

# Seminar series in Spectral graph analysis and applications

Spring, 2011

## Introduction

The seminars will cover topics in spectral graph theory based on a selection of papers. The overview material is adapted from lecture notes by prof. Daniel A. Spielman at Yale University, <http://www.cs.yale.edu/homes/spielman/>

## 1 Summary of seminar, February 4<sup>th</sup>

### 1.1 Graph matrices

We study the graph  $G = (V, E, w)$  with a set of nodes  $V$ , a set of edges  $E$  and weights  $w$  corresponding to the edges. For the special case of an unweighted graph, all weights are one. The graph is studied through the adjacency matrix  $A_G$ , the degree matrix  $D_G$  and the laplacian matrix  $L_G$ .

$$A_G(i, j) = \begin{cases} w_{i,j} & i, j \in E \\ 0 & otherwise \end{cases} \quad (1)$$

$$D_G(i, i) = \sum_j A_G(i, j) \quad (2)$$

$$L_G = D_G - A_G \quad (3)$$

## 1.2 Connectivity

We noted that  $L_G$  is positive semidefinite and showed further that constant vectors are eigenvectors corresponding to the eigenvalue  $\lambda_1 = 0$ .

**Theorem 1.** *Let  $L_G$  be a laplacian matrix of the graph  $G(E, V)$  with eigenvalues  $\lambda_1 = 0 \leq \lambda_2 \leq \dots \leq \lambda_n$ . Then  $\lambda_2 > 0$  if and only if  $G$  is connected.*

## 1.3 Algebraic connectivity, $\lambda_2$

Through the previous and the next theorem two properties of the second eigenvalue of the laplacian matrix can be stated:

1.  $\lambda_2 = 0 \Rightarrow$  Disconnected graph.
2.  $\lambda_2 > k > 0$ . The graph is better connected with high than with low values of  $k$ .

**Theorem 2.** *Let  $S$  be a subset of the vertices in a graph  $G$ .  $S \subset V$ . The relative size of the subset is given by  $\sigma = \frac{|S|}{|V|}$ . We name the boundary of the subset*

$$\delta(S) \{(i, j) \in E : i \in S, j \notin S\}.$$

*Then the following inequality holds:*

$$|\delta S| \geq \lambda_2 |S| (1 - \sigma)$$

## 2 Summary of seminar, February 11<sup>th</sup>

We introduced two related measures of the size of a cut  $S$  in a graph: Conductance and sparsity. Conductance is given by:

$$\phi(S) = d(V) \frac{|\delta(S)|}{d(S)d(\bar{S})},$$

where  $d(S) = \sum_{i \in S} d(i)$  is the sum of the degrees in  $S$ . Sparsity is defined similarly as:

$$sp(S) = \frac{|\delta(S)|}{\min(d(S), d(\bar{S}))}$$

The conductance of the entire graph is defined as the minimum conductance over all possible cuts and often named the Cheeger constant. Sparsity of the graph can be expressed in the same manner. We noted that conductance and the sparsity of a graph never differs by more than a factor of two. We interpreted the measures as describing how easy a graph may be cut into two almost equally large parts.

We looked at an example with the ring graph with  $N$  nodes. The graph is easily divisible into two similar sized parts. We illustrated this by showing that the conductance of the ring graph goes to zero as  $N$  gets bigger.

### 2.1 Cheegers inequality

The general lower bound of  $\lambda_2$  is known as Cheegers inequality. Spielman (and others) describe this bound in terms of the normalized laplacian  $N$  rather than the regular laplacian from equation (3). The bound is expressed in terms of the graph conductance:

**Theorem 3.** *Let  $G$  be a graph with normalized laplacian  $N_G = D_G^{-\frac{1}{2}} L_G D_G^{-\frac{1}{2}}$  with corresponding eigenvalues  $\{\eta_i\}$  and let  $\phi_G$  be the conductance of the graph. Then*

$$\eta_2 \geq \frac{\phi_G^2}{8}.$$

We started looking at the details of the quite technical proof. The proof is in several stages. We mainly exploit the inequality given by any vector  $x$  orthogonal to the first eigenvector (constant) of the Laplacian,  $\frac{x^T N_G x}{x^T x} \geq \eta_2$ , and the fact that the elements of such a vector may be rearranged in any manner since the order of the vector elements only refer to the node labeling. The size of set  $S$  and  $\bar{S}$  itself is introduced via  $d(i)$  and it was shown that there exists an  $S$  satisfying the given inequality. For the (many) details of the proof one is referred to the webpage of prof. Spielman.

### 3 Summary of seminar February 25<sup>th</sup>

Most of the algebraic graph theory is about describing the graph through the eigenvalue spectrum on the graph laplacian. In this seminar we will look at some relations between the laplacian matrix and the adjacency matrix. The eigenvalues of the adjacency matrix will become useful in the definition of expander graphs.

#### 3.1 Complete and d-regular graphs

A complete graph  $K_n = \{E, V\}$  is a graph of  $n$  nodes where all nodes are connected to all other nodes. We proved that the eigenvalue spectrum of  $L_{K_n}$  simply consists of 0 with multiplicity one (as shown in seminar 1) and  $n$  with multiplicity  $n - 1$ .

We looked at a more general group of graphs, the d-regular graphs where every node has degree exactly  $d$ . If we let  $G$  be such a graph then the laplacian of  $G$  is simply:

$$L_G = D_G - A_G = dI - A_G,$$

which gives us the following simple relation between the eigenvalues of  $A_G \{\alpha_i\}$  and the eigenvalues of  $L_G$ :

$$\alpha_i = d - \lambda_i.$$

Note that the eigenvalues  $A$  is ordered  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$ .

For a general graph  $G$  we proved that

$$\alpha_1 \leq d_{max}$$

where  $d_{max}$  is the maximum degree in  $G$ .

#### 3.2 Expander graphs

In short, expander graphs are sparse, but well connected graphs. More precisely: Expanders are d-regular graphs with  $\lambda_2$  bounded away from zero by some constant. We showed that this is equivalent to

$$|\alpha_i| \leq \epsilon d$$

for  $i \geq 2$ . These graphs turn out to be important in many applications. For an in depth discussion on some of the application I refer to Spielman's homepage.

#### 3.3 Expander approximations of complete graphs

We introduced a graph inequality

$$A \succeq B$$

being equivalent to

$$v^T L_A v \geq v^T L_B v, \quad \forall v.$$

We defined further how a graph  $H$  is an  $\epsilon$ -approximation of a graph  $G$  when:

$$(1 - \epsilon)H \succeq G \succeq (1 + \epsilon)H.$$

Note that for a weighed graph  $H$ ,  $cH$  is defined to be the graph  $H$  with all the weights multiplied by the constant  $c$ .

Our main result from this discussion was that the  $d$ -regular expander  $G$  is an  $\epsilon$ -approximation of the scaled complete graph  $H = \frac{d}{n}K_n$ . This implies that  $\|L_G - L_H\| \leq \epsilon d$ . We interpreted graph approximations as making a graph sparse while keeping the properties of the graph. In particular the properties related to conductance.

### 3.4 Tanner's theorem

We stated without proof tanners theorem which describes how a subset of nodes in an expander graph always has many neighbors.

**Theorem 4.** *Let  $G = (V, E)$  be a  $d$ -regular graph that  $\epsilon$ -approximates  $\frac{d}{n}K_n$ .*

$$N(S) \geq \frac{|S|}{\epsilon^2(1 - \alpha) + \alpha} \quad \forall \quad S \subseteq V.$$

Where  $|S| = \alpha n$  and  $N(S)$  is the set of nodes that are neighbors of nodes in  $S$ .  $\alpha$  can also be seen as the probability that a given node in  $V$  is in the subset  $S$

## 4 Summary of seminar March 8<sup>th</sup>

### 4.1 Random walks

Let  $G = (V, E, w)$  be a weighted graph. The weights on the graph represent how easy it is to walk on the given edge. The probability of being at node  $u$  on time  $t$  of the walk is given by  $p_t(u)$ . Hence assuming that the walk starts at a given node the walk may be described as

$$p_{t+1}(u) = \sum_{v:(u,v) \in E} \frac{w(u,v)}{d(v)} p_t(v),$$

where  $d(v)$  is the weighted degree. We saw that the equivalent matrix expression is:

$$p_{t+1} = A_G D_G^{-1} p_t$$

where  $p_t$  this time is a vector.

### 4.2 Lazy random walks

We also introduced the lazy random walk, which has a 50/50 probability of staying put on each step. This lazy walk is guaranteed to converge to a steady state. On matrix form it given by:

$$p_{t+1} = W_G p_t,$$

with

$$W_G = \frac{1}{2}(I + A_G D_G^{-1}).$$

We studied some properties of  $W_G$  and stated that it had a unique largest eigenvalue  $\omega_1 = 1$ .

### 4.3 Node centrality and stable distribution

In the limit (a very long walk) each node is visited with a probability of its weighted degree. This vector of probabilities is named  $\pi$ .

$$\pi = \frac{d}{\sum_j d(j)}$$

This is the stable distribution of the walk, thus

$$W_G \pi = \pi.$$

. We ended by stating a theorem about the convergence properties of the graph.

**Theorem 5.** *Assume that a walk on  $G = (V, E, w)$  starts on node  $a \in V$ . For every node  $b$  we make bound on how far  $p_t(b)$  can be from  $\pi(b)$ .*

$$|p_t(b) - \pi(b)| \leq \sqrt{\frac{d(b)}{d(a)}} \omega_2^t$$

where  $\omega_2$  is the second largest eigenvalue of  $W_G$ .

We compared this to the theory on the convergence of the power method on  $W_G$ .

## 5 Summary of seminar March 15<sup>th</sup>

In this seminar we addressed some properties of planar graphs. These graphs have numerous applications. (Note that the pixel grid in an image is a planer graph).

### 5.1 Planar graphs

A planar graph can roughly be defined as a graph that may be represented in the plane such that non of the edges cross. A planar graph with maximum number of edges form a triangle grid.

### 5.2 Graph embeddings

A graph embedding is a spatial representation of a graph. Note that the normal definition of a graph,  $G = (V, E, w)$  contains no such spatial information.

A common way of representing a graph in the plane is by its spectral embedding. This means that spatial coordinates are given by the second and third eigenvectors of the laplacian matrix.

### 5.3 Planear separators

We have stated earlier that the second eigenvalue of the graph laplacian decides if a sparse cut is present in the graph. For planer graphs we may improve the upper bound on  $\lambda_2$  by the following theorem:

**Theorem 6.** *Let  $G$  be a planar graph with  $n$  nodes of maximum degree  $d$  and with  $\lambda_2$  defined as in the previous seminars. Then*

$$\lambda_2 \leq \frac{8d}{n}.$$

Note that in a spectral embedding of a planer graph a possible bottleneck in the graph (sparse cut) is placed in the center of the first axis.

## 6 Summary of seminar March 22<sup>th</sup>

The seminar was based on a paper on applications of spectral graph properties in classification of graphs [3]. The main problem addressed was how to compare and cluster a set of several graphs of different size and shape.

### 6.1 Eigen modes

The basis for the analysis in this paper is the eigen decomposition of the adjacency matrix of a graph,

$$A_G = \Phi \Lambda \Phi^T.$$

The eigen modes of the graph is defined to be represented as subgraphs with the same node set, but with adjacency matrices

$$S_G^\omega = \phi_G^\omega \phi_G^{\omega T}.$$

The eigen mode numbering is given by  $\omega = 1, 2 \dots m$  where  $m$  is less than the graph size. We define  $\Phi_G$  to be the  $m$  first columns of  $\Phi$ .

### 6.2 Spectral patterns

A spectral pattern is extracted from each graph by measuring graph properties of each eigen mode. The graph properties have been discussed in previous seminars e.g. total graph degree, size of subgraph perimeter and Cheeger number. For a given graph property of the  $m$  eigen modes of a graph constructs a pattern vector  $x_G \in \mathbb{R}^m$ .

When comparing  $n$  different graphs, the data matrix

$$X_i = [x_{G_1} x_{G_2} \dots x_{G_n}]$$

spans a  $m$  dimensional patterns space for a graph property  $i$ .

### 6.3 Clustering in pattern space

The sorting or clustering of the graphs is performed in a lower dimensional approximation of the patterns space. The dimensional reduction is performed by projecting the data matrix to the plane using principal components, independent components or other projection methods.

## 7 Summary of seminar March 29<sup>th</sup>

The material in this seminar was from a paper by Shi and Malik [4]. In the paper they introduce normalized cuts. We linked this to our previous seminars on sparsity and conductance.

### 7.1 Normalized cuts

We have in previous seminars defined a cut  $S$  to be the total weight of the edges going from one subregion of the graph to its complimentary part. In the notation of this paper that is:

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v),$$

where  $A$  and  $B$  are subsets of the vertex set  $V$ .  $A \cup B = V$   $A \cap B = \emptyset$ . Similarly the association of a vertex subset is given by:

$$assoc(A) = \sum_{u \in A, v \in V} w(u, v),$$

and represents the total connection from the nodes in  $A$  to the rest on the graph.

A general graph will typically have a minimum cut containing only one ore a few nodes. Many applications are looking for cuts that will divide the graph into larger subregions. We have seen this described by graph sparsity. In the paper by Shi and Malik a similar measure us introduced called normalized cuts:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A)} + \frac{cut(A, B)}{assoc(B)}.$$

Solving for a minimal  $Ncut$  turns out to be a NP-complete. We described how a relaxation of the optimization terminates in a eigenvalue problem quite similar to the one we have seen with the normalized laplacian.

$$D^{-\frac{1}{2}}LD^{-\frac{1}{2}}z = \lambda z$$

where  $y = D^{\frac{1}{2}}z$  is expected to be a binary indicator vector representing the indices for the two sets.

### 7.2 Comparison to other cuts

The normalized cut may at first glaze seam similar to the cut made by thresholding the Fiedler vector (the second eigenvector of the regular laplacian  $L$ ). In the setting of Shi and Malic, this cut is expressed:

$$\min_S \frac{cut(S, \bar{S})}{|S|} + \frac{cut(S, \bar{S})}{|\bar{S}|}$$

where  $|S|$  is the number of nodes in  $S$  and not the total degree of the nodes in  $S$ . We may summarize by stating that at least three different types of cuts is solvable by relaxation into a eigenvalue problem:

1. Normalized cut: Cuts the graph into two similar sized regions such that the edge weight between the regions is low and the clustering within the region is high. Solved using the normalized laplacian

$$D^{-\frac{1}{2}}LD^{-\frac{1}{2}}z = \lambda z$$

2. Average cut: Cuts the graph into two similar sized regions such that the edge weight between the regions is low. Solved using the ordinary laplacian:

$$Lz = \lambda z$$

3. Average association: Cuts the graph into two similar sized regions such that the clustering in each region is high. Solved using the adjacency matrix:

$$Az = \lambda z$$

### 7.3 Cuts in image segmentation

The normalized cuts opens a framework for image segmentation based on different pixel similarity measures. We looked at some examples basted on a intensity graph with weights given by:

$$w_{i,j} = \exp\left(\frac{-\|F(i) - F(j)\|^2}{\sigma_I}\right) \cdot \begin{cases} \exp\left(\frac{-\|X(i) - X(j)\|^2}{\sigma_I}\right) & \text{for } \|X(i) - X(j)\| < r \\ 0 & \text{otherwise} \end{cases}$$

where  $F$  is a vectorized edition of the image,  $X$  is a vector of the spatial coordinates in the same image and  $r$  is a cutoff radius. The general observation from the examples was that the method had a tendency to cut the image into two equal sized regions, which did not always suit the application.

## 8 Plan for seminar April 5<sup>th</sup>

Based on the normalized cut introduced in the previous seminar we now discuss a generalization of the spectral cuts in graphs. In a paper by Inderjit, Guan and Kulis [2] they show that a normalized cut and a k-means clustering are mathematically equivalent. This equivalence is used to derive a segmentation method faster and more accurate than the methods in [4].

### 8.1 Kernel k-means

The normal k-means clustering of a set of vectors  $a_1, a_2 \dots a_n$  into clusters  $\pi_1, \pi_2 \dots \pi_k$  may be expressed as a minimization of

$$\sum_{c=1}^k \sum_{a \in \pi_c} \|a_i - m_c\|, \quad \text{where} \quad m_c = \frac{\sum_{a_i \in \pi_c} a_i}{|\pi_c|}.$$

In such a setting the clusters are constructed based on euclidian distances in hyperspaces and only linear separators between the clusters are considered.

To allow nonlinear separators kernel k-means introduce a nonlinear kernel function  $\phi$  and replaces the vectors  $a_i$  in the objective function by  $\phi(a_i)$ .

### 8.2 Trace maximization

We will show how both the kernel k-means optimization and the normalized cut problem may be represented as a matrix trace maximization. Using a suitable selection of kernel functions we will see that the two problems are mathematically equivalent.

### 8.3 Multilevel algorithm

In [2] they exploit the mentioned equivalence in order to solve normalized cut problems with a k-means algorithm. They further improve the performance by doing the segmentation in a multilevel style. The input graph is shrunk to a small size base graph which is segmented. Further, the k-means method is applied to increasing sizes of the graph, updating the initial coarse segmentation.

We will look at some examples from the paper describing the performance of this multilevel algorithm.

## 9 Plan for seminar May 3<sup>rd</sup>

This seminar will mainly be on work by Bae and Tai [1] on efficient segmentation using min-cut/ max-flow models. This involves introducing yet another family of graphs, s/t-graphs (probably familiar to most of us)

### 9.1 Mumford-Shah model

The Mumford-Shah model search for a segmentation of the image domain  $\Omega$  in to segments  $\{\Omega_i\}_{i=1}^n$  by minimizing the energy

$$E(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \mu \int_{\Omega \setminus \cup_i \Gamma_i} |\nabla u|^2 dx + \sum_{i=1}^n \nu \int_{\Gamma_i} ds.$$

The model enforces similarities and smoothness within each region as well as smoothness of the boundaries  $\Gamma_i$  between the regions. Robust minimization of the energy in the Mumford-Shah model has proven to be hard in a variational setting due to low convexity and other difficulties. Bae and Tai uses a simplification of the energy with constant intensities within each region:

$$E(u, \phi) = \sum \int_{\Omega} (u - u_0)^2 dx + \nu \int_{\Omega} |\nabla \phi| dx \quad (4)$$

Here,  $\phi$  is a a levelset function ( $\phi=i$  in  $\Omega_i$ ),  $u$  is an approximation image with constant intensity in each region.

### 9.2 s/t-graphs

Seeing the graph as a pipeline system gives rise to graph flow problems. The graph weights are then considered as edge capacities. If one node is considered as source and one as terminal many optimization problems can be described as maximizing the total flow from source to terminal. A classical theorem from Ford and Fulkerson stated that the maximum flow corresponds to the minimum cut in the graph such that the source and the terminal are separated. Several efficient algorithms for flow maximization are available.

### 9.3 Graph solutions of energy minimization

A descritization of the Mumford-Shah energy (and many others) may be solved as a s/t-graph problem. by assigning coefficient of the energy terms to each dimension of the graph. We will se how a single cut in a graph may give several segments of an image via the levelset function.

## References

- [1] E. Bae and X.C. Tai. Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. *Scale space and variational methods in computer vision*, pages 1–13, 2009.
- [2] Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Weighted graph cuts without eigenvectors a multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1944–1957, 2007.
- [3] Bin Luo, Richard C. Wilson, and Edwin R. Hancock. Spectral embedding of graphs. *Pattern Recognition*, 36(10):2213 – 2230, 2003.
- [4] Jianbo Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888 –905, August 2000.