CSC 2421H : Graphs, Matrices, and Optimization	Lecture 4 : 10 1 2018
Random Walk	
Lecturer: Sushant Sachdeva	*Scribe: Junwei Sun*

HW: The proof of statement is left as exercise for the student

1 Remark from last class

If at step t, your distribution is given by $\mathbf{p}_t$, then the next distribution $\mathbf{p}_{t+1}$ is given by:

$$\mathbf{p}_{t+1} = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}_t, \tag{1}$$

where:
$\mathbf{A}$: weighted adjacency matrix
$\mathbf{D}$: diagonal weighted degree matrix

The state transition can also be expressed as:

$$\mathbf{p}_{t+1}(x) = \sum_{y:(x,y)\in E} \frac{w(x,y)}{\mathbf{D}(y)} * \mathbf{p}_t(y)$$

2 Stationary State

Definition 2.1. *If distribution $\pi \in \mathbb{R}^{\mathbf{v}}$ is said to be stationary distribution for G if $\mathbf{A}\mathbf{D}^{-1}\pi = \pi$*

Lemma 2.2. *Any undirected graph has a stationary distribution*

Proof. Given any undirected graph G, let

$$\pi = \frac{1}{\mathbf{1}^\top \mathbf{D}\mathbf{1}}\mathbf{D}\mathbf{1}$$

This is the stationary distribution for G since

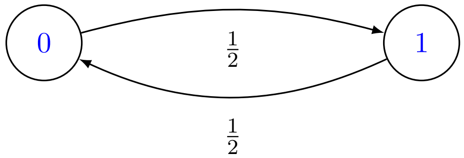
$$\mathbf{A}\mathbf{D}^{-1}\pi = \frac{1}{\mathbf{1}^\top \mathbf{D}\mathbf{1}}\mathbf{A}\mathbf{1} = \frac{1}{\mathbf{1}^\top \mathbf{D}\mathbf{1}}\mathbf{D}\mathbf{1} = \pi$$

□

Claim: If G is connected, π is unique

Remark: Even if G is connected, it is not true that for any $\mathbf{p_0}$, $\mathbf{p}_t \to \pi$

Example:



The stationary state $\pi = (\frac{1}{2}, \frac{1}{2})$ but the random walk will alternate between Vertex 0 and Vertex 1

3 Positive Semi-Definite(PSD)

Definition 3.1. *A symmetric matrix $\boldsymbol{M}$ is positive semi-definite(psd) if $\forall$ $\boldsymbol{x}$, $\mathbf{x}^\top \mathbf{M} \mathbf{x} \geq 0$*

Theorem 3.2. *the following statements are equivalent:*
(1) $\boldsymbol{M}$ is psd
(2) All eigenvalues of $\boldsymbol{M}$ are non-negative
(3) There exist an matrix $\boldsymbol{A}$ such that $\mathbf{M} = \mathbf{A}\mathbf{A}^\top$

Lemma 3.3. *If $\boldsymbol{M}$ is psd, then for all matrices $\boldsymbol{C}$, $\mathbf{C}^\top \boldsymbol{MC}$ is psd*

Proof. $\forall$ $\mathbf{x}$, $\mathbf{x}^\top \mathbf{C}^\top \mathbf{M} \mathbf{C} \mathbf{x} = (\mathbf{C}\mathbf{x})^\top \mathbf{M} (\mathbf{C}\mathbf{x}) \geq 0$ since $\mathbf{M}$ is psd □

Notation: $\mathbf{M}$ is psd $\Leftrightarrow \mathbf{M} \succeq 0$

Lemma 3.4. *$\mathcal{L} \succeq 0$ where $\mathcal{L}$ is the laplacian matrix of some graph*

Remark: $\mathcal{L} \succeq 0$ implies $\mathbf{N} \succeq 0$ since $\mathbf{N} = \mathbf{D}^{-\frac{1}{2}} \mathcal{L} \mathbf{D}^{-\frac{1}{2}}$

Lemma 3.5. $\mathcal{L} \preccurlyeq 2\boldsymbol{D} \Leftrightarrow \boldsymbol{N} \preccurlyeq 2\boldsymbol{I} \Leftrightarrow \lambda_i(\mathbf{N}), \nu_i \leq 2$

HW: If $\mathbf{A} \succeq \mathbf{B}$, then $\lambda_i(\mathbf{A}) \geq \lambda_i(\mathbf{B})$

4 Lazy random walk

4.1 Lazy random walk matrix

At each step, the lazy random walk will do the following

$$\begin{cases} \text{with probability}\frac{1}{2} & \text{stay at the current vertex} \\ \text{with probability}\frac{1}{2} & \text{take a usual random step} \end{cases}$$

Lazy Random Walk Transition Matrix $\mathbf{W} = \frac{1}{2}(\mathbf{I} + \mathbf{A}\mathbf{D}^{-1})$
We know that the normalized Laplacian($\mathbf{N}$) can be expressed as:

$$
\begin{aligned}
\mathbf{N} &= \mathbf{D}^{-\frac{1}{2}}\mathcal{L}\mathbf{D}^{-\frac{1}{2}} \\
&= \mathbf{D}^{-\frac{1}{2}}(\mathbf{D}-\mathbf{A})\mathbf{D}^{-\frac{1}{2}} \\
&= \mathbf{I}-\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}
\end{aligned}
$$

applied this result to the lazy walk transition matrix

$$
\begin{aligned}
\mathbf{W} &= \frac{1}{2}\mathbf{I}+\frac{1}{2}\mathbf{A}\mathbf{D}^{-1} \\
&= \frac{1}{2}\mathbf{I}+\frac{1}{2}\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}} \\
&= \frac{1}{2}\mathbf{I}+\frac{1}{2}\mathbf{D}^{\frac{1}{2}}(\mathbf{I}-\mathbf{N})\mathbf{D}^{-\frac{1}{2}} \\
&= \mathbf{I}-\frac{1}{2}\mathbf{D}^{\frac{1}{2}}\mathbf{N}\mathbf{D}^{-\frac{1}{2}}
\end{aligned}
$$

Thus,we can express the lazy random walk transition matrix as:

$$
\mathbf{W} = \mathbf{I}-\frac{1}{2}\mathbf{D}^{\frac{1}{2}}\mathbf{N}\mathbf{D}^{-\frac{1}{2}} \tag{2}
$$

4.2 Eigenpair for lazy random walk matrix

Lemma 4.1. *If*(ν_i, ψ_i) *is an eigenpair for* $\boldsymbol{N}$*, i.e* $\boldsymbol{N}\psi_i = \nu_i\psi_i \Leftrightarrow (1-\frac{1}{2}\nu_i, D^{\frac{1}{2}}\psi_i)$ *is an eigenpair for* $\boldsymbol{W}$

Proof.

$$
\begin{aligned}
\mathbf{W}\mathbf{D}^{\frac{1}{2}}\psi_i &= (\mathbf{I}-\frac{1}{2}\mathbf{D}^{\frac{1}{2}}\mathbf{N}\mathbf{D}^{-\frac{1}{2}})\mathbf{D}^{\frac{1}{2}}\psi_i \\
&= \mathbf{D}^{\frac{1}{2}}\psi_i-\frac{1}{2}\nu_{\mathbf{i}}\mathbf{D}^{\frac{1}{2}}\psi_i
\end{aligned}
$$

□

Because of lamma 4.1 and lemma 3.5, we can obtain the following corollary
corollary: $0\leq \lambda_i(\mathbf{W}) \leq 1$
Warning: W is not symmetric. Thus, its eigenvector need not be orthogonal

5 Convergence of Lazy Random Walk

5.1 Finding an expression for $\mathbf{p}_t$

State transition from $\mathbf{p}_t$ to $\mathbf{p}_{t+1}$ in a lazy random is given by :

$$
\mathbf{p}_{t+1} = \mathbf{W}\mathbf{p}_t
$$

When t = 0:

$$\mathbf{p}_1 = \mathbf{W}\mathbf{p}_0$$

We know that

$$\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0 = \sum_{i=1}^{n} \alpha_i \psi_i \Leftrightarrow \mathbf{p}_0 = \sum_{i=1}^{n} \alpha_i \mathbf{D}^{\frac{1}{2}} \psi_i$$

Thus, we can express $\mathbf{p}_1$ as:

$$\mathbf{p}_1 = \mathbf{W}\mathbf{p}_0 = \sum_{i=1}^{n} \alpha_i (\mathbf{W}\mathbf{D}^{\frac{1}{2}} \psi_i) = \sum_{i=1}^{n} \alpha_i (1 - \frac{\nu_i}{2}) \mathbf{D}^{\frac{1}{2}} \psi_i$$

Iterating the process above, we obtain:

$$\mathbf{p}_t = \sum_{i=1}^{n} \alpha_i (1 - \frac{\nu_i}{2})^{\top} \mathbf{D}^{\frac{1}{2}} \psi_i$$

Claim: If G is connected $\Leftrightarrow \nu_2 > 0$
Remark: the claim above implies the following:

$$\forall i \neq 1, 0 \leq 1 - \frac{\nu_i}{2} < 1$$

5.2 Given ϵ, finding step t such that $\mathbf{p}_t$ is ϵ closed to the stationary distribution

At arbitrary vertex u, we have the following:

$$\begin{aligned} \mathbf{1_u^{\top} P_t} - \mathbf{1_u^{\top}} \pi &= \mathbf{1_u^{\top} P_t} - \frac{\mathbf{1_u^{\top} D 1}}{\mathbf{1^{\top} D 1}} \\ &= \sum_{i=1}^{n} \alpha_i (1 - \frac{\nu_i}{2})^{\top} \mathbf{1_u^{\top}} \mathbf{D}^{\frac{1}{2}} \psi_{\mathbf{i}} - \frac{\mathbf{1_u^{\top} D 1}}{\mathbf{1^{\top} D 1}} \end{aligned} \tag{3}$$

We know that:

$$\psi_1 = \frac{(\mathbf{D}^{\frac{1}{2}}\mathbf{1})}{||\mathbf{D}^{\frac{1}{2}}\mathbf{1}||} = \frac{(\mathbf{D}^{\frac{1}{2}}\mathbf{1})}{\sqrt{\mathbf{1}^{\top}\mathbf{D}\mathbf{1}}}$$

multiplied both side with $\mathbf{1_u^{\top}} \mathbf{D}^{\frac{1}{2}}$, we get

$$\begin{aligned} \mathbf{1_u^{\top}} \mathbf{D}^{\frac{1}{2}} \psi_1 &= \mathbf{1_u^{\top}} \mathbf{D}^{\frac{1}{2}} \frac{(\mathbf{D}^{\frac{1}{2}}\mathbf{1})}{||\mathbf{D}^{\frac{1}{2}}\mathbf{1}||} \\ &= \frac{(\mathbf{1_u^{\top} D 1})}{\sqrt{\mathbf{1}^{\top}\mathbf{D}\mathbf{1}}} \end{aligned} \tag{4}$$

We can express $\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0$ as following

$$\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0 = \sum_{\mathbf{i=1}}^{\mathbf{n}} \alpha_{\mathbf{i}}\psi_{\mathbf{i}}$$

multiply each side with $\psi_1^\top$

$$\psi_1^\top \mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0 = \alpha_{\mathbf{1}}$$

This gives us:

$$\alpha_1 = \frac{(\mathbf{1}^\top \mathbf{D}^{\frac{1}{2}})\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0}{\sqrt{\mathbf{1}^\top \mathbf{D}\mathbf{1}}}$$

because $\mathbf{p}_0$ is a probability vector and sum up to 1, we have,

$$= \frac{1}{\sqrt{\mathbf{1}^\top \mathbf{D}\mathbf{1}}} \tag{5}$$

Now, we can further simplify (3) as:

$$\begin{aligned}
\mathbf{1}_\mathbf{u}^\top \mathbf{p}_\mathbf{t} - \mathbf{1}_\mathbf{u}^\top \pi &= \alpha_1(1 - \frac{\nu_1}{2})^\top \psi_1 + \sum_{i=2}^{n} \alpha_i (1 - \frac{\nu_i}{2})^\top \mathbf{1}_\mathbf{u}^\top \mathbf{D}^{\frac{1}{2}} \psi_i - \frac{\mathbf{1}_\mathbf{u}^\top \mathbf{D}\mathbf{1}}{\mathbf{1}^\top \mathbf{D}\mathbf{1}} \\
&= \sum_{i=2}^{n} \alpha_i (1 - \frac{\nu_i}{2})^\top 1_u^\top \mathbf{D}^{\frac{1}{2}} \psi_i
\end{aligned}$$

Combining the result above, we have the following

$$\begin{aligned}
|\mathbf{1}_\mathbf{u}^\top \mathbf{p}_t - \mathbf{1}_\mathbf{u}^\top \pi| &\le \sum_{i=2}^{n} |\alpha_i \mathbf{1}_\mathbf{u}^\top \mathbf{D}^{\frac{1}{2}} \psi_i| (1 - \frac{\nu_i}{2})^\top \\
&\le (1 - \frac{\nu_2}{2})^\top \sum_{i=2}^{n} |\alpha_i \mathbf{1}_\mathbf{u}^\top \mathbf{D}^{\frac{1}{2}} \psi_i| \\
&\le (1 - \frac{\nu_2}{2})^\top \sqrt{\sum_{i=2}^{n} \alpha_i^2 \sum_{i=2}^{n} (\mathbf{1}_\mathbf{u}^\top \mathbf{D}^{\frac{1}{2}} \psi_i)^2}
\end{aligned} \tag{6}$$

Now let's try to simplify the term inside the square root, starting with $\sum_{i=2}^{n} \alpha_i^2$

$$\begin{aligned}
\sum_{i=2}^{n} \alpha_i^2 &\le ||\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0||_2^2 \\
&\le \mathbf{1}_\mathbf{v}^\top \mathbf{D}^{-1} \mathbf{1}_\mathbf{v} \\
&= \frac{1}{D(v)}
\end{aligned} \tag{7}$$

Now let's simplify $\sum_{i=2}^{n} (1_u^\top D^{\frac{1}{2}} \psi_i)^2$
We shall start with finding an expression for $\mathbf{D}^{\frac{1}{2}}\mathbf{1}_\mathbf{u}$

Claim:

$$\mathbf{D}^{\frac{1}{2}}\mathbf{1_u} = \sum_{i=1}^{n}(\mathbf{1_u}^\top\mathbf{D}^{\frac{1}{2}}\psi_i)\psi_j$$

Proof. Let's express $\mathbf{D}^{\frac{1}{2}}\mathbf{1_u}$ with eigenvector and eigenvalue

$$\mathbf{D}^{\frac{1}{2}}\mathbf{1_u} = \sum_{\mathbf{i=1}}^{\mathbf{n}}\beta_\mathbf{i}\psi_\mathbf{i}$$

$$\psi_\mathbf{j}^\top\mathbf{D}^{\frac{1}{2}}\mathbf{1_u} = \beta_\mathbf{j}$$

$$(\psi_\mathbf{j}^\top\mathbf{D}^{\frac{1}{2}}\mathbf{1_u})^\top = \mathbf{1_u}^\top\mathbf{D}^{\frac{1}{2}}\psi_j$$

$$\begin{aligned}
||\mathbf{D}^{\frac{1}{2}}\mathbf{1_u}||_2^2 &= (\mathbf{D}^{\frac{1}{2}}\mathbf{1_u})^\top(\mathbf{D}^{\frac{1}{2}}\mathbf{1_u}) \\
&= (\sum_{i=1}^{n}\beta_i\psi_i)^\top(\sum_{j=1}^{n}\beta_j\psi_j) \\
&= \sum_{i=1}^{n}\beta_i^2 \\
&= \sum_{i=1}^{n}(\mathbf{1_u}^\top\mathbf{D}^{\frac{1}{2}}\psi_i)^2
\end{aligned} \tag{8}$$

□

With the results above, (6) can be further simplified as following:

$$\begin{aligned}
|\mathbf{1_u}^\top\mathbf{p_t} - \mathbf{1_u}^\top\pi| &\leq (1-\nu_2)^\top\sqrt{||\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0||_2^2||\mathbf{D}^{\frac{1}{2}}\mathbf{1_u}||_2^2} \\
&= ||\mathbf{D}^{\frac{1}{2}}\mathbf{1_u}||\cdot||\mathbf{D}^{-\frac{1}{2}}\mathbf{p}_0||(1-\frac{\nu_2}{2})^\top \\
&= \sqrt{\frac{D_u}{D_v}}(1-\frac{\nu_2}{2})^\top \\
&\leq \sqrt{\frac{D_u}{D_v}}e^{-\nu_2 t/2}
\end{aligned} \tag{9}$$

Theorem 5.1. *Given an undirected unweighted graph G and ϵ, it suffices for the lazy random walk to take t steps to get ϵ closed to stationary distribution.*
In other word, if

$$t \geq \frac{2}{\nu_2}\log(\frac{n}{\epsilon})$$

then

$$|\mathbf{1_u}^\top\mathbf{p}_t - \mathbf{1_u}^\top\pi| \leq \epsilon$$

Interpretation: On a graph with n vertex, if the lazy random walk want to be $\frac{\epsilon}{2}$ closed to the stationary distribution, then it would satisfy the following relation.

$$\begin{aligned} T(\frac{\epsilon}{2}) &= \frac{2}{\nu_2} log(\frac{2n}{\epsilon}) \\ &= \Theta(\frac{1}{\nu_2}) + T(\epsilon) \end{aligned}$$

5.3 Application of the theorem on some examples

K_n:complete graph with n vertices
Lazy random walk mixes in $\Omega(\log n)$ steps

$$\begin{aligned} \mathcal{L}_{K_n} &= n\mathbf{I} - \mathbf{1}\mathbf{1}^\top \\ \lambda_i(L) &= \begin{cases} 0 & i = 1 \\ n & o/w \end{cases} \\ \nu_2(N_{K_n}) &= 1 \end{aligned}$$

the theorem gives that the lazy random walk would mix in $O(\log n)$. In this case, the bound given by the theorem is tight

R_n:n-ring graph
The second eigenvalue of n-ring graph is given by:

$$\nu_2(N_{K_n}) = \theta(\frac{1}{n^2})$$

thus, the theorem gives that the lazy random walk mixes in $O(n^2 \log n)$ steps.
the lazy random walk on the n-ring graph can be defined as following:

$$x_{i+1} = \begin{cases} x_i & \text{with probability} \frac{1}{2} \\ x_i + 1 & \text{with probability} \frac{1}{4} \\ x_i - 1 & \text{with probability} \frac{1}{4} \end{cases}$$

$$E[x_{i+1}|x_i] = x_i$$

$$E[x_{i+1}^2|x_i] = x_i^2 + \frac{1}{2}$$

thus, the Lazy Random Walk is mixed in $\Omega(n^2)$ steps. In this case, the bound given by the theorem is off up to logn