Spectral graph theory: Applications of Courant-Fischer*

Steve Butler

September 2006

Abstract

In this second talk we will introduce the Rayleigh quotient and the Courant-Fischer Theorem and give some applications for the normalized Laplacian. Our applications will include structural characterizations of the graph, interlacing results for addition or removal of subgraphs, and interlacing for weak coverings. We also will introduce the idea of "weighted graphs".

1 Introduction

In the first talk we introduced the three common matrices that are used in spectral graph theory, namely the adjacency matrix, the combinatorial Laplacian and the normalized Laplacian. In this talk and the next we will focus on the normalized Laplacian. This is a reflection of the bias of the author having worked closely with Fan Chung who popularized the normalized Laplacian.

The adjacency matrix and the combinatorial Laplacian are also very interesting to study in their own right and are useful for giving structural information about a graph. However, in many "practical" applications it tends to be the spectrum of the normalized Laplacian which is most useful.

In this talk we will introduce the Rayleigh quotient which is closely linked to the Courant-Fischer Theorem and give some results. We give some structural results (such as showing that if a graph is bipartite then the spectrum is symmetric around 1) as well as some interlacing and weak covering results. But first, we will start by listing some common spectra.

*Second of three talks given at the *Center for Combinatorics*, Nankai University, Tianjin

Some common spectra

When studying spectral graph theory it is good to have a few basic examples to start with and be familiar with their spectrums. Building a large encyclopedia of spectrums can help to recognize patterns. Here we list several graphs and their spectrums for their normalized Laplacians.

- **Complete graphs:** The complete graph on n vertices, denoted K_n, is the graph which contains all possible edges. The spectrum of the graph is 0 (once) and $n/(n-1)$ ($n-1$ times).
- **Complete multipartite graph:** The complete multipartite graph, denoted $K_{m,n}$, has $m+n$ vertices where the vertices are partitioned into two groups A (with m vertices) and B (with n vertices) and as edges all possible edges connecting A and B. The spectrum of the graph is 0 (once), 1 ($m+n-2$ times) and 2 (once).
- **Cycles:** The cycle on n vertices, denoted C_n, is as its name implies a cycle. The spectrum of the graph is $1-\cos(2\pi k/n)$ for $k=0,1,\ldots,n-1$.
- **Paths:** The path on n vertices, denoted P_n, is as its name implies a path. The spectrum of the graph is $1-\cos\big(\pi k/(n-1)\big)$ for $k=0,1,\ldots,n-1$.
- **Petersen graph:** Perhaps the most famous graph which shows up in countless examples and counterexamples in graph theory. This is a graph on 10 vertices and is illustrated in Figure 1. The spectrum of the graph is 0 (once), 2/3 (five times), 5/3 (four times).
- **Hypercubes:** The n-cube has 2^n vertices which can be represented as all possible strings of length n using 0s and 1s with an edge connecting two strings if and only if they differ in a single entry. The spectrum of the graph is $2k/n$ ($\binom{n}{k}$ times) for $k=0,1,\ldots,n$.

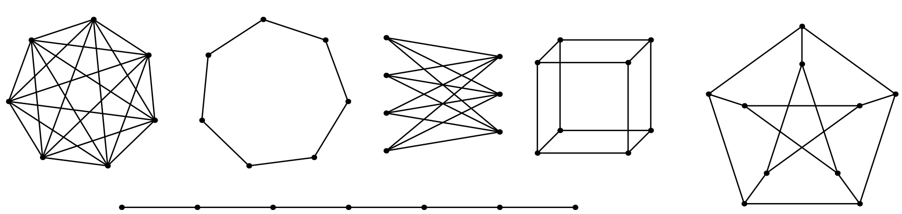

Figure 1: From left to right the graphs are K_7, C_7, $K_{3,4}$, Q_3, the Petersen graph, and on the bottom P_7.

2 The Rayleigh quotient and Courant-Fischer

For a nonzero (real) vector x and (real) matrix M, the Rayleigh quotient of x with M is defined as:

$$R(x) = \frac{x^T M x}{x^T x}.$$

(It is usually understood which matrix we are working with and so we write $R(x)$ instead of $R(x, M)$.)

Comment. In general we should use x^* (the conjugate transpose) instead of x^T for defining the Rayleigh quotient. In our case where we are dealing with a real symmetric matrix we can restrict ourselves only to real vectors and thus drop the conjugation.

Note that if x is an eigenvector of M associated with eigenvalue λ then $R(x) = x^T M x / x^T x = \lambda x^T x / x^T x = \lambda$. From this it is easy to see that $\max_{x \neq 0} R(x) \geq \max_i \lambda_i$, which is to say that the maximum of the Rayleigh quotient is at least as big as the largest eigenvalue of the matrix.

In fact, much more can be said if M is symmetric. In this case we have a full set of orthonormal eigenvectors $\phi_1, \ldots, \phi_n$ which we can use to decompose $x = \sum a_i \phi_i$. We then have that

$$R(x) = \frac{x^T M x}{x^T x} = \frac{\big(\sum_i a_i \phi_i\big)\big(\sum_i a_i \lambda_i \phi_i\big)}{\big(\sum_i a_i \phi_i\big)\big(\sum_i a_i \phi_i\big)} = \frac{\sum_i a_i^2 \lambda_i}{\sum_i a_i^2} \leq \max_i \lambda_i.$$

Combining the two ideas above it follows that for M real and symmetric with eigenvalues $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$,

$$\max_{x \neq 0} R(x) = \lambda_{n-1}, \tag{1}$$

while a similar derivation will give

$$\min_{x \neq 0} R(x) = \lambda_0. \tag{2}$$

Equations (1) and (2) are simple cases of the Courant-Fischer Theorem which generalizes the above statements to not only finding the smallest and largest eigenvalues of a graph but any single particular eigenvalue. In the statement of the theorem below the assumption of being real is not needed if we use conjugation as noted above.

Theorem 1 (Courant-Fischer Theorem)**.** *Let M be a (real) symmetric matrix with eigenvalues $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$. Let $\mathcal{X}^k$ denote a k dimensional subspace of $\mathbf{R}^n$ and $x \perp \mathcal{X}^k$ signify that $x \perp y$ for all $y \in \mathcal{X}^k$. Then*

$$\lambda_i = \min_{\mathcal{X}^{n-i-1}} \big(\max_{x \perp \mathcal{X}^{n-i-1}, x \neq 0} R(x) \big) = \max_{\mathcal{X}^i} \big(\min_{x \perp \mathcal{X}^i, x \neq 0} R(x) \big). \tag{3}$$

From the proof of the Courant-Fischer Theorem it will follow that when looking for eigenvalue λ_i, that $\mathcal{X}^{n-i-1}$ in the first expression is the span of the last $n-i-1$ eigenvectors while $\mathcal{X}^i$ in the second expression is the span of the first i eigenvectors.

We will not include the proof of the Courant-Fischer Theorem here. The interested reader can find a proof in any major linear algebra textbook.

3 Eigenvalues of the normalized Laplacian

Recall from the last lecture that $\mathcal{L} = D^{-1/2}LD^{-1/2} = D^{-1/2}\big(D-A\big)D^{-1/2}$. Then note that for a vector $y = (y_i)$ that

$$y^T Ly = y^T(D-A)y = \sum_i y_i^2 d_i - 2\sum_{i\sim j} y_i y_j = \sum_{i\sim j}(y_i - y_j)^2,$$

where $i{\sim}j$ indicates that i is adjacent to j. In general, $y^T My$ is sometimes referred to as the quadratic form of the matrix because it can always be decomposed into a sum of squares. In our case since all of the coefficients are 1 it follows that $y^T Ly \geq 0$ from which it follows that the eigenvalues of L are nonnegative.

Relating back to the normalized Laplacian we have

$$\frac{x^T\mathcal{L}x}{x^Tx} = \frac{(D^{1/2}y)\mathcal{L}(D^{1/2}y)}{(D^{1/2}y)^T(D^{1/2}y)} = \frac{y^TLy}{y^TDy} = \frac{\sum_{i\sim j}(y_i-y_j)^2}{\sum_i y_i^2 d_i}$$

where we made the substitution $x = D^{1/2}y$ for some vector y. In particular, if we assume that D is invertible (i.e., no isolated vertices) then $D^{1/2}$ maps a k-dimensional subspace to some other k-dimensional subspace. Using this we will then get the following:

$$\begin{aligned}\lambda_i &= \min_{\mathcal{X}^{n-i-1}}\left(\max_{x\perp\mathcal{X}^{n-i-1},x\neq 0}\frac{x^T\mathcal{L}x}{x^Tx}\right) = \min_{\mathcal{X}^{n-i-1}}\left(\max_{D^{1/2}y\perp\mathcal{X}^{n-i-1},D^{1/2}y\neq 0}\frac{y^TLy}{y^Ty}\right)\\ &= \min_{\mathcal{Y}^{n-i-1}}\left(\max_{y\perp\mathcal{Y}^{n-i-1},y\neq 0}\frac{\sum_{i\sim j}(y_i-y_j)^2}{\sum_i y_i^2 d_i}\right).\end{aligned} \tag{4}$$

Similarly,

$$\lambda_i = \max_{\mathcal{Y}^i}\left(\min_{y\perp\mathcal{Y}^i,y\neq 0}\frac{\sum_{i\sim j}(y_i-y_j)^2}{\sum_i y_i^2 d_i}\right). \tag{5}$$

To find the eigenvector we find the y which minimizes or maximizes the above expressions and take $D^{1/2}y$.

Lemma 2. *The eigenvalues of $\mathcal{L}$ are nonnegative.*

This follows easily from noting that the expressions in (4) and (5) are always nonnegative. More specifically we have that 0 is an eigenvalue, and it is easy to see that this corresponds to the eigenvector $D^{1/2}\mathbf{1}$ (in equations (4) and (5) this would correspond to the choice of $y = \mathbf{1}$).

Since λ_0 is always 0 the two important eigenvalues will be λ_{n-1} and λ_1. In the third lecture we will see that these control "expansion" and the closer these eigenvalues are to 1 the more "random-like" our graph will be. From (4) we have that

$$\lambda_{n-1} = \max_{y \neq 0} \frac{\sum_{i\sim j}(y_i - y_j)^2}{\sum_i y_i^2 d_i},$$

while from (5) and the above comments

$$\lambda_1 = \min_{y \perp D\mathbf{1}, y \neq 0} \frac{\sum_{i\sim j}(y_i - y_j)^2}{\sum_i y_i^2 d_i}.$$

If G is not connected it is simple to find a y so that $\lambda_1 = 0$. On the other hand if G is connected since a vector y which minimizes the above expression must have positive and negative entries (to satisfy $y \perp D\mathbf{1}$), there would be an $i\sim j$ so that $y_i > 0$ and $y_j < 0$, thus that term, as well as the complete sum, would be strictly positive. This establishes the following.

Lemma 3. *Given a graph G, $\lambda_1 > 0$ if and only if G is connected. More generally, $\lambda_i = 0$ if and only if there are at least $i + 1$ connected components of G.*

This shows the power of spectral graph theory, eigenvalues are analytic tools while being connected is a structural property, and as the above lemma shows they are closely connected. Similarly we have the following.

Lemma 4. *Given a graph G, $\lambda_{n-1} \leq 2$ and $\lambda_{n-1} = 2$ if and only if G has a bipartite component.*

This makes use of the following simple inequality: $(a - b)^2 \leq 2a^2 + 2b^2$.

$$\lambda_{n-1} = \max_{y \neq 0} \frac{\sum_{i\sim j}(y_i - y_j)^2}{\sum_i y_i^2 d_i} \leq \max_{y \neq 0} \frac{\sum_{i\sim j}(2y_i^2 + 2y_j^2)}{\sum_i y_i^2 d_i} = 2$$

Equality will hold if and only if $(y_i - y_j)^2 = 2y_i^2 + 2y_j^2$ for all $i\sim j$, which is equivalent to saying that $y_i = -y_j$ for all $i\sim j$. So we can then use the eigenvector which achieves the eigenvalue 2 to find a bipartite component by ignoring any vertices which have 0 entry in the eigenvector and then putting all vertices with positive entry in one group and all vertices with negative entry in another group. It is easy to check that there then must be a nontrivial bipartite component.

Comment. In the above argument we used the eigenvector to help locate a bipartite component. While most of our discussion will focus on the eigenvalues of the matrix, there is also important information that can be derived by examining the eigenvalues of the graph.

More generally, let G be a bipartite graph with the vertices in two sets A and B. Then λ is an eigenvalue of $\mathcal{L}$ if and only if $2-\lambda$ is an eigenvalue. This follows by taking the eigenvector for λ, say y, and modifying it to create

$$\overline{y} = \begin{cases} y_i & \text{if } i \in A; \\ -y_i & \text{if } i \in B; \end{cases}$$

then verifying that $\overline{y}$ is an eigenvector of $\mathcal{L}$ associated with eigenvalue $2-\lambda$.

Lemma 5. *If G is not a complete graph then $\lambda_1 \leq 1$.*

For this suppose that k and ℓ are not adjacent and consider the vector $\hat{y}$ which is d_ℓ at k and $-d_k$ at ℓ. Then we have that

$$\lambda_1 = \min_{y \perp D\mathbf{1}, y \neq 0} \frac{\sum_{i\sim j}(y_i - y_j)^2}{\sum_i y_i^2 d_i} \leq \frac{\sum_{i\sim j}(\hat{y}_i - \hat{y}_j)^2}{\sum_i \hat{y}_i^2 d_i} = \frac{d_k d_\ell^2 + d_\ell d_k^2}{d_k d_\ell^2 + d_\ell d_k^2} = 1.$$

Note that earlier we saw that the complete graph had $\lambda_1 = n/(n-1) > 1$. This implies that G is a complete graph if and only if $\lambda_1 > 1$.

4 Interlacing inequalities

In this section and the next we will compare eigenvalues of two graphs. Intuitively we want to be able to say that if two graphs are "very similar" then their spectrums should be close. Before we begin this we will first introduce the idea of weighted graphs.

Weighted graphs

So far we have worked with simple unweighted graphs. For some applications it can be useful to allow weights to be placed on the edges. For example, if we want to work with the spectrum of multigraphs then we can model such graphs by assigning the weight between vertex i and vertex j to be the number of multi-edges.

For another example, suppose we want to follow the flow of water through a system of pipes, since not all pipes may be equal in size we should account for the disparity of flow, which can be done by putting weights on the edges. In this case one reasonable weight to put on the edges is the diameter of the pipe, since a larger pipe should allow for more water flow.

In general, a weighted graph is a graph with a nonnegative weight function w where $w(i,j) = w(j,i)$ (i.e., still undirected at this time), and there is an edge joining i and j if and only if $w(i,j) > 0$. We also allow loops, these correspond to when $w(i,i) > 0$. The degree of vertex i will now be the sum of the weights of the incident edges, i.e.,

$$d_i = \sum_j w(i,j).$$

We then proceed as before by letting $A_{i,j} = w(i,j)$, D the diagonal degree matrix (using our new definition of degree) and $\mathcal{L} = D^{-1/2}(D-A)D^{-1/2}$. This is similar to what we have done before and many of the results easily generalize (indeed everything we have done so far corresponds to the case when the weights are 0 and 1). In particular, we can now write (4) and (5) as

$$\begin{aligned}\lambda_i &= \min_{\mathcal{Y}^{n-i-1}} \left(\max_{y \perp \mathcal{Y}^{n-i-1}, y \neq 0} \frac{\sum_{i\sim j}(y_i - y_j)^2 w(i,j)}{\sum_i y_i^2 d_i} \right) & (6)\\ &= \max_{\mathcal{Y}^i} \left(\min_{y \perp \mathcal{Y}^i, y\neq 0} \frac{\sum_{i\sim j}(y_i - y_j)^2 w(i,j)}{\sum_i y_i^2 d_i} \right). & (7)\end{aligned}$$

Subtracting out a weighted subgraph

Given a simple graph G suppose we remove one edge. This will change the graph and so also change the spectrum. What can we say about the relation between the two different spectrums?

We will show that there is a relation and that the eigenvalues interlace. That is we will show that an eigenvalue of the new graph lies in an interval between two specified eigenvalues of the original graph.

In fact we will show something more general, instead of removing only one edge we will remove some of the weights off a set of edges. First we need to introduce some terminology needed for the statement of our result. Given a weighted graph G a subgraph H of G is a weighted graph where $w_H(i,j) \le w_G(i,j)$ for all i and j, while the graph $G - H$ has weight function $w_{G-H}(i,j) = w_G(i,j) - w_H(i,j)$.

Theorem 6. *Let G be a weighted graph and H a connected subgraph of G with $|V(H)| = t$. If*

$$\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1} \qquad \textit{and} \qquad \theta_0 \le \theta_1 \le \cdots \le \theta_{n-1}$$

are the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(G-H)$ respectively, then for $k=0,1,\ldots,n-1$ we have

$$\lambda_{k-t+1} \ \le \ \theta_k \ \le \ \begin{cases} \lambda_{k+t-1} & H \text{ is loopless and bipartite,} \\ \lambda_{k+t} & \text{otherwise,} \end{cases}$$

where $\lambda_{-t+1}=\cdots=\lambda_{-1}=0$ and $\lambda_n=\cdots=\lambda_{n+t}=2$.

One interesting thing to note is that the result is independent of the amount of weight subtracted. This is one consequence of normalization. Instead of subtracting out a graph we could also add a graph. The following result immediately follows from Theorem 6 working with the graphs $G+H$ and $(G+H)-H=G$.

Corollary 7. *Let G be a weighted graph and H a connected graph on a subset of the vertices of G with $|V(H)|=t$. If*

$$\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1} \qquad \text{and} \qquad \theta_0 \le \theta_1 \le \cdots \le \theta_{n-1}$$

are the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(G+H)$ respectively, then for $k=0,1,\ldots,n-1$ we have

$$\lambda_{k+t-1} \ \ge \ \theta_k \ \ge \ \begin{cases} \lambda_{k-t+1} & H \text{ is loopless and bipartite,} \\ \lambda_{k-t} & \text{otherwise,} \end{cases}$$

where $\lambda_{-t}=\cdots=\lambda_{-1}=0$ and $\lambda_n=\cdots=\lambda_{n+t-1}=2$.

The proof of Theorem 6 is established in three parts. We will do one part here and leave the other two to the interested reader. In particular, we will show that $\theta_k \ge \lambda_{k-t+1}$. Suppose that $\{i_1,j_1\},\{i_2,j_2\},\ldots,\{i_{t-1},j_{t-1}\}$ are edges of a spanning subgraph of H, and let $\mathcal{Z}=\{e_{i_1}-e_{j_1},e_{i_2}-e_{j_2},\ldots,e_{i_{t-1}}-e_{j_{t-1}}\}$ be a set of vectors where e_j is the vector which is 1 in the jth position and 0 otherwise. Then using (6) we have

$$\begin{aligned}
\theta_k &= \min_{\mathcal{Y}^{n-k-1}} \left(\max_{y\perp\mathcal{Y}^{n-k-1},y\ne 0} \frac{\sum_{i\sim j}(y_i-y_j)^2 w_{G-H}(i,j)}{\sum_i y_i^2 d_{G-H,i}} \right) \\
&= \min_{\mathcal{Y}^{n-k-1}} \left(\max_{y\perp\mathcal{Y}^{n-k-1},y\ne 0} \frac{\sum_{i\sim j}(y_i-y_j)^2 w_G(i,j) - \sum_{i\sim j}(y_i-y_j)^2 w_H(i,j)}{\sum_i y_i^2 d_{G,i} - \sum_i y_i^2 d_{H,i}} \right) \\
&\ge \min_{\mathcal{Y}^{n-k-1}} \left(\max_{y\perp\mathcal{Y}^{n-k-1},y\perp\mathcal{Z},y\ne 0} \frac{\sum_{i\sim j}(y_i-y_j)^2 w_G(i,j)}{\sum_i y_i^2 d_{G,i} - \sum_i y_i^2 d_{H,i}} \right) \\
&\ge \min_{\mathcal{Y}^{n-k-1}} \left(\max_{y\perp\mathcal{Y}^{n-k-1},y\perp\mathcal{Z},y\ne 0} \frac{\sum_{i\sim j}(y_i-y_j)^2 w_G(i,j)}{\sum_i y_i^2 d_{G,i}} \right) \\
&\ge \min_{\mathcal{Y}^{n-k+t-2}} \left(\max_{y\perp\mathcal{Y}^{n-k+t-2},y\ne 0} \frac{\sum_{i\sim j}(y_i-y_j)^2 w_G(i,j)}{\sum_i y_i^2 d_{G,i}} \right) \ = \ \lambda_{k-t+1}.
\end{aligned}$$

In going from the second to the third line we added the condition that y also be perpendicular to $\mathcal{Z}$ so that we are maximizing over a smaller set. With the condition that $y \perp \mathcal{Z}$ then $y_i = y_j$ for all i, j in H, in particular the second term in the numerator drops out. Going from the third to the fourth line we make the denominator larger. While in going from the fourth to the fifth line we considered a broader optimization that would include the fourth line as a case.

The other two statements in Theorem 6 use (7) and are proved similarly.

As an example of what can be proved using interlacing, it is an easy exercise using Theorem 6 to show that if G is a simple graph on n vertices and more than $n/2$ of the vertices are connected to every other vertex then $n/(n-1)$ is an eigenvalue of G. (This works well because $n/(n-1)$ has high multiplicity in the complete graph and then we "punch out" a subgraph. In general to use interlacing to show a graph contains a particular eigenvalue we must start with a graph which has an eigenvalue with high multiplicity.)

5 Weak coverings and eigenvalues

In addition to removing subgraphs as we did in the previous section we can also "condense" graphs, that is we identify vertices together and then reassign weights appropriately. We call this process a covering, the condensed graph will be referred to as the covered graph and the original graph the covering graph.

Comment. We will discuss weak coverings which make the fewest assumptions on the structure of the covering. By making more assumptions we get a richer structure and more can be said about how eigenvalues relate. For example, one type of covering is known as a strong regular covering and for such a covering both graphs have the *same* eigenvalues. These types of coverings have been extensively studied in algebraic graph theory with great success in being able to determine the eigenvalues of a large graph by computing the eigenvalues of a small graph. One drawback however is that such coverings require working with graphs with nice structures, such as Cayley graphs, and the results are not widely applicable to "real-world" graphs.

We say that G is a weak covering of H if and only if there is some onto mapping $\pi : V(G) \rightarrow V(H)$ such that for all $u, v \in V(H)$,

$$w_H(u, v) = \sum_{\substack{x \in \pi^{-1}(u) \\ y \in \pi^{-1}(v)}} w_G(x, y).$$

From this definition it follows that $d_H(v) = \sum_{x \in \pi^{-1}(v)} d_G(x)$.

Alternatively, we group the vertices of G in some manner then collapse the individual groups of vertices into a single vertex. To find the edge weights of the resulting edges we add the weights of any resulting parallel edges that are formed. An example is shown in Figure 2 (note that the loop formed will be counted twice when tallying the edge weight).

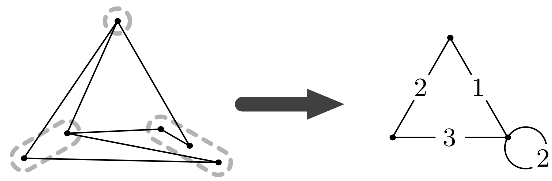

Figure 2: The graph on the left is the covering graph with all edge weights 1 while the graph on the right is the covered graph with edge weights as indicated.

Here we are "weak" in that we are putting few restrictions on the mapping π. However, even with these base assumptions we are still able to relate the eigenvalues of G and H (the covering graph and the covered graph).

Theorem 8. *Let G be a weak cover of H with $|V(G)| = n$ and $|V(H)| = m$, and further let*

$$\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1} \qquad \text{and} \qquad \theta_0 \le \theta_1 \le \cdots \le \theta_{m-1}$$

be the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(H)$. Then for $k = 0, 1, \ldots, m-1$ we have the following

$$\lambda_k \le \theta_k \le \lambda_{k+(n-m)}.$$

To establish this result, for $i = 1, \ldots, m$ let $V_i = \pi^{-1}(v_i)$, i.e., these are the groupings of the vertices of G, and let $\mathcal{Z}_i = \{e_{i_1} - e_{i_2}, e_{i_1} - e_{i_3}, \ldots, e_{i_1} - e_{i_j}\}$ where $V_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_j}\} \subseteq V(G)$. Further we will let $\mathcal{Z} = \bigcup_i \mathcal{Z}_i$. It is easy to check that the dimension of the span of $\mathcal{Z}$ is $n - m$.

$$\begin{aligned} \theta_k &= \max_{\mathcal{Y}^k \subseteq \mathbf{R}^m} \left(\min_{y \perp \mathcal{Y}^k, y \neq 0} \frac{\sum (y_i - y_j)^2 w_H(i,j)}{\sum y_i^2 d_H(i)} \right) \\ &= \max_{\mathcal{Y}^k \subseteq \mathbf{R}^n} \left(\min_{y \perp \mathcal{Y}^k, y \perp \mathcal{Z}, y \neq 0} \frac{\sum (y_i - y_j)^2 w_G(i,j)}{\sum y_i^2 d_G(i)} \right) \end{aligned} \tag{8}$$

What we have done in the second step is used the defining property of weak covers to lift vectors from H up to G so that we still satisfy the same Rayleigh quotient. Our only condition in lifting is that $y_i = y_{\pi(i)}$ which is to say that for a function on H we lift the value at a vertex up to all of the vertices in G which cover it. In particular if $\pi(v_i) = \pi(v_j)$, then we need $y_{\pi(i)} = y_{\pi(j)}$. This last condition is easily achieved by requiring that the lifted vector be perpendicular to $\mathcal{Z}$, and hence the form listed.

Looking at (8) we can take two approaches, the first is to drop the requirement that we remain perpendicular to $\mathcal{Z}$, thus we are minimizing over a larger set and so we have

$$\theta_k \;\geq\; \max_{\mathcal{Y}^k}\left(\min_{y\perp\mathcal{Y}^k, y\neq 0}\frac{\sum(y_i-y_j)^2 w_G(i,j)}{\sum y_i^2 d_G(i)}\right) \;=\; \lambda_k.$$

The other approach is to maximize over some larger set that will also consider the situation given in (8), i.e.,

$$\theta_k \;\leq\; \max_{\mathcal{Y}^{k+n-m}}\left(\min_{y\perp\mathcal{Y}^{k+n-m}, y\neq 0}\frac{\sum(y_i-y_j)^2 w_G(i,j)}{\sum y_i^2 d_G(i)}\right) \;=\; \lambda_{k+n-m}.$$

Combining the two inequalities above establishes Theorem 8.