# Proof of a graph-theoretic lemma

December 12, 2007

**Question:** Let $G$ be a simple undirected $d$-regular graph, and let $H$ be a subset of the vertices of $G$. Consider a random walk starting from a vertex $s$ in $H$. At each step, the walk stays at the current vertex with probability $\frac{1}{2}$ and it moves to each neighbor of the current vertex with probability $\frac{1}{2d}$. If the starting vertex $s$ is chosen uniformly at random from $H$, what is the expected probability that the walk leaves $H$ immediately and takes more than $\ell$ steps to return to a vertex in $H$?

**Answer:** At most $\frac{|G|}{|H|}\frac{1}{\ell}$.

I think the lemma and its proof is interesting for a couple of reasons. First, the upper-bound holds unconditionally for *any* simple undirected $d$-regular graph and any $\ell > 0$. Second, the proof makes elegant use of the stationary distribution of the walk, while in the question, there is no mention of it.

Here is a rough sketch of the proof. As mentioned above, we look at the stationary distribution of the random walk on $G$. By the way the walk is defined, the stationary distribution assigns probability $\frac{1}{|G|}$ to every vertex. Now, we ask the following question: for a given $u$ and $v$ in $H$ and for a given $t > \ell$, what is the probability under the stationary distribution of being on a vertex that lies on a path of length[1] $t$ between $u$ and $v$ that passes only through vertices not in $H$? This probability equals $t$ times the probability of entering such a path of length $t$ from $u$ to $v$ (after the walk has totally mixed). The latter probability is the probability of being at $u$, that is $1/|G|$, times the probability of a walk outside $H$ of length $t$ that starts at $u$ and ends at $v$. Summing over all $u$ and $v$ in $H$ and all $t > \ell$, we find that the probability of walks staying outside $H$ and longer than $\ell$ between some two vertices in $H$ is at most $|G|/\ell$ times the probability under the stationary distribution of being on some vertex that lies on a path of at least $\ell$ between two vertices in $H$. The latter probability, we argue, is at most 1, and that proves our claim. Details follow.

*Proof.* For convenience, let us suppose that we add $d$ self-loops to each vertex in $G$. Then, a random walk on $G$ consists of uniformly choosing one of the $2d$ edges incident to the current vertex at each step. Thus, there are exactly $(2d)^t$ distinct paths of length $t$ starting from a particular vertex.

To model the types of walks that we need to analyze, let us define a new Markov chain $M$. The states of $M$ consist of $H \cup \{(h_{u,v,t}, i) | h_{u,v,t}$ is a path in $G$ of length $t \geq 2$ between vertices $u$ and $v$ in $H$ in which all the vertices other than $u$ and $v$ are in $G - H$, $1 \leq i < t\}$. Let $P$ be the transition matrix for Markov chain $M$. $P$ is defined as follows:

- If $u \in H$, then $P_{u,u} = \frac{1}{2}$
- If $u, v \in H$, $(u, v) \in E(G)$ and $u \neq v$, then $P_{u,v} = \frac{1}{2d}$
- If $h_{u,v,t}$ is a path between two vertices $u$ and $v$ in $H$ of length $t$ that only passes through vertices not in $H$, then $P_{u,(h_{u,v,t},1)} = \frac{1}{(2d)^t}$, $P_{(h_{u,v,t},i-1),(h_{u,v,t},i)} = 1$ for $1 < i < t$ and $P_{(h_{u,v,t},t-1),v} = 1$.
- All unspecified transition probabilities are 0

These probabilities are defined so that for any two vertices $u$ and $v$ in $H$, the probability of taking a walk in $M$ from $u$ to $v$ is exactly the same as the probability of taking a walk in $G$ from $u$ to $v$.

Let $\pi_M$ be the stationary distribution of $M$. From the above observation, it follows that for all $u \in H$, $\pi_M(u) = \frac{1}{|G|}$, the same probability that a random walk in $G$ assigns to $u$. Also, note that for a given $u$ and $v$ in $H$:

$$\sum_{h_{u,v,t}} \pi_M((h_{u,v,t}, 1)) = \pi_M(u) p_{u,v,t} = \frac{p_{u,v,t}}{|G|} \tag{1}$$

where $p_{u,v,t}$ is the probability of a random walk in $G$ of length $t$ that starts from $u$ and ends at $v$ without passing through any other vertex in $H$. Also, for each of the paths $h_{u,v,t}$, it is true that $\pi_M((h_{u,v,t}, 1)) = \pi_M((h_{u,v,t}, i))$ for all $1 \leq i < t$ because of how the transition probabilities of $M$ are defined. Therefore:

$$\frac{p_{u,v,t}}{|G|} = \sum_{h_{u,v,t}} \pi_M((h_{u,v,t}, 1)) = \frac{1}{t-1} \sum_{1 \leq i < t} \sum_{h_{u,v,t}} \pi_M((h_{u,v,t}, i)) \tag{2}$$

Summing over all $t > \ell$,

$$\sum_{t>\ell} p_{u,v,t} = |G| \sum_{t>\ell} \frac{1}{t-1} \sum_{1 \leq i < t} \sum_{h_{u,v,t}} \pi_M((h_{u,v,t}, i)) \leq \frac{|G|}{\ell} \sum_{1 \leq i < t} \sum_{h_{u,v,t}} \pi_M((h_{u,v,t}, i)) \tag{3}$$

[1]By the length of a walk, we mean the number of steps taken, not the number of edges in $G$ that are followed.

Finally, summing over all $u$ and $v$ in $H$:

$$\sum_{u,v\in H}\sum_{t>\ell} p_{u,v,t} \leq \frac{|G|}{\ell}\sum_{u,v\in H}\sum_{1\leq i<t}\sum_{h_{u,v,t}} \pi_M((h_{u,v,t},i)) \leq \frac{|G|}{\ell}\cdot 1 \tag{4}$$

The last inequality holds because all the $(h_{u,v,t},i)$ states in the sum are disjoint.

Therefore, if $u$ is chosen uniformly at random from $H$ as the starting vertex of a random walk in $G$, the expected probability that it takes more than $\ell$ steps to return to a vertex in $H$ is at most $\frac{\sum_{u\in H}\sum_{v\in H,t>\ell} p_{u,v,t}}{|H|} \leq \frac{|G|}{|H|}\frac{1}{\ell}$.

□

This lemma and its proof is taken from Lemma 4.2 in [GR99].

# References


[GR99] Oded Goldreich and Dana Ron. A sublinear bipartitenesss tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.