Random walks in undirected graphs (lecture notes)

Course: Derandomizing Space-Bounded Computation, Winter 2025, University of Chicago
Instructor: William Hoza (williamhoza@uchicago.edu)

Edited 2025-01-14

Suppose we would like to know whether two vertices, s and t, are in the same connected component of a given simple N-vertex undirected graph G. We are not concerned with time complexity, but we want to keep our space complexity at a minimum. We can use the following "random walk" approach.

1. Initialize $i \leftarrow s$.
2. For $k = 1$ to N^c:
   (a) If $i = t$, halt and output "YES."
   (b) Pick a neighbor j of i uniformly at random, and set $i \leftarrow j$.
3. Output "NO."

If c is a constant, then this algorithm only uses $O(\log N)$ bits of memory above and beyond the space taken up by the read-only input. If s and t are not in the same connected component, then clearly the algorithm rejects. What if s and t are in fact in the same connected component?

Theorem 0.1. *If s and t are in the same connected component and the constant c is chosen large enough, then the algorithm described above outputs "YES" with probability at least* 0.99*.*

We will use spectral methods to prove Theorem 0.1. For any directed N-vertex multigraph G, we can define the *transition probability matrix* $M \in [0,1]^{N \times N}$ by letting $M_{i,j}$ be the probability of going to j if we start at i and take a single random step. In other words, $M_{i,j}$ is the number of edges from i to j divided by the number of outgoing edges of i.

A probability distribution over vertices can be represented by its probability mass function, which we think of as a row vector $\pi \in [0,1]^N$. If we pick a vertex according to π and then take a single random step, we arrive at a vertex distributed according to πM.

Let u denote the uniform distribution over vertices, i.e., $u = (\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N})$. Observe that if G is regular, then u is "stationary," i.e., $uM = u$. Intuitively, one might expect that if we start with some distribution π and we take a random step, then the resulting distribution πM ought to be "closer" to u.[1] The *expansion parameter* of G quantifies how quickly we reach u, or in other words, how quickly a random walk in G "mixes."

Definition 0.2 (Expansion parameter)**.** Let G be a directed regular multigraph on N vertices with transition probability matrix $M \in [0,1]^{N \times N}$. The *expansion parameter* of G, denoted $\lambda(G)$, is defined as

$$\lambda(G) = \max_{\pi} \frac{\|\pi M - u\|_2}{\|\pi - u\|_2},$$

where π ranges over all probability distributions $\pi \in [0,1]^N$.

There are several other equivalent ways to define the expansion parameter. We will use the following characterization.

[1] This isn't always literally true. Can you think of counterexamples?

Lemma 0.3. *Let G be a directed regular multigraph on N vertices with transition probability matrix $M \in [0,1]^{N\times N}$. Then*

$$\lambda(G) = \max_v \frac{\|vM\|_2}{\|v\|_2},$$

where v ranges over all vectors that are orthogonal to u.

Proof. For any probability vector π, the vector $v := \pi - u$ is orthogonal to u, and $vM = \pi M - uM = \pi M - u$. Therefore, $\lambda(G) \le \max_v \|vM\|_2/\|v\|_2$. Conversely, if v is orthogonal to u, then there exists $\varepsilon > 0$ such that $\pi := u + \varepsilon v$ is a probability vector, and

$$\frac{\|\pi M - u\|_2}{\|\pi - u\|_2} = \frac{\|uM + \varepsilon vM - u\|_2}{\|\varepsilon v\|_2} = \frac{\|vM\|_2}{\|v\|_2},$$

so $\lambda(G) \ge \max_v \|vM\|_2/\|v\|_2$. □

We always have $0 \le \lambda(G) \le 1$. In a strongly connected directed regular multigraph with self-loops, there is a noticeable gap between $\lambda(G)$ and 1:

Theorem 0.4 (Regular digraphs have nontrivial expansion)**.** *Let $G = (V, E)$ be a D-regular strongly connected directed multigraph on N vertices in which every vertex has at least one self-loop. Then*

$$\lambda(G) \le 1 - \frac{1}{O(D \cdot N^2)}.$$

Proof. Let $v \in \mathbb{R}^N$ be any unit vector that is orthogonal to u. Let M be the transition probability matrix of G, and let $v' = vM$. The key is to analyze the quantity $\frac{1}{D}\sum_{(i,j)\in E}(v_i - v'_j)^2$. On the one hand, we have

$$\begin{aligned}
\frac{1}{D}\sum_{(i,j)\in E}(v_i - v'_j)^2 &= \frac{1}{D}\sum_{(i,j)\in E} v_i^2 - 2v_iv'_j + (v'_j)^2 \\
&= \|v\|_2^2 + \|v'\|_2^2 - \frac{2}{D}\sum_{(i,j)\in E} v_iv'_j && \text{(Regularity)} \\
&= 1 + \|v'\|_2^2 - 2\sum_{j\in[N]} v'_j \cdot \sum_{(i,j)\in E}\frac{v_i}{D} \\
&= 1 + \|v'\|_2^2 - 2\sum_{j\in[N]} v'_j \cdot v'_j \\
&= 1 - \|v'\|_2^2.
\end{aligned}$$

On the other hand, since v is a unit vector that is orthogonal to u, there exist $i_*, j_* \in [N]$ such that $|v_{i_*} - v_{j_*}| \ge 1/\sqrt{N}$. Since G is strongly connected, there is a path from i_* to j_*, say $(i_* = i_0, i_1, i_2, \ldots, i_r = j_*)$. Since every vertex has at least one self-loop, we have

$$\begin{aligned}
\frac{1}{D}\sum_{(i,j)\in E}(v_i - v'_j)^2 &\ge \frac{1}{D}\sum_{k=1}^{r}\left((v_{i_{k-1}} - v'_{i_k})^2 + (v'_{i_k} - v_{i_k})^2\right) \\
&\ge \frac{1}{2Dr}\cdot\left(\sum_{k=1}^{r}(v_{i_{k-1}} - v'_{i_k}) + (v'_{i_k} - v_{i_k})\right)^2 \\
&= \frac{1}{2Dr}\cdot(v_{i_*} - v_{j_*})^2 \\
&\ge \frac{1}{2DNr}.
\end{aligned}$$

Consequently,

$$\|v'\|_2 \le \sqrt{1 - \frac{1}{2DNr}} \le \sqrt{1 - \frac{1}{2DN^2}} \le 1 - \frac{1}{4DN^2}.$$ □

Now let us use Theorem 0.4 to prove Theorem 0.1.

Proof of Theorem 0.1. Since we only care about what happens in the connected component containing s and t, we may assume without loss of generality that G is connected. Furthermore, since adding self-loops can only increase the amount of time it takes for a random walk to reach t, we may assume without loss of generality that G is regular of degree $D \leq N$ and that every vertex has at least one self-loop. Consequently, by Theorem 0.4, we have $\lambda(G) \leq 1 - \Omega(1/N^3)$.

For any initial probability distribution π, we have

$$\|\pi - u\|_2 = \sqrt{\sum_{i=1}^{N} \left(\pi_i - \frac{1}{N}\right)^2} = \sqrt{\sum_{i=1}^{N} \pi_i^2 - \frac{2\pi_i}{N} + \frac{1}{N^2}} = \sqrt{\frac{1}{N} - \frac{2}{N} + \sum_{i=1}^{N} \pi_i^2} < \|\pi\|_2 \leq 1.$$

Therefore, if we take k random steps, the distribution we reach satisfies

$$\|\pi M^k - u\|_2 \leq \lambda(G)^k \cdot \|\pi - u\|_2 \leq \lambda(G)^k \leq \left(1 - \frac{1}{O(N^3)}\right)^k \leq \exp(-\Omega(k/N^3).$$

In particular, the probability of landing at t satisfies

$$\begin{aligned}(\pi M^k)_t = \frac{1}{N} - (u - \pi M^k)_t &\geq \frac{1}{N} - \|u - \pi M^k\|_2 \\ &\geq \frac{1}{N} - \exp(-\Omega(k/N^3)) \\ &\geq \frac{1}{2N},\end{aligned}$$

provided we choose a suitable value $k = O(N^3 \cdot \log N)$.

So, after taking k steps, there is at least a $1/(2N)$ chance that we arrive at t. Furthermore, conditioned on whatever happened during those first k steps, after we take *another* k steps, there is *again* a $1/(2N)$ chance that we arrive at t. This continues again and again, so if we take kq steps, then the chance that we never visit t is at most $(1 - \frac{1}{2N})^q \leq \exp(-q/(2N)) \leq 0.01$, provided we choose a suitable value $q = O(N)$. □