CS271 Randomness & Computation **Fall 2022**

Lecture 24: November 15

Instructor: Alistair Sinclair

Disclaimer: *These notes have not been subjected to the usual scrutiny accorded to formal publications. They may be distributed outside this class only with the permission of the Instructor.*

24.1 Random Walks

Given an undirected graph $G = (V, E)$, with $n = |V|$ and $m = |E|$, a random walk is a stochastic process that starts from a given vertex, and then selects one of its neighbors uniformly at random to visit next.

```
start at v := s
repeat
  pick neighbor u of v u.a.r
  v := u
```

Random walks have many properties that are useful in designing and analyzing algorithms. Most of them are based on the following quantities.

Definition 24.1 *For any two vertices u, v, the **hitting time** from u to v is defined as*

$$H_{uv} = \mathrm{E}[\#\ \textit{of steps to reach } v \textit{ from } u].$$

Definition 24.2 *For any vertex u, the **cover time** from u is defined as*

$$C_u = \mathrm{E}[\#\ \textit{steps to visit all vertices starting from } u].$$

$C(G)$, the cover time of the graph, is defined as $\max_u C_u$ over all vertices u.

Definition 24.3 *The **mixing time** is (roughly) the number of steps for a random walk to reach its stationary distribution (details in the next class).*

Simple example: (Graph searching) One motivation for studying cover time is graph searching. Several standard deterministic graph search algorithms exist, such as Depth-First Search (DFS). On a connected undirected graph $G = (V, E)$, these algorithms run in linear time $O(|E|)$ and space $O(|V|)$ (in DFS, this is the space needed to maintain the stack of visited nodes).

On the other hand, random walk provides a very simple randomized algorithm that runs in (expected) time $O(|V||E|)$ and space $O(1)$; this algorithm simply performs random walk from some arbitrary vertex u. The expected time until the walk visits all vertices is just the cover time $C(G)$, which (as we shall see later) is bounded above by $2|V||E|$. The space requirement is $O(1)$ because the algorithm just needs to keep track of the current vertex. Note that this algorithm can be used to test if G is connected, since if a vertex v is not visited after $O(|V||E|)$ steps we can conclude that v is unlikely to be reachable from u.

Notes

1. The above two algorithms demonstrate a *time-space tradeoff*, in that the product of space and time remains constant. In fact, it turns out that there is a whole spectrum of randomized algorithms spanning this tradeoff [F97]:

   $$\forall s\ \exists \text{ algorithm using space } s \text{ and time } \tilde{O}(\tfrac{|V||E|}{s}) \text{ that visits all (reachable) vertices w.h.p.}$$

   (The notation $\tilde{O}$ hides logarithmic as well as constant factors.)

2. In an important breakthrough, Reingold [R05] showed that the above random walk algorithm could be *derandomized*, thus obtaining a deterministic $O(1)$ space algorithm for graph searching (phrased more precisely as testing whether there is a path between a given pair of vertices u, v. More correctly this is a $O(\log n)$ space algorithm, since $O(\log n)$ space is needed to record the label of a vertex in an n-vertex graph, and to keep track of the number of steps taken so far.) Since the connectivity problem is a canonical problem solvable in randomized log-space, this seems like a major step towards proving that every randomized log-space algorithm can be derandomized with only a constant factor penalty in space. However, this question remains open (see, e.g., [RTV06] for some interesting approaches). We should also observe that the analogous question for *directed* graphs is much more challenging, since testing connectivity in this case is a complete problem for *non-deterministic* log-space. Thus a deterministic (or even randomized) log-space algorithm for it would show that even non-determinism buys us nothing in the world of log-space algorithms. Note that the naive random walk approach breaks down here because the cover time on directed graphs can be exponentially large.

The bound on the cover time comes from the following theorem, originally due to Aleliunas *et al.* [AKLLR79].

Theorem 24.4 *For any connected graph G, $C(G) \leq 2|E||V|$.*

We will follow a different route due to Chandra *et al.* [CRRST89], which is based on an intriguing connection between random walks and electrical networks. Along the way, we will also obtain bounds on hitting times.

24.2 Hitting time and commute time

To prove theorem 24.4, we start by proving some properties of the hitting time for a vertex. To do this, we will view the graph as an electrical network, where each edge has unit resistance. (For a detailed treatment of the relationship between random walks and electrical networks, see the classic book by Doyle and Snell [DS84].) When a potential difference is applied between nodes of this network from an external source, current flows in the network in accordance with Kirchoff's Laws and Ohm's Law:

- **K1:** The total current into and out of a vertex is equal to zero.
- **K2:** The sum of potential differences around any cycle is zero.
- **Ohm's Law:** The current flowing along any edge is equal to $\frac{\text{pot. difference}}{\text{resistance}}$.

Definition 24.5 *The* ***effective resistance*** *$R(u, v)$ between two nodes u and v is the potential difference between u and v that is required to send one unit of current from u to v.*

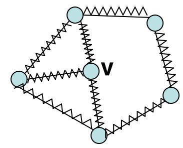

Figure 24.1: Graph as an electrical network

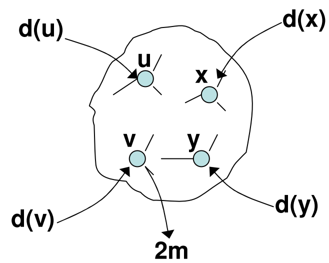

Figure 24.2: Scenario A

Now we prove a few useful lemmas about hitting times.

Consider first **Scenario A**, in which we inject $d(x)$ units of current into each node x, and remove all $2m = \sum_x d(x)$ units of current from node v, where $d(x)$ denotes the degree of x. (See Figure 24.2.)

Lemma 24.6 *The potential difference $\phi_{u,v}$ between v and any other node u in Scenario A satisfies $\phi_{u,v} = H_{u,v}$.*

Proof: Consider any vertex $u \in V$. Using Kirchoff's Laws and Ohm's Law, we have

$$\begin{aligned} d(u) &= \sum_{(u,x)\in E} (\text{current } u \to x) \quad (\mathbf{K1}) \\ &= \sum_{(u,x)\in E} \phi_{u,x} \quad (\mathbf{Ohm}) \\ &= \sum_{(u,x)\in E} (\phi_{u,v} - \phi_{x,v}) \quad (\mathbf{K2}) \\ &= d(u)\phi_{u,v} - \sum_{(u,x)\in E} \phi_{x,v}. \end{aligned}$$

Rearranging thus gives

$$\phi_{u,v} = 1 + \frac{1}{d(u)} \sum_{(u,x)\in E} \phi_{x,v}. \tag{24.1}$$

In addition, we of course have $\phi_{u,u} = 0$.

On the other hand, consider the random walk starting from u. Thinking of just the first step of this walk, we see that the hitting time $H_{u,v}$ satisfies

$$H_{u,v} = 1 + \frac{1}{d(u)} \sum_{(u,x)\in E} H_{x,v},$$

and again $H_{u,u} = 0$. But this is exactly the same system of linear equations as (24.1), satisfied by $\phi_{u,v}$. Since we know that this system has a unique solution (the potentials are uniquely determined by the current flows), we deduce that

$$H_{u,v} = \phi_{u,v} \quad \forall u \in V.$$

■

We now use Lemma 24.6 to deduce a more useful result that expresses the *commute time* $H_{u,v} + H_{v,u}$ precisely in terms of the effective resistance. (Note that the commute time, unlike the hitting time, is a symmetric quantity.)

Lemma 24.7 *For all u,v, the commute time* $H_{u,v} + H_{v,u} = 2mR_{u,v}$.

Proof: The proof makes use of **Scenario B**, which is like Scenario A except that we remove the $2m$ units of current from node u instead of from node v. Denoting potential differences in Scenario B by ϕ', we have, by Lemma 24.6,

$$\phi'_{v,u} = H_{v,u}.$$

Now consider **Scenario C**, which is like Scenario B but with all the currents reversed. Denoting potential differences in this scenario by ϕ'', we thus have

$$\phi''_{u,v} = \phi'_{v,u} = H_{v,u}.$$

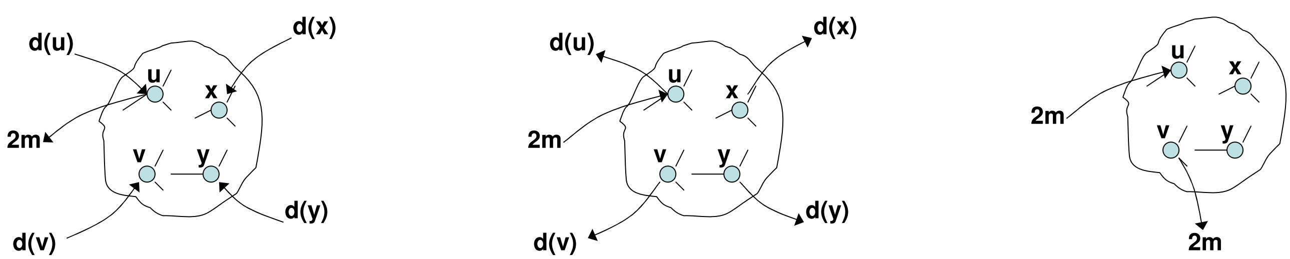

Figure 24.3: Scenarios B, C and D

Finally, consider **Scenario D**, which is just the sum of Scenarios A and C. By linearity, and denoting potential differences in Scenario D by ϕ''', we have

$$\phi'''_{u,v} = \phi_{u,v} + \phi''_{u,v} = H_{u,v} + H_{v,u}.$$

But note that $\phi'''_{u,v}$ is the potential difference required to move $2m$ units of current from u to v, so by definition it is equal to $2mR_{u,v}$. This completes the proof. ■

Examples

1. ***The line graph***
   Consider $n+1$ points on a line. By symmetry, $H_{0n} = H_{n0}$. Also, $H_{0n}+H_{n0} = 2mR_{0n} = 2\times n\times n = 2n^2$. Thus, $H_{0n} = H_{n0} = n^2$. (This is in accordance with the result we obtained earlier using the martingale optional stopping theorem.)

2. ***The "lollipop"***
   As opposed to the previous case, this example illustrates a highly asymmetric situtation in which the values of H_{uv} and H_{vu} are very different from each other. The lollipop is a graph with a line of vertices joined to a clique. Consider an n-vertex lollipop with $n/2+1$ vertices in a line, the last one being a part of a clique of $n/2$ nodes, as shown in Figure 24.4. Let u and v be the endpoints of the line. Lemma 24.7 gives

$$H_{uv} + H_{vu} = 2mR_{uv} = 2 \times \Theta(n^2) \times \Theta(n) = \Theta(n^3).$$

   However, we know from the previous example that $H_{uv} = \Theta(n^2)$, which implies that $H_{vu} = \Theta(n^3)$. In this case, one can easily see the asymmetry between u and v in the number of neighbors. While each random walk starting at u has no option but to go in the direction of v, a random walk starting at v has very little probability of proceeding along the line. Thus, the extra factor of n in this case basically measures the latency of getting started along the line.

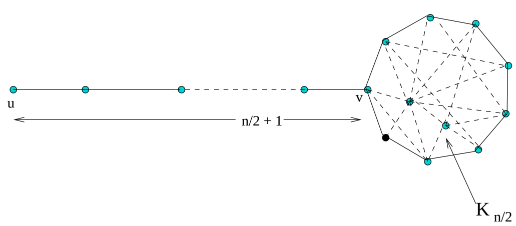

Figure 24.4: The lollipop graph

24.3 The cover time

We now give bounds on the cover time of a graph. The first theorem gives a rather loose upper bound which is completely independent of the structure of the graph except the number of edges. We later give more sensitive bounds in terms of the resistance of the graph.

Our first bound is the one stated earlier as Theorem 24.4:

Theorem 24.4 *For any connected graph G, $C(G) \leq 2|E||V|$*

Proof: Consider any spanning tree T of G. For any vertex u, it is possible to traverse the entire tree and come back to u covering each edge exactly twice. Clearly, the cover time for the vertex u is upper bounded by the expected time for the random walk to visit the vertices of G in this order. Let $u = v_0, v_1, \ldots, v_{2n-2} = u$ denote the vertices in the order they are visited by a particular traversal. Then,

$$C(u) \leq \sum_{i=0}^{2n-2} H_{v_i v_{i+1}} = \sum_{(x,y)\in T} (H_{xy} + H_{yx}).$$

Applying Lemma 24.7, and using the fact that for any two *adjacent* vertices x, y the effective resistance

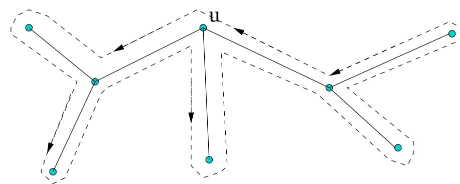

Figure 24.5: Traversal of a spanning tree

$R_{x,y} \leq 1$, we have

$$\begin{aligned} C(G) = \max_{u \in V} C(u) &\leq \sum_{(x,y) \in T} (H_{xy} + H_{yx}) \\ &= 2m \sum_{(x,y) \in T} R_{xy} && \text{(using Lemma 24.7)} \\ &\leq 2|E||V| && (\text{as each } \mathrm{R_{xy}} \leq 1). \end{aligned}$$

This completes the proof. ■

Let us examine the bounds given by the above theorem for a few graphs.

1. ***The line graph***
   For the line graph with $n+1$ vertices and n edges, we get
   $$C(G) \leq 2 \times n \times (n+1) \approx 2n^2$$
   Also, we know that $C(G) \geq H_{0n} = n^2$. Thus, the bound is tight up to constants in this case.

2. ***The lollipop graph***
   Application of the theorem to the lollipop graph gives
   $$C(G) \leq 2 \times \Theta(n^2) \times n = \Theta(n^3).$$
   Again, one can see that it is tight in this case as $C(G) \geq H_{vu} = \Theta(n^3)$. In fact, a result by Feige [F95:1] shows that the lollipop graph is extremal for cover times, i.e., it has the highest cover time amongst all graphs with n vertices. The value of its cover time is shown to be $C(G) = \frac{4}{27}n^3 + o(n^3)$.

3. ***The complete graph***
   For the complete graph on n vertices, we get $C(G) \leq 2|E||V| = \Theta(n^3)$. However, in this case we can easily derive a better bound. At each step, the random walk has a $1/(n-1)$ chance of visiting every vertex other than the current one. If we add self-loops, this can only increase the cover time. Now the problem becomes the same as coupon collecting, with the probability of visiting any vertex being $1/n$ at each step. Thus the cover time is $C(G) = (1+o(1))n \ln n$, which is dramatically less than the above estimate. Incidentally, it has been shown in [F95:2] that this is asymptotically the smallest possible cover time for any n-vertex graph, though the complete graph is not quite extremal here.

The above examples illustrate a possibly counter-intuitive property of cover times (and hitting times): they are not monotonic w.r.t. adding edges to the graph. Starting with the line graph, whose cover time is $\Theta(n^2)$, we can add edges to obtain the lollipop, with larger cover time $\Theta(n^3)$; adding further edges we obtain the complete graph, whose cover time is less than both of these.

Our final theorem gives upper and lower bounds on the cover time of a graph that are within a logarithmic factor of one another. The bounds are in terms of the *resistance* of the graph, defined as $R = \max_{u,v \in V} R_{uv}$.

Theorem 24.8 *For a connected graph G, $|E|R \leq C(G) \leq c(|E|R \log |V|)$, for some universal constant c, where R is the resistance of G.*

Proof: The lower bound follows easily from the facts that $C(G) \geq \max_{u,v} H_{uv}$, and $\max(H_{uv}, H_{vu}) \geq \frac{1}{2}(H_{uv} + H_{vu})$. Together with Lemma 24.7 this gives

$$C(G) \geq \frac{1}{2}(H_{uv} + H_{vu}) = \frac{1}{2}(2mR_{uv}) = |E|R_{uv} \quad \forall u, v \in V.$$

Thus, we get $C(G) \geq |E|R$, as required.

For the upper bound, we divide the random walk into $\ln n$ "epochs" of length $2a|E|R$ each (where $n = |V|$ and a is a constant to be chosen later). Let u be the starting vertex of the random walk. The expected time required for hitting a vertex v during any epoch is at most $\max_x H_{xv}$. Thus, for any epoch i and any vertex v,

$$\begin{aligned} \Pr[v \textit{ is not hit during epoch } i] &\leq \frac{\max_x H_{xv}}{2a|E|R} && \text{(using Markov's inequality)} \\ &\leq \frac{1}{a} && \text{(using } H_{xv} \leq H_{xv} + H_{vx} = 2|E|R_{xv} \leq 2|E|R\text{).} \end{aligned}$$

Thus,

$$\Pr[v \textit{ is not hit during any epoch}] \;\leq\; \left(\frac{1}{a}\right)^{\ln n} \;=\; n^{-\ln a}.$$

By a union bound,

$$\Pr[\textit{some vertex is not hit during any epoch}] \;\leq\; n^{1-\ln a}.$$

Conditioning on whether or not the walk has visited all vertices after all $2a|E|R\ln n$ steps, and using our previous upper bound of $|E||V| \leq n^3$ on the cover time, we have

$$C_u \leq 2a|E|R\ln n + n^{1-\ln a} \times n^3.$$

Finally, choosing a sufficiently large so that the second term is small (say $a = e^4$), we get $C_u \leq c(|E|R\log n)$ for all $u \in V$, and hence $C(G) \leq c(|E|R\log n)$. This completes the proof of the theorem. ∎

We now apply this theorem to the complete graph. Starting at any vertex u, the chance of hitting a given vertex v is $1/(n-1)$ at every step. Thus, for all vertices u and v, $H_{uv} = n - 1$. Also, we have

$$H_{uv} + H_{vu} = 2|E|R_{uv} \Rightarrow 2(n-1) = 2 \times \frac{n(n-1)}{2} \times R_{uv} \Rightarrow R_{uv} = \frac{2}{n}.$$

From Theorem 24.8 we get $C(G) \;\leq\; c(\frac{n(n-1)}{2} \times \frac{2}{n} \times \log n) \;=\; O(n\log n)$, which is tight up to constants.

On the other hand, note that for the lollipop graph we have $R = \Theta(n)$ and $|E| = \Theta(n^2)$, so the upper bound in Theorem 24.8 gives $C(G) \leq O(n^3 \log n)$, which is slightly worse (by a factor of $\log n$) than Theorem 24.4.

References

[AKLLR79] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff, "Random walks, universal traversal sequences, and the complexity of maze traversal," *Proceedings of the 20th IEEE Symposium on Foundations of Computer Science*, 1979, pp. 218–233.

[CRRST89] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky and P. Tiwari, "The electrical resistance of a graph captures its commute and cover times," *Proceedings of the 21st ACM Symposium on Theory of Computing*, 1989, pp. 574–586.

[DS84] P. Doyle and J. Snell, *Random walks and electric networks*, The Mathematical Association of America, vol. 22, 1987.

[F97] U. Feige, "A Spectrum of Time-Space Tradeoffs for Undirected s-t Connectivity," *Journal of Computer and System Sciences*, 54(2), 305–316, 1997.

[F95:1] U. Feige "A tight upper bound on the cover time for random walks on graphs," *Random Structures & Algorithms* **6** (1995), pp. 51–54.

[F95:2] U. Feige "A tight lower bound on the cover time for random walks on graphs," *Random Structures & Algorithms* **6** (1995), pp. 433–438.

[R05] O. Reingold "Undirected st-connectivity in log space," *Proceedings of ACM STOC*, 2005, pp. 376–385.

[RTV06] O. Reingold, L. Trevisan and S. Vadhan, "Pseudorandom walks on regular digraphs and the RL vs. L problem," *Proceedings of ACM STOC*, 2006, pp. 457466.