University of Michigan–Ann Arbor

Department of Electrical Engineering and Computer Science

EECS 498 004 **Advanced Graph Algorithms**, Fall 2021

Lecture 11: Expander: Probabilistic view

October 5, 2021

Instructor: Thatchaphol Saranurak

Scribe: Aditya Anand

1 A probabilistic view of expanders

In this lecture, we shall study yet another characterization of expanders - **expanders are graphs on which random walks mix rapidly.** We shall see that there is a direct connection between the second eigenvalue of the normalized laplacian and the mixing time of a random walk on graphs.

2 Random Walks

Let G be a graph with degree profile $\boldsymbol{d}$. We define a random walk step starting from a given vertex as follows.

Definition 2.1. (Random Walk Step) Given an undirected unweighted graph G and a vertex u, the result of one step of a random walk from u is a uniformly random neighbour of u in G.

The same definition can naturally be extended to weighted graphs.

Definition 2.2. (Random Walk Step in Weighted Graphs) Given an undirected weighted graph G with weight function $w : V(G) \to \mathbb{R}$ and a vertex u, the result of one step of a random walk from u is the random vertex obtained by sampling a single vertex from $N_G(u)$ where $x \in N_G(u)$ is sampled with probability $\frac{w(u,x)}{\sum_{v \in N_G(u)} w(u,v)}$.

Today's lecture is motivated by the following natural questions, which can be better understood with an example.

Example 2.3. Given a cycle of length n, if we start at some vertex u and do a random walk for a sufficiently large number of steps, what is the distribution of vertices we end up with?

Intuitively, we understand that the answer is the uniform distribution. But can we prove this? We can also ask how many steps does it take so that $\Pr[\text{end at } u] = \frac{1 \pm \epsilon}{n}$ for all u - that is, how fast do we converge to the uniform distribution?

For a general graph, these questions become

- Do random walks converge?

Figure 1: Random walk illustration

- If so, to what distribution does the random walk converge?
- What is the rate of convergence - how fast does the random walk "mix" or converge to the final distribution?

The goal of this lecture will be to answer these questions, and understand these in the context of expanders - we will show that expanders have small mixing times.

3 Walk Matrices

Let us now formalize the notion of a random walk. Let $\boldsymbol{p}_t \in \mathbb{R}^V$ denote the probability distribution over vertices after time t. Suppose initially we start at a vertex u, so $\boldsymbol{p}_0 = \mathbf{1}_u$. Then $\boldsymbol{p}_1 = \frac{1}{\deg(u)} \cdot \mathbf{1}_{N(u)}$ where $N(u)$ is the set of neighbors of u. It is clear that

$$\boldsymbol{p}_{t+1}(u) = \sum_{(v,u)\in E} \frac{w(u,v)}{d(v)} \boldsymbol{p}_t(v).$$

Definition 3.1. (Walk matrix) The walk matrix $\boldsymbol{W}$ is the matrix $\boldsymbol{W}$ that satisfies $\boldsymbol{p}_{t+1} = \boldsymbol{W}\boldsymbol{p}_t$. From the above expression, we can see that

$$\boldsymbol{W} = \boldsymbol{A}\boldsymbol{D}^{-1}.$$

After t random walk steps, we have

$$\boldsymbol{p}_t = \boldsymbol{W}\boldsymbol{p}_{t-1} = \boldsymbol{W}^t \boldsymbol{p}_0.$$

We can now formalize the questions from the previous section as follows.

- Convergence - What is the limit of $\boldsymbol{p}_t$ when $t \to \infty$ (if it exists)?
- Rate of convergence - How fast does $\boldsymbol{p}_t$ converge to that limit?

Unfortunately, it turns out that the above limit may not exist for a general graph. In fact, consider a trivial graph consisting of only one edge (u, v). If we do a random walk starting from u, then at every odd step we are at v with probability 1 and at every even step we are at u with probability 1 - from which it is clear that the limit $\boldsymbol{p}_t$ when $t \to \infty$ does not exist.

In fact this limit does not exist for any bipartite or disconnected graph[1] - at every step the total probability mass on one partite set is zero. This suggests that we need a different notion of random walk to ensure convergence for every graph. Does there exist a natural random walk that converges for any graph? The answer is indeed yes, as we shall see in the subsequent section.

4 Lazy Random Walk

We define a lazy random walk step as follows.

Definition 4.1. (Lazy Random Walk Step) Given an undirected unweighted graph G and a vertex u, the result of one step of a lazy random walk is u with probability $\frac{1}{2}$, otherwise it is the result of a random walk step with u.

Let $\boldsymbol{p}_t$ be the probability mass vector after t lazy random walk steps. We now define a new walk matrix $\widetilde{\boldsymbol{W}}$ corresponding to the lazy random walk, so that we have

$$\boldsymbol{p}_{t+1} = \widetilde{\boldsymbol{W}} \boldsymbol{p}_t,$$

What is the matrix W? Since we stay put with probability $\frac{1}{2}$ and do a random walk with probability $\frac{1}{2}$, it follows that

$$\begin{aligned} \widetilde{\boldsymbol{W}} &= \boldsymbol{I}/2 + \boldsymbol{W}/2 \\ &= \boldsymbol{I}/2 + \boldsymbol{A}\boldsymbol{D}^{-1}/2 \end{aligned}$$

Let us now return to our trivial example graph with one edge (u, v) and ask the question if the limit of $\boldsymbol{p}_t = \widetilde{\boldsymbol{W}}^t \boldsymbol{p}_0$ when $t \to \infty$ exists. It is clear that $p_t(u) = p_t(v) = \frac{p_{t-1}(v)+p_{t-1}(u)}{2}$, and hence the lazy random walk converges to $p_t(u) = p_t(v) = \frac{1}{2}$ in just one step.

5 The Stable Distribution

Theorem 5.1. *In a connected undirected graph G, for any $\boldsymbol{p}_0 \in \mathbb{R}^V$,*

$$\lim_{t \to \infty} \widetilde{\boldsymbol{W}}^t \boldsymbol{p}_0 = \frac{1}{\boldsymbol{d}(V)} \cdot \boldsymbol{d}$$

Essentially, the above result says that irrespective of the initial distribution, the lazy random walk converges on any graph. Further, it converges to the same distribution, where each vertex has probability mass proportional to its degree.

Proof. We have

$$\begin{aligned} \boldsymbol{p}_t &= \widetilde{\boldsymbol{W}}^t \boldsymbol{p}_0 \\ &= \boldsymbol{D}^{1/2}\boldsymbol{D}^{-1/2}\widetilde{\boldsymbol{W}}^t \boldsymbol{D}^{1/2}\boldsymbol{D}^{-1/2}\boldsymbol{p}_0 \end{aligned}$$

[1] It can be shown that this random walk does not converge if and only if G is disconnected or bipartite.

$$= \boldsymbol{D}^{1/2}(\boldsymbol{D}^{-1/2}\widetilde{\boldsymbol{W}}\boldsymbol{D}^{1/2})^t\boldsymbol{D}^{-1/2}\boldsymbol{p}_0$$

Recall that

$$\widetilde{\boldsymbol{W}} = \frac{1}{2}(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{D}^{-1})$$

So

$$\boldsymbol{D}^{-1/2}\widetilde{\boldsymbol{W}}\boldsymbol{D}^{1/2} = \frac{1}{2}(\boldsymbol{I} + \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}) = \boldsymbol{I} - \frac{1}{2}\boldsymbol{N}.$$

Now, we write

$$\boldsymbol{p}_t = \boldsymbol{D}^{1/2}(\boldsymbol{I} - \frac{1}{2}\boldsymbol{N})^t\boldsymbol{D}^{-1/2}\boldsymbol{p}_0.$$

Note that $\boldsymbol{N}$ and $\boldsymbol{I} - \frac{1}{2}\boldsymbol{N}$ have the same eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ and since

$$0 = \lambda_1(\boldsymbol{N}) \leq \cdots \leq \lambda_n(\boldsymbol{N}) \leq 2,$$

we have

$$1 = \lambda_1(\boldsymbol{I} - \frac{1}{2}\boldsymbol{N}) \geq \cdots \geq \lambda_n(\boldsymbol{I} - \frac{1}{2}\boldsymbol{N}) \geq 0$$

where

$$\lambda_i(\boldsymbol{I} - \frac{1}{2}\boldsymbol{N}) = 1 - \lambda_i(\boldsymbol{N})/2$$

We can make use of the eigen-decomposition as follows

$$\boldsymbol{D}^{-1/2}\boldsymbol{p}_0 = \sum_i c_i\boldsymbol{v}_i \qquad\qquad \text{where } c_i = \left\langle \boldsymbol{v}_i, \boldsymbol{D}^{-1/2}\boldsymbol{p}_0 \right\rangle$$

Therefore it follows that

$$\begin{aligned}
\boldsymbol{p}_t &= \boldsymbol{D}^{1/2}(\boldsymbol{I} - \frac{1}{2}\boldsymbol{N})^t \sum_i c_i\boldsymbol{v}_i \\
&= \boldsymbol{D}^{1/2} \sum_i \left(\lambda_i(\boldsymbol{I} - \frac{1}{2}\boldsymbol{N})\right)^t c_i\boldsymbol{v}_i \\
&= \boldsymbol{D}^{1/2}c_1\boldsymbol{v}_1 + \boldsymbol{D}^{1/2} \sum_{i\geq 2} \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t c_i\boldsymbol{v}_i
\end{aligned}$$

Since $\lambda_i(\boldsymbol{N}) > 0$ for $i \geq 2$ because G is connected, we have

$$\boldsymbol{p}_t = \boldsymbol{D}^{1/2}c_1\boldsymbol{v}_1 \text{ when } t \to \infty$$

Recall that the first eigenvector $\boldsymbol{v}_1$ of $\boldsymbol{N}$ is $\boldsymbol{v}_1 = \frac{\boldsymbol{d}^{1/2}}{\sqrt{\boldsymbol{d}(V)}}$. So, we have

$$c_1 = \left\langle \frac{\boldsymbol{d}^{1/2}}{\sqrt{\boldsymbol{d}(V)}}, \boldsymbol{D}^{-1/2}\boldsymbol{p}_0 \right\rangle = \left\langle \frac{\boldsymbol{1}}{\sqrt{\boldsymbol{d}(V)}}, \boldsymbol{p}_0 \right\rangle = \frac{1}{\sqrt{\boldsymbol{d}(V)}}$$

and

$$\boldsymbol{p}_t = \boldsymbol{D}^{1/2}\frac{1}{\sqrt{\boldsymbol{d}(V)}}\frac{\boldsymbol{d}^{1/2}}{\sqrt{\boldsymbol{d}(V)}} = \frac{1}{\boldsymbol{d}(V)} \cdot \boldsymbol{d}$$

when $t \to \infty$. $\square$

6 Rate of Convergence

We call $\boldsymbol{\pi} = \lim_{t\to\infty} \widetilde{\boldsymbol{W}}^t \boldsymbol{p} = \frac{1}{d(V)} \cdot \boldsymbol{d}$ the **stable/stationary distribution.**

Observation 6.1.

$$\boldsymbol{\pi} = \widetilde{\boldsymbol{W}}\boldsymbol{\pi}.$$

That is, doing a random walk from the stationary distribution keeps it intact.

Proof. Suppose that $\pi_{t-1}(v) = \frac{d(v)}{d(V)}$ for each vertex v. Fix a vertex x. The total probability mass that lands at x from y after one more random walk step is $\frac{1}{d(y)} \frac{d(y)}{d(V)}$ if $(x, y) \in E(G)$. Now summing over all such y we get $\pi_t(x) = \sum_{y \in N_G(x)} \frac{1}{d(V)} = \frac{d(x)}{d(v)} = \pi_{t-1}(x)$. □

This brings us to our next question on the rate of convergence. How fast does $\boldsymbol{W}^t \boldsymbol{p}$ converge to $\boldsymbol{\pi}$ as a function of t?

Theorem 6.2. *For all $a, b \in V$ and t, suppose $\boldsymbol{p}_0 = \mathbf{1}_a$. Then*

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t.$$

That is, $\lambda_2(\boldsymbol{N})$ dictates how fast $\boldsymbol{p}_t$ converges to the stable distribution $\boldsymbol{\pi}$. On $\Omega(1)$ expanders, we have $\lambda_2(\boldsymbol{N}) \geq \Omega(1)$ and the bound decreases exponentially.

Proof. From previous analysis we have

$$\boldsymbol{p}_t = \widetilde{\boldsymbol{W}}^t \boldsymbol{p}_0 = \boldsymbol{D}^{1/2} c_1 \boldsymbol{v}_1 + \boldsymbol{D}^{1/2} \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t c_i \boldsymbol{v}_i$$

where

$$c_i = \left\langle \boldsymbol{v}_i, \boldsymbol{D}^{-1/2} \boldsymbol{p}_0 \right\rangle$$

So

$$\begin{aligned} \boldsymbol{p}_t(b) = \mathbf{1}_b^\top \boldsymbol{p}_t &= \mathbf{1}_b^\top \boldsymbol{D}^{1/2} c_1 \boldsymbol{v}_1 + \mathbf{1}_b^\top \boldsymbol{D}^{1/2} \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t c_i \boldsymbol{v}_i \\ &= \boldsymbol{\pi}(b) + \mathbf{1}_b^\top \boldsymbol{D}^{1/2} \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t c_i \boldsymbol{v}_i \end{aligned}$$

That is,

$$\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b) = \mathbf{1}_b^\top \boldsymbol{D}^{1/2} \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t c_i \boldsymbol{v}_i$$

and we only need to bound the right hand side. To do this, observe that

$$c_i = \left\langle \boldsymbol{v}_i, \boldsymbol{D}^{-1/2} \boldsymbol{p}_0 \right\rangle = \frac{1}{\sqrt{\boldsymbol{d}(a)}} \boldsymbol{v}_i^\top \mathbf{1}_a$$

also

$$\mathbf{1}_b^\top \boldsymbol{D}^{1/2} \boldsymbol{v}_i = \sqrt{\boldsymbol{d}(b)} \mathbf{1}_b^\top \boldsymbol{v}_i$$

We have

$$\mathbf{1}_b^\top \boldsymbol{D}^{1/2} \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t c_i \boldsymbol{v}_i = \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t \cdot \mathbf{1}_b^\top \boldsymbol{v}_i \boldsymbol{v}_i^\top \mathbf{1}_a$$

Now, we are ready to bound the second term

$$\begin{aligned}
&\left| \sum_i \left(1 - \frac{\lambda_i(\boldsymbol{N})}{2}\right)^t \cdot \mathbf{1}_b^\top \boldsymbol{v}_i \boldsymbol{v}_i^\top \mathbf{1}_a \right| \\
&\leq \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t \sum_i |\mathbf{1}_b^\top \boldsymbol{v}_i| \cdot |\boldsymbol{v}_i^\top \mathbf{1}_a| \\
&\leq \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t \sqrt{\sum_i (\mathbf{1}_b^\top \boldsymbol{v}_i)^2} \cdot \sqrt{\sum_i (\mathbf{1}_a^\top \boldsymbol{v}_i)^2} && \text{by Cauchy-Schwartz} \\
&= \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t \|\mathbf{1}_b\| \cdot \|\mathbf{1}_a\| && \text{as } \boldsymbol{v}_i \text{ form an orthonormal basis} \\
&= \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t
\end{aligned}$$

So we can now conclude that

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t.$$

□

7 Mixing Time

Definition 7.1. We say that a walk after step t has **ϵ-mixed** if

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \epsilon \boldsymbol{\pi}(b)$$

for all vertices b.

In particular, if the walk at step t is ϵ-mixed, the **total variation** between $\boldsymbol{p}_t$ and $\boldsymbol{\pi}$ which is defined as$\|\boldsymbol{p}_t - \boldsymbol{\pi}\|_{TV} = \frac{1}{2}\sum_b |\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \frac{1}{2}\sum_b \epsilon \boldsymbol{\pi}(b) = \frac{\epsilon}{2}$ is at most $\epsilon/2$.

Definition 7.2. (Mixing Time) Fix some random walk process. The **mixing time** $\tau_{\text{mix}}(G, \epsilon)$ of the graph G is given by

$$\tau_{\text{mix}}(G, \epsilon) = \max_{\boldsymbol{p}_0} \min_t \{t \mid \|\boldsymbol{p}_t - \boldsymbol{\pi}\|_{TV} \leq \epsilon\}.$$

In English, mixing time is the minimium amount of steps for total variation to be ϵ small starting with any possible initial distribution.

7.1 Mixing time and eigenvalues: $\tau_{\text{mix}}(G, \epsilon) \approx 1/\lambda_2(\boldsymbol{N}_G)$

The theorem below shows that $\tau_{\text{mix}}(G, \epsilon)$ is just another way to think about $\lambda_2(\boldsymbol{N}_G)$!

Theorem 7.3. *We have that*

$$\Omega\left(\frac{\log(\frac{1}{\epsilon})}{\lambda_2(\boldsymbol{N}_G)}\right) \leq \tau_{\text{mix}}(G, \epsilon) \leq O\left(\frac{\log(\frac{\boldsymbol{d}(V)}{\epsilon \boldsymbol{d}_{\min}})}{\lambda_2(\boldsymbol{N}_G)}\right)$$

where $\boldsymbol{d}_{\min} = \min_u \boldsymbol{d}(u)$.

Proof. We omit the proof of lower bound, and only prove the upper bound below.

First we show that the "worst" initial distribution that takes longest to mix the walk is

$$\boldsymbol{p}_0 = \mathbf{1}_{a^*}$$

for some vertex a^*.

Observation 7.4. *Pick any initial probability distribution $\boldsymbol{p}_0$ on the vertices. Then there exists a vertex a^* such that the probability distribution $\mathbf{1}_{a^*}$ mixes slower - that is, $\tau_{mix}(\boldsymbol{p}_0) < \tau_{mix}(\mathbf{1}_{a^*})$.*

Proof. We get $\|p_t - \boldsymbol{\pi}(t)\|_{TV} = \frac{1}{2}\sum_b |\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)|$. Let $p_t(a, b)$ denote the fraction of mass starting from a landing on b. Then clearly

$$\begin{aligned}
\|p_t - \boldsymbol{\pi}(t)\|_{TV} &= \frac{1}{2}\sum_b |\sum_a p_0(a)\boldsymbol{p}_t(a,b) - \boldsymbol{\pi}(b)| \\
&= \frac{1}{2}\sum_b |\sum_a (p_0(a)\boldsymbol{p}_t(a,b) - p_0(a)\boldsymbol{\pi}(b))| \\
&\leq \frac{1}{2}\sum_b \sum_a |p_0(a)\boldsymbol{p}_t(a,b) - p_0(a)\boldsymbol{\pi}(b)| (\text{Triangle inequality}) \\
&= \frac{1}{2}\sum_b \sum_a p_0(a)|\boldsymbol{p}_t(a,b) - \boldsymbol{\pi}(b)| \\
&= \frac{1}{2}\sum_a p_0(a) \sum_b |\boldsymbol{p}_t(a,b) - \boldsymbol{\pi}(b)| \\
&\leq \frac{1}{2}\max_a \sum_b |\boldsymbol{p}_t(a,b) - \boldsymbol{\pi}(b)|.
\end{aligned}$$

Suppose $a = a^*$ maximizes $\sum_b |\boldsymbol{p}_t(a,b) - \boldsymbol{\pi}(b)|$. Then it is clear that the random walk starting at a^* takes more time to mix than the random walk starting with the initial distribution p_0, and hence we are done.

□

We are now ready to prove the result. Fix a vertex a^* from which we perform the lazy random walk. That is, fix $p_0 = \mathbf{1}_{a^*}$ We have

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a^*)}} \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t .$$

So we ask that is the smallest t where

$$\begin{aligned}
\sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a^*)}} \left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t &\leq \frac{\epsilon \boldsymbol{d}(b)}{\boldsymbol{d}(V)} && \Longleftrightarrow \\
\left(1 - \frac{\lambda_2(\boldsymbol{N})}{2}\right)^t &\leq \epsilon \frac{\sqrt{\boldsymbol{d}(a^*)\boldsymbol{d}(b)}}{\boldsymbol{d}(V)} && \Longleftrightarrow \\
\exp(-\Theta(\frac{t\lambda_2(\boldsymbol{N})}{2})) &\leq \epsilon \frac{\sqrt{\boldsymbol{d}(a^*)\boldsymbol{d}(b)}}{\boldsymbol{d}(V)} && \Longleftrightarrow \text{using } 1 - x \approx \exp(-x) \\
-\Theta(\frac{t\lambda_2(\boldsymbol{N})}{2}) &\leq \ln(\epsilon \frac{\sqrt{\boldsymbol{d}(a^*)\boldsymbol{d}(b)}}{\boldsymbol{d}(V)}) && \Longleftrightarrow \\
t &\geq \Theta(\frac{2}{\lambda_2(\boldsymbol{N})}) \ln(\frac{\boldsymbol{d}(V)}{\epsilon\sqrt{\boldsymbol{d}(a^*)\boldsymbol{d}(b)}}) && \Longleftrightarrow
\end{aligned}$$

from which we get

$$\tau_{\text{mix}}(G, \epsilon) \leq O\left(\frac{\log(\frac{\boldsymbol{d}(V)}{\epsilon \boldsymbol{d}_{\min}})}{\lambda_2(\boldsymbol{N}_G)}\right).$$

□

7.2 Examples

- Let's try to answer the question about cycles from the beginning of the lecture.
  - Why uniform distribution? - we saw that the probability mass on a vertex is proportional to its degree, and hence we have the uniform distribution as the stationary distribution.
  - How fast does the distribution converge? For a cycle G, $\lambda_2(N_G) = \frac{1}{n^2}$. Hence the mixing time is $O(n^2 \log n)$.
- When it is not clear what is $\tau_{\text{mix}}(G, \epsilon)$ or $\lambda_2(\boldsymbol{N}_G)$, knowing one implies $O(\log n)$-approximation of the other.
  - When G is a $\tilde{\Omega}(1)$-expander, we know $\lambda_2(\boldsymbol{N}_G) = \tilde{\Omega}(1)$. So $\tau_{\text{mix}} = \tilde{O}(1)$.
  - When G is a path, we know $\tau_{\text{mix}} \approx n^2$ (why?), so $\lambda_2(\boldsymbol{N}_G) \approx 1/n^2$.
    We can see this from the conductance of path. Another good reason is that the expected transition distance after n steps of a random walk is of order $\sqrt{n}$,[2] whence it takes roughly n^2 steps to move n units of distance away from the starting point.

[2]https://mathworld.wolfram.com/RandomWalk1-Dimensional.html

- When G is a dumbell, $\tau_{\text{mix}} \approx n^2$ (why?), so $\lambda_2(\boldsymbol{N}_G) \approx 1/n^2$
  The idea is that there is roughly $1/n$ of probability for a the current point to move from one clique to an endpoint of the bridge, and it takes another $1/n$ of probability to move from one endpoint of the bridge to the other (so that the point will enter the other clique.)
- When G is a Bolas graph, $\tau_{\text{mix}} \approx n^3$ (why?), so $\lambda_2(\boldsymbol{N}_G) \approx 1/n^3$.
  Similarly, it takes roughly $1/n$ of probability to move from one clique to an endpoint of the bridge, and another $1/n^2$ of probability to move to the other endpoint by previous result for path.

- Actually, the Bolas graph is the *unweighted* graph which is hardest to mix.

Exercise 7.5 (Worst mixing time in unweighted graphs)**.** In the homework, you will show that for any unweighted graph G, $\lambda_2(\boldsymbol{N}_G) \geq \Omega(1/n^3)$. So $\tau_{\text{mix}}(G, 1/\text{poly}(n)) = O(n^3 \log n)$.

8 Applications of Random Walks

8.1 Short path in $\tilde{O}(\sqrt{n})$ time on Expanders

- **Question:** Given a ϕ-expander (undirected, unweighted) $G = (V, E)$ and $s, t \in V$, can we find a "short" path from s to t in sublinear time?

- Algorithm SHORTPATH
  - Perform a lazy random walk from s for $\tau_{\text{mix}}(G, \frac{1}{n^2}) = O(\frac{\log n}{\phi^2})$ steps.
  - Repeat for $O(\sqrt{n} \log n)$ times.
  - Let S be the set of vertices where the walks end.
  - We now do the same walk from t. Let T be the similar set of vertices.
  - If a there is a vertex $v \in S \cap T$, return a path from s to t as the union of paths between s and v and v and t.

  Note that if $S \cap T \neq \emptyset$, then we obtain an (s, t) path of length $O(\tau_{\text{mix}}(G, \frac{1}{n^2}))$. The running time of the algorithm is clearly $\tilde{O}(\sqrt{n} \cdot \tau_{\text{mix}}(G, \frac{1}{n^2}))$.

  Observation 8.1. *$S \cap T \neq \emptyset$ whp.*

  Proof. (Sketch) Essentially birthday paradox. Since the walks have mixed sufficiently, S and T are essentially uniformly random vertices (since G is connected). Thus the probability of their intersection being non-empty is the probability that given n bins, if we independently throw $O(\sqrt{n} \log n)$ balls uniformly at random, two of the balls land in the same bin. To analyze this formally, note that the probability that no two of the balls collide is at most $(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{\sqrt{n} \log n}{n})$ which is at most $e^{-\sum_{i=1}^{\sqrt{n} \log n} \frac{i}{n}} \sim e^{-\log^2 n} \ll \frac{1}{n^c}$ for any constant c. $\square$

  Corollary 8.2. *If G is an $\Omega(1)$-expander, then algorithm* SHORTPATH *returns a $O(\log n)$-length (s, t)-path in $\tilde{O}(\sqrt{n})$ time.*

8.2 Connectivity in $O(\log n)$ bits of space

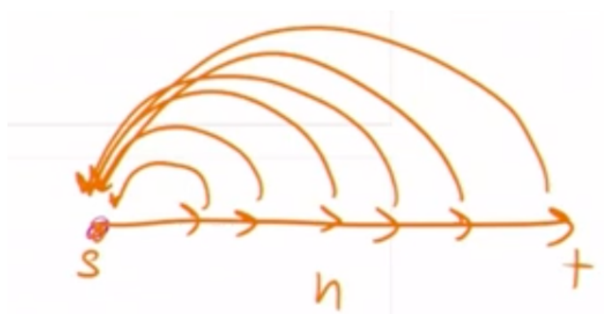
- **Question:** Given a (undirected and unweighted) graph $G = (V, E)$ and $s, t \in V$, is there a $s - t$ path?
- We shall assume that the adjacency lists of G are *read-only* and we can use only $O(\log n)$ bits of *working memory*.
- **The algorithm**
  - Start from s.
  - Perform lazy random walks for $O(n^5 \log^2 n)$ steps.
  - Whenever we visit t, return YES.
  - If we never visit t, return NO.
- If s and t are not connected, clearly the algorithm return NO.

Observation 8.3. *If s and t are in the same connected component, then the random walk will meet t whp.*

Proof. Let C be the connected component containing s and t.
We have $\tau_{\text{mix}}(C, \frac{1}{n^2}) \leq O(n^3 \log n)$ (as C is an unweighted graph - by exercise 7.5)
Therefore after every $O(n^3 \log n)$ steps, we visit t w.p. $\frac{d(t)}{d(C)} \geq \frac{1}{n^2}$. It follows that after $O(n^5 \log^2 n)$ steps, we never visit t w.p. at most $(1 - \frac{1}{n^2})^{O(n^2 \log n)} \leq 1/\text{poly}(n)$. □

Question 8.4. *Is it possible to improve the time to $\tilde{O}(n^2)$ or even $\tilde{O}(n)$ while using $O(\text{polylog}(n))$ space?*

For directed graph, the above algorithm won't work. Indeed, consider the graph in the below figure-in expectation it takes 2^n steps for s to reach t. An interesting open problem is that: is there an (even randomized) s-t connectivity algorithm use $O(n^{0.99})$ space runs in $\text{polylog}(n)$ time? The current best known deterministic algorithm uses $O(n/2^{\sqrt{\log n}})$ space.



8.3 Sampling a Spanning Tree

The problem of sampling (almost uniformly) a spanning tree has been studied extensively in the literature. However, the "right" algorithm has been recently shown last year [3]. We will

[3] https://arxiv.org/abs/2004.07220

only discuss the high level ideas involved in the sampling process and skip the technical details.

- **Question:** Given a connected graph G, sample (almost)-uniformly at random a spanning tree T of G.
- **Key Idea:**
  - Given a spanning tree T, consider the following "flip" operation that
    - Choose one of the $m - (n - 1)$ non-tree edges e
    - Look at the unique cycle C in $T \cup e$
    - Choose e' in C and remove e'.
    - We obtain another spanning tree $T' = T \cup e \setminus e'$.

Figure 2: The flip operation

- Consider the super graph $\mathcal{G}$ where
  * each node is a spanning tree
  * We add an edge between T and T' iff T can be flipped to T'.
- **Key idea (roughly speaking)**: although $\mathcal{G}$ is exponentially big, $\mathcal{G}$ is an expander. So it takes only $O(\log |V(\mathcal{G})|)$ random walk steps for get to sample a uniform spanning tree.

- Algorithm:
  - Start with any spanning tree.
  - Randomly flip for $O(m \log m)$ steps.
  - Return the tree.
- We can simulate the tree flipping using the link-cut tree data structure in $O(\log n)$ time. So the overall running time is $O(m \log^2 n)$.

9 Conclusion

Exercise 9.1. Suppose that the graph G is d-regular. Recall from the previous lecture that the power method for computing $\lambda_2(\boldsymbol{N})$ just computes $\widetilde{\boldsymbol{W}}^{O(\log(n)/\epsilon)}\boldsymbol{x}$ where $\boldsymbol{x}$ is a some random vector.

9.1 Alternate notions of expansion

Recall the various characterizations of expanders we have seen:

1. Robustness view: Robust towards edge deletions
2. Cut view: No sparse cuts, high conductance
3. Flow view: Can embed any degree restricted demand with small congestion
4. Spectral view: Small second eigenvalue of normalized laplacian
5. Probabilistic view: Random walks mix fast

Most of the theory we have seen is about conductance (i.e. edge expansion) in undirected graphs. But what about other notions?

Question 9.2. *Survey the equivalence we have seen so far for vertex expansion and/or directed graphs.*