"Random Walks on Graphs"

Sotiris Nikoletseas
Professor

CEID - ETY Course
2017 - 2018

Random walks on graphs

- Let $G = (V, E)$ a <u>connected, non-bipartite, undirected graph</u> with n vertices. We define a Markov Chain MC_G corresponding to <u>a random walk</u> on the vertices of G, with transition probability:

$$P_{uv} = \begin{cases} \frac{1}{d(u)}, & \text{if } uv \in E(G) \\ 0, & \text{otherwise} \end{cases}$$

where $d(u)$ is the degree of vertex u.

- Since the graph is connected and undirected, MC_G is clearly <u>irreducible</u>. Also, since the graph is non-bipartite, MC_G is <u>aperiodic</u>.

The stationary distribution

- So (from fundamental theorem of Markov Chains) M_G has a unique stationary distribution π.

  <u>Lemma 1:</u> For all vertices $v \in V$ it is $\pi_v = \frac{d(v)}{2m}$, where m is the number of edges of G.

  <u>Proof:</u> From the definition of stationarity, it must be:

  $$\pi_v = [\pi \cdot P]_v = \sum_u \pi_u P_{uv} \quad , \ \forall v \in V$$

  Because of uniqueness, it suffices to verify the claimed solution. Indeed, for all $v \in V$ we have (for the claimed solution value):

  $$\sum_u \pi_u P_{uv} = \sum_{u:uv \in E} \frac{d(u)}{2m} \frac{1}{d(u)} = \frac{1}{2m} \sum_{u:uv \in E} 1 = \frac{1}{2m} d(v) = \pi_v$$

  $\square$

Hitting times / Commute time / Cover time

- <u>Definition:</u> The <u>hitting time h_{uv}</u> is the expected number of steps for random walk starting at vertex u to first reach vertex v.

- <u>Lemma 2:</u> For all vertices $v \in V$, $h_{vv} = \frac{2m}{d(v)}$

  <u>Proof:</u> From fundamental theorem:
  $$h_{vv} = \frac{1}{\pi_v} = \frac{2m}{d(v)} \text{ (from Lemma 1)} \qquad \square$$

- <u>Definition:</u> The <u>commute time between u and v</u> is
  $$CT_{uv} = h_{uv} + h_{vu}$$

- <u>Definition:</u> Let $C_u(G)$ the expected time the walk, <u>starting from u</u>, needs to visit every vertex in G at least once. The <u>cover time</u> of the graph, denoted by $C(G)$, is:
  $$C(G) = \max_u C_u(G)$$

The commute time along an edge

<u>Lemma:</u> For any <u>edge</u> $(u,v) \in E : h_{uv} + h_{vu} \le 2m$

<u>Proof:</u> Consider a new Markov Chain with states the edges of the graph (every edge taken twice as two directed edges), where the transitions occur between adjacent edges. The number of states is clearly $2m$ and the current state is the last (directed) edge visited. The transition matrix is

$$Q_{(u,v)(v,w)} = \frac{1}{d(v)}$$

This matrix is clearly <u>doubly stochastic</u> since not only the rows but also the columns add to 1. Indeed:

$$\sum_{x \in V, y \in \Gamma(x)} Q_{(x,y)(v,w)} = \sum_{u \in \Gamma(v)} Q_{(u,v)(v,w)} = \sum_{u \in \Gamma(v)} \frac{1}{d(v)} = d(v)\frac{1}{d(v)}$$

Proof (continued)

So the stationary distribution is <u>uniform</u>. So if $e = (u, v)$ any edge, then $\pi_e = \frac{1}{2m}$ and $h_{ee} = \frac{1}{\pi_e} = 2m$. In other words, the expected time between successive traversals of edge e is $2m$.

Consider now $h_{uv} + h_{vu}$. This is the expected time to go from u to v and then return back to u. Conditioning on the event that we initially arrived to u from v, then $Q_{(v,u)(v,u)}$ is the time between two successive passages over the edge vu and is an upper bound to the time to go from u to v and back.

But this time is at most $2m$ in expectation. Since the MC is memoryless, we can remove the arrival conditioning and the result holds independently of the vertex we initially arrive to u from.

$\square$

# Electrical networks and random walks

- A <u>resistive electrical network</u> can be seen as an undirected graph. Each edge of the graph is associated to a branch resistance. The electrical flow in the network is governed by two laws:
  - Kirchoff's law for preservation of flow (e.g. all flow that enters a node, leaves it).
  - Ohm's law: the voltage across a resistor equals the product of the resistance times the current through it).

- The <u>effective resistance</u> R_{uv} between nodes u and v is the voltage difference between u and v when current of one ampere is injected into u and removed from v (or injected at v and removed from u).(Thus, the effective resistance is upper bound by the branch resistance but it can be much smaller).

- Given an undirected graph G, let $N(G)$ the electrical network defined over G, associating 1 Ohm resistance to each of the edges.

Commute time and effective resistance

<u>Lemma:</u> For any two vertices u, v in G, the commute time between them is: $CT_{uv} = 2m \cdot R_{uv}$, where m is the number of edges of the graph and R_{uv} the effective resistance between u and v in the associated electrical network $N(G)$.

<u>Proof:</u> Let Φ_{uv} the voltage at u in $N(G)$ with respect to v, where $d(x)$ amperes (degree of x) of current are injected to each node $x \in V$ and all $2m = \sum_x d(x)$ amperes are removed from v. It is:

$$h_{uv} = \Phi_{uv} \qquad (1)$$

Scenario A
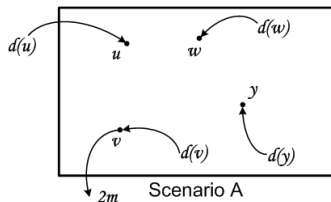
Proof (continued)

Indeed, the voltage difference on the edge uw is $\Phi_{uw} = \Phi_{uv} - \Phi_{wv}$. Using the two laws we get, for all $u \in V - \{u\}$ that:

$$d(u) \overset{K}{=} \sum_{w \in \Gamma(u)} \text{current}(uw) \overset{O}{=} \sum_{w \in \Gamma(u)} \frac{\Phi_{uw}}{\text{resistance}(uw)}$$

$$= \sum_{w \in \Gamma(u)} (\Phi_{uv} - \Phi_{wv}) = d(u) \cdot \Phi_{uv} - \sum_{w \in \Gamma(u)} \Phi_{wv}$$

$$\Rightarrow \Phi_{uv} = 1 + \frac{1}{d(u)} \sum_{w \in \Gamma(u)} \Phi_{wv} \tag{2}$$

Proof (continued)

- On the other hand, from the definition of expectation we have , for all $u \in V - \{v\}$, that:

$$h_{uv} = 1 + \frac{1}{d(u)} \sum_{w \in \Gamma(u)} h_{wv} \qquad (3)$$

Equations (2) and (3) are actually linear systems, with unique solutions (system (2) refers to voltage differences, which are uniquely determined by the current flows). Furthermore, if we identify Φ_{uv} in (2) with h_{uv} in (3), the two systems are identical. This proves that $h_{uv} = \Phi_{uv}$ indeed (as in (1).

- Now note that h_{uv} is the voltage Φ_{uv} at v in $N(G)$ measured w.r.t. u, when currents are injected into all nodes and removed from all other nodes.

Proof (continued)

- Let us now consider a Scenario B, which is like Scenario A except that we remove the $2m$ current units from node u instead of node v.

Scenario B   Scenario C   Scenario D
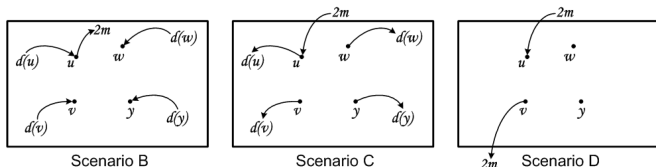
- Denoting the voltage differences in Scenario B by Φ', we have (as in (1)) that
$$\Phi'_{vu} = h_{vu}$$

- Now let us consider a Scenario C, which is like B but with all currents reversed. Denoting the voltage differences in this scenario by Φ'', we have:
$$\Phi''_{uv} = -\Phi'_{uv} = \Phi'_{vu} = h_{vu}$$

Proof (continued)

- Finally, consider a Scenario D, which is just the sum of Scenarios A and C. Denoting Φ''' the voltage differences in D and since the currents (except the $2m$ ones at u, v) cancel out , we have

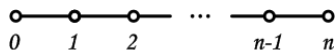$$\Phi'''_{uv} = \Phi_{uv} + \Phi''_{uv} = h_{uv} + h_{vu}$$

But in D, Φ'''_{uv} is the voltage difference between u and v when pushing $2m$ amperes at u and removing them at v, so (by definition of the effective resistance and Ohm's law) we have

$$\Phi'''_{uv} = 2m \cdot R_{uv}$$

$\square$

Examples (I)

The line graph. Consider $n + 1$ points on a line:

$$\begin{array}{ccccccc} \circ & \circ & \circ & \cdots & \circ & \circ \\ 0 & 1 & 2 & & n\text{-}1 & n \end{array}$$

By symmetry, it is $h_{0n} = h_{n0}$. Also (since the effective resistance between 0 and n is clearly n), we have:
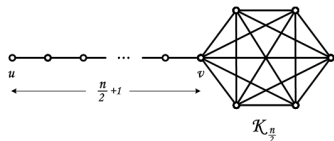
$$h_{0n} + h_{n0} = C_{0n} = 2m \cdot R_{0n} = 2 \cdot n \cdot n = 2n^2, \text{ thus}$$

$$h_{0n} = h_{n0} = n^2$$

We see that in this case the hitting times are symmetric. This is not the case in general.

Examples (II)

The lollipop graph, composed of a line of $\frac{n}{2} + 1$ vertices joined to a $K_{\frac{n}{2}}$ clique, as in the following figure:

Let u and v the endpoints of the line. We have:

$$h_{uv} + h_{vu} = C_{uv} = 2 \cdot m R_{uv} = 2\Theta(n^2) \cdot \Theta(n) = \Theta(n^3)$$

But from line example in the previous slide

$$h_{uv} = \Theta(n^2) \text{ thus } h_{vu} = \Theta(n^3)$$

This asymmetry is due to the fact that, when we start from u, the walk has no option but to go towards v; but when we start from v there is very little probability, i.e. $\Theta\left(\frac{1}{n}\right)$, of proceeding to the line.
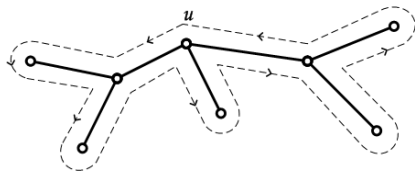
The cover time

We will now give <u>bounds on the cover time</u>. The first one is rather loose since it is independent of the structure of the graph and only takes into account the number of edges:

<u>Theorem.</u> For any connected graph $G(V, E)$, the cover time is:
$$C(G) \leq 2|E||V| = 2 \cdot m \cdot n$$

<u>Proof.</u> 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:

The cover time

Clearly, the cover time from vertex u is upper bounded by the expected time for the walk to visit the vertices of G in this order. Let $u = v_0, v_1, \ldots, v_{2n-2} = u$ denote the visited vertices in such a traversal. Then

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

By the previous lemma on the commute time, we have

$$C(G) = \max_{u \in V} C(u) \leq \sum_{(x,y) \in T} (h_{xy} + h_{yx}) = 2m \sum_{(x,y) \in T} R_{xy}$$
$$\leq 2 \cdot m \cdot n$$

since for any two <u>adjacent</u> vertices x, y the effective resistance is at most $R_{xy} \leq 1$.

(alternatively we can use a previous Lemma stating that the commute time along an <u>edge</u> is at most $2m$, and the tree has $n - 1$ edges).

Examples

1. The <u>line graph</u>. It has $n + 1$ vertices and $m = n$ edges so
   $$C(G) \leq 2 \cdot n(n+1) \simeq 2n^2$$
   Also, we know that $C(G) \geq H_{0n} = n^2$, thus the bound is tight (up to constants) in this case.

2. The <u>lollipop graph</u>. We get $C(G) \leq 2 \cdot \Theta(n^2) \cdot n = \Theta(n^3)$. Again $C(G) \geq H_{vu} = \Theta(n^3)$ so the bound is tight.

3. The <u>complete graph</u>. We set $C(G) \leq 2 \cdot \Theta(n^2) \cdot n = \Theta(n^3)$. But from coupon collectors, the cover time is actually $C(h) = (1 + o(1))n \ln n$, thus it is much smaller than the upper bound.

<u>Comment:</u> This shows a rather counter-intuitive property of cover times (and hitting times): they are <u>not monotonic</u> w.r.t. adding edges to the graph!

A stronger bound

<u>Theorem</u>(proof in the book). Let the resistance of a graph G be $R = \max\limits_{u,v \in V} R_{u,v}$. For a connected graph G its cover time is:

$$m \cdot R \leq C(G) \leq c \cdot m \cdot R \cdot \log n \qquad \text{for some constant } c.$$

<u>Examples:</u>
a) In the <u>complete</u> graph, the probability of hitting a given vertex v, when starting at any vertex u, is $\frac{1}{n-1}$ so, $\forall u, v \in V,\ h_{uv} = n - 1$. Also, we have $h_{uv} + h_{vu} = 2mR_{uv} \Rightarrow 2(n-1) = 2\frac{n(n-1)}{2}R_{uv} \Rightarrow R_{uv} = \frac{2}{n}$ so we get $C(G) \leq c\frac{n(n-1)}{2}\frac{2}{n}\log n = O(n \log n)$, which is tight up to constants.

b) In the <u>lollipop</u> graph, $R = \Theta(n)$ and $m = \Theta(n^2)$, so the upper bound we get is $C(G) \leq O(n^3 \log n)$ which is worse (by a logarithmic factor) from the looser bound.