Entropic Characterization and Time Evolution of Complex Networks

Cheng Ye

Doctor of Philosophy

University of York

Computer Science

April, 2016

Abstract

In this thesis, we address problems encountered in complex network analysis using graph theoretic methods. The thesis specifically centers on the challenge of how to characterize the structural properties and time evolution of graphs. We commence by providing a brief roadmap for our research in Chapter 1, followed by a review of the relevant research literature in Chapter 2. The remainder of the thesis is structured as follows.

In Chapter 3, we focus on the graph entropic characterizations and explore whether the von Neumann entropy recently defined only on undirected graphs, can be extended to the domain of directed graphs. The substantial contribution involves a simplified form of the entropy which can be expressed in terms of simple graph statistics, such as graph size and vertex in-degree and out-degree. Chapter 4 further investigates the uses and applications of the von Neumann entropy in order to solve a number of network analysis and machine learning problems. The contribution in this chapter includes an entropic edge assortativity measure and an entropic graph embedding method, which are developed for both undirected and directed graphs.

The next part of the thesis analyzes the time-evolving complex networks using physical and information theoretic approaches. In particular, Chapter 5 provides a thermodynamic framework for handling dynamic graphs using ideas from algebraic graph theory and statistical mechanics. This allows us to derive expressions for a number of thermodynamic functions, including energy, entropy and temperature, which are shown to be efficient in identifying abrupt structural changes and phase transitions in real-world dynamical systems. Chapter 6 develops a novel method for constructing a generative model to analyze the structure of labeled data, which provides a number of novel directions to the study of graph time-series. Finally, in Chapter 7, we provide concluding remarks and discuss the limitations of our methodologies, and point out possible future research directions.

Contents

List of Figures

List of Tables

Acknowledgements

My deepest and most heartfelt gratitude goes first and foremost to my supervisor Professor Edwin Hancock. It has been a great honour and joy to work with him over the past four years. His continuous guidance, immense knowledge and great enthusiasm and motivation in science have helped me overcome enormous difficulties encountered during my Ph.D. studies and research. I will always remain deeply indebted to him for his patient proofreading and editing, regular meetings and prompt email replies that helped me complete this thesis in timely fashion.

I would also like to express my special thanks to my assessor Professor Richard Wilson for his rigorous assessment of my work. I could not have successfully published research work without the advice received from him. My sincere gratitude also goes to all my colleagues at the Department of Computer Science. I would like to thank our collaborators in Brazil, Luciano da Costa, César Comin, Thomas Peron, Filipi Silva, etc. for sharing the experimental data with us, and also for providing a warm hospitality when I was visiting their research institute in São Carlos.

I am grateful to all my friends, my previous housemates and flatmates in the UK, for their great help and support that made me have a wonderful and enjoyable time during my stay at York.

Lastly and most importantly, I would like to express my sincere gratitude to all my family members for their enormous support in my study and my life. My special thanks go to my father and my mother Xiaoping Ye, for their remarkable efforts in my upbringing and education. They have always taught me to be an honest and humble person, which I believe will be the key to helping me develop further as a researcher and as a human being. I am also grateful to my girlfriend, Changqiong Wang, who has been supporting me unconditionally during these years. We have frequently discussed problems, exchanged opinions and shared experiences together, and this thesis could not have been completed without her great help.

Declaration

I declare that the research described in this thesis is original work, which I undertook at the University of York during 2012 - 2016. Except where stated, all of the work contained within this thesis represents the original contribution of the author.

Some parts of this thesis have been published in conference proceedings and journals; where items were published jointly with collaborators, the author of this thesis is responsible for the material presented here.

Journal Papers

- Cheng Ye, César H. Comin, Thomas K. DM. Peron, Filipi N. Silva, Francisco A. Rodrigues, Luciano da F. Costa, Andrea Torsello, and Edwin R. Hancock: Thermodynamic characterization of networks using graph polynomials. In *Physical Review E*, Volume 92, Issue 3, Number 032810, 2015.

- Cheng Ye, Richard C. Wilson, César H. Comin, Luciano da F. Costa, and Edwin R. Hancock: Approximate von Neumann entropy for directed graphs. In *Physical Review E*, Volume 89, Issue 5, Number 052804, 2014.

- Filipi N. Silva, César H. Comin, Thomas K. DM. Peron, Francisco A. Rodrigues, Cheng Ye, Richard C. Wilson, Edwin Hancock, Luciano da F. Costa: Concentric network symmetry. In *Information Sciences*, Volume 333, Pages 61-80, 2016.

- Filipi N. Silva, César H. Comin, Thomas K. DM. Peron, Francisco A. Rodrigues, Cheng Ye, Richard C. Wilson, Edwin Hancock, Luciano da F. Costa: On the Modular Dynamics of Financial Market Networks. In *ArXiv e-prints*, Number 1501.05040, 2015.

Conference Papers

- Cheng Ye, Richard C. Wilson, Edwin R. Hancock: Thermodynamics of Dynamic Complex Networks. In Presentations of *International School and Conference on Network Science (NetSci)*, 2015, Zaragoza, Spain.

- Cheng Ye, Richard C. Wilson, Edwin R. Hancock: Thermodynamics of Time Evolving Networks. In Proceedings of *10th IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition (GbR)*, 2015, Beijing, China. Pages 315-324.

- Cheng Ye, Richard C. Wilson, Edwin R. Hancock: An Entropic Edge Assortativity Measure. In Proceedings of *10th IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition (GbR)*, 2015, Beijing, China. Pages 23-33.

- Cheng Ye, Richard C. Wilson, César H. Comin, Luciano da F. Costa, Edwin R. Hancock: Approximating the Von Neumann Entropy for Directed Graphs. In Presentations of *European Conference on Complex Systems (ECCS)*, 2014, Lucca, Italy.

- Cheng Ye, Richard C. Wilson, Edwin R. Hancock: Graph Characterization from Entropy Component Analysis. In Proceedings of *22nd International Conference on Pattern Recognition (ICPR)*, 2014, Stockholm, Sweden. Pages 3845-3850.

- Cheng Ye, Richard C. Wilson, Edwin R. Hancock: Entropic Graph Embedding via Multivariate Degree Distributions. In Proceedings of *IAPR joint International Workshop in Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, 2014, Joensuu, Finland. Pages 163-172.

- Cheng Ye, Richard C. Wilson, Edwin R. Hancock: Heterogeneity Index for Directed Graphs. In Proceedings of *15th International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2013, York, UK. Pages 424-431.

- Cheng Ye, Richard C. Wilson, César H. Comin, Luciano da F. Costa, Edwin R. Hancock: Entropy and Heterogeneity Measures for Directed Graphs. In Proceedings of *2nd International Workshop on Similarity-Based Pattern Analysis and Recognition (SIMBAD)*, 2013, York, UK. Pages 219-234.

Chapter 1

Introduction

In this chapter, we provide a roadmap for the research presented in this thesis. Specifically, we give the motivation of our research, i.e., the reason why we are interested in the study of the characterization and time evolution of complex networks. To commence, we address the problems encountered in the network science literature and show what the state-of-the-art solutions for these problems are. Then, we present our research goals of the thesis, which are to develop novel methods for characterizing the structural complexity and time evolution of complex networks. We provide a brief overview of how we will accomplish these research goals, i.e., the novel contributions we will make in the remainder of the thesis. Finally, this chapter concludes with an outline of the rest of the thesis.

1.1   Problems in Network Science

The study of networks has been one of the fundamental branches of discrete mathematics. In 1736, Leonhard Euler has published the solution to a historically notable mathematics problem, namely the Seven Bridges of Königsberg [94]. This work is often cited as the foundation of graph theory, which has continued to develop into a substantial body of knowledge in the following centuries. Recently, however, there has been a new movement in network research, with a considerable interest in studying the statistical properties of large-scale complex networks. This is due to the fact that they play a crucial role in revealing essential features of the structure, function and dynamics of many large-scale systems in biology, physics and the social sciences [3] [28] [34]. In fact, complex networks provide convenient models for complex systems. Specifically, a complex network is a diagrammatic representation of a complex system. It consists of vertices, which
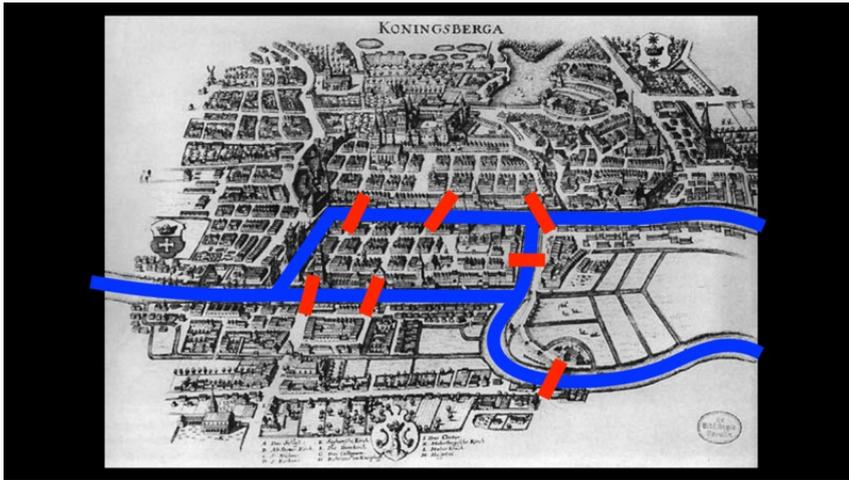
Figure 1.1: The problem of Seven Bridges of Königsberg: finding a way to walk around the town that would cross each bridge once and only once. Picture taken from Wikipedia https://en.wikipedia.org/wiki/K%C3%B6nigsberg.

represent the components of the system, and edges that connect pairs of vertices, and which represent the interconnections between the components. For instance, a social network can be represented by a network whose vertices represent individuals and whose edges are the social relationships between individuals.

To render such representations tractable, it is essential to have to hand methods for characterizing their salient properties. One way of viewing complex networks is as graphs whose connectivity properties deviate from those of regular graphs [34]. Whereas regular graphs can be thought of as simple, complex networks are highly non-regular in structure. Structural complexity is therefore perhaps the most important characteristic of a complex network as it reveals the way in which vertices and edges are arranged in the network, providing a significant influence on the network function and performance [47]. Computationally efficient measures for assessing structural complexity are therefore an imperative tool in the analysis of complex networks.

Graph theory offers an attractive route to such complexity measures since it provides effective tools for characterizing network structure together with their intrinsic complexity. This approach has led to the design of several practical methods for characterizing the global and local structure of undirected graphs [95] [35]. A good recent review of the state-of-the-art can be found in the collection of papers edited by Dehmer and Mowshowitz [39].

However, while the problem of characterizing the structural complexity of undirected graphs is well studied, there is relatively little literature aimed at studying the structural

features of directed graphs. One of the reasons for this is that the graph theory underpinning directed graphs is less developed than that for undirected graphs. In the real world, those complex networks represented by directed graphs are perhaps even more common than those represented by undirected graphs. For example, the World Wide Web is a directed network in which vertices represent web pages and edges are the hyperlinks between pages. Another common example is furnished by citation networks in which the vertices are scholarly papers while the edges are the citations between them. One recent exception is the work of Riis [109] who has extended the computation of entropy to directed graphs, using the concepts of guessing number and shortest index code. The author shows that the entropy is the same as the guessing number and is bounded by the graph size and the shortest index code length. Berwanger et al. [16] have proposed a new parameter for the complexity of infinite directed graphs by measuring the extent to which cycles in graphs are intertwined. Recently, Escolano et al. [45] have extended the heat diffusion-thermodynamic depth approach for undirected networks to directed networks and thus obtain a means of quantifying the complexity of structural patterns encoded by directed graphs. To fill this important gap in the literature, one of the research goals in this thesis is to develop novel and effective methods for quantifying the structural complexity for directed networks. To this end, we explore whether a number of characterizations developed only on undirected graphs can be extended to the domain of directed graphs, using some recent results from spectral graph theory.

Turning attention to a substantial branch in machine learning, namely the statistical pattern recognition, we note that the feature-vector-based methods are often used since they provide powerful and flexible tools for representing patterns. On the other hand, graph-based representations, which are employed in the area of structural pattern recognition, have proved to be an elegant means of implementing pattern characterization tasks, due to the fact that they have a number of advantages over feature vectors. For instance, graphs can represent more structural information of a pattern than feature vectors [108]. With such structural features to hand, especially multi-dimension ones, then problems such as graph embedding, clustering and classification can be addressed using standard machine learning and pattern recognition techniques. Motivated by the utility of both feature-vector-based methods and graph-based representations, in the thesis we seek a way to preserve the advantages of these two approaches. In particular, we aim to develop a method based on information theory to extract multi-dimensional graph features, which

can be served as a long-vector for characterizing the structure of both undirected and directed graphs. We also show that by performing principal component analysis (PCA) on the feature vectors for samples, we manage to embed populations of graphs into a low-dimensional feature space.

Over the recent years, there has been a vast amount of effort expended on the problems of how to represent networks, and from this representation derive succinct characterizations of network structure and in particular how this structure evolves with time [62] [4] [2]. Broadly speaking the representations and the resulting characterizations are goal-directed, and have centered around ways of capturing network substructure using clusters, or notions such as hubs and communities [94] [47] [50] [39]. Here the underlying representations are based on the connectivity structure of the network, or statistics that capture the connectivity structure such as degree distributions [5] [6].

A more principled approach is to try to characterize the properties of networks using ideas from statistical physics [64] [67]. Here the network can be succinctly described using a partition function, and thermodynamic characterizations of the network such as entropy, total energy and temperature can be derived from the partition function [89] [41] [53]. However, to embark on this type of analysis, the microstates of the network system must be specified and a clear interpretation of the network thermodynamics provided. This approach has provided some deep insights into network behaviour. For instance, in the work [18], the Bose-Einstein partition function is used to model a Bose gas on a network, and the process of Bose condensation and its quantum mechanical implications have been studied. This model has also been extended to understand processes such as super-symmetry in networks [17].

Although the bulk of existing network theory is concerned with static networks, most networks are in reality dynamic in nature. Specifically, networks grow and evolve with the addition of new components and connections, or the rewiring of connections from one component to another [2]. Motivated by the need to fill this gap in the literature and to augment the methods available for understanding the evolution of time-varying networks, in this thesis we aim to establish a thermodynamic framework for analyzing the structural evolution of time-varying networks. We also aim to develop a novel method for constructing a generative model to analyze the underlying average connectivity structure of dynamic networks (or graph time-series). We explore how this model can be fitted to the graph time-series data using an information theoretic approach that aims at minimizing

a description length criterion, with the von Neumann entropy encoding the complexity of the model. We then present a new fixed-point iteration scheme to locate the optimal structure of the generative model. This method explores a number of new perspectives on the study of time-evolving networks.

1.2 Research Goals

To address these problems encountered in network science and to make contributions to developing novel and efficient graph-based methods, the research goals in this thesis are the following:

- To explore whether a number of graph complexity measures previously defined only on undirected graphs, can be extended to the domain of directed graphs. Specifically, we aim to develop the directed analogues of the von Neumann entropy [99] and Estrada's heterogeneity index [46], by making use of some recent results from spectral graph theory concerning the construction of the normalized Laplacian matrix for directed graphs [31].

- To explore a number of uses and applications based on the development of the approximate von Neumann entropy of both undirected and directed graphs. In particular, we aim to investigate whether these applications can be used to quantify the entropic assortative mixing properties of networks and to deal with network analysis and machine learning problems, such as structural pattern recognition.

- To develop a novel method for characterizing the evolution of time-varying complex networks. Specifically, we aim to adopt a thermodynamic representation of network structure computed from a) a polynomial (or algebraic) characterization of graph structure and b) a statistical mechanical approach that associate the microscopic configurations with the eigenstates of the normalized Laplacian spectrum of a network respectively. Both approaches allow us to derive expressions for a number of thermodynamic functions including the energy, entropy and temperature.

- To develop a novel method for constructing a generative model to analyze the struc-

ture of a set of labeled graphs. In particular, the vertex set is fixed and the set of possible connections between vertices change between samples in the graph data. The generative model aims to represent these changes with a Gaussian probability distribution for the connection weights on each individual edge. We fit this model to the sample graph data by minimizing a description length criterion, with the von Neumann entropy controlling the complexity of the fitted model structure and the Gaussian log-likelihood controlling the mean edge weights and variances. Then, we aim to obtain the optimal model structure by processing a new fix-point iteration scheme.

1.3 Contributions

To achieve these research goals in this thesis, we make the following novel contributions:

- The first contribution in this thesis will be the development of a novel entropy measure for assessing the structural complexity of directed graphs. To commence, we use the recently defined directed graph normalized Laplacian matrix to extend the analysis of von Neumann entropy from undirected graphs to directed graphs. We will then show how this entropy measure can be approximated by simple graph characteristics, such as graph size and vertex in-degree and out-degree statistics. We further find approximations to the von Neumann entropy that apply to both weakly and strongly directed graphs. Moreover, we will define an analogous directed version of the graph heterogeneity index, which quantifies the structural heterogeneity properties of directed graphs. We will demonstrate experimentally the usefulness of these complexity measures in characterizing structure from both artificial and empirical data.

- The second substantial contribution will be the development of an entropic edge assortativity measure as well as an entropic graph embedding method, both of which are dependent on the von Neumann entropy development. First, commencing from the approximate entropy expression, we derive a quantity which measures the local entropic contribution associated with each edge in a graph. We will show how this quantity can be used to define a novel entropic edge assortativity measure for both

undirected and directed graphs. Moreover, we will proceed to show that this local measure in fact encodes a number of properties of the intrinsic structural properties of a graph, leading to the possibility of analyzing how the von Neumann entropy is distributed on the graph. We aim to explore whether such entropy distribution can be used to solve structural pattern recognition problems. To do this, we show that this entropy distribution can be in fact encoded as a multivariate array, which captures the structure of the graph in terms of an entropic measure of complexity. As a result, such array can be viewed as a sample of entropy histograms from various graphs, allowing us to embed populations of graphs into a low-dimensional feature space, by using PCA or other machine learning techniques. These structural features can be used to implement the tasks of graph clustering and classification, by employing suitable algorithms in machine learning techniques. Then, we will undertake experiments to compare the graph classification accuracy of the proposed embedding method with a number of state-of-the-art embedding and kernel methods, in order to investigate the efficiency of our embedding method.

- The third contribution we will make in this thesis is the development of the thermodynamic representations used for characterizing the evolution of time-varying complex networks. Based on the idea that most of the aggregate thermodynamic variables such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives, we aim to establish a link between a characteristic polynomial (or algebraic) characterization of a network and the Boltzmann partition function. This allows us to derive a number of thermodynamic quantities for the network, including the average energy and entropy. Assuming that the system does not change volume, we can also compute the temperature, defined as the rate of change of entropy with energy. On the other hand, we will show that the approximate von Neumann entropy can also be interpreted as the thermodynamic entropy of a network, when we associate the microscopic configurations of the network with the eigenstates of the normalized Laplacian spectrum. This interpretation gives us a new approach to developing additional thermodynamic functions, such as internal energy and temperature. We will show all these thermodynamic measures are based on simple graph features including graph size, number of edges and degree statistics. Finally, the experimental results will reveal that both

thermodynamic representations provide an efficient tool for detecting abrupt changes and characterizing different stages in financial and biological network evolution.

- The final substantial contribution in the thesis will be the development of a generative model that captures the underlying connectivity structure of a set of labeled graphs. Specifically, in the graph data the weights of the connectivity between vertices in the graph change while the vertex number and label do not. So we only concentrate on the edge patterns present in those graphs and we adopt a Gaussian probability distribution to represent the connectivity changes on each individual edge. This yields a probabilistic framework which describes the likelihood of the observed data given the model structure. Then, we will explore how this structure can be fitted to the sample graph data by minimizing a description length criterion. We will show how this problem can be solved numerically by adopting a new fixed-point iteration scheme which locates the elements of the optimal weighted adjacency matrix of the model structure. This structure is initialized using the mean weighted adjacency matrix for the sample graphs, and then is optimized at each step to best fit the data by an adjustment that is determined by the von Neumann entropy. Finally, we will undertake experiments to evaluate the properties of the model learning method and also to explore the practical utility of the generative model on real-world data, and this shows the generative model reveals new perspectives in analyzing graph time-series data.

1.4 Thesis Structure

The remainder of the thesis is structured as follows. In Chapter 2, we review the research literature related to the work presented in the thesis. In Chapter 3, we give the development of the approximate von Neumann entropy and heterogeneity index of directed graphs. Chapter 4 presents the work of applications of the von Neumann entropy, including the entropic edge assortativity measure and entropic graph embedding methods. Chapter 5 provides two different thermodynamic frameworks for analyzing the structural evolutions of time-varying networks. In Chapter 6, we develop a method for learning a generative model that captures the underlying connectivity structure of a set of sample

graphs, and explore how this model can be used to better analyze graph time-series data. Finally, Chapter 7 concludes the work reported in the thesis and provides possible future research directions.

Chapter 2

Literature Survey

A *complex network* can be formally defined as a diagrammatic representation of a complex system. It consists of vertices, which represent the components of the system, and edges that connect pairs of vertices, and which represent the interconnections between the components. Generally speaking, what makes the networks "complex" is that they often display non-trivial topological properties, which are uncommon in "simple" networks such as regular graphs. In other words, complex networks are so huge that it is difficult to understand their global characteristics by analyzing the local properties of individual vertices or edges [123] [21] [73]. Therefore, in order to render such networks tractable, it is essential to have to hand methods for characterizing their salient properties. In this thesis, our research aims at developing effective methods for characterizing complex networks with different structure and understanding the time evolution of dynamic networks. In the light of this aim, we review relevant research literature on these network science challenges in this chapter.

2.1   Graph Theory in Natural Sciences

One way of viewing complex networks is as graphs whose topological features deviate from those of regular graphs [34]. Whereas regular graphs can be thought of as simple, complex networks are highly non-regular in structure. It is for this reason that graph theory and its applications play a vital role in diverse research fields, including biochemistry, electrical engineering, computer science and operations research [101]. In the following, we briefly explain how the graph-theoretic applications contribute to these areas.

- **Biochemistry:** Due to the complexity of the control mechanisms involved, and the
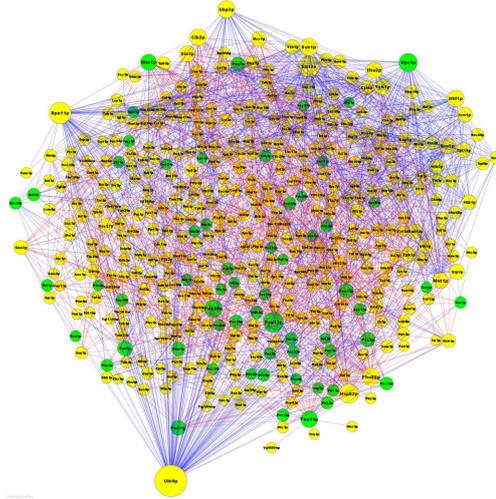
Figure 2.1: The peroxisomal protein-protein interaction network. The green circles represent peroxisomal core proteins and the yellow ones are their direct neighbours. The red and blue lines connecting circles indicate the protein-protein interactions are detected by either binary or cluster assay respectively. Picture by Luciani and Bazzoni [83].

large number of possible interactions, graph-based models are becoming increasingly popular in bioinformatics and chemoinformatics. Specifically, graphs can be used to model cellular networks, whose cellular components are represented by vertices and whose cellular interactions are edges in the graph. A good example is furnished by the protein-protein interaction (PPI) networks, whose structure is often represented by undirected graphs. The proteins are modeled by vertices and edges are drawn between two vertices if the corresponding proteins physically bind [1]. Such models allow scientists to analyze the similarity between proteins and to predict the characteristics of the cells.

- **Telecommunications engineering:** A communications network is often formally defined as a collection of terminals and telecommunications links. Graphs have proved to be an important application in modeling communications networks, with vertices representing terminals and edges indicating telecommunications between terminals [42]. One of the objectives in telecommunications engineering is to solve the problem of how accurately and how effectively the symbols of communication can be transmitted on the telecommunications network.

- **Computer vision:** Over the recent years, graph-based methods have been widely employed to overcome serious challenges in computer vision, such as object recognition, correspondence matching and image segmentation. Overall, hierarchical image features are regularly modeled by tree structures or directed acyclic graphs, whose vertices represent image abstractions and whose edges indicate the spatial relations or mappings between them [70].

- **Operations research:** Graph-based methods are also widely used as a powerful tool for implementing operations research problems, such as modeling transport networks, activity networks and theory of games. One of the most successful applications is furnished by the planning and scheduling of large complicated projects in operations research [117].

- **Network analysis:** In recent decades, a variety of network science techniques have been developed to analyze the behaviour of networked systems. Traditionally, such systems are modeled as random graphs, i.e., the classical Erdős-Rényi model [43]. Specifically, this model defines a random graph as a number of labeled vertices randomly connected by edges with a given probability. The random graph theory has proved to provide many important results that are relevant to complex networks [3]. However, the network science has witnessed a shift in focus in recent years, centering on the challenge of modeling networks as mechanical systems and studying their statistical properties [94]. For instance, Watts and Strogatz [124] have proposed a model that can be used to simulate network "small-world" effects. Moreover, the degree distribution of a large number of realistic networks has been shown to follow a power-law distribution, indicating that those networks are in fact "scale-free". To model such networks, the Barabási-Albert model has been developed to generate random "scale-free" graphs using a preferential attachment mechanism [13].

2.2   Graph Characterization

We focus on one of the most challenging problems in network science, namely the characterization of graph structure. The key to this problem is to extract a set of characteristics

(a) British Isles　　　　　　　　　　　　　　(b) Europe
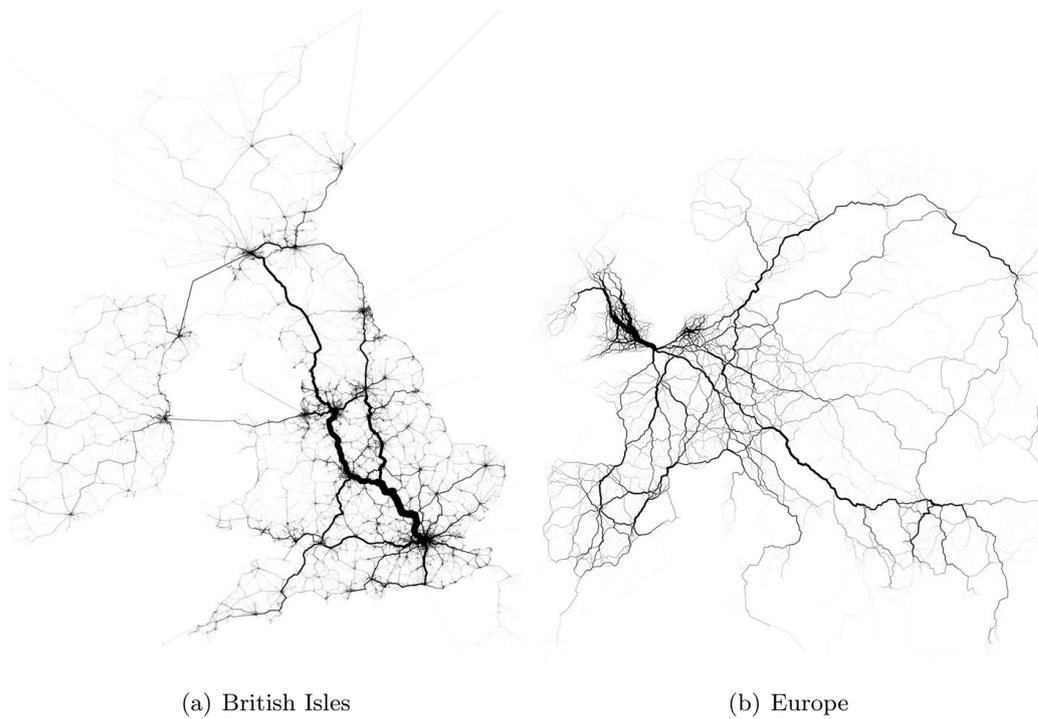
Figure 2.2: Transport networks generated from large-scale online data (over one million Tweet-based trips) in August 2011. The thickness of a transport line in the network is proportional to the volume of Tweets sent along its path. Picture by Eric Fischer https://www.flickr.com/photos/walkingsf/albums/72157628993413851.

from graph structure which capture both the individual structural pattern and the variations between structure classes. Broadly speaking, graph theorists have adopted three different routes to addressing the graph characterization problem, namely a) topological features, b) spectral or algebraic graph theory and c) complexity measures.

2.2.1 Topological Features

One of the most widely encountered problems in network analysis is how to find the most "influential" or "functional" vertex in a network. The conceptually simplest centrality measure is the vertex degree, which is the number of edges connected to the vertex. The degree information gives many useful insights about the topology of a graph. The degree distribution function characterizes the spread of vertex degree in a graph, and gives the probability that a vertex has a particular degree. In the work [3], it has been revealed that in a random graph, the vertex degree generally follows a Poisson distribution, while the degree distribution of most realistic networks significantly differs from that of a random graph.

Another effective vertex centrality measure is called the closeness, which is dependent on the total distance of the vertex to all other vertices in a graph. In effect, the closeness centrality measures the speed of the information flow from a given vertex to others on a graph [48], as a high closeness vertex has a small total distance from all other reachable vertices to it. The betweenness centrality, on the other hand, quantifies the frequency of a vertex lies on the shortest path between two other vertices. In other words, it gives how effectively that a vertex exerts over the interactions of other vertices in a network [52]. Other important centrality measures include the Katz centrality and Google's PageRank [72].

Although the vertices play an important role in understanding the graph topology, many useful insights also reside in the information conveyed by paths on a graph. The length of the longest geodesic path between any pair of vertices in a graph is called the graph diameter. Chung and Lu [32] have studied that when the vertex connecting probability is large, most Erdős-Rényi random graphs are likely to have small diameters. This is because the graph diameter is proportional to the ratio between the logarithms of vertex number and average degree [3]. In other words, the diameter is a measurement of the linear size of a graph.

Another well-studied topological feature is the clustering coefficient. In many empirical

networks it can be observed that if vertices a and b are connected, b and c are connected, then a and c are highly likely to become connected. For instance, in a social network, individuals often create cliques with a high density of ties. In other words, the "friend of your friend is likely also to be your friend" [94]. Mathematically, this inherent tendency to clustering is quantified by [124]

$$\text{Clustering Coefficient} = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}}.$$

In this formula, a connected triplet is a connected substructure consisting of three vertices and two edges. The factor 3 in the numerator is needed because of the fact that each triangle is counted three times in the formula. In principle, the clustering coefficient quantifies triangle density in a network by computing the average probability that two vertices that share the same neighbours are connected.

2.2.2 Spectral and Algebraic Graph Theory

An alternative approach to probing graph structure is to extract vertex permutation invariants straightforwardly from a number of graph matrix representations. Two of the most effective routes adopted by graph theorists include spectral graph theory and algebraic graph theory [20]. These two approaches are intimately related. Both commence from a matrix representation of a graph. In the case of spectral graph theory, it is the eigenvalues and eigenvectors of the matrix that are of interest [99] [25]. In algebraic graph theory, a characteristic polynomial is computed from the determinant of the identity matrix minus a multiple of the matrix. The coefficients of this polynomial are determined by symmetric polynomials of the matrix eigenvalues and they provide many useful graph invariants.

Generally speaking, spectral graph theory studies the properties of the spectra of graph matrix representations, such as the adjacency matrix, the Laplacian matrix and their normalized forms. Although it is generally difficult to characterize a graph by its spectra, some essential structural characteristics of the graph can still be deduced from them [30]. Graph eigenvalues have proved to play a crucial role in the fundamental understanding of graphs. From the Perron-Frobenius theorem, we can immediately draw the conclusion that the largest eigenvalue of the adjacency matrix has multiplicity one if the graph is connected. It is interesting to note that for the Laplacian matrix, this property corresponds to its smallest eigenvalue, i.e., zero-eigenvalue. More generally, the multiplicity

of 0 as an eigenvalue of the Laplacian matrix is equivalent to the number of connected components in the graph [31]. The second smallest Laplacian eigenvalue also plays a special role in a number of problems in graph theory, including the graph expansion and the maximum cut problem. The difference between the second and the first eigenvalues, which is called the eigenvalue gap, measures the graph connectivity properties [30]. Moreover, the Perron-Frobenius theorem also implies that the left eigenvector corresponding to the largest eigenvalue of the transition matrix, is the unique stationary distribution of a random walk on a graph, which is a very important result in random walk theory.

Overall, spectral methods have been exploited directly and with great effect in graph characterization and machine learning. Much of this is due to the close links between graph spectra and random walks on graphs. However, there has been less interest in the algebraic approach. This may be something of an oversight, since there are strong links between algebraic graph theory and number theory, and results from algebraic graph theory can be used to construct important invariants that can be used to probe network structure.

More explicitly, algebraic graph theory aims at using linear algebra, group theory and graph invariants to study the properties of structural patterns. The coefficients of the characteristic polynomial of a graph matrix can be taken as graph characteristics, and have been shown closely related to the graph spectra [106]. For example, the coefficients of the Laplacian characteristic polynomial are related to the number of spanning trees and spanning forests in a graph. In particular, for certain graphs in (a, b)-linear classes, the coefficients can be simply determined by the number of vertices in the graph [96]. Moreover, the Laplacian matrix can be used to construct a zeta function, which can be viewed as an analogue of the Riemann zeta function from number theory [111]. This zeta function, is in fact the moment generating function for the heat kernel, and its derivative at origin can be used to measure the number of spanning trees contained in a graph [127]. The Ihara zeta function, which is derived from a characteristic polynomial for the oriented line graph of a network, can be used to determine the distribution of prime cycles of various length in the network and is also closely linked to the dynamics of a discrete time quantum walk on the network [113] [105] [104]. This latter type of representation has been shown to lift some of the problems in cospectrality of networks encountered if conventional spectral methods are used.

2.2.3   Complexity Measures

Quantifying the intrinsic complexity of graphs is a problem of fundamental practical importance, not only in network analysis but also in other areas such as pattern recognition and control theory [48]. Existing approaches are based either on randomness complexity or statistical complexity. The difference between these two approaches is that randomness complexity aims to quantify the degree of randomness or disorganization of a combinatorial structure, while statistical complexity aims to distinguish a combinatorial structure using statistical features such as vertex degree statistics, edge density or the Laplacian spectrum. Historically, most early work in this area falls into the randomness class, while recent work is statistically based and aims to compute entropic measures of complexity.

2.2.3.1   Randomness Complexity

Graph-based entropy measures have proved to be an elegant approach to quantifying the randomness complexity of graphs. More explicitly, this approach aims at applying an entropy function to a probability distribution defined on a graph. For example, an effective way to accomplish this is to assign a probability distribution over the components of the partition of a graph based on graph structural features [39]. The most commonly adopted complexity measure is Shannon's entropy function, but several different families of entropy functions have also been developed.

The concept of graph entropy is first introduced by Rashevsky [103], who has named this measure as the "topological information content" in his early work. Specifically, Rashevsky has used a number of graph invariants, including the vertex number and the degree sequence, to develop entropic measures that quantify the structural complexity of graphs. Commencing from Rashevsky's topological information content, Mowshowitz [91] has further developed graph structural information measures based on the automorphism partitioning and the chromatic decomposition of the vertices in a graph. Some interesting results can be drawn regarding this entropy measure. For instance, graphs whose automorphism group consists of the identity alone have the maximal entropy while the minimal entropy is reached for complete graphs. The properties of this entropy have been further investigated in the work [39].

One of the earliest and classical contributions in randomness complexity is Körner's entropy associated with a graph [75]. The original motivation of this measure is to compute the entropy of an information source with ambiguous alphabet. Extending this idea

to graphs, the Körner's entropy is defined as the minimal cross entropy between the vertex packing polytope and the vertex probability distribution [45]. Unfortunately, as this complexity measure is posed as a coding optimization problem in information theory, it cannot be used as a quantity to reflect the graph structural properties. Another drawback of this approach is that it is not applicable to more general unweighted graphs. These shortcomings seriously limit the direct use of Körner's entropy in the field of network analysis.

There are a number of alternative methods that are based on Shannon's entropy for quantifying the complexity of a graph. In general, Shannon's entropy function can be directly applied to a probability distribution whose values are assigned by functions that capture the structural characteristics of a graph. Recently, the normalized Laplacian spectrum has been shown to provide a complexity level characterization via definition of the von Neumann entropy (or quantum entropy) associated with a density matrix [99] [4]. By mapping between discrete Laplacians and quantum states [25], provided that the discrete Laplacian [19] [20] is scaled by the inverse of the volume of the graph, a density matrix is obtained whose entropy can be computed using the spectrum of the discrete Laplacian. This measure can distinguish between different structures in extremal graph theory. In particular, star graphs yield the maximal entropy value and the entropy reaches its minimum for regular graphs.

Recently, based on the development of the graph topological information functionals, Dehmer has proposed a generalized framework for defining graph entropies [38]. Specifically, by using graph topological characteristics to map vertices to positive reals, the so-called "information functionals" can be defined. Then, a probability distribution can be derived from such functionals, which can further be used to compute the graph entropy. Clearly, with various topological information functionals to hand, different interpretations of the graph entropy can be defined. The resulting entropies have proved to be useful in classifying graphs with various structural patterns. Moreover, this generalized framework offers an attractive route to defining the complexity trace, which can be used to characterize phase transitions in graph structure. Finally, the computational complexity of the graph entropy is polynomial, which shows the efficiency of this generalized framework.

An important extension of this framework is the work [40]. Motivated by the fact that the classical structure descriptors based on the adjacency matrix spectrum are often misleading because of the large number of isospectral graphs, this work aims to tackle

this problem and focuses on the spectrum of other certain graph matrix representations in order to define a number of novel descriptor families. Such descriptors then allow a novel graph entropy that quantifies the complexity of the underlying graph topology with a given information functional to be defined. The experimental results suggest that these novel complexity measures are effective in characterizing different chemical structures and thus, have broad applications in fields related to structural chemistry such as drug design and medical chemistry [39].

By applying an entropy function to the normalized degree correlation matrix, the so-called "off-diagonal" complexity measure has been proposed recently [33]. The central idea commences from the observation that a biased edge distribution entropy, whose extremal value is reached for a power-law distribution, can be defined on small "scale-free" graphs. Extending this approach to the distribution of correlations between degrees of pairs of vertices, a discrete graph entropy can be defined and computed. This complexity takes on the value zero for both regular lattices and complete graphs and has small values for random graphs and large values for complex structures.

2.2.3.2 Statistical Complexity

The main drawback of randomness complexity is that it does not capture properly the correlations between vertices [50]. Statistical complexity aims to overcome this problem by measuring regularities beyond randomness, and does not necessarily grow monotonically with randomness. It is natural to realize that both completely random systems and completely ordered ones should have a minimal statistical complexity. The first randomness complexity measure introduced, namely the Kolmogorov complexity [74] of an object, is quantified by the length of the shortest algorithm required to reproduce the object. On the other hand, the statistical counterpart of Kolmogorov complexity, the logical depth devised by Bennett [15], is a measure of complexity based on the algorithmic information and computational complexity of an algorithm which can be used to recreate a given piece of information. In essence, the logical depth complexity measure differs from its randomness counterpart in that it is based on the notion of a process rather than a measure.

A set of important recent additions to the statistical complexity literature are the graph spectral methods. In fact, there exist strong links between the eigenvalue-based and polynomial-based approaches and many practical graph structure characterizations have been developed based on such connections. For example, Luo et al. [84] have defined the

eigenmodes using the leading eigenvectors of the graph adjacency matrix, and that can be used to compute vectors of spectral properties. Then, graphs can be embedded in a pattern space via those vectors. The method has proved to be efficient in overcoming problems such as graph clustering and object identification. It is also known that the spectrum of the graph Laplacians can be used as an elegant means of characterizing the topological structure of graphs. For instance, Wilson et al. [126] focus on the Laplacian spectral decomposition and show how the coefficients of the permutation invariant polynomials that are computed from the elements of the spectral matrix for the graph Laplacians, can be used as features that capture the metric structure of graphs. Another important example is furnished by Estrada's network heterogeneity index [46]. In effect, this index gauges differences in degree for all pairs of connected vertices and is dependent on vertex degree statistics and graph size. The expression for the index can be expressed in terms of the Laplacian matrix of graphs. The lower bound of this quantity is zero, which occurs for a regular graph while the upper bound is equal to one, which is obtained for a star graph.

A finer characterization is provided by the Ihara zeta function [66], which is a natural extension of the Riemann zeta function in number theory, and is built out of prime cycles in a graph. Specifically, the expression for the Ihara zeta function of a graph is closely related to the quasi characteristic polynomial of the adjacency matrix of the associated oriented line graph, whose polynomial coefficients are determined by the cycle structure of the graph. Commencing from this observation, Ren et al. [106] have developed a novel method for characterizing unweighted graphs by using the polynomial coefficients determined by the Ihara zeta function. They also show how the graph characterization based on the Ihara coefficients can be extended from unweighted to weighted graphs. Experimental evaluations suggest that the Ihara coefficients in fact outperform the graph spectral methods in terms of distinguishing structures that belong to various structural classes.

The thermodynamic depth is a measure which takes on low values for both random and ordered systems. Hence, it is interesting to explore whether the idea of thermodynamic depth can be extended to graphs in order to obtain novel complexity quantities. However, this requires a definition of the macroscopic states of the graph. To overcome this challenge, Escolano et al. [45] [44] show that by establishing a link between heat kernels and Birkhoff polytopes on a graph, a time-evolving complexity measure can be obtained. Working in

the domain of structural pattern recognition, Xiao et al. [127] have explored how the heat kernel trace can be used as a means of characterizing the structural complexity of graphs. They have also used the derivative of the zeta function at origin as a characterization for distinguishing different types of graphs. Moreover, they show how the heat-content can be used to develop a series of graph invariants. In common with the symmetric polynomials, the heat-content coefficients are permutation invariants that depend on both the eigenvalues and eigenvectors of the Laplacian matrix associated with a graph.

2.3   Graph Causality

Recently, there has been a vast amount of effort expended on the problems of understanding the causal relationships between complex network components, such as the economic agents in financial markets [87] [77] [76]. In general, most current literature aims at studying correlation-based networks. In fact, there exist a large number of distinct relationships between economic components in a financial market. By adopting appropriate filtration methods, the most influential correlations can be reserved for constructing the financial market network, which is used for further statistical analyses [122].

Broadly speaking, the correlation-based networks are obtained via two dominant relationship filtration approaches, namely a) hierarchical clustering and b) thresholding [37]. For hierarchical methods, the networks can be constructed from the Minimum Spanning Tree (MST) [86], which has proved to be one of the earliest and most important filtration techniques. More explicitly, the filtering procedure linked to MST and SL (Single Linkage, which is a hierarchical clustering mechanism that is closely related to MST) allows the elements in the financial market to be arranged into a hierarchical structure, and has proved to provide an efficient way to improve portfolio optimization [121]. Another effective method for building correlation-based networks is the Planar Maximally Filtered Graph (PMFG) [122], which is a generalized framework of the MST, and manages to maintain a higher amount of information, with less strict topological constraints. Although the hierarchical clustering methods are effective in uncovering the nested structure of stock correlations in a financial market, they have a main drawback, namely the strict topological constraints on the network, which make it difficult to reflect the statistical significance of correlations.

On the other hand, for the thresholding methods, a correlation-based network can either be created by simply retaining the n-largest correlations between assets [97], or

more generally, by examining the correlations between economic agents. In other words, the connection between two components is dependent on whether or not their correlation exceeds a predetermined threshold value [76] [65] [61]. So determining the threshold value is critical in constructing threshold networks. In particular, a lower threshold leads to the observation that groups of economic elements gradually merge to form larger groups, and finally merge into the whole financial market [37]. Conversely, with an increasing threshold, the market progressively disintegrates into smaller fragments of economic sectors. This shows the main weakness of thresholding methods is that the appropriate threshold value is difficult to determine, and the resulting networks may not be able to show the nested structure in the financial market. The advantage of this approach, one the other hand, is that threshold-based network are robust to correlation uncertainty.

Turning attention to the causal inference of graphs, there are two extensively-used methods for deriving the relationships in a network, namely a) cross-correlation and b) Granger causality. Cross-correlations of financial time-series are of great interest at both theoretical and practical levels [114]. This is because they contain information about the way how the time-series influences each other over time, which provides a useful route to reflecting the price changes in the stock market [76]. On the other hand, the Granger causality concept, originally defined by Wiener [125] and Granger [57], has attracted a great deal of interest in the econometrics literature since the 1960s. In principle, Granger causality represents a causal relationship between two time-series, i.e., if by including the past information of one time-series, the prediction error of another time-series can be reduced, then we say the first time-series Granger-causes the second one [129]. In other words, the Granger causality gives the information whether one time-series can help forecast another. Recently, the original Granger causality notation has been extended to multivariate cases: the conditional Granger causality analysis [55] has been developed for multivariate data and makes the Granger causality interpretation more straightforward.

2.4 Graph Evolution

Until recently, one fundamental field of graph theory that has broad applications in network analysis, has received marginal attention, namely evolutionary graph theory. In fact, many real-world complex systems such as the citation networks, communications networks, neural networks and financial networks give rise to structures that change with time. In order to analyze such graphs, efficient tools for understanding their time-dependent struc-

ture and function are required. In general, graph evolution can be approached from both macroscopic and microscopic directions [58].

On the one hand, the macroscopic approach aims at studying how the global parameters of a dynamic graph evolve from one snapshot to another. This can be accomplished by directly employing a number of graph characterizations that are developed on static graphs to each snapshot, and then analyzing the time evolution of these characterizations. Specifically, statistical thermodynamics can be combined with both graph theory and kinetics to provide a practical framework for handling highly structured and highly interactive time-evolving complex systems [89]. By using a random walk that maximizes the Ruelle-Bowens free-energy rate on weighted graphs, a novel centrality measure can be computed, and this has been successfully applied to both connected and disconnected large-scale networks [41]. Recently, it has been demonstrated that the subgraph centrality can be interpreted as a partition function of a network [49], and as a result the entropy, internal energy and the Helmholtz free energy can be defined using spectral graph theory. The authors have also argued that the thermodynamic quantities are intimately related to the complex network dynamics. This approach combines the theoretical tools developed for studying graph spectra in the context of statistical mechanics of complex networks and clearly point out the potentials of the current approach to study real-world time-varying networks.

On the other hand, at the microscopic level, it is the birth-death dynamics of individual vertex or edge in the graph evolution that are under study. In the classical Barabási-Albert model [13], dynamic properties are ascribed to a preferential attachment mechanism for graph growth, i.e., adding new vertices which connect to the existing vertices in the graph. The preferential attachment mechanism particularly shows that the connection probability of an existing vertex and the newly added one is proportional to its degree. In effect, this mechanism describes the dynamics as the addition of new vertices. However, graph evolution can also be shaped by a number of extra microscopic events, including the removal of vertices and the addition, removal and rewiring of edges [3]. Based on this observation, Grindrod and Higham [58] have introduced a tractable framework for modeling evolving graphs. To do this, they propose a novel range-dependent birth-death mechanism, which allows a variety of evolutionary behaviours to be modeled. The resulting dynamic graph model is set up as a discrete-time Markov chain, and an analogous continuous-time framework can also be developed. This model has proved to be efficient in investigating the

evolutionary processes that take place on an evolving graph.

2.5   Summary

In this chapter, we have provided a research literature survey on a) applications of graph theory in natural sciences, b) different methods for characterizing graph structure and c) approaches to studying and modeling graph dynamics. We have also analyzed the strengths and deficiencies of the existing methods. To summarize this chapter, we draw a number of conclusions based on the relevant research literature survey.

Complex network analysis has proved to play an increasingly significant role in natural sciences recently, and graph theory offers an interesting route to characterizing the structural features of networks. Although topological characteristics provide a straightforward meaning concerning the graph structure, their direct use in structural characterization is limited because common graph topological features do not take topological scales into account. Therefore, more recent literature aims at characterizing graph structure using spectral or algebraic graph theory and complexity measures. Specifically, graph-based entropy measures have proved to be an elegant approach to quantifying the structural complexity of graphs. Unfortunately, although many entropy functions have been adopted to analyze the complexity of undirected graphs, such as the Körner's entropy, Shannon entropy and von Neumann entropy, there is relatively little literature aimed at studying the structural features of directed graphs. One of the reasons for this is that the graph theory underpinning directed graphs is less developed than that for undirected graphs. Motivated by the need to fill this gap in the literature, in Chapter 3 we explore whether the von Neumann entropy, and other complexity measures, which have been defined only on undirected graphs, can be extended to the domain of directed graphs. We also show how to simplify the computation of these measures and evaluate their properties on both artificial and real-world network data. In Chapter 4 we further explore some uses and applications based on the von Neumann entropy expression. In particular, the approximate entropy formula gives a distribution of how the von Neumann entropy spreads on a graph. Commencing from the edge entropy and entropy distribution, we suggest a number of novel graph structure characterizations for both undirected and directed graphs, including an entropic assortativity measure and an entropic graph embedding method.

Although the bulk of existing network theory is concerned with static networks, most realistic networks are in reality dynamic in nature. Turning attention to the literature

relevant to the time-evolving networks, we note that many methods do not take into account the graph structural changes between snapshots in the graph time-series, such as the vertex attribute change, degree change and edge weight change. In order to make sufficient use of such information and to propose novel methods for studying the evolution of dynamic networks, in Chapter 5 we present a thermodynamic representation of graphs based on the link between characteristic polynomial and the partition function. Moreover, commencing from the link between network microscopic configurations and its Laplacian eigenstates, we explore whether the von Neumann entropy can be used as a thermodynamic entropy, and also whether we can develop expressions for other thermodynamic functions, including the internal energy and temperature accordingly. Finally, Chapter 6 develops a novel method for analyzing the time evolution of time-evolving networks, by learning a generative model that best captures the underlying connectivity structure present in a set of labeled graphs.

Chapter 3

Von Neumann Entropy of Directed Complex Networks

This chapter is motivated by the need to establish novel and effective methods for measuring the structural complexity of directed graphs. In particular, we first explore whether the von Neumann entropy previously defined only on undirected graphs [99] can be extended to the domain of directed graphs. To do this we make use of some recent results from spectral graph theory concerning the construction of the normalized Laplacian matrix for directed graphs [31]. We then show how to extend the heterogeneity index, which is developed for undirected graphs and dependent on vertex degrees as well as graph size, to the domain of directed graphs. We illustrate the usefulness of these graph complexity measures defined in this chapter on both artificial and real-world datasets, including structures from protein databases and high energy physics theory citation networks.

3.1 Preliminaries

Before introducing the development of the von Neumann entropy, in this section we provide some basic definitions and notations in graph theory that will be used throughout the thesis.

A *graph* is an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a *vertex* set $\mathcal{V}$ together with an *edge* set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. If $(u, v) \in \mathcal{E}$, i.e., vertices u and v are connected by edge (u, v), we say that u is *adjacent* to v.

An *undirected* graph is a graph in which edges have no orientation, which means that edge (u, v) is identical to edge (v, u). On the other hand, a *directed* graph is a graph with

an orientation on each edge. In particular, an edge (u, v) is considered to be directed from u to v, where u is called the *starting* vertex and v is called the *end* vertex. In the remaining of this thesis we will use the notation $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent both undirected and directed graphs.

The *adjacency matrix* A of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as

$$A_{uv} = \begin{cases} 1 & \text{if } (u, v) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

For undirected graphs, the *degree* at vertex u is defined as the number of vertices adjacent to u

$$d_u = \sum_{v \in \mathcal{V}} A_{uv} = \sum_{v \in \mathcal{V}} A_{vu}.$$

Similarly, for directed graphs, the *in-degree* and *out-degree* at vertex u are respectively given as

$$d_u^{in} = \sum_{v \in \mathcal{V}} A_{vu}, \; d_u^{out} = \sum_{v \in \mathcal{V}} A_{uv}.$$

A *path* in a graph is a sequence of edges that connect a sequence of distinctive vertices. When there is a path between every pair of vertices in a graph, the graph is said to be *connected*.

For a directed graph, it is *strongly connected* if every vertex in the graph is reachable from any other vertex. In other words, there are directed paths that contain any two vertices in the directed graph.

The *(combinatorial) Laplacian matrix* L of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as $L = D - A$ where D is the degree matrix with the degrees of the vertices of the graph along the diagonal and zeros elsewhere. The elementwise expression of L is given as

$$L_{uv} = \begin{cases} -1 & \text{if } (u, v) \in \mathcal{E} \\ d_u & \text{if } u = v \\ 0 & \text{otherwise.} \end{cases}$$

In the remainder of the thesis, we will apply a number of concepts from physics to the domains of graph theory and network science to study the properties of complex networks. In order to have a deeper understanding of these ideas, here we give a brief introduction of these physics definitions.

In information theory, the *entropy* (or *Shannon entropy*), is defined as the average value of information contained in a message, which is transmitted via a channel between transmitter and receiver.

In classical statistical mechanics, the *Gibbs entropy*, named after Josiah Willard Gibbs, is interpreted as the statistical entropy of the distribution of microstates that defines the macroscopic state of a system. On the other hand, the *von Neumann entropy*, named after John von Neumann, is the quantum analogue of the classical Gibbs entropy in the field of quantum mechanics.

Another significant concept in statistical mechanics is the *partition function*, which describes the statistical properties of a system in thermodynamic equilibrium. More importantly, the partition function can be used to derive most of the aggregate thermodynamic functions of the system, including the total energy, entropy and temperature.

3.1.1 Datasets Overview

In this subsection we give an overview of the datasets will be used for experiments in this thesis. In particular, we will use a large number of different datasets: the first two are synthetically generated artificial networks; while the other datasets are extracted from real-world complex systems, including databases from both biological and financial domains.

- *Preferential Attachment Network Dataset.* Consists of 10 directed networks evolved under preferential attachment. Each network starts from a fully connected seed network of 5 vertices. At each time step, a new vertex is added to the network. This vertex connects to vertices already in the network with a probability proportional to the steady state probability of a random walk taking place on the network. See Antiqueira et al. [8] for details about the model.

- *Random Directed Graph Dataset.* Contains a large number of directed graphs which are randomly generated according to one of three different directed random graph models, namely a) the classical Erdős-Rényi model, b) the "small-world" model, introduced by Watts and Strogatz [124], and c) the Barabási-Albert model [13]. The different directed graphs in the database are created using a variety of model parameters, e.g., the graph size and the connection probability in the Erdős-Rényi model, the edge rewiring probability [47] in the "small-world" model and the number of added connections at each time step [47] in the Barabási-Albert model.

- *Protein Dataset.* Is extracted from the protein database previously used by Riesen and Bunke [107]. It consists of over 200 graphs, representing proteins from the Protein Data Bank [16], labeled with their corresponding enzyme class labels from the BRENDA enzyme database [112]. The database consists of six classes (labeled EC 1,$\cdots$, EC 6), which represent proteins out of the six enzyme commission top level hierarchy (EC classes). The proteins are converted into graphs by first replacing the secondary structure elements of a protein with vertices, and then constructing a 3-nearest neighbour graph for the secondary structure elements. The graphs are thus directed.

- *Citation Network Dataset.* Is the Arxiv HEP-TH (high energy physics theory) citation network. This is an evolving citation graph Gehrke et al. [54] extracted from the e-print arXiv. The directed network represents the citations within a dataset of 27770 papers by 352807 directed edges. If a paper u cites paper v, then the graph contains a directed edge from vertex u to vertex v. Since there is no information about papers that are not included in the database, we do not consider such papers in the network. The data covers papers in the period from January 1993 to April 2003 (124 months). It begins within a few months of the inception of the arXiv, and thus represents essentially the complete history of its HEP-TH section [80].

- *MUTAG Dataset.* Consists of graphs representing 188 mutagenic aromatic and heteroaromatic nitro compounds assayed for mutagenicity on bacterium Salmonella typhimurium. We delete the labels of vertices and edges of each compound in order to transform these weighted graphs to unweighted ones.

- *NCI1 and NCI109 Dataset.* Are two subsets of the National Cancer Institute (NCI) database, consisting of graphs representing chemical compounds screened for activity against non-small cell lung cancer and ovarian cancer cell line respectively. Each subset contains two classes, labeled by active or inactive anti-cancer screen.

- *D&D Dataset.* Contains 1178 proteins, with 691 enzymes and 487 non-enzymes. Each protein is represented by a graph, in which the vertices are amino acids while

edges are the connections between acids.

- *COIL Dataset.* Contains object recognition data collected by Nene et al. [92], in which each 3D object consists of 72 images collected from equally spaced changes in viewing direction over 360 degrees. For each image, we establish a 3-nearest neighbour graph on the extracted feature points, i.e., each feature point have three directed edges going to its nearest neighbour points, thus the graph is directed and the out-degree of all vertices is 3. There are two subsets in this database, one contains the directed graphs extracted from 4 different 3D objects while the other contains graphs from 8 objects.

- *NYSE Stock Market Network Dataset.* Is extracted from a database consisting of the daily prices of 3799 stocks traded on the New York Stock Exchange (NYSE). This data has been well analyzed in [118], which has provided an empirical investigation studying the role of communities in the structure of the inferred NYSE stock market. The authors have also defined a community-based model to represent the topological variations of the market during financial crises. Here we make use of a similar representation of the financial database. Specifically, we employ the correlation-based network to represent the structure of the stock market since many meaningful economic insights can be extracted from the stock correlation matrices [14] [22] [27]. To construct the dynamic network, 347 stocks that have historical data from January 1986 to February 2011 are selected [100] [118]. Then, we use a time window of 28 days and move this window along time to obtain a sequence (from day 29 to day 6004) in which each temporal window contains a time-series of the daily return stock values over a 28-day period. We represent trades between different stocks as a network. For each time window, we compute the cross-correlation coefficients between the time-series for each pair of stocks, and create connections between them if the maximum absolute value of the correlation coefficient is among the highest 5% of the total cross correlation coefficients. This yields a time-varying stock market network with a fixed number of 347 vertices and varying edge structure for each of 5976 trading days.

- *Drosophila Melanogaster Gene Network Dataset.* Is extracted from DNA microarrays that contain the transcriptional profiles for nearly one-third of all predicted fruit fly (Drosophila melanogaster) genes through the complete life cycle, from fertilization to adult. The data is sampled at 66 sequential developmental time points. The fruit fly life cycle is divided into four stages, namely the embryonic (samples 1-30), larval (samples 31-40) and pupal (samples 41-58) periods together with the first 30 days of adulthood (samples 59-66). Early embryos are sampled hourly and adults are sampled at multiday intervals according to the speed of the morphological changes. At each time point, by comparing each experimental sample to a reference pooled mRNA sample, the relative abundance of each transcript can be measured, which can further be used as a gene's expression level [9]. To represent this gene expression measurements data using a time-evolving network, the following steps are followed [119]. At each developmental point the 588 genes that are known to play an important role in the development of the Drosophila are selected. These genes are the vertices of the network. The edges are established based on the distribution of the gene expression values, which can be modeled as a binary pair-wise Markov Random Field (MRF) whose parameter indicates the strength of undirected interactions between two genes. In other words, two genes are connected when their model parameter exceeds a threshold. This dataset thus yields a time-evolving Drosophila gene-regulatory network with a fixed number of 588 vertices, sampled at 66 developmental time points.

3.2 Approximate Von Neumann Entropy for Directed Graphs

In this section, we develop a novel entropy measure for assessing the structural complexity of directed graphs. Although there are many existing alternative measures for quantifying the structural properties of undirected graphs, there are relatively few corresponding measures for directed graphs. To fill this gap in the literature, we explore an alternative technique that is applicable to directed graphs. We commence by using Chung's generalization of the Laplacian of a directed graph to extend the computation of von Neumann entropy from undirected to directed graphs. We provide a simplified form of the entropy

which can be expressed in terms of simple vertex in-degree and out-degree statistics. More-over, we find approximate forms of the von Neumann entropy that apply to both weakly and strongly directed graphs, and that can be used to characterize network structure.

One natural way of capturing the structural complexity of directed graphs is to use simple statistics that quantify the balance of in-degree and out-degree at different vertices. A similar but largely heuristic approach has been used to characterize undirected graphs in terms of vertex degree. In fact Han et al.'s work [59] puts this work on a firmer footing by showing how simple vertex degree statistics can be used to approximate the von Neumann entropy for undirected graphs. This is a natural step since in information theory, entropy is a measure of unpredictability or information content in a random variable [66]. By extending this definition to graphs we arrive at a natural way of characterizing their structural complexity. In particular, we can use ideas related to random walks on directed graphs to compute their entropy, and these lead naturally to a characterization in terms of vertex in-degree and out-degree statistics.

Our work commences from Passerini and Severini's postulate [99] that the combinato-rial Laplacian can be scaled by the sum of vertex degrees in the graph, and the resulting matrix interpreted as the scaled density matrix of an undirected graph. As a result, it is possible to compute the von Neumann entropy of a graph from the eigenvalues of the as-sociated combinatorial Laplacian. We extend this work to directed graphs, using Chung's definition of the normalized Laplacian of a directed graph [31]. According to this defini-tion, the directed normalized Laplacian matrix is Hermitian, so Passerini and Severini's density matrix interpretation still holds in the domain of directed graphs. Furthermore, the directed graph von Neumann entropy is essentially the Shannon entropy associated with the normalized Laplacian eigenvalues. Following Han et al. [59] we again approx-imate the Shannon entropy by its quadratic counterpart, with the result that the von Neumann entropy can be simplified in terms of simple in-degree and out-degree statistics. Specifically, the resulting entropy expression depends on the in-degree and out-degree of pairs of vertices connected by edges. To further simplify this expression, we consider graphs that are either weakly or strongly directed, i.e., those in which there are large or small proportions of bidirectional edges, and develop corresponding approximations of the von Neumann entropy. The approximations accord with our physical intuition concerning in-degree and out-degree on vertices and connecting edges.

3.2.1 Von Neumann Entropy of Undirected Graphs

In this section, our aim is to propose a novel entropy measure for characterizing the complexity of directed graphs. We commence from the von Neumann entropy development for undirected graphs. Then, we extend the development to the domain of directed graphs by making use of Chung's definition of the Laplacian for directed graphs. This leads to an expression for the von Neumann entropy in terms of the in-degree and out-degree statistics of vertices. We then provide approximations of the von Neumann entropy for both strongly directed graphs where there are few bidirectional edges and weakly directed graphs where there are few edges that are unidirectional.

Passerini and Severini [99] have argued that the combinatorial Laplacian can be interpreted as the density matrix of an undirected graph. Therefore, it is possible to define the von Neumann entropy of a graph and calculate it from the eigenvalues of the associated combinatorial Laplacian. In order to gain new insights about the meaning of the von Neumann entropy of a graph, we now show how to obtain a simplified expression for this entropy that can be written in terms of the degrees of the vertices. We commence by summarizing the approximation of the undirected graph von Neumann entropy presented by Han et al. [59], and then develop this further to illustrate the limitations of the approximations used.

Although Passerini and Severini have used the traditional Laplacian in their calculations, in order to simplify matters we use the *normalized Laplacian matrix* $\tilde{L} = D^{-1/2}LD^{-1/2}$ here. In our analysis the choice of normalization is not an important detail since both Laplacian and normalized Laplacian matrices make valid density matrices. Furthermore, the scaling of the eigenvalues does not affect the functional dependence of the entropy with the degree. In particular, the largest eigenvalue of the Laplacian matrix is bounded by twice the largest vertex degree in a graph, while the normalized Laplacian matrix has eigenvalues between 0 and 2. With this choice of density matrix, the von Neumann entropy of the undirected graph is the Shannon entropy associated with the normalized Laplacian eigenvalues, i.e.,

$$H_{VN}^U = -\sum_{i=1}^{|\mathcal{V}|} \frac{\tilde{\lambda}_i}{|\mathcal{V}|} \ln \frac{\tilde{\lambda}_i}{|\mathcal{V}|}, \qquad (3.1)$$

where $\tilde{\lambda}_i, i = 1, \cdots, |\mathcal{V}|$ are the eigenvalues of the normalized Laplacian matrix $\tilde{L}$.

Commencing from this definition and making use of the quadratic approximation to the Shannon entropy (i.e., $-x \ln x \approx x(1-x)$, which holds well when x is close to 0 or 1),

Han et al. [59] approximate the von Neumann entropy by

$$H_Q^U = \sum_{i=1}^{|\mathcal{V}|} \frac{\tilde{\lambda}_i}{|\mathcal{V}|}(1 - \frac{\tilde{\lambda}_i}{|\mathcal{V}|}).$$

For undirected graphs this quadratic approximation allows the von Neumann entropy to be expressed in terms of the trace of the normalized Laplacian (which is equal to the sum of the normalized Laplacian eigenvalues) and the trace of the squared normalized Laplacian (which is equal to the sum of the squares of the normalized Laplacian eigenvalues), with the result that

$$H_{VN}^U = \frac{Tr[\tilde{L}]}{|\mathcal{V}|} - \frac{Tr[\tilde{L}^2]}{|\mathcal{V}|^2}. \tag{3.2}$$

For undirected graphs, the two traces appearing in the above expression are given in terms of statistics for the degree of vertices in the graph [59], with the result that

$$H_{VN}^U = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \frac{1}{d_u d_v}. \tag{3.3}$$

This formula contains two measures of graph structure, the first one is the number of vertices of graph, while the second one is based on degree statistics for pairs of vertices connected by edges. Moreover, the computational complexity of this expression is quadratic in graph size, which is much simpler than that of the original entropy.

The accuracy of the above expression depends on the veracity of the quadratic approximation to the Shannon entropy $-x \ln x \approx x(1 - x)$. This approximation is known to hold well when either $x \to 0$ or $x \to 1$, which guarantees the accuracy of the quadratic entropy since $\frac{\tilde{\lambda}_i}{|\mathcal{V}|} \to 0$ when the graph size is very large.

A more precise expression for the von Neumann entropy can be obtained by making a second-order Taylor series approximation for the Shannon entropy with expansion point x_0 at the mean value of $\frac{\tilde{\lambda}}{|\mathcal{V}|}$, i.e.,

$$x_0 = \frac{\sum_{i=1}^{|\mathcal{V}|} \frac{\tilde{\lambda}_i}{|\mathcal{V}|}}{|\mathcal{V}|} = \frac{Tr[\tilde{L}]}{|\mathcal{V}|^2}.$$

The second-order Taylor expansion for $x \ln x$ about the expansion point x_0 is

$$x \ln x \approx -x(-\ln x_0 - \frac{x}{2x_0}) - \frac{x_0}{2}.$$

Substituting this series approximation for the Shannon entropy with expansion point

$$x_0 = \frac{Tr[\tilde{L}]}{|\mathcal{V}|^2} = \frac{1}{|\mathcal{V}|}$$

53

into the expression for the von Neumann entropy Eq. (3.1), we obtain

$$H_T^U = \ln |\mathcal{V}| - \frac{1}{2|\mathcal{V}|} \sum_{(u,v)\in\mathcal{E}} \frac{1}{d_u d_v}. \tag{3.4}$$

As a result, the Taylor series approximation to the von Neumann entropy at the expansion point $x_0 = \frac{1}{|\mathcal{V}|}$ and the quadratic approximation are related by

$$H_T^U = \frac{|\mathcal{V}|}{2} H_Q^U + \ln |\mathcal{V}| + \frac{1 - |\mathcal{V}|}{2}.$$

In other words, the two entropies are related by an offset and a scale, which are related to the number of vertices in the graph. Since we are concerned in applying the von Neumann entropy for characterizing the structure of graphs, the differences caused by the influence of graph size do not matter in our analysis. Therefore, both expressions can be used. Throughout the chapter we use the simpler expression given by H_Q^U.

3.2.2 Normalized Laplacian Matrix of Directed Graphs

The transition matrix P of a graph is a matrix describing the transitions of a Markov chain on the graph. On a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, P is given as

$$P_{uv} = \begin{cases} \frac{1}{d_u^{out}} & \text{if } (u,v) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

It is interesting to note that for a strongly connected graph, the transition matrix P is column-stochastic. Moreover, according to the Perron-Frobenius theorem, on a strongly connected graph, P has a unique left eigenvector ϕ with $\phi(u) > 0$, for all $u \in \mathcal{V}$, which satisfies $\phi P = \rho \phi$ where ρ denotes the eigenvalue of P. The theorem also implies that if P is aperiodic, the eigenvalues of P have absolute values smaller than the leading eigenvalue $\rho = 1$. Thus any random walk on a directed graph will converge to a unique stationary distribution if the graph satisfies the properties of strong connection and aperiodicity. We normalize ϕ such that $\sum_{i=1}^{|\mathcal{V}|} \phi(i) = 1$, this normalized vector corresponds to the unique stationary distribution. Therefore, the probability of a random walker being at vertex u is the sum of all incoming probabilities of vertices v satisfying $(v,u) \in \mathcal{E}$, i.e., $\phi(u) = \sum_{v,(v,u)\in\mathcal{E}} \phi(v) P_{vu}$. Based on the properties of the random walk on a directed graph, we assume that the eigenvector component $\phi(u)$ is proportional to the in-degree of the corresponding vertex d_u^{in}, i.e.,

$$\frac{\phi(u)}{\phi(v)} \approx \frac{d_u^{in}}{d_v^{in}} \tag{3.5}$$

From this, we derive

$$
\begin{aligned}
\frac{\phi(u)}{d_u^{in}} &\approx \frac{\phi(v)}{d_v^{in}} \\
&= \frac{\phi(1) + \phi(2) + \cdots + \phi(|\mathcal{V}|)}{d_1^{in} + d_2^{in} + \cdots + d_{|\mathcal{V}|}^{in}} \\
&= \frac{1}{vol(\mathcal{G})},
\end{aligned}
$$

where $vol(\mathcal{G})$ is the volume of the graph, defined as the sum of all vertex in-degree or out-degree. To illustrate the plausibility of the above assumption, we note that

$$
\begin{aligned}
\phi(u) &= \sum_{v,(v,u)\in\mathcal{E}} \phi(v) P_{vu} \\
&= \sum_{v,(v,u)\in\mathcal{E}} \frac{d_v^{in}}{vol(\mathcal{G})} \cdot \frac{A_{vu}}{d_v^{out}} \\
&= \frac{1}{vol(\mathcal{G})} \sum_{v,(v,u)\in\mathcal{E}} \frac{d_v^{in}}{d_v^{out}} \\
&= \frac{d_u^{in}}{vol(\mathcal{G})} \left\langle \frac{d_v^{in}}{d_v^{out}} \right\rangle_{v,(v,u)\in\mathcal{E}}.
\end{aligned}
$$

This implies that the approximation in Eq. (3.5) holds only when the neighbourhood of vertex u has similar out-degree and in-degree. Although this condition may seem to be a strong requirement, we will undertake experiments to analyze how the local average vertex degree ratio

$$
r_u = \left\langle \frac{d_v^{in}}{d_v^{out}} \right\rangle_{v,(v,u)\in\mathcal{E}} \tag{3.6}
$$

of u affects the accuracy of our suggested approximate von Neumann entropy (provided later) and the result reveals that this ratio indeed does not cause a significant error.

As stated in Chung [31], if we let $\Phi = diag(\phi(1), \phi(2), \cdots)$, then the normalized Laplacian matrix of a directed graph can be defined as

$$
\tilde{L} = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^T \Phi^{1/2}}{2}, \tag{3.7}
$$

where I is the identify matrix.

Clearly, the normalized Laplacian matrix is Hermitian, i.e., $\tilde{L} = \tilde{L}^T$ where $\tilde{L}^T$ denotes the conjugated transpose of $\tilde{L}$.

3.2.3 Von Neumann Entropy of Directed Graphs

Following the development of the von Neumann entropy of undirected graphs, the true von Neumann entropy for a directed graph can be computed using the Shannon entropy

55

associated with the eigenvalues of its normalized Laplacian matrix. Unfortunately, for large graphs this is not a viable proposition since the time required to solve the eigensystem is cubic in the number of vertices. To overcome this problem we extend the analysis of Han et al. [59] from undirected to directed graphs. To do this we again make use of the quadratic approximation to the Shannon entropy in order to obtain a simplified expression for the von Neumann entropy of a directed graph, which can be computed in a time that is quadratic in the number of vertices. Our starting point is the quadratic approximation to the von Neumann entropy in terms of the traces of normalized Laplacian and the squared normalized Laplacian, which is given in Eq. (3.2)

$$H_{TVN}^D = \frac{Tr[\tilde{L}]}{|\mathcal{V}|} - \frac{Tr[\tilde{L}^2]}{|\mathcal{V}|^2}. \tag{3.8}$$

To simplify this expression a step further, we repeat the computation of the traces for the case of a directed graph. This is not a straightforward task, and requires that we distinguish between the in-degree and out-degree of vertices. We first consider Chung's expression for the normalized Laplacian of directed graphs and write

$$
\begin{aligned}
Tr[\tilde{L}] &= Tr[I - \frac{\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}}{2}] \\
&= Tr[I] - \frac{1}{2}Tr[\Phi^{1/2}P\Phi^{-1/2}] - \frac{1}{2}Tr[\Phi^{-1/2}P^T\Phi^{1/2}].
\end{aligned}
$$

Since the matrix trace is invariant under cyclic permutations, we have

$$
\begin{aligned}
Tr[\tilde{L}] &= Tr[I] - \frac{1}{2}Tr[P\Phi^{-1/2}\Phi^{1/2}] - \frac{1}{2}Tr[P^T\Phi^{1/2}\Phi^{-1/2}] \\
&= Tr[I] - \frac{1}{2}Tr[P] - \frac{1}{2}Tr[P^T].
\end{aligned}
$$

The diagonal elements of the transition matrix P are all zeros, hence we obtain

$$Tr[\tilde{L}] = Tr[I] = |\mathcal{V}|,$$

which is exactly the same as in the case of undirected graphs.

Next we turn our attention to $Tr[\tilde{L}^2]$:

$$
\begin{aligned}
Tr[\tilde{L}^2] &= Tr[I^2 - (\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}) + \frac{1}{4}(\Phi^{1/2}P\Phi^{-1/2}\Phi^{1/2}P\Phi^{-1/2} \\
&\quad + \Phi^{1/2}P\Phi^{-1/2}\Phi^{-1/2}P^T\Phi^{1/2} + \Phi^{-1/2}P^T\Phi^{1/2}\Phi^{1/2}P\Phi^{-1/2} \\
&\quad + \Phi^{-1/2}P^T\Phi^{1/2}\Phi^{-1/2}P^T\Phi^{1/2})] \\
&= Tr[I^2] - Tr[P] - Tr[P^T] + \frac{1}{4}(Tr[P^2] + Tr[P\Phi^{-1}P^T\Phi] \\
&\quad + Tr[P^T\Phi P\Phi^{-1}] + Tr[P^{T^2}]) \\
&= |\mathcal{V}| + \frac{1}{2}(Tr[P^2] + Tr[P\Phi^{-1}P^T\Phi]),
\end{aligned}
$$

which is different to the result obtained in the case of undirected graphs.

To continue the development we first partition the edge set $\mathcal{E}$ of the graph $\mathcal{G}$ into two disjoint subsets $\mathcal{E}_1$ and $\mathcal{E}_2$, where $\mathcal{E}_1 = \{(u,v)|(u,v) \in \mathcal{E} \bigwedge (v,u) \notin \mathcal{E}\}$, $\mathcal{E}_2 = \{(u,v)|(u,v) \in \mathcal{E} \bigwedge (v,u) \in \mathcal{E}\}$ that satisfy the conditions $\mathcal{E}_1 \bigcup \mathcal{E}_2 = \mathcal{E}$, $\mathcal{E}_1 \bigcap \mathcal{E}_2 = \emptyset$. Then according to the definition of the transition matrix, we find

$$Tr[P^2] = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} P_{uv} P_{vu} = \sum_{(u,v) \in \mathcal{E}_2} \frac{1}{d_u^{out} d_v^{out}}.$$

Using the fact that $\Phi = diag(\phi(1), (2), \cdots)$ we have

$$Tr[P\Phi^{-1}P^T\Phi] = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} P_{uv}^2 \frac{\phi(u)}{\phi(v)} = \sum_{(u,v) \in \mathcal{E}} \frac{\phi(u)}{\phi(v) d_u^{out2}}.$$

Using Eq. (3.5), we can approximate the von Neumann entropy of a directed graph in terms of the in-degree and out-degree of the vertices as follows

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \left\{ \sum_{(u,v) \in \mathcal{E}} \left(\frac{1}{d_u^{out} d_v^{out}} + \frac{d_u^{in}}{d_v^{in} d_u^{out2}} \right) - \sum_{(u,v) \in \mathcal{E}_1} \frac{1}{d_u^{out} d_v^{out}} \right\} \qquad (3.9)$$

or, equivalently,

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \left\{ \sum_{(u,v) \in \mathcal{E}} \frac{d_u^{in}}{d_v^{in} d_u^{out2}} + \sum_{(u,v) \in \mathcal{E}_2} \frac{1}{d_u^{out} d_v^{out}} \right\}. \qquad (3.10)$$

It is interesting to identify the structures that give extremal values of our approximate entropy. By inspection of Eq. (3.9) and Eq. (3.10), when the terms in the curly brackets reach their largest value, the von Neumann entropy takes on its minimum value. This occurs when the structure is a circle graph, in which each vertex has only one outgoing edge and one incoming edge. On the other hand, when the terms in the curly brackets take on their smallest value, the entropy is maximal. This occurs when there are no bidirectional edges in the graph. Vertices that have outgoing edges have no incoming edges. A typical example of this type of structure is a star graph.

However, there are a number of assumptions concerning the admissible structure of the graph which underpin the definition of directed graph Laplacian and hence the derivation of our approximate entropy. Two of these are in conflict with the two conditions giving extremal values of the entropy. The condition that the walks on the graph are aperiodic is not consistent with cycle structure, while the condition that they are strongly connected is not consistent with the star graph case. Nonetheless, even if the assumptions underpinning the formula break down in these two cases, it is empirically interesting that is behaves in this way.

The maximum and minimum von Neumann entropies corresponding to these cases are as follows. For a circle directed graph, all vertices have the same out-degree and in-degree equal to 1, then

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \cdot |\mathcal{V}| = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|}.$$

Turning attention to the case of a star graph, the centre vertex has out-degree (in-degree) $|\mathcal{V}| - 1$, and the remaining vertices have in-degree (out-degree) 1. In this case

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \cdot 0 = 1 - \frac{1}{|\mathcal{V}|}.$$

So if we do not take into account the requirements of the entropy definition on directed graphs, i.e., when we analyze in a quantitative manner, the approximate von Neumann entropy suggested in Eq. (3.10) gives the minimum value for the ring graph, which is the simplest regular graph. It takes on its maximum value for star graphs. The latter structure can be viewed as the most complex, since it has the greatest difference in vertex out-degree and in-degree.

To proceed the development, we can simplify the approximate von Neumann entropy expression according to the relative sizes of the sets $\mathcal{E}_1$ and $\mathcal{E}_2$, to provide approximations to the von Neumann entropy which are specific to weakly and strongly directed graphs.

For weakly directed (WD) graphs, $|\mathcal{E}_1| \ll |\mathcal{E}_2|$, i.e., few of the edges are not bidirectional, we can ignore the summation over $\mathcal{E}_1$ in Eq. (3.9). Re-writing the remaining terms in curly brackets in terms of a common denominator and then dividing numerator and denominator by $d_u^{out} d_v^{out}$ we obtain

$$H_{VN}^{WD} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{\frac{d_u^{in}}{d_u^{out}} + \frac{d_v^{in}}{d_v^{out}}}{d_u^{out} d_v^{in}} \right\}. \tag{3.11}$$

The term $1 - \frac{1}{|\mathcal{V}|}$ tends to unity as the graph size becomes large. In the summation, the numerator is given in terms of the sum of the ratios of in-degree and out-degree of the vertices. Since the directed edges cannot start at a sink (a vertex of zero out-degree), the ratios do not become infinite.

On the other hand, for strongly directed (SD) graphs, there are few bidirectional edges, i.e., $|\mathcal{E}_2| \ll |\mathcal{E}_1|$, and we can ignore the summation over $\mathcal{E}_2$ in Eq. (3.10), giving the approximate entropy for strongly directed graphs

$$H_{VN}^{SD} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{d_u^{in}}{d_v^{in} d_u^{out^2}} \right\}. \tag{3.12}$$

Both the weakly and strongly directed forms of the von Neumann entropy H_{VN}^{WD} and H_{VN}^{SD} contain two terms. The first is the graph size while the second one depends on the in-degree and out-degree statistics of each pair of vertices connected by an edge. Moreover, the computational complexity of these expressions is quadratic in the graph size.

There are a number of points to note concerning the development above. First, the normalized Laplacian matrix of directed graphs denoted by $\tilde{L}$ in Eq. (3.7) satisfies Passerini and Severini's conditions [99] for the density matrix. Moreover, we have shown that $\tilde{L}$ is Hermitian, so its eigenvalues are all real. Hence theoretically, the density matrix interpretation of Passerini and Severini [99] can be extended to directed graphs. Secondly, when the out-degree and in-degree are the same at all vertices, the von Neumann entropies for directed and undirected graphs are identical.

To conclude this section, it is worth discussing the role of sinks in our analysis. A sink is a vertex with several incident edges, but no outgoing edges. Hence they are characterized by zero out-degree. One obvious problem with our formulation is that our expression for the von Neumann entropy of a weakly directed graph, which is given in Eq. (3.9), will become singular when vertex v is a sink, i.e., $d_v^{out} = 0$. However, in the case of weakly directed graphs, the likelihood of sink vertices is small, since the number of bidirectional edges is large. We can reach the same conclusion by recalling that the graph represents a Markov chain with equal transition probabilities on the vertices. If the chain is irreducible and aperiodic, then the convergence to a stationary distribution is guaranteed. Otherwise, the final distribution may not be stable or may depend on the initial conditions. In particular, if the irreducibility condition is not true, then the Perron-Frobenius theorem does not hold and we cannot construct the Laplacian in that case, or at least it is not clear if the theorems in Chung's paper hold. So if we demand that the Markov chain is irreducible, this means the graph is strongly connected and there are no sinks in the graph.

3.3   Heterogeneity Index for Directed Graphs

In this section we propose a novel measure for quantifying the structural heterogeneity properties of directed graphs. We extend the development of Estrada's heterogeneity index [46] from undirected graphs to directed graphs in order to define an analogous directed heterogeneity measure. We then provide a normalized form of the directed heterogeneity index which can be expressed in terms of simple vertex in-degree and out-degree statistics. We show this measure has the lower bound 0 and the upper bound 1, for regular graphs

and star graphs respectively.

To commence, we follow Estrada's work [46] and establish a local assessment to quantify the relative irregularity associated with a single edge in a directed graph. Specifically, on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, Estrada uses the following quantity to measure the variation in degrees at vertices connected by an edge:

$$\omega_{uv}^U = [f(d_u) - f(d_v)]^2,$$

where $f(d)$ is a function of the vertex degree d. To extend this evaluation to directed graphs, we assess the difference in the out-degree of the starting vertex and the in-degree of the end vertex of a directed edge (u, v) and write

$$\omega_{uv}^D = [f(d_u^{out}) - f(d_v^{in})]^2.$$

This local degree heterogeneity measure takes on a value zero when the out-degree of the starting vertex is the same as the in-degree of the end vertex. On the other hand, this local measure should increase as the difference between the two degrees increases. Thus we can select $f(d) = d^{-1/2}$. This is mainly because this function allows us to measure the degree difference in both quantity and magnitude. For instance, suppose (u_1, v_1) and (u_2, v_2) are two edges in the directed graph and we have $d_{u_1}^{out} = 1$, $d_{v_1}^{in} = 11$, $d_{u_2}^{out} = 90$, $d_{v_2}^{in} = 100$. The simple vertex degree difference cannot distinguish between these two edges as $d_{v_1}^{in} - d_{u_1}^{out} = d_{v_2}^{in} - d_{u_2}^{out} = 10$. However, using $f(d) = d^{-1/2}$ we obtain $\omega_{u_1 v_1}^D = 0.488$ while $\omega_{u_2 v_2}^D = 2.93 \cdot 10^{-5}$, which indicates that (u_1, v_1) has relatively greater edge irregularity than (u_2, v_2).

Then the local heterogeneity assessment associated with the irregularity of the edge (u, v) in a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is given by

$$\omega_{uv}^D = \left\{ \frac{1}{\sqrt{d_u^{out}}} - \frac{1}{\sqrt{d_v^{in}}} \right\}^2.$$

To compute the global heterogeneity measure of a directed graph we sum the local measure over all the edges in the graph to obtain

$$
\begin{aligned}
\Omega^D &= \sum_{(u,v) \in \mathcal{E}} \left\{ \frac{1}{\sqrt{d_u^{out}}} - \frac{1}{\sqrt{d_v^{in}}} \right\}^2 \\
&= \sum_{(u,v) \in \mathcal{E}} \left\{ \frac{1}{d_u^{out}} + \frac{1}{d_v^{in}} \right\} - 2 \sum_{(u,v) \in \mathcal{E}} \frac{1}{\sqrt{d_u^{out} d_v^{in}}}.
\end{aligned}
\tag{3.13}
$$

Clearly, the directed heterogeneity index should take on the minimum value when the directed graph is regular, i.e., all the vertices have the same in-degree and out-degree. In

contrast, the index is maximal when the graph is a star graph, i.e., there exists a central vertex such that all the other vertices connect and only connect to it. We calculate the lower and upper bounds of Ω^D according to these constraints.

For a regular directed graph, suppose all the vertices have the same in-degree and out-degree d_0, then we have

$$\Omega^D = \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{1}{d_0} + \frac{1}{d_0} \right\} - 2 \sum_{(u,v)\in\mathcal{E}} \frac{1}{d_0} = 0.$$

On the other hand, for a star graph, the central vertex has out-degree (in-degree) $|\mathcal{V}| - 1$ and all the other vertices have in-degree (out-degree) 1. Then,

$$
\begin{aligned}
\Omega^D &= \sum_{i=1}^{|\mathcal{V}|} \left(\frac{1}{|\mathcal{V}| - 1} + 1 \right) - 2 \sum_{i=1}^{|\mathcal{V}|} \frac{1}{\sqrt{|\mathcal{V}| - 1}} \\
&= \frac{|\mathcal{V}|(|\mathcal{V}| - 2\sqrt{|\mathcal{V}| - 1})}{|\mathcal{V}| - 1} \\
&\approx |\mathcal{V}| - 2\sqrt{|\mathcal{V}| - 1},
\end{aligned}
$$

when the graph size $|\mathcal{V}|$ is large enough.

We hence have the following lower and upper bounds for the directed heterogeneity index

$$0 \leq \Omega^D = \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{1}{d_u^{out}} + \frac{1}{d_v^{in}} - \frac{2}{\sqrt{d_u^{out} d_v^{in}}} \right\} \leq |\mathcal{V}| - 2\sqrt{|\mathcal{V}| - 1}.$$

Based on the above analysis we define the following normalized heterogeneity index

$$\tilde{\Omega}^D = \frac{\Omega^D}{|\mathcal{V}| - 2\sqrt{|\mathcal{V}| - 1}} = \frac{\sum_{(u,v)\in\mathcal{E}} \left\{ \frac{1}{d_u^{out}} + \frac{1}{d_v^{in}} - \frac{2}{\sqrt{d_u^{out} d_v^{in}}} \right\}}{|\mathcal{V}| - 2\sqrt{|\mathcal{V}| - 1}}. \tag{3.14}$$

This index is defined over the interval $[0, 1]$ and is zero for regular directed graphs, one for star graphs. Clearly, the normalized index depends on two terms, the first one is the graph size while the second one is based on the statistics of vertex in-degree and out-degree. To summarize, the directed heterogeneity index captures the structural irregularity of each edge in a directed graph by measuring the relative difference of degrees at both the starting vertex and the end vertex. Moreover it accounts for the effect of graph size, so in principle it provides an efficient means of quantifying a directed graph's structural complexity.

To conclude this section, it is worth discussing the choice of using simple degree pair d_u^{out} and d_v^{in} to calculate the relative irregularity associated with a single directed edge (u, v). Recall the approximate von Neumann entropy of the strongly directed graphs,

which is given in Eq. (3.12):

$$H_{VN}^{SD} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \sum_{(u,v) \in \mathcal{E}} \left\{ \frac{d_u^{in}}{d_v^{in} d_u^{out2}} \right\}.$$

This entropy measure contains the term $d_u^{out} d_v^{in}$ as an invariant in the denominator, since this combination of degrees is of importance in representing the structural characteristics of the associated edge in a directed graph. Moreover, in some directed graphs, there may exist edges that end at sink vertices (with no outgoing edges) and edges that start from source vertices (with no incoming edges), as a result, such a combination of vertex degrees becomes the only choice that can be employed to compute the degree-based irregularity of these edges.

3.4 Experiments

We have derived an expression for the von Neumann entropy of a directed graph, and have provided approximations that apply to both weakly and strongly directed graphs. We have also developed a novel version of Estrada's heterogeneity index to measure the heterogeneous characteristics of directed graphs quantitatively. In this section, we explore whether these graph complexity measures can be used to characterize different directed graph structure patterns and determine changes in the structure of dynamic directed graphs.

3.4.1 Directed Von Neumann Entropy

We confine our attention to two principal tasks. The first one is to explore whether the complexity measures can be used to distinguish different types of directed graphs. The second is to use the complexity measures to detect abrupt changes in the structure of networks that evolve with time. For most experiments, we normalize the entropy measures studied, including the approximate von Neumann entropy in Eq. (3.10) together with its approximations for both weakly and strongly directed graphs given in Eq. (3.11) and Eq. (3.12) respectively. The normalization is with respect to the graph size, and this removes some of the size dependence. Specifically, we compute the quantity

$$
\begin{aligned}
Y_{VN}^D &= |\mathcal{V}| \cdot \left| H_{VN}^D - (1 - \frac{1}{|\mathcal{V}|}) \right| \\
&= \frac{1}{2|\mathcal{V}|} \left\{ \sum_{(u,v) \in \mathcal{E}} \frac{d_u^{in}}{d_v^{in} d_u^{out2}} + \sum_{(u,v) \in \mathcal{E}_2} \frac{1}{d_u^{out} d_v^{out}} \right\}
\end{aligned}
\tag{3.15}
$$

as a normalized quantity which captures variations in the in-degree and out-degree statistics in the same manner as the approximate von Neumann entropy H_{VN}^D. Similarly, the corresponding normalized quantities for the weakly and strongly directed approximations H_{VN}^{WD} and H_{VN}^{SD} are given as

$$Y_{VN}^{WD} = \frac{1}{2|\mathcal{V}|} \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{\frac{d_u^{in}}{d_u^{out}} + \frac{d_v^{in}}{d_v^{out}}}{d_u^{out} d_v^{in}} \right\} \tag{3.16}$$

and

$$Y_{VN}^{SD} = \frac{1}{2|\mathcal{V}|} \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{d_u^{in}}{d_v^{in} d_u^{out2}} \right\}. \tag{3.17}$$

It is important to note that these normalized quantities and the original entropy measures have opposite monotonicity properties. In other words when the normalized entropy Y_{VN}^D decreases, the approximate von Neumann entropy H_{VN}^D increases.

In this section we use four different datasets, namely *Preferential Attachment Network Dataset*, *Random Directed Graph Dataset*, *Protein Dataset* and *Citation Network Dataset*. The first two datasets are synthetically generated artificial networks, while the other two are extracted from real-world systems.

An important point to note concerning these datasets is that in the development of the directed von Neumann entropy, to keep our development simple and straightforward, we require the directed graph under study is strongly connected, but here the graphs used for experiments do not always guarantee the strong connectivity, which implies that the graphs may have more than one strongly connected components. To compute the von Neumann entropy for such graphs, we follow the method used by Bai and Hancock [12]. According to their work, suppose a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of two connected components $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$. Then, the entropy (denoted by H here) of $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is computed as the entropy of the disjoint union graph of $\mathcal{G}_1$ and $\mathcal{G}_2$, i.e.,

$$H(\mathcal{G}) = H(\mathcal{G}_{DU}) = \pi_1 H(\mathcal{G}_1) + \pi_2 H(\mathcal{G}_2),$$

where $\mathcal{G}_{DU} = \mathcal{G}_1 \cup \mathcal{G}_2 = \{\mathcal{V}_1 \cup \mathcal{V}_2, \mathcal{E}_1 \cup \mathcal{E}_2\}$. Moreover, π_1 and π_2 indicate the entropy weights which are determined by the size of the connected components, i.e., $\pi_1 = \frac{|\mathcal{V}_1|}{|\mathcal{V}_1|+|\mathcal{V}_2|}$ and $\pi_2 = \frac{|\mathcal{V}_2|}{|\mathcal{V}_1|+|\mathcal{V}_2|}$ respectively. Following this idea, the normalized von Neumann entropy for a directed graph can be simply computed as the sum of the entropy for each strongly connected component. The weight of each component is not needed as the quantity Y_{VN}^D already takes into account the graph size. In other words, by using this strategy, our

suggested approximate von Neumann entropy can also apply to directed graphs that are not strongly connected.

An alternative way to compute the entropy of a graph consisting of connected components would be first construct a product graph of the components and then compute the entropy of the resulting product graph. Unfortunately, constructing a product graph is computationally burdensome. Furthermore, the number of vertices for the product graph is significantly greater than that of the disjoint union graph. So in order to simplify matters, in our analysis we will use the former entropy computing method.

It is also worth noting that in the *Citation Network Dataset*, citation networks do not contain bidirectional edges (a paper cannot cite any paper that has not yet been written). As a result they are strongly directed graphs that contain a number of sink vertices. According to our previous analysis, these sink vertices may lead to situations where the directed graph von Neumann entropy is not well defined. However, from the strongly directed von Neumann entropy approximation obtained in Eq. (3.12), we find for a directed edge (u, v) the denominator term $d_v^{in} d_u^{out^2}$ is only related to the out-degree of the starting vertex u and the in-degree of the end vertex v. This means that the sink vertices do not make the expression singular, so the strongly directed von Neumann entropy approximation can still be computed in a valid manner on citation networks.

We investigate the difference between the previously defined undirected graph von Neumann entropy and its directed analogue in order to analyze how these entropies correlate. To do this, we select the directed graphs in the *Preferential Attachment Network Dataset* and compute their normalized entropies using Eq. (3.15), we then drop all the edge orientations to make the graphs undirected and compute their corresponding entropies using the following normalized quantity

$$Y_{VN}^{U} = \frac{1}{|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \frac{1}{d_u d_v}. \tag{3.18}$$

Figure 3.1 shows the mean of the normalized entropies and their difference versus graph size for both directed and the corresponding undirected graphs. The main feature to note is that as time evolves, the difference between the two normalized entropies maintains small, which suggests that the directed and undirected graph von Neumann entropies have consistence on graphs that are well correlated. From the plot, at some particular time, the directed entropy (blue solid line) fluctuates significantly while the corresponding undirected one (red dotted line) does not, as a result, the difference between them becomes
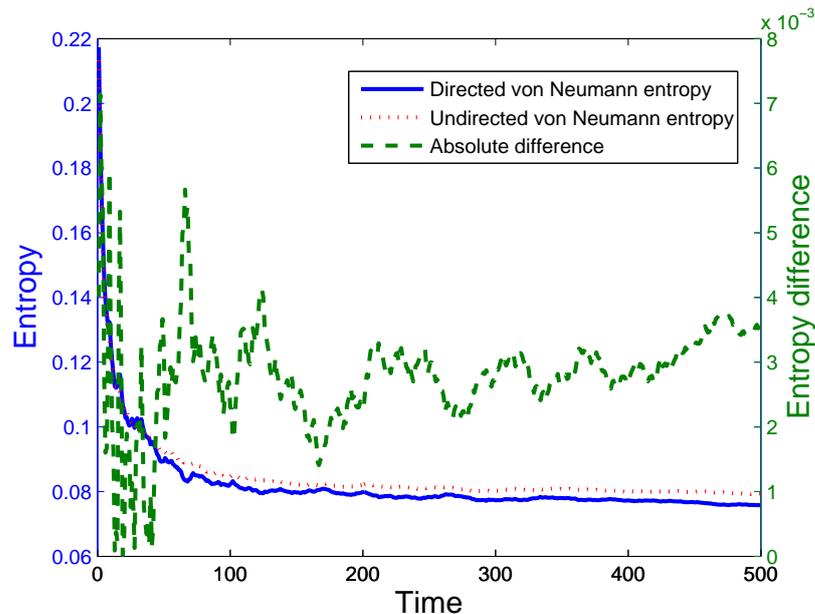
Figure 3.1: Comparing approximate von Neumann entropies for undirected and directed graphs.

Datasets	Wiki-Vote	p2p-G05	p2p-G06	p2p-G08	p2p-G09	Arxiv HEP-TH
Graph size	8297	8846	8717	6301	8114	27751
Min. ratio	0.0213	0.0303	0.0177	0.0208	0.0182	0.0035
Max. ratio	24.7500	22.4583	12.0000	9.2222	19.7000	46.6667
Average ratio	0.1984	0.4663	0.4501	0.4335	0.4196	0.4874
True entropy	0.0128	0.0343	0.0361	0.0374	0.0357	0.0430
Approx. entropy	0.0142	0.0489	0.0412	0.0387	0.0419	0.0585

Table 3.1: Average vertex degree ratio and relative error for real-world network data.

particularly large. This implies that by dropping the edge directions, the undirected graph obtained loses some of the structural information residing in the directed graph, and the undirected graph von Neumann entropy also fails to capture such structure characteristics.

In the previous development we made use of the assumption that the local average vertex degree ratio r_u, which is computed using Eq. (3.6), is close to unity in order to develop our approximate expression for the von Neumann entropy. In order to explore whether this assumption is empirically valid, we explore the dependence of the approximate entropy on the average value of the vertex degree ratio. To this end we compute the average

of the local degree ratios over all vertices in a graph, i.e.,

$$\bar{r} = \frac{1}{|\mathcal{V}|} \sum_{u \in \mathcal{V}} r_u, \tag{3.19}$$

where r_u is the degree ratio for vertex u. We investigate empirically how this global ratio affects the accuracy of the approximate von Neumann entropy.

We commence by studying some real-world directed networks and compare their normalized approximate von Neumann entropies Eq. (3.15) with normalized true entropies, which are computed using the formula

$$Y_{TVN}^D = |\mathcal{V}| \cdot \left| H_{TVN}^D - (1 - \frac{1}{|\mathcal{V}|}) \right|. \tag{3.20}$$

The networks under study include the Wikipedia vote network [79], the Gnutella peer-to-peer networks from August 5 to 9, 2002, which are a sequence of snapshots of the Gnutella peer-to-peer file sharing network [110] and the Arxiv HEP-TH citation network in the *Citation Network Dataset*. Table 3.1 gives the network size, the minimum and maximum values of the local degree ratio, and the average degree ratio. The table also lists the values of both the true entropy and approximate entropy. For each network studied, the average vertex degree ratio is always between 0 and 1, although locally the degree ratio differs significantly. The main feature to note is that the difference between the true and approximate entropy is relatively small, even though we are dealing with large networks.

To take this analysis a step further, we use the random directed graphs in the *Random Directed Graph Dataset*. These graphs are generated according to three different models, and we use them to investigate the degree to which the approximate entropy deviates from the true value for different types of structure. We generate 1000 graphs for each model. For each graph we compute the relative error in the normalized approximate entropy, i.e., $|Y_{VN}^D - Y_{TVN}^D|/Y_{TVN}^D$. We then calculate the mean and standard deviation of the relative error, and explore the dependence on the global vertex degree ratio defined in Eq. (3.19).

Figure 3.2 shows the mean and standard deviation (standard deviation shown as an error bar) of the relative error as a function of the global vertex degree ratio. The statistics needed for this plot are accumulated over graphs whose average vertex degree ratio falls into a fixed interval. From the plot, the relative error is negligible (less than 0.2%) for graphs with global vertex degree ratios ranging between 0.4 and 1.1. Moreover, it takes on its minimum value when the ratio is equal to unity. This is as expected since
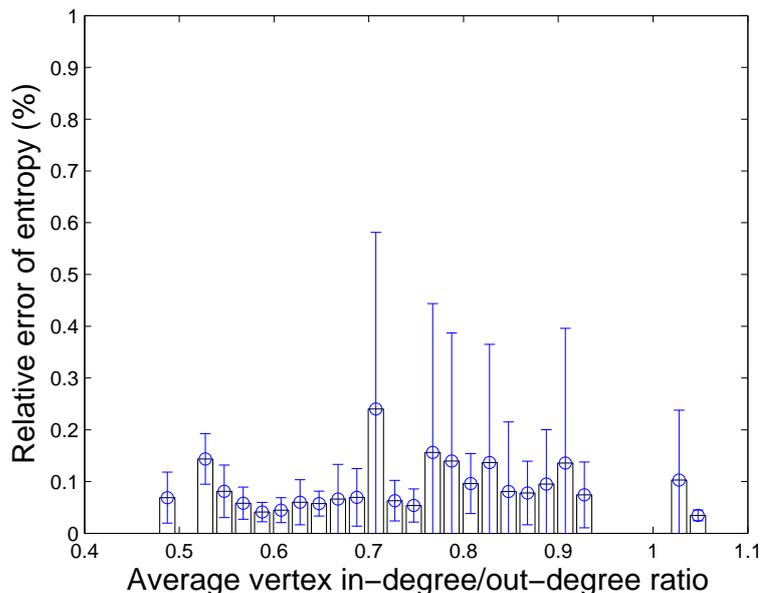
Figure 3.2: Mean and standard deviation of relative error of normalized entropy for graphs with different global vertex in-degree/out-degree ratios.

our development of the approximate von Neumann entropy expression is based on the assumption given in Eq. (3.5), which shows that local vertices have the similar in-degree and out-degree. Therefore the experimental results demonstrate that the approximate von Neumann entropy does not deviate too far from the true value even when the global vertex degree ratio is not close to unity and thus our assumption appears empirically valid.

Then, we use the *Preferential Attachment Network Dataset* to examine the accuracy of the approximations of the entropy for weakly and strongly directed graphs. In other words, we verify that the simplified expressions approximate well the true values of von Neumann entropy. In fact, the evolving directed graphs in the *Preferential Attachment Network Dataset* are strongly directed as the number of unidirectional edges is significantly greater than that of bidirectional edges. To obtain weakly directed graphs, we choose a large number of pairs of vertices that are connected by unidirectional edges in these strongly directed graphs, and change the unidirectional connections to bidirectional ones.

In Figs. 3.3(a) and 3.3(b) we show the mean of the normalized entropies versus graph size for the directed graphs in the *Preferential Attachment Network Dataset*. Here we have computed the approximate entropies for weakly and strongly directed graphs Y_{VN}^{WD} and Y_{VN}^{SD} using Eq. (3.16) and Eq. (3.17) respectively. We compare their values with the normalized approximate entropy Y_{VN}^{D} given in Eq. (3.15) and the normalized true entropy
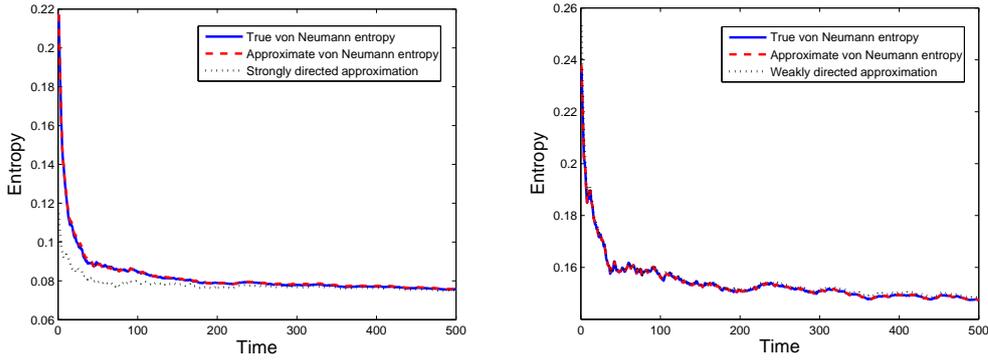
(a) Comparing the weakly directed approxima-
tion, the normalized approximate entropy and
the normalized true entropy

(b) Comparing the strongly directed approx-
imation, the normalized approximate entropy
and the normalized true entropy

Figure 3.3: Approximations to the von Neumann entropy.

Y_{TVN}^D defined in Eq. (3.20).

From both plots, as the network evolves, all these quantities decrease gradually to a value close to zero, which implies that the true von Neumann entropy and its approximations increase monotonically until a plateaux value of unity is reached. It is also worth noting that the difference between these entropies is negligible, thus we can deduce that the approximate von Neumann entropy we suggested (red dashed line) approximates the true von Neumann entropy (blue solid line) very well.

Figure 3.4 shows scatter plots of the weakly and strongly directed approximations Y_{VN}^{WD} and Y_{VN}^{SD} versus the approximate entropy Y_{VN}^D for sets of weakly directed and strongly directed graphs. We select the relevant sets of graphs from the *Preferential Attachment Network Dataset* using a fixed time interval, which gives 50 samples for strongly and weakly directed graphs respectively.

From Fig. 3.4, the scatter plots of the weakly (strongly) directed approximations Y_{VN}^{WD} (Y_{VN}^{SD}) are much closer to the true values for the weakly (strongly) directed graphs Y_{VN}^D. Thus we conclude that the true value of von Neumann entropy and the simplified weakly (strongly) directed form we suggested are approximately equivalent on weakly (strongly) directed graphs.

Next we aim to explore whether the von Neumann entropy can be used to distinguish directed graphs with different structural properties. To this end we have generated graphs with different parameter settings and have explored the dependence of the von Neumann

68
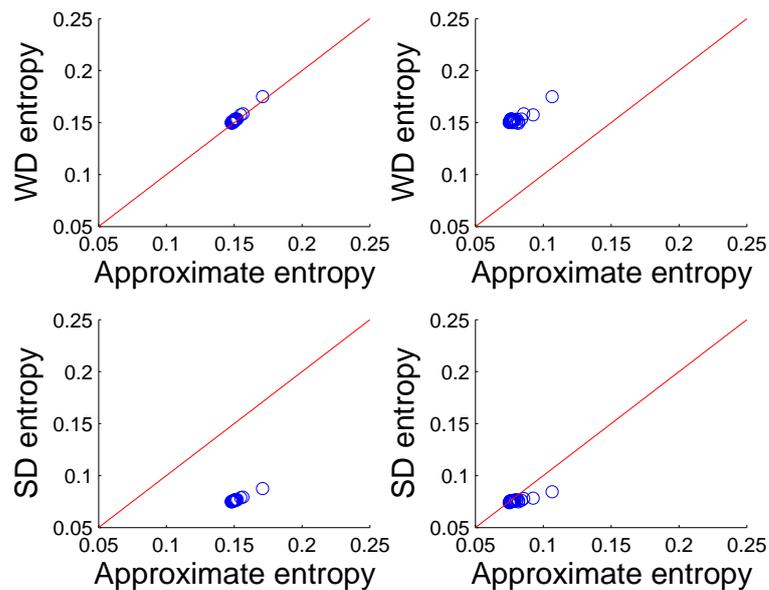
Figure 3.4: Comparing approximate von Neumann entropies for weakly directed and strongly directed graphs.

entropy on these parameters.

We commence by considering the Erdős-Rényi model, where the two parameters are the graph size n (or number of vertices) and the vertex connection probability p. We vary these parameters and randomly generate a number of directed graphs at each setting. We compute the mean and standard deviation of the normalized approximate von Neumann entropy Y_{VN}^D from Eq. (3.15) over samples with the same parameter settings.

Figure 3.5(a) shows the normalized approximate von Neumann entropy (mean and standard deviation as an error bar) for the Erdős-Rényi model, with $n = 20, 30, 50, 100$ as a function of p varying from 0.1 to 0.9. Figure 3.5(b) plots the same data for $p = 0.1, 0.2, 0.3, 0.9$ as a function of n varying from 20 to 100. From the plots, the mean value of the normalized entropy decreases gradually, which implies that the von Neumann entropy increases with both the graph size and the vertex connection probability. This result is as expected since in an Erdős-Rényi network, the structure becomes more complex when there are both a large number of vertices in the network (n is large) and there are a large number of random edges in the network (p is large). When the probability p is small, the standard deviation of entropy is particularly large. This is because for a network with a fixed size, a smaller number of directed edges in the network leads to a greater uncertainty of how these edges are connected. As a result there is significant variance in the network

(a) Vertex connection probability

(b) Graph size
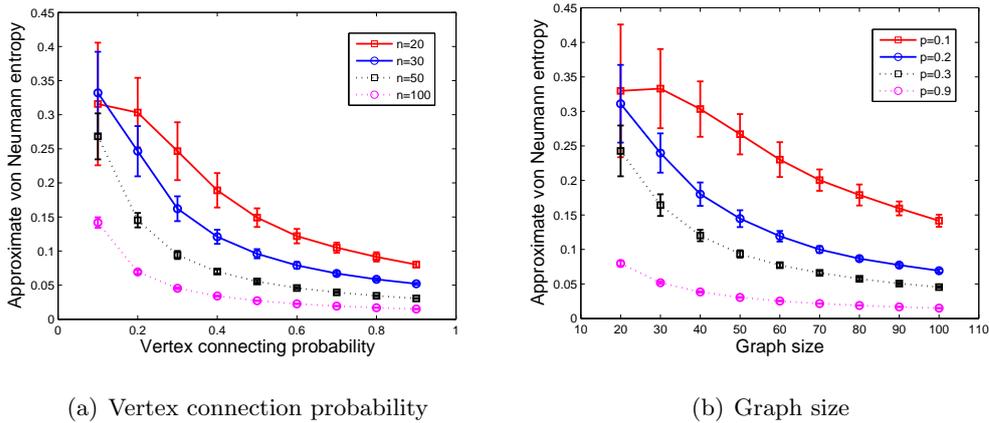
Figure 3.5: Mean and standard deviation of normalized approximate von Neumann entropy as a function of model parameters for Erdős-Rényi graphs.

entropy.

We then proceed to apply the von Neumann entropy to the directed graphs in the *Random Directed Graph Dataset* to investigate whether different topologies can be distinguished. Figure 3.6 shows the mean value of the normalized approximate von Neumann entropy as a function of graph size (again, standard deviation as an error bar). For a given graph size, the difference in mean entropy for different models is much larger than the standard deviation of the entropy within each model. This suggests that the variance in von Neumann entropy due to different parameter settings is much smaller than that due to differences in structure, which means that different network models have different values of von Neumann entropy for a given size.

Next we verify whether the von Neumann entropy can be used to determine the enzyme class of the protein graphs. Here in order to better visualize the result, we use the original approximate von Neumann entropy expression Eq. (3.10) instead of the normalized entropy Eq. (3.15). In Fig. 3.7 we show a histogram of the von Neumann entropy for the graphs in the database. The different line styles represent different enzyme classes in the database. Four classes of proteins (EC 3, EC 4, EC 5 and EC 6) show some separation. Another interesting feature is that class EC 1 is located between and is also overlapped with class EC 3 and EC 6. Because of the larger population of EC 1, the overlap is in fact relatively small. Unfortunately, class EC 2, on the other hand, cannot be easily separated as it is mixed with classes EC 3 and EC 5.

The above experiments in this section show that the directed von Neumann entropy
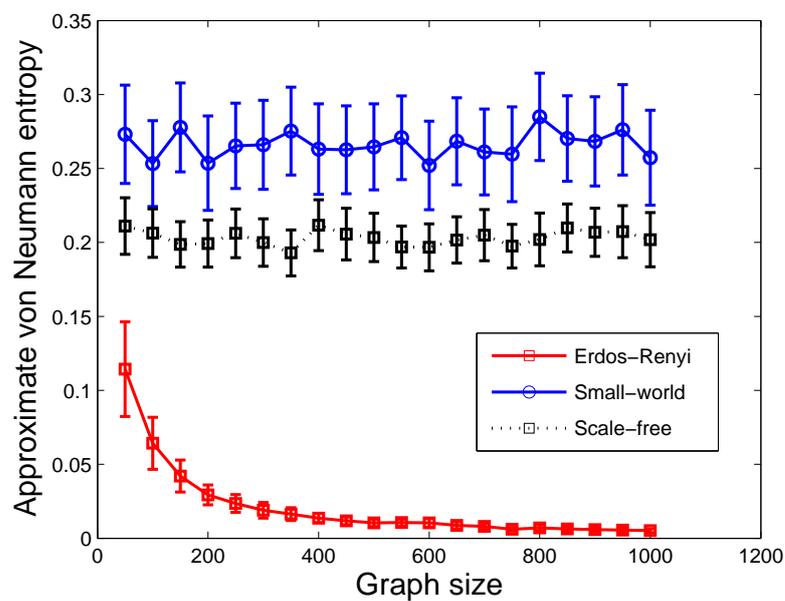
Figure 3.6: Mean and standard deviation of normalized approximate von Neumann entropy as a function of graph size for different models of directed graphs.
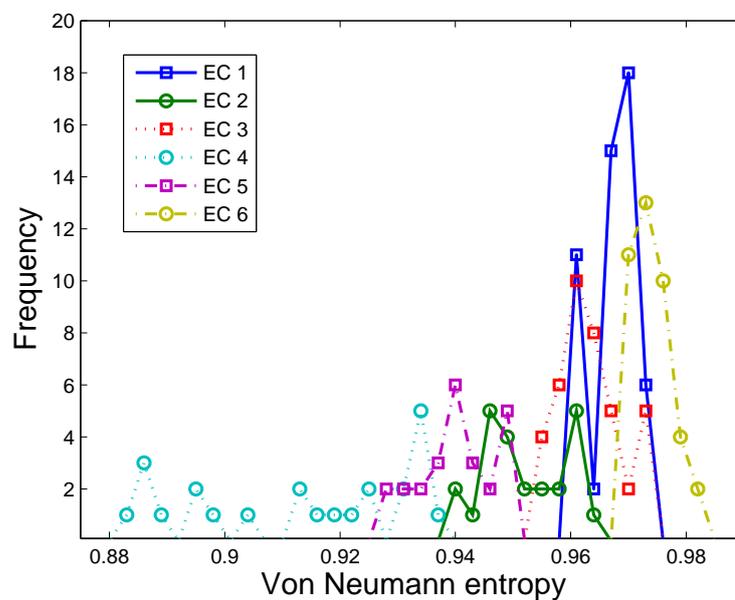
Figure 3.7: Histograms of approximate von Neumann entropy for different enzyme classes of proteins.

can be efficiently used to distinguish different types of directed graphs from both artificial and real-world data since it captures differences in structural features of directed networks.

Then, we explore whether the von Neumann entropy and its simplifications for strongly and weakly directed graphs can be used to detect changes in the structure of a citation network that evolves over time. In this context, it is important to note that a high impact (or highly cited) paper may cause a much more significant change in the network structure than a paper with an average citation profile, since such paper usually leads to a paradigm change in the subject it concerns.

We convert the Arxiv HEP-TH citation network to an evolving directed graph and explore whether the directed von Neumann entropy can be used to detect changes in graph structure caused by the publication of high impact papers.

As noted earlier, the *Citation Network Dataset* is hermetic in the sense that it does not contain any citation information related to papers that fall outside its coverage. Thus the citation graph grows from a single vertex to a graph consisting of 27770 vertices with 352807 directed edges. Occasionally a newly published paper may cite a number of papers that are not in the current citation network, i.e., these papers do not cite any papers in the dataset and are only cited by other papers in it. In this case we regard the newly published paper as a primary paper and the cited papers as its secondary papers. The primary paper and the secondary papers are thus introduced into the network at the same time epoch.

There are 25001 primary papers in the dataset, and we label them from 1 to 25001 according to the time at which they first appear in the citation network. Hence these ordinal labels index the epoch at which papers appear in the database and can be viewed as a time sequence, i.e., the citation network begins at $t = 1$ (January 1993) and ends at $t = 25001$ (April 2003).

The impact of a paper on a citation network is not reflected immediately after it is published. Instead, the influence develops and is sustained for a period of time. This is because after the publication of a high impact paper, a large number of subsequent papers will cite it (in the citation graph, the corresponding vertex will have a large in-degree). As a result its influence will be sustained until the most recent paper has cited it. In order to capture the impact of a paper, we use the rate of change of the directed von Neumann entropy.

To use this quantity to measure the citational influence of a paper, suppose a primary
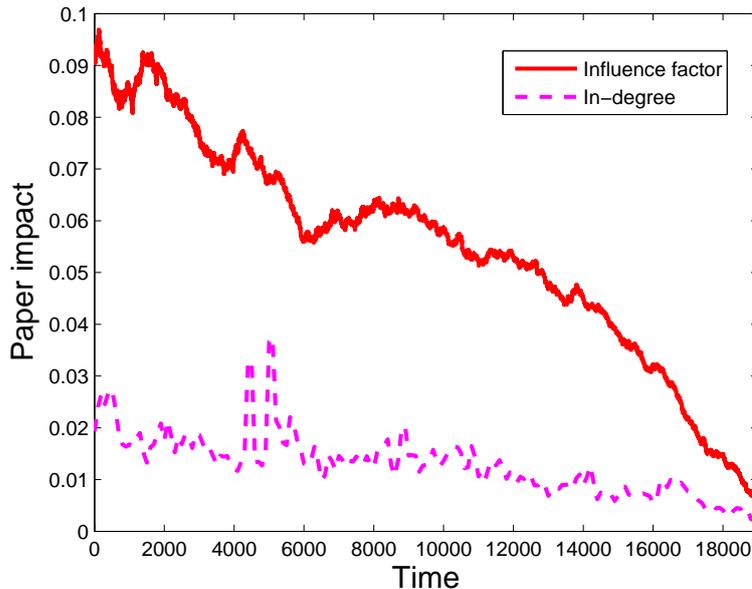
Figure 3.8: Citational influence factor and vertex in-degree as a function of time for the Arxiv HEP-TH citation network.

paper u is published at time t_0, and its impact is sustained for a period of length N, which means the impact ends at time t_N. We define the citational influence factor ξ_u of paper u as the mean value of the relative change in the normalized strongly directed approximation Y_{VN}^{SD}, which is given in Eq. (3.17), over the relevant influence period $t_0, t_1, \cdots, t_N$, i.e.,

$$\xi_u \triangleq \frac{\sum_{i=1}^{N} \left\{ Y_{VN}^{SD}(t_i) - Y_{VN}^{SD}(t_0) \right\}}{N \cdot Y_{VN}^{SD}(t_0)}. \tag{3.21}$$

From the dataset we find that most papers have an influence period between 5000 to 6000 (measured in terms of change in sequence number). Thus we take the average and fix $N = 5500$. At the beginning of the citation sequence the volume of data is not sufficient for reliable analysis. We thus start the analysis at $t = 5000$ instead of $t = 1$ and terminate at $t = 24000$ which gives a sequence length of 19000.

In Fig. 3.8, we plot both the influence factor and the in-degree distribution for primary papers against time. The main feature to note is that although the influence factor fluctuates, it decreases gradually to a value close to zero. This is because as time evolves, the citation network size increases rapidly, reducing the potential relative impact of more recent papers. Another important feature of this figure is that our influence factor can be used to reveal the changes in structure caused by influential papers. In the plot at epochs close to $t = 2000, 4500, 14000$, we see some significant fluctuations in the influence factor

curve, which represent significant changes in network structure. Turning our attention to the in-degree distribution, there are peaks at epochs around $t = 2000, 4500, 14000$, which means that papers published at these times are cited heavily. Thus we combine these observations and suggest that the influential papers can create significant changes in the structure of the evolving citation network.

To take this analysis a step further, we modify the original citation data and explore how the influence factors change. To this end, we select papers from a period of time and delete most of their citation connections. Figures 3.9(a) and 3.9(b) show the analysis before and after modifying the data if we delete connections in the time interval a) $t \in [4000, 6000]$ and b) $t \in [8000, 12000]$. As a result, the revised influence factors show a sharp drop in values, but after a transient time return to the behavior of the original curve. Thus there are significant differences in the network structure when high impact papers are published, and the directed graph von Neumann entropy can capture such differences.

3.4.2 Directed Heterogeneity Index

We proceed to analyze whether the directed heterogeneity index can be used to determine the topological structure of directed graphs efficiently. To do this, we commence by using this new index to distinguish different types of directed graphs in both *Random Directed Graph Dataset* and *Protein Dataset*, then we use it to detect abrupt changes in the structure of an evolving network, i.e., the *Citation Network Dataset*.

Figure 3.10(a) shows the directed heterogeneity index (mean and standard deviation as an error bar) for the Erdős-Rényi model, with $n = 20, 30, 50, 100$ and as a function of p varying from 0.2 to 0.9. Figure 3.10(b) plots the same data for $p = 0.1, 0.3, 0.9$ as a function of n varying from 30 to 100. From the plots, the directed heterogeneity index decreases gradually with both the graph size and the vertex connecting probability. This result is as expected since in an Erdős-Rényi network, the vertex out-degree and in-degree distribution becomes more uniform (less heterogeneous) when there are a large number of vertices in the network (n is large) and there are more edges that randomly connect two vertices in the network (p is large). When the probability p is small, the standard deviation of the directed heterogeneity index is particularly large. This is because for a network with a fixed size, a smaller number of directed edges in the network leads to a greater deviation in vertex out-degree and in-degree distribution (more heterogeneous). As a result there is significant variance in the network heterogeneous characteristics.
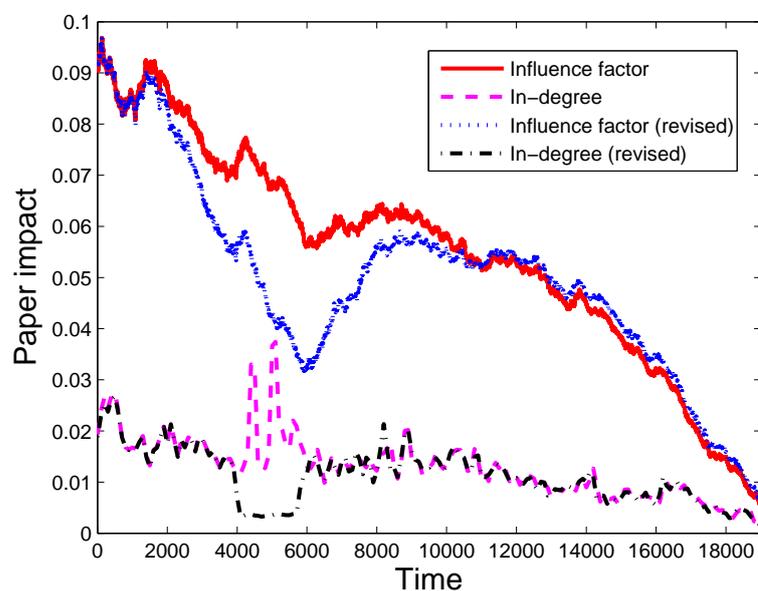
(a) Between 4000 and 6000
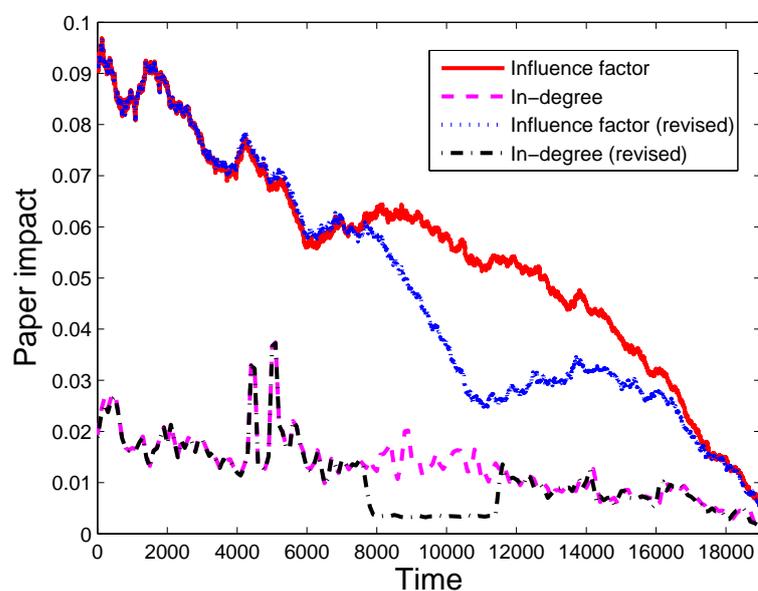
(b) Between 8000 and 12000

Figure 3.9: Citational influence factor and vertex in-degree as a function of time for modified citation networks when citation data are deleted during different time periods.

(a) Vertex connection probability
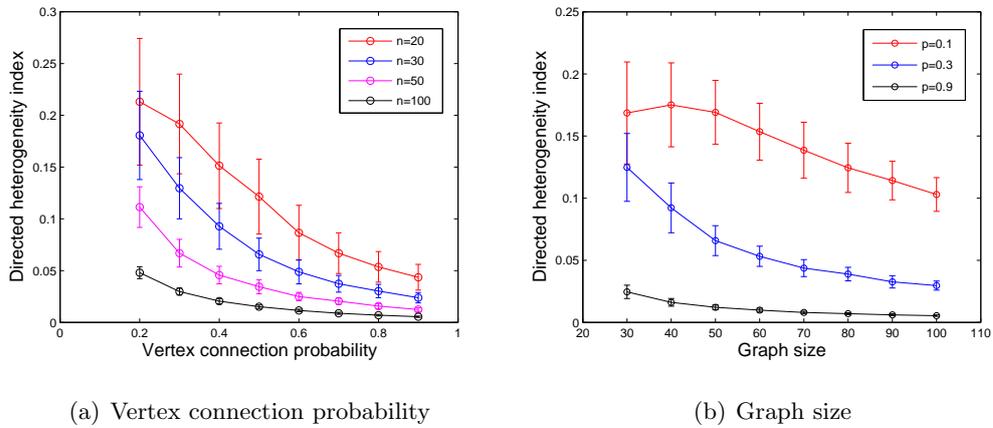
(b) Graph size

Figure 3.10: Mean and standard deviation of directed heterogeneity index as a function of model parameters for Erdős-Rényi graphs.
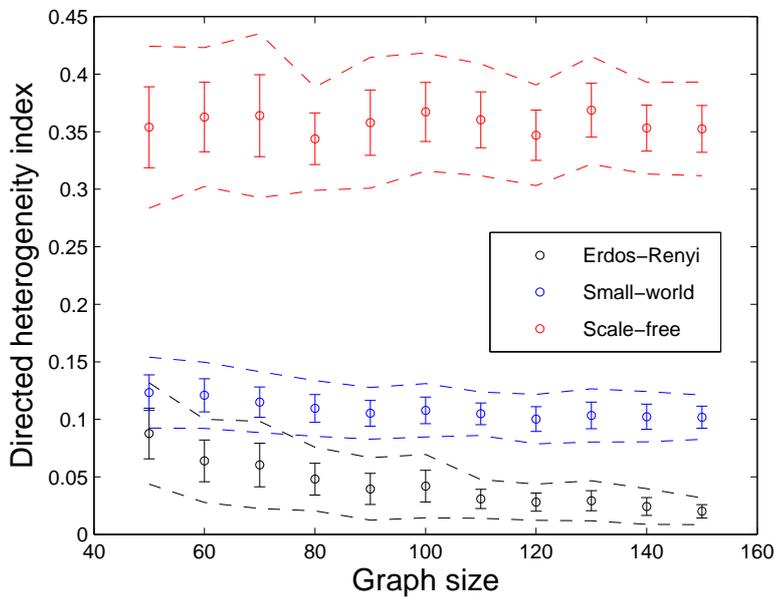
Figure 3.11: Mean, variance and standard deviation of directed heterogeneity index as a function of graph size for different models of directed graphs.
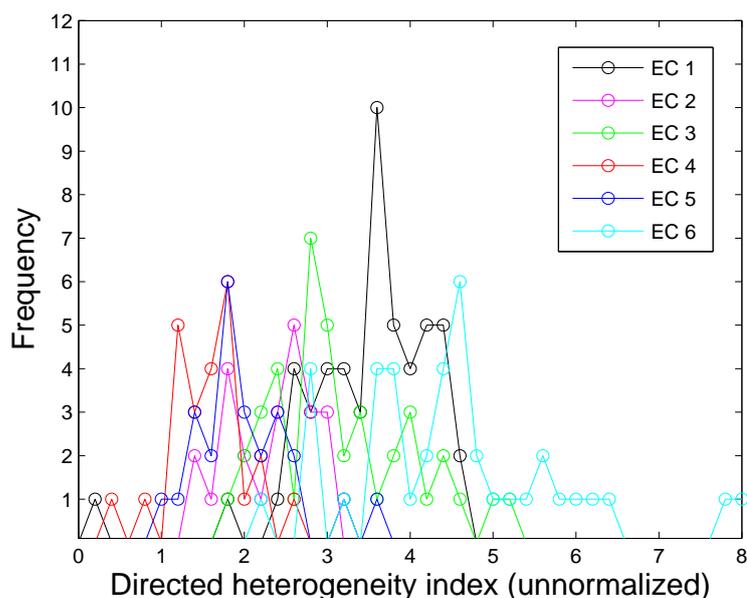
Figure 3.12: Histograms of unnormalized directed heterogeneity index for different enzyme classes of proteins.

Figure 3.11 shows both the mean and variance of the directed heterogeneity index as functions of graph size (again, standard deviation as an error bar). The mean values of directed heterogeneity index for Barabási-Albert and "small-world" graphs almost do not change with graph size, although the mean value of Erdős-Rényi graphs decreases slightly, which conforms to the result shown in Fig. 3.10. Moreover, for a given size of the graph the difference in mean heterogeneity index for different models is much larger than the standard deviation of heterogeneity index within each model. The variance in heterogeneity index due to different parameter settings is much smaller than that due to differences in structure. This means that different network models have different values of directed heterogeneity index for a given size.

Next we aim to verify whether the directed heterogeneity index can be used to determine the enzyme class of the protein graphs. Here in order to obtain clearer visual results, we use the quantity Ω^D given in Eq. (3.13) instead of the normalized directed heterogeneity index given in Eq. (3.14). In Fig. 3.12 we show a histogram of the directed heterogeneity index for the protein graphs in the *Protein Dataset*. The different coloured lines represent different enzyme classes in the database. The result is satisfying as four classes of proteins (EC 1, EC 3, EC 4 and EC 6) show some separation. Unfortunately, classes EC 2 and EC 5, on the other hand, cannot be easily separated as they are mixed
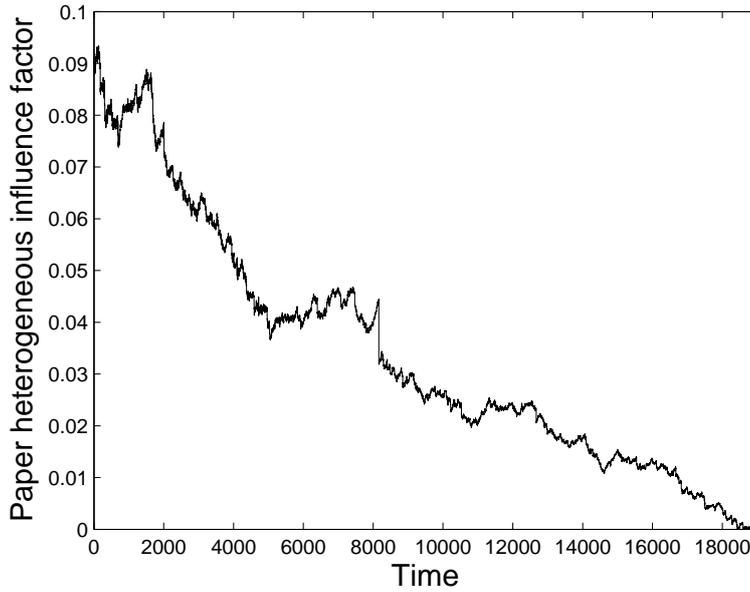
Figure 3.13: Citational heterogeneous influence factor as a function of time for the Arxiv HEP-TH citation network.

with classes EC 3 and EC 4.

The experiments show that the directed heterogeneity index can be efficiently used to distinguish different types of directed graphs from both artificial and real-world data since it captures the relative irregularity of each edge and the heterogeneous characteristics of a directed network.

Then, we investigate whether the directed heterogeneity index can be used to detect changes in the structure of a citation network that evolves over time. Similarly, we define the relative heterogeneous influence factor ξ'_u of paper u as the mean value of the relative change in quantity $\tilde{\Omega}^D$ which is given in Eq. (3.14) over the relevant influence period $t_0, t_1, \cdots, t_N$, i.e.,

$$\xi'_u \triangleq \frac{\sum_{i=1}^{N} \left\{ \tilde{\Omega}^D(t_i) - \tilde{\Omega}^D(t_0) \right\}}{N \cdot \tilde{\Omega}^D(t_0)}. \tag{3.22}$$

In Fig. 3.13, we plot the heterogeneous influence factor for primary papers against time. The main feature to note is that although the influence factor fluctuates, it decreases gradually to a value close to zero. This is because at the beginning, the citation network size is relatively small, any added paper may change the paper citation distribution significantly, then the rate of change in the citation network's directed heterogeneity index is great. However, as time evolves, a rapidly increasing number of papers are in-
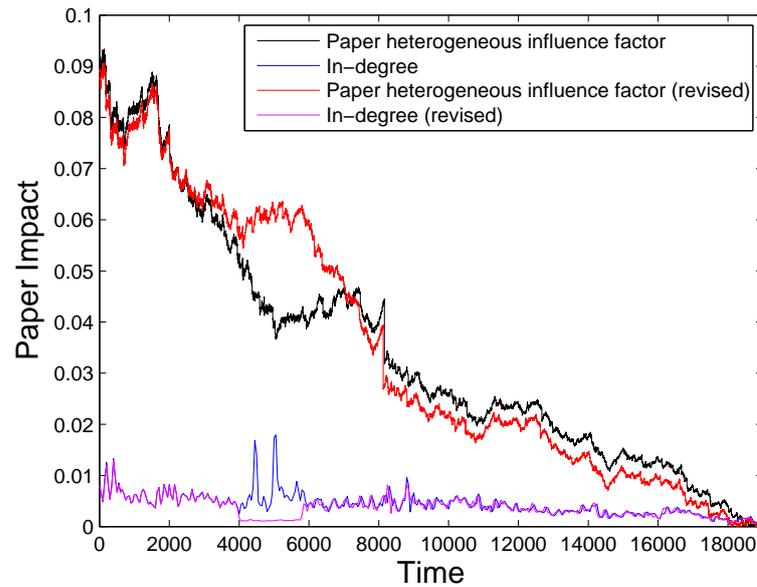
Figure 3.14: Citational heterogeneous influence factor as a function of time for modified citation networks when citation data are deleted between 4000 and 6000 time epochs.

troduced to the citation network, and most of these papers are of average citation profile, making the paper citation distribution more uniform, and that leads to the decrease of the relative heterogeneous impact of new papers.

We then modify the original citation data and explore how the relative heterogeneous influence factors change. To this end, we select papers from some period of time and delete most of their citation connections. Figure 3.14 shows the analysis before and after modifying the data if we delete connections in the time interval $t \in [4000, 6000]$. It clearly displays the revised influence factors show a sharp increase in values, but after a transient time return to the similar behavior of the original curve. This is not unexpected as after modifying, all the papers published between time 4000 and 6000 have citation profiles much lower than the average, making the network heterogeneous characteristics significantly different from the original data. As a result, the rate of change in citation network's directed heterogeneity index becomes greater and the difference is sustained for a certain period of time. Thus there are significant differences in the network structure when both high impact and low impact papers are published as they change the paper citation heterogeneous features, and the suggested directed heterogeneity index can capture such differences.

3.5 Summary

This chapter is motivated by the aim to develop novel and effective measures for quantifying the structural complexity of directed graphs. We have made a number of novel contributions. First we have shown how to compute the von Neumann entropy of a directed graph using Chung's definition of the normalized Laplacian matrix. We simplify the calculation of von Neumann entropy by replacing the Shannon entropy by the quadratic entropy. From this starting point, we have developed approximations to the entropy that can be computed using vertex in-degree and out-degree statistics. Moreover, we present specific approximations of the von Neumann entropy that apply to both strongly and weakly directed graphs, according to whether or not the majority of edges are unidirectional edges. We have also followed Estrada's idea and have constructed a directed version of heterogeneity index that can be used to measure the heterogeneous characteristics of directed graphs.

To evaluate both complexity measures and analyze their properties, we have undertaken experiments on both artificial and real-world network data. These experiments demonstrate that both the approximate von Neumann entropy and the heterogeneity index for directed graphs can be used to distinguish different classes of directed graphs (proteins) and analyze the structural changes of time-evolving networks (Arxiv HEP-TH citation network). Moreover, we show that the entropy characterization is not unduly limited by the approximations made in deriving it.

The work reported in this chapter can clearly be extended in a number of different ways. Firstly, we acknowledge that we have explored a relatively limited quantity of empirical data. It would for example be interesting to see if the method can be used to detect network anomalies and disturbances. Another interesting line of investigation would be to explore whether the method can be applied to the PageRank matrix, since it too is based on random walks on directed graphs. Finally, we plan to explore whether this work can be extended to edge-weighted graphs, labeled graphs and hypergraphs.

Chapter 4

Uses and Applications of Von Neumann Entropy

In this chapter, we explore a number of applications of the approximate von Neumann entropy for network analysis problems involving both undirected and directed graphs. First, commencing from the graph von Neumann entropy expression, we derive a local entropic contribution associated with each edge in the graph and use this to propose a novel entropic edge assortativity measure for both undirected and directed graphs. Then, we use this local entropic measure to analyze how the von Neumann entropy is distributed over edges in a graph. We show that this measure encodes a number of properties of the intrinsic structural properties of a graph, leading to the possibility of characterizing graphs of different structure. Specifically, we explore whether the von Neumann entropy distribution of a graph can be used to solve structural pattern recognition problems. To this end, we commence from the von Neumann entropy approximations and we calculate the entropic probability distribution for both undirected and directed graphs. Then, we extract multi-dimensional features that can be used to effectively represent the statistical information of the structural characteristics of those graphs.

4.1 An Entropic Edge Assortativity Measure

Assortativity or assortative mixing is the tendency of a network's vertices to connect to others with similar characteristics, and this has been shown to play a vital role in the structural properties of complex networks. Specifically, assortativity is often formalized as a correlation between the degree distinction of two vertices in a graph. This interpretation

of assortativity indicates that a network is highly assortative if the high degree vertices tend to be connected to other vertices with high degrees. Most of the existing assortativity measures have been developed on the basis of vertex degree information. However, there is a significant amount of additional information residing in the edges in a network, such as the edge directionality and weights. Recently, Foster et al. [51] have pointed out that the classification based on network assortativity is not always efficient for undirected networks. They further show that the fundamental feature of edge direction in a network also plays an important role, and thus propose a set of four directed assortativity measures based on vertex in-degree and out-degree combinations. Moreover, the von Neumann entropy has proved to be an efficient entropic complexity level characterization of the structural and functional properties of both undirected and directed networks. Hence, in this section we aim to explore whether we can combine these two methods and propose a novel edge assortativity measure which quantifies the entropic preference of edges to form connections between similar vertices in undirected and directed graphs. We use this measure to analyze how the entropy is distributed over edges. We also show that this quantity encodes a number of properties of the intrinsic structural properties of a graph, leading to the possibility of characterizing graphs of different structure.

4.1.1 Entropy Contribution for Undirected Edges

To commence, we recall the approximate von Neumann entropy of undirected graphs, which is given in Eq. (3.3)

$$H_{VN}^U = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{|\mathcal{V}|^2} \sum_{(u,v) \in \mathcal{E}} \frac{1}{d_u d_v}.$$

This approximation clearly contains two measures of graph structure. The first term measures the effect of graph size and the second term of this formula simply calculates the sum of each edge contribution to the whole entropy of a graph. This leads to the possibility of defining a normalized local entropic measure for a single edge in the graph.

To this end, we normalize the entropy expression with respect to the total number of edges in the graph in order to obtain a normalized edge entropy contribution, i.e.,

$$\gamma_{uv}^U = \frac{1}{|\mathcal{V}||\mathcal{E}|d_u d_v}. \tag{4.1}$$

For an arbitrary graph, this normalized local entropic measure clearly avoids graph size bias and gives the von Neumann entropy contribution associated with each edge in the graph.

4.1.2 Entropy Contribution for Directed Edges

Turning attention to the case of directed graphs, the approximations to the directed graph von Neumann entropy are expressed in Eq. (3.9) and Eq. (3.10):

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \left\{ \sum_{(u,v)\in\mathcal{E}} \left(\frac{1}{d_u^{out} d_v^{out}} + \frac{d_u^{in}}{d_v^{in} d_u^{out2}} \right) - \sum_{(u,v)\in\mathcal{E}_1} \frac{1}{d_u^{out} d_v^{out}} \right\}$$

and

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \left\{ \sum_{(u,v)\in\mathcal{E}} \frac{d_u^{in}}{d_v^{in} d_u^{out2}} + \sum_{(u,v)\in\mathcal{E}_2} \frac{1}{d_u^{out} d_v^{out}} \right\}.$$

In particular, when the graph is weakly directed, we ignore the summation over $\mathcal{E}_1$ in Eq. (3.9) in order to obtain the approximate entropy given in Eq. (3.11)

$$H_{VN}^{WD} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{\frac{d_u^{in}}{d_u^{out}} + \frac{d_v^{in}}{d_v^{out}}}{d_u^{out} d_v^{in}} \right\}.$$

It is natural to realize that in our analysis an undirected graph is equivalent to a weakly directed graph, since their von Neumann entropy expressions Eq. (3.3) and Eq. (3.11) are equivalent if we consider each undirected edge as a bidirectional one.

On the other hand, we can similarly obtain the approximate entropy for strongly directed graphs:

$$H_{VN}^{SD} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \left\{ \frac{d_u^{in}}{d_v^{in} d_u^{out2}} \right\},$$

which is given in Eq. (3.12). This approximation clearly sums the entropy contribution from each directed edge, which is based on the in and out-degree statistics of the vertices connected by the edge. In other words, we can compute a normalized local entropy measure for each directed edge in the strongly directed graph. To do this, we remove the term $1 - \frac{1}{|\mathcal{V}|}$ and normalize the remaining term with respect to the number of edges in the graph so that we obtain

$$\gamma_{uv}^D = \frac{d_u^{in}}{|\mathcal{V}||\mathcal{E}| d_v^{in} d_u^{out2}} \tag{4.2}$$

as the von Neumann entropy contribution for the edge (u, v).

4.1.3 Entropic Edge Assortativity Measure for Undirected Graphs

In this subsection, we propose a novel assortativity measure for both undirected and directed graphs based on the von Neumann entropy contributions associated with undirected and directed edges respectively. This method provides useful underpinning at the use of entropy in determining graph structure. For instance, a high entropic edge assortativity

indicates that edges with large entropy contributions associate preferentially and form some high entropy clusters in a graph. In contrast, a negative assortativity results from edges with high and low entropies that connect to each other.

Generally speaking, the assortativity is defined as a network indicator which measures to what extent the vertices are connected in a network. The most commonly used network assortativity is the one determined by the degree of vertices in the network, which is introduced by Newman [93]. Moreover, many other vertex characteristics can also be used to define the network assortativity, including the vertex centrality measures such as closeness and betweenness [47]. However, the vertex-based assortativity alone cannot be served as a sufficient network analysis tool since it does not consider the great amount of information contained in the network edges. Our goal is therefore to construct an edge-based network assortativity, using both the vertex and edge information, i.e., degree and edge entropy respectively.

We first consider undirected graphs. Mathematically, the traditional degree assortativity is given as a function of the degrees of pairs of linked vertices in an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [93]:

$$C_D = \frac{|\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} d_u d_v - [|\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} \frac{d_u + d_v}{2}]^2}{|\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} \frac{d_u^2 + d_v^2}{2} - [|\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} \frac{d_u + d_v}{2}]^2} \in [-1, 1]. \qquad (4.3)$$

When $C_D = 1$, the network is said to be perfectly assortative, when $C_D = 0$ the network is non-assortative, and when $C_D = -1$ the network is completely disassortative.

Similarly, for a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a set of four degree assortativity measures can also be defined according to Foster et al. [51]. Let $\alpha, \beta \in \{\text{in}, \text{out}\}$ be the directionality index for an edge at a vertex (i.e., whether it is incoming or outgoing). Then the directed degree assortativity measures are

$$C_D(\alpha, \beta) = \frac{|\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} [(d_u^\alpha - \bar{d}_u^\alpha)(d_v^\beta - \bar{d}_v^\beta)]}{\sigma^\alpha \sigma^\beta}, \qquad (4.4)$$

where $\bar{d}_u^\alpha = |\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} d_u^\alpha$ and $\sigma^\alpha = \sqrt{|\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} (d_u^\alpha - \bar{d}_u^\alpha)^2}$; $\bar{d}_v^\beta$ and σ^β are similarly defined.

With these definitions to hand, in order to obtain the edge-based network assortativity, we simply use the edge entropy information to replace the vertex degree in the formulas above, which means that the edge assortativity is quantified by the Pearson correlation coefficient of edge entropies in a network.

To proceed, we show how the edge entropy information is computed on both undirected and directed graphs. Suppose $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an undirected graph, or equivalently,
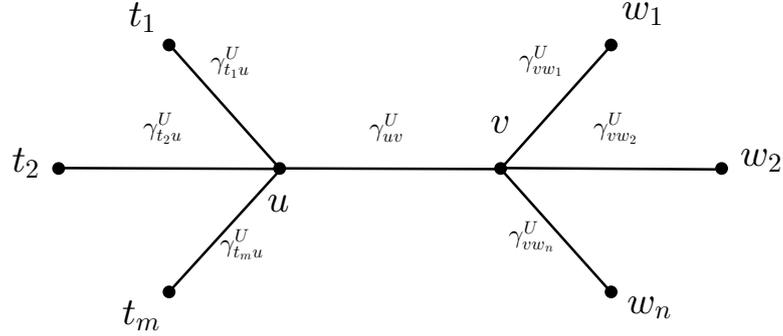
Figure 4.1: Illustration of calculation of quantities Γ_{uv}^U and Γ_{vu}^U associated with an undirected edge (u, v).

a weakly directed graph, then for an edge (u, v), we define the entropy contribution associated with the end vertex u of this edge Γ_{uv}^U as the summation of the entropies on the edges connected with u except the edge (u, v), i.e., $\Gamma_{uv}^U = \sum_{(t,u) \in \mathcal{E}, t \neq v} \gamma_{tu}^U$. The entropy contribution associated with another end vertex v is therefore $\Gamma_{vu}^U = \sum_{(v,w) \in \mathcal{E}, w \neq u} \gamma_{vw}^U$.

Then, we define the undirected edge entropic assortativity as the Pearson correlation coefficient of the entropy contributions associated with the two end vertices connected by an edge in the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with the result that

$$C_E^U = \frac{\sum_{(u,v) \in \mathcal{E}} (\Gamma_{uv}^U - \bar{\Gamma}_{uv}^U)(\Gamma_{vu}^U - \bar{\Gamma}_{vu}^U)}{\sigma_u^U \sigma_v^U}, \tag{4.5}$$

where $\bar{\Gamma}_{uv}^U = |\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} \Gamma_{uv}^U$ and $\sigma_u^U = \sqrt{\sum_{(u,v) \in \mathcal{E}} (\Gamma_{uv}^U - \bar{\Gamma}_{uv}^U)^2}$; $\bar{\Gamma}_{vu}^U$ and σ_v^U are similarly defined. Clearly, this edge assortativity index provides a novel way to understand the entropic preference of edges to form connections between similar vertices in a graph.

4.1.4   Entropic Edge Assortativity Measure for Directed Graphs

We turn our attention to the domain of directed graphs. Here we mainly focus on the strongly directed graphs, i.e., graphs in which there are few bidirectional edges. Assume $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a strongly directed graph, then for a directed edge starting from vertex u, ending at vertex v, we define the edge entropic assortativity as the Pearson correlation coefficient between the edge entropy contribution Γ_{uv}^u associated with all the outgoing edges
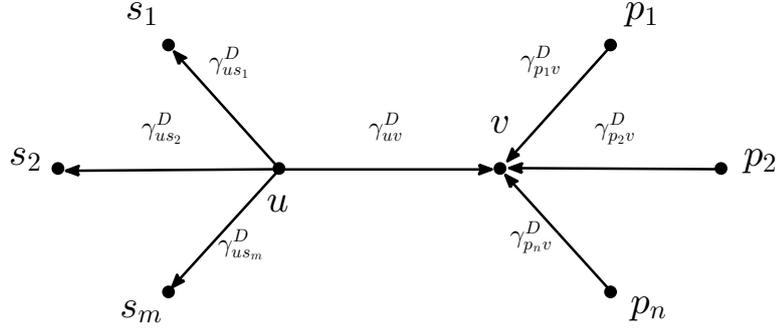
Figure 4.2: Illustration of calculation of quantities Γ_{uv}^D and Γ_{vu}^D associated with a directed edge (u, v).

of vertex u (exclude edge (u, v)) and the contribution Γ_{uv}^v associated with all the incoming connections of vertex v (except edge (u, v)). The reason we use such definition is that this expression conforms to the structure of the approximate von Neumann entropy for strongly directed graphs given in Eq. (3.12). Mathematically, we have $\Gamma_{uv}^u = \sum_{(u,s) \in \mathcal{E}, s \neq v} \gamma_{us}^D$ and $\Gamma_{uv}^v = \sum_{(p,v) \in \mathcal{E}, p \neq u} \gamma_{pv}^D$.

Therefore the edge entropic assortativity coefficient for strongly directed graphs is given by

$$C_E^D = \frac{\sum_{(u,v) \in \mathcal{E}} (\Gamma_{uv}^u - \bar{\Gamma}_{uv}^u)(\Gamma_{uv}^v - \bar{\Gamma}_{uv}^v)}{\sigma_u^D \sigma_v^D}, \tag{4.6}$$

where $\bar{\Gamma}_{uv}^u = |\mathcal{E}|^{-1} \sum_{(u,v) \in \mathcal{E}} \Gamma_{uv}^u$ and $\sigma_u^D = \sqrt{\sum_{(u,v) \in \mathcal{E}} (\Gamma_{uv}^u - \bar{\Gamma}_{uv}^u)^2}$; $\bar{\Gamma}_{uv}^v$ and σ_v^D are similarly defined. This measure is bounded between -1 and 1: a high coefficient of a graph indicates that most of the directed edges in the graph start from the vertex with outgoing edges that have high entropy contributions, and point to the vertex with incoming edges with high entropy contributions. Conversely, a negative coefficient results from most of the directed edges connect two vertices that have significantly different von Neumann edge entropy contributions.

4.2   Entropic Graph Embedding via Multivariate Degree Distributions

Structural complexity measures and embedding have both been extensively and separately employed for the problems of graph clustering and classification. In this section we aim to explore whether entropy component analysis (ECA) can be used as a means of combining these two fundamental approaches. Specifically we develop a novel method that embeds both undirected and directed graphs into a feature space based on the graph edge entropy distribution. The method commences from a characterization based on the distribution of the von Neumann entropy of a graph with the degree configurations associated with edges. Based on this analysis we identify the local entropy contribution associated with each edge in a graph, and thus obtain a simple entropic characterization of graph structure, based on a histogram in which the bins are indexed by vertex degree and the bin-contents is the total entropy contribution associated with the edges that connect vertices of specific degree. This distribution of entropy with vertex degree can be encoded as a multivariate array, which captures the structure of the graph in terms of an entropic measure of complexity. The matrix can hence be viewed as a sample of entropy histograms from different graphs. By performing PCA on a sample of histograms, we embed populations of graphs into a low-dimensional space. We apply this method to the problem of graph classification, and compare the classification results of our new method with some alternative state-of-the-art pattern recognition methods on bioinformatics data.

4.2.1   Undirected Graph Embedding via Von Neumann Entropy Distribution

In our analysis, we propose a novel graph embedding method based on the idea of kernel ECA [69], which is a technique that transforms data to a space spanned by those kernel PCA axes contributing most to the entropy estimate of the data. Specifically, commencing from an approximation of the von Neumann entropy of an undirected graph, we analyze the contribution originating from each edge. We show that this contribution is determined by the reciprocal of the product of the degrees of the two vertices defining the edge. Based on this observation we explore the bivariate distribution of entropy with different combinations of vertex degree that define edges in a graph. In practice this distribution can be computed by constructing a two-dimensional histogram whose bins are indexed

by the degrees of the connected vertices and whose contents accumulate the edge entropy contribution over the entire graph. The contents of the histogram can be represented by a matrix whose contents can be encoded as a long vector, which serves as a feature vector for the graph.

One of the problems that potentially limits this approach is that the vertex degree is unbounded. Hence, the size of the histogram can become large and it can become populated by a large number of empty bins. This renders the analysis of the feature vector unstable. In order to keep the vector length constant for graphs with a large variance in vertex degree, we use the vertex degree probability distribution to construct the cumulative distribution function (CDF), from which we can determine the m-quantiles, which divide the ordered vertex degree data into m essentially equal-sized parts. This allows us to relabel each vertex with a specific quantile point label $1, 2, \cdots, m$. As a result, the length of our proposed feature vector is not affected by the variance of the degree distribution.

We recall that for an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the approximate von Neumann entropy expression that computed from the second-order Taylor series approximation for the Shannon entropy, which is given in Eq. (3.4)

$$H_T^U = \ln |\mathcal{V}| - \frac{1}{2|\mathcal{V}|} \sum_{(u,v) \in \mathcal{E}} \frac{1}{d_u d_v}.$$

This approximation clearly contains two measures of graph structure, one is the graph size while the second is related to the degree statistics of vertices connected by edges. Specifically, the second term of this formula simply calculates the sum of each edge's degree-based entropy contribution in a graph, leading to the possibility of defining a normalized local entropic measure for a single edge in the graph.

To this end, we normalize this approximate entropy with respect to the total number of edges contained in the graph and thus obtain the expression of the normalized local entropic measure for an edge (u, v),

$$\gamma'^U_{uv} = \frac{\ln |\mathcal{V}|}{|\mathcal{E}|} - \frac{1}{2|\mathcal{E}||\mathcal{V}|d_u d_v}. \tag{4.7}$$

This local measure represents precisely the von Neumann entropy contribution of each single edge in the graph, since the sum of these measures over all edges leads to the value of the approximate von Neumann entropy in Eq. (3.4)

$$\sum_{(u,v) \in \mathcal{E}} \gamma'^U_{uv} = H_T^U.$$

Moreover, this normalized local entropic measure avoids the graph size bias, which means that for an arbitrary graph, it is the degree-based edge entropic measure, not the vertex number or edge number of the graph that distinguishes the entropy contribution of a single edge.

Our graph embedding method is based on the statistical information of edges in the graph. In other words, we compute the sum of the normalized local entropic measures of edges with the same degree status, i.e., the degrees of two vertices of an edge, and thus obtain a two-dimensional histogram which represents the edge-based entropy distribution of the graph. Since our idea heavily relies on the vertex degree statistics of graphs, it is natural to realize that different degree distributions may lead to completely different histogram bin numbers, rendering the instability of our proposed feature vectors.

In order to overcome such problem, here we introduce the definitions of cumulative distribution function and quantiles. Suppose a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has $|\mathcal{V}|$ vertices with an ordered degree sequence $d_1 \leq d_2 \leq \cdots \leq d_{|\mathcal{V}|}$, then the degree distribution is the probability distribution of these degrees over the entire graph. The corresponding cumulative distribution function is then given by

$$F_x(d_i) = p(x \leq d_i),$$

where $i = 1, 2, \cdots, |\mathcal{V}|$. This function describes the probability that a given degree x takes on a number less than or equal to degree d_i.

Broadly speaking, quantiles are the points taken at regular intervals from the cumulative distribution function of vertex degrees. Specifically, they divide the ordered degree data $d_1, d_2, \cdots, d_{|V|}$ into a number of essentially equal-sized data subsets. Since the vertex degree is always a non-negative integer, these quantiles are thus can be viewed as new degree labels. In our analysis, we let the number of subsets be m, so the m-quantiles can be obtained as follows

$$Q_j^U = \underset{d_i}{\operatorname{argmin}} \left\{ F_{Q_j^U}(d_i) - \frac{j}{m} \right\},$$

where $i = 1, 2, \cdots, |\mathcal{V}|$ and $j = 1, 2, \cdots, m$. These new degrees satisfy $Q_1^U \leq Q_2^U \leq \cdots \leq Q_m^U$ and in fact, $Q_m^U = d_{|\mathcal{V}|}$.

With these ingredients, we can give each vertex in the graph a new degree label. To do this, for a given vertex u, we first examine its original degree d_u, if d_u satisfies that $Q_{k-1}^U < d_u \leq Q_k^U$, then vertex u is labelled with $q_u = k$ (here we define $Q_0^U = 0$). With all the vertices in the graph having new degree labels from 1 to m, we then can

simply construct a two-dimensional edge-based entropy histogram whose bin number in each dimension is fixed to m.

We construct a $m \times m$ zero matrix M^U whose elements are the histogram bin contents, and where the row and column indices represent the new degree labels of vertices and run from 1 to m. For instance, the entry M^U_{12} accumulates the entropy contribution for all the edges that connect vertices with relabeled degree labels 1 and 2.

To compute the bin contents we proceed as follows. First, we calculate the normalized local contribution to the entropy from each edge in Eq. (4.7), then we accumulate the sum over all edges that have the same degree combinations. We store this accumulated sum in the corresponding element of the matrix M. The elementwise accumulation is formally given as

$$M^U_{ij} = \sum_{\substack{q_u=i, q_v=j \\ i \leq j, (u,v) \in \mathcal{E}}} \left\{ \frac{\ln |\mathcal{V}|}{|\mathcal{E}|} - \frac{1}{2|\mathcal{E}||\mathcal{V}|d_u d_v} \right\},$$

where $i, j = 1, 2, \cdots, m$ and $q_u = i$ denotes that vertex u, with original degree d_u is assigned a new degree label i. It is worth noting that since we consider only undirected graphs here, there is no direction information on each edge. As a result the matrix is symmetric since there is no difference between the elements M^U_{ij} and M^U_{ji}. So for convenience, we do not store the elements in the lower triangle below the main diagonal i.e., the matrix M^U is upper triangular. To extract a feature vector from M^U, as all the entries below the main diagonal of M^U are zeros, we can simply list all the upper triangular elements row by row, with the result that

$$V^U = (M^U_{11}, M^U_{12}, \cdots, M^U_{1m}, M^U_{22}, M^U_{23}, \cdots, M^U_{mm})^T.$$

Clearly, this feature vector has length

$$m + (m-1) + (m-2) + \cdots + 1 = m(m+1)/2.$$

To summarize, our graph embedding method is based on the bivariate distribution of von Neumann entropy contribution with vertex degree for edges in a graph. Moreover, since the von Neumann entropy quantifies the structural complexity of a graph, our proposed feature vector represents statistical information concerning the local structural properties in the graph.

4.2.2 Directed Graph Embedding via Von Neumann Entropy Distribution

We proceed to explore whether we can extract multi-dimensional structural features from directed graphs, and hence apply standard techniques from pattern recognition and machine learning to embed, cluster and classify data in the form of samples of directed graphs. We commence by computing the von Neumann entropy associated with each edge in a directed graph. This leads us to a four-dimensional characterization of directed graph structure, which depends on the distribution of entropy with the in and out-degrees of pairs of vertices connected by directed edges. We represent this distribution by a four-dimensional histogram, which can be encoded as a long-vector for the analysis purposes. To curb the size of the histogram, we show how to requantize the bin-contents using quantiles of the four cumulative degree distributions.

We recall the approximate von Neumann entropy expression for strongly directed graphs

$$H_{VN}^{SD} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2} \sum_{(u,v) \in \mathcal{E}} \left\{ \frac{d_u^{in}}{d_v^{in} d_u^{out2}} \right\}.$$

which is given in Eq. (3.12).

In the previous section, we have shown how the von Neumann entropy contribution for each edge in a graph can be derived. Specifically, for an edge (u,v) in graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we compute

$$\gamma_{uv}^D = \frac{d_u^{in}}{|\mathcal{E}||\mathcal{V}|d_v^{in}d_u^{out2}} \tag{4.8}$$

as the entropy contribution. If this edge is bidirectional, i.e., $(u,v) \in \mathcal{E}_2$, then we add an addition entropy contribution to γ_{uv}^D

$$\gamma_{uv}' = \frac{1}{|\mathcal{E}_2||\mathcal{V}|d_u^{out}d_v^{out}}.$$

This local measure represents the entropy associated with each directed edge since for arbitrary directed graphs, we have $\sum_{(u,v) \in \mathcal{E}} \gamma_{uv}^D + \sum_{(u,v) \in \mathcal{E}_2} \gamma_{uv}' = H_{VN}^D$ and for strongly directed graphs, we also have $\sum_{(u,v) \in \mathcal{E}} \gamma_{uv}^D = H_{VN}^{SD}$. Moreover, this measure avoids the bias caused by graph size, which means that it is the edge entropy contribution determined by the in and out-degree statistics, and neither the vertex number nor edge number of the graph that distinguishes a directed edge.

The directed graph characterization is based on the statistical information converged by the distribution of directed edge entropy with the in and out-degrees of the starting and

end vertices. We represent this distribution of entropy using a four-dimensional histogram over the in and out-degrees of the two vertices.

As noted above, one potential problem is that the bin-contents can become sparse in a high-dimensional histogram. To overcome this problem we again use the cumulative distribution function. Suppose a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has $|\mathcal{V}|$ vertices which have been sorted according to in-degree (or out-degree) in the sequence $d_1^{in} \leq d_2^{in} \leq \cdots \leq d_{|\mathcal{V}|}^{in}$. Let $p(x = d_i^{in})$ be the in-degree probability distribution of the graph. The corresponding cumulative distribution function for the in-degree is given by

$$F_x(d_i^{in}) = p(x \leq d_i^{in}),$$

where $i = 1, 2, \cdots, |\mathcal{V}|$. This function describes the probability that a given in-degree x takes on a number less than or equal to d_i^{in}.

Quantiles are intervals of equal size over the cumulative distribution function. They divide the ordered data $d_1^{in}, d_2^{in}, \cdots, d_{|V|}^{in}$ into a number of equal-sized data subsets. Since vertex degree is always a non-negative integer, the quantiles can thus be viewed as new quantization of the degree based on its statistical distribution. We define the degree quantiles over the cumulative distribution of degree for the entire sample of graphs under study, and produce requantized versions of the individual entropy histograms for each individual graph. Suppose the number of quantiles in each dimension of the degree distribution is fixed to be m. Then, for example, the m-quantiles of the in-degree distribution can be obtained as follows

$$Q_j^D = \underset{d_i^{in}}{\operatorname{argmin}} \left\{ F_{Q_j^D}(d_i^{in}) - \frac{j}{m} \right\},$$

where $i = 1, 2, \cdots, |\mathcal{V}|$ and $j = 1, 2, \cdots, m$. Clearly, these degree quantiles satisfy $Q_1^D \leq Q_2^D \leq \cdots \leq Q_m^D$ and in fact, $Q_m^D = d_{|\mathcal{V}|}^{in}$.

With the sample degree quantiles to hand, we assign each vertex degree quantile labels. We first examine the original in-degree d_u^{in} of a vertex u, if d_u^{in} satisfies the condition that $Q_{k-1}^D < d_u^{in} \leq Q_k^D$ (here $Q_0^D = 0$), then its in-degree quantile is $q_u^{in} = k$. The corresponding out-degree quantile labels can also be obtained in the same manner. Since all the vertices in the graph have in-degree and out-degree quantile labels ranging from 1 to m, we can then simply construct the directed edge entropy histogram whose size in each dimension is fixed to m. The histogram is stored as a four-dimensional array.

To do this, we first construct a $m \times m \times m \times m$ array M^D whose elements represent the histogram bin-contents, and whose indices represent the degree quantile labels of the

vertices. For instance, the element M_{1234}^D accumulates the entropy contribution for all the directed edges starting from vertices with out-degree quantile label 1 and in-degree quantile label 2, pointing to vertices with out-degree quantile label 3 and in-degree quantile label 4. We then compute the bin-contents by summing the directed edge entropy contributions over the sample graph. The histogram bins contain all directed edges having the same quantile label combinations. We store the accumulated sum in the corresponding element of array M^D. The elementwise accumulation is formally given as

$$M_{ijkl}^D = \sum_{\substack{q_u^{out}=i, q_u^{in}=j \\ q_v^{out}=k, q_v^{in}=l \\ (u,v)\in\mathcal{E}}} \left\{ \frac{d_u^{in}}{|\mathcal{E}||\mathcal{V}|d_v^{in}d_u^{out2}} \right\}.$$

If the graph contains bidirectional edges, we additionally accumulate the following quantity

$$M_{ijkl}' = \sum_{\substack{q_u^{out}=i, q_u^{in}=j \\ q_v^{out}=k, q_v^{in}=l \\ (u,v)\in\mathcal{E}_2}} \left\{ \frac{1}{|\mathcal{E}_2||\mathcal{V}|d_u^{out}d_v^{out}} \right\},$$

where $i, j, k, l = 1, 2, \cdots, m$. To extract a feature vector from M^D, we can simply list all the elements in the array, with the result that

$$V^D = (M_{1111}^D, M_{1112}^D, \cdots, M_{111m}^D, M_{1121}^D, M_{1122}^D, \cdots, M_{mmmm}^D)^T$$

Clearly, this feature vector has length m^4.

It is worth pausing to consider the case of strongly directed graphs. For such graphs, from Eq. (3.12) the directed edge entropy does not depend on d_v^{out}. As a result the dimensionality of the corresponding histogram can be reduced from four to three by ignoring the third index k in M_{ijkl}^D. This leads to a new feature vector with length m^3. In the following discussion, to distinguish between these two types of feature vectors, we name the former full-form (FF) while the latter strongly-directed (SD).

When accumulated in this way we effectively count directed edges with the same configurations of degree quantile labels, and weight them according to their entropy. If the different quantile labels were independent, we would expect a uniform histogram. However, structure in the individual sample graphs due to preferred combinations of vertex in-degree and out-degree will give rise to a non-uniform distribution. To some extent, the quantization of the distribution of entropy with degree according to quantile labels, may dilute this structure due to merging adjacent degree bins. However, the directed edge entropy contribution is based on the original vertex in and out-degree statistics, and the

m-quantiles play a role in diminishing the bias caused by different populations of directed graphs. Therefore the entropic representation can still be effective in capturing statistical information concerning the local structural properties in the graph. By embedding graphs into a space spanned by feature vectors, it provides a theoretically principled and efficient tool for graph characterization tasks, which captures the graph characteristics at both the statistical and structural levels.

4.3 Experiments

In this chapter, we have proposed a novel edge assortativity characterization for quantifying the assortative mixing properties for both undirected and directed graphs based on the von Neumann entropy associated with edges. We have also developed an entropic graph embedding method that can be applied to both undirected and directed graphs. To evaluate these methods and analyze their properties, in this section we employ them to solve a number of classification problems and compare their performance with several state-of-the-art methods.

4.3.1 Entropic Edge Assortativity

We first explore whether the entropic edge assortativity measure can reveal more useful features of the graph structure than the traditional degree-based ones. To this end, we confine our attention to two main tasks. We first apply the entropic edge assortativity measure to a number of real-world complex networks to show that it can effectively reflect to what extent the vertices are connected preferentially in a network. We then demonstrate one advantage of this novel assortativity characterization, namely that it is more efficient in distinguishing between different classes of complex networks than the traditional degree-based measures.

We commence by comparing the performance of traditional degree assortativity coefficients and our novel entropic edge assortativity measure on real-world collaboration citation networks. These include the Arxiv Astro Physics, Condensed Matter, General Relativity, High Energy Physics and High Energy Physics Theory networks [81]. Table 4.1 gives the network size, edge number and value of both the degree and edge assortativity measures. From the table, all the coauthership networks have positive degree assortativity coefficients. This is a reasonable result since productive authors prefer to collaborate.

Datasets	HepTh	HepPh	GrQc	CondMat	AstroPh
Network size	9877	12008	5242	23133	18772
Edge number	51971	237010	28980	186936	396160
Degree assort.	0.2674	0.6322	0.6592	0.1339	0.2051
Edge entropy assort.	0.2012	0.6035	0.3910	0.3435	0.5458

Table 4.1: Degree assortativity coefficients and entropic edge assortativity measures of real-world undirected complex networks.

However, the traditional assortativity coefficient has difficultly in distinguishing between CA-HepPh and CA-GrQc networks as their values are similar. The entropic edge assortativity coefficient, on the other hand, is able to characterize these two networks. One of the reasons for this is that the edge assortativity measure can capture not only the degree properties of vertices, but also the underlying entropic structural complexity associated with the edges in a network.

Next we show that the edge assortativity measure is more efficient than the traditional degree-based measure in classifying graphs that belong to different random graph models. To do this we employ both assortativity measures to the *Random Undirected Graph Dataset*, which contains a large number of undirected random graphs that are generated according to one of three graph models, namely a) the Erdős-Rényi model [43], b) the "small-world" model [124], and c) the Barabási-Albert model [13]. The different graphs in the database are generated using a variety of model parameters, e.g., the graph size and the connection probability in the Erdős-Rényi model, the edge rewiring probability in the "small-world" model and the number of added connections at each time step in the Barabási-Albert model.

In Fig. 4.3 we plot the mean value of the degree assortativity coefficients computed from Eq. (4.3) and the entropic edge assortativity measures computed using Eq. (4.5) as a function of graph size (standard deviation as an error bar). In the left panel, all three classes of graphs tend to have zero assortative mixing when the graph size becomes very large, and it is difficult to separate the "small-world" and Barabási-Albert graphs. Turning our attention to the right panel, the difference in mean edge assortativity coefficients for different models is much larger than the standard deviation of the coefficients for the different models, even when the graph size is large. This suggests that the variance in the edge assortativity measure due to different parameter settings is much smaller than

(a) Degree assortativity coefficients   (b) Entropic edge assortativity measures
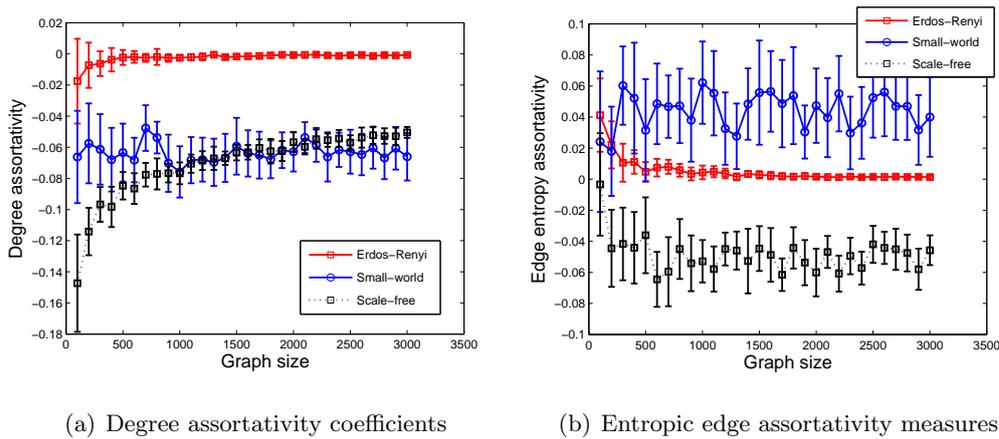
Figure 4.3: Mean and standard deviation of degree and entropic edge assortativity measures for different models of undirected graphs.

that due to differences in structure. This indicates that different network models have different values of edge assortativity coefficients for a given size. This accords with our expectations since the entropy itself is sensitive to the different graph models.

For directed graphs, we first provide a comparison of our entropic directed edge assortativity measure, and the four degree assortativity coefficients that can be computed from the four combinations of in and out-degree on the two vertices of an edge. We commence with a study on some real-world networks, and these include the Wikipedia vote network, provided by Leskovec et al. [79], the Gnutella peer-to-peer networks from August 5 to 9, 2002, which are a sequence of snapshots of the Gnutella peer-to-peer file sharing network [110] and the Arxiv HEP-TH citation network [80]. Table 4.2 gives the network size, edge number and the values of in/in-degree, in/out-degree, out/in-degree, out/out-degree and edge assortativity measures.

There are a number of observations to note concerning this data. In the Wikipedia vote network, a person who receives many votes is more likely to vote for a person who also obtains a large number of votes, rather than voting for individuals who vote many times. In the file sharing networks, computers that receive a great number of documents preferentially share files with one-another. Computers that send many files are unlikely to share files with computers that receive many documents. For the citation network, important papers are those cited most heavily and this can be reflected accurately by the degree assortativity measures. Although when taken in combination the four types of directed degree assortativity coefficients are useful in characterizing different networks, it

Datasets	Wiki-Vote	p2p-G05	p2p-G06	p2p-G08	p2p-G09	Arxiv HEP-TH
Network size	7115	8846	8717	6301	8114	27751
Edge number	103689	31839	31525	20777	26013	352807
In/in deg. assort.	0.0051	0.0312	0.0880	0.1079	0.1042	0.0405
In/out deg. assort.	0.0071	-0.0002	0.0322	0.0315	0.0190	0.0055
Out/in deg. assort.	-0.0832	-0.0034	-0.0032	-0.0285	-0.0327	0.0016
Out/out deg. assort.	-0.0161	-0.0017	0.0082	-0.0157	-0.0062	0.0951
Edge entropy assort.	0.0006	0.0053	-0.0092	-0.0038	-0.0055	0.1126

Table 4.2: Degree assortativity coefficients and entropic edge assortativity measures of real-world directed complex networks.

is difficult to use a single measure alone to do this. However, when using the entropic edge assortativity measure developed for directed graphs, networks with different structure are efficiently characterized.

In Fig. 4.4 we plot the values of the entropic edge assortativity coefficients, and compare them to the degree assortativity coefficients obtained with the four different combinations of vertex in and out-degree on an edge computed from Eq. (4.4). Here we use randomly generated data for three different directed graph models contained in the *Random Directed Graph Dataset*, which has been described in Chapter 3. The figure shows the assortativity measures versus graph size, and shows the mean value and standard deviation. The most important feature in the figure is that although the Barabási-Albert networks are easily separated, the Erdős-Rényi and "small-world" networks are overlapped significantly, for each of the four degree assortativity coefficients. However, Fig. 4.4(e) suggests that as the graph size increases, for all three models the mean values of the edge assortativity measures grow slowly and approach zero, with clear separations between them. The result obtained here demonstrates that the edge assortativity measure provides a powerful tool for capturing both the degree properties and the entropic information on edges in a directed network.

4.3.2 Undirected Graph Embedding

In this subsection we provide a comprehensive experimental evaluation of the feature-vector-based graph embedding method suggested. We perform PCA over a set of feature vectors extracted from graphs. This allows us to identify a new basis associated with
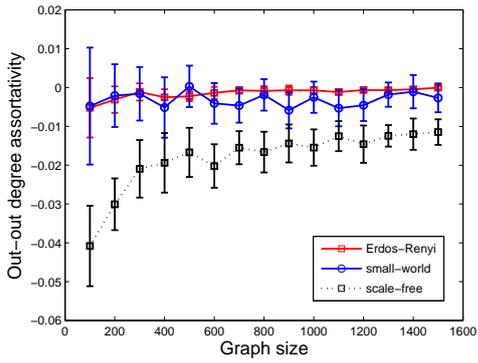
(a) Out-out degree assortativity coefficients
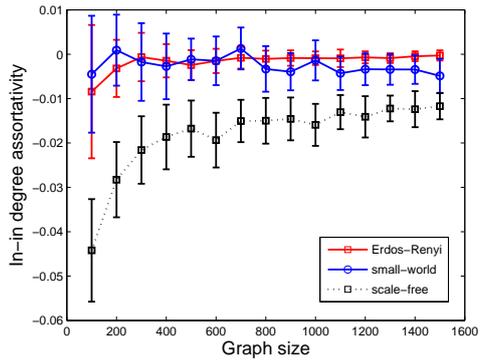
(b) In-in degree assortativity coefficients
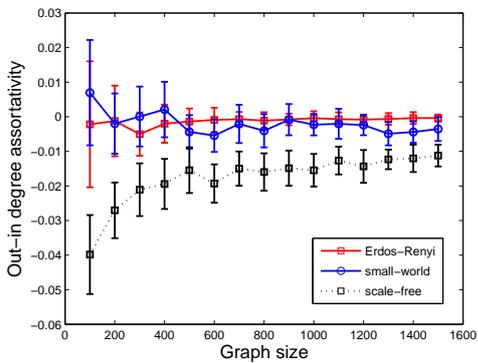
(c) Out-in degree assortativity coefficients
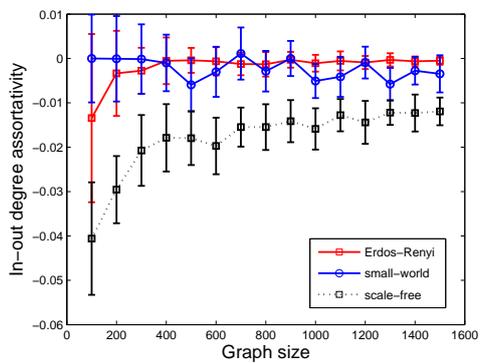
(d) In-out degree assortativity coefficients
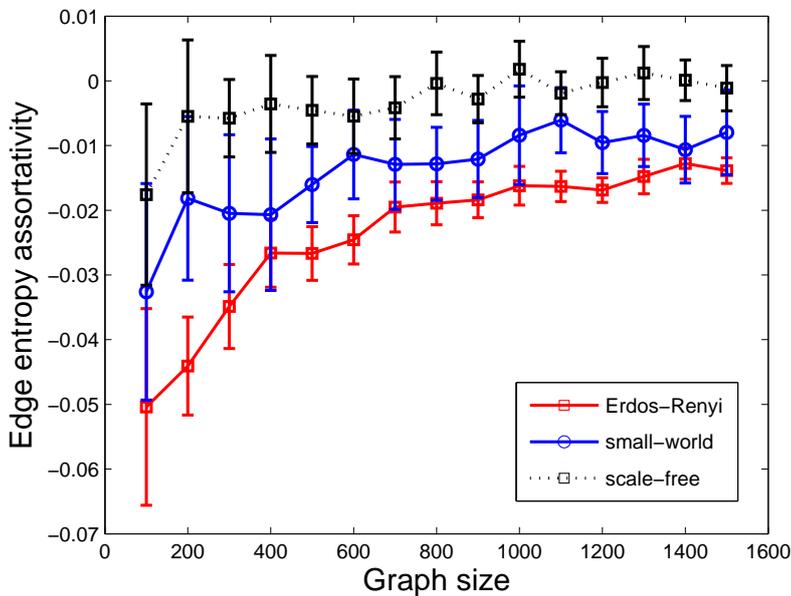
(e) Entropic edge assortativity measures

Figure 4.4: Mean and standard deviation of degree and entropic edge assortativity measures for different models of directed graphs.

Datasets	MUTAG	NCI1	NCI109	D&D
Maximal vertices #	28	111	111	5748
Minimal vertices #	10	3	4	30
Average vertices #	17.9	29.9	29.7	284.3
Graph #	188	4110	4127	1178

Table 4.3: Detailed information of the bioinformatics graph database used in the experiments.

maximum variation in entropy contribution with vertex degree. This is an effective form of ECA since the directions of maximum entropy contribution variation correspond to those where there is maximum variation in degree structure for the edges. Specifically, we apply our method to some real-world bioinformatics database in order to extract the corresponding feature vectors and thus obtain the graph classification results, using the support vector machine (SVM) classifier. We compare the performance with that of several other alternative state-of-the-art approaches proposed in recent literature.

In this subsection, we use the following datasets: *MUTAG Dataset*, *NCI1 and NCI109 Dataset* and *D&D Dataset*, which are extracted from chemical compounds and protein data. Table 4.3 gives detailed information of each dataset, including the number of graphs, the maximal and minimal number of vertices in a graph.

We first investigate how the choice of the number of quantiles m affects the classification performance and the computational complexity of our proposed embedding method on the graphs extracted from the real-world bioinformatics database. To this end, we vary the quantile number m, and for each m, we use our embedding method to extract the corresponding feature vectors. We then report the corresponding classification accuracy and the computation time on the data, using standard vector-based classification algorithms.

In the following evaluation, for our embedding method and all other alternative state of the art approaches, we perform the 10-fold cross-validation using support vector machine classifier associated with the sequential minimal optimization (SMO) [102] and the Pearson VII universal kernel (PUK). All the SMO-SVM parameters are optimized on a Weka platform, and all the experiments are performed on an Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz processor, with 8 GB memory.

In Fig. 4.5(a) we report the average classification rates of 10 runs for SMO-SVM classifier as a function of quantile number m on the *NCI1 and NCI109 Dataset*, while Fig.

(a) Average classification rates　　　　　(b) Experimental runtime
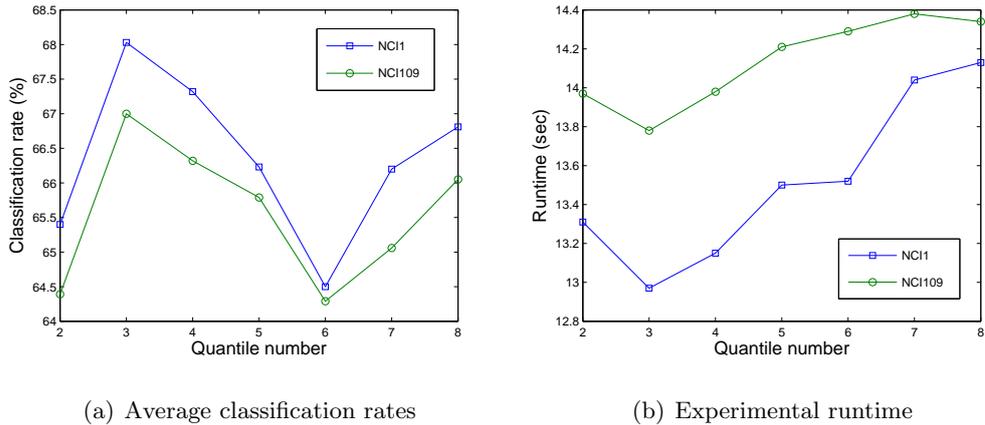
Figure 4.5: Graph classification performance using entropic graph embedding on the *NCI1 and NCI109 Dataset* as a function of quantile number.

4.5(b) gives the relationship between the average computation runtime of the experiments and the number of quantiles on these datasets.

From Fig. 4.5(b), the experimental runtime for both datasets basically has a growing trend as the quantile number m increases, which is as expected since greater quantile number leads to greater size of the feature vector, resulting in the greater computational complexity. Moreover, the entropic embedding method is computationally tractable as the runtime does not increase rapidly even when the size of the feature vector becomes particularly large.

Turning attention to the classication results reported in Fig. 4.5(a), the performance is good on the two subsets in the *NCI1 and NCI109 Dataset*, with the highest classication accuracy over 68% and 67% respectively. Moreover, as the increase of the quantile number, the classication rates on both datasets witness a slight growth, reaching a peak when the quantile number reaches 3, then they drop signicantly. This shows that our method gives the best classication performance when $m = 3$. Furthermore, with this choice of quantile number, the experimental runtime is relatively low, which suggests that our method can achieve a sufficient accuracy without causing expensive computation.

We then fix the number of quantiles to $m = 3$ and show the comparison results between our method and several alternative state of the art learning methods. These methods include three feature-vector-based graph embedding methods, namely a) the coefficient feature vector from Ihara zeta function for graphs (CI) [106], b) graph features from topological and label attributes (GF) [82], and c) the discriminative prototype selection

Datasets	MUTAG	NCI1	NCI109	D&D
CI	80.85	60.05	62.79	DNF
GF	86.57	65.81	65.30	69.92
DP	75.61	60.93	60.23	63.19
RW	87.01	DNF	DNF	DNF
WL	84.57	73.00	73.28	75.63
GC	84.04	67.71	67.32	77.33
ED	**86.18**	**68.05**	**67.04**	**75.39**

Table 4.4: Comparison of graph classification results on bioinformatics graph database (accuracy unit is %).

embedding method (DP) [24] and three graph kernel methods, namely a) the random walk graph kernel (RW) [71], b) the Weisfeiler-Lehman subtree graph kernel (WL) [115] and c) the graphlet count graph kernel (GC) [116].

In the following analysis, the entropic embedding method is denoted as ED (entropy distribution). For the graph embedding methods CI, GF, DP and ED we perform PCA on their corresponding feature vectors to obtain graph features. On the other hand, for graph kernel methods RW, WL and GC, we first compute their kernel matrix, then perform kernel PCA on this matrix in order to embed a graph into a principal component feature space. This allows us to employ any standard machine learning algorithm for graph classification.

For performance comparison, we report the average classification accuracy for each method over the 10-fold cross-validation run 10 times. We also give an evaluation of the runtime of each method, which exclude the time for SVM training since all methods employ the same SMO-SVM classifier. Moreover, In Table 4.5, the accuracy unit is percentage while in Table 4.5, the runtime unit is second and minute, in both tables, the "DNF" in any cell indicates that the computation did not finish (DNF) within a sufficient long time (12 hours, in this experiment) due to the large computational complexity, so the experiment run is aborted.

Table 4.4 shows the accuracy comparison for our proposed method (ED) versus other alternative graph embedding methods and graph kernel methods. The corresponding runtime results are shown in Table 4.5. Comparing to graph embedding methods (CI, GF and DP), the entropic embedding method ED gives the best classification performance. Specifically, on the *NCI1 and NCI109 Dataset* and *D&D Dataset*, ED has the highest

Datasets	MUTAG	NCI1	NCI109	D&D
CI	< 1 s	23 s	23 s	DNF
GF	< 1 s	30 s	30 s	$\approx$ 40 m
DP	< 1 s	55 s	56 s	$\approx$ 60 m
RW	5 s	DNF	DNF	DNF
WL	< 1 s	$\approx$ 2 m	$\approx$ 2 m	$\approx$ 10 m
GC	2 s	$\approx$ 1 m	$\approx$ 1 m	$\approx$ 20 m
ED	**<1 s**	**13 s**	**13 s**	**$\approx$8 m**

Table 4.5: Comparison of experimental runtime on bioinformatics graph database (time unit is minute and second).

classification rates among all graph embedding methods. On the *MUTAG Dataset*, ED also gives the second highest classification rate, only slightly lower than that of GF.

For graph kernel methods, although their classification rates are significantly high on the *NCI1 and NCI109 Dataset*, our embedding method still gives a competitive performance. Furthermore, we note that the random walk graph kernel (RW) did not finish experiments on three datasets, due to one of the main drawbacks of graph kernel methods, namely they tend to have a high computational overhead.

From Table 4.5, we see in general, the graph embedding methods have faster runtime than graph kernel methods. Another interesting feature is that two of the methods did not finish computation within a sufficiently long time on the *D&D Dataset*, this is mainly because some graphs in this dataset have more than 5000 vertices (according to Table 4.3). As a result, the graph embedding and graph kernel methods which rely on the computation over the number of vertices and edges have a significant amount of computational complexity, leading to the infinite experimental runtime. However, the entropic embedding method ED shows a good runtime performance as it finished experimental computation on all four datasets with lower runtime, especially on the *D&D Dataset*. Therefore, from the experimental evaluation, the graph embedding method associated with the entropy distribution has proved to outperform some state of the art methods, as it gives both accurate and computationally efficient graph classification performance.

4.3.3 Directed Graph Embedding

Next we evaluate the experimental performance of the entropic directed graph characterization. Specifically, we first explore the graph clustering performance of the method on a set of random graphs generated from three classical random graph models, which is the *Random Directed Graph Dataset* we have used in Chapter 3. Then we apply our method to some real-world data and report the graph classification results, including the *Protein Dataset* that we have studied in Chapter 3, and the *COIL Dataset*.

To investigate the classification performance of our proposed directed graph characterization, we perform PCA on both FF feature vectors and SD feature vectors extracted from the randomly generated graphs in the *Random Directed Graph Dataset*. These feature vectors are long-vectors formed by concatenating the elements of the four and three-dimensional histograms respectively. Here we select different parameter settings to generate 500 normal directed graphs and 500 additional strongly directed graphs for each of the three random graph models, with graph size ranged between 100 and 150. Moreover, in all the experiments in this section, we choose the number of quantiles $m = 3$, giving all the FF feature vectors with a constant length $m^4 = 81$, while for SD feature vectors, the length is $m^3 = 27$.

Figures 4.6(a), 4.6(c) and 4.6(d) each show that by embedding different random graphs into a feature space spanned by the first three principal components constructed from the feature vectors, the three classes of random graphs display some clear separation between each other. However in Fig. 4.6(b), which is the plot of SD feature vectors extracted from normal directed graphs, the "small-world" graphs and the Barabási-Albert graphs show some overlap. This suggests the FF feature vectors are efficient in distinguishing any normal directed graphs while the SD feature vectors are effective only for strongly directed graphs, which is an expected result. Therefore in the following experiments we use the FF feature vectors in our analysis.

To take this analysis a step further, we evaluate the classification performance of en entropic graph embedding method on the graphs in the *COIL Dataset* and *Protein Dataset*, using standard vector-based clustering and classification algorithms. In the following evaluation, we perform the 10-fold cross-validation using two classifiers, namely support vector machine (SVM) classifier associated with the sequential minimal optimization (SMO) [102] and the Pearson VII universal kernel (PUK), and k-nearest neighbour (k-NN) classifier. All the SMO-SVM and k-NN parameters are optimized for each method on a Weka plat-
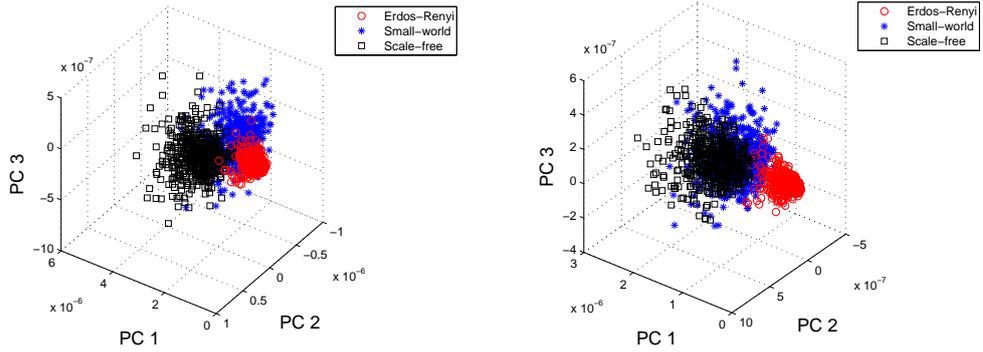
(a) FF feature vectors extracted from normal directed graphs

(b) SD feature vectors extracted from normal directed graphs
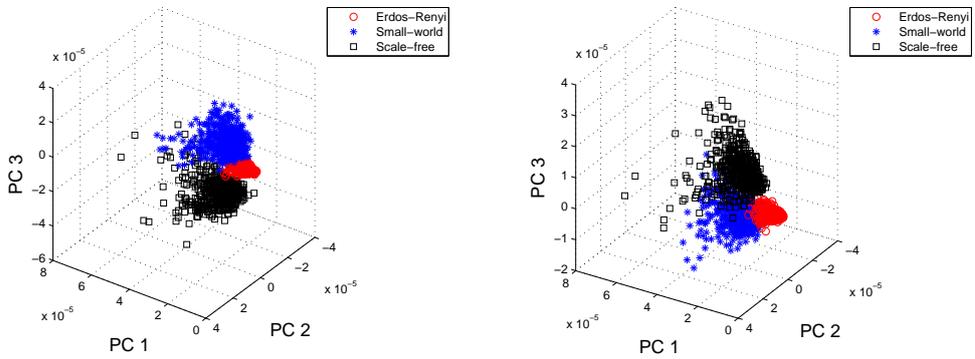
(c) SD feature vectors extracted from SD graphs

(d) FF feature vectors extracted from SD graphs

Figure 4.6: PCA plots of graph classification performance using entropic graph embedding on different models of directed graphs.

(a) 4-object data

(b) 8-object data

(c) Protein data
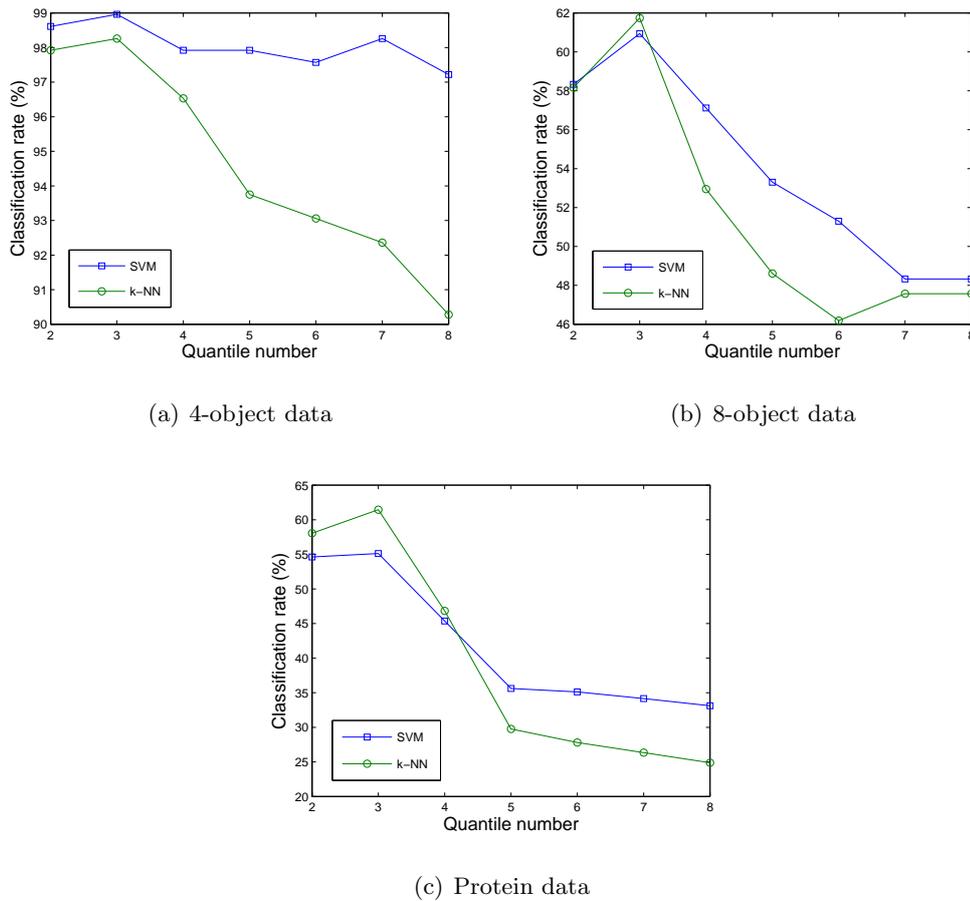
Figure 4.7: Average graph classification rates of entropic graph embedding with both SVM and k-NN classifiers on different datasets as a function of quantile number.

form, and all experiments are performed on an Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz processor, with 8 GB memory.

In Fig. 4.7 we report the average classification rates of 10 runs for both SVM and k-NN classifiers as a function of quantile number m on three different datasets, including the 4-object data and 8-object data in the *COIL Dataset* and *Protein Dataset*. Figure 4.8 gives the relationship between the average runtime and the quantile number of the experiments on these datasets.

From Fig. 4.8, the experimental runtime for all three classification problems grows as the quantile number increases, which is as expected since greater quantile number leads to greater size of the feature vector, resulting in the greater computational complexity. Moreover, the entropic directed graph characterization is computationally tractable as the runtime does not increase rapidly even when the size of the feature vector becomes
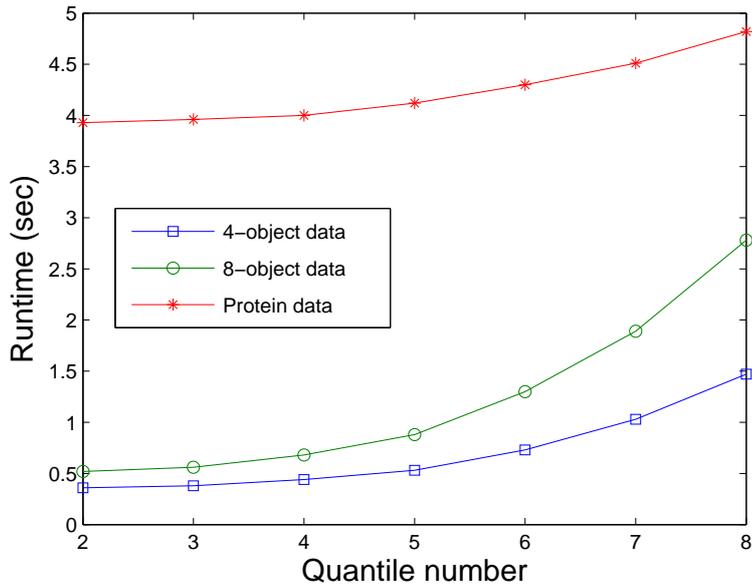
Figure 4.8: Average experimental runtime of entropic graph embedding with both SVM and k-NN classifiers as a function of quantile number.

particularly large.

Turning attention to the classification results reported in Fig. 4.7, the performance is particularly good on 4-object data, with a classification accuracy over 98%, and on 8-object data and 6-class protein database, the accuracy is still acceptable (50% to 60%). Moreover, as the increase of the quantile number, the classification rates for both classifiers on all three datasets witness a slight growth, reaching a peak when the quantile number reaches 3, then they drop significantly. This is because in the graphs of these datasets, all vertices have the same out-degree 3, therefore when $m = 3$ the corresponding feature vectors can precisely preserve the information of the vertex in and out-degree statistics, which guarantees that $m = 3$ gives the best classification performance and any greater quantile number will lead to a decrease of classification accuracy. Furthermore, with this choice of quantile number, the experimental runtime is relatively low, which suggests that our method can achieve a sufficient accuracy without causing expensive computation. Overall, based on these observations we claim that that our directed graph characterization can be both accurate and computationally efficient in clustering and classifying directed graphs when the appropriate parameters are selected.

106

4.4   Summary

It is interesting to note that the network von Neumann entropy analysis is posed at the global level, and does not consider in detail how the entropy is distributed across edges or other network substructures. The aim in this chapter is to provide a number of new characterization of network structure, which captures the distribution of entropy across the edges of a network.

First, we commence from the simplified approximations to the von Neumann entropy for both undirected and directed graphs, which are dependent on the graph size and degree statistics of vertices that are connected. From these approximations we then derive a local measure for quantifying the von Neumann entropy contribution for each edge in the undirected and directed graph respectively. This leads to the possibility of designing a correlation coefficient that measures the average assortative properties of how the entropy contributions that reside in edges are connected in a network, which we name the entropic edge assortativity measure. The resulting expressions for such measures of both undirected and directed graphs are simply related to some graph invariants, including the graph size, number of edges and the vertex degree.

The second contribution we have made in this chapter is to use von Neumann edge entropy distribution to perform network similarity comparisons and also to embed samples of networks into a low-dimensional space using PCA. Since our entropy is defined in terms of the two vertex degree values defining an undirected edge, and the two in-degree and out-degree combinations defining a directed edges, we can use histogram the edge entropy using a two-dimensional degree array for undirected networks and a four-dimensional array for directed networks. Each edge in a network increments the contents of the appropriate bin in the histogram, according to the degree pair in an undirected graph or the degree quadruple for a directed graph. We normalize the resulting histograms and vectorize them to give network feature vectors reflecting the distribution of entropy across the edges of the network. This provides a complexity level characterization of graph structure based on the statistical information residing edge degree distribution.

The work reported in this chapter can be extended in a number of ways. First, it would be interesting to explore how the distribution of the edge entropy contributions in a network can contribute to the development of novel information theoretic divergences, distance measures and relative entropies. Another interesting line of investigation would be to investigate whether this measure can be applied further to weighted graphs and

hypergraphs. In the future, we also intend to explore novel and effective graph kernels defined over the inner products of our entropy distribution feature vectors.

Chapter 5

Thermodynamic Characterization of Time Evolving Networks

In this chapter we present a novel method for characterizing the evolution of time-varying complex networks by adopting a thermodynamic representation of network structure computed from a polynomial (or algebraic) characterization of graph structure. Commencing from a representation of graph structure based on a characteristic polynomial computed from the normalized Laplacian matrix, we show how the polynomial is linked to the Boltzmann partition function of a network. This allows us to compute a number of thermodynamic quantities for the network, including the average energy and entropy. Assuming that the system does not change volume, we can also compute the temperature, defined as the rate of change of entropy with energy. All these thermodynamic variables can be approximated using low-order Taylor series that can be computed using the traces of powers of the normalized Laplacian matrix, avoiding explicit computation of the normalized Laplacian spectrum. These polynomial approximations allow a smoothed representation of the evolution of networks to be constructed in the thermodynamic space spanned by entropy, energy and temperature. We show how these thermodynamic variables can be computed in terms of simple network characteristics, e.g., the total number of vertices and vertex degree statistics for vertices connected by edges.

Another contribution we will make in this chapter is the development of a statistical thermodynamical framework for analyzing the time-varying networks, using ideas from statistical mechanics. To do this, we first explore whether the approximate von Neumann entropy can be used as a thermodynamic entropy of a network, when we associate the network microscopic configurations with the eigenstates of the normalized Laplacian spec-

trum. Then, we derive expressions for additional thermodynamic variables of networks, including the internal energy and temperature. In the experiments, we apply the resulting thermodynamic characterizations to real-world time-varying networks representing complex systems in the financial and biological domains. The study demonstrates that the methods provide an efficient tool for detecting abrupt changes and characterizing different time stages in network evolution.

5.1   Thermodynamic Characterization Using Graph Polynomials

The aim in this section is to establish a link between the characteristic polynomials from algebraic graph theory, and the thermodynamical analysis of networks. For an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, our characterization commences from the Boltzmann partition function $Z(\beta) = Tr[\exp\{-\beta\hat{H}\}]$ where $\hat{H}$ is the Hamiltonian associated with the graph and $\beta = 1/kT$ with k the Boltzmann constant and T the temperature. The Hamiltonian is the total energy operator, which can be defined in a number of ways. For instance, in quantum mechanics the choice dictated by the Schrödinger equation is $\hat{H} = -\nabla^2 + U$, where $-\nabla^2$ is the Laplacian and U the potential energy operator. For a graph, if we specify the vertex potential energy as the negative degree matrix, i.e., $U = -D$ and replace the Laplacian by its combinatorial counterpart $L = D - A$, where A is the adjacency matrix, then $\hat{H} = -A$. This choice of Hamiltonian is often used in the Hückel molecular orbital (HMO) method [36]. An alternative is to assume a graph is immersed in a heat bath with the eigenvalues of its normalized Laplacian matrix as the energy eigenstates. In this case, we set the potential energy operator $U(r,t)$ to zero, and can identify ∇^2 with the graph normalized Laplacian, i.e., $\hat{H} = -\tilde{L} = -D^{-1/2}(D - A)D^{-1/2}$.

With this choice of Hamiltonian and hence partition function, the energy associated with the graph is $E = -\partial \ln Z(\beta)/\partial\beta = -\sum_i p_i \tilde{\lambda}_i$, where $\tilde{\lambda}_i$ denote the eigenvalues of $\tilde{L}$ and $p_i = \exp\{\beta\tilde{\lambda}_i\}/\sum_i \exp\{\beta\tilde{\lambda}_i\}$, i.e., a weighted average of the normalized Laplacian eigenvalues, where the weights associated with the individual eigenvalues are determined by the Boltzmann occupation probabilities. The entropy is given by $S = k\{\ln Z(\beta) + \beta E\}$.

We characterize the graph using the quasi polynomial $R(\beta) = \det(I - \beta\tilde{L})$. We show in our analysis that $Z(\beta) \approx -\ln R(\beta) + |\mathcal{V}|$, where $|\mathcal{V}|$ is the graph size and as a result both the energy and entropy can be expanded as power series in β. The leading coefficients

of the two series are determined by the sum of the reciprocal of the degree products for vertices forming edges and triangles in the graph. The coefficients of the increasing powers of β depend on the frequencies of increasingly large substructure. The higher the degrees of the vertices forming these structures, the smaller the associated weight. Hence high degree structures are energetically more favourable than low degree ones (because they have lower reciprocal of the degree product). Also larger structures are energetically more favourable.

The expressions derived for energy and entropy of the network depend only on the assumed model for Hamiltonian of the system, and the approximations needed to express the partition function in terms of the characteristic polynomial associated with the normalized Laplacian of the network. Hence the energy and entropy can be used as a characterization of structure for any set of networks. However, in our experiments we study the time evolution of networks with fixed numbers of vertices. This is not an entirely uncommon situation, and arises where networks are used to abstract systems with a known set of states or components. In the financial network example, the vertices are stock traded over a 6000-day period, and in the second example the vertices represent genes expressed by fruit flies at different stages in their development. In this set-up we require a natural way of measuring fluctuations in network structure with time.

For a thermodynamic system with freedom to vary its volume, temperature and pressure, the change in internal energy is given by $dE = TdS - P_r dV_o + mdN$ where T is the temperature, P_r the pressure, dV_o the change in volume, m the particle mass and dN the change in the number of particles. When the number of particles and volume are fixed, we have an isochoric process, and the temperature is the rate of change of energy with entropy. With the expressions for these two quantities derived from the partition function, the isochoric temperature is also determined by a simple expression involving the frequencies of edges and triangles of different degree configurations. One way to picture this system is a thermal distribution across the energy states corresponding to the normalized Laplacian eigenvalues. Large changes in temperature are hence associated with a) large changes in the number of triangles compared to the number of edges, and b) when the average degree of the vertices changes significantly. Hence the temperature fluctuation between graphs in a sequence is sensitive to changes in internal structure of the network. We show that our method in fact smooths the time dependence of the thermodynamic characterization, so we present the global thermodynamic analysis in a computationally efficient and tractable

111

way.

So, to summarize we present a method motivated by thermodynamics for characterizing time sequences of networks. Although it is not a model of network evolution, it may provide the building blocks for such model. The approach has some similarities to that reported by Javarone and Armano [67] who use the classical limits of quantum models of gases as analogues to analyze complex networks. However, rather than using the classical Boltzmann distribution and the normalized Laplacian characteristic polynomial as the basis of their model, they base their model on a fermionic system. Finally, we note that the notion of temperature used in our work is not the physical temperature of the system, but a means of gauging fluctuations in network structure with time.

In this section, we provide a detailed development of how we compute thermodynamic quantities for a network, including the thermodynamic entropy, average energy and temperature, commencing from a characteristic polynomial representation of network structure. First, we provide some preliminaries on how graphs can be represented using the normalized Laplacian matrix. We then explain how the Boltzmann partition function can be used to describe the thermalization of the population of the energy microstates of network as represented by its Hamiltonian. The key step in establishing our thermodynamic characterization of network evolution, is to show a relationship between the partition function and the characteristic polynomial for the network. Normally, the thermalization process arises via the analogy of emersing the network in heat bath, with the adjacency matrix eigenvalues playing the role of energy eigenstates and the thermal population of the energy levels being controlled by the Boltzmann distribution. Here we aim to make a connection between the heat bath analogy and an alternative graph representation based on a characteristic polynomial. This is a powerful approach since there are several alternative matrix representations of graphs, and their characteristic polynomials together with the closely related zeta function representations have been extensively studied in graph theory [113] [106] [105]. Our approach therefore allows these potentially rich representations to be investigated from the thermodynamic perspective. Specifically, we show how the partition function can be approximated by the characteristic polynomial associated with the normalized Laplacian matrix for the network. This picture of the heat bath emerges when the Hamiltonian is the negative normalized Laplacian. From this starting point and using the network partition function approximation, we derive the expressions for the network average energy and entropy, and under the assumption of constant volume determine

112

the network temperature by measuring fluctuations in entropy and average energy. We show for networks of approximately constant size, each of these thermodynamic quantities can be computed using simple network statistics, including the number of vertices and vertex degree statistics.

5.1.1 Boltzmann Partition Function

In statistical mechanics, the canonical partition function associated with the Boltzmann factor of a system is

$$Z = \sum_i e^{-\beta E_i},$$

where $\beta = 1/kT$ is proportional to the reciprocal of the temperature T with k the Boltzmann constant, and E_i denotes the total energy of the system when it is in microstate i. Moreover, the partition function can be formalized as a trace over the state space:

$$Z(\beta) = Tr[\exp\{-\beta \hat{H}\}],$$

where $\hat{H}$ is the Hamiltonian operator and $\exp\{\cdots\}$ represents the matrix exponential.

The Hamiltonian operator of a graph may be defined in a number of ways. In quantum mechanics, one choice dictated by the Schrödinger equation is

$$\hat{H} = -\nabla^2 + U.$$

If we set the potential energy operator U to zero, we can identify $-\nabla^2$ with the graph Laplacian in either its combinatorial or normalized form. With this choice we obtain

$$\hat{H} = L$$

or

$$\hat{H} = \tilde{L}. \tag{5.1}$$

Alternatively, we can specify the vertex potential energy operator as the negative degree matrix, i.e., $U = -D$, with the result that

$$\hat{H} = -A.$$

This choice of Hamiltonian is often used in Hückel molecular orbital (HMO) method [36]. Generally, in this case $\hat{H} = c_1 I + c_2 A$ where A is the adjacency matrix of a graph representing the carbon skeleton of the molecule and c_1, c_2 are constants.

In our analysis we let the Hamiltonian operator $\hat{H} = -\tilde{L}$, which is the opposite as shown in Eq. (5.1). As a result, the Boltzmann partition function takes the form

$$Z(\beta) = Tr[\exp\{\beta\tilde{L}\}]. \tag{5.2}$$

Although most of the aggregate thermodynamic variables of the system, such as the average energy and entropy, can be expressed in terms of the partition function or its derivatives, deriving expressions for these variables directly from Eq. (5.2) can be computationally difficult. A more convenient route is to adopt an alternative graph representation based on a characteristic polynomial. In this way we approximate the Boltzmann partition function, so that the computation for thermodynamic variables can be simplified.

It is important to stress that making use of the statistical mechanical analysis usually requires a specification of the microscopic configurations of a thermodynamic system together with a clear physical interpretation of their meaning. In this section, we do not dwell on the microstates of the thermodynamic system arise or how they are populated. Briefly, our Hamiltonian is the negative of the normalized Laplacian, and one physical interpretation of our model would be of a graph immersed in a heat bath with the normalized Laplacian eigenvalues as energy eigenstates. The graph is subject to thermalization via the Boltzmann distribution. Our main concern is though to understand how to approximate the partition function of the resulting system so as to render thermodynamic analysis tractable. Although we do define a Hamiltonian for the system, our basic representation of the graph is in terms of the characteristic polynomial. We show how the characteristic polynomial can be used to approximate the Boltzmann partition function when the graph is immersed in a heat bath. Here the polynomial coefficients are themselves symmetric polynomials of the normalized Laplacian eigenvalues, and the polynomial variable is linked to the temperature of the heat bath. As we will show in our experiments, this approximation effectively smooths the time dependence of the network evolution, by allowing the thermodynamic variables to be approximated by low-order polynomials.

5.1.2 Characteristic Polynomial of Normalized Laplacian Matrix

The characteristic polynomial of the normalized Laplacian matrix $\tilde{L}$ of a graph, denoted by $f_c(x)$, is the polynomial defined by

$$f_c(x) = \det(xI - \tilde{L}),$$

where I indicates the identity matrix and x is the polynomial variable.

At this point it is worth noting that polynomial characterizations are also central to the definition of various types of zeta function of a graph. For instance, the determinant expression for the reciprocal of the Ihara zeta functions of a graph [106] is

$$\zeta^{-1}(x) = \det(I - xB)$$

where B is the Hashimoto's edge adjacency operator on the oriented line graph of graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. By replacing the Hashimoto operator with the normalized Laplacian operator $B = \tilde{L}$, we immediately obtain

$$\zeta^{-1}(x) = \det(I - x\tilde{L}).$$

Therefore the characteristic polynomial of the normalized Laplacian matrix and the above zeta function of graph G are related by

$$f_c(x) = x^{|\mathcal{V}|} \det(I - \frac{1}{x}\tilde{L}) = x^{|\mathcal{V}|}\zeta^{-1}(\frac{1}{x})$$

where $|\mathcal{V}|$ is the number of vertices in the graph.

Here we use $R(x)$ to denote the inverse zeta function determinant $\det(I - \frac{1}{x}\tilde{L})$ and refer to it as the quasi characteristic polynomial of the normalized Laplacian matrix. To show that $R(x)$ can be employed as an efficient tool for approximating the Boltzmann partition function in Eq. (5.2), we first note that for a square matrix M, the determinant can be calculated by

$$\det(M) = \exp\{Tr[\ln M]\}.$$

Thus, we have

$$R(x) = \exp\{Tr[\ln(I - \frac{1}{x}\tilde{L})]\}.$$

Recalling the classical Mercator series for the matrix logarithm of $I + M$

$$\ln(I + M) = M - \frac{M^2}{2} + \frac{M^3}{3} - \cdots, \quad \rho(M) < 1,$$

where $\rho(M)$ indicates the spectral radius of M, which is equal to the largest absolute value of the eigenvalues of M. Since the normalized Laplacian matrix has eigenvalues between 0 and 2 [30], the matrix Mercator series holds if and only if $\rho(\frac{1}{x}\tilde{L}) < 1$, i.e., $|\frac{1}{x}| < \frac{1}{2}$.

To develop these ideas a step further, if we let $\frac{1}{x} = \beta$, the quasi characteristic polynomial of the normalized Laplacian matrix can then be expressed as

$$R(\beta) = \exp\{Tr[-\beta\tilde{L} - \frac{1}{2}\beta^2\tilde{L}^2 - \frac{1}{3}\beta^3\tilde{L}^3 - \cdots]\}. \tag{5.3}$$

Moreover, using the first-order MacLaurin formula to expand the matrix exponential, i.e.,

$$\exp M = I + M + \frac{M^2}{2!} + \frac{M^3}{3!} + \cdots,$$

where M is an arbitrary square matrix, we can immediately re-write the Boltzmann partition function Eq. (5.2) in the following way:

$$Z(\beta) = Tr[I + \beta \tilde{L} + \frac{1}{2!}\beta^2 \tilde{L}^2 + \frac{1}{3!}\beta^3 \tilde{L}^3 + \cdots]. \tag{5.4}$$

By comparing the expressions in Eq. (5.3) and Eq. (5.4), the Boltzmann partition function can then be calculated from the quasi characteristic polynomial of the normalized Laplacian matrix as follows:

$$
\begin{aligned}
Z(\beta) &= Tr[I] + Tr[\beta \tilde{L} + \frac{1}{2!}\beta^2 \tilde{L}^2 + \cdots] \\
&= |\mathcal{V}| - \ln R(\beta) + \epsilon(\beta),
\end{aligned}
$$

where $\epsilon(\beta)$ denotes the residual. More explicitly, the residual is computed by

$$
\begin{aligned}
\epsilon(\beta) &= \sum_{n=3}^{\infty} (\frac{1}{n!} - \frac{1}{n})\beta^n Tr[\tilde{L}^n] \\
&= -\sum_{n=3}^{\infty} \frac{\beta^n}{n} \left[1 - \frac{1}{(n-1)!}\right] Tr[\tilde{L}^n] \\
&= -\frac{\beta^3}{6} Tr[\tilde{L}^3] - \frac{5\beta^4}{24} Tr[\tilde{L}^4] - \cdots.
\end{aligned}
$$

As a result, when $|\beta|$ takes on a small value, we have

$$\lim_{\beta \to 0} \frac{\epsilon(\beta)}{\ln R(\beta)} = 0,$$

i.e., $r(\beta) = o[\ln R(\beta)]$. This implies that the partition function is approximately equal to the negative of natural logarithm of the quasi characteristic polynomial plus a constant:

$$Z(\beta) \approx -\ln R(\beta) + |\mathcal{V}|. \tag{5.5}$$

To conclude this subsection, it is worth discussing the validity of the above approximation. We have shown that the requirements a) $|\beta| < \frac{1}{2}$ and b) $\epsilon(\beta) = o[\ln R(\beta)]$ are essential to making this approximation valid, which implies that the value of β must be small. Later we will provide an empirical analysis showing that this condition is well satisfied for a number of real-world complex networks.

5.1.3 Thermodynamic Variables of Complex Networks

For thermodynamics, a thermodynamic state of a system can be fully described by an appropriate set of principal parameters known as thermodynamic variables. These include the average energy, entropy and temperature. In this subsection, we give a detailed development showing how these thermodynamic state variables are derived from the approximate partition function and how they can be computed via simple network statistics.

To commence, we recall that given a partition function $Z(\beta)$, the average energy E of a system $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is obtained by taking the partial derivative of the logarithm of the partition function with respect to β, i.e.,

$$E(\mathcal{G}) = -\frac{\partial \ln Z(\beta)}{\partial \beta}. \tag{5.6}$$

Moreover, the thermodynamic entropy S is obtained by

$$S(\mathcal{G}) = k\{\ln Z(\beta) + \beta E(\mathcal{G})\}, \tag{5.7}$$

where k denotes the Boltzmann constant.

The thermodynamic temperature T, measures fluctuations in network structure with time. More specifically, suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ represent the structure of a time-varying system at two consecutive epochs t_1 and t_2 respectively. For a thermodynamic system of constant number of particles, we recall the fundamental thermodynamic relation $dE = TdS - P_r dV_o$, where P_r and V_o denote the pressure and volume respectively. The volume is a concept generally considered in the context of ideal gases and many thermodynamic processes could result in a change in volume. Here we consider the network under study $\mathcal{G}$ as a closed system and from $\mathcal{G}_1$ to $\mathcal{G}_2$ it undergoes a constant-volume process (isochoric process) during which the system volume remains constant.

It is important to stress that this equation holds and is valid for both reversible and irreversible processes for a closed system, since E, T, S, P_r and V_o are all state functions and are independent of thermodynamic path. As a result, for the path from $\mathcal{G}_1$ to $\mathcal{G}_2$ we have $dV_o = 0$ and $dE = TdS$. For example, when an ideal gas undergoes an isochoric process, and the quantity of gas remains constant, then the energy increment is proportional to the increase in temperature and pressure. As a result, the reciprocal of the temperature T is the rate of change of entropy with average energy, subject to the condition that the volume and number of particles are held constant, i.e.,

$$\frac{1}{T(\mathcal{G}_1, \mathcal{G}_2)} = \frac{dS}{dE} = \frac{S_1 - S_2}{E_1 - E_2}. \tag{5.8}$$

This definition can be applied to evolving complex networks which do not change significantly in size during their evolution.

To further develop the temperature expression, we first compute the change in entropy

$$
\begin{aligned}
S_1 - S_2 &= k\{\ln Z_1(\beta) + \beta E_1(\mathcal{G})\} - k\{\ln Z_2(\beta) + \beta E_2(\mathcal{G})\} \\
&= k\{\ln \frac{Z_1}{Z_2} + \beta(E_1 - E_2)\}.
\end{aligned}
$$

Note, that in our development, the partition function is approximated by $Z(\beta) \approx -\ln R(\beta) + |\mathcal{V}|$. Therefore, we have

$$
\begin{aligned}
\ln \frac{Z_1}{Z_2} &\approx \ln \frac{|\mathcal{V}| - \ln R_1}{|\mathcal{V}| - \ln R_2} \\
&= \ln |\mathcal{V}| + \ln(1 - \frac{1}{|\mathcal{V}|} \ln R_1) - \ln |\mathcal{V}| - \ln(1 - \frac{1}{|\mathcal{V}|} \ln R_2) \\
&= \ln(1 - \frac{1}{|\mathcal{V}|} \ln R_1) - \ln(1 - \frac{1}{|\mathcal{V}|} \ln R_2).
\end{aligned}
$$

The term $\frac{1}{|\mathcal{V}|} \ln R$ is close to zero since we assume that $|\beta|$ is small. As a result, using the Mercator series, we obtain $\ln(1 - \frac{1}{|\mathcal{V}|} \ln R) \approx -\frac{1}{|\mathcal{V}|} \ln R$, leading to the result that

$$
\begin{aligned}
\ln \frac{Z_1}{Z_2} &\approx -\frac{1}{|\mathcal{V}|} \ln R_1 + \frac{1}{|\mathcal{V}|} \ln R_2 \\
&= \frac{1}{|\mathcal{V}|} \ln \frac{R_2}{R_1} = \frac{1}{|\mathcal{V}|} \ln(1 + \frac{R_2 - R_1}{R_1}) \\
&\approx \frac{1}{|\mathcal{V}|} \cdot \frac{R_2 - R_1}{R_1},
\end{aligned}
$$

where $R_2 - R_1$ is the difference between the values for the quasi characteristic polynomial $R(\beta)$ at times t_1 and t_2.

Next, we calculate the energy

$$
\begin{aligned}
E(\mathcal{G}) &\approx -\frac{\partial \ln(|\mathcal{V}| - \ln R)}{\partial \beta} \\
&= -\frac{1}{|\mathcal{V}| - \ln R} \cdot \frac{\partial(|\mathcal{V}| - \ln R)}{\partial \beta} \\
&= \frac{1}{|\mathcal{V}| - \ln R} \cdot \frac{\partial \ln R}{\partial \beta} \\
&= -\frac{1}{|\mathcal{V}| - \ln R} \cdot \sum_{n=1}^{\infty} \beta^{n-1} Tr[\tilde{L}^n].
\end{aligned}
$$

Since the value for β is always small, then $\ln R(\beta) \ll |\mathcal{V}|$, and as a result the average energy expression is

$$
E(\mathcal{G}) = -\frac{1}{|\mathcal{V}|} \sum_{n=1}^{\infty} \beta^{n-1} Tr[\tilde{L}^n]. \tag{5.9}
$$

As a result, the difference between network energy E at times t_1 and t_2, is

$$E(\mathcal{G}_1) - E(\mathcal{G}_2) = E_1 - E_2 = -\frac{1}{|\mathcal{V}|}[f_{p,1}(\beta) - f_{p,2}(\beta)] \qquad (5.10)$$

where the polynomial function $f_p(\beta) = \sum_{n=1}^{\infty} \beta^{n-1} Tr[\tilde{L}^n]$.

Then, we compute the temperature using Eq. (5.8), with the result that

$$\frac{1}{T(\mathcal{G}_1, \mathcal{G}_2)} = \frac{k\{\ln \frac{Z_1}{Z_2} + \beta(E_1 - E_2)\}}{E_1 - E_2}$$

$$\approx k\beta - k \cdot \frac{\frac{R_2}{R_1} - 1}{f_{p,1} - f_{p,2}}.$$

Both the quasi characteristic polynomial $R(\beta)$ and the polynomial $f_p(\beta)$ can be expanded as power series, expressed as sums of traces of the powers of the normalized Laplacian matrix of the network. Expanding the two polynomials to third order requires the following traces:

$$Tr[\tilde{L}] = |\mathcal{V}|,$$

$$Tr[\tilde{L}^2] = |\mathcal{V}| + J,$$

$$Tr[\tilde{L}^3] = |\mathcal{V}| + 3J - K$$

where

$$J = \sum_{u,v} \frac{A_{uv}}{d_u d_v}$$

and

$$K = \sum_{u,v,w} \frac{A_{uv} A_{vw} A_{wu}}{d_u d_v d_w}$$

respectively [59] [128]. Expanding $R(\beta)$ to third order, we find

$$\frac{R_2}{R_1} = \frac{\exp\{Tr[-\beta\tilde{L}_2 - \frac{\beta^2}{2}\tilde{L}_2^2 - \frac{\beta^3}{3}\tilde{L}_2^3]\}}{\exp\{Tr[-\beta\tilde{L}_1 - \frac{\beta^2}{2}\tilde{L}_1^2 - \frac{\beta^3}{3}\tilde{L}_1^3]\}}$$

$$= \exp\left\{\beta(Tr[\tilde{L}_1] - Tr[\tilde{L}_2]) + \frac{\beta^2}{2}(Tr[\tilde{L}_1^2] - Tr[\tilde{L}_2^2]) + \frac{\beta^3}{3}(Tr[\tilde{L}_1^3] - Tr[\tilde{L}_2^3])\right\}$$

$$= \exp\left\{\frac{\beta^2}{2}(J_1 - J_2) + \frac{\beta^3}{3}[3(J_1 - J_2) - (K_1 - K_2)]\right\}.$$

Similarly, for $f_p(\beta)$ we obtain

$$f_{p,1} - f_{p,2} = \beta(J_1 - J_2) + \beta^2[3(J_1 - J_2) - (K_1 - K_2)].$$

As a result, the reciprocal of the temperature is given by

$$\frac{1}{T(\mathcal{G}_1, \mathcal{G}_2)} = k\beta + k \cdot \frac{1 - \exp\left\{\frac{\beta^2}{2}(J_1 - J_2) + \frac{\beta^3}{3}[3(J_1 - J_2) - (K_1 - K_2)]\right\}}{\beta(J_1 - J_2) + \beta^2[3(J_1 - J_2) - (K_1 - K_2)]}.$$

Since $T = 1/k\beta$, the second term on the right-hand side must vanish. As a consequence, we have that

$$\frac{\beta^2}{2}(J_1 - J_2) + \frac{\beta^3}{3}[3(J_1 - J_2) - (K_1 - K_2)] = 0.$$

First, when $J_1 - J_2 = K_1 - K_2 = 0$, i.e., graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are identical, $T = 1/k\beta$ holds. In other words, there are no structural differences between graphs $\mathcal{G}_1$ and $\mathcal{G}_2$. The second trivial solution is obtained by $\beta = 0$, implying that the temperature $T = 1/k\beta$ goes to infinity. Finally, the nontrivial solution is

$$\beta = -\frac{3(J_1 - J_2)}{6(J_1 - J_2) - 2(K_1 - K_2)},$$

which leads to the following expression for the temperature

$$T(\mathcal{G}_1, \mathcal{G}_2) = \frac{1}{k\beta} = -\frac{2}{k} + \frac{2}{3k} \cdot \frac{K_1 - K_2}{J_1 - J_2}. \tag{5.11}$$

Here $J_1 - J_2$ and $K_1 - K_2$ represent the change in quantities J and K when graph $\mathcal{G}_1$ evolves to $\mathcal{G}_2$ respectively:

$$J_1 - J_2 = \sum_{u_1, v_1 \in \mathcal{V}_1} \frac{A_{u_1 v_1}}{d_{u_1} d_{v_1}} - \sum_{u_2, v_2 \in \mathcal{V}_2} \frac{A_{u_2 v_2}}{d_{u_2} d_{v_2}}$$

$$K_1 - K_2 = \sum_{u_1, v_1, w_1 \in \mathcal{V}_1} \frac{A_{u_1 v_1} A_{v_1 w_1} A_{w_1 u_1}}{d_{u_1} d_{v_1} d_{w_1}} - \sum_{u_2, v_2, w_2 \in \mathcal{V}_2} \frac{A_{u_2 v_2} A_{v_2 w_2} A_{w_2 u_2}}{d_{u_2} d_{v_2} d_{w_2}}.$$

The temperature measures fluctuations in the internal structure of the time-evolving network, and depends on the ratio of total change of degree statistics for vertices that form triangles and for vertices connected by edges in the network. This is a direct consequence of the fact that we have truncated our series expansion of the partition function with third order. If we had continued the expansion to higher order, then the temperature would reflect this and contain terms in the numerator and denominator corresponding to changes in the number of cliques of size larger than three. By adjusting temperature in this way, we take account of fluctuations from the expected value of temperature $T = 1/k\beta$. When combined with the polynomial approach, this has the effect of smoothing the time dependence of the thermodynamic representation.

Finally, in order to calculate the network average energy, we substitute the obtained β into Eq. (5.9) and again remove the terms that have powers larger than 3, with the result that

$$E(\mathcal{G}) = -\frac{1}{|\mathcal{V}|}[|\mathcal{V}| + \beta(|\mathcal{V}| + J) + \beta^2(|\mathcal{V}| + 3J - K)]. \tag{5.12}$$

Similarly, for the thermodynamic entropy, we have

$$
\begin{aligned}
S(\mathcal{G}) &= k\{\ln Z(\beta) + \beta E(\mathcal{G})\} \\
&\approx k\{\ln(|\mathcal{V}| - \ln R) + \beta E\} \\
&\approx k\left\{\ln|\mathcal{V}| - \frac{1}{|\mathcal{V}|}\ln R + \beta E\right\} \\
&= k\left\{\ln|\mathcal{V}| - \frac{1}{|\mathcal{V}|}\sum_{n=1}^{\infty}(1 - \frac{1}{n})\beta^n Tr[\tilde{L}^n]\right\},
\end{aligned}
$$

and expanding to third order,

$$
S(G) = k\ln|\mathcal{V}| - \frac{k}{|\mathcal{V}|}\left[\frac{\beta^2}{2}(|\mathcal{V}| + J) + \frac{2\beta^3}{3}(|\mathcal{V}| + 3J - K)\right]. \tag{5.13}
$$

In order to obtain a better understanding of these network thermodynamic measures, it is interesting to explore how the average energy and entropy are bounded for graphs of a particular size, and in particular which topologies give the maximum and minimum values of the energy and entropy (we consider connected graphs only).

From Eq. (5.12) and Eq. (5.13), when the quantity J is minimal and quantity K reaches its maximum value, both the energy and the entropy reach their maximum values. This occurs when each pair of graph vertices is connected by an edge, and this means that the graph is complete. On the other hand, when J and K respectively take on their maximum and minimum values, the energy and entropy reach their minimum values. This occurs when the structure is a string.

The maximum and minimum average energies and entropies corresponding to these cases are as follows. For a complete graph $\mathcal{K}_n$, in which each vertex has degree $n - 1$, we have that

$$
E(\mathcal{K}_n) = -\left[1 + \frac{n}{n-1}\beta + \frac{n^2}{(n-1)^2}\beta^2\right]
$$

and

$$
S(\mathcal{K}_n) = k\ln n - k\left[\frac{n}{2(n-1)}\beta^2 + \frac{2n^2}{3(n-1)^2}\beta^3\right].
$$

Turning our attention to the case of a string $\mathcal{P}_n$ $(n \geq 2)$, in which the two terminal vertices have degree 1 while the remainder have degree 2, we have that

$$
E(\mathcal{P}_n) = -\left[1 + \frac{3n+1}{2n}\beta + \frac{5n+3}{2n}\beta^2\right]
$$

and

$$
S(\mathcal{P}_n) = k\ln n - k\left[\frac{(3n+1)}{4n}\beta^2 + \frac{(5n+3)}{3n}\beta^3\right].
$$

As a result, the average energy and entropy of graphs with N vertices are bounded as follows:

$$-\left[1 + \frac{3N+1}{2N}\beta + \frac{5N+3}{2N}\beta^2\right] \leq E(\mathcal{G}) \leq -\left[1 + \frac{N}{N-1}\beta + \frac{N^2}{(N-1)^2}\beta^2\right]$$

$$k\ln N - k\left[\frac{(3N+1)\beta^2}{4N} + \frac{(5N+3)\beta^3}{3N}\right] \leq S(\mathcal{G}) \leq k\ln N - k\left[\frac{N\beta^2}{2(N-1)} + \frac{2N^2\beta^3}{3(N-1)^2}\right]$$

where the lower bounds are achieved by strings, while the upper bounds are obtained for complete graphs.

There are a number of points to note concerning the development above. One of the most fundamental aspects of the presented thermodynamic measurements is the interplay between quantities J and K. The first represents the direct connections of vertices (also known as generalized Randić indices [29]), while the second is related to the number of triangles. Both measurements are weighted by their joint degrees.

To provide a deeper intuition concerning the physical meaning of our thermodynamic analysis in terms of changes in graph structure, we provide some examples. We commence by considering a regular graph with n vertices in which each vertex has the same degree m ($n \cdot m$ must be an even number). In this case, the quantity J is the sum of existing edges weighted by the network average degree m:

$$J_{reg} = \frac{n}{m}.$$

This result holds for both trees and cyclic multi-dimensional lattices. On the other hand, the calculation of K depends on the nature of the connections for the regular networks. For lattices connecting vertices at distance $\delta = 1$ (first neighbourhood) and for all trees, $K_{reg} = 0$ (since there are no triangles). For other regular networks the value of K depends on the number of triangles in the network ν, i.e.,

$$K_{reg} = \frac{6\nu}{m^3}.$$

The multiplicative factor 6 is needed as the summation in the equation of K considers each edge (u, v) two times, also because the summation is taken over all edges, and each triangle is counted three times. Moreover, when the regular network is a lattice of neighbourhood distance $\delta \geq 2$,

$$K_{reg}(\delta) = \frac{2N\nu(\delta)}{m^3},$$

where $\nu(\delta)$ is the number of triangles of each repeated element. Finally, for the cyclic 1D-lattice with connection distance δ, the number of triangles each vertex participates is

122

given by $\nu(\delta) = 3(\delta - 1)[(\delta - 1) + 1]/2 = 3\delta(\delta - 1)/2$, the average degree is $m = 2\delta$, thus the quantity K is evaluated as follows:

$$K_{lat-1D}(\delta) = \frac{3(\delta - 1)}{8\delta^3}.$$

As noted earlier, this analysis is based on a power series expansion of the partition function up to the third order. Clearly, to develop a realistic thermodynamic model for structures in which triangles are absent by reason of construction, then the expansion should be taken to higher order. Unfortunately, this renders analysis of the traces appearing in the partition function in terms of degree statistics intractable [59] [128]. An alternative would be to use the Ihara zeta function [113] as a network characterization. Here the underlying characteristic is computed from the adjacency matrix of the oriented line graph for a network. The polynomial coefficients are related to the numbers of prime cycles of varying length in a network [106].

To summarize, in this section we have proposed a novel method for characterizing the evolution of complex networks by employing thermodynamic variables. Specifically, we commence from a quasi characteristic polynomial of the normalized Laplacian matrix of a network and show this polynomial can be used as a tool for approximating the Boltzmann partition function on the network, when we identify Hamiltonian operator with the normalized Laplacian operator. Then, using the approximate network partition function, we develop the expressions for the network average energy and entropy. The thermodynamic temperature measures fluctuations via the changes in the connectivity pattern of the network, and is determined by the distribution of vertex degree. We show that these thermodynamic variables are expressed in terms of simple network features, including the number of vertices and the degree statistics for connected vertices.

5.2 Thermodynamic Characterization Using Statistical Mechanics

In this section, we provide a detailed development of expressions for network thermodynamic variables, including the entropy, internal energy and temperature. To do this, we first show that the approximate von Neumann entropy can be interpreted as the thermodynamic entropy of a network, when we associate the microscopic configurations of a network with the eigenstates of the normalized Laplacian spectrum. We then develop an

expression for network internal energy and determine the network temperature by measuring fluctuations in entropy and internal energy. Each of these thermodynamic variables can be computed using simple graph statistics, including the number of vertices and edges together with vertex degree statistics.

5.2.1 Thermodynamic Variables of Complex Networks

This section aims at developing a thermodynamic characterization of network structure. We commence by assuming that at any instant in time a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, is statistically distributed across an ensemble of $|\mathcal{V}|$ microstates. The probability that the system occupies a microstate indexed s is given by $p_s = \tilde{\lambda}_s / \sum_{s=1}^{|\mathcal{V}|} \tilde{\lambda}_s$, where $\tilde{\lambda}_s$, $s = 1, 2, \cdots, |\mathcal{V}|$ are the eigenvalues of the normalized Laplacian matrix of graph $\mathcal{G}$. Noting that the trace of a matrix is the sum of its eigenvalues, we have $\sum_{s=1}^{|\mathcal{V}|} \tilde{\lambda}_s = Tr[\tilde{L}] = |\mathcal{V}|$, so the microstate occupation probability is simply $p_s = \tilde{\lambda}_s / |\mathcal{V}|$.

We define the thermodynamic entropy of a network using the Shannon formula that is exclusively dependent on the probabilities of the microstates:

$$S = -k \sum_{s=1}^{|\mathcal{V}|} p_s \ln p_s = -k \sum_{s=1}^{|\mathcal{V}|} \frac{\tilde{\lambda}_s}{|\mathcal{V}|} \ln \frac{\tilde{\lambda}_s}{|\mathcal{V}|}, \tag{5.14}$$

where k is the Boltzmann constant and is set to be 1 to simplify matters.

It is clear that the thermodynamic entropy Eq. (5.14) and the undirected von Neumann entropy we have introduced in Eq. (3.1) take the same form. Both depend on the graph size and the eigenvalues of the normalized Laplacian matrix. It is reasonable to suggest that the von Neumann entropy can be interpreted as the thermodynamic entropy of a complex network. Therefore we use the approximate von Neumann entropy formula of undirected graphs as the network entropy:

$$S = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{|\mathcal{V}|^2} \sum_{(u,v) \in \mathcal{E}} \frac{1}{d_u d_v}.$$

The internal energy of a network is defined as the mean value of the total energy, i.e., the sum of all microstate energies, each weighted by its occupation probability:

$$E = \sum_{s=1}^{|\mathcal{V}|} p_s E_s, \tag{5.15}$$

where E_s is the energy of microstate s. Here we take the internal energy to be the total number of edges in the graph i.e., $E = |\mathcal{E}|$. From the properties of the Laplacian and

normalized Laplacian matrices, we have $|\mathcal{E}| = Tr[L] = Tr[D^{1/2}\tilde{L}D^{1/2}] = Tr[D\tilde{L}]$. This
can be achieved if we set the microstate energies to be $E_s = |\mathcal{V}| \cdot \tilde{\lambda}_s^{-1} \cdot d_s$, i.e., proportional
to the vertex degrees.

Suppose that graphs $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ represent the structure of a time-
varying complex network at two consecutive epochs t_1 and t_2 respectively. The reciprocal
of the thermodynamic temperature T is the rate of change of entropy with internal energy,
subject to the condition that the volume and number of particles are held constant, i.e.,
$1/T = dS/dE$. This definition can be applied to evolving complex networks which do not
change size during their evolution.

We write the change of the entropy S between graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ as

$$dS = S(\mathcal{G}_2) - S(\mathcal{G}_1) \approx \sum_{(u,v) \in \mathcal{E}_1, \mathcal{E}_2} \frac{d_u \Delta_v + d_v \Delta_u + \Delta_u \Delta_v}{d_u(d_u + \Delta_u)d_v(d_v + \Delta_v)},$$

where Δ_u is defined as the difference between degree of vertex u in graphs $\mathcal{G}_2$ and $\mathcal{G}_1$; Δ_v is
similarly defined. The change in internal energy, is equal to the change in the total number
of edges: $dE = E(\mathcal{G}_2) - E(\mathcal{G}_1) = |\mathcal{E}_2| - |\mathcal{E}_1| = \Delta|\mathcal{E}|$. Hence the reciprocal temperature T
is:

$$\frac{1}{T(\mathcal{G}_1, \mathcal{G}_2)} = \sum_{(u,v) \in \mathcal{E}_1, \mathcal{E}_2} \frac{d_u \Delta_v + d_v \Delta_u + \Delta_u \Delta_v}{\Delta|\mathcal{E}|d_u(d_u + \Delta_u)d_v(d_v + \Delta_v)}. \tag{5.16}$$

The temperature measures fluctuations in the internal structure of the time evolving net-
work, and depends on two properties of the network. The first is the overall or global
change of the number of edges, while the second property is a local one which measures
the change in degree for pairs of vertices connected by edges. Both quantities measure
fluctuations in network structure, but at different levels of detail. The temperature is
greatest when there are significant differences in the global number of edges, and smallest
when there are large local variations in edge structure which do not result in an overall
change in the number of edges.

To summarize, in this section we have proposed a novel method for characterizing the
evolution of complex networks by employing thermodynamic variables. Specifically, we
show that the network von Neumann entropy can be used as a thermodynamic charac-
terization, provided that the eigenstates of the normalized Laplacian matrix define the
network's microstates together with their occupation probabilities. Moreover, the internal
energy depends on the number of edges in the network. The thermodynamic temperature
measures fluctuations via the change in the number of edges and vertex degree changes.

5.2.2 Comparative Study of Evolving Network Analysis Methods

Finally, it is interesting to have a comparative analysis of our network thermodynamics methods and other state-of-the-art network evolution analysis approaches. Generally speaking, most existing methods for analyzing time evolution of complex networks have centered on studying structural measures of static networks, and then applying these quantities to each snapshot of the time-varying network in order to understand the evolutionary patterns. For instance, Holme et al. [63] have analyzed the time evolution of a number of well-known network features, including clustering coefficient, degree-degree correlations, average geodesic length and reciprocity of a large-scale online social network. Moreover, in the work [78], the authors have analyzed how the social networks of Flickr and Yahoo!360 evolve over different time periods using measures such as network density and average distance between vertices in the network components. Although such methods have proved to be efficient in reflecting the time evolution of some structural properties of evolving networks, they have a significant drawback, namely the lacking use of structure information between temporal networks at two consecutive time steps, e.g., the vertex degree change and edge number change.

In order to overcome this problem and to incorporate the missing structure information, a number of alternative techniques to capture the structure and evolution of networks have been proposed. For instance, Palla et al. [98] have developed a method for investigating the time dependence of the overlapping communities on a social network, using clique percolation method. Specifically, they take into consideration both the group size and age, and propose a measure for quantifying the relative overlap between two states of the same community at different time steps. Also they have developed a new network indicator called stationarity in order to quantify the changing rate of communities based on their size and age. In this way the authors have managed to exploit the community structure information between subsequent states of a time-evolving network.

Comparing to the existing evolving network analysis approaches, our thermodynamics analysis provides an advantageous approach in that the thermodynamic quantities, especailly the temperature, fully exploit the information related to the structural changes of networks at subsequent time steps. More importantly, our approach only uses a number of simple but important network characteristics, i.e., vertex degree statistics, edge number and degree information of some simple substructures such as triangles. This yields a low computational complexity to our thermodynamic analysis.

5.3 Experiments

We have derived two sets of expressions for the thermodynamic entropy, energy and temperature of time-evolving complex networks. In this section, we explore whether the resulting characterizations can be employed to provide a useful tool for better understanding the evolution of dynamic networks. Specifically, we aim at applying the novel thermodynamic method to a number of real-world time-evolving networks in order to analyze whether abrupt changes in structure or different stages in network evolution can be efficiently characterized. In this section, to simplify the calculation, we let the Boltzmann constant $k = 1$.

5.3.1 Thermodynamic Characterization From Graph Polynomial

We use two different datasets, *NYSE Stock Market Network Dataset* and *Drosophila Melanogaster Gene Network Dataset*, both of which are extracted from real-world complex systems.

We commence by examining whether the network Boltzmann partition function given in Eq. (5.2) is well approximated by the normalized Laplacian quasi characteristic polynomial Eq. (5.3), as expected from Eq. (5.5). To this end, we first create a large number of random graphs distributed according to two different models, namely a) the classical Erdős-Rényi model [43] and b) the Barabási-Albert model [13]. We randomly generate 500 graphs for each of the two models using a variety of model parameters. For instance, for the Erdős-Rényi model, the graph size is between 30 and 1000 and the connection probability is $p \in [0.1, 0.9]$; for the Barabási-Albert model, the graph size has the same range and the average vertex degree is bounded between 1 and 20. Then, for each random graph, we compute both the partition function $Z(\beta)$ and the quasi characteristic polynomial $-\ln R(\beta) + N$ for three different values of β. The result is shown as the scatter plot in Fig. 5.1.

The most striking feature in this figure is that although β takes on different values, the vast majority of the corresponding data points are close to the diagonal line $y = x$. This result empirically proves that the partition function $Z(\beta)$ is always very accurately approximated by the characteristic polynomial $-\ln R(\beta) + |\mathcal{V}|$ for different types of random graphs, as shown in Eq. (5.5).

Next, we investigate the relationship between the thermodynamic variables developed and the structural changes of networks. Specifically, we aim at exploring how the temper-
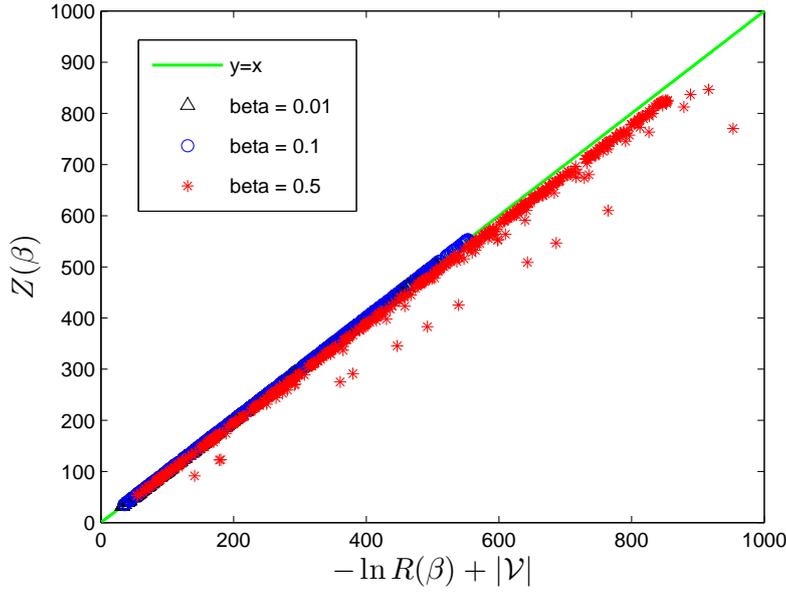
Figure 5.1: Scatter plot of Boltzmann partition function associated with normalized Laplacian operator and the normalized Laplacian quasi characteristic polynomial for different β for Erdős-Rényi and Barabási-Albert random graphs.

ature fluctuates when a graph experiences various degrees of evolutionary change. To this end, we commence by constructing a complete graph with 80 vertices, and randomly deleting its edges with a probability $p \in [0, 0.2]$. Then, we start from the same complete graph, and with probability $p + \Delta p$, we again delete edges in the graph randomly. Using these two random graphs, we compute the temperature according to Eq. (5.11). We repeat the process for different values of $\Delta p \in [0.1, 0.6]$ (100 realizations each), which indicate the different degrees of structural change during graph evolution. We then repeat the analysis for graphs with 150 vertices and 300 vertices respectively and produce a plot showing the mean and standard deviation (shown as error bar) of the temperature against Δp for a large number of random graphs with different sizes.

The most important feature in Fig. 5.2(a) is that as Δp increases, the mean values of the temperature for all three graph sizes grow. Moreover, the variance of temperature also increases gradually with the increase of Δp. This is because the variance of the ratio $(K_1 - K_2)/(J_1 - J_2)$ becomes large when there is a dramatic structural change in the time-evolving network, resulting in a significant change of the value of temperature. Moreover, when Δp remains small, the temperature remains relatively stable. This result agrees with expression for temperature in Eq. (5.11). Slight evolutionary changes lead to

(a) Edge deletion probability

(b) Degree difference
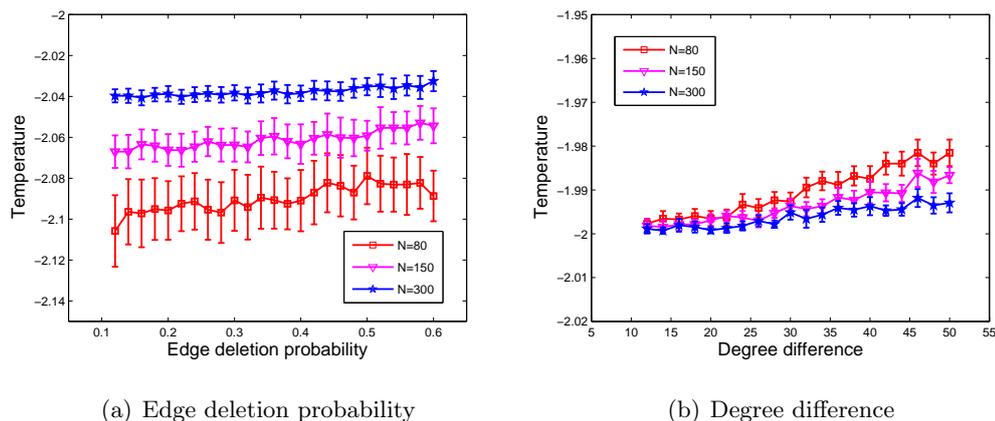
Figure 5.2: Mean and standard deviation of temperature as a function of model parameters for random graphs with different graph sizes.

a small value of $(K_1 - K_2)/(J_1 - J_2)$, the value of temperature then stabilizes at -2.

In order to demonstrate that fluctuations in temperature are caused by structural changes in the arrangement of edges in a network, rather than by difference in edge number between two networks, we provide the following empirical analysis. We first create a regular graph of 80 vertices with degree $m = 10$, and create a second regular graph that has the same graph size, but with a greater degree $m + \Delta m$. Thus, the temperature due to fluctuations between these two networks can be computed. For each $\Delta m = 12, 14, \ldots, 50$, we again produce 100 realizations of the graphs. We then plot the mean and standard deviation of temperature against Δm for different graph sizes in Fig. 5.2(b). For random graphs with various vertex number, although there are some fluctuations, the temperature is almost constant despite the fact that the degree difference varies significantly. This is because there is no significant change in the internal structure of the network during such an evolution. This result confirms that the thermodynamic characterizations are effective in capturing the changes in internal structure of time evolving networks.

The value of the temperature deserves further comment. In this experiment T is always negative, this is because the first term in the temperature expression Eq. (5.11) has a minus sign. It is worth stressing that this sign appears naturally from the temperature development and it does not mean the temperature is negative physically.

We explore whether the thermodynamic measures can be used as an effective tool for better understanding the evolution of realistic complex networks. To commence, we explore the evolutionary behaviour of the financial network by applying our thermodynamic
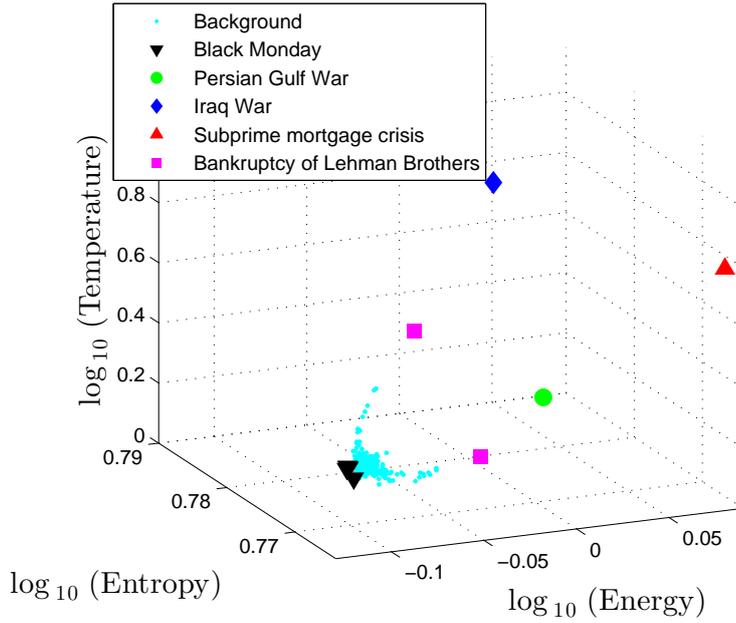
Figure 5.3: Scatter plot of the time-evolving financial network in the thermodynamic space spanned by temperature, average energy and entropy.

characterization method to the dynamic networks in the *NYSE Stock Market Network Dataset.* At each time step, we compute the average energy, entropy and temperature according to Eq. (5.12), Eq. (5.13), and Eq. (5.11) respectively. This allows us to investigate how these network thermodynamic variables evolve with time and whether critical events can be detected in the network evolution.

Figure 5.3 is a 3D scatter plot showing the thermodynamic variables for the time-evolving stock correlation network. It represents a thermodynamic space spanned by average energy, entropy and temperature. The thermodynamic distribution of networks clearly shows a strong manifold structure. The outliers, on the other hand, indicate singular global events. Examples include Black Monday (black downward-pointing triangles) [26], the Persian Gulf War and Iraq War (green circles and blue diamonds respectively), and the subprime mortgage crisis (red upward-pointing triangles) together with the bankruptcy of Lehman Brothers (magenta squares).

The individual time-series for different thermodynamic variables, i.e., temperature, energy and entropy are shown in Fig. 5.4. There are a number of important observations. First, most of the significant fluctuations in the individual time-series of thermodynamic variables successfully correspond to some realistic serious financial crises, e.g., Black Monday [26], Friday the 13th mini-crash [68], September 11 attacks and the bankruptcy of
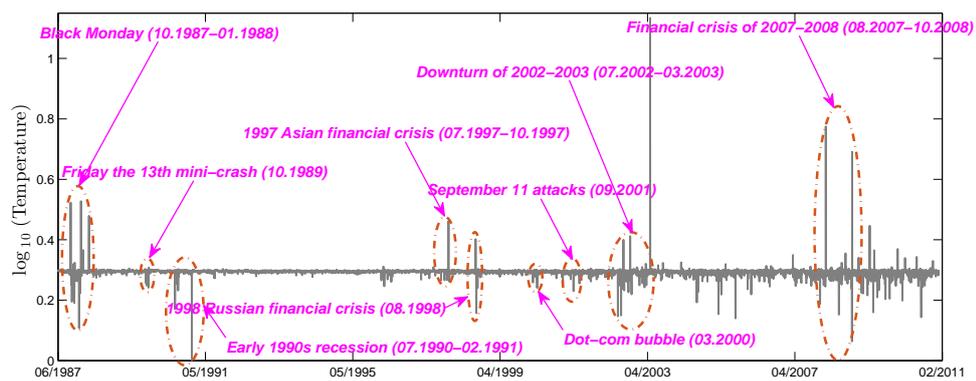
(a) Temperature
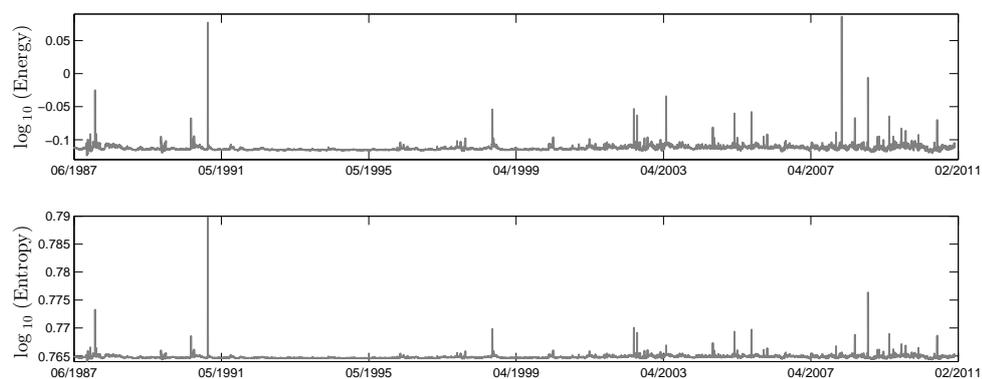
(b) Energy and entropy

Figure 5.4: Thermodynamic characterizations as a function of time for the time-evolving financial network.

Lehman Brothers [90]. The reason for this is that the stock market network experiences dramatic structural changes when a financial crisis occurs. For instance, during the dot-com bubble period [7], a significant number of Internet-based companies were founded, leading to a rapid increase of both stock prices and market confidence. This considerably modified both the inter-relationships between stocks and the resulting structure of the entire market, which can be captured by the thermodynamic characterization. Another interesting feature in the figure is that the stock correlation network structure becomes considerably unstable after entering the 21st century, compared to that before year 2000. Particularly, there are a great number of significant fluctuations in all three time-series in recent years, which is due to the outbreak of the global recession and financial crisis that began in 2007.

To see more clearly the detail of how the thermodynamic variables change over time during the different financial crises, in Fig. 5.5 we show all three thermodynamic variable time-series for the nine global events identified in Fig. 5.4. From Fig. 5.5(a), the most striking observation is that almost all of the largest peaks and troughs can find their realistic financial crisis correspondences, which shows the thermodynamic characterization is sensitive to network structural changes. Also, different global events exhibit different detailed behaviours. For example, both wars (Persian Gulf and Iraq) dramatically change the network structure in a short time, which are shown as a sharp trough and a peak respectively in the corresponding time-series. Moreover, the September 11 attacks clearly have a persistent influence on the stock market since the network temperature fluctuates significantly after the attack. The reason for this is that different financial crises affect the stock network structure in different ways. Specifically, some crises lead the degree-products for both triangles K and edges J increase or decrease simultaneously (Black Monday, Iraq War, the subprime mortgage crisis, etc.), and as a result $(K_1 - K_2)/(J_1 - J_2)$ is positive and the temperature increases. In contrast, some events lead to the result that K and J change in a different direction, which means that $(K_1 - K_2)/(J_1 - J_2)$ is negative and the temperature decreases accordingly, such as Persian Gulf War, the minicrash on October 27, 1997 and the dot-com bubble climax.

We now compare our thermodynamic representation with a number of methods from the spectral analysis of graphs, namely the heat kernel signature [120] and the wave kernel signature [10]. Figure 5.6 shows 3D scatter plots obtained from the PCA of network characterizations delivered by these two methods respectively. Both plots show a compact
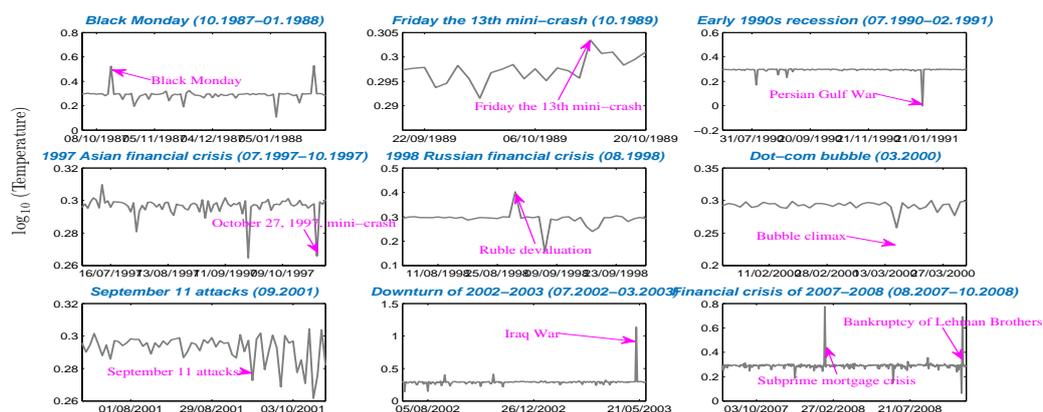
(a) Temperature
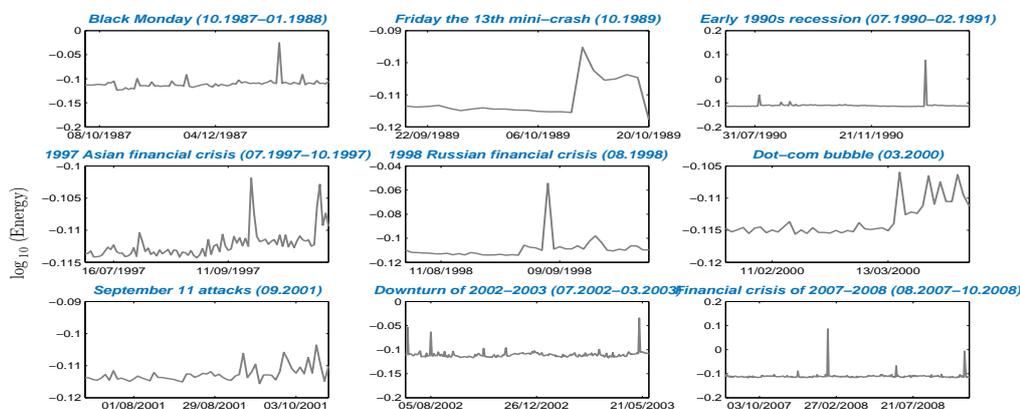
(b) Energy
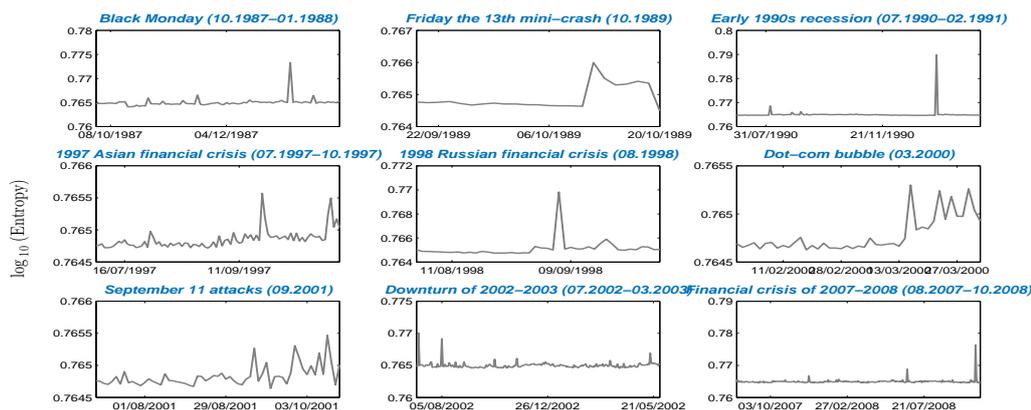
(c) Entropy

Figure 5.5: Individual time-series of thermodynamic characterizations for nine different global events that have been identified in Fig. 5.4.

(a) Heat kernel signature

(b) Wave kernel signature
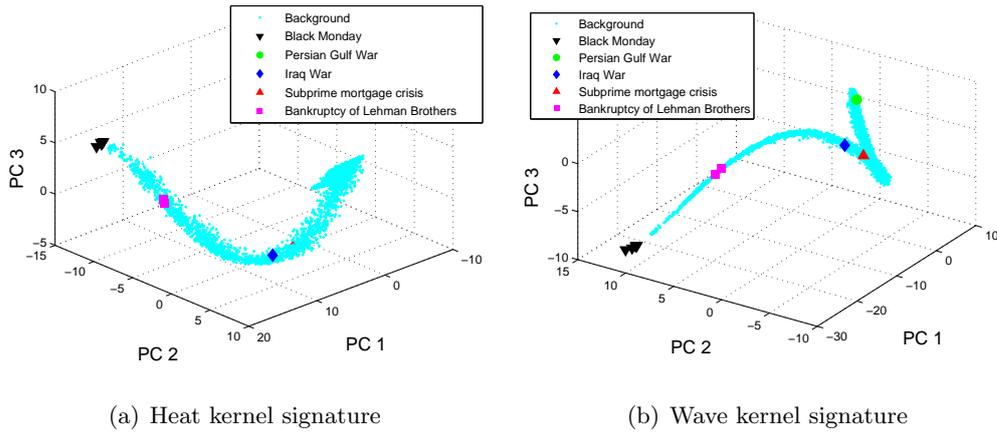
Figure 5.6: PCA plots of the time-evolving financial network characterization delivered by different kernel signature methods.

manifold structure. However, only the Black Monday (black triangles) can be identified. The critical points representing other financial events such as the subprime mortgage crisis and the bankruptcy of Lehman Brothers, do not deviate from the manifold structure, which means that these events cannot be detected. This illustrates that the thermodynamic characterization provides an effective method for analyzing financial network evolution, which smooths the manifold structure while preserving information concerning significant changes in network structure.

We now focus on two different financial crises in more detail, and explore how the stock market network structure changes with time according to the thermodynamic variables. In Fig. 5.7 we show a set of points indicating the path of the stock network in the entropy-energy space with time during a) Black Monday and b) the Lehman Brothers bankruptcy. The colour bar beside each plot represents the date in the time-series. The top panel shows that before Black Monday (blue and green triangles), the network structure remains stable. Neither the network entropy nor the average energy change significantly. However, during Black Monday (from day 116), the network undergoes a considerable change in structure since the entropy decreases dramatically and energy increases significantly. After the crisis, the stock correlation network gradually returns to its normal state. A different behaviour can be observed concerning the Lehman Brothers bankruptcy which is shown in the bottom panel. The stock network undergoes a significant crash in which the network structure undergoes a significant change, as signalled by a large increase in both network energy and entropy. More importantly, the crash is followed by a quick recovery. Hence, our

(a) Black Monday

(b) Bankruptcy of Lehman Brothers
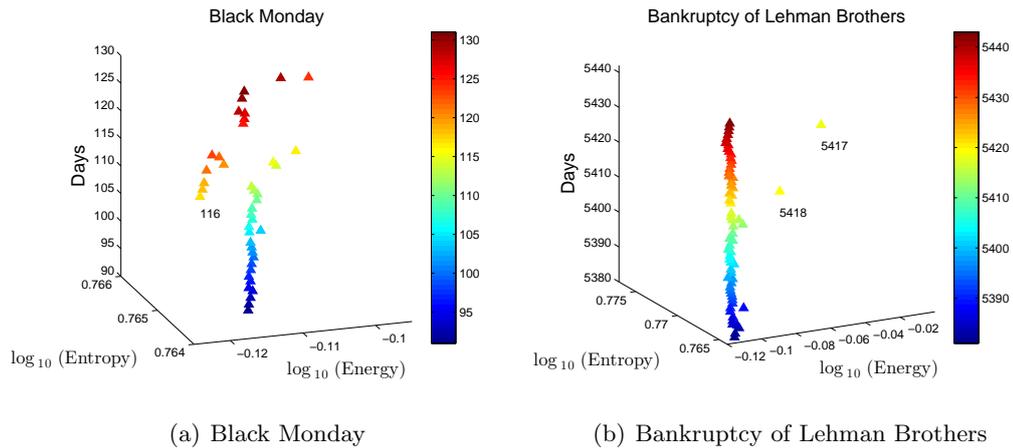
Figure 5.7: Path of the time-evolving financial network in the entropy-energy-time space during different financial crises.

thermodynamic representation can be used to both characterize and distinguish between different financial crises.

In Fig. 5.8 we provide a normalized histogram of β for this time-evolving stock correlation network. The most striking feature is that the vast majority of this parameter is between -0.6 and -0.4. This result shows empirically, that for real-world complex networks, the approximation between the Boltzmann partition function and the quasi characteristic polynomial of normalized Laplacian matrix Eq. (5.5) is valid.

We now turn our attention to the fruit fly network, i.e., the Drosophila gene regulatory network contained in the *Drosophila Melanogaster Gene Network Dataset.* In Fig. 5.9, we again show a 3D scatter plot of the time-varying thermodynamic variable space. Unlike the NYSE data for the stock, here the data points do not display a clear manifold in the thermodynamic space. This is because there are only 66 time epochs in the time-series of the gene regulatory network. Nevertheless, some critical morphological changes can still be identified, such as the egg hatching (black triangle), molt (magenta circle) and pupation (blue diamond). More importantly, the red triangle, representing the most significant morphological change, namely the emergence of the adult is separated by the greatest distance from the remainder of the developmental samples. This indicates that the thermodynamic characterization successfully captures the evolutionary changes in the underlying dynamic network.

Figure 5.10 shows the separate time-series of temperature, energy and entropy for the fruit fly network. Also shown in this figure are a number of critical evolutionary
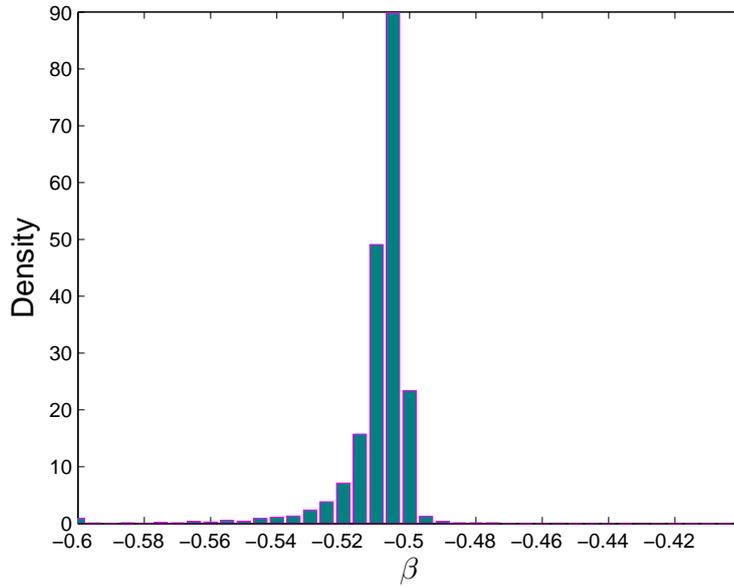
Figure 5.8: Normalized histogram of $\beta = 1/kT$, for the time-evolving financial network.
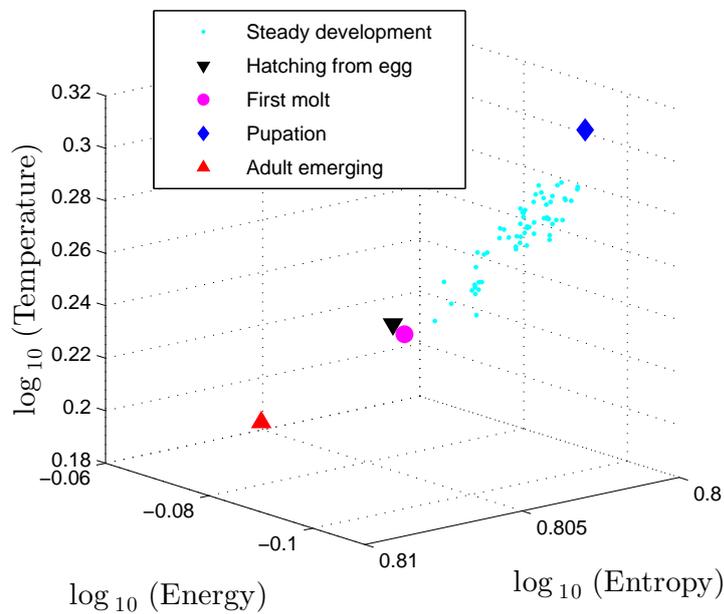
Figure 5.9: Scatter plot of the time-evolving gene network in the thermodynamic space spanned by temperature, average energy and entropy.
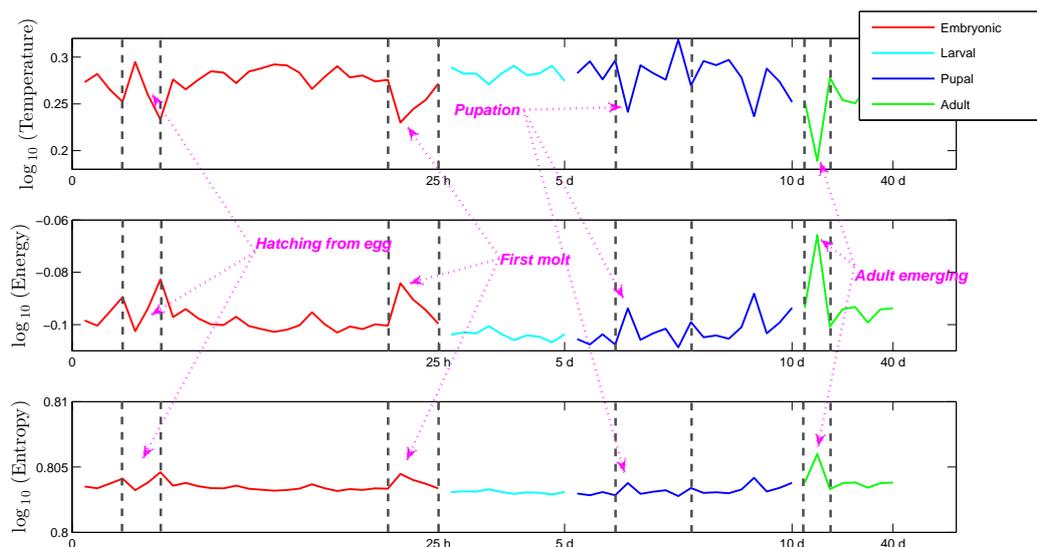
Figure 5.10: Temperature, average energy and entropy as a function of time for the time-evolving gene network.

events, which are indicated by arrows and four developmental stages, which are shown in different colours. In the plot, the early development of embryo, which is represented using the red line (embryonic period) shows some significant fluctuations. This is attributable to strong and rapidly changing gene interactions, because of the need for rapid organism development. Moreover, in the pupal stage, there are also considerable fluctuations. This is attributable to the fact that during this period, the pupa undergoes a number of significant pupal-adult transformations. As the organism evolves into an adult, the gene interactions which control its growth begin to slow down. Hence the green line (adulthood) remains stable (after the adult emerges).

We again provide a comparison between our thermodynamic representation and the heat kernel signature together with the wave kernel signature analyses on this biological data. To this end, we apply PCA to the network characterizations delivered by these two methods and obtain the 3D scatter plots in Fig. 5.11. Comparing to Fig. 5.9, it is difficult to distinguish the time points when significant morphological changes take place between those representing steady evolutionary development. This observation confirms that the thermodynamic characterization is not only effective in the financial domain, but also provides some useful insights to analyze the biological data.

Finally, in Fig. 5.12 we show a normalized histogram of β for the Drosophila gene

137

(a) Heat kernel signature

(b) Wave kernel signature
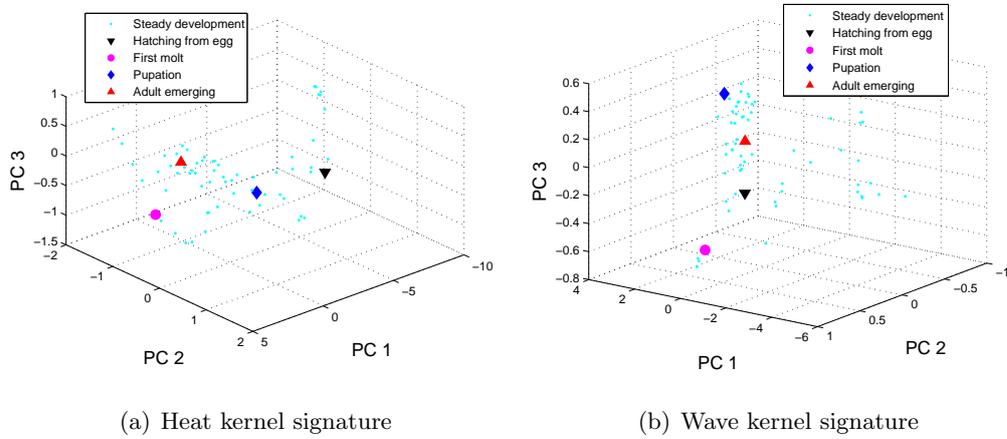
Figure 5.11: PCA plots of the time-evolving gene network characterization delivered by different kernel signature methods.
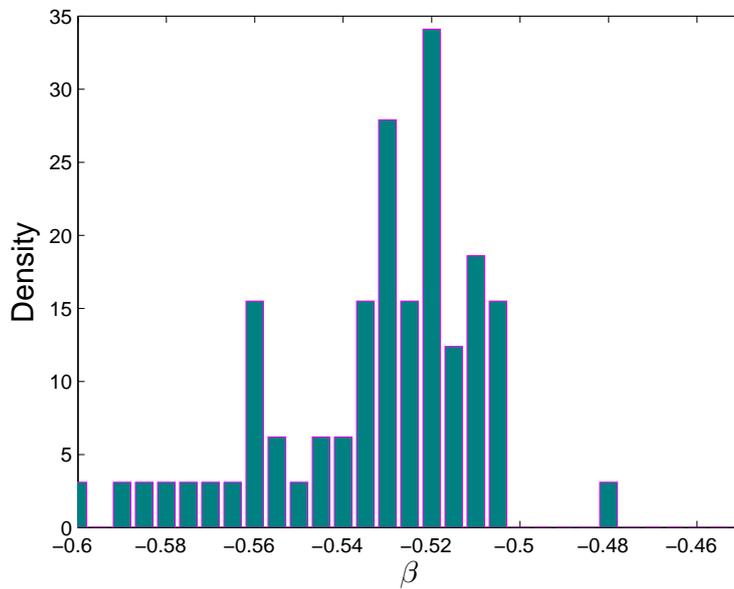
Figure 5.12: Normalized histogram of $\beta = 1/kT$, for the time-evolving gene network.
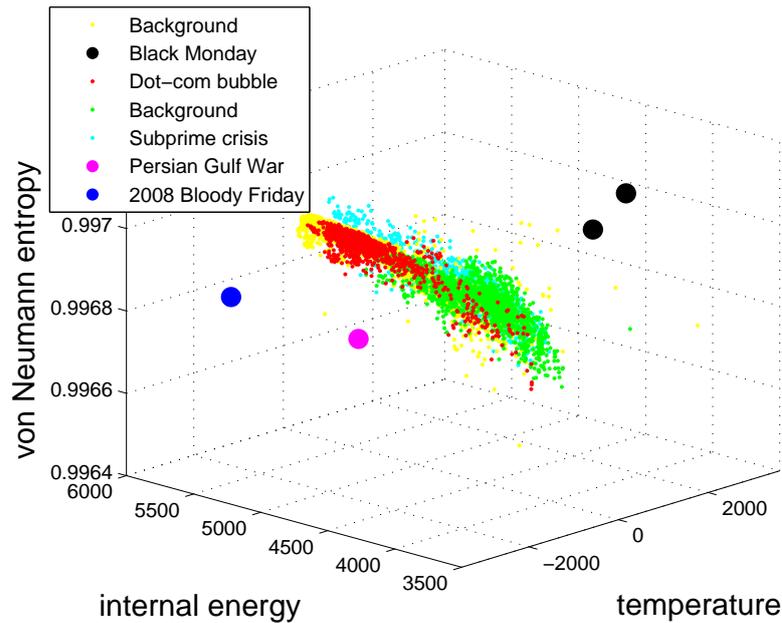
Figure 5.13: Scatter plot of the time-evolving financial network in the thermodynamic space spanned by temperature, average energy and entropy.

regulatory network. The main conclusion from the plot is that result β is most densely distributed over the interval $[-0.6, -0.45]$, empirically showing that $|\beta|$ takes on a small value such that $r(\beta) = o[\ln R(\beta)]$, which again confirms the validity of the approximation obtained in Eq. (5.5).

5.3.2 Thermodynamic Characterization Using Statistical Mechanics

We apply our thermodynamic characterization method obtained from statistical mechanics to the dynamic networks in the *NYSE Stock Market Network Dataset* and *Drosophila Melanogaster Gene Network Dataset*. At each time step we compute the entropy, internal energy and temperature according to Eq. (3.3), Eq. (5.15) and Eq. (5.16) respectively. This allows us to investigate how these network thermodynamic variables evolve with time and whether some critical events can be detected in the network evolution. These include financial crises or crashes in the stock market, and the essential morphological transformations that occur in the development of the Drosophila.

The plot in Fig. 5.13 is a 3D scatter plot showing the thermodynamic variables for the time-evolving stock correlation network. It represents a thermodynamic space spanned by entropy, internal energy and temperature. The plots shown in Fig. 5.14 are the individual times series for the different thermodynamic variables. The most important feature here
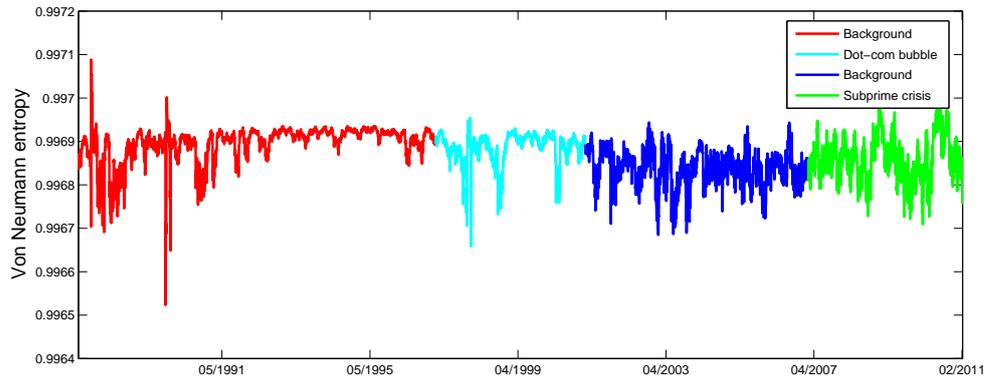
(a) Von Neumann entropy
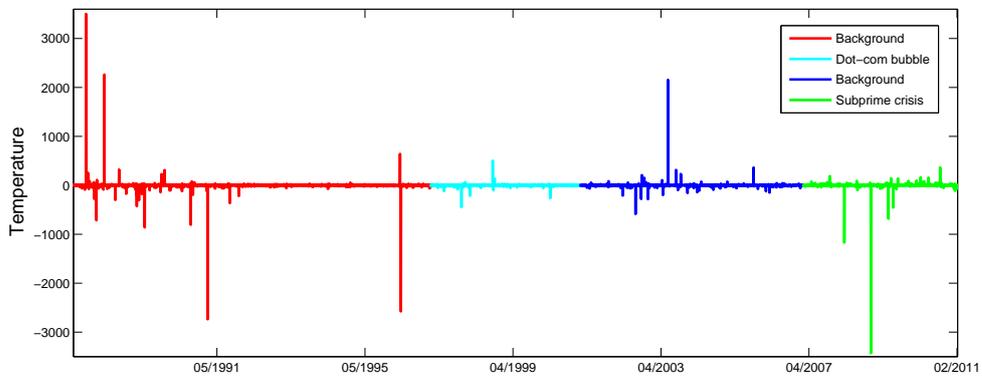
(b) Energy
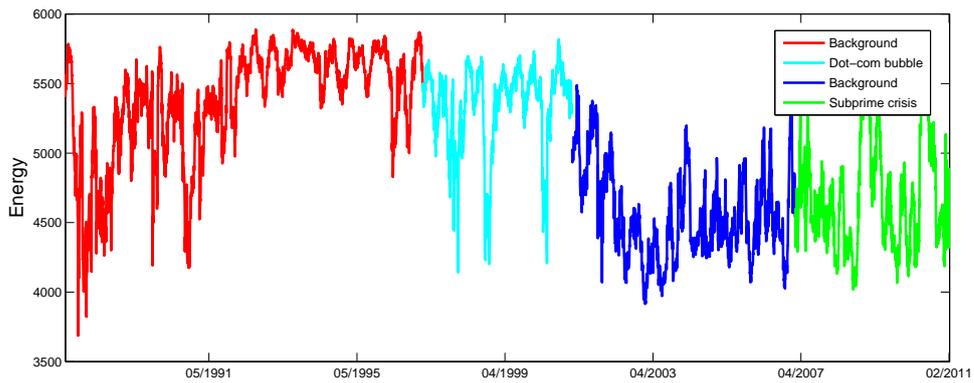
(c) Temperature

Figure 5.14: Individual time-series of thermodynamic characterizations as a function of time of the time-evolving financial network.

is that the thermodynamic distribution of networks in Fig. 5.13 shows a strong manifold structure with different phases of network evolution occupying different volumes of the thermodynamic space. There are though outliers, and these appear as peaks and troughs in the individual time series (in Figs. 5.14(a), 5.14(b) and 5.14(c)). The outliers indicate significant global events. Examples include Black Monday, the outbreak of the Persian Gulf War (17 January 1991) and the 24 October 2008 stock market crash. Another interesting feature in Fig. 5.13 is that the dot-com bubble period (approximately from 1997 to 2000) which is represented by red dots, is separated from the background data points and occupies a distinct region in the thermodynamic space. The reason for this is that during the dot-com bubble period, a significant number of Internet-based companies were founded, leading to a rapid increase of both stock prices and market confidence. This considerably changed both the inter-relationships between stocks and the resulting structure of the entire market.

In Fig. 5.15 we show the trace of the stock network on the entropy-energy plane during Black Monday, the Asian Financial Crisis and the Lehman Brothers bankruptcy respectively. The number beside each data point represents the day number in the time-series. From the left-hand panel we observe that before Black Monday, the network structure remains stable, neither the network entropy nor the internal energy change significantly. However, during Black Monday (day 115 and 116), the network experiences a considerable change in structure since the entropy increases dramatically. After the crisis, the network entropy slowly decreases, and the stock correlation network gradually returns to its normal state. A similar observation can also be made concerning the 1997 Asian Financial Crisis which is shown in the middle panel. The stock network again undergoes a significant crash in which the network structure undergoes a significant change, as signalled by a large drop in network entropy. The crash is followed by a slow recovery. For the Lehman Brothers bankruptcy in the right-hand panel, it is interesting to note that as the time-series evolves, both the network entropy and the internal energy continue to grow gradually. This illustrates a very different pattern. Hence, our thermodynamic representation can be used to both characterize and distinguish between different financial crises.

We now turn our attention to the fruit fly network, i.e., the Drosophila gene regulatory network contained in *Drosophila Melanogaster Gene Network Dataset.* In Fig. 5.16, we again show the 3D scatter plot of the time-varying thermodynamic variables space. Figure 5.17 shows the entropy, energy and temperature time-series. The four developmental stages

141

(a) Black Monday

(b) Asian Financial Crisis

(c) Bankruptcy of Lehman Brothers
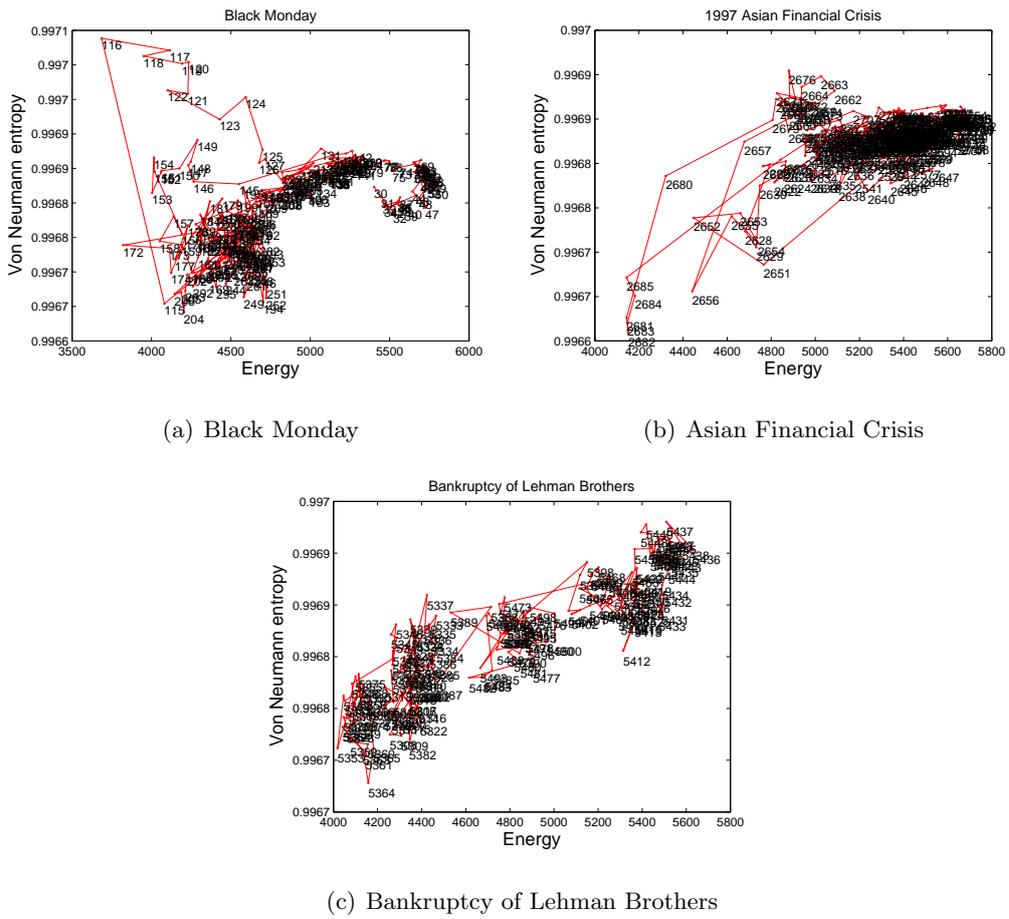
Figure 5.15: Path of the time-evolving financial network in the entropy-energy plane during different financial crises.
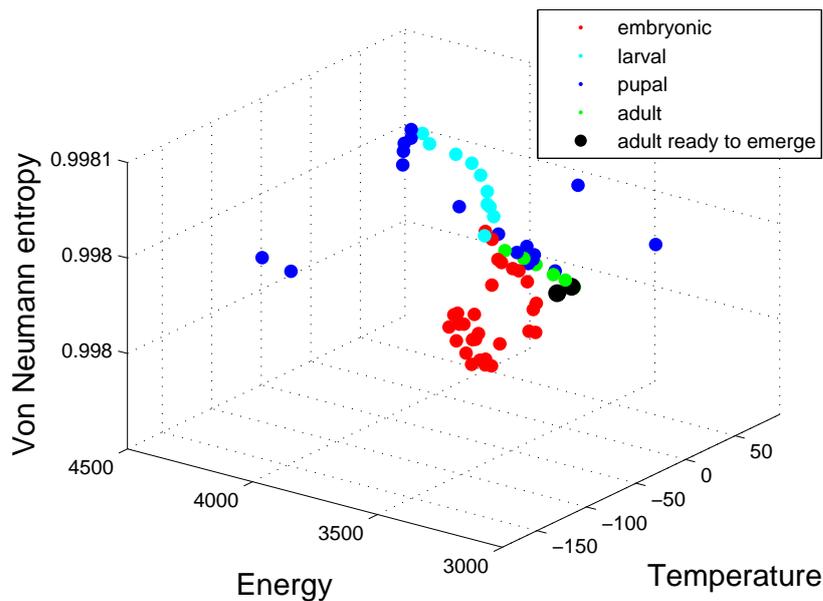
Figure 5.16: Scatter plot of the time-evolving gene network in the thermodynamic space spanned by temperature, average energy and entropy.

are shown in different colours. The different stages of evolution are easily distinguished by the thermodynamic variables. For instance from Fig. 5.17(a), since the early development of an embryo, the red dots (embryonic period) show significant fluctuations. This is attributable to strong and rapidly changing gene interactions, because of the need for rapid development. Moreover, in the pupal stage, the data are relatively sparsely distributed in the thermodynamic space. This is attributable to the fact that during this period, the pupa undergoes a number of significant pupal-adult transformations. As the organism evolves into an adult, the gene interactions which control its growth begin to slow down. Hence the green points (adulthood) remain stable. Finally, the black data points are well separated from the remainder of the developmental samples, and correspond to the time when the adult emerges.

In this section we have undertaken experiments on a number of realistic time-varying complex systems in order to analyze whether the thermodynamic characterizations we have developed are efficient in studying the evolution of dynamic networks. The experimental results of both the stock correlation network of NYSE and the Drosophila gene regulatory network demonstrate that the thermodynamic entropy, internal energy together with temperature provide a powerful tool for detecting abrupt events and characterizing different stages in the network evolution.

143
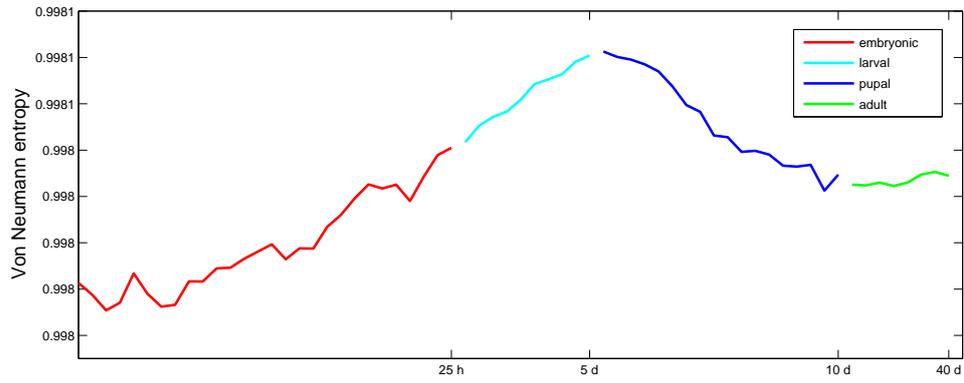
(a) Von Neumann entropy
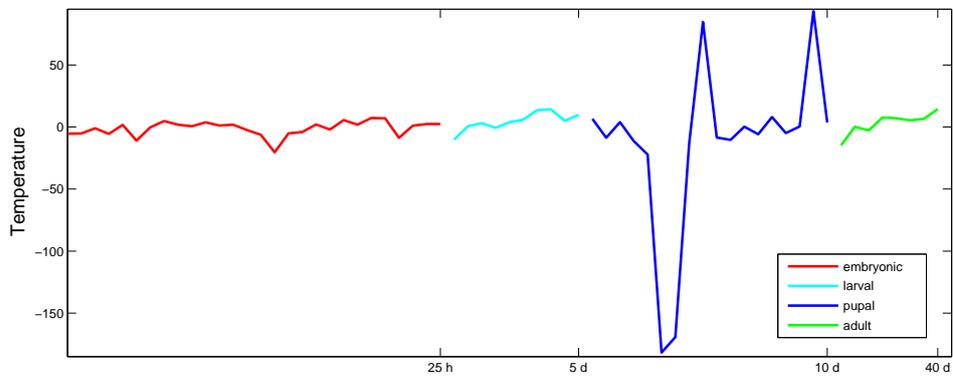
(b) Energy
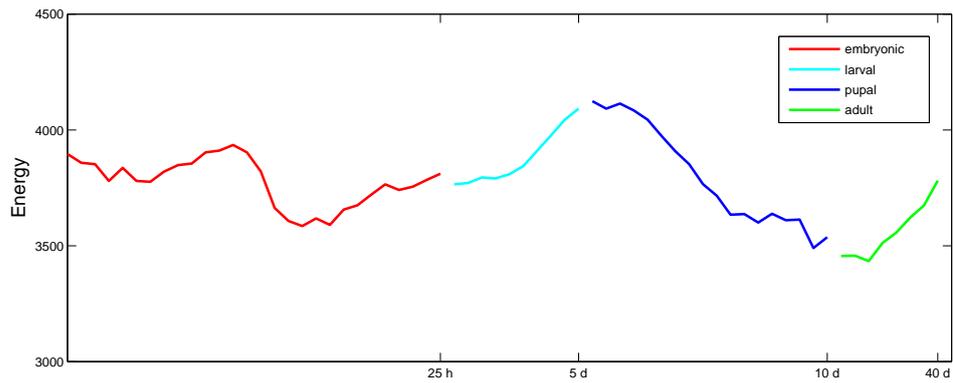
(c) Temperature

Figure 5.17: Individual time-series of thermodynamic characterizations as a function of time of the time-evolving gene network.

5.4 Summary

In this chapter, we have adopted two different thermodynamic representations of network structure in order to visualize and understand the evolution of time-varying networks. We first show how a characteristic polynomial can be used to approximate the Boltzmann partition function of a network. We commence from a quasi characteristic polynomial computed from the normalized Laplacian matrix of a graph and show how this polynomial is linked to the Boltzmann partition function of the graph, when the graph Hamiltonian is defined by the normalized Laplacian operator. This allows us to derive a thermodynamic representation of network structure which can be used to visualize and understand the evolution of time-varying networks. Under the assumption that the network is of constant volume, we provide approximate expressions for a number of thermodynamic network variables, including the entropy, average energy and temperature.

Moreover, based on statistical thermodynamics and commencing from the fact that the microscopic configurations of a network can be defined as the normalized Laplacian eigenstates, the approximate von Neumann entropy can be interpreted as the thermodynamic entropy of a network. The internal energy depends on the number of edges in the network. The thermodynamic temperature measures fluctuations via changes in the number of edges and individual vertex degree changes.

We evaluate both thermodynamic characterizations experimentally using data representing a variety of real-world complex systems, taken from the financial and biological domains. The experimental results demonstrate that the thermodynamic variables are efficient in analyzing the evolutionary properties of dynamic networks, including the detection of abrupt changes and phase transitions in structure or other distinctive periods in the evolution of time-varying complex networks.

The method we have suggested in this chapter does though appear to have some limitations. For instance it does appear sensitive to random fluctuations in network structure, not associated with identifiable events in the time-series studied. Also critical events do not necessarily give rise to unique patterns.

In the future, it would be interesting to see what features the network thermodynamic variables reveal in additional domains, such as human functional magnetic resonance imaging data. Another interesting line of investigation would be to explore if the thermodynamic framework can be extended to the domains of dynamic directed networks, edge-weighted networks, labeled networks and hypergraphs. Finally, it would be intrigu-

ing to investigate whether partition functions from different quantum statistics, such as Bose-Einstein partition function and Fermi-Dirac partition function, together with Ihara zeta function can be applied to network science to provide a way to probe larger structure.

Chapter 6

Analyzing Graph Time Series Using a Generative Model

In this chapter, we present a novel method for constructing a generative model to analyze the structure of labeled data. Given a set of sample graphs, which is a graph time-series in our analysis, we aim to learn a so-called "supergraph" that best describes the underlying average connectivity structure present in the data. In the samples the vertex set is fixed and labeled and the set of possible connections between vertices change with time. The supergraph represents these changes with a Gaussian probability distribution for the connection weights on each individual edge. This structure is fitted to the sample data by minimizing a description length criterion, with the von Neumann entropy controlling the complexity of the fitted model structure and the Gaussian log-likelihood controlling the mean edge weights and variances. We further show this fitting process can be optimized by using a new fixed-point iteration scheme which locates the elements of the optimal weighted adjacency matrix of the supergraph. We show the iteration process is in fact governed by the partial derivative of the von Neumann entropy. In the experiments, by applying the proposed method to graph time-series data, we show the resulting generative model provides an effective tool for analyzing the underlying connectivity structure of time-evolving networks in the financial domain, and in particular for locating critical events and distinct time epochs in their evolution.

6.1 Probabilistic Framework

This chapter focuses on the challenge of learning a generative model which best captures the underlying edge connectivity in a set of labeled graphs. Broadly speaking, there have been two main approaches to characterizing edge structure variations in graphs, namely a) graph spectral methods and b) probabilistic methods. Although the spectral approach is simple and effective in developing generative models based on the Laplacian eigenvectors, the method has a serious weakness, namely the lack of stability of the Laplacian spectrum under perturbations in graph structure [60]. The probabilistic approach, on the other hand, is potentially more robust, but the method requires that the accurate correspondence information which can be inferred from the available graph structure is to hand [85]. It is important to stress that the graph under study is an ordered collection consisting of a vertex set, an edge set and a vertex label set which maps the vertices to a set of labels. The vertex labels of the graphs give important information on the vertex correspondences between data, which plays a central role in problems such as graph matching. In our analysis, we focus on a simpler case where the vertex number and vertex label information do not change between sample graphs. In other words, we are dealing with graphs whose vertex correspondences are to hand, and particularly are concentrating on the edge patterns present in those graphs.

In this section, we develop a likelihood function of the observed data given the available model. To commence the development, we define some notations that will be used throughout this chapter. Let $\mathbf{G} = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_t, \cdots, \mathcal{G}_N\}$ represent the sample graph dataset under study, and $\mathcal{G}_t$ is used to denote the t-th sample graph in the dataset. The generative model, or the supergraph, which we aim to learn from the sample data is denoted by $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$, with vertex set $\tilde{\mathcal{V}}$ and edge set $\tilde{\mathcal{E}}$.

We are dealing with labeled graphs. Each vertex in a graph has a unique label. In our application involving the *NYSE Stock Market Network Dataset*, there are stocks trading in the New York Stock Exchange market, which are represented by labeled vertices in the network. The vertex indices are denoted by lower-case letters including u, v, a, b, α and β, and we will interchange these vertex indices with the vertex labels.

We represent the connectivity structure of the sample graph $\mathcal{G}_t$ using a weighted adjacency matrix ${}^tW$ whose (u, v)-th entry ${}^tW_{uv}$ indicates the connectivity between vertices u and v in the graph, and clearly, we have ${}^tW_{uv} \in [0, 1]$. Similarly, we use the matrix $\tilde{W}$ to represent the structure of the supergraph $\tilde{\mathcal{G}}$.

Having introduced the necessary formalism, we then develop the probabilistic framework for the generative model learning method. To commence, we require the posterior probability of the observed sample graphs given the structure of the generative model $p(\mathbf{G}|\hat{\mathcal{G}})$. Then, the problem of finding the optimal supergraph can be posed in terms of seeking the structure $\tilde{\mathcal{G}}$ that satisfies the condition

$$\tilde{\mathcal{G}} = \underset{\hat{\mathcal{G}}}{\operatorname{argmax}}\, p(\mathbf{G}|\hat{\mathcal{G}}).$$

We follow the standard approach to constructing the likelihood function, which has been previously used in the work [85] and [60]. This involves factorizing the likelihood function over the observed data graphs and making use of the assumption that each individual edge in the sample graph is conditionally independent of each other, given the structure of the supergraph. As a result, we have

$$p(\mathbf{G}|\tilde{\mathcal{G}}) = \prod_t p(\mathcal{G}_t|\tilde{\mathcal{G}}) = \prod_t \prod_u \prod_v p(^tW_{uv}|\tilde{W}_{uv}), \qquad (6.1)$$

where $t = 1, 2, \cdots, N$, and u and v represent vertex indices. Moreover, $p(^tW_{uv}|\tilde{W}_{uv})$ is the probability that the connectivity between u and v in the sample graph $\mathcal{G}_t$ is equal to $^tW_{uv}$, given that the edge (u, v) in the supergraph $\tilde{\mathcal{G}}$ has connectivity $\tilde{W}_{uv}$.

To proceed, a model for the observation density $p(^tW_{uv}|\tilde{W}_{uv})$ is required. Luo and Hancock [85] have shown that when dealing with the unweighted case where the individual edge connectivity of both the supergraph $\tilde{A}_{uv}$ and the sample graph $^tA_{uv}$ is either 0 or 1, the probability distribution $p(^tA_{uv}|\tilde{A}_{uv})$ can be modeled by a Bernoulli distribution. The idea behind this model is that the connectivity of a particular edge in the data graph is derived from that of the same edge in the supergraph through a Bernoulli distribution. In their work, such model has proved to be useful in solving inexact graph matching problems. Recently, Martin et al. [88] have used a similar model to develop methods for inferring structure for uncertain networks and particularly, for solving community detection problems. Han et al. [60] also adopt this model for the purpose of constructing a generative prototype for a set of graphs, and show that the prototype is effective in implementing tasks of graph classification, graph clustering and generating new sample graphs.

Since we are dealing with graphs whose edge connectivity is weighted, i.e., takes on a value between 0 and 1. As a result the Bernoulli distribution is not appropriate to model the observation density in our analysis. To overcome this problem, we note that according

to the central limit theorem, the distribution of the arithmetic mean of a large number of independent random variables is approximately a Gaussian distribution, regardless of the underlying distribution. Therefore, to simplify matters, here we model the distribution $p(^tW_{uv}|\tilde{W}_{uv})$ by adopting a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ of the connection weights whose mean is the weight for the edge (u, v) in the supergraph, i.e., $\mu = \tilde{W}_{uv}$ and whose variance is σ^2. It is worth noting that the choice of σ does not make a significant difference in our development (we will show this later). Even if we choose an inappropriate σ which makes the generated edge connectivity exceed the valid weight interval from 0 to 1, we could use data re-normalization techniques to guarantee that the connectivity is on a scale of 0 to 1.

Finally, with the observation density model to hand, we write

$$p(^tW_{uv}|\tilde{W}_{uv}) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-(^tW_{uv}-\tilde{W}_{uv})^2/2\sigma^2}.$$

With these ingredients, the likelihood function given in Eq. (6.1) then becomes

$$p(\mathbf{G}|\tilde{\mathcal{G}}) = \prod_t \prod_u \prod_v \frac{1}{\sqrt{2\pi}\sigma} \, e^{-(^tW_{uv}-\tilde{W}_{uv})^2/2\sigma^2}. \tag{6.2}$$

To optimize the supergraph $\tilde{\mathcal{G}}$, we maximize this likelihood function with respect to the elements of the weighted adjacency matrix $\tilde{W}_{uv}$. This can be accomplished by straightforwardly employing a maximum-likelihood estimation (MLE) method. However, this leads to the result that the generative structure of the graph data is simply the mean of the data graphs, i.e., the weighted adjacency matrix of the supergraph $\tilde{W}$ is obtained by taking the average of the data graph adjacency matrices $^tW_{uv}$. Clearly, such structure does not capture sufficient structural properties of the observed data graphs and thus cannot represent the underlying connectivity structure of the sample graphs. So a more meaningful and effective method for estimating the generative model is required.

6.2 Minimum Description Length Coding

To locate the optimal supergraph. we adopt an information theoretic approach and use a two-part minimum description length (MDL) criterion. Underpinning MDL is the principle that the best and most probable explanation of the data is the one that gives the shortest code length of the observed data. To formalize this idea, we encode and transmit the data $\mathcal{G}_t$ together with the hypothesis $\tilde{\mathcal{G}}$, leading to a two-part message whose total length is

given by

$$\mathcal{L}(\mathbf{G}, \tilde{\mathcal{G}}) = \mathcal{L}(\mathbf{G}|\tilde{\mathcal{G}}) + \mathcal{L}(\tilde{\mathcal{G}}),$$

where $\mathcal{L}(\mathbf{G}|\tilde{\mathcal{G}})$ is the code length of the data graphs given the supergraph and $\mathcal{L}(\tilde{\mathcal{G}})$ is the code length of the estimated supergraph. Determining the most likely supergraph structure can be viewed as seeking the one that minimizes the total code length of the likelihood function. To this end, we take into account the total code length and apply the MDL principle to the model, this allows us to construct a supergraph representation that trades off goodness-of-fit with the sample graphs against the complexity of the model.

To apply the two-part MDL principle, we commence by computing the code length of the data graphs given the supergraph. This can be achieved by simply using the average of the negative logarithm of the likelihood function given in Eq. (6.2), with the result that

$$
\begin{aligned}
\mathcal{L}(\mathbf{G}|\tilde{\mathcal{G}}) &= -\frac{1}{N} \ln p(\mathbf{G}|\tilde{\mathcal{G}}) \\
&= -\frac{1}{N} \sum_t \sum_u \sum_v \left\{ \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{({}^tW_{uv} - \tilde{W}_{uv})^2}{2\sigma^2} \right\},
\end{aligned}
\tag{6.3}
$$

where N is the length of the observed sample data $\mathbf{G}$.

Next, we compute the code length of the supergraph structure. Traditionally, the complexity of a model is closely related to the number of parameters in the model. However, this quantity does not provide a good complexity measure for graphs since the true graph complexity cannot be accurately reflected by information such as the vertex number or edge number in the graph. To overcome such problem, we adopt a more meaningful measure of graph complexity, namely the von Neumann entropy which we have thoroughly studied in Chapter 3, to encode the complexity of the supergraph structure.

Recall that the approximate von Neumann entropy expression of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an unweighted adjacency matrix, which is given in Eq. (3.3)

$$H_{VN}^U = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{|\mathcal{V}|^2} \sum_{(u,v)\in\mathcal{E}} \frac{1}{d_u d_v}.$$

Clearly, this approximation can be simply extended to weighted graphs. As a result, for the supergraph $\tilde{\mathcal{G}}$ with weighted adjacency matrix $\tilde{W}$, we have the supergraph complexity code length as follows,

$$\mathcal{L}(\tilde{\mathcal{G}}) = 1 - \frac{1}{|\tilde{\mathcal{V}}|} - \frac{1}{|\tilde{\mathcal{V}}|^2} \sum_{(u,v)\in\tilde{\mathcal{E}}} \frac{\tilde{W}_{uv}}{w_u w_v}, \tag{6.4}$$

where $w_u = \sum_{(u,v)\in\tilde{\mathcal{E}}} \tilde{W}_{uv}$ is the weighted degree of vertex u, which is defined as the sum of the connectivity weights of the edges connected to u and w_v is similarly defined. In

effect, the complexity of the supergraph depends on two factors. The first is the order of the supergraph, i.e., the number of the vertices while the second is based on the degree statistics of the vertices in the supergraph.

Finally, by adding together the two contributions to the total code length, the overall code length of the likelihood function is

$$
\begin{aligned}
\mathcal{L}(\mathbf{G}, \tilde{\mathcal{G}}) &= \mathcal{L}(\mathbf{G}|\tilde{\mathcal{G}}) + \mathcal{L}(\tilde{\mathcal{G}}) \\
&= -\frac{1}{N} \sum_t \sum_u \sum_v \left\{ \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{(^t W_{uv} - \tilde{W}_{uv})^2}{2\sigma^2} \right\} \\
&\quad + 1 - \frac{1}{|\tilde{\mathcal{V}}|} - \frac{1}{|\tilde{\mathcal{V}}|^2} \sum_{(u,v)\in\tilde{\mathcal{E}}} \frac{\tilde{W}_{uv}}{w_u w_v}.
\end{aligned}
\tag{6.5}
$$

To recover the supergraph we must optimize the above code length criterion with respect to the weighted adjacency matrix $\tilde{W}$. This can be done in a number of ways. These include gradient descent and soft assign [56]. However here we use a simple fixed-point iteration scheme. To proceed with the development of a useful optimization scheme we must compute the partial derivative of the code length criterion $\mathcal{L}(\mathbf{G}|\tilde{\mathcal{G}})$ given in Eq. (6.3) with respect to the elements of the weighted adjacency matrix $\tilde{W}_{ab}$. First, we compute the partial derivative of the code length of the sample graphs,

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{G}|\tilde{\mathcal{G}})}{\partial \tilde{W}_{ab}} &= \frac{\partial}{\partial \tilde{W}_{ab}} \left(-\frac{1}{N} \sum_t \sum_u \sum_v \left\{ \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{(^t W_{uv} - \tilde{W}_{uv})^2}{2\sigma^2} \right\} \right) \\
&= \frac{1}{N\sigma^2} \sum_t (\tilde{W}_{ab} - {}^t W_{ab}).
\end{aligned}
$$

Then, we need to compute the partial derivative of the code length of the supergraph complexity, which is given in Eq. (6.4), with respect to $\tilde{W}_{ab}$. This is not a straightforward task, and requires that we distinguish between different terms that contain the elements $\tilde{W}_{ab}$ in the summation $\sum_{(u,v)\in\tilde{\mathcal{E}}} \frac{\tilde{W}_{uv}}{w_u w_v}$.

Specifically, the connectivity of an edge (a,b) in the supergraph $\tilde{W}_{ab}$ contributes both to the degrees w_a and w_b. Hence, the partial derivative is determined by three categories of edges in the summation term. The first is the edge (a,b) itself as both w_a and w_b in $\frac{\tilde{W}_{ab}}{w_a w_b}$ contain the element $\tilde{W}_{ab}$. The second category is the edges connected to a, excluding the edge (a,b), i.e., (a,β) and $\beta \neq b$. This is because w_a is needed in the computation of the terms $\frac{\tilde{W}_{a\beta}}{w_a w_\beta}$, which further contribute to the summation $\sum_{(u,v)\in\tilde{\mathcal{E}}} \frac{\tilde{W}_{uv}}{w_u w_v}$. Similarly, the

third category is the edges connected to b, excluding the edge (a, b). As a result, we write,

$$
\frac{\partial}{\partial \tilde{W}_{ab}} \left(\sum_{(u,v) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{uv}}{w_u w_v} \right) = \frac{\partial}{\partial \tilde{W}_{ab}} \left(\frac{\tilde{W}_{ab}}{w_a w_b} + \sum_{\substack{(a,\beta) \in \tilde{\mathcal{E}}, \\ \beta \neq b}} \frac{\tilde{W}_{a\beta}}{w_a w_\beta} + \sum_{\substack{(\alpha,b) \in \tilde{\mathcal{E}}, \\ \alpha \neq a}} \frac{\tilde{W}_{\alpha b}}{w_\alpha w_b} \right)
$$

$$
= \frac{w_a w_b - (w_a + w_b) \tilde{W}_{ab}}{w_a^2 w_b^2} - \frac{1}{w_a^2} \sum_{\substack{(a,\beta) \in \tilde{\mathcal{E}}, \\ \beta \neq b}} \frac{\tilde{W}_{a\beta}}{w_\beta} - \frac{1}{w_b^2} \sum_{\substack{(\alpha,b) \in \tilde{\mathcal{E}}, \\ \alpha \neq a}} \frac{\tilde{W}_{\alpha b}}{w_\alpha}
$$

$$
= \frac{1}{w_a w_b} - \frac{1}{w_a^2} \sum_{(a,\beta) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{a\beta}}{w_\beta} - \frac{1}{w_b^2} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{\alpha b}}{w_\alpha},
$$

where β denote the neighbour vertices of a and α are the neighbours of b.

With these ingredients to hand, we finally have the result of the partial derivative of the total code length of the likelihood function with respect to the structure of the supergraph,

$$
\frac{\partial \mathcal{L}(\mathbf{G}, \tilde{\mathcal{G}})}{\partial \tilde{W}_{ab}} = \frac{1}{N\sigma^2} \sum_t (\tilde{W}_{ab} -^t W_{ab}) - \frac{1}{|\tilde{\mathcal{V}}|^2} \left\{ \frac{1}{w_a w_b} - \frac{1}{w_a^2} \sum_{(a,\beta) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{a\beta}}{w_\beta} - \frac{1}{w_b^2} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{\alpha b}}{w_\alpha} \right\}.
$$
(6.6)

Determining the optimal generative model requires us to find the values of $\tilde{W}_{ab}$ that make Eq. (6.6) equal to zero. However, this leads to an implicit equation that has no easy solution.

6.3 Fixed-point Iteration Scheme

An alternative technique is to use a fixed-point iteration scheme to find the approximation to the solution $\tilde{W}_{ab}$. To set up this scheme, we set the above derivative to zero, and re-organize the resulting equation to obtain an update equation of the form $\tilde{W}_{ab}^{(n+1)} = g(\tilde{W}_{ab}^{(n)})$, where $g(\cdots)$ is the iteration function and n is iteration number. There is of course no unique way of doing this, and for convergence the iteration function $g(\tilde{W}_{ab})$ must have a derivative of magnitude less than unity at the fixed point corresponding to the required solution. One such scheme is

$$
\tilde{W}_{ab}^{(n+1)} = \frac{1}{N\sigma^2} \sum_t {}^t W_{ab} + \frac{1}{|\tilde{\mathcal{V}}|^2} \left\{ \frac{1}{w_a^{(n)} w_b^{(n)}} - \frac{1}{w_a^{(n)2}} \sum_{(a,\beta) \in \tilde{\mathcal{E}}^{(n)}} \frac{\tilde{W}_{a\beta}^{(n)}}{w_\beta^{(n)}} - \frac{1}{w_b^{(n)2}} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}^{(n)}} \frac{\tilde{W}_{\alpha b}^{(n)}}{w_\alpha^{(n)}} \right\}.
$$
(6.7)

The update process is governed by two terms. The first is computed from the local windowed mean of the sample structures $\frac{1}{N\sigma^2} \sum_t {}^t W_{ab}$, while the second term is a step away

from the local windowed mean determined by the partial derivative of the von Neumann entropy. This latter update term depends on the local pattern of vertex degrees.

The convergence properties of the above fixed-point scheme are clearly critical. In general, a fixed-point iteration process $x_{n+1} = g(x_n), \quad n = 0, 1, 2, \cdots$, will converge to the fixed point $x^\star = g(x^\star)$ provided that $|g'(x^\star)| < 1$, and $g(x)$ has exactly one fixed point in $[a, b]$ and the sequence x_n is initialized with $x_0 \in [a, b]$. In our case, from Eq. (6.7), we have

$$\tilde{W}_{ab}^{(n+1)} = g(\tilde{W}_{ab}^{(n)}),$$

where

$$g(\tilde{W}_{ab}^{(n)}) = \frac{1}{N\sigma^2} \sum_t {}^t W_{ab} + \frac{1}{|\tilde{\mathcal{V}}|^2} \left\{ \frac{1}{w_a^{(n)} w_b^{(n)}} - \frac{1}{w_a^{(n)2}} \sum_{(a,\beta) \in \tilde{\mathcal{E}}^{(n)}} \frac{\tilde{W}_{a\beta}^{(n)}}{w_\beta^{(n)}} - \frac{1}{w_b^{(n)2}} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}^{(n)}} \frac{\tilde{W}_{\alpha b}^{(n)}}{w_\alpha^{(n)}} \right\}.$$

Computing the derivative of $g(\tilde{W}_{ab})$ gives

$$
\begin{aligned}
g'(\tilde{W}_{ab}) &= \frac{1}{|\tilde{\mathcal{V}}|^2} \left\{ \frac{\partial \frac{1}{w_a w_b}}{\partial \tilde{W}_{ab}} - \frac{\partial \left(\frac{1}{w_a^2} \sum_{(a,\beta) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{a\beta}}{w_\beta} \right)}{\partial \tilde{W}_{ab}} - \frac{\partial \left(\frac{1}{w_b^2} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{\alpha b}}{w_\alpha} \right)}{\partial \tilde{W}_{ab}} \right\} \\
&= \frac{1}{|\tilde{\mathcal{V}}|^2} \left\{ -\frac{w_a + w_b}{w_a^2 w_b^2} - \frac{w_b - \tilde{W}_{ab}}{w_a^2 w_b^2} - \frac{w_a - \tilde{W}_{ab}}{w_a^2 w_b^2} + \frac{2}{w_a^3} \sum_{(a,\beta) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{a\beta}}{w_\beta} \right. \\
&\quad \left. + \frac{2}{w_b^3} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{\alpha b}}{w_\alpha} \right\} \\
&= \frac{2}{|\tilde{\mathcal{V}}|^2} \left\{ \frac{\tilde{W}_{ab} - w_a - w_b}{w_a^2 w_b^2} + \frac{1}{w_a^3} \sum_{(a,\beta) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{a\beta}}{w_\beta} + \frac{1}{w_b^3} \sum_{(\alpha,b) \in \tilde{\mathcal{E}}} \frac{\tilde{W}_{\alpha b}}{w_\alpha} \right\}.
\end{aligned}
$$

Since $\tilde{W}_{ab} \leq w_a$ and $\tilde{W}_{ab} \leq w_b$, and the vertex weighted degree is normally not small, the derivative satisfies the condition $g'(\tilde{W}_{ab}) \in [-1, 0]$ for all $\tilde{W}_{ab} \in [0, 1]$, and we are assured convergence of the fixed-point iteration scheme. Since the derivative of $g(\tilde{W}_{ab})$ is negative in sign, the convergence pattern is cobweb, i.e., from alternating sides of the fixed point.

There are a number of important points to note concerning the above analysis. First, our goal is to develop a generative model (or supergraph) that can be used to best explain the structural variations present in a set of sample graphs. To this end, we commence from a probabilistic framework which describes the likelihood of the observed data given the model structure. We then pose the problem of determining the optimal model structure as one of minimizing a code length criterion. To solve this problem numerically, we develop a simple fixed-point iteration scheme for optimizing the weighted adjacency matrix of the supergraph. The supergraph is initialized using the mean weighted adjacency matrix for

the sample graphs, and then optimized at each time step to best fit the data in a window, the adjustment is determined by the von Neumann entropy.

6.4 Experiments

It is interesting to note that the generative model provides a number of new directions to the study of time-varying complex networks. When applying this method to the graph time-series data, the resulting generative model can be used as a more efficient structure representation of the time-varying network as it captures more structural information present in the time-series data. Moreover, the observed data graphs in the time-series can be viewed as the samples generated from the model, so the graphs with significantly different structural patterns are expected to be the outliers that are generated with a relatively low probability from the model. This allows us to detect abrupt structural changes in the network evolution by measuring the similarity between the sample graph at each time point and the model graph.

To evaluate the properties of the generative model and explore its practical utilities on real-world data, in this section we report experimental results on time-evolving financial networks representing stock trading in the *NYSE Stock Market Network Dataset*. We first examine the validity of the proposed model learning method by exploring its convergence properties. Then, we compare the data graph structure with the supergraph learned from a time window of fixed length for a number of financial crisis time-series, and the result shows the supergraph is able to smooth the time-series data and more importantly, is more effective in locating critical events and distinct time epochs in financial crises. Also shown in the experiments is that by comparing the generative structure learned from different time-series data, we can better visualize and understand the structural difference of the stock market network in different time periods.

6.4.1 Convergence

The first part of our experimental investigation aims to explore the convergence properties of the fixed-point iteration scheme. To this end, we first investigate whether the choice of Gaussian distribution adopted for modeling the observation density $p(^tW_{uv}|\tilde{W}_{uv})$ is valid. Figure 6.1 gives the normalized histogram of the edge connectivity of two randomly selected edges in the time-evolving financial network and the probability density function curve of

(a) $u = 147, v = 318$　　　　　　　(b) $u = 48, v = 289$
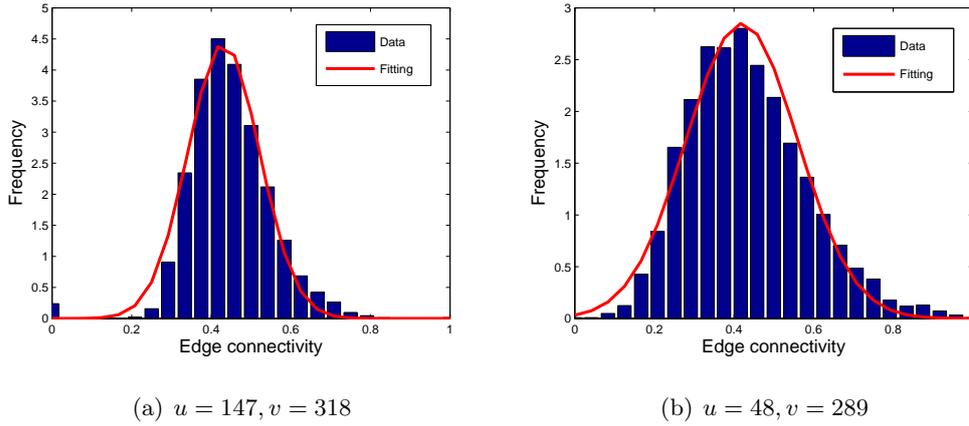
Figure 6.1: Normalized histogram of the connectivity of two randomly selected edges (u, v) in the time-evolving financial network and the Gaussian distribution fitting.

a Gaussian distribution whose mean is equal to the mean value of the corresponding edge connectivity and whose standard deviation is set to be a small value (0.1 here). Clearly, both plots show the normal "bell curve" or Gaussian distribution curve fits the data well, which means that using Gaussian distribution to model the probability of the observed data graph edge connectivity given the corresponding supergraph edge connectivity is effective.

In Fig. 6.2 we show the histogram of both mean and standard deviation of the connectivity of each edge in the time-evolving stock correlation network. We see the mean connectivity of most edges in the evolving network take values between 0.4 and 0.5, and the standard deviation of edge connectivity is densely populated around 0.1. This again shows that the use of Gaussian distribution is empirically valid.

To better visualize how the fixed-point iteration converges, we test the iteration process on a number of time-series graph data of different financial crises, including the Black Monday, Friday the 13th mini-crash and September 11 attacks. For each financial crisis time-series, we initialize the supergraph $\tilde{W}_0$ with the mean data graph, which is obtained by taking the average of the structure of the sample graphs in the time-series. Then we process the iteration scheme $\tilde{W}_{ab}^{(n+1)} = g(\tilde{W}_{ab}^{(n)})$, $n = 0, 1, 2, \cdots$, which is given in Eq. (6.7) and observe how the complexity of the supergraph, the average log-likelihood of the sample graphs and the total code length of the likelihood function vary with the iteration number. Here, the standard deviation of the Gaussian distribution is set to be $\sigma = 0.1$, and at each iteration step, we perform a feature scaling normalization method to guarantee

(a) Mean

(b) Standard deviation
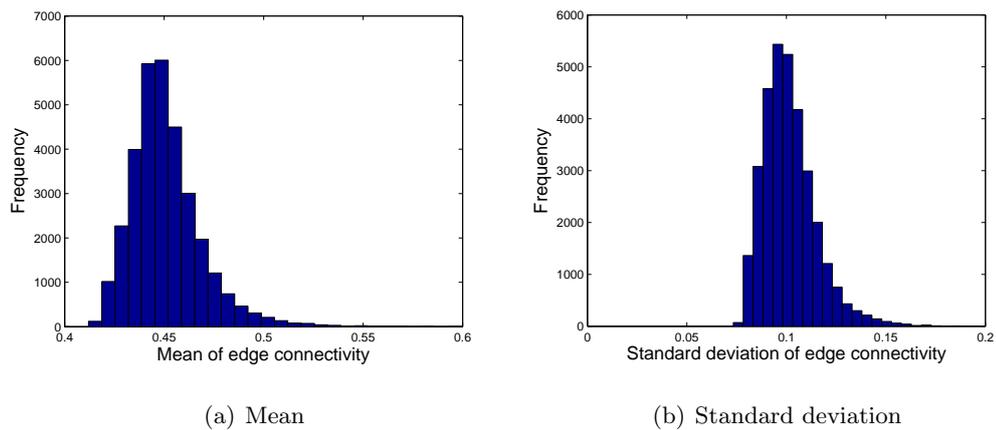
Figure 6.2: Histogram of the mean and standard deviation of the edge connectivity in the time-evolving financial network.

that the elements of the weighted adjacency matrix of the updated supergraph $\tilde{W}_{ab}^{(n)}$ is in the interval $[0, 1]$.

Figure 6.3 shows how various properties of the fixed-point iteration scheme for a number of financial crisis time-series data change during updating. From the plots in the first column, as the iteration processes, the supergraph entropy of all three time-series fluctuates and gradually converges to a value that is lower than the initial entropy. This observation indicates that the structure of the generative model becomes less complex as the iteration scheme processes. The reason for this is that the supergraph is being optimized to best summarize the structural variations present in the time-series data during updating. Then, some negligible structural information contained in the data is discarded from the generative structure, which makes the supergraph less complex. Another interesting feature to note in the plots is that the entropy convergence of three time-series differs from each other, which is a reasonable result as the supergraph is a structure representation that best explains the data graphs, then different financial crisis time-series yield different generative structures. In the second and third columns, the plots show the average log-likelihood function gradually increases as the increase of the number of iteration, while the total code length reduces. This is an expected observation since the goal of the proposed method is to maximize the probability of the observed data graphs given the model structure and also to minimize the total code length of the likelihood function.

Another interesting feature to note in the figure is that in both cases, the convergence of the fixed-point iteration process is oscillating from side-to-side of the fixed point. This
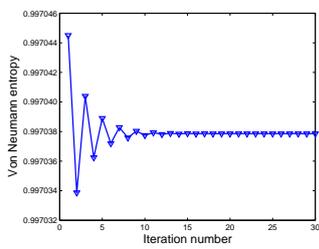
(a) Von Neumann entropy for Black Monday
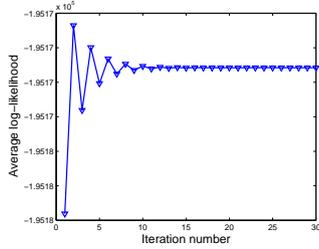
(b) Average log-likelihood for Black Monday
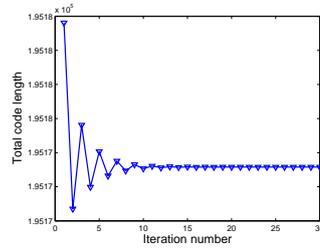
(c) Total code length for Black Monday
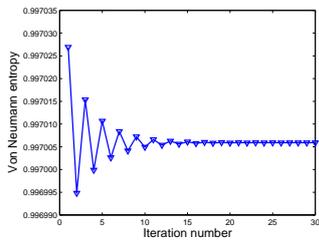
(d) Von Neumann entropy for Friday the 13th mini-crash
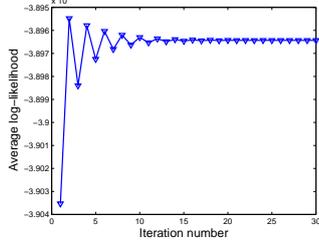
(e) Average log-likelihood for Friday the 13th mini-crash
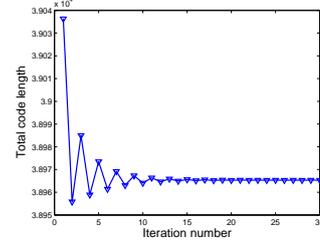
(f) Total code length for Friday the 13th mini-crash
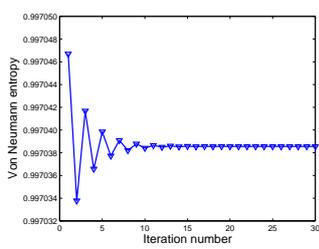
(g) Von Neumann entropy for September 11 attacks
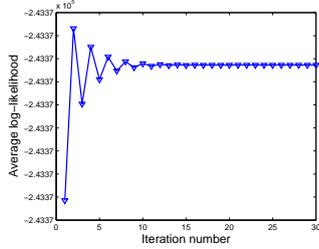
(h) Average log-likelihood for September 11 attacks
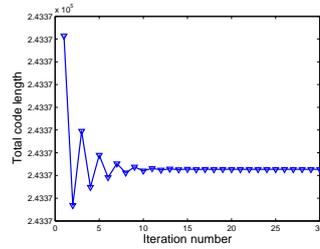
(i) Total code length for September 11 attacks

Figure 6.3: Convergence properties of fixed-point iteration process as a function of iteration number for data graphs of different financial crises.
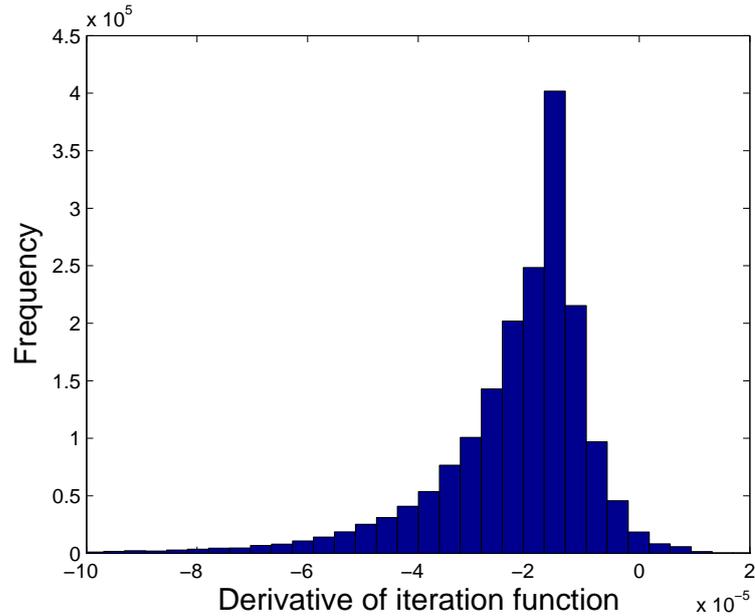
Figure 6.4: Histogram of the derivative of the iteration function for all three iteration processes.

cobweb pattern is a consequence of the fact that $g'(\tilde{W}_{ab}) \in [-1, 0]$. To verify this, in Fig. 6.4 we show a histogram of the values of the derivative function $g'(\tilde{W}_{ab})$ for all the elements $\tilde{W}_{ab}$ obtained during the iteration processes. The result clearly conforms to our expectation as most derivative function values lie in the interval $[-1, 0]$. In particular, the values for the derivative function are close to zero, which means that the iteration process converges close to quadratically to the fixed point.

6.4.2 Event Detection

The second experimental goal is to explore whether the generative model can be used to provide a more efficient representation of the network structure in terms of capturing and detecting the structural changes during network evolution. To this end, for each time epoch t in the time-series data, we view the graphs from $t - \frac{N}{2}$ to $t + \frac{N}{2}$ as the observed sample graphs, where $N + 1$ is the length of the time-series data. We then learn a generative model from this set of sample graphs, which can be viewed as a new network representation at t. We compare this new structure representation with the original data graph at each time epoch in the financial crisis time-series, by using the graph approximate von Neumann entropy.

In Fig. 6.5 we plot the approximate von Neumann entropy of both the data graph and

159
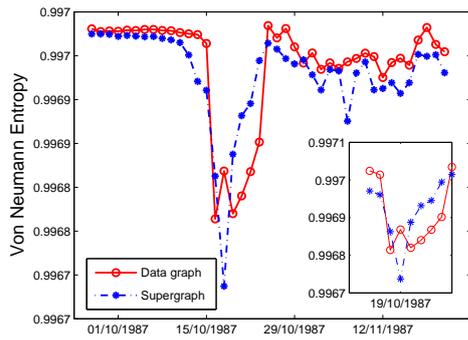
(a) Black Monday
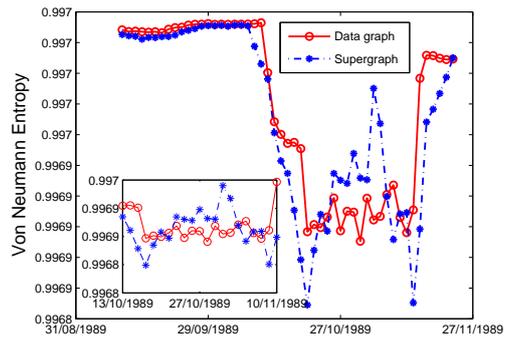
(b) Friday the 13th mini-crash
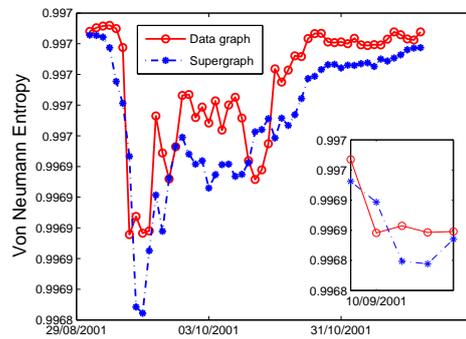
(c) September 11 attacks

Figure 6.5: Von Neumann entropy of supergraph and data graph as a function of time for different financial crises.

the supergraph for a number of different financial events time-series, including the Black Monday, Friday the 13th mini-crash and the September 11 attacks. It is worth pointing out that, to construct the generative structure at each t, we use the sample graphs from $t-5$ to $t+5$ as the data. This is because two weeks (10 trading days) is assumed to be a proper time length that can be used to effectively represent how long the effects of stock correlations last in the stock market. Moreover, the supergraph structure at each time point is obtained after 15 iterations since we have observed from Fig. 6.3 that generally, the structure of the supergraph converges after 15 steps of updating.

From the three plots, the von Neumann entropy of the supergraph generally follows the trend of that of the data graph in the time-series, implying that the supergraph is effective in capturing the graph structural properties in the data. More importantly, the generative model clearly gives a better graph structure representation than the sample graph since the supergraph entropy curve eliminates some of the random fluctuations observed in that of the sample graph. For example, from the inset plot in Fig. 6.5(a), at the day when Black Monday takes place, i.e., 19th October, 1987, the von Neumann entropy of the data graph clearly shows some unexpected fluctuations. However, the supergraph entropy reaches its minimum with a significant decrease, precisely representing that the stock correlation network experiences significant structural changes at that day. The similar behaviours can be observed in Figs. 6.5(b) and 6.5(c) as well. Although the entropy of the sample graph can be used to locate the approximate time period of the financial crises, the supergraph entropy identifies the critical dates (13th October, 1989 and 11th September, 2001 respectively) more precisely. Overall, Fig. 6.5 shows that the supergraph we have learned using the generative model can be used as an efficient tool for summarizing a time-series of sample graphs and more importantly, for identifying significant structural changes during the network evolution.

Next we investigate how the time-series data length N influences the properties of the generative model. To this end, we select the time-series data of the dot-com bubble financial crisis, and repeat the above experimental process to build a set of supergraphs using various choices of data length. We again compare the von Neumann entropy of the data graph and the supergraph and report the result in Fig. 6.6. From the upper-left plot, when the data length N is particularly small (4 days), the supergraph entropy curve basically has the identical trend with that of the data graph. In contrast, when N is large (50 days and 100 days), we see from the bottom plots that the supergraph

161
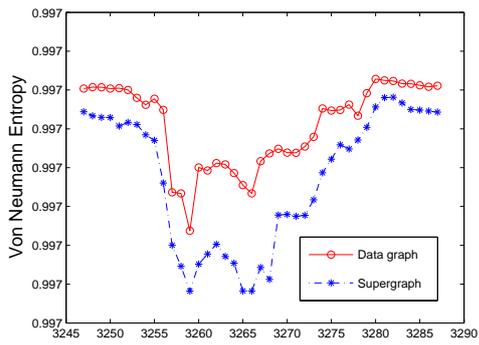
(a) $N = 4$
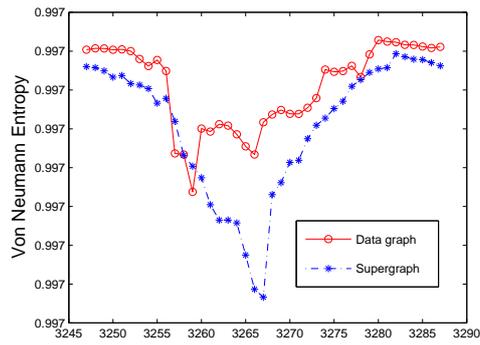
(b) $N = 10$
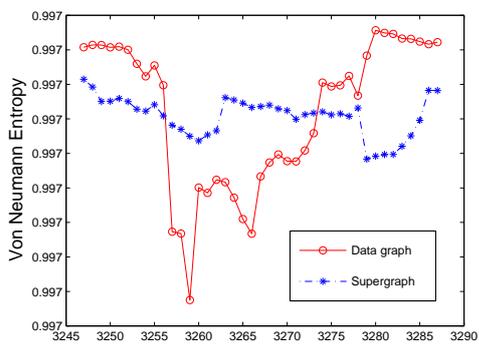
(c) $N = 50$
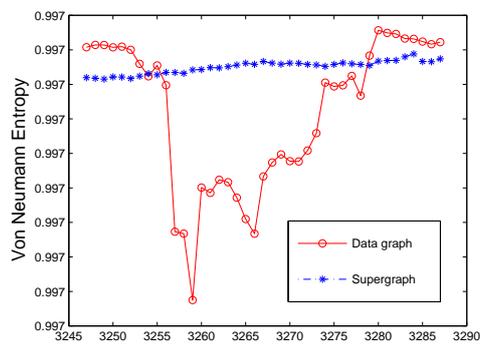
(d) $N = 100$

Figure 6.6: Von Neumann entropy of supergraph and data graph as a function of time for various choices of data length for the dot-com bubble financial crisis.

entropy curve does not show fluctuations when critical event occurs, so it is difficult to represent the underlying network structure using the generative model. However, when choosing an appropriate data length, such as 10 days (upper-right plot), the supergraph entropy removes some of the random fluctuations and precisely identifies the dot-com bubble peak time, i.e., 10th March, 2000. This observation shows that an appropriate data length N plays a central role in the effectiveness of the generative model. On the one hand, a short data length would result in some unexpected fluctuations in the structure of the supergraph, which means the generative structure over-summarizes the structural information contained in the time-series data. On the other hand, a large data length could lose a great amount of structural information during the generative model learning, so the resulting supergraph does not appear to be sensitive to structural changes during network evolution.

6.4.3 Time Series Structure

From our development of the generative model, the supergraph represents a generative structure over a time-series of sample graphs. So it is expected that the graph whose structure significantly differs from that of the supergraph, can be viewed as a sample generated from the generative model with a relatively low probability. In contrast, graphs that have similar structure with the supergraph are the samples generated from the model with higher probabilities. In other words, the graphs corresponding to critical events and periods in the financial time-series should be highly dissimilar to the generative structure computed from the time-series data. To verify this expectation, in Fig. 6.7 we plot the shortest path kernel [23] and the Jensen-Shannon divergence kernel [11] between the sample graph at each time epoch and the supergraph for the entire time period in the financial data. Unlike the experiments shown above, here the supergraph is constructed over the whole collection of the financial data graphs.

The shortest path kernel compares graphs based on the shortest path length of all pairs of vertices. Mathematically, given a pair of graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ that are Floyd-transformed into $\mathcal{S}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{S}_2 = (\mathcal{V}_2, \mathcal{E}_2)$. Borgwardt and Kriegel [23] define the shortest-path graph kernel on $\mathcal{S}_1$ and $\mathcal{S}_2$ as

$$k_{SP}(\mathcal{S}_1, \mathcal{S}_2) = \sum_{(u_1,v_1)\in\mathcal{E}_1} \sum_{(u_2,v_2)\in\mathcal{E}_2} k_{walk}^{(1)}[(u_1, v_1), (u_2, v_2)],$$

where $k_{walk}^{(1)}$ is a positive definite kernel on edge walks of length 1.

The Jensen-Shannon divergence kernel is a non-extensive information theoretic kernel, which is defined using the von Neumann entropy and mutual information computed from the structures being compared. In the work [11], the Jensen-Shannon divergence between graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ is expressed as

$$D_{JS}(\mathcal{G}_1, \mathcal{G}_2) = H_{VN}^U(\mathcal{G}_U) - \frac{H_{VN}^U(\mathcal{G}_1) + H_{VN}^U(\mathcal{G}_2)}{2},$$

where $\mathcal{G}_U$ denotes the disjoint union graph of $\mathcal{G}_1$ and $\mathcal{G}_2$. Then, the Jensen-Shannon divergence kernel is computed as

$$k_{JSD}(\mathcal{G}_1, \mathcal{G}_2) = \exp\{-D_{JS}(\mathcal{G}_1, \mathcal{G}_2)\}.$$

It is interesting to note that the Jensen-Shannon divergence kernel is dependent on the individual von Neumann entropies of graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ as the composite entropy $H_{VN}^U(\mathcal{G}_U)$ can be computed from $H_{VN}^U(\mathcal{G}_1)$ and $H_{VN}^U(\mathcal{G}_2)$.

The most important feature in the figure is that most of the significant troughs can be used to identify serious real-world financial crises. For instance, the Black Monday is the deepest one in both plots and the September 11 attacks also gives a sharp drop in the kernel curve. The reason for this is that during financial crisis, the stock correlation network undergoes significant structural changes, making the graph structure different from that of the generative model, which is a summarized structure over the entire financial time-series. Such difference in graph topology can be efficiently captured by the shortest path kernel and the information theoretic kernel. Then, the similarity between the financial-crisis graph and the generative model becomes extremely low.

Another interesting observation to note in Fig. 6.7 is that both similarity measures exhibit very different patterns before and after a so-called "critical point" time epoch, which is around July and August, 2002. In particular, before the critical point, the kernels generally remain stable with time, except for a small number of fluctuations indicating the financial crises. This implies that the stock correlation network is able to return to its normal state in a short time after each financial crisis from 1987 to 2002. In contrast, after the critical point, the kernels become extremely unstable, which means that the network structure fluctuates significantly and becomes difficult to recover from the crises over the last decade.

To better visualize and study the difference between the stock correlation network structure before and after the critical point, in Fig. 6.8 we show the communities (symbolized by different colours) in the generative model network constructed from the financial
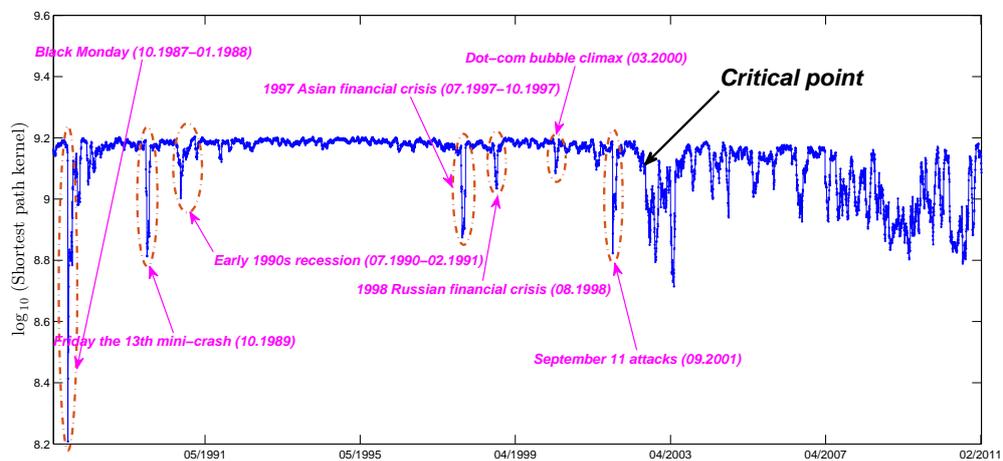
(a) Shortest path kernel
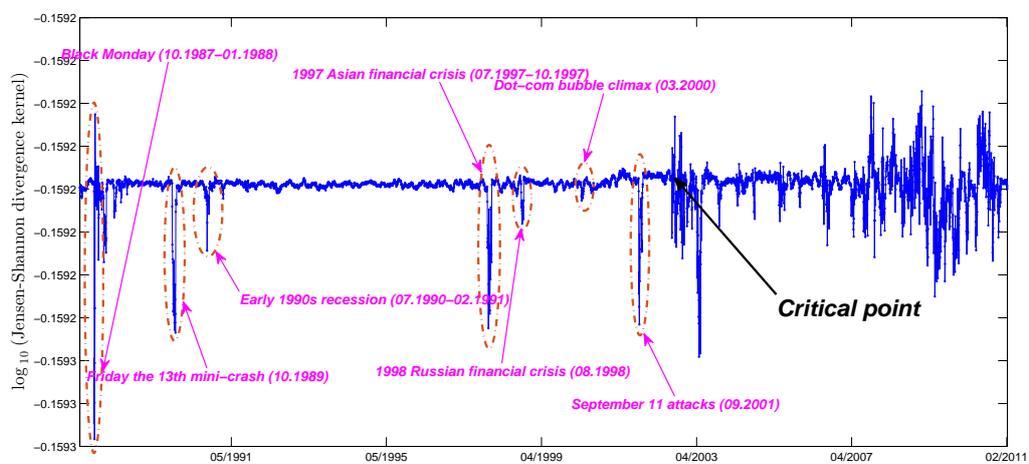
(b) Jensen-Shannon divergence kernel

Figure 6.7: Kernels between sample graph and supergraph as a function of time for the time-evolving financial network.

(a) Before critical point   (b) After critical point
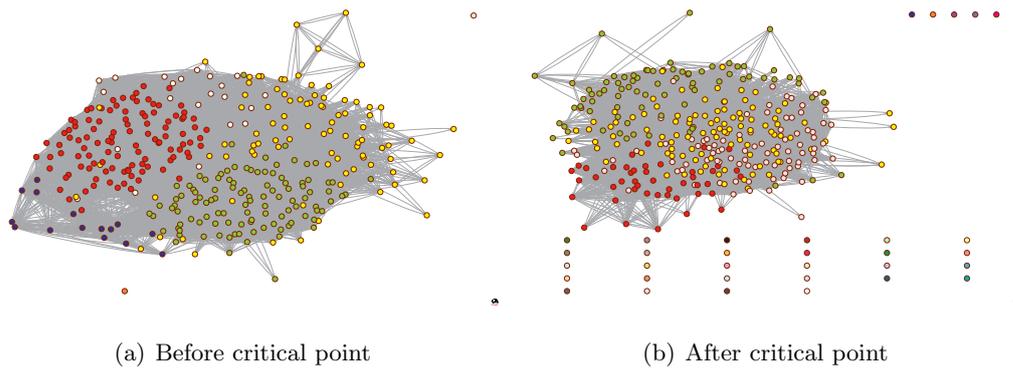
Figure 6.8: Network representation of the communities (symbolized by different colours) in the generative structure constructed from the time-evolving financial network before and after the critical point.

time-series data before and after the critical point respectively. Moreover, Fig. 6.9 gives the adjacency matrix representation of the communities (indicated by squares) found in the generative model. Specifically, the community detection technique adopted here is the Louvain method, whose aim is to optimize the value for network modularity, which is a measurement of the density of connections inside communities compared to that between communities in a network. Moreover, in order to simplify the computation, the generative models before and after the critical point are represented by unweighted adjacency matrices, which are obtained by performing the same thresholding method we have used in the experiments in Chapter 5.

Both figures show before the critical point, the underlying connectivity structure of the stock correlation network contains a small number of large communities, and the network is almost strongly connected as the number of disconnected components is particularly small. In contrast, after the critical point, we see the generative structure becomes significantly different as the number of disconnected components in the network becomes significantly greater. These observations can be mathematically verified by computing the modularity of these two structure, which are 0.3311 and 0.1860 respectively. In effect, the network modularity is bounded between -1 and 1, and a network with a high modularity has modules in which the vertices are densely connected between each other but sparse connections between vertices in different modules. Based on these observations, we draw the conclusion that during 1987 to 2002, most stocks in the market have the tendency to merge into larger-sized groups in which the stocks are densely correlated internally.

(a) Before critical point

(b) After critical point
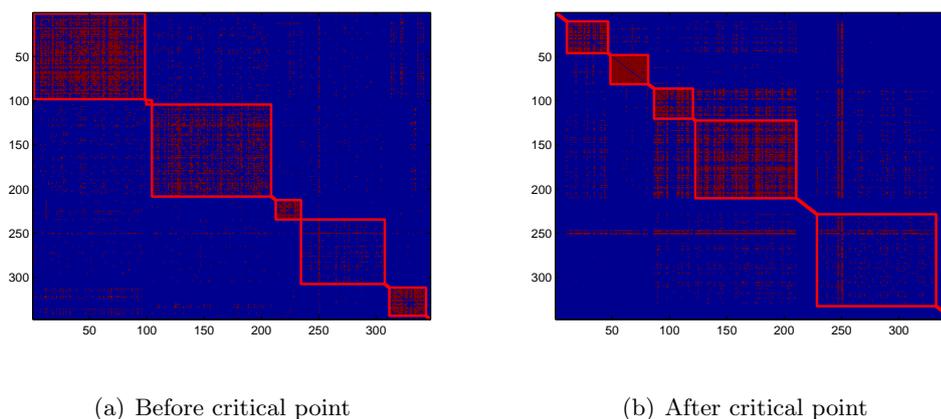
Figure 6.9: Matrix representation of the communities (indicated by squares) in the generative structure constructed from the time-evolving financial network before and after the critical point.

After 2002, some stocks in the market are more likely to remove the connections with the large-sized stock groups, which may be related to the fact that such structure could lower the risk of stock price falling caused by the crash of the important stocks, i.e., core vertices in the network.

6.5 Summary

To conclude, this chapter is motivated by the need to develop efficient tools for analyzing time-evolving network data. To this end, we have suggested a novel method for learning a generative model from a set of sample graphs in which the weights of the connectivity between vertices in the graph change between samples while the vertex number and label do not. We concentrate on the edge patterns present in those graphs and represent the connectivity changes on each individual edge with a Gaussian probability distribution, which is characterized by the structure of the generative model (or supergraph). This chapter then explores how this structure can be fitted to the sample graph data using an information theoretic approach with an MDL criterion, whose model complexity is encoded by the graph von Neumann entropy. To solve the data-fitting problem, we present a new fixed-point iteration scheme which optimizes the structure of the generative model. In the experiments, we show our proposed method provides a number of new directions to the time-evolving network analysis.

There are a number of ways in which the work reported here can be extended. First, in our analysis the edge connectivity between vertices is essentially the cross-correlation coefficient. It would be interesting to explore whether the Granger causality, which has been widely used to quantify the causal relationships between economic entities in the literature of econometrics, can reveal more information of the connections between stocks and provide better understanding of the connectivity structure. Moreover, it would be natural to consider that whether the method proposed for analyzing time-series data can contribute to modeling the dynamic of time-evolving networks. To do this, we could apply a Markov chain model to our generative model and investigate whether we can seek the evolutionary rules that govern the network dynamics.

Chapter 7

Conclusion

This chapter summarizes the main contributions in this thesis and analyzes the limitations of the methods we have developed. Also discussed in this chapter are the possible solutions for those limitations and potential future research directions.

7.1 Contributions

In this thesis, the overall research goal is to develop effective methods for characterizing the structural properties of complex networks and analyzing the time evolution of time-evolving networks. To this end, we have suggested a von Neumann entropy measure for directed graphs, and have explored its uses and applications for solving network analysis and machine learning problems such as graph classification and pattern recognition. In the second part of this thesis, we have proposed a novel thermodynamic framework for studying the properties of time-evolving complex networks. We have also developed a novel method for learning a generative model to capture the structure of a set of labeled graphs, which can be used to analyze graph time-series data.

The first substantial contribution in this thesis is the extension of the recently defined von Neumann entropy measure from undirected graphs to the domain of directed graphs. Graph-based entropy measures, such as Shannon entropy, von Neumann entropy and Körner's entropy, have proved to be an efficient tool for characterizing the graph structural complexity. However, despite that many entropy measures have been developed for the purpose of quantifying the complexity of undirected graphs, their corresponding methods in the directed graph domain have been less explored. This is due to the fact that graph theory underpinning directed graphs is less developed than that for undirected

graphs. Motivated by the need to overcome this problem, we have shown how to compute the approximate von Neumann entropy for directed graphs, by making use of some recent spectral graph theory results. We have also explored how the entropy expression can be further simplified on both weakly directed graphs, where most edges are bidirectional, and strongly directed graphs in which most edges are unidirectional. The resulting expressions are dependent on the vertex number as well as the degree statistics of connected vertices in the graph. We have shown the computational complexity of the approximate entropy is quadratic in graph size, which is lower than that of the original entropy computation.

We have also extended Estrada's heterogeneity index to directed graphs in order to quantify the vertex degree heterogeneous characteristics of directed graphs. The normalized version of this measure is bounded between 0 and 1, where the lower bound is reached for regular graphs and the upper bound is obtained for star graphs. We have demonstrated experimentally that both the entropy measure and the heterogeneity index are useful in characterizing directed graphs with different structure and detecting structural changes in evolving graphs.

The second substantial contribution in this thesis is to explore a number of uses and applications based on the development of the approximate von Neumann entropy. It is interesting to note that the entropy approximation can be expressed in terms of a sum of edge-dependent entropic contributions. As a result, a local measure that quantifies the von Neumann entropy associated with each edge in a graph can be defined. Combining this definition with the traditional degree assortativity measure, we have suggested a novel edge assortativity measure which quantifies the entropic preference of edges to form connections between similar vertices in both undirected and directed graphs. We have shown that this entropic edge assortativity measure is more effective in distinguishing the structural characteristics of graphs than the traditional degree-based analogue.

Another useful entropy application also commences from the approximation of the von Neumann entropy, whose expression gives the information of how the von Neumann entropy is distributed over edges in a graph. Such distribution of edge-based entropy contributions clearly encodes a number of properties of the intrinsic structural properties of a graph, allowing us to obtain a simple entropic characterization of graph structure, based on a histogram whose bins are indexed by vertex degree and whose bin-contents is the sum of entropy contributions associated with edges connecting vertices with specified degrees. We have shown how such histogram can further be encoded as a feature vector,

which captures the structural information of a graph in an entropic manner. In effect, we have proposed a graph embedding method that embeds both undirected and directed graphs into a low-dimensional feature space. This allows us to employ standard machine learning techniques to implement the tasks of graph classification and pattern recognition. The experiments have demonstrated that comparing to a number of state-of-the-art graph embedding methods, our method is more effective in terms of classification accuracy. Moreover, comparing to kernel methods, our method is particularly computationally fast with a competitive classification performance.

The third substantial contribution in this thesis is the development of a novel thermodynamic framework for visualizing and analyzing the structural changes during the evolution of time-evolving complex networks. The starting point of the idea is the link between a characteristic polynomial computed from the Laplacian matrix and the Boltzmann partition function of a graph. We have shown that when the graph Hamiltonian is equal to the negative Laplacian, the characteristic polynomial can be used to approximate the partition function, and a number of graph thermodynamic functions, such as average energy, entropy can be obtained. We have further investigated that assume the dynamic graph undergoes a constant-volume process, the expression for the temperature can also be obtained, which is defined as the change of energy with entropy. It is important to stress that all the thermodynamic variables are dependent on simple graph characteristics such as the number of vertices and degree statistics.

Turning attention to statistical mechanics, we note that the von Neumann entropy can be interpreted as a thermodynamic entropy when we associate the microscopic configurations of a network with the eigenstates of the normalized Laplacian spectrum. Furthermore, if we define that the microstate energy is proportional to the vertex degree, the internal energy of a network can be simply expressed as the number of edges. Moreover, the thermodynamic temperature is again the rate of change of energy with entropy, subject to the condition that the volume and number of particles are held constant. In the experiments, we have applied both thermodynamic representations to empirical network data in financial and biological areas. The results have suggested that both thermodynamic characterizations are useful in analyzing the structural evolutionary properties of dynamic networks. In particular, abrupt structural changes caused by critical events, such as financial crises in the financial data and morphological transformations in the biological data, can be well identified. Also different time stages in the time-varying networks are well

characterized in the thermodynamic space spanned by energy, entropy and temperature.

The final substantial contribution in this thesis aims at developing a novel method for constructing a generative model to analyze labeled graph data where the vertex set is fixed and the edge set changes between samples. This is accomplished by two steps. The first step is to propose a probabilistic framework for modeling the probability distribution of the observed sample graphs given the model structure, with a Gaussian distribution capturing the change of connection weights on each individual edge. The second step is then to fit the model structure to the sample graph data by adopting an information theoretic approach with a two-part MDL criterion in which the model complexity is reflected by the von Neumann entropy. This allows us to pose the problem of determining the optimal model structure as one of minimizing the code length criterion. To solve this problem numerically, we have developed a new fixed-point iteration scheme which gives an updating mechanism for locating the elements of the optimal weighted adjacency matrix of the model structure. Experimental evaluations have shown that by employing our proposed method to graph time-series data, the generative model provides a number of new directions to the study of the time-evolving complex networks.

7.2   Limitations

Although the methods we have developed in this thesis have proved to be effective in analyzing the structural characteristics of complex networks, and have outperformed some of the state-of-the-art graph characterization measures, there are a number of limitations with the methods proposed.

First, from the development of the graph von Neumann entropy measure, we note that the analysis commences from the definition that the von Neumann entropy of a graph is essentially the Shannon entropy associated with the normalized Laplacian eigenvalues. This clearly leads to a disadvantage of the resulting entropy measure, namely it is less powerful in characterizing cospectral graphs. In other words, it is difficult to use the approximate von Neumann entropy alone to distinguish cospectral graphs with different structural patterns. This shortcoming clearly suppresses the utility of this entropic characterization in a number of applications such as pattern recognition and object classification.

Another limitation with the development of the directed graph von Neumann entropy is that a sink vertex, which has no out-going edges connected to it, will make the entropy approximation formula of the weakly directed graphs singular. In our development, we

have discussed that in a weakly directed graph, the number of unidirectional edges is small, so the likelihood of the existence of sink vertices is also small. However, the possibility cannot be completely avoided, and it may diminish the applicability of the proposed entropy measure on a large number of networks. One way to avoid this problem, which we have addressed in the development of the von Neumann entropy, is to demand that the graph under study must satisfy the strong connectivity property, thus it can be guaranteed that there are no sinks in the graph. This requirement may seem to be a limiting constraint, however, we have stated that by summing up the entropy for each strongly connected component in a graph, the suggested entropy approximation is applicable to directed graphs that are not strongly connected.

Turning attention to the time-evolving network analysis, we have shown experimentally that, the thermodynamic characterizations, including the energy, entropy and temperature, which are computed from the approximate Boltzmann partition function, are efficient in identifying abrupt changes and phase transitions in structure and detecting other distinctive periods in the evolution of time-varying financial and biological networks. However, these thermodynamic functions do appear to have some limitations. For example, from the individual time-series of the thermodynamic variables, there exist some unexpected random fluctuations, which are not associated with any identifiable events in the time-series studied. Another limitation is that various critical events in the time-series do not necessarily give rise to unique patterns, which means that it is difficult to identify these events on the thermodynamic variable plots.

Finally, one limitation with the development of the generative model for the financial time-series analysis is that we have used the cross-correlation coefficients to measure the similarity of the stock returns as a function of the time shift between pairs of stock return time-series. The use of the cross-correlation is based on the fact that in realistic economics, the evolution of stock prices can be driven by common economic factors. However, it should be noted that such analysis can only be applied to measure the pairwise correlation between time-series. More generally, the causal relationship between a pair of time-series could be directed, or indirect, mediated by a third set of variables, or a combination of both. In this case, the pairwise analysis would lead to spurious results as it does not capture the whole covariance structure for multivariate data.

7.3 Future Work

In this section we point out the possible solutions to the limitations we have discussed and also provide some potential future research directions.

First, to solve the problem that the graph von Neumann entropy measure is less efficient in characterizing cospectral graphs, we turn to the probabilistic kernel methods based on the information divergence. Graph kernels have recently evolved into a rapidly developing branch of pattern recognition. Broadly speaking, there are two main advantages of the kernel methods, namely a) kernel methods can bypass the need for constructing an explicit high-dimension feature space when dealing with high-dimensional data and b) kernel methods allow standard machine learning techniques to be applied to complex data, which bridges the gap between structural and statistical pattern recognition. In particular, a kernel function is a symmetric and a positive definite function that maps two patterns to a real value, which is used to quantify the similarity between patterns.

On the other hand, an information divergence measures the dissimilarity between a pair of probabilistic distributions. Thus, it is natural to construct a link between kernels and information divergences, and define novel kernel functions based on particular information divergence accordingly. It is interesting to note that from the development of the entropic graph embedding method in the thesis, a local edge entropic measure which quantifies the von Neumann entropy contribution associated with each edge in the graph has been defined. Therefore, the information of the local measure of all edges in a graph can be viewed as a distribution of the von Neumann entropy on the graph. Such distribution can further be used to compute the information divergence between structures, which provides a new direction to the development of graph kernels. For instance, we could adopt the Jensen-Shannon divergence, which is a mutual information dissimilarity measure between probability distributions in terms of the difference between their associated entropies. In effect, the Jensen-Shannon divergence between a pair of graphs can be simply defined as the von Neumann entropy of the graph union minus the average of the two individual graph entropies. This allows us to define the Jensen-Shannon divergence graph kernel and diffusion kernel respectively, for both undirected and directed graphs. These graph kernels clearly provide a more powerful tool for characterizing graphs with various structural patterns.

Then, to further reduce the problem caused by the existence of the sink vertices in the development of the directed graph von Neumann entropy, i.e., to relax the strong

requirement of strong connectivity of directed graphs, it is worth in the future looking at the PageRank matrix instead of the Laplacian matrix. Specifically, the PageRank matrix allows a random walk to "teleport" or "dangle" on a directed graph, so it does not suffer from the problem of sink vertex and strong connectivity. Hence, it would be promising to explore whether a number of novel directed graph complexity measures can be defined and computed from the PageRank matrix of directed graphs.

Next, in order to improve the thermodynamic characterizations so that they can become more effective in identifying critical events and significant time stages in the evolution of time-varying networks, in the future we could turn our attention to the quantum physics. In particular, we would be interested in exploring whether partition functions from different quantum statistics, such as Bose-Einstein partition function and Fermi-Dirac partition function, can be used to replace the Maxwell-Boltzmann partition function which we have adopted in this thesis, for the purpose of providing a more efficient way to probe dynamic network structure.

In the future, it would also be intriguing to explore whether the Granger causality can reveal more causal information between time-series than the cross-correlation coefficients in the dynamic graph analysis. Put simply, a time-series is said to Granger-cause another one if the prediction of the latter time-series can be improved by containing the knowledge of the first one. A main advantage of Granger causality over the cross-correlation methods is that it can be used to measure not only the direct correlation, but also the indirect causal relationship (through a mediator variable) between a pair of time-series. Specifically, to efficiently identify the indirect causality between pairs of time-series, we could adopt the conditional Granger causality method, in which the mediator variable can either be an individual time-series or a set of time-series. Applying the Granger causality method to the financial and biological data, we would be interested in investigating whether the thermodynamic representations and the generative model could provide a more powerful tool for analyzing the time evolution of these dynamical systems.

In this thesis, although we have suggested a number of novel methods for characterizing and understanding the dynamics of time-evolving complex networks, the method for modeling such dynamics is lacking. One possible way to do this in the future would be to commence from the generative structure which is learned from a time-series of graphs and then to apply an appropriate Markov chain on this structure so as to control the evolution of the dynamic graph. The idea behind this suggestion is that since the generative model

is the underlying structure of a graph time-series, then the sample graphs in the time-series can be viewed as a sequence of outcomes which are generated from the generative model using some time-dependent rules that best describe the graph dynamics. Then, the problem is to propose a method for learning these evolutionary rules, which could be modeled by non-stochastic or stochastic processes. It would be particularly interesting to explore whether the model parameters learned from the time-series graph data are related to some graph characteristics, such as degree statistics and entropy measures, etc.

Finally, we acknowledge that we have explored a relatively limited quantity of empirical data. The methodologies reported in this thesis can clearly be extended to additional domains, such as human functional magnetic resonance imaging data and online social network data, to explore what features can be revealed in these complex systems. Furthermore, most graph characterization measures we have suggested in this thesis are based on unweighted graphs. In the future, it would be interesting to see whether these measures, can be generalized to the domains of weighted graphs (undirected and directed), labeled graphs and hypergraphs.

List of Symbols

$\mathcal{G}$ Graph

$\mathcal{V}$ Vertex set of a graph

$\mathcal{E}$ Edge set of a graph

u Vertex index

A Adjacency matrix of a graph

d_u Degree of vertex u

d_u^{in} In-degree of vertex u

d_u^{out} Out-degree of vertex u

L (Combinatorial) Laplacian matrix of a directed/undirected graph

D Degree matrix of a graph

$\tilde{L}$ Normalized Laplacian matrix of a directed/undirected graph

$\tilde{\lambda}$ Eigenvalue of normalized Laplacian matrix

H_{VN} Von Neumann entropy of a graph

r_u Local average degree ratio of vertex u

I Identity matrix

ω_{uv} Heterogeneity measure of edge (u, v)

Ω Heterogeneity index of a graph

$\tilde{\Omega}$ Normalized heterogeneity index of a graph

γ_{uv} Entropic contribution of edge (u, v)

Γ_{uv} Sum of entropic contributions associated with edge (u, v)

C_D Degree assortativity measure

C_E Entropic edge assortativity measure

Q_j Quantile index

V Feature vector

Z Boltzmann partition function

$\hat{H}$ Hamiltonian operator of a graph

E Thermodynamic energy

S Thermodynamic entropy

T Thermodynamic temperature

R Quasi characteristic polynomial

ζ Ihara zeta function of a graph

k Boltzmann constant

ϵ Residual

$\mathcal{K}$ Complete graph

$\mathcal{P}$ String graph

Δ_u Degree change at vertex u

$\mathbf{G}$ Graph time series dataset

$\tilde{\mathcal{G}}$ Generative model structure

$\tilde{W}$ Weighted adjacency matrix of $\tilde{\mathcal{G}}$

w_u Weighted degree of vertex u

$\mathcal{N}$ Gaussian distribution

$\mathcal{L}$ Code length

Abbreviations

PCA   Principal component analysis

PPI   Protein-protein interaction

MST   Minimum spanning tree

SL   Single linkage

PMFG   Planar maximally filtered graph

WD   Weakly directed

SD   Strongly directed

ECA   Entropy component analysis

CDF   Cumulative distribution function

SVM   Support vector machine

MUTAG   Mutagenicity

NCI   National Cancer Institute

SMO   Sequential minimal optimization

PUK   Pearson VII universal kernel

k-NN   k-nearest neighbour

COIL   Columbia object image library

HMO   Hückel molecular orbital

NYSE New York Stock Exchange

MDL Minimum description length

MLE Maximum-likelihood estimation

References

[1] T. Aittokallio and S. Schwikowski. Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, 7:243–255, 2006.

[2] R. Albert and A. L. Barabási. Topology of evolving networks: Local events and universality. *Physical Review Letters*, 85(24):5234–5237, 2000.

[3] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

[4] K. Anand and G. Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E*, 80(045102), 2009.

[5] K. Anand, G. Bianconi, and S. Severini. Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Physical Review E*, 83(036109), 2011.

[6] K. Anand, D. Krioukov, and G. Bianconi. Entropy distribution and condensation in random networks with a given degree distribution. *Physical Review E*, 89(062807), 2014.

[7] K. Anderson, C. Brooks, and A. Katsaris. Speculative bubbles in the S&P 500: was the tech bubble confined to the tech sector? *Journal of Empirical Finance*, 17(3):345–361, 2010.

[8] L. Antiqueira, F. A. Rodrigues, and L. da F. Costa. Modeling connectivity in terms of network activity. *Journal of Statistical Mechanics: Theory and Experiment*, 0905(4706), 2009.

[9] M. N. Arbeitman, E. E. Furlong, F. Imam, E. Johnson, B. H. Null, B. S. Baker, M. A. Krasnow, M. P. Scott, R. W. Davis, and K. P. White. Gene expression during the life cycle of Drosophila melanogaster. *Science*, 297(5590):2270–2275, 2002.

References

[10] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. *IEEE International Conference on Computer Vision (ICCV) - Workshop on Dynamic Shape Capture and Analysis (4DMOD)*, pages 1626–1633, 2011.

[11] L. Bai and E. R. Hancock. Graph kernels from the Jensen-Shannoni divergence. *Journal of Mathematical Imaging and Vision*, 47:60–69, 2013.

[12] L. Bai and E. R. Hancock. Fast depth-based subgraph kernels for unattributed graphs. *Pattern Recognition*, 50:233–245, 2016.

[13] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[14] S. Battiston and G. Caldarelli. Systemic risk in financial networks. *Journal of Financial Managements Markets and Institutions*, 1:129–154, 2013.

[15] C. H. Bennett. On the nature and origin of complexity in discrete, homogeneous, locally-interacting systems. *Foundations of Physics*, 16:585–592, 1986.

[16] D. Berwanger, E. Gradel, L. Kaiser, and R. Rabinovich. Entanglement and the complexity of directed graphs. *Theoretical Computer Science*, 463:2–25, 2012.

[17] G. Bianconi. Quantum statistics in complex networks. *Physical Review E*, 66(056123), 2002.

[18] G. Bianconi and A. L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86(5632), 2001.

[19] N. Biggs. Algebraic graph theory. *Cambridge Tracts in Math*, 67, 1974.

[20] N. Biggs. *Algebraic Graph Theory*. Cambridge: Cambridge University Press, 1993.

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424:175308, 2006.

[22] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccichè, N. Vandewalle, and R. N. Mantegna. Networks of equities in financial markets. *European Physical Journal B*, 38:363–372, 2004.

[23] K. M. Borgwardt and H. P. Kriegel. Shortest-path kernels on graphs. *Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 74–81, 2005.

[24] E. Z. Borzeshi, M. Piccardi, K. Riesen, and H. Bunke. Discriminative prototype selection methods for graph embedding. *Pattern Recognition*, 46:1648–1657, 2013.

[25] S. Braunstein, S. Ghosh, and S. 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.

[26] E. S. Browning. Exorcising ghosts of Octobers past. *The Wall Street Journal (Dow Jones & Company)*, page C1C2, 2007.

[27] G. Caldarelli, S. Battiston, D. Garlaschelli, and M. Catanzaro. Emergence of complexity in financial networks. *Lecture Notes in Physics*, 650:399–423, 2004.

[28] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009.

[29] M. Cavers, S. Fallat, and S. Kirkland. On the normalized Laplacian energy and general Randić index of graphs. *Linear Algebra and its Applications*, 433:172–190, 2010.

[30] F. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[31] F. Chung. Laplacians and the Cheeger inequailty for directed graphs. *Annals of Combinatorics*, 9:1–19, 2005.

[32] F. Chung and L. Lu. The diameter of sparse random graphs. *Advances in Applied Mathematics*, 26:257–279, 2001.

[33] J. C. Claussen. Offdiagonal complexity: A computationally quick complexity measure for graphs and networks. *Physica A*, 375:365–373, 2006.

[34] L. da F. Costa, O. Oliveira, G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antiqueira, M. P. Viana, and L. E. Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.

References

[35] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

[36] C. A. Coulson, B. O'Leary, and R. B. Mallion. *Hückel Theory for Organic Chemists*. Academic Press, 1978.

[37] C. Curme, M. Tumminello, R. N. Mantegna, H. E. Stanley, and D. Y. Kenett. Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, 14, 2014.

[38] M. Dehmer. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation*, 201:82–94, 2008.

[39] M. Dehmer, A. Mowshowitz, and F. Emmert-Streib. *Advances in Network Complexity*. Wiley-Blackwell, 2013.

[40] M. Dehmer, L. Sivakumar, and K. Varmuza. Uniquely discriminating molecular structures using novel eigenvalue-based descriptors. *Communications in Mathematical and in Computer Chemistry*, 67:147–172, 2012.

[41] J. C. Delvenne and A. S. Libert. Centrality measures and thermodynamic formalism for complex networks. *Physical Review E*, 83(046117), 2011.

[42] S. Deswal and A. Singhrova. Application of graph theory in communication networks. *International Journal of Application or Innovation in Engineering & Management*, 1:66–70, 2012.

[43] P. Erdős and A. Rényi. On random graphs, i. *Publicationes Mathematicae*, 6:290–297, 1959.

[44] F. Escolano, B. Bonev, and E. R. Hancock. Heat flow-thermodynamic depth complexity in directed networks. *Structural, Syntactic, and Statistical Pattern Recognition*, pages 190–198, 2012.

[45] F. Escolano, E. R. Hancock, and M. A. Lozano. Heat diffusion: Thermodynamic depth complexity of networks. *Physical Review E*, 85(036206), 2012.

[46] E. Estrada. Quantifying network heterogeneity. *Physical Review E*, 82(066102), 2010.

[47] E. Estrada. *The Structure of Complex Networks: Theory and Applications.* Oxford University Press, 2011.

[48] E. Estrada, M. Fox, D. Higham, and G. L Oppo. *Network science: Complexity in nature and techenology.* Springer-Verlag London, 2010.

[49] E. Estrada and N. Hatano. Statistical-mechanical approach to subgraph centrality in complex networks. *Chemical Physics Letters*, 439:247–251, 2007.

[50] D. Feldman and J. Crutchfield. Measures of statistical complexity: Why? *Physics Letters A*, 238:244–252, 1998.

[51] J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24):10815–10820, 2010.

[52] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[53] A. Fronczak, P. Fronczak, and J. A. Holyst. Thermodynamic forces, flows, and Onsager coefficients in complex networks. *Physical Review E*, 76(061106), 2007.

[54] J. Gehrke, P. Ginsparg, and J. M. Kleinberg. Overview of the 2003 KDD cup. *SIGKDD Explorations*, 5(2):149–151, 2003.

[55] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77:304–324, 1982.

[56] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):377–388, 1996.

[57] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[58] P. Grindrod and D. J. Higham. Evolving graphs: dynamical models, inverse problems and propagation. *Proceedings of the Royal Society A*, 466:753–770, 2009.

[59] L. Han, F. Escolano, E. R. Hancock, and R. C. Wilson. Graph characterizations from von Neumann entropy. *Pattern Recognition Letters*, 33:1958–1967, 2012.

References

[60] L. Han, R. C. Wilson, and E. R. Hancock. Generative graph prototypes from information theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(10), 2015.

[61] R. H. Heiberger. Stock network stability in time of crisis. *Physica A*, 393:376–381, 2014.

[62] R. v. d. Hofstad. *Random Graphs and Complex Networks.* Eindhoven University of Technology, 2010.

[63] P. Holme, C. R. Edling, and F. Liljeros. Structure and time evolution of an internet dating community. *Social Networks*, 26:155–174, 2004.

[64] K. Huang. *Statistical mechanics.* New York: Wiley, 1987.

[65] W. Huang, X. Zhuang, and S. Yao. A network analysis of the Chinese stock market. *Physica A*, 388:2956–2964, 2009.

[66] S. Ihara. Information theory for continuous systems. *World Scientific*, 1993.

[67] M. A. Javarone and G. Armano. Quantum-classical transitions in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(04):P04019, 2013.

[68] D. Jenkins. *Handbook of Airline Economics.* Aviation Week A Division of Mc, 2002.

[69] R. Jenssen. Kernel entropy component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:847–860, 2010.

[70] A. Kandel, H. Bunke, and M. Last. *Applied Graph Theory in Computer Vision and Pattern Recognition.* Springer Publishing, 2007.

[71] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. *International Conference on Machine Learning*, pages 321–328, 2003.

[72] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43, 1953.

[73] J. Kim and T. Wilhelm. What is a complex graph? *Physica A: Statistical Mechanics and its Applications*, 387:2637–2652, 2008.

[74] A. N. Kolmogorov. On tables of random numbers. *Theoretical Computer Science*, 207:387–395, 1998.

[75] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. *6th Prague conference on information theory*, pages 411–425, 1973.

[76] L. Kullmann, J. Kertész, and K. Kaski. Time-dependent cross-correlations between different stock returns: A directed network of influence. *Physical Review E*, 66(026125), 2002.

[77] L. Kullmann, J. Kertész, and R. N. Mantegna. Identification of clusters of companies in stock indices via potts super-paramagnetic transitions. *Physica A*, 287:412–419, 2000.

[78] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. *Proceedings of the Twelfth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 611–617, 2006.

[79] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1361–1370, 2010.

[80] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187, 2005.

[81] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (ACM TKDD)*, 1(1), 2007.

[82] G. Li, M. Semerci, B. Yener, and M. J. Zaki. Effective graph classification based on topological and label attributes. *Statistical Analysis and Data Mining*, 5(4):265–283, 2012.

[83] D. Luciani and G. Bazzoni. From networks of protein interactions to networks of functional dependencies. *BMC Systems Biology*, 6, 2012.

[84] B. Luo, Wilson R. C., and E. R. Hancock. Spectral embedding of graphs. *Pattern Recognition*, 36:2213–2230, 2003.

References

[85] B. Luo and E. R. Hancock. Structural graph matching using the EM algorithm and singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1120–1136, 2001.

[86] R. N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B*, 11:193–197, 1999.

[87] R. N. Mantegna and H. E. Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, 1999.

[88] T. Martin, B. Ball, and M. Newman. Structural inference for uncertain networks. *Physical Review E*, 93(012306), 2016.

[89] D. C. Mikulecky. Network thermodynamics and complexity: a transition to relational systems theory. *Computers & Chemistry*, 25:369–391, 2001.

[90] C. Mollenkamp. Lehman files for bankruptcy. *Wall Street Journal*, 2008.

[91] A. Mowshowitz. Entropy and the complexity of graphs. ii. the information content of digraphs and infinite graphs. *Bulletin of Mathematical Biology*, 30:225–240, 1968.

[92] A. K. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (coil-20). *Technical Report*, February(CUCS-005-96), 1996.

[93] M. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(208701), 2002.

[94] M. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[95] M. Newman. *Networks: an introduction.* Oxford University Press, 2010.

[96] C. Oliveira, N. Abreu, and S. Jurkiewicz. The characteristic polynomial of the Laplacian of graphs in (a,b)-linear classes. *Linear Algebra and its Applications*, 356:113–121, 2002.

[97] J. P. Onnela, K. Kaski, and J. Kertész. Clustering and information in correlation based financial networks. *The European Physical Journal B*, 38:353–362, 2004.

[98] G. Palla, A. L. Barabási, and T. Vicsek. Quantifying social group evolution. *Nature*, 446:664–667, 2007.

[99] F. Passerini and S. Severini. The von Neumann entropy of networks. *Inthernational Journal of Agent Technologies and Systems*, pages 58–67, 2008.

[100] T. K. DM. Peron and F. A. Rodrigues. Collective behavior in financial markets. *Europhysics Letters*, 96(48004), 2011.

[101] S. Pirzada and A. Dharwadker. Applications of graph theory. *The journal of the Korean Society for Industrial and Applied Mathematics*, 11(4), 2007.

[102] J. Platt. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.* MIT Press, 1999.

[103] N. Rashevsky. Life, information theory, and topology. *The bulletin of mathematical biophysics*, 17:229–235, 1955.

[104] P. Ren, T. Aleksić, D. Emms, R. C. Wilson, and E. R. Hancock. Quantum walks, Ihara zeta functions and cospectrality in regular graphs. *Quantum Information Processing*, 10:405–417, 2011.

[105] P. Ren, T. Aleksic, R. C. Wilson, and E. R. Hancock. A polynomial characterization of hypergraphs using the Ihara zeta function. *Pattern Recognition*, 44(9):1941–1957, 2011.

[106] P. Ren, R. C. Wilson, and E. R. Hancock. Graph characterization via Ihara coefficients. *IEEE Transactions on Neural Networks*, 22:233–245, 2011.

[107] K. Riesen and H. Bunke. Iam graph database repository for graph based pattern recognition and machine learning. *Structural, Syntactic, and Statistical Pattern Recognition*, 5342:287–297, 2008.

[108] K. Riesen, M. Neuhaus, and H. Bunke. Graph embedding in vector space by means of prototype selection. *Graph-Based Representations in Pattern Recognition*, 4538:383–393, 2007.

[109] S. Riis. Graph entropy, network coding and guessing games. *The Computing Research Repository*, 0711(4175), 2007.

[110] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal*, 6(1), 2002.

References

[111] S. Rosenberg. *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds.* Cambridge: Cambridge University Press, 1997.

[112] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg. Brenda, the enzyme database: updates and major new developments. *Nucleic Acids Research*, 32(D431-D433), 2004.

[113] G. Scott and C. Storm. The coefficients of the Ihara zeta function. *Involve - a journal of mathematics*, 1(2), 2008.

[114] A. Sensoya, S. Yuksela, and M. Erturk. Analysis of cross-correlations between financial markets after the 2008 crisis. *Physica A*, 392:5027–5045, 2013.

[115] N. Shervashidze, P. Schweitzer, E. J. V. Leeuwen, and K. M. Borgwardt. Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*, 1:1–48, 2010.

[116] N. Shervashidze, S. V. N. Vishwanathan, T. H. Petri, K. Mehlhorn, and K. M. Borgwardt. Efficient graphlet kernels for large graph comparison. *12th International Conference on Artificial Intelligence and Statistics*, pages 488–495, 2009.

[117] S. G. Shirinivas, S. Vetrivel, and N. M. Elango. Applications of graph theory in compter science: An overview. *International Journal of Engineering Science and Technology*, 2:4610–4621, 2010.

[118] F. N. Silva, C. H. Comin, T. K. DM. Peron, F. A. Rodrigues, C. Ye, R. C. Wilson, E. R. Hancock, and L. da F. Costa. On the modular dynamics of financial market networks. *ArXiv e-prints*, (1501.05040), 2015.

[119] L. Song, M. Kolar, and E. P. Xing. KELLER: estimating time-varying interactions between genes. *Bioinformatics*, 25(12):128–136, 2009.

[120] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Eurographics Symposium on Geometry Processing*, 28(5), 2009.

[121] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32:235–258, 2008.

[122] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102:10421–10426, 2005.

[123] M. van Steen. *Graph Theory and Complex Networks: An Introduction.* Maarten Van Steen, United States, 2010.

[124] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[125] N. Wiener. *The theory of prediction. Modern Mathematics for the Engineer.* McGraw-Hill, New York, 1956.

[126] R. C. Wilson, E. R. Hancock, and B. Luo. Pattern vectors from algebraic graph theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1112–1124, 2005.

[127] B. Xiao, E. R. Hancock, and R. C. Wilson. Graph characteristics from the heat kernel trace. *Pattern Recognition*, 42:2589–2606, 2009.

[128] C. Ye, R. C. Wilson, C. H. Comin, L. da F. Costa, and E. R. Hancock. Approximate von Neumann entropy for directed graphs. *Physical Review E*, 89(052804), 2014.

[129] C. Zou and J. Feng. Granger causality vs. dynamic Bayesian network inference: a comparative study. *BMC Bioinformatics*, 10(122), 2009.