Combinatorial Network Analysis - Final Lecture

Vanni Noferini (Aalto University)

Aalto University

December 7, 2021

Summary of previous lecture

Last week, we expressed the minimum 2-cut problem as the integer optimization problem

$$\min v^T L v : v \in \{-1, 1\}^n, v \neq \pm e.$$

Summary of previous lecture

Last week, we expressed the minimum 2-cut problem as the integer optimization problem

$$\min v^T L v : v \in \{-1, 1\}^n, v \neq \pm e.$$

What makes this problem difficult is the constraint.

Summary of previous lecture

Last week, we expressed the minimum 2-cut problem as the integer optimization problem

$$\min v^T L v : v \in \{-1, 1\}^n, v \neq \pm e.$$

What makes this problem difficult is the constraint.

The relaxed problem when $0 \neq v \in \mathbb{R}^n$ is instead easy: its minimum is 0 (the least eigenvalue of L) and its minimizer is (any nonzero multiple) of e (the eigenvector).

Fiedler's idea

Then the Czech mathematician M. Fiedler had the idea to harden the relaxed constraint a bit, so to minimise over all real vectors v that are orthogonal to e.

Fiedler's idea

Then the Czech mathematician M. Fiedler had the idea to harden the relaxed constraint a bit, so to minimise over all real vectors v that are orthogonal to e.
This exploits the

Theorem (Courant-Fischer)

Suppose that M is a real symmetric matrix having least eigenvalue λ_n with eigenvector w. Then,

$$\min_{\substack{v^T v=1 \\ w^T v=0}} v^T M v = \lambda_{n-1}$$

$$\mathrm{argmin}_{\substack{v^T v=1 \\ w^T v=0}} v^T M v = u$$

where λ_{n-1} is the second least eigenvalue and u is a normalized eigenvector of its.

Fiedler's idea - 2

When applied to the case $M = L$, the Courant-Fischer theorem implies that

$$\min_{\substack{v^T v = 1 \\ e^T v = 0}} v^T L v = \lambda_{n-1}$$

where λ_{n-1} is the spectral gap. Morever, the minimiser is the Fiedler vector f. Some remarks:

Fiedler's idea - 2

When applied to the case $M = L$, the Courant-Fischer theorem implies that

$$\min_{\substack{v^T v = 1 \\ e^T v = 0}} v^T L v = \lambda_{n-1}$$

where λ_{n-1} is the spectral gap. Morever, the minimiser is the Fiedler vector f. Some remarks:

1. $f^T e = 0$ implies that f must have both negative and positive entries.

Fiedler's idea - 2

When applied to the case $M = L$, the Courant-Fischer theorem implies that

$$\min_{\substack{v^T v = 1 \\ e^T v = 0}} v^T L v = \lambda_{n-1}$$

where λ_{n-1} is the spectral gap. Morever, the minimiser is the Fiedler vector f. Some remarks:

1. $f^T e = 0$ implies that f must have both negative and positive entries.
2. This suggests to assign $i \in V_1$ if $f_i > 0$ and $i \in V_2$ if $f_i < 0$.

Fiedler's idea - 2

When applied to the case $M = L$, the Courant-Fischer theorem implies that

$$\min_{\substack{v^T v=1 \\ e^T v=0}} v^T L v = \lambda_{n-1}$$

where λ_{n-1} is the spectral gap. Morever, the minimiser is the Fiedler vector f. Some remarks:

1. $f^T e = 0$ implies that f must have both negative and positive entries.
2. This suggests to assign $i \in V_1$ if $f_i > 0$ and $i \in V_2$ if $f_i < 0$.
3. The method is coherent even if one multiplies f by a negative number (up to swtiching V_1, V_2).

Fiedler's idea - 2

When applied to the case $M = L$, the Courant-Fischer theorem implies that

$$\min_{\substack{v^T v=1 \\ e^T v=0}} v^T L v = \lambda_{n-1}$$

where λ_{n-1} is the spectral gap. Morever, the minimiser is the Fiedler vector f. Some remarks:

1. $f^T e = 0$ implies that f must have both negative and positive entries.
2. This suggests to assign $i \in V_1$ if $f_i > 0$ and $i \in V_2$ if $f_i < 0$.
3. The method is coherent even if one multiplies f by a negative number (up to swtiching V_1, V_2).
4. What if $f_i = 0$? We could assign them randomly or, if there are not many, try manually all the possibilities and pick the best.

Back to the example of last time

Last week, we solved by hand the 2-cut problem for $V(G) = [4]$ and $E(G) = \{12, 21, 23, 32, 24, 42, 34, 43\}$. The actual optimum was for $V_1 = \{1\}$.

Back to the example of last time

Last week, we solved by hand the 2-cut problem for $V(G) = [4]$ and $E(G) = \{12, 21, 23, 32, 24, 42, 34, 43\}$. The actual optimum was for $V_1 = \{1\}$.

Let us try Fiedler's clustering method using MATLAB.

Another example

Again on MATLAB

Comments on the previous example

- Fiedler's method got close to, but was not quite able to, identify the two cycles (which is the minimum cut: cut function is 1 and cannot be 0 as G is connected). The first entry of f was negative, but quite small.

Comments on the previous example

- Fiedler's method got close to, but was not quite able to, identify the two cycles (which is the minimum cut: cut function is 1 and cannot be 0 as G is connected). The first entry of f was negative, but quite small.
- This can be understood by thinking of the constraint $v^T e = 0$. If v had entries in $\{-1, 1\}$ and G had an even number of vertices, this would imply $\#V_1 = \#V_2$.

Comments on the previous example

- Fiedler's method got close to, but was not quite able to, identify the two cycles (which is the minimum cut: cut function is 1 and cannot be 0 as G is connected). The first entry of f was negative, but quite small.
- This can be understood by thinking of the constraint $v^T e = 0$. If v had entries in $\{-1, 1\}$ and G had an even number of vertices, this would imply $\#V_1 = \#V_2$.
- In Fiedler's method, there is somehow a balance between minimum cut and equal distribution of vertices. This side effect is unwanted, and is the price we pay to make the problem computationally more tractable!

Induced subgraphs

If $G = (V, E)$ is a simple graph and $V_1 \subseteq V$, then the subgraph induced by V_1 is $G_1 = (V_1, E_1)$ where E_1 is the subset of E containing all edges in E whose endvertices are both in V_1.

A theorem by Fiedler

Theorem (Fiedler)

Let $G = (V, E)$ be a simple connected graph and suppose that

1. *the Fiedler vector f does not have any zero entry;*
2. *in the partition $V = V_1 \cup V_2$ prescribed by Fiedler's clustering algorithm both V_1 and V_2 contain at least two nodes*

Denote by G_1 and G_2 the subgraphs induced by V_1 and V_2 respectively. Then, both G_1 and G_2 are connected.

We will give a proof due to J. Demmel.

Demmel's proof of Fiedler's theorem - I

Suppose for a contradiction that G_1 is not connected (the proof for G_2 is the same). Then, up to graph isomorphism, the graph Laplacian of G has the form

$$L = \begin{bmatrix} L_{11} & 0 & -A_{13} \\ 0 & L_{22} & -A_{23} \\ -A_{13}^T & -A_{23}^T & L_{33} \end{bmatrix};$$

here, the three blocks correspond to: one connected component (or isolated node) in G_1, the rest of G_1, and G_3.

Demmel's proof of Fiedler's theorem - I

Suppose for a contradiction that G_1 is not connected (the proof for G_2 is the same). Then, up to graph isomorphism, the graph Laplacian of G has the form

$$L = \begin{bmatrix} L_{11} & 0 & -A_{13} \\ 0 & L_{22} & -A_{23} \\ -A_{13}^T & -A_{23}^T & L_{33} \end{bmatrix};$$

here, the three blocks correspond to: one connected component (or isolated node) in G_1, the rest of G_1, and G_3.

Let us partition the Fiedler vector coherently as

$$f = \begin{bmatrix} x \\ y \\ -z \end{bmatrix}$$

with $x, y, z > 0$ componentwise.

Demmel's proof of Fiedler's theorem - II

Moreover, denoting by ϕ the spectral gap, we have the equations:

$$Lf = \phi f \Leftrightarrow \begin{cases} L_{11}x + A_{13}z = \phi x, \\ L_{22}y + A_{23}z = \phi y, \\ -A_{13}^T x - A_{23}^T y - L_{33}z = -\phi z. \end{cases}$$

Demmel's proof of Fiedler's theorem - II

Moreover, denoting by ϕ the spectral gap, we have the equations:

$$Lf = \phi f \Leftrightarrow \begin{cases} L_{11}x + A_{13}z = \phi x, \\ L_{22}y + A_{23}z = \phi y, \\ -A_{13}^T x - A_{23}^T y - L_{33}z = -\phi z. \end{cases}$$

Cauchy interlacing theorem: if M is a symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and S is any $m \times m$ principal (i.e. obtained by selecting the same subset of rows and columns) submatrix of M, with eigenvalues $\mu_1 \geq \cdots \geq \mu_m$, then, for all $j = 1, \ldots, m$, it holds that

$$\lambda_{n-m+j} \leq \mu_j \leq \lambda_j.$$

Demmel's proof of Fiedler's theorem - II

Moreover, denoting by ϕ the spectral gap, we have the equations:

$$Lf = \phi f \Leftrightarrow \begin{cases} L_{11}x + A_{13}z = \phi x, \\ L_{22}y + A_{23}z = \phi y, \\ -A_{13}^T x - A_{23}^T y - L_{33}z = -\phi z. \end{cases}$$

Cauchy interlacing theorem: if M is a symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and S is any $m \times m$ principal (i.e. obtained by selecting the same subset of rows and columns) submatrix of M, with eigenvalues $\mu_1 \geq \cdots \geq \mu_m$, then, for all $j = 1, \ldots, m$, it holds that

$$\lambda_{n-m+j} \leq \mu_j \leq \lambda_j.$$

In the special case $j = m - 1$ and $M = L$, this implies

$$\phi \leq \mu_{m-1}.$$

In other words, any principal submatrix of L cannot have more than one eigenvalue strictly less than the spectral gap.

Demmel's proof of Fiedler's theorem - III

Rayleigh's theorem: denoting by $\mu_{\min}(L_{11})$ the smallest eigenvalue of L_{11},

$$\mu_{\min}(L_{11}) = \min_{v \neq 0} \frac{v^T L_{11} v}{v^T v} \leq \frac{x^T L_{11} x}{x^T x}$$

which implies that

$$x^T x \mu_{\min}(L_{11}) \leq x^T L_{11} x = \phi x^T x - x^T A_{13} z < \phi x^T x.$$

Demmel's proof of Fiedler's theorem - III

Rayleigh's theorem: denoting by $\mu_{\min}(L_{11})$ the smallest eigenvalue of L_{11},

$$\mu_{\min}(L_{11}) = \min_{v \neq 0} \frac{v^T L_{11} v}{v^T v} \leq \frac{x^T L_{11} x}{x^T x}$$

which implies that

$$x^T x \mu_{\min}(L_{11}) \leq x^T L_{11} x = \phi x^T x - x^T A_{13} z < \phi x^T x.$$

Why the last step? Because $A_{13} \geq 0$, and $\neq 0$ (else, G is disconnected!). Since $z > 0$, then $-A_{13} z \leq 0$ and $\neq 0$. On the other hand, $x > 0$, and hence, $-x^T A_{13} z < 0$. Thus, $\mu_{\min}(L_{11}) < \phi$.

Demmel's proof of Fiedler's theorem - IV

In the very same way we can prove

$$y^T y \mu_{\min}(L_{22}) \leq y^T L_{22} x = \phi y^T y - y^T A_{23} z < \phi y^T y$$

implying $\mu_{\min}(L_{22}) < \phi$.

Demmel's proof of Fiedler's theorem - IV

In the very same way we can prove

$$y^T y \mu_{\min}(L_{22}) \leq y^T L_{22} x = \phi y^T y - y^T A_{23} z < \phi y^T y$$

implying $\mu_{\min}(L_{22}) < \phi$. Hence,

$$\begin{bmatrix} L_{11} & 0 \\ 0 & L_{22} \end{bmatrix}$$

is a principal submatrix of L having at least two eigenvalues strictly smaller than ϕ: this contradicts Cauchy's interlacing theorem.

Clustering via the adjacency matrix

Here is a variant of Fiedler's method: since for a simple graph A is also symmetric, we could try to maximize $v^T A v$ with the costraint that v is orthogonal to the Perron-Frobenius eigenvector. This would lead us to consider the eigenvector s associated with the second largest eigenvalue of A, and again (since, like before, s must have both negative and positive entries) we can use the signs of s to construct a partition.

Clustering via the adjacency matrix

Here is a variant of Fiedler's method: since for a simple graph A is also symmetric, we could try to maximize $v^T Av$ with the costraint that v is orthogonal to the Perron-Frobenius eigenvector. This would lead us to consider the eigenvector s associated with the second largest eigenvalue of A, and again (since, like before, s must have both negative and positive entries) we can use the signs of s to construct a partition.

Let us go back to MATLAB and test this algorithm on the second of our prior examples.

Directed graphs

Clustering using A was able to identify the minimum cut in our example, but not always it is better than Fiedler's original method.

Directed graphs

Clustering using A was able to identify the minimum cut in our example, but not always it is better than Fiedler's original method.

However, it has the advantage of being generalisable to directed graphs: in that case, one starts from a singular value decomposition $A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$.

Directed graphs

Clustering using A was able to identify the minimum cut in our example, but not always it is better than Fiedler's original method.

However, it has the advantage of being generalisable to directed graphs: in that case, one starts from a singular value decomposition $A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$.

The leading singular vectors u_1, v_1 are the Perron-Frobenius eigenvectors of AA^T and A^TA, resp., and hence can be taken to be positive. Therefore, the next singular vectors u_2, v_2 (being orthogonal to u_1, v_1 resp.) must contain both negative and positive entries!

Directed graphs

Clustering using A was able to identify the minimum cut in our example, but not always it is better than Fiedler's original method.

However, it has the advantage of being generalisable to directed graphs: in that case, one starts from a singular value decomposition $A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$.

The leading singular vectors u_1, v_1 are the Perron-Frobenius eigenvectors of AA^T and $A^T A$, resp., and hence can be taken to be positive. Therefore, the next singular vectors u_2, v_2 (being orthogonal to u_1, v_1 resp.) must contain both negative and positive entries!

If $A \neq A^T$, we then have two distinct partitions. That given by u_2 corresponds to clustering authorities (sources followed by many nodes), that given by v_2 corresponds to clustering hubs (targets that follow many nodes).