5 Random Walks on Graphs

A random walk on a graph consists of a sequence of vertices generated from a start vertex by randomly selecting an edge, traversing the edge to a new vertex, and repeating the process. Under some simple conditions, the probability that the walk is at a given vertex at time t converges to a limit independent of the start state as t goes to infinity. This limit is called the stationary probability of the vertex. The stationary probability of a vertex equals the fraction of time the walk spends at the vertex in the long term.

The study of random walks is a classical subject with many applications. An important recent one is in defining the pagerank of pages on the World Wide Web by their stationary probabilities. For undirected graphs, there is an analogy between random walks and electrical circuits which we develop first. Then we consider the special case of random walks on a lattice. Finally we consider random walks on directed graphs

5.1 Electrical networks and random walks

An electrical network is a connected, undirected graph in which each edge *xy* has a resistance $r_{xy} > 0$. In what follows it will be easier to deal with conductance defined by $c_{xy} = \frac{1}{r_{xy}}$ rather then resistance. Associated with an electrical network is a random walk on the underlying graph defined by assigning a probability to each edge incident to a vertex. Let $p_{xy} = \dfrac{c_{xy}}{c_x}$ be the probability assigned to the walk selecting edge *xy* at vertex *x* where $c_x = \sum_y c_{xy}$. Note that although $c_{xy} = c_{yx}$, p_{xy} and p_{yx} may not be equal due to the required normalization so that the probabilities at each vertex sum to one. We shall soon see that there is a relationship between current flowing in an electrical network and a random walk on the underlying graph.

We denote by P the matrix whose $(x, y)^{\text{th}}$ entry is p_{xy}, the probability of a transition from x to y. The matrix P is called the *transition probability matrix*. Suppose we start the random walk at a vertex x. At the start, the probability mass at x is 1 and 0 at every other vertex. At time one, for every state y, the probability of being at y is the probability of going from x to y, namely p_{xy}. Generally, for any state y, the probability of being at y at time $t+1$, is sum over each state x of being at x at time t and taking the transition from x to y at the $t+1^{\text{st}}$ step. If $\mathbf{p}$ is a row vector with a component for each node giving the probability mass at the node at time t, and $\mathbf{q}$ is a similar vector at time $t+1$, then the above can be written in matrix notation as

$$\mathbf{p}P = \mathbf{q}.$$

If the stationary probability (call it $\mathbf{f}$) exists, repeatedly multiplying the probability distribution by P yields the stationary probability distribution $\mathbf{f}$ where

$$\mathbf{f}P = \mathbf{f}.$$

A graph, such as a bipartite graph, in which the greatest common divisor of all cycle lengths is greater than one is said to be *periodic*. A graph that is connected and not periodic has a unique stationary probability. In a periodic graph the probability distribution will cycle among a set of distributions instead of converging to a stationary probability distribution. For example, with bipartite graphs the probability distribution will alternate between two distributions on odd and even numbered steps.

The stationary probability (assuming it exists) is given by $f_x = \frac{c_x}{c}$ where $c = \sum_x c_x$. It is easy to check that this $\mathbf{f}$ satisfies the condition $\mathbf{f}P = \mathbf{f}$:

$$(\mathbf{f}P)_x = \sum_y \frac{c_{yx}}{c_y}\frac{c_y}{c} = \sum_y \frac{c_{xy}}{c} = \frac{c_x}{c} = f_x .$$

Note that if each edge has resistance one, then the value of $c_x = \sum_y c_{xy}$ is d_x where d_x is the degree of *x*. In this case, $c = \sum_x c_x$ equals *2m* where *m* is the total number of edges and the stationary probability is $\frac{1}{2m}(d_1, d_2, \cdots, d_n)$. This means that for undirected graphs, the stationary probability of each vertex is proportional to its degree and that every edge is traversed in each direction with the same probability of $\frac{1}{2m}$.

A random walk associated with an electrical network has the important property that given the stationary probability, the probability $f_x p_{xy}$ of traversing the edge *xy* from vertex *x* to vertex *y* is the same as the probability $f_y p_{yx}$ of traversing the edge in the reverse direction from vertex *y* to vertex *x*. This follows from the manner in which probabilities were assigned and the fact that the conductance $c_{xy} = c_{yx}$.

$$f_x p_{xy} = \frac{c_x}{c}\frac{c_{xy}}{c_x} = \frac{c_{xy}}{c} = \frac{c_{yx}}{c} = \frac{c_y}{c}\frac{c_{yx}}{c_y} = f_y p_{yx} .$$

Harmonic functions

Harmonic functions will be useful in developing the relationship between electrical networks and random walks on undirected graphs. Given an undirected graph, designate certain vertices as boundary vertices and the remaining vertices as interior vertices. The value of the function at the boundary vertices is fixed to some boundary condition and this imparts values to the interior vertices. A harmonic function g on the vertices is one in which the value of g at any interior vertex x is a weighted average of the value at all its adjacent vertices y, where the weights p_{xy} sum to 1 over all y. So, g is harmonic if and only if $g_x = \sum_y g_y p_{xy}$. From the fact that $\mathbf{f}P = \mathbf{f}$, we see that the function $g_x = \frac{f_x}{c_x}$ is harmonic:

$$g_x = \tfrac{f_x}{c_x} = \tfrac{1}{c_x}\sum_y f_y p_{yx} = \tfrac{1}{c_x}\sum_y f_y \tfrac{c_{xy}}{c_y} = \sum_y g_y p_{xy}$$

An harmonic function takes on its maximum and minimum on the boundary. This is an immediate consequence of the fact that the value at any interior point is a weighted average of the adjacent values. There is a unique harmonic function satisfying a given set of equations and boundary condition. For suppose there were two solutions f(x) and g(x). The difference of two solutions is itself harmonic. Since h(x)=f(x)-g(x) is harmonic and has value zero on the boundary, by the maximum principle it has value zero everywhere. Thus f(x)=g(x).

The important connection will be that voltage at a node in an electrical network is also a harmonic function. We show below that there is essentially a unique harmonic function for a graph.

The analogy between electrical networks and random walks

Choose two vertices a and b. For reference purposes let the voltage v_b equal zero. Attach a current source between a and b so that the voltage v_a equals one. Fixing the voltages at v_a and v_b induces voltages at all other vertices along with a current flow through the edges of the network. The analogy between electrical networks and random walks is the following. First having fixed the voltages at the vertices v_a and v_b, the voltage at an arbitrary vertex *x* equals the probability of reaching *a* from *x* before reaching *b*. Second, if the voltage v_a is adjusted so that the current flowing into vertex *a* is one, then the current flowing through an edge is the net frequency in which a random walk from *a* to *b* traverses the edge.

Probabilistic interpretation of voltages

Before showing that the voltage at an arbitrary vertex *x* equals the probability of reaching *a* from *x* before reaching *b*, we first show that the voltages form a harmonic function. Let *x* and *y* be adjacent vertices and let i_{xy} be the current flowing through the edge from x to y. By Ohm's law,

$$i_{xy} = \frac{v_x - v_y}{r_{xy}} = (v_x - v_y)c_{xy}$$

By Kirchoff's Law the currents flowing out of any vertex sum to zero.

$$\sum_y i_{xy} = 0$$

Replacing currents in the above sum by the voltage difference times conductance yields

$$\sum_y (v_x - v_y)c_{xy} = 0$$

or

$$v_x \sum_y c_{xy} = \sum_y v_y c_{xy}$$

Observing that $\sum_y c_{xy} = c_x$ and that $p_{xy} = \frac{c_{xy}}{c_x}$ yields $v_x c_x = \sum_y v_y p_{xy} c_x$. So, $v_x = \sum_y v_y p_{xy}$. Thus, the voltage at each vertex *y* is a weighted average of the voltages at the adjacent vertices. Hence the voltages are harmonic.

Now let p_x be the probability that a random walk starting at vertex *x* reaches *a* before *b*. Then $p_a = v_a = 1$ and $p_b = v_b = 0$. Furthermore, the probability of reaching *a* from *x* before reaching *b* is the sum over all *y* adjacent to *x* of going to *y* and then reaching a from *y* before reaching *b*. That is

$$p_x = \sum_y p_{xy} p_y$$

Thus p_x is the same harmonic function as the voltage function v_x and v and p satisfy the same boundary conditions (a,b form the boundary). Thus, they are identical functions. The probability of reaching *a* from *x* before reaching *b* is the voltage v_x.

Probabilistic interpretation of current

Set the current into the network at *a* to have value one. We then show that the current i_{xy} is the net frequency with which a random walk from *a* to *b* goes through the edge *xy* before reaching b. Let u_x be the expected number of visits to vertex *x* on a walk from *a* to *b* before reaching *b*. Now $u_b = 0$. For $x \neq a, b$ since every time we visit x, we must come to x from some vertex *y*, the number of visits to *x* is the sum over all *y* of the number of visits u_y to *y* times the probability p_{xy} of going from y to x. Thus

$$u_x = \sum_y u_y p_{yx}$$

It follows that $\frac{u_x}{c_x}$ is harmonic (with a,b as the boundary). Now, $\frac{u_b}{c_b} = 0$. Setting the current into *a* to one fixed the value of v_a. Set u_a so that $\frac{u_a}{c_a} = v_a$. Since $\frac{u_x}{c_x}$ and v_x satisfy the same harmonic conditions, they are the same harmonic function. Let the one amp correspond to one walk. Note that if our walk starts at *a* and ends at *b*, the expected value of the difference between the number of times the walk leaves *a* and enters *a* must be one.

Next we need to show that the current i_{xy} is the net frequency with which a random walk traverses edge *xy*.

$$i_{xy} = (v_x - v_y)c_{xy} = \left(\frac{u_x}{c_x} - \frac{u_y}{c_y}\right)c_{xy} = u_x \frac{c_{xy}}{c_x} - u_y \frac{c_{xy}}{c_y} = u_x p_{xy} - u_y p_{yx}$$

The quantity $u_x P_{xy}$ is the expected number of times the edge *xy* is traversed from *x* to *y* and the quantity $u_y P_{yx}$ is the expected number of times the edge *xy* is traversed from *y* to *x*. Thus, the current i_{xy} is the expected net number of traversals of the edge *xy* from *x* to *y*.

Effective Resistance and Escape Probability

Set $v_a = 1$ and $v_b = 0$. Let i_a be the current flowing into the network at vertex *a* and out at vertex *b*. Define the *effective resistance* r_{eff} between *a* and *b* to be $r_{eff} = \frac{v_a}{i_a}$ and the *effective conductance* c_{eff} to be $c_{eff} = \frac{1}{r_{eff}}$. Define the *escape probability* to be the probability that a random walk starting at *a* reaches *b* before returning to *a*. We now show that $\frac{c_{eff}}{c_a}$ is the escape probability.

$$i_a = \sum_y (v_a - v_y) c_{ay}$$

Since $v_a = 1$

$$\begin{aligned} i_a &= \sum_y (1 - v_y) \frac{c_{ay}}{c_a} c_a \\ &= c_a \left[\sum_y \frac{c_{ay}}{c_a} - \sum_y v_y \frac{c_{ay}}{c_a} \right] \\ &= c_a \left[1 - \sum_y p_{ay} v_y \right] \\ &= c_a p_{escape} \end{aligned}$$

where p_{escape} is the probability that starting at *a* the walk reaches *b* before retuning to *a*. For each y adjacent to the vertex a, P_{ay} is the probability of the walk going from vertex a to vertex y. v_y is the probability of a walking starting at y going to a before reaching b. Thus $\sum_y P_{ay} v_y$ is the probability of a walk starting at *a* returning to *a* before reaching *b* and $1 - \sum_y P_{ay} v_y$ is the probability of a walk starting at *a*, reaching *b* before returning to *a*.

Since $v_a = 1$ and $c_{eff} = \frac{i_a}{v_a}$, it follows that $i_a = c_{eff}$. Thus $c_{eff} = c_a p_{escape}$ or $p_{escape} = \frac{c_{eff}}{c_a}$.

For a finite graph the escape probability will always be non zero. Now consider an infinite graph such as a lattice and a random walk starting at some vertex *a*. Form a series of finite graphs by merging all vertices at distance *d* from *a* into a single vertex *b* for larger and larger values of *d*. The limit as *d* goes to infinity of p_{escape} is the probability that the random walk will never return to *a*. If $p_{escape} = 0$, then eventually any random walk will return to *a*. If $p_{escape} \neq 0$, then a fraction of the walks never return. Thus, the escape probability terminology.

Random walks on undirected graphs

We now focus our discussion on random walks on undirected graphs with uniform edge weights. At each vertex, the random walk is equally likely to take any edge. This corresponds to an electrical network in which all edge resistances are one. Assume the graph is connected. If it is not, the analyses below can be applied to each connected component separately. We consider questions such as what is the expected time for a random walk starting at a vertex *u* to reach a target vertex *v*, what is the expected time until the random walk returns to the vertex it started at, and what is the expected time to reach every vertex?

Hitting time

The *hitting time* h_{uv} is the expected time of a random walk starting at vertex *u* to reach vertex *v*. Sometimes a more general definition is given where the hitting time is the expected time to reach a vertex *v* by randomly selecting the start vertex according to some probability distribution.

One interesting fact is that adding edges to a graph may either increase or decrease h_{uv} depending on the particular situation. An edge can shorten the distance from *u* to *v* thereby decreasing h_{uv} or the edge could increase the probability of a random walk going to some far off portion of the graph thereby increasing h_{uv}. Another interesting fact is that hitting time is not symmetric. The expected time to reach a vertex *v* from a vertex *u* in an undirected graph may be radically different from the time to reach *u* from *v*.

We start with two technical lemmas. The first lemma states that the expected time to traverse a chain of *n* vertices is $\Theta\left(n^2\right)$.

Lemma 5.1: The expected time for a random walk starting at one end of a chain of *n* vertices to reach the other end is $\Theta\left(n^2\right)$.

Proof: **C**onsider walking from vertex 1 to vertex *n* in a graph consisting of a single path of *n* vertices. Let h_{ij}, i<j, be the hitting time of reaching *j* starting from *i*. Now $h_{1,2} = 1$ and

$$h_{i,i+1} = \tfrac{1}{2}\times 1 + \tfrac{1}{2}\left(1 + h_{i-1,i} + h_{i,i+1}\right) \quad 2 \le i \le n-1 .$$

Solving for $h_{i,i+1}$ yields the recurrence

$$h_{i,i+1} = 2 + h_{i-1,i}$$

Solving the recurrence yields

$$h_{i,i+1} = 2i - 1.$$

To get from 1 to *n*, go from 1 to 2, 2 to 3, etc. Thus

$$\begin{aligned} h_{1,n} &= \sum_{i=1}^{n-1} h_{i,i+1} = \sum_{i=1}^{n-1}(2i-1) \\ &= 2\sum_{i=1}^{n-1} i - \sum_{i=1}^{n-1} 1 \\ &= 2\tfrac{n(n-1)}{2} - (n-1) \\ &= (n-1)^2 \end{aligned}$$

■

The next lemma shows that the time spent at vertex *i* by a random walk from vertex 1 to vertex *n* in a chain of *n* vertices is 2(i-1) for $2 \le i \le n-1$.

Lemma 5.2: Consider a random walk from vertex 1 to vertex *n* in a chain of *n* vertices. Let t(i) be the expected time spent at vertex i. Then

$$t(i) = \begin{cases} n-1 & i = 1 \\ 2(n-i) & 2 \le i \le n-1 \\ 1 & i = n \end{cases}$$

Proof: Now $t(n) = 1$ since the walk stops when it reaches vertex n. Half of the time when the walk is at vertex $n-1$ it goes to vertex *n*. Thus $t(n-1) = 2$. For $3 \le i \le n-1$, $t(i) = \frac{1}{2}\left[t(i-1) + t(i+1)\right]$ and $t(1)$ and $t(2)$ satisfy $t(1) = \frac{1}{2}t(2) + 1$ and $t(2) = t(1) + \frac{1}{2}t(3)$. Solving for t(i+1) for $3 \le i \le n-1$ yields t(i+1)=2t(i)-t(i-1) which has solution t(i)=2(n-i) for $3 \le i \le n-1$. Then solving for $t(1)$ and $t(2)$ yields $t(2) = 2(n-2)$ and $t(1) = n-1$. Thus, the total time spent at vertices is

$$n - 1 + 2(1 + 2 + \cdots + n - 2) + 1 = n^2 - 2n + 2 = (n-1)^2 + 1$$

which is one more than h_{1n} and thus is correct.

■

Next we show that adding edges to a graph might either increase or decrease the hitting time h_{uv}. Consider the graph consisting of a single path of *n* vertices. Add edges to this graph to get the graph in Fig. 5.2 consisting of a clique of size n/2 connected to a path of n/2 vertices. Then add still more edges to get a clique of size *n*. Let *u* be the vertex at the midpoint of the original path and let *v* be the other endpoint of the path consisting of n/2 vertices as shown in Fig. 5.2. In the first graph $h_{uv} = \Theta\left(n^2\right)$. In the second graph $h_{uv} = \Theta\left(n^3\right)$. To see this note that staring at *u* the walk will go down the chain towards *v* and return to *u* *n* times before reaching *v* for the first time. At *u* with probability only $1/n$ will the walk start back down the chain. If the walk goes into the clique, it spends *n* steps before returning to *u* and on average will do this *n* times before starting down the chain again. In the third graph, which is the clique of size *n*, $h_{uv} = \Theta\left(n \log n\right)$. Thus, adding edges first increased h_{uv} from n^2 to n^3 and then decreased it to $n \log n$.

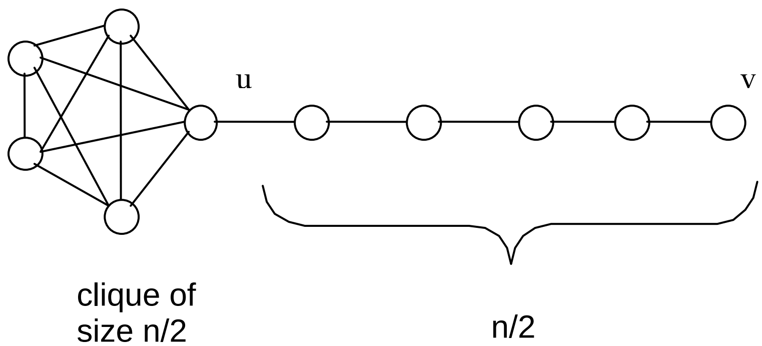


Figure 5.2: Graph illustrating that adding edges to a graph can either increase or decrease hitting time.

Hitting time is not symmetric even in the case of undirected graphs. In the above graph, the expected time, h_{uv}, of a random walk from *u* to *v* (where *u* is the vertex of attachment and *v* is the end vertex) is $\Theta(n^3)$. However, reversing *u* and *v*, h_{vu} is $\Theta(n^2)$.

Next we ask what is the maximum that the hitting time could be. We first show that if vertices *u* and *v* are connected by an edge, that the expected time h_{uv} of a random walk from *u* to *v* plus the expected time h_{vu} from *v* to *u* is less than twice the number of edges.

Lemma 5.3: If vertices *u* and *v* are connected by an edge, then $h_{uv} + h_{vu} \le 2m$.

Proof: In a random walk on an undirected graph the probability of traversing an arbitrary edge in either direction is independent of the edge and the direction. To see this, note that if it were true, the probability of arriving at a vertex is proportional to the degree of the vertex, and hence the probability of leaving on a given edge is the same independent of which edge in the graph it is. This means that the probability of traversing an edge in a given direction is $\frac{1}{2m}$ and thus the expected time between transversals of the directed edge (*u,v*) is $2m$. Thus, if we traverse edge (*u,v*), the expected time to traverse a path from *v* back to *u* and then traverse the edge (*u,v*) again is *2m*. But since a random walk is a memory less process we can drop the condition that we started by traversing the edge (*u,v*). Hence the expected time from *v* to *u* and back to *v* is at most *2m*. Note the path we follow went from *v* to *u* and then may have returned to either *u* or *v* several time before going through the edge (*u*,*v*). Thus, the $\le$ sign since we may have gone from *v* to *u* to *v* without going through the edge (*u*,*v*).

■

Notice that the proof relied on the fact that there was an edge from *u* to *v* and thus the theorem is not necessarily true for arbitrary *u* and *v*. When *u* and *v* are not connected by an edge consider a path from *u* to *v*. The path is of length at most *n*. Consider the time it takes to reach each vertex on the path in the order they appear. Since the vertices are connected by an edge, reaching the next vertex on the path takes time at most equal to the number of edges in the graph by the above theorem. Thus the total expected time is at $\Theta\left(n^3\right)$. This result is asymptotically tight since the bound is met by the graph of Fig. 5.2 consisting of a clique of size *n*/2 and a path of length *n*/2.

Commute time

The *commute time*, *commute*(*u*,*v*), is the expected time of a random walk starting at *u* arriving back at *u* hitting *v* at least once on the way. Think of going from home to office and returning home. **SAME SYMBOL AS WAS USED FOR CONDUCTANCE ALSO IN THM 6.6 WE USE u AND v FOR VERTICES BUT IN PROOF USE i AND j. Sometime a and b.**

Theorem 5.1: Given an undirected graph consider the electrical network where each node of the graph is replaced by a one ohm resistor. Given vertices *u* and *v*, the commute time *commute*(*u*,*v*) equals $2mr_{uv}$ where r_{uv} is the effective resistance from *u* to *v* and *m* is the number of edges in the graph.

Proof: Insert at each vertex *i* a current equal to the degree d_i of vertex *i*. The total current inserted is *2m* where *m* is the number of edges and thus extract a current of *2m* from vertex *j*. Let v_{ij} be the voltage difference from *i* to *j*. The current into *i* divides into the d_i resistors at node *i*. The current in each resistor is proportional to the voltage across it. Let *k* be a vertex adjacent to *i*. Then the current through the resistor between *i* and *k* is $v_{ij} - v_{kj}$, the voltage drop across the resister. The sum of the currents out of *i* through the resisters must equal d_i the current injected into *i*.

$$d_i = \sum_{\substack{\text{k adj}\\ \text{to i}}} (v_{ij} - v_{kj})$$

Noting that v_{ij} does not depend on *k*, we write

$$d_i = \sum_{\substack{\text{k adj}\\ \text{to i}}} v_{ij} - \sum_{\substack{\text{k adj}\\ \text{to i}}} v_{kj} = d_i v_{ij} - \sum_{\substack{\text{k adj}\\ \text{to i}}} v_{kj}.$$

Solving for v_{ij}

$$v_{ij} = 1 + \sum_{\substack{\text{k adj}\\ \text{to i}}} \tfrac{1}{d_i} v_{kj} = \sum_{\substack{\text{k adj}\\ \text{to i}}} \tfrac{1}{d_i}(1 + v_{kj}) \qquad \text{Eq. 5.1}$$

Now the expected time from *i* to *j* is the average time over all paths from *i* to *k* adjacent to *i* and then on from *k* to *j*. This is given by

$$h_{ij} = \sum_{\substack{\text{k adj} \\ \text{to i}}} \tfrac{1}{d_i}(1 + h_{kj}) \qquad \text{Eq. 5.2}$$

Subtracting Eq. 5.2 from Eq. 5.1, we get $v_{ij} - h_{ij} = \sum_{\substack{\text{k adj} \\ \text{to i}}} \tfrac{1}{d_i}(v_{kj} - h_{kj})$. Thus the function $v_{ij} - h_{ij}$ is harmonic. Designate vertex *j* has the only exterior vertex. The value of $v_{ij} - h_{ij}$ at j, namely $v_{jj} - h_{jj}$, is zero. So it must be zero everywhere. Thus, the voltage v_{ij} equals the expected time h_{ij} from *i* to *j*.

To complete the proof, note that $h_{ij} = v_{ij}$ is the voltage from *i* to *j* when currents are inserted at all nodes in the graph and extracted from *j*. If we extract the current from *i* instead of *j*, then the voltages change and $v_{ji} = h_{ji}$. Finally, reverse all currents in this latter step. The voltages change again and for the new voltages $-v_{ji} = h_{ji}$. Since $-v_{ji} = v_{ij}$, we get $h_{ji} = v_{ij}$.

Thus when we insert a current at each node equal to its degree and extract the current from *j*, the voltage v_{ij} in this set yo equals h_{ij} When we extract the current from *i* instead of *j* and then reverse all currents, the voltage v_{ij} in this new set up equals h_{ji}. By super position of currents $v_{ij} = h_{ij} + h_{ji}$. All currents cancel except the *2m* amps injected at *i* and withdrawn at *j*. Thus, $2mR_{ij} = v_{ij} = h_{ij} + h_{ji} = c_{ij}$. Thus $commute_{ij} = 2mR_{ij}$

■

Corollary 5.1: For any *n*-vertex graph and for any vertices *u* and *v*, the commute time c_{uv} is less than or equal to n^3

Proof: By Theorem 5.1 the commute time is given by the formula $commute_{ij} = 2mR_{ij}$ where m is the number of edges. In an *n* vertex graph there exists a path from *u* to *v* of length at most *n*. This implies $r_{uv} \le n$ since the resistance can not be greater than that of any path from *u* to *v*. Since the number of edges is at most $\binom{n}{2}$

$$c_{uv} = 2mr_{uv} \le 2\binom{n}{2} n = n^3$$

■

Again adding edges to a graph may increase or decrease the commute time. To see this, consider the graph consisting of a chain of n vertices, the graph of Fig. 5.2, and the clique on *n* vertices.

Cover times

The *cover time* $\text{cover}(u,G)$ is the expected time of a random walk starting at vertex *u* in the graph G to reach each vertex at least one. We write $\text{cover}(u)$ when G is understood. The cover time of an undirected graph G, denoted cover(G), is
$\text{cover}(G) = \max_u \text{cover}(u,G).$

For cover time of an undirected graph increasing the number of edges in the graph may increase or decrease the cover time depending on the situation. Again consider three graphs, a chain of length n which has cover time $\Theta(n^2)$, the graph in Figure 5.2 which has cover time $\Theta(n^3)$, and the complete graph on n vertices which has cover time $\Theta(n\log n)$. Thus adding edges to the chain of length n to create the graph in Figure 5.2 increases the cover time from n^2 to n^3 and then adding even more edges to obtain the complete graph reduces the cover time to nlogn.

Note: The cover time of a clique is $n\log n$ since that is the time to select (with high probability) every integer out of *n* integers, drawing integers at random. This is the coupon collector problem. The cover time for a straight line is $\Theta(n^2)$ since it is the same as the discovery time. For the graph in Fig. 5.2, the cover time is $\Theta(n^3)$ since one takes the maximum over all start states and $\text{cover}(u,G) = \Theta\left(n^3\right)$..

Theorem 5.2: Let G be a graph with *n* vertices and *m* edges. The time for a random walk to cover all vertices of a graph G is bounded above by $2m(n-1)$.

Proof: Consider a depth first search (dfs) of the graph G starting from vertex *i* and let T be the resulting dfs spanning tree of G. The dfs covers every vertex. Consider the expected time to cover every vertex in the order visited by the depth first search. Clearly this bounds the cover time of G starting from vertex i.

$$\text{cover}(i,G) \le \sum_{(u,v)\in T} h_{uv} .$$

Since $(u,v)\in T$ are adjacent Lemma 6.5 implies $h_{uv} \le 2m$. Since there are *n*-1 edges in the dfs tree $\text{cover}(i) \le 2m(n-1)$. Since this holds for all starting vertices i, $\text{cover}(G) \le 2m(n-1)$

■

The theorem gives the correct answer, n^3, for the n/2 clique with the n/2 tail. It gives an upper bound of n^3 for the n-clique where the actual cover time is $n\log n$.

Let r_{uv} be the effective resistance from u to v. Define the resistance r(G) of a graph G by $r(G) = \max_{u,v}(r_{uv})$.

Theorem 5.3: Let G be an undirected graph with *m* edges. Then the cover time for G is bounded by the following inequality

$$mr(G) \le \text{cover}(G) \le 2e^3 mr(G)\ln n + n$$

where e=2.71 is Euler's constant.

Proof: By definition $r(G) = \max_{u,v}(r_{uv})$. Let *u* and *v* be the vertices of G for which r_{uv} is maximum. Then $r(G) = r_{uv}$. By Theorem 5.1, $\text{commute}_{uv} = 2mr_{uv}$ Clearly the commute time, c_{uv}, from *u* to *v* and back to *u* is less than twice the $\max(h_{uv}, h_{vu})$ which is clearly less than the cover time of G. Putting these facts together

$$mr(G) = mr_{uv} = \tfrac{1}{2}\text{commute}_{uv} \le \max(h_{uv}, h_{vu}) \le \text{cover}(G)$$

For the second inequality note that for any *u* and *v*, $h_{uv} \le 2mr(G)$. By Markov inequality the probability that *v* is not reached in $2mr(G)e^3$ steps is at most $\frac{1}{e^3}$. Thus, the probability that *v* has not been reached in $2mr(G)\log n$ steps is at most $\frac{1}{n^3}$. Summing over *n* vertices, the probability that all vertices are not visited is bounded by $\frac{1}{n^2}$. If one or more vertices have not been reached after *2mr(G)logn* steps, continue walking until all have been reached. By Cor 5.1 this requires at most n^3 additional steps. Hence the expected length of the walk is $2e^3 mr(G)\ln n + n$.

■

Return time

The *return time* is the expected time of a walk starting at *u* returning to *u*.

Random walks in Euclidean space

Many physical processes such as Brownian motion are modeled by random walks. Random walks in Euclidean *d*-space consisting of fixed length steps are really random walks on a *d*-dimensional lattice and are a special case of random walks on graphs. In a random walk on a graph, at each time unit an edge from the current vertex is selected at random and the walk proceeds to the adjacent vertex. We begin by studying random walks on lattices.

Random walks on lattices

We now apply the analogy between random walks and current to lattices. Consider a random walk on a finite segment $-n, \cdots -1, 0, 1, 2, \cdots, n$ of a one dimensional lattice starting from the origin. Is the walk certain to return to the origin or is there some probability that it will escape, i.e., reach the boundary before returning? The probability of reaching the boundary before returning is called the escape probability. We shall be interested in this quantity as *n* goes to infinite.

Convert the lattice to an electrical network by replacing each edge with a one ohm resister. Then the probability of a walk starting at the origin reaching *n* or –*n* before returning to the origin is the escape probability given by

$$p_{escape} = \frac{c_{eff}}{c_a}$$

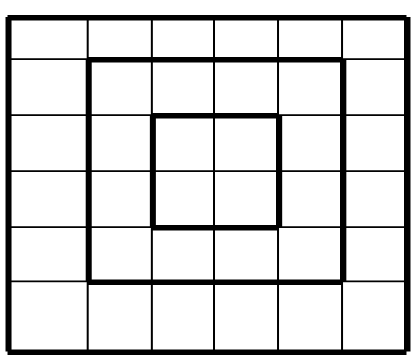
where c_{eff} is the effective conductance between the origin and the boundary points and c_a is the sum of the conductance's at the origin. In a *d*-dimensional lattice $c_a = 2d$ assuming that the resistors are of value one. For the *d*-dimensional lattice

$$p_{escape} = \frac{1}{2dr_{eff}}$$

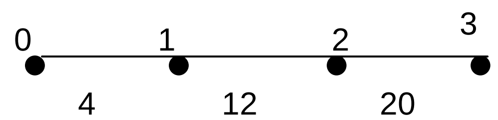
In one dimension, the electrical network is just two series connections of n one ohm resistors connected in parallel. So, r_{eff} goes to infinity and so the escape probability goes to zero as n goes to infinity. So the walk in the unbounded one dimensional lattice will return to the origin with probability one.

Two dimensions.

For the 2-dimensional lattice, consider larger and larger squares about the origin for the boundary and consider the limit of r_{eff} as the squares get larger.



Shorting the resistors on each square can only reduce r_{eff}. Shorting the resistors results in the linear network shown below.



As the paths get longer the number of resistors in parallel also increases. So the resistor between node i and $i+1$ is really made up of $O(i)$ unit resistors in parallel. The effective resistance of $O(i)$ resistors in parallel is $1/O(i)$. So, we have

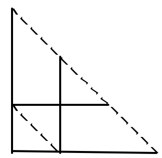
$$r_{eff} \geq \tfrac{1}{4}+\tfrac{1}{12}+\tfrac{1}{20}+\cdots = \tfrac{1}{4}(1+\tfrac{1}{3}+\tfrac{1}{5}+\cdots) = \Theta(\ln n)$$
.

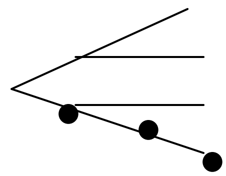
Since the lower bound on the effective resistance goes to infinity, the escape probability goes to zero for the 2 dimensional lattice.

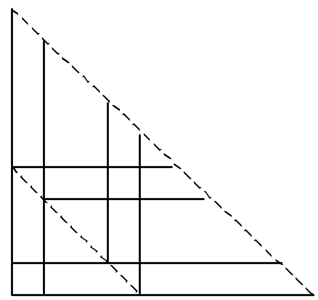
Three dimensions.

In three dimensions, the resistance along any path to infinity grows to infinity but the number of paths in parallel also grows to infinity. It turns out that r_{eff} remains finite and thus there is a non zero escape probability.

The construction used in 3 dimensions is easier to explain first in two dimensions. Draw dotted diagonal lines at $x+y=2^n-1$. Consider two paths that start at the origin. One goes up and the other goes to the right. Each time a path encounters a dotted diagonal line, split the path into two, one which goes right and the other up. Where two paths cross, split the vertex into two keeping the paths separate. By a symmetry argument splitting a vertex does not change the resistance of the network. Remove all resistors except those on these paths. The resistance of the original network is less than that of the tree produced by this process since removing a resistor is equivalent to increasing its resistance to infinity.







The resistance to infinity in this two dimensional example is $\frac{1}{2}+\frac{1}{4}2+\frac{1}{8}4+\cdots=\infty$.
In the analogous three dimensional construction, paths go up, to the right and out of the plane of the paper. They split three ways at planes given by $x+y+z=2^n-1$. Segments of the paths between splits are of length 1, 2, 4, 8, etc and thus the resistance of the segments are equal to the lengths. The resistance out to infinity for the tree is

$$\frac{1}{3}+\frac{1}{9}2+\frac{1}{27}4+\cdots=\frac{1}{3}\left(1+\frac{2}{3}+\frac{4}{9}+\cdots\right)=\frac{1}{3}\frac{1}{1-\frac{2}{3}}=1$$

The resistance of the three dimensional lattice is less. Thus, in three dimensions there is a nonzero probability of escape. The upper bound on r_{eff} gives the lower bound

$$p_{escape}=\frac{1}{2dr_{eff}}\geq\frac{1}{6}.$$

A lower bound on r_{eff} gives an upper bound on p_{escape}. To get the lower bound, short all resistors on surfaces of boxes at distances $1,2,3,\cdots$, etc. Then

$$r_{eff}\geq\frac{1}{6}\left[1+\frac{1}{9}+\frac{1}{25}+\cdots\right]\geq\frac{1.23}{6}\geq 0.2$$

This gives

$$p_{escape}=\frac{1}{2dr_{eff}}\leq\frac{5}{6}$$

.

Exercises

Exercise 5.1:

(a) Give an example of a graph, with cycles of more than one length, for which the greatest common divisor of all cycle lengths is three.
(b) Prove that a graph is bipartite if and only if it has no odd length cycle.
(c) Show that for the random walk on a bipartite graph (with any edge weights), the stationary probabilities do not exist.

Exercise 5.2: (a) What is the set of possible harmonic functions on a graph if there are only interior vertices and no external vertices that supply the boundary condition?
(b) Let q_x be the stationary probability of vertex *x* in a random walk on an undirected graph and let d_x be the degree of vertex *x*. Show that $\frac{q_x}{d_x}$ is a harmonic function.
(c) If there are multiple harmonic functions when there are no boundary conditions why is the stationary probability of a random walk on an undirected graph unique?
(d) What is the stationary probability of a random walk on an undirected graph?

Exercise 5.3: Given a graph consisting of a single path of five vertices numbered 1 to 5, what is the probability of reaching vertex 1 before vertex 5 when starting at vertex 4.

Exercise 5.4: Prove that reducing the value of a resistor in a network cannot increase the effective resistance.

Exercise 5.5: Prove that the escape probability $p_{escape} = \frac{c_{eff}}{c_a}$ must be less than one.

Exercise 5.6: What is probability of returning to start vertex on a random walk on an infinite planar graph?

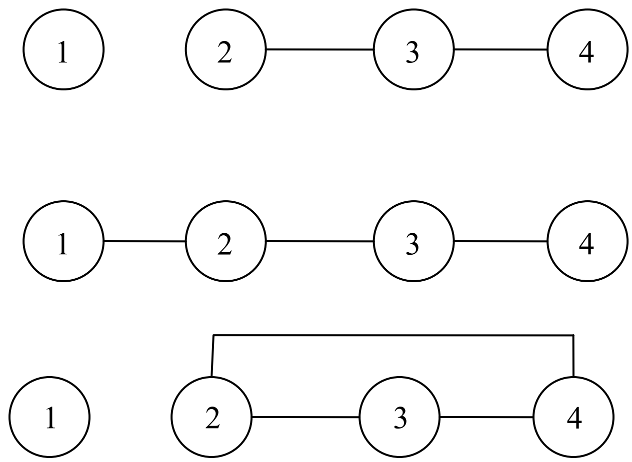
Exercise 5.7: Create a model for a graph similar to a three dimensional lattice in the way that a planar graph is similar to a two dimensional lattice. What is probability of returning to the start vertex in your model?

Exercise 5.8: What is the hitting time h_{uv} for two adjacent vertices on a cycle of length *n*? What is the hitting time if the edge *u*,*v* is removed?

Exercise 5.XXX: In general how do you solve a homogenous difference equation with constant coefficients such as $t(i+1)-2t(i)+t(i-1)=0$? Describe the space of all solutions to a general homogeneous difference equation with constant coefficients.

Exercise 5.9: Consider the set of integers $\{1,2,\cdots,n\}$. How many draws *d* with replacement are necessary so that every integer is drawn?

Exercise 5.10: Show that adding an edge can either increase or decrease hitting time by calculating h_{24} for the following three graphs.



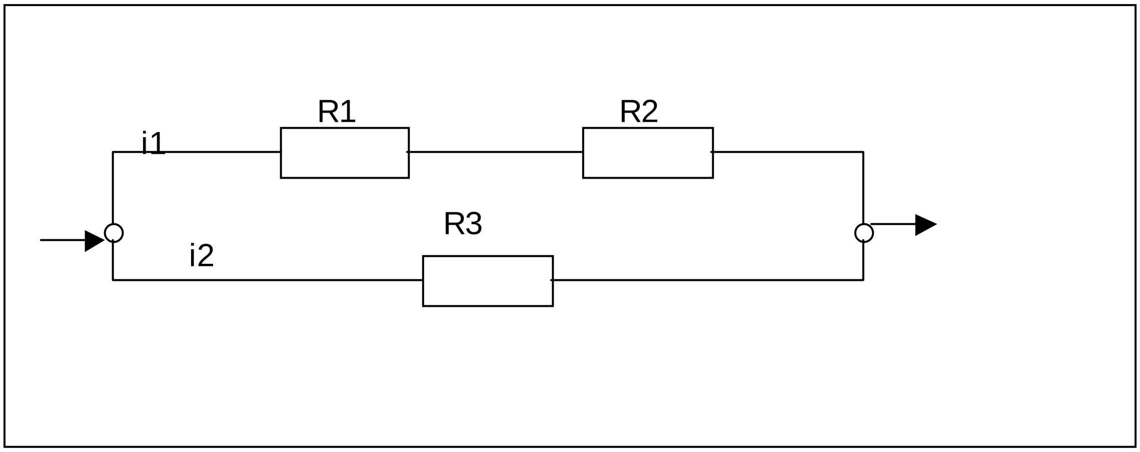
Exercise 5.11: Prove that two independent random walks on a two dimensional lattice will hit with probability one.

Exercise 5.12: Determine by simulation the escape probability for the 3-dimensional lattice.

Exercise 5.13: What is the escape probability for a random walk starting at the root of a binary tree?

Exercise 5.14: Consider a random walk on the positive half line, that is the integers $1,2,3,\cdots$. At the origin, always move right one step. At all other integers move right with probability 2/3 and left with probability 1/3. What is the escape probability?

Exercise 5.15: Consider the electrical resistive network of Figure 5.1 consisting of vertices connected by resistors. Kirchoff's law states that the currents at each node sum to zero. Ohm's law states that the voltage across a resistor equals the product of the resistance times the current through it. Using these laws calculate the effective resistance of the network.



Exercise 5.16: The energy dissipated by the resistance of edge *xy* in an electrical network is given by $i_{xy}^2 r_{xy}$. The total energy dissipation in the network is $E = \frac{1}{2}\sum_{x,y} i_{xy}^2 r_{xy}$ where the $\frac{1}{2}$ accounts for the fact that the dissipation in each edge is counted twice in the summation. Show that the actual current distribution is that distribution satisfying Ohm's law that minimizes energy dissipation.

Exercise 5.17: Prove that increasing any resistance in a network increases the effective resistance.

Notes

When revising read "Link Evolution:Analysis and Algorithms" Chien, et al

EXPLAIN HOW TO DO A RANDOM WALK on the space of all graphs.
Useful if we generate a graph with certain properties and want random graph

Should we add material on
Hidden Markov Models
Detecting bursts
Expanders
Derandomization
Monte Carlo methods

References

Aldous, D.J., "Reversible Markov chains and random walks on graphs, 1994 unpublished

Kemeny, Snell, and Knapp, Denumerable Markov Chains, The University Series in Higher Mathematics, Van Nostrand, Princeton NJ 1966

5.1 Electrical networks and random walks

Doyle and Snell, “Random walks and Electric Networks, The Mathematical Association of America, 1984

Motwani and Raghavan pp127-160

Grinstead and Snell pp471-

Wilf