Lecture 5: The Cheeger Constant and the Spectral Gap

Radu Balan

Department of Mathematics, AMSC, CSCAMM and NWC
University of Maryland, College Park, MD

March 1, 2018

Eigenvalues of Laplacians
$\Delta, L, \tilde{\Delta}$

Today we discuss the spectral theory of graphs. Recall the Laplacian matrices:

$$\Delta = D - A \quad , \quad \Delta_{ij} = \begin{cases} d_i & if \quad i = j \\ -1 & if \quad (i,j) \in \mathcal{E} \\ 0 & otherwise \end{cases}$$

$$L = D^{-1}\Delta \quad , \quad L_{i,j} = \begin{cases} 1 & if \quad i = j \text{ and } d_i > 0 \\ -\frac{1}{d(i)} & if \quad (i,j) \in \mathcal{E} \\ 0 & otherwise \end{cases}$$

$$\tilde{\Delta} = D^{-1/2}\Delta D^{-1/2} \quad , \quad \tilde{\Delta}_{i,j} = \begin{cases} 1 & if \quad i = j \text{ and } d_i > 0 \\ -\frac{1}{\sqrt{d(i)d(j)}} & if \quad (i,j) \in \mathcal{E} \\ 0 & otherwise \end{cases}$$

Eigenvalues of Laplacians

$\Delta, L, \tilde{\Delta}$

Today we discuss the spectral theory of graphs. Recall the Laplacian matrices:

$$\Delta = D - A \quad , \quad \Delta_{ij} = \begin{cases} d_i & \textit{if} \quad i = j \\ -1 & \textit{if} \quad (i,j) \in \mathcal{E} \\ 0 & \textit{otherwise} \end{cases}$$

$$L = D^{-1}\Delta \quad , \quad L_{i,j} = \begin{cases} 1 & \textit{if} \quad i = j \text{ and } d_i > 0 \\ -\frac{1}{d(i)} & \textit{if} \quad (i,j) \in \mathcal{E} \\ 0 & \textit{otherwise} \end{cases}$$

$$\tilde{\Delta} = D^{-1/2}\Delta D^{-1/2} \quad , \quad \tilde{\Delta}_{i,j} = \begin{cases} 1 & \textit{if} \quad i = j \text{ and } d_i > 0 \\ -\frac{1}{\sqrt{d(i)d(j)}} & \textit{if} \quad (i,j) \in \mathcal{E} \\ 0 & \textit{otherwise} \end{cases}$$

Remark: $D^{-1}, D^{-1/2}$ are the pseudoinverses.

Eigenvalues of Laplacians

$\Delta, L, \tilde{\Delta}$

What do we know about the set of eigenvalues of these matrices for a graph G with n vertices?

1. $\Delta = \Delta^T \geq 0$ and hence its eigenvalues are non-negative real numbers.
2. $eigs(\tilde{\Delta}) = eigs(L) \subset [0, 2]$.
3. 0 is always an eigenvalue and its multiplicity equals the number of connected components of G,

$$dim\,ker(\Delta) = dim\,ker(L) = dim\,ker(\tilde{\Delta}) = \#connected\ \ components.$$

Eigenvalues of Laplacians

$\Delta, L, \tilde{\Delta}$

What do we know about the set of eigenvalues of these matrices for a graph G with n vertices?

1. $\Delta = \Delta^T \geq 0$ and hence its eigenvalues are non-negative real numbers.
2. $eigs(\tilde{\Delta}) = eigs(L) \subset [0, 2]$.
3. 0 is always an eigenvalue and its multiplicity equals the number of connected components of G,

$$dim\, ker(\Delta) = dim\, ker(L) = dim\, ker(\tilde{\Delta}) = \#connected\ \ components.$$

Let $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ be the eigenvalues of $\tilde{\Delta}$. Denote

$$\lambda(G) = \max_{1 \leq i \leq n-1} |1 - \lambda_i|.$$

Note $\sum_{i=1}^{n-1} \lambda_i = trace(\tilde{\Delta}) = n$. Hence the average eigenvalue is about 1. $\lambda(G)$ is called *the absolute gap* and measures the spread of eigenvalues away from 1.

The absolute spectral gap
$\lambda(G)$

The main result in [8]) says that for connected graphs w/h.p.:

$$\lambda_1 \geq 1 - \frac{C}{\sqrt{\text{Average Degree}}} = 1 - \frac{C}{\sqrt{p(n-1)}} = 1 - C\sqrt{\frac{n}{2m}}.$$

Theorem (For class $\mathcal{G}_{n,p}$)

Fix $\delta > 0$ and let $p > (\frac{1}{2} + \delta)\log(n)/n$. Let $d = p(n-1)$ denote the expected degree of a vertex. Let $\tilde{G}$ be the giant component of the Erdös-Rényi graph. For every fixed $\varepsilon > 0$, there is a constant $C = C(\delta, \varepsilon)$, so that

$$\max(|1-\lambda_1|, \lambda_{n-1} - 1) = \lambda(\tilde{G}) \leq \frac{C}{\sqrt{d}} = C\sqrt{\frac{n}{2m}}$$

with probability at least $1 - Cn\exp(-(2-\varepsilon)d) - C\exp(-d^{1/4}\log(n))$.

Connectivity threshold: $p \sim \frac{\log(n)}{n}$.

The absolute spectral gap

$\lambda(G)$

The main result in [8] says that for connected graphs w/h.p.:

$$\lambda_1 \geq 1 - \frac{C}{\sqrt{\text{Average Degree}}} = 1 - \frac{C}{\sqrt{p(n-1)}} = 1 - C\sqrt{\frac{n}{2m}}.$$

Theorem (For class $\Gamma^{n,m}$)

Fix $\delta > 0$ and let $m > \frac{1}{2}(\frac{1}{2} + \delta) n \log(n)$. Let $d = \frac{2m}{n}$ denote the expected degree of a vertex. Let $\tilde{G}$ be the giant component of the Erdös-Rényi graph. For every fixed $\varepsilon > 0$, there is a constant $C = C(\delta, \varepsilon)$, so that

$$\max(|1-\lambda_1|, \lambda_{n-1} - 1) = \lambda(\tilde{G}) \leq \frac{C}{\sqrt{d}} = C\sqrt{\frac{n}{2m}}$$

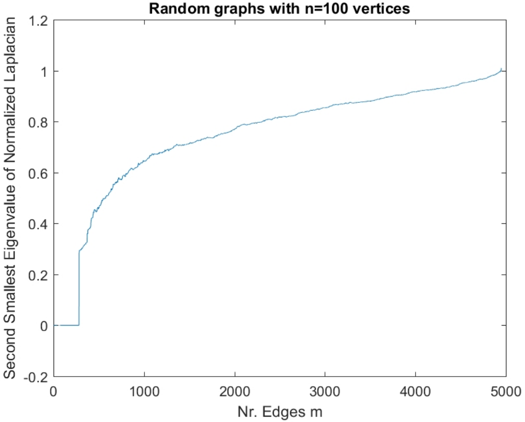
with probability at least $1 - Cn\exp(-(2-\varepsilon)d) - C\exp(-d^{1/4}\log(n))$.

Connectivity threshold: $m \sim \frac{1}{2} n \log(n)$.

Random graphs

λ_1 for random graphs

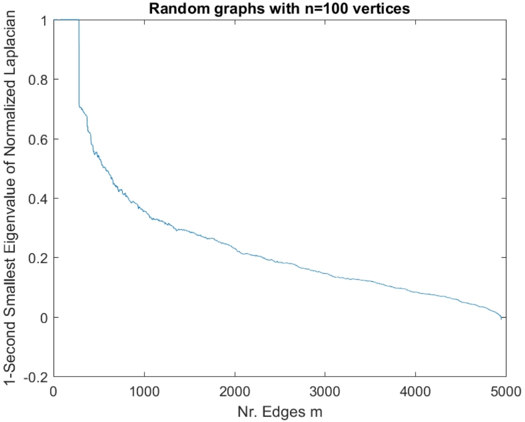
Results for $n = 100$ vertices: $\lambda_1(\tilde{G}) \approx 1 - \frac{C}{\sqrt{m}}$.



Random graphs

$1 - \lambda_1$ for random graphs

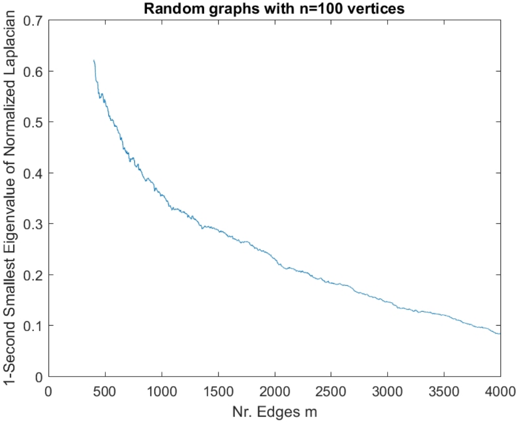
Results for $n = 100$ vertices: $1 - \lambda_1(\tilde{G}) \approx \frac{C}{\sqrt{m}}$.



Random graphs

$1-\lambda_1$ for random graphs

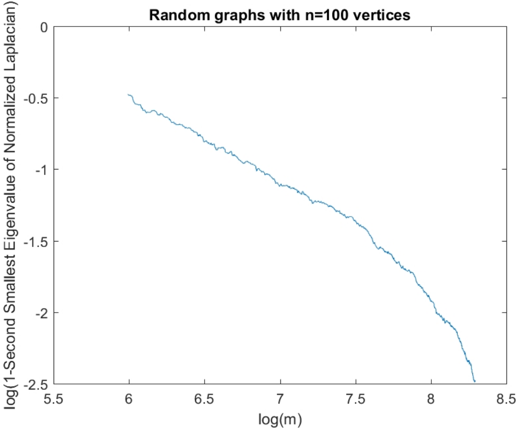
Results for $n=100$ vertices: $1-\lambda_1(\tilde{G}) \approx \frac{C}{\sqrt{m}}$. Detail.



Random graphs

$log(1-\lambda_1)$ vs. $log(m)$ for random graphs

Results for $n = 100$ vertices: $log(1-\lambda_1(\tilde{G})) \approx b_0 - \frac{1}{2}log(m)$.



AR(1) Prediction Model

Construction of weight matrix

See the high-res (FishReef 640x360) and low-res (10x10) movies.
Model: Use current frame ($x^{(t)}$) to predict the increment ($x^{(t+1)} - x^{(t)}$):

$$x^{(t+1)} = x^{(t)} + Wx^{(t)} + \nu^{(t)}$$

with constraints: $W = W^T$, $diag(W) = 0$, (potentially also) $W_{i,j} \geq 0$.

AR(1) Prediction Model

Construction of weight matrix

See the high-res (FishReef 640x360) and low-res (10x10) movies.
Model: Use current frame ($x^{(t)}$) to predict the increment ($x^{(t+1)} - x^{(t)}$):

$$x^{(t+1)} = x^{(t)} + Wx^{(t)} + \nu^{(t)}$$

with constraints: $W = W^T$, $diag(W) = 0$, (potentially also) $W_{i,j} \geq 0$.
The MLE/LSE (least-Squares) estimator minimizes:

$$W = \operatorname*{argmin}_{\substack{W = W^T \\ diag(W) = 0 \\ W_{i,j} \geq 0}} \sum_{t=1}^{T} \| Wx^{(t)} - (x^{(t+1)} - x^{(t)}) \|_2^2$$

AR(1) Prediction Model
CVX code

Criterion is expanded as:

$$J(W) = trace(WRW^T) - 2trace(WQ) + r_0 \quad , \quad R = \sum_t x^{(t)}(x^{(t)})^T$$

and then rewritten as

$$J(w) = w^T R_2 w - q^T w + r_0$$

where w is a 4950-long vector ($n = 100$).

AR(1) Prediction Model

CVX code

Criterion is expanded as:

$$J(W) = trace(WRW^T) - 2trace(WQ) + r_0 \quad , \quad R = \sum_t x^{(t)}(x^{(t)})^T$$

and then rewritten as

$$J(w) = w^T R_2 w - q^T w + r_0$$

where w is a 4950-long vector ($n = 100$).

To appreciate the convenience of CVX:

```
cvx_begin
      variable w(n2) nonnegative
      minimize (w' * R2 * w - q' *w + r0)
cvx_end
```

AR(1) Prediction Model

Weight matrix visualisation: $W_{i,j} \geq 0$.

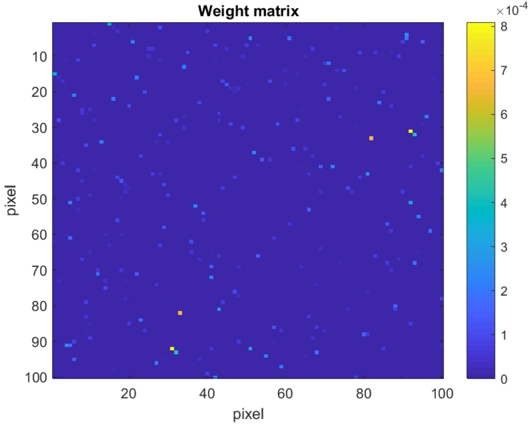


Figure: Weight matrix W for the 10x10 low resolution movie. Entries are color coded according to colormap.

AR(1) Prediction Model

Cumulative count of 3-cliques

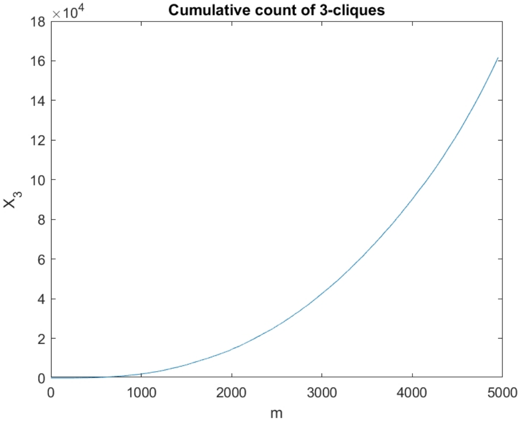


Figure: Cumulative count of 3-cliques, X_3, for $W = W^T$, $diag(W) = 0$, $W_{i,j} \geq 0$, order according to $W_{i,j}$

AR(1) Prediction Model

Analysis of the cumulative count of 3-cliques

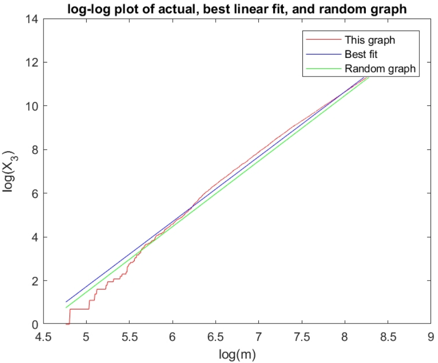


Figure: $log(X_3)$ vs. $log(m)$ for: the cumulative count of 3-cliques, best linear fit $a_0 log(m) + b_0$, $3log(m) + b_3$ (random graph)

Best linear fit:

$$y = 2.974 log(m) - 13.146.$$

Random graph:

$$\mathbb{E}[X_3] = \frac{4(n-2)}{3n^2(n-1)^2} m^3.$$

AR(1) Prediction Model

Cumulative count of 4-cliques

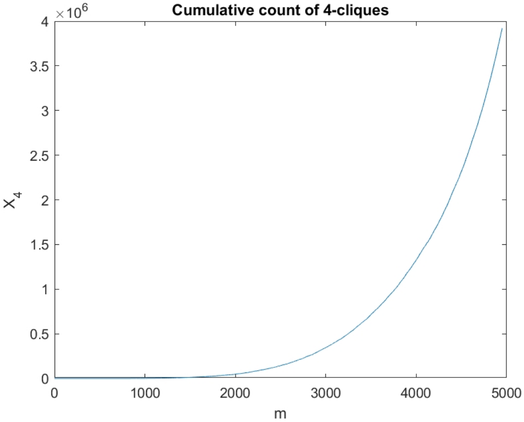


Figure: Cumulative count of 4-cliques, X_4, for $W = W^T$, $diag(W) = 0$, $W_{i,j} \geq 0$

AR(1) Prediction Model

Analysis of the cumulative count of 4-cliques

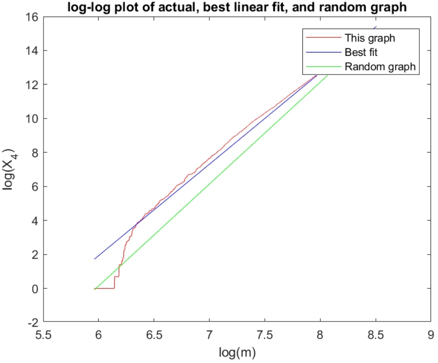


Figure: $log(X_4)$ vs. $log(m)$ for: the cumulative count of 4-cliques, best linear fit $a_0 log(m) + b_0$, $6log(m) + b_4$ (random graph)

Best linear fit:

$$y = 5.383 log(m) - 30.376.$$

Random graph:

$$\mathbb{E}[X_4] = \frac{8(n-2)(n-3)}{3n^5(n-1)^5} m^6.$$

AR(1) Prediction Model

Number of connected components

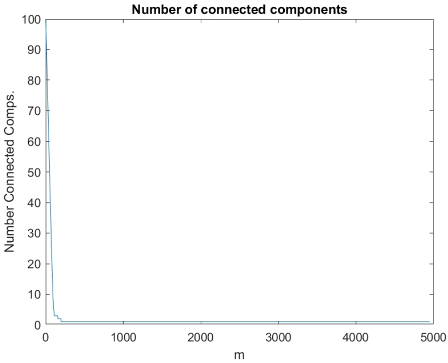


Figure: N_c vs. m, for $W = W^T$, $diag(W) = 0$ and $W_{i,j} \geq 0$

First connected graph at $m = 199$.
Connectivity threshold for random graphs:

$$m_c = \frac{1}{2} n \log(n) = 230.$$

AR(1) Prediction Model

Second Smallest Eigenvalue

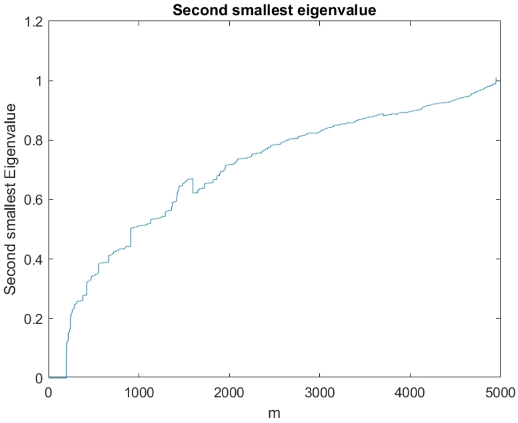


Figure: Distribution of the 2^{nd} smallest eigenvalue of normalized graph Laplacian, for $W = W^T$, $diag(W) = 0$ and $W_{i,j} \geq 0$

AR(1) Prediction Model

Analysis of the absolute spectral gap

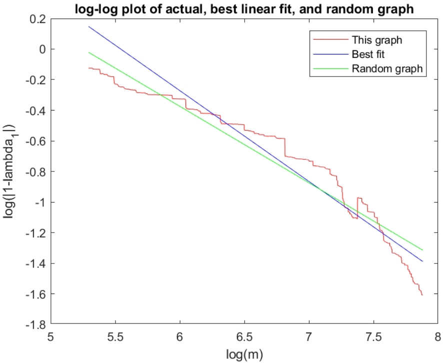


Figure: $log(\lambda(G)) = log(1-\lambda_1)$ vs. $log(m)$; best linear fit; random graph: $c - \frac{1}{2}log(m)$

Best linear fit:

$$y = -0.594 log(m) + 3.292.$$

Random graph:

$$log(1-\lambda_1) \approx c - \frac{1}{2} log(m).$$

(LSE: $c = 2.625$.)

AR(1) Prediction Model

Weight matrix visualisation: No sign constraints.

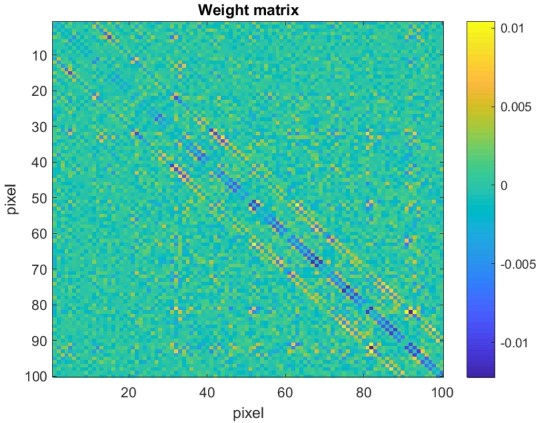


Figure: Weight matrix W for the 10x10 low resolution movie. Entries are color coded according to colormap.

AR(1) Prediction Model

Cumulative count of 3-cliques

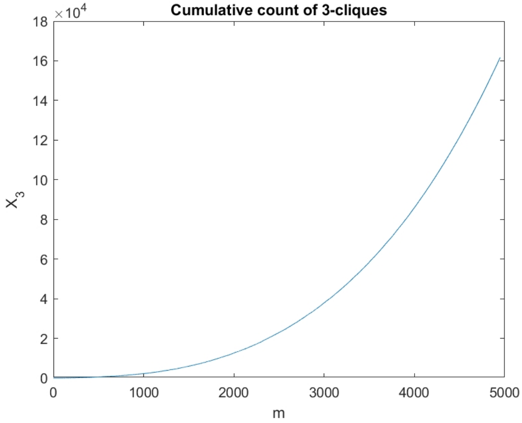


Figure: Cumulative count of 3-cliques, X_3, for $W = W^T$, $diag(W) = 0$, order according to $|W_{i,j}|$

AR(1) Prediction Model

Analysis of the cumulative count of 3-cliques

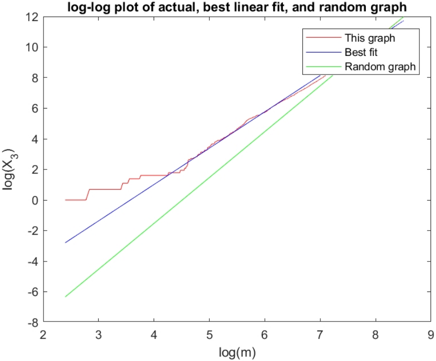


Figure: $log(X_3)$ vs. $log(m)$ for: the cumulative count of 3-cliques, best linear fit $a_0 log(m) + b_0$, $3log(m) + b_3$ (random graph)

Best linear fit:

$$y = 2.377 log(m) - 8.496.$$

Random graph:

$$\mathbb{E}[X_3] = \frac{4(n-2)}{3n^2(n-1)^2} m^3.$$

AR(1) Prediction Model

Cumulative count of 4-cliques

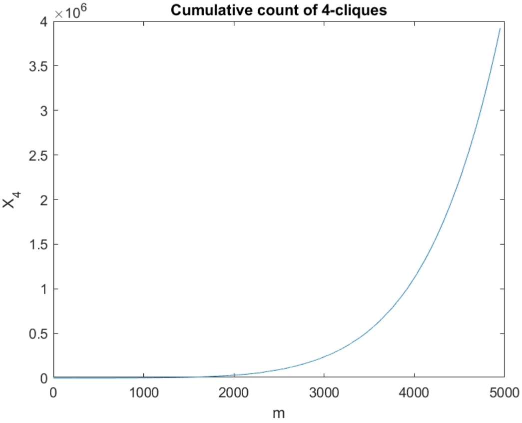


Figure: Cumulative count of 4-cliques, X_4, for $W = W^T$, $diag(W) = 0$

AR(1) Prediction Model

Analysis of the cumulative count of 4-cliques

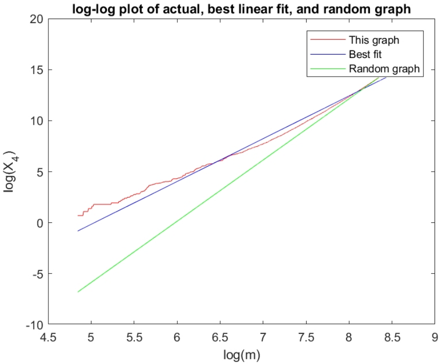


Figure: $log(X_4)$ vs. $log(m)$ for: the cumulative count of 4-cliques, best linear fit $a_0 log(m) + b_0$, $6log(m) + b_4$ (random graph)

Best linear fit:

$$y = 4.195 log(m) - 21.136.$$

Random graph:

$$\mathbb{E}[X_4] = \frac{8(n-2)(n-3)}{3n^5(n-1)^5} m^6.$$

AR(1) Prediction Model

Number of connected components

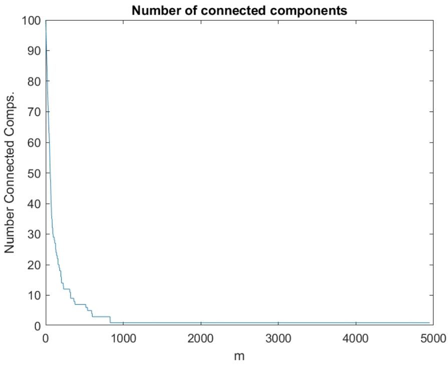


Figure: N_c vs. m, for $W = W^T$ and $diag(W) = 0$

First connected graph at $m = 833$.
Connectivity threshold for random graphs:

$$m_c = \frac{1}{2} n \log(n) = 230.$$

AR(1) Prediction Model

Second Smallest Eigenvalue

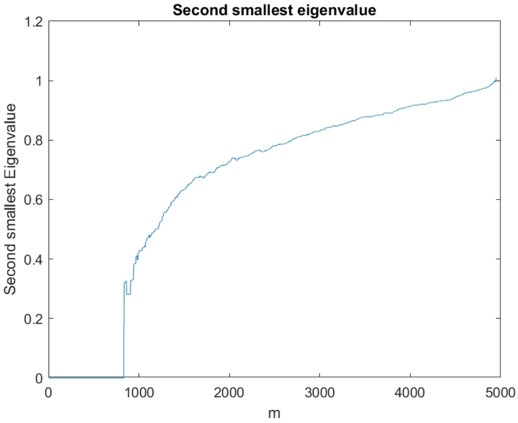


Figure: Distribution of the 2^{nd} smallest eigenvalue of normalized graph Laplacian, for $W = W^T$, $diag(W) = 0$

AR(1) Prediction Model

Analysis of the absolute spectral gap

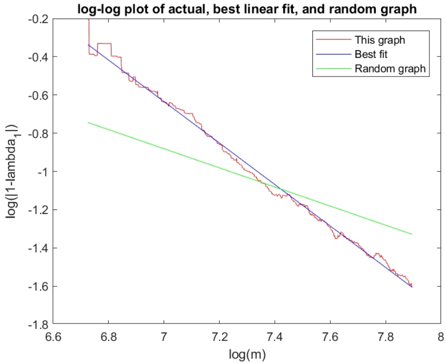


Figure: $log(\lambda(G)) = log(1 - \lambda_1)$ vs. $log(m)$; best linear fit; random graph: $c - \frac{1}{2}log(m)$

Best linear fit:

$$y = -1.084log(m)+6.958.$$

Random graph:

$$log(1-\lambda_1) \approx c - \frac{1}{2}log(m).$$

(LSE: $c = 2.618$.)

The absolute spectral gap
Proof

How to obtain such estimates? Following [4]:
First note: $\lambda_i = 1 - \lambda_i(D^{-1/2}AD^{-1/2})$. Thus

$$\lambda(G) = \max_{1 \leq i \leq n-1} |1 - \lambda_i| = \|D^{-1/2}AD^{-1/2}\| = \sqrt{\lambda_{max}((D^{-1/2}AD^{-1/2})^2)}$$

Ideas:

1. For $X = D^{-1/2}AD^{-1/2}$, and any positive integer $k > 0$,

$$\lambda_{max}(X^2) = \left(\lambda_{max}(X^{2k})\right)^{1/k} \leq \left(trace(X^{2k})\right)^{1/k}$$

2. (Markov's inequality)

$$Prob\{\lambda(G) > t\} = Prob\{\lambda(G)^{2k} > t^{2k}\} \leq \frac{1}{t^{2k}}\mathbb{E}[trace(X^{2k})].$$

The absolute spectral gap
Proof (2)

Consider the easier case when $D = dI$ (all vertices have the same degree):

$$\mathbb{E}[(X^{2k})] = \frac{1}{d^{2k}}\mathbb{E}[trace(A^{2k})].$$

The expectation turns into numbers of $2k$-cycles and loops. Combinatorial kicks in ...

The absolute spectral gap
Proof (2)

Consider the easier case when $D = dI$ (all vertices have the same degree):

$$\mathbb{E}[(X^{2k})] = \frac{1}{d^{2k}}\mathbb{E}[trace(A^{2k})].$$

The expectation turns into numbers of $2k$-cycles and loops. Combinatorial kicks in ...

Remark

An additional ingredient sometimes: Bernstein's "trick" for $X \geq 0$,

$$Prob\{X \leq t\} = Prob\{e^{-sX} \geq e^{-st}\} \leq \min_{s \geq 0} \frac{\mathbb{E}[e^{-sX}]}{e^{-st}}$$

$$= \min_{s \geq 0} e^{st} \int_0^\infty e^{-sx} p_X(x) dx$$

The absolute spectral gap
Proof (2)

Consider the easier case when $D = dI$ (all vertices have the same degree):

$$\mathbb{E}[(X^{2k})] = \frac{1}{d^{2k}}\mathbb{E}[trace(A^{2k})].$$

The expectation turns into numbers of $2k$-cycles and loops. Combinatorial kicks in ...

Remark

An additional ingredient sometimes: Bernstein's "trick" for $X \geq 0$,

$$Prob\{X \leq t\} = Prob\{e^{-sX} \geq e^{-st}\} \leq \min_{s \geq 0} \frac{\mathbb{E}[e^{-sX}]}{e^{-st}}$$

$$= \min_{s \geq 0} e^{st} \int_0^\infty e^{-sx} p_X(x) dx$$

(the "Laplace" method). It gives exponential decay instead of $\frac{1}{t}$ or $\frac{1}{t^2}$.

The Cheeger constant
Partitions

Fix a graph $G = (\mathcal{V}, \mathcal{E})$ with n vertices and m edges. We try to find an optimal partition $\mathcal{V} = A \cup B$ that minimizes a certain quantity.
Here are the concepts:

1. For two disjoint sets of vertices A abd B, $E(A,B)$ denotes the set of edges that connect vertices in A with vertices in B:
$$E(A,B) = \{(x,y) \in \mathcal{E} \quad , \quad x \in A \; , \; y \in B\}.$$
2. The *volume* of a set of vertices is the sum of its degrees:
$$vol(A) = \sum_{x \in A} d_x.$$
3. For a set of vertices A, denote $\bar{A} = \mathcal{V} \setminus A$ its complement.

The Cheeger constant
h_G

The *Cheeger constant* h_G is defined as

$$h_G = \min_{S \subset \mathcal{V}} \frac{|E(S, \bar{S})|}{min(vol(S), vol(\bar{S}))}.$$

Remark

It is a min edge-cut problem: This means, find the minimum number of edges that need to be cut so that the graph becomes disconnected, while the two connected components are not too small.
There is a similar min vertex-cut problem, where $E(S, \bar{S})$ is replaced by $\delta(S)$, the set of boundary points of S (the constant is denoted by g_G).

Remark

The graph is connected iff $h_G > 0$.

The Cheeger inequalities

h_G and λ_1

See [2](ch.2):

Theorem

For a connected graph

$$2h_G \geq \lambda_1 > 1 - \sqrt{1 - h_G^2} > \frac{h_G^2}{2}.$$

Equivalently:

$$\sqrt{2\lambda_1} > \sqrt{1 - (1 - \lambda_1)^2} > h_G \geq \frac{\lambda_1}{2}.$$

Why is it interesting: finding the exact h_G is a NP-hard problem.

The Cheeger inequalities
Proof of upper bound

Why the upper bound: $2h_G \geq \lambda_1$?
All starts from understanding what λ_1 is:

$$\Delta 1 = 0 \to \tilde{\Delta} D^{1/2} 1 = 0$$

Hence the eigenvector associated to $\lambda_0 = 0$ is

$$g^0 = (\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_n})^T.$$

The eigenpair (λ_1, g^1) is given by a solution of the following optimization problem:

$$\lambda_1 = \min_{h \perp g^0} \frac{\langle \tilde{\Delta} h, h \rangle}{\langle h, h \rangle}$$

In particular any h so that $\langle h, g^0 \rangle = \sum_{k=1}^n h_k \sqrt{d_k} = 0$ satisfies

$$\langle \tilde{\Delta} h, h \rangle \geq \lambda_1 \|h\|^2.$$

The Cheeger inequalities
Proof of upper bound (2)

Assume that we found the optimal partition $(A = S, B = \bar{S})$ of $\mathcal{V}$ that minimizes the edge-cut.
Define the following particular n-vector:

$$h_k = \begin{cases} \frac{\sqrt{d_k}}{vol(A)} & if \quad k \in A = S \\ -\frac{\sqrt{d_k}}{vol(B)} & if \quad k \in B = \mathcal{V} \setminus S \end{cases}$$

One checks that $\sum_{k=1}^{n} h_k \sqrt{d_k} = 1 - 1 = 0$, and $\|h\|^2 = \frac{1}{vol(A)} + \frac{1}{vol(B)}$.
But:

$$\langle \tilde{\Delta} h, h \rangle = \sum_{(i,j):A_{i,j}=1} \left(\frac{h_i}{\sqrt{d_i}} - \frac{h_j}{\sqrt{d_j}}\right)^2 = |E(A,B)| \left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right)^2.$$

Thus:

$$2h_G = \frac{2|E(A,B)|}{min(vol(A), vol(B))} \geq |E(A,B)| \left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right) \geq \lambda_1.$$

Min-cut Problems

Initialization

The proof of the upper bound in Cheeger inequality reveals a "good" initial guess of the optimal partition:

1. Compute the eigenpair (λ_1, g^1) associated to the second smallest eigenvalue;
2. Form the partition:

$$S = \{k \in \mathcal{V} \ \ , \ \ g_k^1 \geq 0\} \ \ , \ \ \bar{S} = \{k \in \mathcal{V} \ \ , \ \ g_k^1 < 0\}$$

Min-cut Problems
Weighted Graphs

The Cheeger inequality holds true for weighted graphs, $G = (\mathcal{V}, \mathcal{E}, W)$.

- $\Delta = D - W$, $D = diag(w_i)_{1 \leq i \leq n}$, $w_i = \sum_{j \neq i} w_{i,j}$
- $\tilde{\Delta} = D^{-1/2} \Delta D^{-1/2} = I - D^{-1/2} W D^{-1/2}$
- $eigs(\tilde{\Delta}) \subset [0, 2]$
- $h_G = \min_S \frac{\sum_{x \in S, y \in \bar{S}} W_{x,y}}{min(\sum_{x \in S} W_{x,x}, \sum_{y \in \bar{S}} W_{y,y})}$
- $2h_G \geq \lambda_1 \geq 1 - \sqrt{1 - h_G^2}$
- Good initial guess for optimal partition: Compute the eigenpair (λ_1, g^1) associated to the second smallest eigenvalue of $\tilde{\Delta}$; set:

$$S = \{k \in \mathcal{V} \quad , \quad g_k^1 \geq 0\} \quad , \quad \bar{S} = \{k \in \mathcal{V} \quad , \quad g_k^1 < 0\}$$

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.
- F. Chung, L. Lu, V. Vu, The spectra of random graphs with Given Expected Degrees, Internet Math. **1**(3), 257–275 (2004).
- 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.

C. Hoffman, M. Kahle, E. Paquette, Spectral Gap of Random Graphs and Applications to Random Topology, arXiv: 1201.0425 [math.CO] 17 Sept. 2014.

J. Leskovec, J. Kleinberg, C. Faloutsos, Graph Evolution: Densification and Shrinking Diameters, ACM Trans. on Knowl.Disc.Data, **1**(1) 2007.