COMS 4995-3: Advanced Algorithms — March 6, 2017

Lecture 13 – Spectral Graph Algorithms

Instructor: *Alex Andoni* — Scribe: *Srikar Varadaraj*

1 Introduction

Today's topics:

- Finish proof from last lecture
- Example of random walk, stationary distributions
- Linear Algebra overview - Spectral theorem
- Define Rayleigh quotient, prove a property

2 Shortest Augmenting Path Algorithm - Max Flow

- To Recap, the shortest augmenting path algorithm always chose the shortest remaining augmenting path in the residual graph G_f. We want to bound its running time.
- Every iteration takes time $O(m)$ by BFS.
- We need to answer: How many iterations do we need?

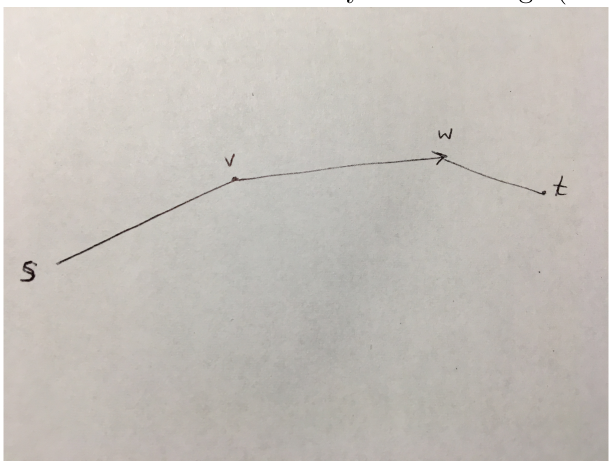
Definition 1. *Fix the current flow f. The residual graph is denoted as G_f. We define $d_f(s, v)$ = distance from s to v in G_f.*

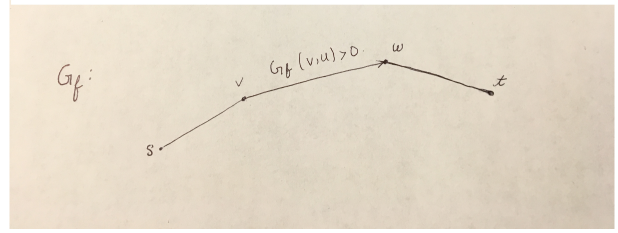
Claim 2. *Let P be the shortest $s \to t$ augmenting path in G_f. Let f' be f after augmenting with the path P. Let $d'(s, v) = d_{f'}(s, v)$.*
Then, we claim that $d'(s, v) \geq d(s, v) \forall v$.

Proof. If $d_f(s, t) \geq n$ then we are done trivially.
We would like to prove that $d_f(s, t)$ increases by 1 every time we augment by the shortest path. However, we will not be able to show this. Instead, we assert and prove the following lemma:

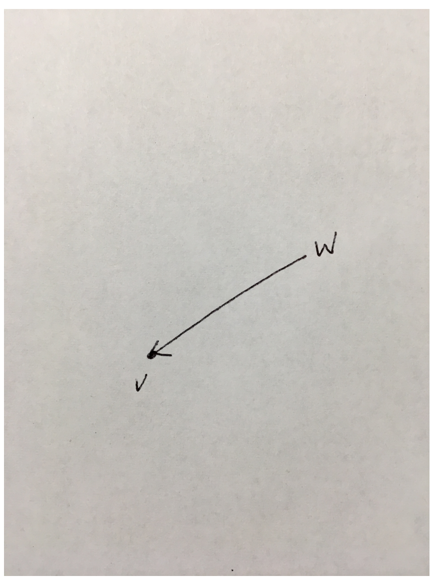
Lemma 3. *We use $\leq \frac{nm}{2}$ augmenting paths until we are done, and every edge is saturated.*

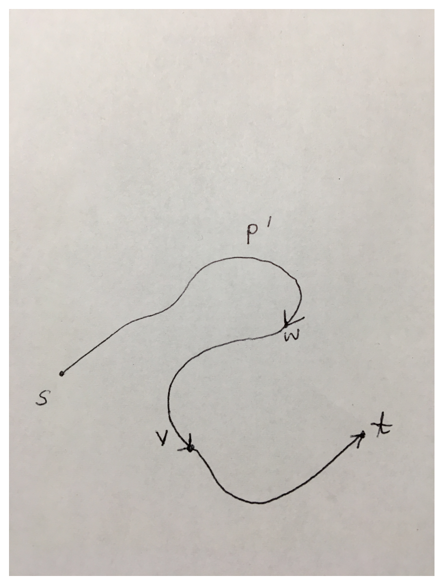
- We want to find out how many times an edge $(v - w)$ in the graph can be saturated.





- Let $d(s, v)$ = distance to v from s before saturation.
- Before augmentation, $d(s, w) = d(s, v) + 1$.
- Note that before any edge $v - w$ is saturated again, the following situation must occur:





- Now consider distances $d'(s, w)$ and $d'(s, v)$.
- Then, $d'(s, v) \geq d'(s, w) + 1 = d(s, v) + 2$ because P' is saturated.
- So, distance increases by 2 through any augmenting path.
- Therefore, every edge $v - w$ can be saturated $\leq \frac{n}{2}$ times. Since there are m total edges, the total number of augmenting paths is upper bounded by $\frac{mn}{2}$. After these many augmenting paths, all the edges must be completely saturated.

□

The above procedure must be repeated for every edge, so the running time is $O(m * \frac{mn}{2}) = O(m^2 n)$, as required.

- If we do "more work" per iteration of the algorithm, we can decrease the number of iterations. The best known current algorithm is: $\tilde{O}(m\sqrt{n})$, where $\tilde{O}(n) = n(logn)^{O(1)}$.
- For capacity $U = O(1)$, the best known algorithm is $\tilde{O}(m^{10/7})$.

3 Spectral Graph Theory

Observation 4. *The adjacency matrix of an undirected graph G, denoted by A_G, has $A_{ij} = 1$ iff $\exists$ edge $i \to j$. The adjacency matrix of an undirected graph is symmetric.*

We only consider undirected graphs in this lecture since the theory we develop works well for them. Note that for directed matrices, the adjacency matrix is not symmetric.

3.1 Motivation

We begin the discussion on adjacency matrices and spectral graph theory with a few definitions and a motivating example.

Definition 5. *The Diffusion operation D is defined on a graph to have the following properties:*

- *Fix a mass vector $x \in \mathbb{R}^n$, which assigns a weight or mass, to every vertex in a graph G at time $t = 0$.*
- *Time discrete step.*
- *$\forall i \in [n]$, x_i is distributed equally amongst its neighbors. So, if vertices with indices $a_1, ..., a_k$ are adjacent to the vertex with index i, then $x_{a_j} \to x_{a_j} + \frac{x_i}{k}$, where x_{a_j} denotes the mass at vertex with index a_j.*
- *Note that we assume no vertex has an edge to itself.*

Definition 6. *Matrix D_G is defined for a graph G as:*

$$\begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{bmatrix}$$

Here d_i is the degree of node x_i.

Example 7. *Consider the graph* $G : 1-2-3-4-5$.
A_G *is:*

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

D_G *is:*

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $x = (0, 1, 0, 0, 0)^T$ *at time* $t = 0$. *We apply the diffusion operator* D *repeatedly.*
At time $t = 1$, $x_1 = (\frac{1}{2}, 0, \frac{1}{2}, 0, 0)$. *Now, note that* $A_G D_G^{-1} x = x_1$.
Also, time $t = 2$, $x_2 = (0, \frac{1}{2} + \frac{1}{4}, 0, \frac{1}{4}, 0) = (0, \frac{3}{4}, 0, \frac{1}{4}, 0)$.
$D_G^{-1} x_1 =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{4} \\ 0 \\ 0 \end{bmatrix}$$

And so, $A_G D_G^{-1} x$*:*

$$\begin{bmatrix} 0 \\ \frac{3}{4} \\ 0 \\ \frac{1}{4} \\ 0 \end{bmatrix}$$

which is simply x_2.
Hence, we see that $A_G D_G^{-1} x_t = x_{t+1}$
Also note that the distribution $x^* = (\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8})$ *is a "stationary" distribution in the sense that* $x^* = A_G D_G^{-1} x^*$.

The example above motivates our discussion on spectral graph theory, since knowing properties of the matrices A_G and D_G^{-1} easily gives us the value of x_t, as well as properties of stationary distributions (which we did not discuss in depth). We begin with a review of the basic linear algebra required.

3.2 Spectral Theorem and Spectral Decomposition

Definition 8. v *is an eigenvector of the matrix* $M_{n \cdot n}$ *if* $Mv = \lambda v$ *for some* $\lambda \in \mathbb{R}$. λ *is referred to an eigenvalue of the matrix* M.

Observation 9. $Mv = \lambda v \iff (M - \lambda I)v = 0 \iff det(M - \lambda I) = 0$. *Hence, all the eigenvalues of* M *satisfy the equation* $|M - \lambda I| = 0$.

- $|M - \lambda I| = 0$ *is a polynomial of degree n.*
- *This implies that there are n eigenvalues counted with multiplicity.*

The following theorem gives us a special property of symmetric matrices, which is especially relevant to adjacency matrices of undirected graphs.

Theorem 10. *Spectral Theorem*
For symmetric matrix M, $\exists n$ orthonormal eigenvectors $v_1, ..., v_n$ such that $\|v_i\| = 1$ and $v_i^T v_j = 0 \; \forall i \neq j$. The eigenvector v_i has corresponding eigenvalue λ_i.

Some observations:

- If $\lambda_i = \lambda_{i+1}$, then $\frac{v_i + v_{i+1}}{\sqrt{2}}$ is also an eigenvector with eigenvalue λ_i.
- The Spectral Theorem implies that there exists a Spectral decomposition:

$$M = \sum_{i=1}^{n} \lambda_i v_i v_i^T \tag{1}$$

3.3 Rayleigh quotient and properties

Definition 11. *The Rayleigh quotient is defined as* $R(x) = \frac{x^t M x}{\|x\|^2}$

Observation 12. *Notice that $\forall$ eigenvectors v_i,*

$$R(v_i) = \frac{v_i^T M v_i}{1} = v_i^T \lambda_i v_i = \lambda_i v_i^T v_i = \lambda_i \|v_i\| = \lambda_i \tag{2}$$

Let $x^* = argmax\ R(x)$, or a vector that maximizes the Rayleigh quotient.

Theorem 13. *Let the eigenvalues be ordered as: $\lambda_n \geq \lambda_{n-1} \geq \lambda_{n-2} ... \geq \lambda_1$. Then, $R(x^*) = \lambda_n$.*

Proof. Every vector $v = \sum_{i=1}^{n} a_i v_i$, since $\{v_i\}$ forms a basis. Then,

$$R(v) = \frac{(v^T)M(v)}{\|v_i\|} = \frac{(\sum_{i=1}^{n} a_i v_i)^T M (\sum_{i=1}^{n} a_i v_i)}{a_1^2 + ... + a_n^2} = \frac{(\sum_{i=1}^{n} a_i v_i)^T (\sum_{i=1}^{n} a_i \lambda_i v_i)}{a_1^2 + ... + a_n^2} = \frac{\sum_{i=1}^{n} \lambda_i a_i^2}{a_1^2 + ... + a_n^2} \leq \frac{\sum_{i=1}^{n} \lambda_n a_i^2}{a_1^2 + ... + a_n^2} = \lambda_n \tag{3}$$

Hence, $R(v) \leq \lambda_n$. Similarly, we can also conclude that $R(v) \geq \lambda_1$. $\square$

Hence, the eigenvalues of the matrix M give us deep properties regarding the Rayleigh quotient. We will see in later lectures how the Rayleigh Quotient is useful and relate it to properties of graphs.