lambda_1, \ldots, \lambda_n$ with $1 = \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$. The (absolute value of the) second eigenvalue λ_2 determines how long a random walk in G we need so that the distribution of the final location is close to uniform:

Theorem 1 *Let G be a d-regular, undirected graph on n vertices with second eigenvalue λ_2, and let $\mathbf{p}$ correspond to an arbitrary probability distribution over the vertices of G. Then for any $t > 0$*

$$\left\| G^t \cdot \mathbf{p} - \mathbf{1} \right\|_2 \leq |\lambda_2|^t.$$

Proof Write $\mathbf{p} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$, where the $\{\mathbf{v}_i\}$ are the eigenvectors of G (sorted according to decreasing absolute value of their eigenvalues); recall $\mathbf{v}_1 = \mathbf{1}$. We have $\alpha_1 = 1$; this follows since $\alpha_1 = \langle \mathbf{p}, \mathbf{1} \rangle \, / \, \|\mathbf{1}\|_2^2 = (1/n)/(1/n) = 1$. We thus have

$$G^t \cdot \mathbf{p} = G^t \cdot \mathbf{1} + \sum_{i=2}^{n} \alpha_i G^t \cdot \mathbf{v}_i = \mathbf{1} + \sum_{i=2}^{n} \alpha_i \left(\lambda_i\right)^t \mathbf{v}_i$$

and so, using the fact that the $\{\mathbf{v}_i\}$ are orthogonal,

$$\begin{aligned} \left\| G^t \cdot \mathbf{p} - \mathbf{1} \right\|_2^2 &= \sum_{i=2}^{n} \alpha_i^2 \left(\lambda_i\right)^{2t} \cdot \|\mathbf{v}_i\|_2^2 \\ &\leq \lambda_2^{2t} \cdot \sum_{i=2}^{n} \alpha_i^2 \cdot \|\mathbf{v}_i\|_2^2 \\ &\leq \lambda_2^{2t} \cdot \|\mathbf{p}\|_2^2 \ \leq \ \lambda_2^{2t} \cdot \|\mathbf{p}\|_1^2 = \lambda_2^{2t}. \end{aligned}$$

The theorem follows. ■

It remains to show a bound on $|\lambda_2|$.

Theorem 2 *Let G be a d-regular, connected, undirected graph on n vertices with at least one self-loop at each vertex and $d \leq n$. Then $|\lambda_2| \leq 1 - \frac{1}{\mathsf{poly}(n)}$.*

Proof Let $\mathbf{u} = (u_1, \ldots, u_n)$ be a unit eigenvector corresponding to λ_2, and recall that $\mathbf{u}$ is orthogonal to $\mathbf{1} = (1/n, \ldots, 1/n)$. Let $\mathbf{v} = G\mathbf{u} = \lambda_2 \mathbf{u}$. We have

$$\begin{aligned} 1 - \lambda_2^2 = \|\mathbf{u}\|_2^2 \cdot (1 - \lambda_2^2) &= \|\mathbf{u}\|_2^2 - \|\mathbf{v}\|_2^2 \\ &= \|\mathbf{u}\|_2^2 - 2\,\|\mathbf{v}\|_2^2 + \|\mathbf{v}\|_2^2 \\ &= \|\mathbf{u}\|_2^2 - 2\,\langle G\mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|_2^2 \\ &= \sum_i u_i^2 - 2 \sum_{i,j} G_{i,j}\, u_j v_i + \sum_j v_j^2 \\ &= \sum_{i,j} G_{i,j}\, u_i^2 - 2 \sum_{i,j} G_{i,j}\, u_j v_i + \sum_{i,j} G_{i,j}\, v_j^2 \\ &= \sum_{i,j} G_{i,j}\, (u_i - v_j)^2, \end{aligned}$$

using the fact that G is a symmetric, doubly stochastic matrix for the second-to-last equality. Since $\mathbf{u}$ is a unit vector orthogonal to $\mathbf{1}$, there exist i, j with $u_i > 0 > u_j$ and such that at least one of u_i or u_j has absolute value at least $1/\sqrt{n}$, meaning that $u_i - u_j \geq 1/\sqrt{n}$. Since G is connected, there is a path of length D, say, between vertices i and j. Renumbering as necessary, let $i = 1$, $j = D+1$, and let the vertices on the path be $2, \ldots, D$. Then

$$\begin{aligned} \frac{1}{\sqrt{n}} \leq u_1 - u_{D+1} &= (u_1 - v_1) + (v_1 - u_2) + (u_2 - v_2) + (v_2 - u_3) + \cdots + (v_D - u_{D+1}) \\ &\leq |u_1 - v_1| + \cdots + |v_D - u_{D+1}| \\ &\leq \sqrt{(u_1 - v_1)^2 + \cdots + (v_D - u_{D+1})^2} \cdot \sqrt{2D} \end{aligned}$$

(using Cauchy-Schwarz for the last inequality). But then

$$\sum_{i,j} G_{i,j}(u_i - v_j)^2 \geq \frac{1}{d} \cdot \left((u_1 - v_1)^2 + \cdots + (v_D - u_{D+1})^2\right) \geq \frac{1}{2dnD},$$

using $G_{i,i} \geq 1/d$ (since every vertex has a self-loop) and $G_{i,i+1} \geq 1/d$ (since there is an edge from vertex i to vertex $i+1$). Since $D \leq n-1$, we get $1 - \lambda_2^2 \geq 1/4dn^2$ or $|\lambda_2| \leq 1 - 1/8dn^2$, and the theorem follows. ■

We can now analyze the algorithm for undirected connectivity. Let us first specify the algorithm more precisely. Given an undirected graph G and vertices s, t, we want to determine if there is a path from s to t. We restrict our attention to the connected component of G containing s, add at least one self-loop to each vertex in G, and add sufficiently many additional self-loops to each vertex in order to ensure regularity. Then we take a random walk of length $\ell = 16dn^2 \log n \geq 2 \cdot (1 - |\lambda_2|)^{-1} \log n$ starting at vertex s, and output 1 if we are at vertex t at the end of the walk. (Of course, we do better if we output 1 if the walk ever passes through vertex t; our analysis does not take this into account.) By Theorem 1,

$$\left\| G^\ell \cdot \mathbf{e}_s - \mathbf{1} \right\|_2 \leq |\lambda_2|^\ell \leq 1/n^2.$$

If t is in the connected component of s, the probability that we are at vertex t at the end of the walk is at least $\frac{1}{n} - \frac{1}{n^2} \geq 1/2n$. We can, of course, amplify this by repeating the random walk sufficiently many times.