A Probabilistic Excursion in Spectral Graph Theory

Garrett J. Kepler
Washington State University
Co.L.A.Nu.T.S.
April 7th, 2025

Notation

Definition

Notation

Definition

Let $G = (V, E)$ be an undirected graph on n nodes. The *Laplacian* of G is the matrix $L = L(G) \in M_n$ defined as follows:

$$L_{ij} = \begin{cases} \text{degree of node } i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and node } i \sim \text{node } j \\ 0 & \text{if } i \neq j \text{ and node } i \nsim \text{node } j \end{cases}$$

with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Notation

Definition

Let $G = (V, E)$ be an undirected graph on n nodes. The *Laplacian* of G is the matrix $L = L(G) \in M_n$ defined as follows:

$$L_{ij} = \begin{cases} \text{degree of node } i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and node } i \sim \text{node } j \\ 0 & \text{if } i \neq j \text{ and node } i \nsim \text{node } j \end{cases}$$

with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Properties:

Notation

Definition

Let $G = (V, E)$ be an undirected graph on n nodes. The *Laplacian* of G is the matrix $L = L(G) \in M_n$ defined as follows:

$$L_{ij} = \begin{cases} \text{degree of node } i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and node } i \sim \text{node } j \\ 0 & \text{if } i \neq j \text{ and node } i \nsim \text{node } j \end{cases}$$

with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Properties:

- L is symmetric, positive-semidefinite $\implies$ L has nonnegative eigenvalues

Notation

Definition

Let $G = (V, E)$ be an undirected graph on n nodes. The *Laplacian* of G is the matrix $L = L(G) \in M_n$ defined as follows:

$$L_{ij} = \begin{cases} \text{degree of node } i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and node } i \sim \text{node } j \\ 0 & \text{if } i \neq j \text{ and node } i \nsim \text{node } j \end{cases}$$

with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Properties:

- L is symmetric, positive-semidefinite $\implies$ L has nonnegative eigenvalues
- 0 is an eigenvalue of L corresponding to the all ones vector $\mathbb{1}$

Why spectra?

Why spectra?

Theorem

For a Laplacian L with eigenvalues λ_i:

- $\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$
- $\lambda_n \leq \max\{d_i + d_j : i \sim j\}$
- $\lambda_2 \geq \frac{n}{\Delta}$

$\vdots$

Why spectra?

Theorem

For a Laplacian L with eigenvalues λ_i:

- $\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$
- $\lambda_n \leq \max\{d_i + d_j : i \sim j\}$
- $\lambda_2 \geq \frac{n}{\Delta}$

$\vdots$

Laplacian spectrum has applications in:

- numerous clustering methods
- network characterization
- signal processing

$\vdots$

Motivation

Motivation

Computing eigenvalues is expensive! In fact, finding exactly the eigenvalues of the $n \times n$ Laplacian is generally $\mathcal{O}(n^3)$.

Motivation

Computing eigenvalues is expensive! In fact, finding exactly the eigenvalues of the $n \times n$ Laplacian is generally $\mathcal{O}(n^3)$.

Question

- How do I minimize cost? i.e. how little information do I need to reconstruct the spectra?

Motivation

Computing eigenvalues is expensive! In fact, finding exactly the eigenvalues of the $n \times n$ Laplacian is generally $\mathcal{O}(n^3)$.

Question

- How do I minimize cost? i.e. how little information do I need to reconstruct the spectra?
- How can I at least approximate my spectra efficiently?

Motivation

Computing eigenvalues is expensive! In fact, finding exactly the eigenvalues of the $n \times n$ Laplacian is generally $\mathcal{O}(n^3)$.

Question

- How do I minimize cost? i.e. how little information do I need to reconstruct the spectra?
- How can I at least approximate my spectra efficiently?
- What information do I need to approximate my spectra?

Motivation

Computing eigenvalues is expensive! In fact, finding exactly the eigenvalues of the $n \times n$ Laplacian is generally $\mathcal{O}(n^3)$.

Question

- How do I minimize cost? i.e. how little information do I need to reconstruct the spectra?
- How can I at least approximate my spectra efficiently?
- What information do I need to approximate my spectra?

Definition

Given a Laplacian L with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ we define the j-th moment of the empirical spectral measure μ as

$$\mu(\lambda^j) \equiv \int \lambda^j d\mu(x)$$

Definition

Given a Laplacian L with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ we define the j-th moment of the empirical spectral measure μ as

$$\mu(\lambda^j) \equiv \int \lambda^j d\mu(x)$$

What is typically exploited is the equivalence to the trace of L^j:

$$\mu(\lambda^j) \equiv \int \lambda^j d\mu(x) = \frac{1}{n} \sum_i \lambda_i^j = \frac{1}{n} tr(L^j)$$

Definition

Given a Laplacian L with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ we define the j-th moment of the empirical spectral measure μ as

$$\mu(\lambda^j) \equiv \int \lambda^j d\mu(x)$$

What is typically exploited is the equivalence to the trace of L^j:

$$\mu(\lambda^j) \equiv \int \lambda^j d\mu(x) = \frac{1}{n}\sum_i \lambda_i^j = \frac{1}{n} tr(L^j)$$

In other words, to estimate the moments of the empirical spectral measure (and hence the distribution of eigenvalues of L), all we need is an estimation of $tr(L^j)$!

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:
    1. Draw i.i.d. vectors $x_1, \ldots, x_m \in \{1, -1\}^n$

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:
    1. Draw i.i.d. vectors $\mathrm{x}_1, \ldots, \mathrm{x}_m \in \{1, -1\}^n$
    2. Compute $\tilde{T} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{x}_i^T A \mathrm{x}_i$

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:
    1. Draw i.i.d. vectors $x_1, \dots, x_m \in \{1, -1\}^n$
    2. Compute $\tilde{T} = \frac{1}{m} \sum_{i=1}^m x_i^T A x_i$
  - For some random draw $x \in \{1, -1\}^n$,

$$\begin{aligned}\mathbb{E}[\tilde{T}] &= \mathbb{E}[x^T A x] \\ &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[x_i x_j A_{ij}] \\ &= \sum_{i=1}^n A_{ii}\end{aligned}$$

i.e. that $\mathbb{E}[\tilde{T}] = tr(A)$

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:
    1. Draw i.i.d. vectors $x_1, \ldots, x_m \in \{1, -1\}^n$
    2. Compute $\tilde{T} = \frac{1}{m} \sum_{i=1}^{m} x_i^T A x_i$
  - For some random draw $x \in \{1, -1\}^n$,

$$\mathbb{E}[\tilde{T}] = \mathbb{E}[x^T A x]$$

$$= \sum_{i=1}^{n} A_{ii}$$

i.e. that $\mathbb{E}[\tilde{T}] = tr(A)$

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:
    1. Draw i.i.d. vectors $\mathrm{x}_1, \dots, \mathrm{x}_m \in \{1, -1\}^n$
    2. Compute $\tilde{T} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{x}_i^T A \mathrm{x}_i$
  - For some random draw $\mathrm{x} \in \{1, -1\}^n$,

$$\begin{aligned} \mathbb{E}[\tilde{T}] &= \mathbb{E}[\mathrm{x}^T A \mathrm{x}] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[x_i x_j A_{ij}] \\ &= \sum_{i=1}^{n} A_{ii} \end{aligned}$$

- Matrix multiplication is cheaper! ($\mathcal{O}(n^2)$)
- Hutchinson-Girard Estimators (1991 & 1987):
  - Process:
    1. Draw i.i.d. vectors $x_1, \ldots, x_m \in \{1, -1\}^n$
    2. Compute $\tilde{T} = \frac{1}{m} \sum_{i=1}^{m} x_i^T A x_i$
  - For some random draw $x \in \{1, -1\}^n$,

$$\begin{aligned}\mathbb{E}[\tilde{T}] &= \mathbb{E}[x^T A x] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}[x_i x_j A_{ij}] \\ &= \sum_{i=1}^{n} A_{ii}\end{aligned}$$

i.e. that $\mathbb{E}[\tilde{T}] = tr(A)$

Theorem

Set $m = \mathcal{O}(\frac{1}{\epsilon^2})$. Then, the trace estimate $\tilde{T}$ for the Laplacian L satisfies:

$$(1-\epsilon)tr(L) \leq \tilde{T} \leq (1+\epsilon)tr(L)$$

Theorem

Set $m = \mathcal{O}(\frac{1}{\epsilon^2})$. Then, the trace estimate $\tilde{T}$ for the Laplacian L satisfies:

$$(1-\epsilon)tr(L) \leq \tilde{T} \leq (1+\epsilon)tr(L)$$

These bounds can actually be improved by combining brute force computation and Huthcinson's approach in Hutch++ [Musco et al].

Theorem (Hutch++)

Consider the following algorithm on a Laplacian L (Hutch++):

Theorem (Hutch++)

Consider the following algorithm on a Laplacian L (Hutch++):

1. *Find approximate span for top k eigenvectors Q*

Theorem (Hutch++)

Consider the following algorithm on a Laplacian L (Hutch++):

1. *Find approximate span for top k eigenvectors* Q
2. *Approximate* $\tilde{P} = tr(L(I - QQ^T))$ *using Hutchinson's.*

Theorem (Hutch++)

Consider the following algorithm on a Laplacian L (Hutch++):

1. *Find approximate span for top k eigenvectors* Q
2. *Approximate* $\tilde{P} = tr(L(I - QQ^T))$ *using Hutchinson's.*
3. *Return* $\tilde{T} = tr(LQQ^T) + \tilde{P}$

Theorem (Hutch++)

Consider the following algorithm on a Laplacian L (Hutch++):

1. *Find approximate span for top k eigenvectors Q*
2. *Approximate* $\tilde{P} = tr(L(I - QQ^T))$ *using Hutchinson's.*
3. *Return* $\tilde{T} = tr(LQQ^T) + \tilde{P}$

For a given L with $m = \mathcal{O}\left(\frac{\log(1/\delta)}{\epsilon}\right)$, *Hutch++ returns* $\tilde{T}$ *satisfying*

$$(1 - \epsilon)tr(L) \leq \tilde{T} \leq (1 + \epsilon)tr(L)$$

with probability $1 - \delta$

Typical framework:

Typical framework:

- Estimate: estimate cumulative density function of the distribution via moments of the distribution of eigenvalues

Typical framework:

- Estimate: estimate cumulative density function of the distribution via moments of the distribution of eigenvalues
- Reconstruct: recover the cumulative density function from the estimation step

Definition

Let G be a connected, undirected, weighted graph on n nodes and m edges.

Definition

Let G be a connected, undirected, weighted graph on n nodes and m edges.A *tree* in G is a connected subgraph with no cycles.

Definition

Let G be a connected, undirected, weighted graph on n nodes and m edges.A *tree* in G is a connected subgraph with no cycles.A *spanning tree* is a tree such that every node in G is also in the tree.

Definition

Let G be a connected, undirected, weighted graph on n nodes and m edges.A *tree* in G is a connected subgraph with no cycles.A *spanning tree* is a tree such that every node in G is also in the tree.A *forest* is a union of disjoint trees. A *spanning forest* is a forest such that every node in G is also in the forest.

Definition

Let G be a connected, undirected, weighted graph on n nodes and m edges.A *tree* in G is a connected subgraph with no cycles.A *spanning tree* is a tree such that every node in G is also in the tree.A *forest* is a union of disjoint trees. A *spanning forest* is a forest such that every node in G is also in the forest.A *rooted spanning forest* is a spanning forest such that on node per tree is picked to represent the tree and labeled as “root”.

Theorem (Kirchoff's Matrix Tree Theorem)

Let T be the set of all spanning trees of a graph G with Laplacian L. Then

$$Z = |T| = \frac{det(L[i])}{n} = \frac{1}{n}\prod_{i \neq 1} \lambda_i$$

where $L[i]$ is the matrix obtained by removing the i-th row and column from L.

Definition

Consider the following distribution over T:

$$\mathbb{P}[\mathcal{T} = t] = \frac{1}{Z} \prod_{(i,j) \in t} w_{ij}$$

where w_{ij} is the edgeweight of (i, j) in the spanning tree t. The random object $\mathcal{T}$ is called a *uniform spanning tree*.

Definition

Let ϕ be a forest with edge set $E(\phi)$ and roots $\rho(\phi)$.A Kirchoff Forest is a distribution on rooted spanning forests with the following probability mass function:

$$\mathbb{P}[\Phi_q = \phi] = \frac{q^{|\rho|}}{Z_q} \prod_{(i,j)\in E(\phi)} w_{ij}$$

where Z_q is a normalization constant and q is a paramter controlling the average number of roots in the random forest (and thus the number of trees).

Estimation process:

Estimation process:

- Take k independent Kirchoff Forests wth paramter q.

Estimation process:

- Take k independent Kirchoff Forests wth paramter q.
- Keep track of where each node gets sent at each iteration.

Estimation process:

- Take k independent Kirchoff Forests wth paramter q.
- Keep track of where each node gets sent at each iteration.
- Count the number of times each node ended up as the root at the k-th forest and divide by the number of nodes.

Estimation process:

- Take k independent Kirchoff Forests wth paramter q.
- Keep track of where each node gets sent at each iteration.
- Count the number of times each node ended up as the root at the k-th forest and divide by the number of nodes.

In expectation, this process equals the trace of the k-th moment of L!

Preliminaries

Preliminaries

Definition

A matrix $A \in M_n$ is *stochastic* if every row sums to 1. Likewise, A is *positive stochastic* if every row sums to 1 and every entry is positive.

Preliminaries

Definition

A matrix $A \in M_n$ is *stochastic* if every row sums to 1. Likewise, A is *positive stochastic* if every row sums to 1 and every entry is positive.

Viewing a stochastic matrix as a representation of a distribution over a state space is straight forward:

- Label a finite set of potential states $1, \ldots, n$.

Preliminaries

Definition

A matrix $A \in M_n$ is *stochastic* if every row sums to 1. Likewise, A is *positive stochastic* if every row sums to 1 and every entry is positive.

Viewing a stochastic matrix as a representation of a distribution over a state space is straight forward:

- Label a finite set of potential states $1, \ldots, n$.
- Assign each matrix entry A_{ij} the corresponding transition probability $\mathbb{P}[i \to j]$.

Preliminaries

Definition

A matrix $A \in M_n$ is *stochastic* if every row sums to 1. Likewise, A is *positive stochastic* if every row sums to 1 and every entry is positive.

Viewing a stochastic matrix as a representation of a distribution over a state space is straight forward:

- Label a finite set of potential states $1, \ldots, n$.
- Assign each matrix entry A_{ij} the corresponding transition probability $\mathbb{P}[i \to j]$.

Definition

A polynomial $p(x)$ is said to *preserve nonnegativity* if for a nonnegative matrix $A \in M_n$, $p(A) \geq 0$.

Definition

A polynomial $p(x)$ is said to *preserve nonnegativity* if for a nonnegative matrix $A \in M_n$, $p(A) \geq 0$.

- In 1978, Loewy and London put forth the idea that the cone of polynomials that preserve the nonnegativity of matrices may play an important role in the solution to the Nonnegative Inverse Eigenvalue Problem (NIEP).

Definition

A polynomial $p(x)$ is said to *preserve nonnegativity* if for a nonnegative matrix $A \in M_n$, $p(A) \geq 0$.

- In 1978, Loewy and London put forth the idea that the cone of polynomials that preserve the nonnegativity of matrices may play an important role in the solution to the Nonnegative Inverse Eigenvalue Problem (NIEP).
- In 2024, Kepler, Brannen and Clark showed that a polynomial preserves the nonnegativity of a matrix if and only if the polynomial preserves the nonnegativity of kA for every positive scalar k and positive stochastic matrix A.

Definition

A polynomial $p(x)$ is said to *preserve nonnegativity* if for a nonnegative matrix $A \in M_n$, $p(A) \geq 0$.

- In 1978, Loewy and London put forth the idea that the cone of polynomials that preserve the nonnegativity of matrices may play an important role in the solution to the Nonnegative Inverse Eigenvalue Problem (NIEP).
- In 2024, Kepler, Brannen and Clark showed that a polynomial preserves the nonnegativity of a matrix if and only if the polynomial preserves the nonnegativity of kA for every positive scalar k and positive stochastic matrix A.

Theorem

The largest $t \in \mathbb{R}^+$ such that

$$\sum_{k=0}^{n-1} x^k - tx^m + \sum_{k=2m-n+1}^{2m} x^k$$

preserves nonnegativity for all $m \geq n$ is $t = 2$.

Stochastic Matrix Inequalities

Proof:

- Form a graph from the positive stochastic A.

Stochastic Matrix Inequalities

Proof:

- Form a graph from the positive stochastic A.
- Since A is positive, there exists a path of *every* length from any node i to any other node j.

Stochastic Matrix Inequalities

Proof:

- Form a graph from the positive stochastic A.
- Since A is positive, there exists a path of *every* length from any node i to any other node j.
- With each successive power k of A, we collect the products of edgeweights along paths of length k.

Stochastic Matrix Inequalities

Proof:

- Form a graph from the positive stochastic A.
- Since A is positive, there exists a path of *every* length from any node i to any other node j.
- With each successive power k of A, we collect the products of edgeweights along paths of length k.
- Pick some nodes i and j with a path of length $m - \ell$ between them. Let $\rho^{m-\ell}$ be the product of their edgeweights.

Stochastic Matrix Inequalities

Proof:

- Form a graph from the positive stochastic A.
- Since A is positive, there exists a path of *every* length from any node i to any other node j.
- With each successive power k of A, we collect the products of edgeweights along paths of length k.
- Pick some nodes i and j with a path of length $m - \ell$ between them. Let $\rho^{m-\ell}$ be the product of their edgeweights.

Proof (continued):

Proof (continued):

- As we take longer walks on our graph heading toward a length $m + \ell$, we account for the off-path weight with say w.

Proof (continued):

- As we take longer walks on our graph heading toward a length $m + \ell$, we account for the off-path weight with say w.
- Plugging our matrix into our polynomial, we obtain the ij-th entry of $p(A)$ to have in its sum

$$\rho^{m-\ell} - tw\rho^{m} + w^{2}\rho^{m+\ell} = \rho^{m-\ell}\left(1 - t(w\rho^{\ell}) + (w\rho^{\ell})^{2}\right).$$

which is nonnegative for all $0 < t \leq 2$.

Proof (continued):

- As we take longer walks on our graph heading toward a length $m+\ell$, we account for the off-path weight with say w.
- Plugging our matrix into our polynomial, we obtain the ij-th entry of $p(A)$ to have in its sum

$$\rho^{m-\ell} - tw\rho^{m} + w^2\rho^{m+\ell} = \rho^{m-\ell}\left(1 - t(w\rho^{\ell}) + (w\rho^{\ell})^2\right).$$

which is nonnegative for all $0 < t \leq 2$.

- There's ample room for matrix, graph, and probability theory to work together!