Notes on Mihail's combinatorial analysis of expanders

Arnab Bhattacharyya

December 12, 2007

Mihail's *Conductance and Convergence of Markov Chains – A Combinatorial Treatment of Expanders* [Mih89] describes purely combinatorially why expanders are rapidly mixing. The following will basically be a rewriting of Section 2 of the paper in my words. I have also used the notes in [Bol] as reference.

Let G be a d-regular directed graph with n vertices. That is, for every vertex, both the in-degree and the out-degree are exactly equal to d. We will denote by $\Gamma(u)$ the set $\{v|(u,v) \in E(G)\}$. A random walk on G is defined by the transition matrix P where:

$$P_{i,j} = \begin{cases} \frac{1}{2} & \text{if } i = j \\ \frac{1}{2d} & \text{if } i \neq j \text{ and } (i,j) \in E(G) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The above definition of P makes sure that the stationary distribution of the walk is the uniform distribution. This is so since $\pi \cdot \frac{1}{2} + \pi \cdot d \cdot \frac{1}{2d} = \pi$. As a side remark, if we only required that the in-degree and the out-degree of any vertex in G be at most d (instead of exactly equal to d), then P would be defined as:

$$P_{i,j} = \begin{cases} 1 - \frac{|\Gamma^{-1}(i)|}{2d} & \text{if } i = j \\ \frac{1}{2d} & \text{if } i \neq j \text{ and } (i,j) \in E(G) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

in order to maintain the property that the stationary distribution has uniform weight on all vertices. This definition is used by Goldreich and Ron in their paper on testing bipartiteness on bounded-degree graphs. However, in the following, I will continue to assume for convenience that G is exactly d-regular.

Our goal is to show that if G has a certain combinatorial property, namely no small cuts, then the random walk rapidly approaches the stationary distribution. Formally, suppose π_0 is the distribution that has all its weight on a single vertex, and $\pi_t = \pi_{t-1}P$ for $t > 0$. So, π_t is the distribution of the random walk on the vertices at time t. Let u be the uniform distribution on the vertices. Finally, let $e_t = \pi_t - u$ be the error vector. Note that $\sum_i e_t(i) = 0$ and that $e_t = \pi_t - u = (\pi_{t-1} - u)P = e_{t-1}P$. We want to show that e_t approaches the zero vector and the rapidity of the approach is related to the lack of small cuts in G.

To characterize e_t's distance from 0, we are going to look at the L_2^2 norm of e_t. Let us see how $\|e_t\|_2^2$ evolves with t:

$$\begin{aligned} \|e_{t+1}\|_2^2 = \sum_i (e_{t+1}(i))^2 &= \sum_i \left(\frac{e_t(i)}{2} + \sum_{j \in \Gamma^{-1}(i)} \frac{e_t(j)}{2d} \right)^2 \\ &= \sum_i \left(\sum_{(i,j) \in E(G)} \frac{e_t(i) + e_t(j)}{2d} \right)^2 \\ &= \sum_i \left(\mathop{\mathbb{E}}_{j \in \Gamma^{-1}(i)} \left[\frac{e_t(i) + e_t(j)}{2} \right] \right)^2 \end{aligned}$$

$$\leq \sum_i \mathop{\mathbb{E}}_{j \in \Gamma^{-1}(i)} \left[\frac{(e_t(i) + e_t(j))^2}{4} \right]$$

$$= \sum_{(i,j) \in E(G)} \frac{(e_t(i) + e_t(j))^2}{4d} \tag{3}$$

So,

$$\begin{aligned} \|e_t\|_2^2 - \|e_{t+1}\|_2^2 &\geq \sum_i (e_t(i))^2 - \sum_{(i,j) \in E(G)} \frac{(e_t(i) + e_t(j))^2}{4d} \\ &= \sum_{(i,j) \in E(G)} \frac{(e_t(i))^2 + (e_t(j))^2}{2d} - \sum_{(i,j) \in E(G)} \frac{(e_t(i) + e_t(j))^2}{4d} \\ &= \sum_{(i,j) \in E(G)} \frac{(e_t(i) - e_t(j))^2}{4d} \end{aligned} \tag{4}$$

Since the last expression is always positive, the L_2^2-norm of e_t continually decreases. Moreover, the decrease is greater if there are many neighbors i and j such that $e_t(i)$ and $e_t(j)$ are very different.

Next, we want to show that if G has no small cuts, then there are a lot of such neighbors. Specifically, we define the notion of *conductance*:

$$\alpha_G = \min_{A \subseteq V(G), |A| \leq n/2} \frac{|e(A, \bar{A})|}{d|A|} \tag{5}$$

where $e(A, \bar{A})$ is the set of edges $\{(i,j) \in E(G) | i \in A, j \notin A\}$. A small α_G implies that there is a small cut of G. α_G is, thus, sort of like the minimal ratio between the surface area and the volume of a subgraph of G. Our goal is to show that the $\sum_{(i,j) \in E(G)} \frac{(e_t(i) - e_t(j))^2}{4d}$ of equation (4) is positively correlated with α_G, where the only fact we know about e_t is that $\sum_i e_t(i) = 0$.

For convenience, let us label the vertices such that $e_t(1) \geq e_t(2) \geq \cdots \geq e_t(n)$. Also, define a new vector f_t to be $f_t(i) = e_t(i) - e_t(m)$ for all $i \in V(G)$, where m is set to $\lceil n/2 \rceil$. It is clear that $\sum_{(i,j) \in E(G)} (f_t(i) - f_t(j))^2 = \sum_{(i,j) \in E(G)} (e_t(i) - e_t(j))^2$. For technical reasons that will become apparent momentarily, we will work with the translated vector f_t instead of e_t. Note that $f_t(i) \geq 0$ for $i \leq m$, $f_t(m) = 0$, and $f_t(i) \leq 0$ for $i > m$. We can now intuitively understand why $\sum_{(i,j) \in E(G)} (f_t(i) - f_t(j))^2$ should increase with α_G. Suppose that for some $k < m$, $f_t(k) - f_t(k+1)$ is relatively large. If we let A_k be the set of vertices $\{1, 2, \ldots, k\}$, by definition of α_G and by the fact that $k \leq n/2$, we know that there are at least $d\alpha_G k$ edges in $e(A_k, \bar{A}_k)$. That is, there are at least $d\alpha_G k$ edges (i,j) such that $i \leq k$ and $j > k$. And because of the way we have ordered the vertices, for each such edge (i,j), $f_t(i) - f_t(j)$ is at least $f_t(k) - f_t(k+1)$; so, there are a substantial number of edges with relatively large difference in f_t between their endpoints. This observation translates to a lower bound for $\sum_{(i,j) \in E(G)} (f_t(i) - f_t(j))^2$.

We implement this intuition in the following way. Define the vector u_t as follows:

$$u_t(i) = \begin{cases} f_t(i) & \text{if } 1 \leq i \leq m \\ 0 & \text{if } i > m \end{cases} \tag{6}$$

and v_t as:

$$v_t(i) = \begin{cases} 0 & \text{if } 1 \leq i \leq m \\ f_t(i) & \text{if } i > m \end{cases} \tag{7}$$

So, $f_t = u_t + v_t$. Now note that for any edge (i,j), $(f_t(i) - f_t(j))^2 = (u_t(i) - u_t(j) + v_t(i) - v_t(j))^2 \geq (u_t(i) - u_t(j))^2 + (v_t(i) - v_t(j))^2$. Our plan will be to lower-bound $\sum_{(i,j)} (u_t(i) - u_t(j))^2$ and $\sum_{(i,j)} (v_t(i) - v_t(j))^2$ in terms of α_G, using the above intuition as our guide.

Let us start with $\sum_{(i,j)}(u_t(i)-u_t(j))^2$. Observe that by the Cauchy-Schwarz inequality:

$$\left(\sum_{(i,j)\in E(G)} |u_t(i)^2-u_t(j)^2|\right)^2 = \left(\sum_{(i,j)\in E(G)} (|u_t(i)-u_t(j)|)(u_t(i)+u_t(j))\right)^2$$
$$\leq \sum_{(i,j)\in E(G)} (u_t(i)-u_t(j))^2 \sum_{(i,j)\in E(G)} (u_t(i)+u_t(j))^2$$
$$\leq \sum_{(i,j)\in E(G)} (u_t(i)-u_t(j))^2 \cdot 2 \sum_{(i,j)\in E(G)} (u_t(i)^2+u_t(j)^2)$$
$$= 4d\cdot \|u_t\|_2^2 \cdot \sum_{(i,j)\in E(G)} (u_t(i)-u_t(j))^2 \tag{8}$$

so that:

$$\sum_{(i,j)\in E(G)} (u_t(i)-u_t(j))^2 \geq \frac{\left(\sum_{(i,j)\in E(G)} |u_t(i)^2-u_t(j)^2|\right)^2}{4d\|u_t\|_2^2} \tag{9}$$

The rest roughly follows our plan.

$$\sum_{(i,j)\in E(G)} |u_t(i)^2-u_t(j)^2|$$
$$= \sum_{(i,j)\in E(G);i<j} (u_t(i)^2-u_t(j)^2) + \sum_{(j,i)\in E(G);i<j} (u_t(i)^2-u_t(j)^2)$$
$$= \sum_{(i,j)\in E(G);i<j} \sum_{k=i}^{j-1}(u_t(k)^2-u_t(k+1)^2) + \sum_{(j,i)\in E(G);i<j} \sum_{k=i}^{j-1}(u_t(k)^2-u_t(k+1)^2)$$
$$= \sum_{k=1}^{n-1}(u_t(k)^2-u_t(k+1)^2)(e(A_k,\bar{A}_k)+e(\bar{A}_k,A_k))$$
$$= 2\sum_{k=1}^{n-1}(u_t(k)^2-u_t(k+1)^2)e(A_k,\bar{A}_k) \tag{10}$$

where the last line follows by d-regularity of G. Using the fact that $u_t(i)=0$ for $i\geq m$, (10) implies that

$$\sum_{(i,j)\in E(G)} |u_t(i)^2-u_t(j)^2| = 2\sum_{k=1}^{m-1}(u_t(k)^2-u_t(k+1)^2)e(A_k,\bar{A}_k)$$
$$\geq 2\sum_{k=1}^{m-1}(u_t(k)^2-u_t(k+1)^2)d\alpha_G k$$
$$= 2d\alpha_G \sum_{k=1}^{m-1} u_t(k)^2 = 2d\alpha_G\|u_t\|_2^2 \tag{11}$$

So, by the inequality (9), we have that $\sum_{(i,j)\in E(G)}(u_t(i)-u_t(j))^2 \geq d\alpha_G^2\|u_t\|_2^2$. Similarly, it is also true that $\sum_{(i,j)\in E(G)}(v_t(i)-v_t(j))^2 \geq d\alpha_G^2\|v_t\|_2^2$. Adding the two inequalities, we get that:

$$\sum_{(i,j)\in E(G)} (u_t(i)-u_t(j))^2 + (v_t(i)-v_t(j))^2 \geq d\alpha_G^2(\|u_t\|_2^2+\|v_t\|_2^2) = d\alpha_G^2\|f_t\|_2^2 \tag{12}$$

But for any edge (i,j), $(f_t(i)-f_t(j))^2 = (u_t(i)-u_t(j)+v_t(i)-v_t(j))^2 \geq (u_t(i)-u_t(j))^2+(v_t(i)-v_t(j))^2$. So, in fact, $\sum_{(i,j)\in E(G)}(f_t(i)-f_t(j))^2 \geq d\alpha_G^2\|f_t\|_2^2$. Finally, to relate back to e_t, we note that for any edge (i,j),

$(f_t(i)-f_t(j))^2=(e_t(i)-e_t(j))^2$ and that $\|f_t\|_2^2=\sum_i(e_t(i)-e_t(m))^2=\|e_t\|_2^2-2e_t(m)\sum_i e_t(i)+ne_t(m)^2=\|e_t\|_2^2+ne_t(m)^2\geq\|e_t\|_2^2$ since $\sum_i e_t(i)=0$. So, we have that $\sum_{(i,j)\in E(G)}(e_t(i)-e_t(j))^2\geq d\alpha_G^2\|e_t\|_2^2$.

Therefore, going back to equation (4):

$$\|e_{t+1}\|_2^2\leq\left(1-\frac{\alpha_G^2}{4}\right)\|e_t\|_2^2 \tag{13}$$

In particular, since $\|e_0\|_2^2\leq 2$,

$$\|e_t\|_2^2\leq 2\left(1-\frac{\alpha_G^2}{4}\right)^t \tag{14}$$

Also, since $|e_t|_1\leq\sqrt{n}\|e_t\|_2$, we have that $|e_t|_1\leq\sqrt{2n}\left(1-\frac{\alpha_G^2}{4}\right)^{t/2}$. So, if we want the L_1 difference between the random walk distribution and the uniform distribution to be smaller than ϵ, the random walk needs to run for only $O(\frac{1}{\alpha_G^2}(\log\frac{1}{\epsilon}+\frac{1}{2}\log 2n))$ steps. So, if G is an expander, meaning that α_G is a constant bounded away from 0, the random walk on G rapidly mixes. (The technical definition of rapid mixing is that the mixing time is $\text{polylog}(n)\log\frac{1}{\epsilon}$.) Of course, expansion in the sense defined here coincides with the linear algebra definition of expansion.

References

[Bol] Béla Bollobás. Volume estimates and rapid mixing. In Silvio Levy, editor, *Flavors of Geometry*, volume 31, pages 151–194. MSRI Publications.

[Mih89] Milena Mihail. Conductance and convergence of markov chains – a combinatorial treatment of expanders. In *FOCS*, pages 526–531, 1989.