Randomized Algorithms 2021A – Lecture 1 (second part)
Random Walks on Graphs*

Robert Krauthgamer

1 Random Walks on Graphs

Let $G = (V, E)$ be an undirected graph on n vertices. Throughout, we shall assume that G is connected.

A random walk on G is the following random process that proceeds in discrete steps. Start at some initial vertex $v_0 \in V$, then at each time step, pick a random neighbor (same as random incident edge) of the current vertex and move to that vertex.

Formally, for each vertex $v \in V$ let $N(v) \subset V$ be the set of its neighbors, and let $\deg(v) = |N(v)|$ be its degree. Now define random variables $X_0, X_1, \ldots$ where $X_0 = v_0$, and for each $t \geq 0$, set X_{t+1} to each $w \in N(X_t)$ with probability $1/\deg(X_t)$.

Remark: Given X_t, we know the distribution of future steps $(X_{t+1}, X_{t+2}, \ldots)$ and it will not change if we are also given any additional information about earlier steps $(X_{t-1}, X_{t-2}, \ldots)$. This is called a Markovian process.

Potential usage: We will see how random walks can be used to design various algorithms. For example, to check if $u, v \in V$ are connected, we could start a random walk at u and see if it reaches v within a reasonable amount of time. We need to analyze the probability to reach v, but implementing the walk surely requires very little storage!

2 Hitting Time

The *hitting time* from vertex u to vertex v, denoted H_{uv}, is the expected number of steps for a random walk that starts at u until it hits v. Formally, define the random variable $T = \min\{t \geq 0 : X_t = v\}$ and let $H_{uv} = \mathbb{E}[T]$.

Notice that H_{uv} depends on G, but it is not a random variable (despite capital letter notation). Notice also that it is not symmetric, i.e., in some cases $H_{uv} \neq H_{vu}$.

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

Example: Consider an n-clique, i.e., $G = K_n$. Then $H_{uv} = n-1$ for all $u \neq v$, because T is a geometric random variables with parameter $p = 1/(n-1)$. And by definition $H_{uu} = 0$ (for every G).

Lemma 1: We have the directed triangle inequality

$$\forall u, v, w \in V, \qquad H_{uw} \leq H_{uv} + H_{vw}.$$

Proof: Was seen in class, using one random walk that starts at u.

Exer 1: Let $G = K_{n_1,n_2}$ be a complete bipartite graph with n_1 and n_2 vertices. Analyze H_{uv} for all possible $u, v \in V$.

Exer 2: Let G be a path on n vertices. Give an explicit formula for H_{uv} for all possible $u, v \in V$, and show in particular that $H_{uv} = O(n^2)$.

Hint: Denote the vertices $1, 2, \ldots, n$, and write linear equations $H_{uv} = 1 + \frac{1}{2}H_{u-1,v} + \frac{1}{2}H_{u+1,v}$ and solve these $\binom{n}{2}$ equations over $\binom{n}{2}$ variables. A simpler version is to consider h_{uv} only for $u < v$ (the other case follows by symmetry), express each $H_{uv} = H_{u,u+1} + H_{u+1,u+2} + \cdots + H_{v-1,v}$, and now the earlier equations give us $n-1$ equations using $n-1$ variables.

We will soon see that the hitting time is always (for every connected G) bounded by a polynomial in n. The next exercise shows this is not true for directed graphs.

Exer 3: Show that for every graph G and every start vertex $u \in V$,

$$\max_{v \in V} H_{uv} \geq \tfrac{1}{2}n.$$

Can you improve the leading constant $\frac{1}{2}$? Or alternatively prove that this bound is tight, by showing graphs G and $v \in V$ (for every n) for which $\max_{v \in V} H_{uv} \leq \frac{1}{2}n$? We saw that for a clique this bound is $n-1$.

Exer 4: Consider the analogous definitions of random walks and hitting time for *directed* graphs, and show (that for every n) there exists a directed graph on n vertices and two vertices u, v such that $H_{uv} = 2^{\Omega(n)}$.

3 Commute Time

The *commute time* between vertices u and v is defined as $C_{uv} = H_{uv} + H_{vu} = C_{vu}$. It can be viewed as the expected time for a random walk that starts at u, to return to u after at least one visit to v. It is sometimes viewed as a symmetric version of the hitting time.

Lemma 2: We have the triangle inequality

$$\forall u, v, w \in V, \qquad C_{uw} \leq C_{uv} + C_{vw}.$$

The proof follows immediately from Lemma 1.

Theorem 3: For all $(u, v) \in E$, we have $C_{uv} \leq 2|E|$.

We will prove it in the next class, for now let's see some consequences.

Corollary 4: For all $u, v \in V$, we have $C_{uv} \leq 2(n-1)|E| < n^3$ (recall G is connected).

Proof: Follows from Lemma 2 (the triangle inequality) along a shortest path between u and v, and then applying Theorem 3.

4 Undirected Connectivity

Undirected st-connectivity (USTCON): In this problem, the input is a undirected graph G and two vertices s, t and the goal is to determine if s, t are the in the same connected component (equivalently, there is a path between them).

Theorem 5 [Aleliunas, Karp, Lipton, Lovasz, and Rackoff, 1979]: $USTCON \in RL$, i.e., USTCON can be solved by a randomized algorithm (Turing machine) that uses $O(\log n)$ bits of space and has one-sided error.

We did not see the proof, only briefly discussed it.

Remark: It was a big open problem to solve USTCON in deterministic logarithmic space, and Reingold proved it in 2005.

Exer 5: Show similarly how to decide whether all of G is connected (i.e., G has only one connected component) in randomized log-space.