

Very short intro to Spectral Graph Theory

Edo Liberty
Algorithms in Data mining

1 Introduction

Up until this point, we encountered a few different data types. These were either arbitrary elements, or elements in a metric space, vectors in a normed space, or matrices. Another very common type of data is a Graph. In many different scenarios a graph is the most natural representation of the data. Example include communication networks, transportation routes, online social networks etc. In these cases, a lot of information is captured by the connections (and their strengths) between the different elements. For example, the graph describing the physical layout of the internet infrastructure defines how well messages can pass or how robust it is to congestion. Or, friendship connections between users in a social network can offer insights to their topics of interest.

In this class we will review three ways to represent a graph as a matrix. The Adjacency matrix, the Markov transition matrix, and the Graph Laplacian. Relating properties of graphs to properties of the matrices which represent them is researched in a subfield of computer science called Spectral Graph Theory. Here, we will barely scratch the surface of this vast literature. The aim is to provide some intuitions and access to the terminology.

2 The adjacency matrix

One of the most natural matrices associated with a graph $G(V, E)$ is the adjacency matrix. It is defined by $A(i, j) = 1$ if $\{i, j\} \in E$ and zero else.

Fact 2.1. *Let α_1 be the largest eigenvalue of A . Then, $\alpha_1 \leq \max_i \deg(i)$*

Proof. Given in class. □

Fact 2.2. *Let α_1 be the largest eigenvalue of A . Then, $\alpha_1 \geq \frac{1}{n} \sum_i \deg(i)$*

Proof. Given in class. □

Fact 2.3. *Let β_1 be the largest eigenvalue of B which is the adjacency graph of G with one node removed. Then, $\beta_1 \leq \alpha_1$.*

Proof. Given in class. □

Fact 2.4. Let α_1 be the largest eigenvalue of A . Then, the graph G is $\lfloor \alpha_1 \rfloor + 1$ colorable.

Proof. Given in class. □

3 The Markov transition matrix

The adjacency matrix also defines the diffusion or Markov transition matrix of G . Let $D = \text{diag}(A\mathbf{1}_n)$, that is $D(i, i) = \sum_j A(i, j)$. We define the transition matrix $M = AD^{-1}$. The columns of the matrix M sum up to 1. The entries of M can be seen as giving transitional probability between nodes in the graph. That is, let p_t denote the distribution over nodes in time t then:

$$p_{t+1} = Mp_t$$

Let us see that there exists a stationery distribution p_∞ .

$$p_\infty = \lim_{t \rightarrow \infty} p_t = \lim_{t \rightarrow \infty} M^t p_0$$

Our goal is to understand the structure of $\lim_{t \rightarrow \infty} M^t$.

Fact 3.1. Let d be the vector such that $d(i) = \text{deg}(i) / \sum_j \text{deg}(i)$. If the graph G is connected and irreducible then $\lim_{t \rightarrow \infty} M^t = d\mathbf{1}^T$.

Before we show that let us see what that means.

$$p_\infty = \lim_{t \rightarrow \infty} p_t = \lim_{t \rightarrow \infty} M^t p_0 = d\mathbf{1}^T p_0 = d$$

That is, regardless of the original distribution, after a sufficient number of steps we arrive at a terminal distribution which remains unchanged. Surprisingly, the probability of being at node i is relative only to its degree $\text{deg}(i)$ and does not depend on the structure of the graph.

Let's turn to proving fact 3.1. First, note that M is congruent to the symmetric matrix $D^{-1/2}AD^{-1/2}$

$$M \sim D^{-1/2}MD^{1/2} = D^{-1/2}AD^{-1/2} \equiv \tilde{A}$$

Where \tilde{A} is the normalized adjacency matrix. That means that M and \tilde{A} share the same eigenvalues and that their eigenvectors are related.

$$f^T M = \lambda f^T \rightarrow f^T D^{1/2} D^{-1/2} M D^{1/2} = \lambda f^T D^{1/2} \rightarrow (f^T D^{1/2}) \tilde{A} = \lambda (f^T D^{1/2})$$

Which means that λ is also an eigenvalue of \tilde{A} which corresponds to eigenvector $f^T D^{1/2}$.

Fact 3.2. The top eigenvalue of M is 1 and the all other eigenvalues of M are strictly smaller than 1 if the graph G is connected and irreducible (which we will not define here).

Proof. Given in class. □

Due to the above and since $\mathbf{1}^T M = \mathbf{1}^T$ we know that the top eigenvalue of \tilde{A} is 1 and its top eigenvector is proportional to $D^{1/2}\mathbf{1}$. Since it must be normalized (eigenvectors have norm 1) it must be $d^{1/2}$ where that means that $d^{1/2}(i) = \sqrt{\text{deg}(i)/\sum_j \text{deg}(i)}$. Moreover, since all other eigenvalues of \tilde{A} are strictly smaller than 1 we get that

$$\lim_{t \rightarrow \infty} \tilde{A}^t = \mathbf{1}^t \cdot (d^{1/2})(d^{1/2})^T$$

Combining with the fact that:

$$M^t = D^{1/2} \tilde{A}^t D^{-1/2}$$

we get that

$$\lim_{t \rightarrow \infty} M^t = D^{1/2} \tilde{A}^t D^{-1/2} = D^{1/2} (d^{1/2})(d^{1/2})^T D^{-1/2} = d\mathbf{1}^T.$$

4 The graph Laplacian

The most natural matrix representing a graph (in my opinion) is the Graph Laplacian Matrix which is defined as:

$$L = D - A$$

The graph Laplacian is positive semidefinite and symmetric matrix which encapsulates many different properties of the graph.

Fact 4.1. *The graph Laplacian L is positive semidefinite.*

Proof. Let us define $L_{i,j}$ as an $n \times n$ matrix whose entries are all zero apart for $L_{i,j}(i,i) = L_{i,j}(j,j) = 1$ and $L_{i,j}(i,j) = L_{i,j}(j,i) = -1$. We have that $L = \sum_{\{i,j\} \in E} L_{i,j}$. That means that if $L_{i,j}$ is positive then so is L . computing $x^T(L_{i,j})x = (x_i - x_j)^2 \geq 0$ for all x . Therefore, $x^T L x \geq 0$ for all x which completes the proof. □

The vector $\mathbf{1}_n$ exhibits $\mathbf{1}_n^T L \mathbf{1}_n = 0$. That means that L has at least one eigenvalue 0.

Fact 4.2. *If the graph G is connected then all other eigenvalues of L are strictly positive.*

Proof. Given in class □

Fact 4.3. *The multiplicity of the eigenvalue 0 is equal to the number of connected components in G .*

Proof. Given in class □

The second eigenvector of L , the one corresponding to the second smallest eigenvalue (λ_2) is an extremely important parameter of any graph. It tells us how well connected it is. We unfortunately cannot explain this here. The corresponding vector is called the Fiedler vector and is heavily used in spectral clustering. I will try to explain more in class if we have time.