COMS 4995-2: Advanced Algorithms (Spring'20)	Feb 18, 2020
Lecture 9: Random walks, and the largest eigenvalue	
Instructor: *Alex Andoni*	Scribes: *Xingyu Zhu, Yiyang Zeng*

1 Introduction

In the previous lecture,we introduced the random walk, spectral graph theory and rayleigh quotient, today we continue the topic of random walk, we will apply spectral decomposition and drive some properties of diffusion and eigenvalues, and we will investigate the relationship between eigenvalue and graph connectivity.

2 Remainder

- we can apply spectral decomposition to a symmetric graph matrix M : M $= \sum_{i=1}^{n} \lambda_i v_i v_i^T$, where $v_1, v_2, ..., v_n$ are n orthonormal eigenvectors such that $||v_i|| = 1$ and $v_i^T v_j = 0\ \forall i \neq j$.
- rayleigh quotient is defined as R(x)= $\dfrac{x^t M x}{||x||^2}$ and R$(v_i) = \lambda_i\ \forall v_i$.

3 Diffusion/ Random walk

Let $x^0 (\in R^n) =$ distribution over vertices in G.

$$x^t = W(x^{t-1}) = AD^{-1} x^{t-1} (t \geq 1)$$

$$x^t = AD^{-1} \cdot AD^{-1} \cdots x^0 = (AD^{-1})^t \cdot x^0 \tag{1}$$

If (AD^{-1}) is symmetric (eg, when D is diagonal of the same number, equivalently all degrees are exactly equal)
e.g. when d-regular:

$$AD^{-1} = \frac{1}{d} A, M = AD^{-1} = \sum \lambda_i v_i v_i^T$$

$$M^2 = (\sum_i \lambda_i v_i v_i^T) \cdot (\sum_j \lambda_j v_j v_j^T)$$

$$= \sum_i \lambda_i v_i (v_i^T v_i) v_i^T \lambda_i = \sum_{i=1}^{n} \lambda_i^2 v_i v_i^T$$

In general:

$$M^t = \sum_{i=1}^{n} \lambda_i^t v_i v_i^T (\text{by induction})$$

when AD^{-1} is not symmetric, we can analyze a related matrix.

$$\begin{aligned} x^t &= AD^{-1} \cdot AD^{-1} \cdots x^0 = (AD^{-1})^t \cdot x^0 \\ &= AD^{-1/2}D^{-1/2}AD^{-1/2} \cdots D^{-1/2}x^0 \end{aligned}$$

Definition 1. $\hat{A} = D^{-1/2}AD^{-1/2}$

Note: $\hat{A_{i,j}} = \frac{A_{i,j}}{\sqrt{d_i}\sqrt{d_j}}$

Definition 2. $y^t = D^{-1/2}x^t$

From the definition 1,2 and equation(1),we have:

$$y^t = \hat{A}^t \cdot y^0. \tag{2}$$

Apply spectral decomposition for $\hat{A}$:

$$\begin{aligned} \hat{A} &= \sum \lambda_i v_i v_i^T, \hat{A}^t = \sum \lambda_i^t v_i v_i^T \\ y^0 &= \sum_{i=1}^{n} \alpha_i v_i, \alpha_i \in R, ||y^0||_2^2 = \sum \alpha_i^2 \\ y^t = \hat{A}^t y^0 &= (\sum_{i=1}^{n} \lambda_i^t v_i v_i^T)(\sum_{j=1}^{n} \alpha_j v_j) \\ &= \sum_{i,j} \lambda_i^t v_i v_i^T \alpha_j v_j \\ &= \sum_{i} \lambda_i^t \alpha_i v_i \end{aligned}$$

Claim 3.

$$\begin{aligned} &\textit{If } \lambda_i < 1 \implies \textit{term } \lambda_i^t \alpha_i v_i \textit{ disappears as } t \to \infty \\ &\textit{If } \lambda_i > 1 \implies \lambda_i^t \textit{ diverges} \end{aligned}$$

Thus, intuitively, we must have:

$$\begin{aligned} &1)|\lambda_i| \leq 1 \quad \forall i \\ &2)|\lambda_i| = 1 \quad \exists i \end{aligned}$$

Theorem 4. $\lambda_1 = 1$

Proof. 1) $\lambda_1 \geq 1$ (recall $\lambda_1 = \max \ R(x)$ $(x \neq 0)$)

Let $x = (\sqrt{d_1}, \sqrt{d_2}, ..., \sqrt{d_n})^T$:

$$
\begin{aligned}
R(x) &= \frac{x^T \cdot \widehat{A} \cdot x}{\|x\|_2^2} \\
&= \frac{x^T \cdot D^{-1/2} \cdot A \cdot D^{-1/2} \cdot x}{\sum_{i=1}^n d_i} \\
&= \frac{\mathbb{1}_n^T \cdot A \cdot \mathbb{1}}{\sum_{i=1}^n d_i} \\
&= \frac{\mathbb{1}_n \cdot [d_1, d_2, ..., d_n]^T}{\sum_{i=1}^n d_i} \\
&= 1
\end{aligned}
$$

$\implies \lambda_1 \geq R(x) = 1$
2) $\lambda_1 \leq 1 \Leftrightarrow R(x) \leq 1 \ \forall x \neq 0$

$$
\begin{aligned}
R(x) &= \frac{x^T \widehat{A} x}{\|x\|_2^2} \\
&= \frac{\sum_{ij \in E} \frac{A_{ij}}{\sqrt{d_i}\sqrt{d_j}} \cdot x_i x_j}{\|x\|_2^2} \\
&= \frac{\sum_{ij \in E} \frac{x_i}{\sqrt{d_i}} \cdot \frac{x_j}{\sqrt{d_j}}}{\|x\|_2^2} \\
&\leq \frac{(\sum_{ij \in E} (\frac{x_i}{\sqrt{d_i}})^2)^{1/2} \cdot (\sum_{ij \in E} (\frac{x_j}{\sqrt{d_j}})^2)^{1/2}}{\|x\|_2^2} (Cauchy - Schwartz) \\
&= \frac{(\sum_i \frac{x_i^2}{d_i} \cdot d_i)^{1/2} \cdot (\sum_j \frac{x_j^2}{d_j} \cdot d_j)^{1/2}}{\|x\|_2^2} \\
&= 1
\end{aligned}
$$

Cauchy-Schwartz: $p^T \cdot q \leq \|p\|^2 \cdot \|q\|^2$ □

Remarks:

1. since $v_1 = argmax\ R(x) \Rightarrow v_1 = (\sqrt{d_1}, \sqrt{d_2}, ..., \sqrt{d_n})$ (unique iff $\lambda_2 < 1$)
2. $R(x) = 1$ iff CS is tight for x iff $\exists \alpha > 0$ s.t. $\frac{x_i}{\sqrt{d_i}} = \frac{x_j}{\sqrt{d_j}} \cdot \alpha$, $\frac{x_i}{\sqrt{d_i}} \cdot \alpha = \frac{x_j}{\sqrt{d_j}}$ (vector p is a rescale of vector q)
   $\Rightarrow \alpha = +1$
   if $\alpha = +1 \Rightarrow \frac{x_i}{\sqrt{d_i}} = \frac{x_j}{\sqrt{d_j}} \ \forall ij \in E$
   Similarly $R(x) \geq -1$ always (using $p \cdot q \geq -\|p\| \cdot \|q\|$), and $R(x) = -1$ iff the above holds with

$\alpha = -1$ ($\dfrac{x_i}{\sqrt{d_i}} = -\dfrac{x_j}{\sqrt{d_j}}$)
Applying to our diffusion/random walk example: Recall $y^t = \sum \alpha_i \lambda_i^t v_i$.
Given $\lambda_1 = 1$ if $|\lambda_2|, ..., |\lambda_n| < 1 \Rightarrow \lim_{t\to\infty} y^t = \alpha_1 \cdot \lambda_1^t \cdot v_1 = \alpha_1 v_1$
where $v_1 = \dfrac{1}{\sum_{d_i}}(\sqrt{d_1}, \sqrt{d_2}, ..., \sqrt{d_n})$ is the unique eigenvector with λ_1)
Note that also $\alpha_1 = v_1^T \cdot y^0 = \sum_i x_i^0 = 1$ (by the assumption that x^0 is distribution). Hence $\lim_{t\to\infty} x^t = D^{1/2} v_1 = \begin{pmatrix} d_1 \\ d_2 \\ ... \\ d_n \end{pmatrix} \cdot \dfrac{1}{\sum d_i}$ is the unique stationary distribution in this case.

Theorem 5. $\lambda_2 < 1 \Leftrightarrow G$ *connected*

Proof. G is disconnected $\Rightarrow \lambda_2 = 1$

Let's say A has two components: comp1 and rest as below:

$$A = \begin{pmatrix} comp1 & 0 \\ 0 & rest \end{pmatrix}$$

comp1 will act independently of rest meaning each component has its own eigenvalues and eigenvectors, so each component will have its own top eigenvalue,which is 1 as proven above. Say comp1 is composed of k nodes {1,2,...,k}, we can write $v_1 = (\sqrt{d_1}, ..., \sqrt{d_k}, 0, ..., 0), v_2 = (0, ..., 0, \sqrt{d_{k+1}}, ..., \sqrt{d_n})$.

Prove in other direction: If G is connected $\Rightarrow \lambda_2 < 1$

take $v_2 \perp v_1$ ($v_1 = \sqrt{d_1}, ..., \sqrt{d_n}$), prove by contradiction:

$$\begin{aligned} \text{suppose } R(x = v_2) = 1 &\Rightarrow \frac{x_i}{\sqrt{d_i}} = \frac{x_j}{\sqrt{d_j}} \ \forall (i,j) \\ &\Rightarrow \frac{x_i}{\sqrt{d_i}} = \beta \text{ where } \beta = \frac{x_1}{\sqrt{d_1}} \ \ \forall i \\ &\Rightarrow \text{x proportional to } v_1 \text{ (contradicted with } x \perp v_1) \end{aligned}$$

where the middle step is since there's a path from node 1 to each other node, and we can apply the above equality to each edge on this path. ∎