

Analysis and quantitative evaluation of large networks using
spectral distribution

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

In analyzing the properties and structure of a network, the spectrum can be a very important tool. However, since computing the full eigendecomposition is expensive, the analysis of large networks have been done by using only a part of the full spectrum. Recently, a method for calculating the entire spectrum distribution has been developed, but its application has not yet been deeply considered, and it has been limited to qualitative evaluation by visualizing the spectrum distribution.

In this study, we will analyze the distribution of the overall spectrum of the theoretically created graphs and the real-world network. Then, we will discuss the relation between the properties of the network and the spectrum. As a result, we find that graphs in which vertices are adjacent to each other with similar properties have sharp spectra found in complex networks. Furthermore, we illustrate that graphs with randomness in vertex adjacencies will have the semicircular spectra found in random graphs. We also suggest applying the concepts of entropy and divergence of discrete probability distributions to quantitative evaluation and comparison of spectral distributions. A fast way to search for similar spectra using cosine distance is also discussed.

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。近年スペクトルの分布全体を計算する手法が考案されたが、応用方法などはまだ深く考えられておらず、スペクトル分布を視覚化して定性的な評価をするにとどまっている。

本研究では、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、性質が近い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こと分かった。さらに、頂点同士の隣接にランダム性があるグラフは、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。また、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エントロピーやダイバージェンス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。cosine distance を用いて類似のスペクトルを高速に検索する方法についても言及する。

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
3	Conclusion	3
	References	4

Chapter 1

Introduction

1.1 Complex Network

1.2 von Neumann entropy of Network

The Kullback-Leibler Divergence [9] is the most popular divergence. However, the KLD is asymmetric. Several symmetrizations [13] were proposed including the Jeffreys divergence [8] and the Jensen-Shannon Divergence [11].

The square root of JS Divergence is called Jensen-Shannon distance which has been proved to be a valid distance metric [6].

Von Neumann entropy, Quantum Relative entropy, Quantum Jensen-Shannon divergence [3, 10]

Inspired by quantum information theory, the concept of von Neumann graph entropy (VNGE) was introduced [2].

Some approximate algorithms have been developed to compute Von Neumann Graph Entropy [4, 15].

Other network distance metric, NetLSD, was developed [14, 15].

Motif multiplicity corresponds to eigenvalues multiplicity [12, 5].

There are fast online search tools for DOS [1, 7]

Chapter 2

Preliminaries

2.1 General Notation

Chapter 3

Conclusion

References

- [1] 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.
- [2] 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.
- [3] Jop Briët and Peter Harremoës. Properties of classical and quantum jensen-shannon divergence. *Physical review A*, 79(5):052311, 2009.
- [4] 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.
- [5] 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.
- [6] Dominik Maria Endres and Johannes E Schindelin. A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7):1858–1860, 2003.
- [7] 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.
- [8] 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.
- [9] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- [10] 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.
- [11] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.
- [12] Ranjit Mehatari and Anirban Banerjee. Effect on normalized graph laplacian spectrum by motif attachment and duplication. *Applied Mathematics and Computation*, 261:382–387, 2015.

- [13] Frank Nielsen. On the jensen–shannon symmetrization of distances relying on abstract means. *Entropy*, 21(5):485, 2019.
- [14] 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.
- [15] 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.