Convergence of Random Walks in Undirected Graphs

$G = (V, E)$ a connected graph.

We have seen various properties of random walks on G

(i) Stationary distribution depends only on degrees.

$$\pi_i = \frac{deg(i)}{2m}.$$

(ii) Cover time is $\leq 2m(n-1)$

(iii) $h_{s,t} + h_{t,s} = 2m \, R^{eff}_{s,t}$

Made connection to electrical networks.

We are interested in convergence of the random walk to its stationary distribution.

In order to make this formal we define the following.

<u>Defn:</u> Given two distributions π and π' on V their total variational distance is $\|\pi-\pi'\|_1$.

We say that a Markov chain converges to its stationary distribution if $\|\pi^t-\pi\|_1 \to 0$ as $t \to \infty$ where π is the stationary distribution.

We will be interested in the rate of convergence. That is, starting from some arbitrary π^0, how long does it take to get ε-close to π?

Recall that a chain/walk may be periodic in which case it does not converge. We will then work with the lazy walk.

We will focus on random walks in undirected graphs and spectral analysis based convergence. And then relate spectral analysis with combinatorial property of a graph.

So far we have been working with distributions over states as <u>row</u> vectors. Since we will work with eigen values and eigen vectors a lot it is useful to switch to $p^{(0)}, p^{(1)}, \ldots, p^{(t)}$ as <u>column</u> <u>vectors</u>. Recall P was the probability transition matrix of the chain. In order to work with column vectors we will use P^T.

Thus $p^{(t+1)} = P^T p^{(t)}$

where $p^{(t)}$ is the distribution of the chain at time step t.

Consider random walk on undirected graph $G = (V, E)$ with $V = [n]$.

We let A be the adjacency matrix of G. Let D be the $n \times n$ diagonal matrix with $D_{ii} = \deg(i)$. Then D^{-1} will be the matrix with $D^{-1}_{ii} = \frac{1}{\deg(i)}$.

Recall the transition matrix P was defined as $P_{ij} = \frac{1}{\deg(i)}$.

Hence $P^T_{ij} = \frac{1}{\deg(j)}$.

We can see that $P^T = A D^{-1}$.

We call $W = AD^{-1}$ the walk matrix.

For the lazy random walk the walk matrix is $\frac{1}{2}(I + AD^{-1})$.

Suppose G is d-regular. Then W is symmetric. Symmetric matrices have substantial structure and there is a beautiful and powerful spectral theory for them with many applications.

Review of some linear algebra

For $n \times n$ matrix $A \in R^{n \times n}$ a vector $\bar{v} \in R^n$ is an eigen vector of A if $\exists\, \lambda \in R$ s.t.

$$A\bar{v} = \lambda \bar{v}.$$

λ is an eigen value.

Eigen values are solutions to the polynomial $\det(A - \lambda I) = 0$.

In general this polynomial may not have real roots and hence A may not have eigen vectors/values.

However if A is viewed as a $\mathbb{C}^{n\times n}$ matrix then we have complex eigen values/vectors since every univariate polynomial over $\mathbb{C}$ can be factorized.

Symmetric matrices are however special and have substantial structure. This is captured by the Spectral theorem which has many applications.

Spectral theorem

Let A be a $n \times n$ symmetric matrix. Then A has n real eigen values $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$ and corresponding eigen vectors $\bar{x}_1, \ldots, \bar{x}_n$ which form an orthonormal basis of $\mathbb{R}^n$. And $A = V^T D V$ where D is a diagonal matrix with $D_{ii} = \lambda_i$ and $V = [\bar{x}_1 \ \bar{x}_2 \ \cdots \ \bar{x}_n]$.

A useful characterization of the eigenvectors is obtained via the Raleigh coefficient.

Given A and $\bar{x} \in \mathbb{R}^n$

Consider $x^t A \bar{x}$ which is the quadratic form induced by A.

Note that $x^t A x = \sum_{1 \le i, j \le n} A_{ij} x_i x_j$

When A is a symmetric matrix we get $\sum_{i=1}^{n} A_{ii} x_i^2 + 2 \sum_{i<j} A_{ij} x_i x_j$.

Consider the problem of

$$\max_{\|x\|=1} / \min \; x^t A x$$

which is same as

$$\max_{\bar{x} \neq 0} / \min \; \frac{x^t A x}{x^t x}$$

Turns out that if A is symmetric then

$$\lambda_1 = \min_{\bar{x} \neq 0} \frac{x^t A x}{x^t x}$$

$$\lambda_n = \max_{\bar{x} \neq 0} \frac{x^t A x}{x^t x}$$

In fact one can characterize all eigen values.

$$\lambda_k = \min_{V_k} \max_{\substack{\bar{x} \in V_k \\ \bar{x} \neq 0}} \frac{x^t A x}{x^t x}$$

where V_k is a. k-dimensional subspace of $\mathbb{R}^n$.

One can derive this characterization from the spectral theorem or directly.

Suppose $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ are orthonormal unit vectors that are eigen values of A and let $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$.

Let $\bar{v} \in \mathbb{R}^n$ and $\|\bar{v}\| = 1$ be any unit vector.

Then $\bar{v} = \sum_{i=1}^{n} \alpha_i \bar{x}_i$ where $\sum_{i=1}^{n} \alpha_i^2 = 1$

Consider

$$\min_{V_n} \max_{\substack{\bar{v} \in V_n \\ \|\bar{v}\|=1}} v^t A \bar{v}$$

$$= \max_{\substack{v \in R^n \\ \|\bar{v}\|=1}} v^t A \bar{v}$$

$$v^t A \bar{v} = \sum_{i=1}^{n} \lambda_i \alpha_i^2 \quad \text{where } \bar{v} = \sum_{i=1}^{n} \alpha_i \bar{x}_i$$

hence $\max_{\substack{\|\bar{v}\|=1 \\ \bar{v} \in R^n}} v^t A \bar{v} = \lambda_n.$

similarly $\min_{\substack{\|\bar{v}\|=1 \\ \bar{v} \in R^n}} = \lambda_1.$

etc.

Defn: A real symmetric matrix A is positive semi-definite (psd) $(A \succeq 0)$ if $\bar{v}^t A \bar{v} \geq 0 \ \forall \ \bar{v} \in \mathbb{R}^n$.

Theorem: Symmetric matrix A is psd if one of the following conditions holds.

1. $\bar{v}^t A \bar{v} \geq 0 \ \forall \ \bar{v} \in \mathbb{R}^n$.
2. All eigenvalues of A are non-negative $0 \leq \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$.
3. $A = W^t W$ for some $W \in \mathbb{R}^{n \times n}$.

Proof: $(3) \Rightarrow (1)$

$$
\begin{aligned}
v^t A v &= v^t W^t W v \\
&= u^t \cdot u \geq 0.
\end{aligned}
$$

1 $\Rightarrow$ 2

Recall for symmetric A

$$\lambda_1 = \min_{\substack{\bar{v} \in \mathbb{R}^n \\ \|\bar{v}\| = 1}} v^t A \bar{v}.$$

Hence $\lambda_1 \geq 0$.

2 $\Rightarrow$ 3

A is symmetric

$\Rightarrow A = V^t D V$ where

$D_{ii} = \lambda_i$ if $\lambda_i \geq 0 \ \forall i$

then $D = \left(D^{(1/2)}\right)^T D^{(1/2)}$

where $D^{1/2}_{ii} = \sqrt{\lambda_i}$.

Hence $A = V^t \left(D^{1/2}\right)^T D^{1/2} V = W^t W$.

$\square$.

Back to Random Walks

We will first consider d-regular graphs.

$W = AD^{-1}$ is symmetric.

Note that W is doubly stochastic

By spectral theorem all eigen values are real. Let $\alpha_1 \geq \alpha_2 \ldots \geq \alpha_n$ be the eigen values of W.

Claim: $\alpha_1 = 1$ and $\alpha_n \leq -1$.

Proof: Exercise.

Exercise: Show that $\alpha_n = -1 \Rightarrow G$ is bipartite. Otherwise $\alpha_n < 1$.

Claim: Consider $W = \frac{1}{2}(I + AD^{-1})$.

Eigenvalues are $\underbrace{\frac{1+\alpha_1}{2}}_{=1}, \frac{1+\alpha_2}{2}, \ldots, \underbrace{\frac{1+\alpha_n}{2}}_{\geq 0}$.

Convergence Analysis

By spectral theorem

W can be written as

$$\sum_{i=1}^{n} \alpha_i \; \bar{v}_i \oplus \bar{v}_i$$

where $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$ are the eigen vectors of W normalized to be unit vectors and $\oplus$ denotes outer product.

$\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$ are orthonormal.

We want to know $W^t p^{(0)}$ where $p^{(0)}$ is the starting distribution.

Since $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ are orthonormal we can write $p^{(0)}$ as $\sum_{i=1}^{n} c_i \bar{v}_i$

where $c_i = \langle p^{(0)}, \bar{v}_i \rangle$

Then $W^t p^{(0)} = \sum_{i=1}^{n} c_i \alpha_i^t p^{(0)}$

Recall $\alpha_1 = 1$ and $\bar{v}_1 = \frac{\bar{1}}{\sqrt{n}}$ (since we normalize to unit vectors).

Spectral gap $\beta = \min\{1-\alpha_2, 1-|\alpha_n|\}$

Note $0 \leq \beta \leq 1$

Suppose $\beta > 0 \Rightarrow \alpha_2 < 1$ and $\alpha_n < -1$

$\Rightarrow 1 < \alpha_2, \ldots, \alpha_n < -1$

$\Rightarrow W^t p^{(0)} \longrightarrow c_1 \bar{v}_1$ as $t \to \infty$

since $\alpha_1 = 1$.

$c_1 = \langle p^{(0)}, \bar{v}_1 \rangle = \frac{1}{\sqrt{n}}$

since $\sum p_i^{(0)} = 1$ and $\bar{v}_1 = \frac{\bar{1}}{\sqrt{n}}$.

$\Rightarrow \bar{c}_1 \bar{v}_1 = \frac{\bar{1}}{n}$ which is the

uniform distribution, and the stationary distribution.

Rate of Convergence

Claim: Mixing time is $O\left(\frac{\ln n}{\gamma}\right)$

Proof:

$$W^t \bar{p}_0 = \bar{\pi} + \sum_{i=2}^{n} c_i \alpha_i^t \bar{v}_i$$

Hence

$$d_{TV}(W^t \bar{p}_0, \bar{\pi}) = \| W^t \bar{p}_0 - \bar{\pi} \|_1$$

$$= \| \sum_{i=2}^{n} c_i \alpha_i^t \bar{v}_i \|_1$$

$$\leq \sqrt{n} \left\| \sum_{i=2}^{n} c_i \alpha_i^t \bar{v}_i \right\|_2$$

by Cauchy-Schwartz;

Since $\bar{v}_i$ are othonormal.

$$\left\| \sum_{i=2}^{n} c_i \alpha_i^t \bar{v}_i \right\|_2 \leq \sum_{i=2}^{n} c_i^2 \alpha_i^{2t}$$

$$\leq (1-\beta)^{2t} \sum_{i=2}^{n} c_i^2$$

$$\sum_{i=2}^{n} c_i^2 \leq \sum_{i=1}^{n} c_i^2 = \|\bar{p}_0\|_2^2 \leq \|p_0\|_1^2 \leq 1.$$

Hence $d_{TV}(W^t \bar{p}_0 , \bar{\Pi}) \leq (1-\beta)^{2t} \sqrt{n}$

$\Rightarrow$ want $(1-\beta)^{2t} \sqrt{n} \leq \frac{1}{4}$

$2t \ln(1-\beta) \geq \quad \ln 4\sqrt{n}$

$\Rightarrow \quad t = \Omega\left(\frac{\ln n}{\beta}\right)$ suffices..

<u>Lazy</u> <u>random</u> <u>walk</u>

$W = \frac{1}{2}(I + A)$

Recall $\alpha'_1 = 1 \quad \alpha'_2 = \frac{1+\alpha_2}{2}, \ldots,$

$$\alpha'_n = \frac{1+\alpha_n}{2} \geq 0.$$

$$\Rightarrow \beta = \min(1-\alpha'_2, 1-\alpha'_n)$$

$$= 1-\alpha'_2 = \frac{1}{2}(1-\alpha_2).$$

<u>Example</u>: $G = C_n$ the n cycle.

What are eigen values of AD^{-1}?

Can show that they are

$$\alpha_i = \cos \frac{2\pi(i-1)}{n}.$$

If n is even $\Rightarrow$ $\alpha_n = -1$

spectral gap is 0. Hence need to use lazy walk.

What about α_2? $\cos \frac{2\pi}{n}$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \dots$$

as $x \to 0$ $\cos x \sim 1 - \frac{x^2}{2}$.

$\Rightarrow$ $\cos \frac{2\pi}{n}$ as $n \to \infty$ $\rightsquigarrow$ $1 - \frac{2\pi^2}{n^2}$.

$\Rightarrow$ $\alpha_2 \geq 1 - \frac{2\pi^2}{n^2}$

$\Rightarrow$ $\beta \leq \frac{1}{n^2}$.

$\Rightarrow$ Convergence time is $\Omega(n^2)$.

Not surprising.

General graphs

not necessarily regular.

Lazy walk.

$$W = \frac{1}{2}(I + AD^{-1}).$$

is walk matrix.

W is not symmetric so cannot use dual theorem directly.

Consider normalized adjacency matrix

$$\mathcal{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

$$\mathcal{A}_{ij} = \frac{1}{\sqrt{dg(i)}\sqrt{dg(j)}}. \qquad \text{symmetric}$$

We can write

$$W \text{ as } D^{\frac{1}{2}}\underbrace{\left(\frac{1}{2}\left(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right)\right)}_{\text{Symmetric}} D^{-\frac{1}{2}}$$

W is <u>similar</u> to a symmetric matrix.

<u>Defn:</u> X, Y are $n \times n$ matrices.
X is <u>similar</u> to Y if
there exists a <u>non-singular</u> matrix
B s.t. $X = BYB^{-1}$

The action of X can be understood via action of Y.

<u>Claim</u>: Eigen <u>values</u> of X, Y same. Hence same spectrum. Eigen vectors may be different.

<u>Proof</u>: Suppose

$$Y \bar{v} = \lambda \bar{v}$$

Let $\bar{u} = B \bar{v} \implies B^{-1} \bar{u} = \bar{v}$

then
$$\begin{aligned} X \bar{u} &= (B Y B^{-1}) \bar{u} \\ &= B (Y \bar{v}) = B (\lambda \bar{v}) \\ &= \lambda \bar{u} \end{aligned}$$

$\implies \lambda$ is eigen value of X.

$\square$.

<u>Corollary</u>: If X is similar to symmetric matrix then all eigen values are real. Eigen vectors span $\mathbb{R}^n$ even though they may not be orthonormal.

Now back to W.

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

$$W \text{ is similar to } \frac{1}{2}\left(I + \underbrace{D^{-\frac{1}{2}} A D^{-\frac{1}{2}}}_{\substack{\text{normalized} \\ \text{adjacency} \\ \text{matrix.}}}\right)$$

Eigenvalues of A are

$$1 \geq \alpha_1 \geq \alpha_2 \cdots \geq \alpha_n \geq -1.$$

$\Rightarrow$ Eigen values of W are

$$\alpha_i' = \frac{1+\alpha_i}{2} \qquad i=1 \text{ to } n$$

If $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$ are eigen vectors of $\frac{1}{2}(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}})$ then eigen vectors of W are

$$D^{\frac{1}{2}} \bar{v}_i \qquad i=1 \text{ to } n.$$

Note $\bar{v}_1, \ldots, \bar{v}_n$ are orthonormal.

$\Rightarrow$ $D^{\frac{1}{2}} v_i$ $i=1$ to n are linearly indep and span R^n.

Now we want to understand $W^t \bar{p}_0$ where $\bar{p}_0$ is the starting distribution.

$$\bar{p}_0 = \sum_{i=1}^{n} c_i D^{1/2} \bar{v}_i \quad \text{for some } c_1, \dots, c_n$$

since the eigenvectors of W span $\mathbb{R}^n$.

Therefore $W^t \bar{p}_0 = \sum_{i=1}^{n} c_i W^t (D^{1/2} v_i)$

recall $D^{1/2} \bar{v}_i$ is eigenvector of W with eigenvalue $\frac{1+\lambda_i}{2}$

$$\Rightarrow W^t \left(D^{1/2} \bar{v}_i\right) = \left(\frac{1+\alpha_i}{2}\right)^t D^{1/2} v_i$$

$$\Rightarrow W^t \bar{p}_0 = \sum_{i=1}^{n} c_i \left(\frac{1+\alpha_i}{2}\right)^t D^{1/2} \bar{v}_i$$

Recall $1 = \alpha_1 \geq \alpha_2 \cdots \geq \alpha_n \geq -1$

$$\underbrace{1 = \frac{1+\alpha_1}{2}}_{\beta_1} \geq \underbrace{\frac{1+\alpha_2}{2}}_{\beta_2} \cdots \geq \underbrace{\frac{1+\alpha_n}{2}}_{\beta_n} \geq 0.$$

If $\beta_2 < 1$ then

$$W^t \bar{p}_0 = c_1 D^{1/2} \bar{v}_1 + \sum_{i=2}^{n} \beta_i^t D^{1/2} \bar{v}_i$$

Since $\beta_2 < 1 \Rightarrow \beta_2, \cdots, \beta_n \leq 1$

$$W^t \bar{p}_0 \rightarrow c_1 D^{1/2} \bar{v}_1 \text{ as } t \rightarrow \infty$$

$\Rightarrow$ Converges to $c_1 D^{1/2} \bar{v}_1$

Stationary distribution.

What is $c_1 D^{1/2} \bar{v}_1$?

Recall $\bar{p}_0 = \sum_{i=1}^{n} c_i D^{1/2} \bar{v}_i$

$$= D^{1/2} \sum_{i=1}^{n} c_i \bar{v}_i$$

$\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$ orthonormal.

$$c_1 = \langle D^{-1/2} \bar{p}_0, \bar{v}_1 \rangle$$

Recall $\bar{v}_1$ is first eigenvector of

$\frac{1}{2}(I + D^{-1/2} A D^{1/2})$ which is ?

Claim: $\bar{v}_1 = \dfrac{D^{\frac{1}{2}}\bar{1}}{\|D^{1/2}\bar{1}\|} = \dfrac{1}{\sqrt{2m}} D^{1/2}\bar{1}$

Therefore $c_1 = \langle D^{-1/2}\bar{p}_0, \bar{v}_1 \rangle$

$$= \left\langle D^{-1/2}\bar{p}_0, \frac{1}{\sqrt{2m}} D^{1/2}\bar{1} \right\rangle$$

$$= \frac{1}{\sqrt{2m}}.$$

$\Rightarrow$ $W^t \bar{p}_0 \longrightarrow c_1 D^{1/2} \bar{v}_1$

$$= \frac{1}{\sqrt{2m}} \cdot D^{1/2}\left(\frac{1}{\sqrt{2m}} \cdot D^{1/2}\bar{1}\right)$$

$$= \frac{1}{2m} D\bar{1} = \frac{1}{2m}\begin{bmatrix} deg(1) \\ deg(2) \\ \vdots \\ deg(n) \end{bmatrix}.$$

Hence lazy walk converges to stationary distribution whatever is the starting distribution. Note lazy walk only assumes connectivity (Since $\lambda_2 < 1$ which is true if G is connected).

<u>Mixing time?</u>

$\bar{\pi} = \frac{1}{2m} D \bar{1}$ is stationary distribution

$$\bar{p}_t = W^t \bar{p}_0$$
$$= \bar{\pi} + \sum_{i=2}^{n} c_i \left(\frac{1+\lambda_i}{2}\right)^t D^{1/2} \bar{v}_i$$

Want to know how long before

$$\| \bar{p}_t - \bar{\pi} \|_1 \leq \frac{1}{4}.$$

$$\bar{p}_t - \bar{\pi} = \sum_{i=2}^{n} c_i \left(\frac{1+\alpha_i}{2}\right)^t D^{1/2} \bar{v}_i$$

$$\Rightarrow D^{-1/2}(\bar{p}_t - \bar{\pi}) = \sum_{i=2}^{n} c_i \left(\frac{1+\alpha_i}{2}\right)^t \bar{v}_i$$

$\| \bar{u} \|_1 \leq \sqrt{n} \, \| u \|_2$ by Cauchy-Schwartz.

$$\Rightarrow \| D^{-1/2}(\bar{p}_t - \pi) \|_2^2 \leq \left\| \sum_{i=2}^{n} c_i \left(\frac{1+\alpha_i}{2}\right)^t \bar{v}_i \right\|^2$$

$$= \sum_{i=2}^{n} c_i^2 \left(\frac{1+\alpha_i}{2}\right)^{2t}$$

Since $\bar{v}_1, \ldots, \bar{v}_n$ are orthonormal.

$$= \sum_{i=2}^{n} c_i^2 \left(\frac{1+\alpha_i}{2}\right)^{2t}$$

$$\leq (1-\beta)^{2t} \cdot \sum_{i=2}^{n} c_i^2$$

Note $\sum_{i=2}^{n} c_i^2 \leq \sum_{i=1}^{n} c_i^2 = \| D^{-1/2} \bar{p}_0 \|_2^2$

$$\leq \frac{1}{d_{min}}$$

d_{min} is min degree.

$$\Rightarrow \| D^{-1/2} (\bar{p}_t - \bar{\pi}) \|_2^2 \leq \frac{1}{d_{min}} \cdot (1-\beta)^{2t}$$

$$\| D^{-1/2} (\bar{p}_t - \bar{\pi}) \|_2^2 \geq \frac{1}{d_{max}} \| p_t - \bar{\pi} \|_2^2$$

$$\Rightarrow \quad \|\bar{p}_t - \pi\|_2^2 \leq \frac{d_{max}}{d_{min}} \cdot (1-\beta)^{2t}$$

$$\Rightarrow \quad \|\bar{p}_t - \pi\|_1 \leq \sqrt{n \frac{d_{max}}{d_{min}}} \; (1-\beta)^t$$

Therefore how large should t be

s.t.

$$\sqrt{n \frac{d_{max}}{d_{min}}} \; (1-\beta)^t \leq \frac{1}{4}$$

$$\frac{d_{max}}{d_{min}} \leq n.$$

$$\Rightarrow \quad t \simeq \Omega\left(\frac{\ln n}{\beta}\right).$$ would suffice

$\square$

To get TDV $\leq \varepsilon$

need $\Omega\left(\frac{\ln\left(\frac{n}{\varepsilon}\right)}{\beta}\right)$.

$\square$.

Next lecture:

Which graphs have spectral gap constant? This will ensure that random walk converges in $O(\log n)$ steps!