Lecture 6: Graph Embeddings. Spectral Analysis

Radu Balan

University of Maryland

March 24, 2020

Graphs

The overarching problem is the following:

Main Problem

Given a graph find a low-dimensional representation of the graph, also called a graph embedding*.*

Graphs

The overarching problem is the following:

Main Problem

Given a graph find a low-dimensional representation of the graph, also called a graph embedding*.*

As we shall see there are a several results that ultimately reduce the problem to a spectral analysis.
Input data: an adjacency matrix A or a weight matrix W, of size $n \times n$.
Outcome: $\{x_1, \cdots, x_n\} \subset \mathbb{R}^d$ a set of d-vectors so that:

- For unweighted graphs: If node i is connected to node j, then $\|x_i - x_j\|$ should be smaller than, say $\|x_k - x_p\|$ if nodes k and p are not connected.
- For weighted graphs: The larger the weight $W_{i,j}$ the smaller the distance $\|x_i - x_j\|$.

Visualization Problem

Consider a graph $(\mathcal{V}, \mathcal{E}, W)$ with n vertices and $m = |\mathcal{E}|$ edges. We want a 2-dimensional (planar) visualization of this graph.
Idea (due to Hall '70): Let $\{x(1), x(2), \cdots, x(n)\} \subset \mathbb{R}^2$ denote a collection of n points in 2D-plane. Points are chosen so to minimize the weighted sum of edge lengths:

$$J = \sum_{(k,j)\in\mathcal{E}} W_{k,j} \|x(k) - x(j)\|^2$$

This is similar to the Dirichlet energy except that each $x(k)$ is a 2D vector instead of a scalar value. J can be rewritten more compactly using the $2 \times n$ matrix X whose columns are the vectors $\{x(1), \cdots, x(n)\}$:

$$X = \left[\begin{array}{cccc} x(1) & x(2) & \cdots & x(n) \end{array} \right]$$

The Objective Function J

Explicit expansion of criterion J:

$$J = \sum_{(k,j)\in\mathcal{E}} W_{k,j}\|x(k)-x(j)\|^2 = \frac{1}{2}\sum_{k,j=1}^{n} W_{k,j}\|x(k)-x(j)\|^2 =$$

$$= \sum_{k,j=1}^{n} W_{k,j}\|x(k)\|^2 - \sum_{k,j=1}^{n} W_{k,j}x(k)^T x(j) =$$

$$= \sum_{k=1}^{n} D_{k,k}\left(x(k)^T x(k)\right) - \sum_{k,j} W_{k,j}\left(x(j)^T x(k)\right)$$

Remark $x(k)^T x(k) = (X^T X)_{k,k}$ and $x(j)^T x(k) = (X^T X)_{j,k}$. Next we write the sums in a more compact form using trace and matrix multiplication notations.

Traces and Commutation Relation

For a square matrix $M \in \mathbb{R}^{r\times r}$, its *trace* is defined as

$$trace(M) = \sum_{k=1}^{r} M_{k,k}$$

that is the sum of its diagonal elements.
For two matrices $A \in \mathbb{R}^{p\times q}$, $B \in \mathbb{R}^{q\times p}$:

$$trace(AB) = \sum_{k=1}^{p} (AB)_{k,k} = \sum_{k=1}^{p}\sum_{j=1}^{q} A_{k,j}B_{j,k} =$$
$$= \sum_{j=1}^{q}\sum_{k=1}^{p} B_{j,k}A_{k,j} = \sum_{j=1}^{q} (BA)_{j,j} = trace(BA).$$

This identity, $trace(AB) = trace(BA)$, allows to introduce an inner product on spaces of matrices of same size similar to the dot product betwen vectors of same size: If $U, V \in \mathbb{R}^{p\times q}$ then

$$\langle U, V\rangle = trace(U^T V) = trace(V^T U).$$

The Objective Function J - cont.

Return to J:

$$J = \sum_{k=1}^{n} D_{k,k}\left(x(k)^T x(k)\right) - \sum_{k,j} W_{k,j}\left(x(j)^T x(k)\right) =$$

$$= \sum_{k=1}^{n} D_{k,k}(X^T X)_{k,k} - \sum_{k,j} W_{k,j}(X^T X)_{j,k} = trace(DX^T X) - trace(WX^T X) =$$

$$= trace((D-W)X^T X) = trace(X \cdot \Delta \cdot X^T)$$

where the graph Laplacian $\Delta = D - W$ is defined in terms of the diagonal matrix $D = diag(W \cdot 1)$ and the weight matrix W.

Constraints

The objective is to minimize $J = X \cdot \Delta \cdot X^T$ over the $2 \times n$ matrix X. The global minimum is reached for instance by $X = 0$. This says that all points scrum in one location (the origin). To avoid this phenomenon we introduce constraints. First, each row of X should have norm 1. However there is a non-informative solution given by the constant matrix $\frac{1}{\sqrt{n}} 1_{2\times n}$: $\Delta 1_{n\times 2} = 0$. To avoid this case we ask that each row of X to be orthogonal to the constant vector 1, i.e. $X \cdot 1 = 0$. Lastly, to make sure the first row of X does not repeat in the second row, we ask them to be linearly independent. Even stronger, we ask the rows of X to be orthogonal vectors in $\mathbb{R}^n$. A compact form of these three conditions (normalization and orthogonalities):

$$XX^T = I_2 \quad , \quad X \cdot 1 = 0$$

Optimization Problem

Putting everything together, we obtain the optimization problem

$$\min_{\substack{X \in \mathbb{R}^{2,n} \\ X1 = 0 \\ XX^T = I_2}} trace(X \Delta X^T)$$

Optimization Problem

Putting everything together, we obtain the optimization problem

$$\min_{\substack{X \in \mathbb{R}^{2,n} \\ X1 = 0 \\ XX^T = I_2}} \quad trace(X\Delta X^T)$$

Luckily there is an easy algorithm to solve this problem. It is based on computing eigenpairs of the graph Laplacian Δ.

Graph Visualization Spectral Algorithm

Algorithm (Graph Visualization Spectral Algorithm)

Input: An adjacency matrix A or a weight matrix W.

1. *Compute the graph Laplacian* $\Delta = D - A$, *or* $\Delta = D - W$.
2. *Compute the lowest three eigenpairs* (e_1, λ_1), (e_2, λ_2), (e_3, λ_3), *where* $\Delta e_k = \lambda_k e_k$, $\|e_k\| = 1$, *and* $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3$.
3. *Construct the* $2 \times n$ *matrix* X

$$X = \begin{bmatrix} e_2^T \\ e_3^T \end{bmatrix}.$$

Output: Columns of matrix X are the n 2-dimensional vectors $\{x(1), \cdots, x(n)\}$.

Examples

See the Matlab simulations: circulant matrix case; perturbations.

Why the eigenpairs optimize the criterion?

The significant result: The Courant-Fisher criterion and Rayleight quotient.

Theorem

Assume T is a real symmetric $n \times n$ matrix. Then:

1. *All eigenvalues of T are real numbers.*
2. *There are n eigenvectors that can be normalized to form an orthonormal basis for $\mathbb{R}^n$.*
3. *The largest (principal) eigenpair (e_{max}, λ_{max}) and the smallest eigenpair (e_{min}, λ_{min}) satisfy*

$$(e_{max}), \lambda_{max}) = (arg) \max_{x \neq 0} \frac{\langle Tx, x \rangle}{\langle x, x \rangle} \quad , \quad (e_{min}, \lambda_{min}) = arg \min_{x \neq 0} \frac{\langle Tx, x \rangle}{\langle x, x \rangle}$$

Why the eigenpairs optimize the criterion? -cont

Theorem

Assume T is a real symmetric $n \times n$ matrix. Then:

4. *Assume $(e_1, ..., e_k)$ are the eigenvectors associated to the largest k eigenvalues. Then*

$$(e_{k+1}, \lambda_{k+1}) = arg \max_{x \neq 0, \langle x, e_1 \rangle = \cdots = \langle x, x_k \rangle = 0} \frac{\langle Tx, x \rangle}{\langle x, x \rangle}$$

5. *Assume $(e_{n-k+1}, ..., e_n)$ are the eigenvectors associated to the smallest k eigenvalues. Then*

$$(e_{n-k}, \lambda_{n-k}) = arg \min_{x \neq 0, \langle x, e_{n-k+1} \rangle = \cdots = \langle x, x_n \rangle = 0} \frac{\langle Tx, x \rangle}{\langle x, x \rangle}$$

Spectral Analysis
Basic Properties

We previously introduced: the Adjacency matrix A, the Degree matrix D, the (unnormalized symmetric) graph Laplacian matrix $\Delta = D - A$, the normalized Laplacian matrix $\tilde{\Delta} = D^{\dagger/2} \Delta D^{\dagger/2}$, and the normalized asymmetric Laplacian matrix $L = D^{\dagger} \Delta$.

Spectral Analysis
Basic Properties

We previously introduced: the Adjacency matrix A, the Degree matrix D, the (unnormalized symmetric) graph Laplacian matrix $\Delta = D - A$, the normalized Laplacian matrix $\tilde{\Delta} = D^{\dagger/2}\Delta D^{\dagger/2}$, and the normalized asymmetric Laplacian matrix $L = D^{\dagger}\Delta$.

We denote: n the number of vertices (also known as the *size* of the graph), m the number of edges, $d(v)$ the degree of vertex v, $d(i,j)$ the distance between vertex i and vertex j (length of the shortest path connecting i to j), and by *Diam* the diameter of the graph (the largest distance between two vertices = "longest shortest path").

Spectral Analysis

Basic Properties

Theorem (1)

1. $\Delta = \Delta^T \geq 0$, $\tilde{\Delta} = \tilde{\Delta}^T \geq 0$ *are positive semidefinite matrices.*
2. $eigs(\tilde{\Delta}) = eigs(L) \subset [0, 2]$*:*

$$0 = \lambda_{min}(\tilde{\Delta}) = \lambda_{min}(L) \leq \lambda_{max}(\tilde{\Delta}) = \lambda_{max}(L) \leq 2$$

3. 0 *is always an eigenvalue of* Δ, $\tilde{\Delta}$, L *with same multiplicity. Its multiplicity is equal to the number of connected components of the graph.*
4. *For the unnormalized graph Laplacian* Δ*:*

$$0 = \lambda_{min}(\Delta) \leq \lambda_{max}(\Delta) \leq 2 \max_v d(v),$$

i.e. the lagest eigenvalue of Δ *is bounded by twice the largest degree of the graph.*

Spectral Analysis
Basic Properties

Theorem (2)

Let $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1} \leq 2$ *be the eigenvalues of* $\tilde{\Delta}$ *(or* L*), that is* $eigs(\tilde{\Delta}) = \{\lambda_0, \lambda_1, \cdots, \lambda_{n-1}\} = eigs(L)$. *Then:*

1. $\sum_{i=0}^{n-1} \lambda_i \leq n$.
2. $\sum_{i=0}^{n-1} \lambda_i = n - \#\text{isolated vertices}$.
3. $\lambda_1 \leq \frac{n}{n-1}$.
4. $\lambda_1 = \frac{n}{n-1}$ *if and only if the graph is complete (i.e. any two vertices are connected by an edge).*
5. *If the graph is not complete then* $\lambda_1 \leq 1$.
6. *If the graph is connected then* $\lambda_1 > 0$. *If* $\lambda_i = 0$ *and* $\lambda_{i+1} \neq 0$ *then the graph has exactly* $i + 1$ *connected components.*
7. *If the graph is connected (no isolated vertices) then* $\lambda_{n-1} \geq \frac{n}{n-1}$.

Spectral Analysis

Smallest nonnegative eigenvalue

Theorem

Assume the graph is connected. Thus $\lambda_1 > 0$. Denote by D its diameter and by $d_{max}, \bar{d}, d_H$ the maximum, average, and harmonic avergae of the degrees $(d_1, \cdots, d_n)$:

$$d_{max} = \max_j d_j \quad , \quad \bar{d} = \frac{1}{n}\sum_{j=1}^{n} d_j \quad , \quad \frac{1}{d_H} = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{d_j}.$$

Then

1. $\lambda_1 \geq \frac{1}{nD}$.
2. *Let* $\mu = \max_{1 \leq j \leq n-1} |1 - \lambda_j|$. *Then*

$$1 + (n-1)\mu^2 \geq \frac{n}{d_H}(1 - (1+\mu)(\frac{\bar{d}}{d_H} - 1)).$$

Spectral Analysis
Smallest nonnegative eigenvalue

Theorem

[continued]

3. *Assume $D \geq 4$. Then*

$$\lambda_1 \leq 1 - 2\frac{\sqrt{d_{max} - 1}}{d_{max}}(1 - \frac{2}{D}) + \frac{2}{D}.$$

Spectral Analysis
Comments on the proof

"Ingredients" and key relations:

1. Let $f = (f_1, f_2, \cdots, f_n) \in \mathbb{R}^n$ be a n-vector. Then:

$$\langle \Delta f, f \rangle = \sum_{x \sim y} (f_x - f_y)^2$$

where $x \sim y$ if there is an edge between vertex x and vertex y (i.e. $A_{x,y} = 1$).
This proves positivity of all operators.

2. Last time we showed $eigs(\tilde{\Delta}) = eigs(L)$ because $\tilde{\Delta}$ and L are similar matrices.

3. 0 is an eigenvalue for Δ with eigenvector $1 = (1, 1, \cdots, 1)$. If multiple connected components, define such a 1 vector for each component (and 0 on rest).

4. $\lambda_{max}(\tilde{\Delta}) = 1 - \lambda_{min}(D^{-1/2}AD^{-1/2}) \leq 1 + |\lambda_{min}(D^{-1/2}AD^{-1/2})|$.

Spectral Analysis

Comments on the proof - 2

$$\lambda_{max}(D^{-1/2}AD^{-1/2}) = \max_{\|f\|=1} \langle D^{-1/2}AD^{-1/2}f, f\rangle = \max_{\|f\|=1} \sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}$$

$$\lambda_{min}(D^{-1/2}AD^{-1/2}) = \min_{\|f\|=1} \langle D^{-1/2}AD^{-1/2}f, f\rangle$$

$$|\lambda_{min,max}(D^{-1/2}AD^{-1/2})| \leq \max_{\|f\|=1} \left|\langle D^{-1/2}AD^{-1/2}f, f\rangle\right| = \max_{\|f\|=1} \left|\sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}\right|$$

Next use Cauchy-Schwartz to get

$$\left|\sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}\right| \leq \sum_i \frac{f_i^2}{d_i} \sum_j A_{i,j} = \sum_i f_i^2 = \|f\|^2 = 1.$$

Thus $\lambda_{max}(\tilde{\Delta}) \leq 2$. Similarly $\lambda_{max}(\Delta) \leq 2(\max_i d_i)$.

Spectral Analysis
Comments on the proof - 2

$$\lambda_{max}(D^{-1/2}AD^{-1/2}) = \max_{\|f\|=1} \langle D^{-1/2}AD^{-1/2}f, f \rangle = \max_{\|f\|=1} \sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}$$

$$\lambda_{min}(D^{-1/2}AD^{-1/2}) = \min_{\|f\|=1} \langle D^{-1/2}AD^{-1/2}f, f \rangle$$

$$|\lambda_{min,max}(D^{-1/2}AD^{-1/2})| \leq \max_{\|f\|=1} \left| \langle D^{-1/2}AD^{-1/2}f, f \rangle \right| = \max_{\|f\|=1} \left| \sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}} \right|$$

Next use Cauchy-Schwartz to get

$$\left| \sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}} \right| \leq \sum_i \frac{f_i^2}{d_i} \sum_j A_{i,j} = \sum_i f_i^2 = \|f\|^2 = 1.$$

Thus $\lambda_{max}(\tilde{\Delta}) \leq 2$. Similarly $\lambda_{max}(\Delta) \leq 2(\max_i d_i)$.

5. If the graph is connected, $trace(\tilde{\Delta}) = n = \sum_{i=0}^{n-1} \lambda_i$. Since $\lambda_0 = 0$ we

Spectral Analysis
Comments on the proof - 2

$$\lambda_{max}(D^{-1/2}AD^{-1/2}) = \max_{\|f\|=1} \langle D^{-1/2}AD^{-1/2}f, f\rangle = \max_{\|f\|=1} \sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}$$

$$\lambda_{min}(D^{-1/2}AD^{-1/2}) = \min_{\|f\|=1} \langle D^{-1/2}AD^{-1/2}f, f\rangle$$

$$|\lambda_{min,max}(D^{-1/2}AD^{-1/2})| \leq \max_{\|f\|=1} \left|\langle D^{-1/2}AD^{-1/2}f, f\rangle\right| = \max_{\|f\|=1} \left|\sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}\right|$$

Next use Cauchy-Schwartz to get

$$\left|\sum_{i,j} A_{i,j} \frac{f_i}{\sqrt{d_i}} \frac{f_j}{\sqrt{d_j}}\right| \leq \sum_i \frac{f_i^2}{d_i} \sum_j A_{i,j} = \sum_i f_i^2 = \|f\|^2 = 1.$$

Thus $\lambda_{max}(\tilde{\Delta}) \leq 2$. Similarly $\lambda_{max}(\Delta) \leq 2(\max_i d_i)$.

5. If the graph is connected, $trace(\tilde{\Delta}) = n = \sum_{i=0}^{n-1} \lambda_i$. Since $\lambda_0 = 0$ we

Spectral Analysis

Special graphs: Cycles and Complete graphs

Cycle graphs: like a regular polygon.

Remark: Adjacency matrices are circulant, and so are Δ, $\tilde{\Delta} = L$.

Spectral Analysis

Special graphs: Cycles and Complete graphs

Cycle graphs: like a regular polygon.
Remark: Adjacency matrices are circulant, and so are Δ, $\tilde{\Delta} = L$.

Then argue the FFT forms a ONB of eigenvectors. Compute the eigenvalues as FFT of the generating sequence.

Spectral Analysis
Special graphs: Cycles and Complete graphs

Cycle graphs: like a regular polygon.
Remark: Adjacency matrices are circulant, and so are Δ, $\tilde{\Delta} = L$.

Then argue the FFT forms a ONB of eigenvectors. Compute the eigenvalues as FFT of the generating sequence.

Consequence: The normalized Laplacian has the following eigenvalues:

1. For cycle graph on n vertices: $\lambda_k = 1 - cos\frac{2\pi k}{n}$, $0 \leq k \leq n-1$.
2. For the complete graph on n vertices:

$$\lambda_0 = 0\ ,\ \ \lambda_1 = \cdots = \lambda_{n-1} = \frac{n}{n-1}.$$

References

- B. Bollobás, **Graph Theory. An Introductory Course**, Springer-Verlag 1979. **99**(25), 15879–15882 (2002).
- F. Chung, **Spectral Graph Theory**, AMS 1997.
- F. Chung, L. Lu, The average distances in random graphs with given expected degrees, Proc. Nat.Acad.Sci. 2002.
- R. Diestel, **Graph Theory**, 3rd Edition, Springer-Verlag 2005.
- P. Erdös, A. Rényi, On The Evolution of Random Graphs
- G. Grimmett, **Probability on Graphs. Random Processes on Graphs and Lattices**, Cambridge Press 2010.
- J. Leskovec, J. Kleinberg, C. Faloutsos, Graph Evolution: Densification and Shrinking Diameters, ACM Trans. on Knowl.Disc.Data, **1**(1) 2007.