

Analysis and Quantitative Evaluation of Large Networks
Using Spectral Distributions
スペクトル分布を用いた大規模ネットワークの分析と定量的
評価について

by

Naoki Murakami
村上直輝

A Senior Thesis
卒業論文

Submitted to
the Department of Information Science
the Faculty of Science, the University of Tokyo
on January 29, 2021
in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science

Thesis Supervisor: Hiroshi Imai 今井浩
Professor of Information Science

ABSTRACT

To analyze the properties and structure of a network, one of the very important tools is the spectrum, or eigenvalue distribution. However, since computing the full eigendecomposition is expensive, large networks have been analyzed by using only part of the spectrum. Recently, a method for calculating the entire eigenvalue distribution has been developed, but the analysis and application of the spectrum have not yet been deeply considered.

In this paper, we firstly discuss the quantitative evaluation of spectra and its applications. For quantitative evaluation and comparison of spectral densities, we suggest applying the concepts of divergence and metric of discrete probability distributions to spectral distributions. This allows an intuitive comparison based on the structure of networks. As an application, a fast search algorithm for similar spectra using cosine distance is also discussed. Secondly, we analyze the distribution of the overall spectrum of theoretical graphs and real-world networks. Then, we discuss the relation between the network properties and the spectrum. As a result, for sparse networks, there is a large correspondence between the local structure of the network, called a motif, and the distribution of eigenvalues. In addition, we find that a graph in which vertices connected to neighbors have a sharp spectrum found in complex networks. Furthermore, we illustrate that graphs with vertices connected randomly have the semicircular spectra found in Erdős-Rényi random graphs.

This study explains a quantitative evaluation and application of spectra and shows that the shape of the spectra can give us information about the network properties. This means that the spectral analysis is an effective tool for clarifying the network properties.

論文要旨

ネットワークの性質や構造を分析する上で、スペクトルは非常に重要な指標となり得る。しかし、大規模ネットワークについては、計算量の問題からスペクトルの一部を利用した分析しか行われていなかった。近年スペクトル全体を計算する手法が考案されたが、スペクトルの分析や応用などはまだ深く考えられていない。

本研究では、はじめにスペクトルの定量的評価とその応用方法について議論する。スペクトル分布の定量的な評価や比較のために、離散確率分布に対するダイバージェンスや距離の概念を流用することを提案する。これにより、ネットワークの構造に基づいた直感的な比較が可能になる。応用方法として cosine distance を用いた類似のスペクトルを高速に検索する手法についても言及する。次に、理論的に得られるグラフと実世界のネットワークのスペクトル全体の分布を分析し、ネットワークの性質とスペクトルの対応について考察する。結果として、平均次数が小さい疎なネットワークについては、モチーフと呼ばれるネットワークの局所的な構造と固有値の分布が大きく対応していることが分かった。また、近くの頂点同士が隣接するようなグラフは、複雑ネットワークに見られるような鋭いスペクトルを持つ。さらに、頂点同士の隣接にランダム性があるグラフは、Erdős-Rényi ランダムグラフに見られる半円状のスペクトルになることを例示する。

本研究はスペクトルの定量的評価や応用法について提案するとともに、スペクトルの形状からネットワークの性質についての情報が得られることを示した。これによりスペクトルの分析がネットワークの性質解明に有効な手段であることがわかる。

Acknowledgements

First of all, I really appreciate supports by Prof. Imai to proceed with my research, to speak in seminar, and to write the thesis. His helpful advice from insight has taught me how to work on research. I am also grateful to Assist. Prof. Hiraishi for giving me suggestions and taking care of me. Finally, I would like to thank all members of Imai Laboratory. When I got stuck in my research, they always gave me helpful technical advice.

Contents

1	Introduction	1
1.1	Complex Network	1
1.2	von Neumann entropy of Network	1
2	Preliminaries	2
2.1	General Notation	2
2.2	The Laplacian and Eigenvalues	2
2.3	Matrix and Spectra	3
2.4	Stochastic Trace Estimator	3
2.5	Estimation of bilinear form $\mathbf{z}^T f(\mathbf{M}) \mathbf{z}$	5
2.6	Stochastic Lanczos Quadrature(SLQ)	6
3	the Sum of Powers of Eigenvalues of Network (SPENet)	7
3.1	Algorithms	7
3.1.1	STE algorithm	7
3.1.2	SLQ algorithm	7
3.2	Conditions of k	8
3.3	Error bounds	9
3.4	Experiments	9
3.4.1	Parameter Settings	9
4	Conclusion	10
	References	11

Chapter 1

Introduction

1.1 Complex Network

1.2 von Neumann entropy of Network

The Kullback-Leibler Divergence [13] is the most popular divergence. However, the KLD is asymmetric. Several symmetrizations [17] were proposed including the Jeffreys divergence [12] and the Jensen-Shannon Divergence [15].

The square root of JS Divergence is called Jensen-Shannon distance which has been proved to be a valid distance metric [8].

Von Neumann entropy, Quantum Relative entropy, Quantum Jensen-Shannon divergence [5, 14]

Inspired by quantum information theory, the concept of von Neumann graph entropy (VNGE) was introduced [4].

Some approximate algorithms have been developed to compute Von Neumann Graph Entropy [6, 20].

Other network distance metric, NetLSD, was developed [19, 20].

Motif multiplicity corresponds to eigenvalues multiplicity [16, 7].

There are fast online search tools for DOS [3, 9]

Chapter 2

Preliminaries

2.1 General Notation

Let $G = (V, E, w)$ be a wighted undirected graph where $V = (v_0, v_2, \dots, v_{n-1})$ is a set of vertices, $E \subseteq V \times V$ is a set of undirected edges, and $w : E \rightarrow \mathbb{R}$ is a weight function. The adjacency matrix \mathbf{A} is a $n \times n$ matrix of G , (i.e., $A_{ij} = w(v_i, u_i)$).

The laplacian matrix \mathbf{L} and the normalized laplacian matrix \mathcal{L} are defined by the weight function w and the degree $d_v := \sum_u w(u, v)$.

$$L(u, v) = \begin{cases} d_v - w(u, v) & \text{if } u = v, \\ -w(u, v) & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathcal{L}(u, v) = \begin{cases} 1 - \frac{w(u, v)}{d_v} & \text{if } u = v, \\ -\frac{w(u, v)}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

Let \mathbf{M} denote a real symmetric matrix that can be \mathbf{A} , \mathbf{L} , or \mathcal{L} .

Symbol	Description
$n \in \mathbb{N}$	$ V $:the number of vertices
$m \in \mathbb{N}$	$ E $:the number of edges
$\mathbf{M} \in \mathbb{R}^{n \times n}$	Real symmetric matrix which can be \mathbf{A} , \mathbf{L} , or \mathcal{L}
$\mathbf{A} \in \mathbb{R}^{n \times n}$	Adjacency matrix of a graph G
$\mathbf{L} \in \mathbb{R}^{n \times n}$	Laplacian matrix of a graph G
$\mathcal{L} \in \mathbb{R}^{n \times n}$	Normalized Laplacian matrix of a graph G
$n_v \in \mathbb{R}^+$	Number of SLQ random vectors
$s \in \mathbb{R}^+$	Number of Lanczos algorithm steps
$\mathbf{x} \in \mathbb{R}^n$	Random vector which satisfies $\mathbb{E}_{p(\mathbf{x})}[\mathbf{x}\mathbf{x}^T] = \mathbf{I}$

Table 2.1: Summary of notion

2.2 The Laplacian and Eigenvalues

Let \mathbf{D} denote the diagonal matrix, such that

$$\mathbf{D}(u, v) = \begin{cases} d_v & \text{if } u = v, \\ 0 & \text{otherwise.} \end{cases}$$

There are some relationships among the adjacency matrix, laplacian matrix, and normalized laplacian matrix.

By definitions, we can write

$$\begin{aligned}\mathcal{L} &= \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}\end{aligned}$$

2.3 Matrix and Spectra

Firstly, we define some concepts regarding matrix and its eigenvalues and prove some theorems which are used later.

Definition 1. Let $\lambda_0 \leq \lambda_2 \leq \dots \leq \lambda_{n-1}$ be the eigenvalues of a n -dimensional matrix \mathbf{M} . The spectrum $\sigma(\mathbf{M})$ of \mathbf{M} is the set of its overall eigenvalues.

$$\sigma(\mathbf{M}) := \{\lambda_i | i = 0, 1, \dots, n-1\}$$

Definition 2. The spectral radius $\rho(\mathbf{M})$ of a n -dimensional matrix \mathbf{M} is the nonnegative number

$$\rho(\mathbf{M}) := \sup_{\lambda \in \sigma(A)} |\lambda|.$$

Theorem 1. Let $\lambda_0 \leq \lambda_2 \leq \dots \leq \lambda_{n-1}$ be the eigenvalues of a real symmetric matrix \mathbf{M} and the eigendecomposition of \mathbf{M} be $\mathbf{M} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$. Then $\lambda_0^k, \lambda_2^k, \dots, \lambda_{n-1}^k$ are eigenvalues of \mathbf{M}^k .

Proof.

$$\begin{aligned}\mathbf{M}^k &= (\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T)^k \\ &= (\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1})^k && \mathbf{Q} \text{ is a orthogonal matrix} \\ &= \mathbf{Q} \mathbf{\Lambda}^k \mathbf{Q}^{-1} \\ &= \mathbf{Q} \mathbf{\Lambda}^k \mathbf{Q}^T\end{aligned}$$

□

Theorem 2. Let the eigendecomposition of a real symmetric matrix \mathbf{M} be $\mathbf{M} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$. If $f(x)$ is given by $f(x) = \sum_{k=0}^{\infty} a_k x^k$ with radius of convergence greater than the spectral radius $\rho(\mathbf{M})$ and $f(\mathbf{M})$ is defined by $f(\mathbf{M}) = \sum_{k=0}^{\infty} a_k \mathbf{M}^k$, then

$$f(\mathbf{M}) = \mathbf{Q} f(\mathbf{\Lambda}) \mathbf{Q}^T.$$

Proof. $\mathbf{\Lambda}$ is a diagonal matrix. Therefore $f(\mathbf{\Lambda})$ is easy to calculate:

$$[f(\mathbf{\Lambda})]_{ij} = \begin{cases} f(\lambda_i) & (i = j) \\ 0 & (\text{otherwise}) \end{cases}$$

By applying theorem 1 to each degree, we can get $f(\mathbf{M}) = \mathbf{Q} f(\mathbf{\Lambda}) \mathbf{Q}^T$. □

2.4 Stochastic Trace Estimator

For trace estimation of large implicit matrices, we can use randomized estimator described by Hunchinson [11, 1].

Theorem 3. (Proposition 4.1 [1]) Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a square matrix and $\mathbf{x} \in \mathbb{R}^n$ be a random vector such that $\mathbb{E}_{p(\mathbf{x})}[\mathbf{x} \mathbf{x}^T] = \mathbf{I}$. Then $\mathbb{E}_{p(\mathbf{x})}[\mathbf{x}^T \mathbf{M} \mathbf{x}] = \text{tr}(\mathbf{M})$.

Proof.

$$\begin{aligned}
\mathbb{E}_{p(\mathbf{x})}[\mathbf{x}^T \mathbf{M} \mathbf{x}] &= \mathbb{E}_{p(\mathbf{x})}[\text{tr}(\mathbf{x}^T \mathbf{M} \mathbf{x})] \\
&= \mathbb{E}_{p(\mathbf{x})}[\text{tr}(\mathbf{M} \mathbf{x} \mathbf{x}^T)] && \text{invariance to cyclic permutation} \\
&= \text{tr}(\mathbb{E}_{p(\mathbf{x})}[\mathbf{M} \mathbf{x} \mathbf{x}^T]) && \text{linearity of trace} \\
&= \text{tr}(\mathbf{M} \mathbb{E}_{p(\mathbf{x})}[\mathbf{x} \mathbf{x}^T]) && \text{linearity of expectation} \\
&= \text{tr}(\mathbf{M})
\end{aligned}$$

□

The requirement $\mathbb{E}_{p(\mathbf{x})}[\mathbf{x} \mathbf{x}^T] = \mathbf{I}$ for the random vector \mathbf{x} can be easily satisfied. For example, paractical choices of $p(\mathbf{x})$ include Ramemacher or standard normal distributions with zero mean and one variance.

The following corollary holds.

Corollary 3.1. *If $f(\mathbf{M}) \in \mathbb{R}^{n \times n}$ is a square matrix, and $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_v} \in \mathbb{R}^n$ are random vectors drawn from a distribution $p(\mathbf{x})$ such that $\mathbb{E}_{p(\mathbf{x})}[\mathbf{x} \mathbf{x}^T] = \mathbf{I}$, then*

$$\text{tr}(f(\mathbf{M})) = \mathbb{E}_{p(\mathbf{x})}[\mathbf{x}^T f(\mathbf{M}) \mathbf{x}] \approx \frac{1}{n_v} \sum_{i=0}^{n_v-1} \mathbf{x}_i^T f(\mathbf{M}) \mathbf{x}_i$$

Unit vectors $\mathbf{z}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}$ can be used instead of \mathbf{x}_i . We can rewrite the formula of Corollary 3.1.

Theorem 4. *Let $\mathbf{z}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}$ be a unit vector, then*

$$\text{tr}(f(\mathbf{M})) \approx \frac{n}{n_v} \sum_{i=0}^{n_v-1} \mathbf{z}_i^T f(\mathbf{M}) \mathbf{z}_i$$

Proof.

$$\begin{aligned}
\text{tr}(f(\mathbf{M})) &\approx \frac{1}{n_v} \sum_{i=0}^{n_v-1} \mathbf{x}_i^T f(\mathbf{M}) \mathbf{x}_i \\
&= \frac{1}{n_v} \sum_{i=0}^{n_v-1} \|\mathbf{x}_i\|_2^2 \cdot \mathbf{z}_i^T f(\mathbf{M}) \mathbf{z}_i \\
&\approx \frac{1}{n_v} \sum_{i=0}^{n_v-1} n \cdot \mathbf{z}_i^T f(\mathbf{M}) \mathbf{z}_i && \mathbb{E}_{p(\mathbf{x})}[\|\mathbf{x}_i\|_2^2] \text{ is equal to } n \\
&= \frac{n}{n_v} \sum_{i=0}^{n_v-1} \mathbf{z}_i^T f(\mathbf{M}) \mathbf{z}_i
\end{aligned}$$

□

For Randemacher vectors, $\|\mathbf{x}_i\|_2^2$ is always n . For other random vectors, $\mathbb{E}_{p(\mathbf{x})}[\|\mathbf{x}_i\|_2^2]$ is equal to n as long as $\mathbb{E}_{p(\mathbf{x})}[\mathbf{x} \mathbf{x}^T] = \mathbf{I}$.

For accurate computation, n_v should approaches infinity. However using large n_v is expensive. For practical purposes, we can use small n_v . The convergence rate of such estimator was studied by Avron and Toledo[2].

2.5 Estimation of bilinear form $\mathbf{z}^T f(\mathbf{M})\mathbf{z}$

There are relationships between matrix, orthogonal polynomials, quadrature rules and the Lanczos algorithms. Gene H Golub and Gérard Meurant compiled a book of previous research on such relationships and on estimation and bounds of bilinear form $\mathbf{u}^T f(\mathbf{M})\mathbf{v}$ [10]. In this paper, we focus on a special case that $\mathbf{u} = \mathbf{v} = \mathbf{z}$.

Firstly, the bilinear form $\mathbf{z}^T f(\mathbf{M})\mathbf{z}$ is reformed as a Riemann-Stieltjes integral and approximated by applying Gauss quadrature rule. Then, the Lanczos algorithm is applied for calculation.

The Riemann–Stieltjes integral is a generalization of the Riemann integral. It is defined to be the limit of a sum.

Definition 3. A Riemann–Stieltjes integral of a real valued function f of a real variable on the infinite interval $[a, b]$ with respect to a real function α is denoted by

$$\int_a^b f(x) d\alpha(x).$$

This integral is defined to be the limit, as the length of the subinterval of the partition $\pi = \{a = x_0 < x_1 < \dots < x_n = b\}$ goes to zero, of the approximating sum

$$\sum_{x_i \in \pi} f(c_i)(\alpha(x_{i+1}) - \alpha(x_i)).$$

where c_i is in the i -th subinterval $[x_i, x_{i+1}]$.

Now, we consider the approximation of this integral by Gauss quadrature rule. This is a relation:

$$I[f] = \int_a^b f(x) d\alpha(x) = \sum_{k=0}^{s-1} w_k f(\theta_k) + R[f] \quad (2.1)$$

where $\{w_k\}$ are the unknowns weights and $\{\theta_k\}$ are the unknowns nodes. In the right-hand side, the sum $\sum_{k=0}^{s-1} w_k f(\theta_k)$ is the approximation of the Riemann–Stieltjes integral $I[f]$, and $R[f]$ is the remainder. If $R[p] = 0$ for all polynomials p of degree d and $R[q] \neq 0$ for any polynomials q of degree $d+1$, the rule is called *exact*.

The value of $R[f]$ is known [10, 18]. If the measure $\alpha(x)$ is a positive nondecreasing function and $f \in C^{2s}[a, b]$, then

$$R[f] = \frac{f^{2s}(\xi)}{(2s)!} \int_a^b \left[\prod_{k=0}^{s-1} (x - \theta_k) \right]^2 d\alpha(x) \quad (2.2)$$

for some $\xi \in (a, b)$. Therefore, the Gauss rule is exact for polynomials f of degree less than $2s$.

Finally, we can reform and compute the bilinear form $\mathbf{z}^T f(\mathbf{M})\mathbf{z}$ using the

Gauss quadrature rule.

$$\begin{aligned}
\mathbf{z}^T f(\mathbf{M}) \mathbf{z} &= \mathbf{z}^T \mathbf{Q} f(\mathbf{\Lambda}) \mathbf{Q}^T \mathbf{z} \\
&= \sum_{j=0}^{n-1} f(\lambda_j) \mu_j^2 \\
&= \int_a^b f(t) d\mu(t) && \text{Riemann-Stieltjes integral} \\
&\approx \sum_{k=0}^{s-1} \omega_k f(\theta_k) && \text{the Gauss quadrature rule}
\end{aligned}$$

where $\mu_j = [\mathbf{Q}^T \mathbf{z}]_j$ and $\mu(t)$ is a piecewise constant function defined as

$$\mu(t) = \begin{cases} 0 & \text{if } t < a = \lambda_0 \\ \sum_{j=0}^{i-1} \mu_j^2 & \text{if } \lambda_{i-1} \leq t < \lambda_i, i = 1, \dots, n-1 \\ \sum_{j=0}^{n-1} \mu_j^2 & \text{if } b = \lambda_{n-1} \leq t \end{cases}$$

and $\{\omega_k\}$ are the weights and $\{\theta_k\}$ are the nodes of the s-point Gauss quadrature.

One choice for computing the nodes and the weights of the Gauss quadrature rule is the Lanczos algorithm. Let \mathbf{M} be a real symmetric matrix, \mathbf{w}_0 be an arbitrary starting unit-vector, and \mathcal{K} be the Krylov subspace spanning vectors $\{\mathbf{w}_0, \mathbf{M}\mathbf{w}_0, \dots, \mathbf{M}^{s-1}\mathbf{w}_0\}$. The output of the Lanczos algorithm are an $n \times s$ matrix \mathbf{W} and an $s \times s$ tridiagonal matrix \mathbf{T} , such that $\mathbf{W}^T \mathbf{M} \mathbf{W} = \mathbf{T}$. \mathbf{W} has the orthonormal columns:

$$\begin{aligned}
w_0 &: \text{an initial unit-vector} \\
w_k &= p_k(\mathbf{M})w_0, \quad k = 1, \dots, s-1
\end{aligned}$$

where p_k are the Lanczos polynomials which are orthogonal with respect to the measure $\mu(t)$.

Now, the bilinear form $\mathbf{z}^T f(\mathbf{M}) \mathbf{z}$ is reformed as

$$\mathbf{z}^T f(\mathbf{M}) \mathbf{z} \approx \sum_{k=0}^{s-1} \tau_k^2 f(\theta_k) \tag{2.3}$$

$$\tau_k = \mathbf{Y}_{0,k} = \mathbf{e}_1^T \mathbf{y}_k, \quad \theta_k = \Theta_{k,k}, \quad \mathbf{T} = \mathbf{Y} \mathbf{\Theta} \mathbf{Y}^T \tag{2.4}$$

2.6 Stochastic Lanczos Quadrature(SLQ)

There is a method called stochastic Lanczos quadrature(SLQ) [21], for approximate computing of the trace of functions of large matrices. Firstly, the stochastic trace estimator is used for approximating the trace. Next, the estimation method we discussed in the section 2.5 is applied.

The stochastic trace estimator has been explained in the previous section.

$$\text{tr}(f(\mathbf{M})) = \mathbb{E}_{p(\mathbf{x})}[\mathbf{x}^T f(\mathbf{M}) \mathbf{x}] \approx \frac{n}{n_v} \sum_{i=0}^{n_v-1} \mathbf{z}_i^T f(\mathbf{M}) \mathbf{z}_i \tag{2.5}$$

Now, we need to compute $\mathbf{z}_i^T f(\mathbf{M}) \mathbf{z}_i$. By applying (2.1) to the stochastic trace estimator (2.3), we get Stochastic Lanczos Quadrature estimator:

$$\text{tr}(f(\mathbf{M})) \approx \frac{n}{n_v} \sum_{i=0}^{n_v-1} \left(\sum_{k=0}^{s-1} (\tau_k^i)^2 f(\theta_k^i) \right) \tag{2.6}$$

Chapter 3

the Sum of Powers of Eigenvalues of Network (SPENet)

Our main purpose is computing sum of k -th powers of eigenvalues $\sum_{i=0}^{n-1} \lambda_i^k$ which we call SPENet. We propose two approximation algorithms (STE and SLQ) based on the stochastic trace estimator and the stochastic Lanczos quadrature. For naive algorithms, the computational complexity of SPENet using full eigen-decomposition is $O(n^3)$, while for STE algorithm it is $O(mkn_v)$ and for SLQ algorithm it is $O((ms + ns^2)n_v)$.

Since the laplacian matrix \mathbf{L} and the normalized laplacian matrix \mathcal{L} have zero eigenvalues, we are interested in the case that index of power k is positive. When we consider the case that $\mathbf{M} = \mathbf{A}$, k must be positive integer or zero because \mathbf{A} have negative eigenvalues.

3.1 Algorithms

3.1.1 STE algorithm

The first estimation method for SPENet is based on the stochastic trace estimator(STE), which we apply in our setting for the trace of power of matrix $f(\mathbf{M}) = \mathbf{M}^k$. We are interested in the case where $\mathbf{M} = \mathbf{A}, \mathbf{L}, \text{ or } \mathcal{L}$. In this algorithm, the index of power k must be positive integer.

By using theorem 1 and 4, we can apply the stochastic trace estimator for computing sum of powers of k -th eigenvalues.

$$\begin{aligned} \sum_{i=0}^{n-1} \lambda_i^k &= \text{tr}(\mathbf{M}^k) \\ &= \mathbb{E}_{p(\mathbf{x})}[\mathbf{x}^T \mathbf{M}^k \mathbf{x}] \approx \frac{n}{n_v} \sum_{i=0}^{n_v-1} \mathbf{z}_i^T \mathbf{M}^k \mathbf{z}_i \end{aligned} \quad (3.1)$$

The product of two matrices costs $O(n^3)$, but the product of a vector and a sparse matrix costs only $O(m)$. Therefore, when calculating $\mathbf{z}_i^T \mathbf{M}^k \mathbf{z}_i$, we just calculate the product of a vector and a matrix k times, and the inner product of a vector once from the left. From the above, the cost of the STE algorithm will be $O(mkn_v)$.

3.1.2 SLQ algorithm

The second estimation method for SPENet is based on the stochastic Lanczos quadrature(SLQ), which we apply in our setting for the trace of power of matrix

$f(\mathbf{M}) = \mathbf{M}^k$. We are interested in the case where $\mathbf{M} = \mathbf{A}, \mathbf{L}, \text{ or } \mathcal{L}$. In this algorithm, the index of power k must be positive real number.

By using theorem 1 and (2.4), we can estimate SPENet.

$$\begin{aligned}
\sum_{i=0}^{n-1} \lambda_i^k &= \text{tr}(\mathbf{M}^k) \\
&= \mathbb{E}_{p(\mathbf{x})}[\mathbf{x}^T \mathbf{M}^k \mathbf{x}] \approx \frac{n}{n_v} \sum_{i=0}^{n_v-1} \mathbf{z}_i^T \mathbf{M}^k \mathbf{z}_i \\
&\approx \frac{n}{n_v} \sum_{i=0}^{n_v-1} \left(\sum_{j=0}^{s-1} (\tau_j)^2 (\theta_j)^k \right)
\end{aligned} \tag{3.2}$$

Since we apply s-step Lanczos algorithm for τ_k^i and θ_k^i , the orthogonalization cost inside the Lanczos algorithm is $O(ns^2)$ and the cost of the product of a vector and a matrix s times in the Lanczos algorithm is $O(ms)$. From the above, all the computational cost of SLQ algorithm is $O((ms + ns^2)n_v)$.

Algorithm 1 STE algorithm

Input: a real symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, a positive integer k , and n_v .

Output: Approximate sum of k-th powers of eigenvalues $\sum_{i=0}^{n-1} \lambda_i^k$

```

1: for  $i = 0$  to  $n_v - 1$  do
2:   Generate a random vector  $\mathbf{x}_i$  with the standard normal distribution.
3:   Form unit vector  $\mathbf{z}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}$ .
4:    $\mathbf{v}_i^T \leftarrow \mathbf{z}_i^T$ .
5:   for  $j = 0$  to  $k - 1$  do
6:      $\mathbf{v}_i^T \leftarrow \mathbf{v}_i^T \mathbf{M}$ .
7:   end for
8:    $\Gamma \leftarrow \Gamma + \mathbf{v}_i^T \mathbf{z}_i$ 
9: end for
10:  $\Gamma \leftarrow \frac{n}{n_v} \Gamma$ 
11: return  $\Gamma$ 

```

Algorithm 2 SLQ algorithm

Input: a real symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, a positive real number k , and n_v .

Output: Approximate sum of k-th powers of eigenvalues $\sum_{i=0}^{n-1} \lambda_i^k$

```

1: for  $i = 0$  to  $n_v - 1$  do
2:   Generate a random vector  $\mathbf{x}_i$  with the standard normal distribution.
3:   Form unit vector  $\mathbf{z}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}$ .
4:    $T \leftarrow \text{Lanczos}(\mathbf{M}, \mathbf{z}_i, s)$ .
5:    $[\mathbf{Y}, \boldsymbol{\Theta}] \leftarrow \text{eig}(T)$ 
6:   Compute  $\tau_j = [e_1^T y_j]$  for  $j = 0, \dots, s - 1$ 
7:    $\Gamma \leftarrow \Gamma + \sum_{j=0}^{s-1} (\tau_j)^2 (\theta_j)^k$ 
8: end for
9:  $\Gamma \leftarrow \frac{n}{n_v} \Gamma$ 
10: return  $\Gamma$ 

```

3.2 Conditions of k

In SLQ algorithm, we can generalize the index of power k . If \mathbf{M} is positive definite, we can use not only a positive integer but also a positive real number as

k . This is because that $f(x) = x^k$ is analytic inside a closed interval $[\lambda_0, \lambda_{n-1}]$. However the function $f(x) = x^k$ have a singularity at zero and the laplacian matrix \mathbf{L} and the normalized laplacian matrix \mathcal{L} are positive semi-definite matrix, which have $\lambda_0 = 0$.

To overcome the issue mentioned above, we can use *shifting the spectrum* proposed by Shashanka Ubaru, Jie Chen, and Yousef Saad[21]. The idea is shifting the eigenvalues by replacing \mathbf{M} with $\mathbf{M} + \delta\mathbf{I}$. The shifted matrix have the eigenvalues in the interval $[\lambda_0 + \delta, \lambda_{n-1} + \delta]$. Hence we can obtaining the approximation error bounds of SLQ algorithm and find that SLQ algorithm is practically useful.

If k is a positive real number and $\mathbf{M} = \mathbf{A}$, the SPENet $\sum_{i=0}^{n-1} \lambda_i^k$ is undefined because \mathbf{A} have negative eigenvalues.

	\mathbf{A}	\mathbf{L}	\mathcal{L}
$k \in \mathbb{N}_{\geq 0}$	STE or SLQ	STE or SLQ	STE or SLQ
$k \in \mathbb{R}_{\geq 0}$	undefined	SLQ	SLQ

Table 3.1: Conditions under which the algorithms can be used

3.3 Error bounds

Suppose k is a positive integer and Lanczos step size is s . When $2s - 1$ is greater or equal to the degree of $f(x)$, the error of Gauss quadrature rule is zero.

3.4 Experiments

We evaluate computational time and variance of SPENet using STE and SLQ algorithm. Additionally, we evaluate its relative error against the exact computation of the eigenvalues where allowed by the graph size. We perform our experiments averating 10 times for all experiments unless stated otherwise.

3.4.1 Parameter Settings

Chapter 4

Conclusion

References

- [1] Ryan P Adams, Jeffrey Pennington, Matthew J Johnson, Jamie Smith, Yaniv Ovadia, Brian Patton, and James Saunderson. Estimating the spectral density of large implicit matrices. *arXiv preprint arXiv:1802.03451*, 2018.
- [2] Haim Avron and Sivan Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *Journal of the ACM (JACM)*, 58(2):1–34, 2011.
- [3] Stanislav S Borysov, Bart Olsthoorn, M Berk Gedik, R Matthias Geilhufe, and Alexander V Balatsky. Online search tool for graphical patterns in electronic band structures. *NPJ Computational Materials*, 4(1):1–8, 2018.
- [4] Samuel L Braunstein, Sibasish Ghosh, and Simone Severini. The laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics*, 10(3):291–317, 2006.
- [5] Jop Briët and Peter Harremoës. Properties of classical and quantum jensen-shannon divergence. *Physical review A*, 79(5):052311, 2009.
- [6] Pin-Yu Chen, Lingfei Wu, Sijia Liu, and Indika Rajapakse. Fast incremental von neumann graph entropy computation: Theory, algorithm, and applications. In *International Conference on Machine Learning*, pages 1091–1101, 2019.
- [7] Kun Dong, Austin R Benson, and David Bindel. Network density of states. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1152–1161, 2019.
- [8] Dominik Maria Endres and Johannes E Schindelin. A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7):1858–1860, 2003.
- [9] R Matthias Geilhufe, Stanislav S Borysov, Dmytro Kalpakchi, and Alexander V Balatsky. Towards novel organic high- T_c superconductors: Data mining using density of states similarity search. *Physical Review Materials*, 2(2):024802, 2018.
- [10] Gene H Golub and Gérard Meurant. *Matrices, moments and quadrature with applications*, volume 30. Princeton University Press, 2009.
- [11] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.
- [12] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

- [13] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- [14] PW Lamberti, AP Majtey, A Borrás, Montserrat Casas, and A Plastino. Metric character of the quantum jensen-shannon divergence. *Physical Review A*, 77(5):052311, 2008.
- [15] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.
- [16] Ranjit Mehatari and Anirban Banerjee. Effect on normalized graph laplacian spectrum by motif attachment and duplication. *Applied Mathematics and Computation*, 261:382–387, 2015.
- [17] Frank Nielsen. On the jensen–shannon symmetrization of distances relying on abstract means. *Entropy*, 21(5):485, 2019.
- [18] Josef Stoer and Roland Bulirsch. *Introduction to numerical analysis*, volume 12. Springer Science & Business Media, 2013.
- [19] Anton Tsitsulin, Davide Mottin, Panagiotis Karras, Alexander Bronstein, and Emmanuel Müller. Netlsd: hearing the shape of a graph. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2347–2356, 2018.
- [20] Anton Tsitsulin, Marina Munkhoeva, and Bryan Perozzi. Just slaq when you approximate: Accurate spectral distances for web-scale graphs. In *Proceedings of The Web Conference 2020*, pages 2697–2703, 2020.
- [21] Shashanka Ubaru, Jie Chen, and Yousef Saad. Fast estimation of $\text{tr}(\mathbf{f}(\mathbf{A}))$ via stochastic lanczos quadrature. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1075–1099, 2017.