Statistical Mechanics for Network Structure and Evolution

Jianjia Wang

Doctor of Philosophy

Computer Science

University of York

March 2018

Abstract

In this thesis, we address problems in complex networks using the methods of statistical mechanics and information theory. We particularly focus on the thermodynamic characterisation of networks and entropic analysis on statistics and dynamics of network evolution. After a brief introduction of background and motivation behind the thesis in Chapter 1, we provide a review of relevant literature in Chapter 2, and elaborate the main methods from Chapter 3 to Chapter 6.

In Chapter 3, we explore the normalised Laplacian matrix as the Hamiltonian operator of the network which governs the particle occupations corresponding to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. The relevant partition functions derive the thermodynamic quantities in revealing network structural characterisations. Chapter 4 further decomposes the global network entropy in three statistics on edge-connection components. This decompensation reflects the detailed distribution of entropy across the edges of a network.

Furthermore, Chapter 5 and Chapter 6 provide the theoretical approaches to analyse the dynamic network evolution and the application of the real-world networks. In Chapter 5, we investigate both undirected and directed network evolution using the Euler-Lagrange equation. This variational principle is based on the von Neumann entropy for capturing the topological variations of the time-varying network. Chapter 6 studies the fMRI regional brain interaction networks. We further develop a novel method for characterising networks and offer a high discrimination among patients with suspected Alzheimer's disease. Finally, Chapter 7 concludes the thesis and discusses the limitations of our methodologies, which also supplies the potential research in the future.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express the sincere gratitude to my responsible supervisor Professor Edwin Hancock, for his guidance throughout my PhD study and research at the University of York. His immense knowledge and great enthusiasm inspire me to explore challenges in a professional manner during these four years. His support has gone far beyond that of a PhD supervisor including being a careers advisor, fellow traveller and good friend. I am also grateful for his patient proofreading, editing, regular meetings and supporting my attendance in several overseas conferences.

I would also like to appreciate my assessor Dr William Smith and coauthor Professor Richard Wilson for their impartial assessment and constructive comments on my research. My sincere thanks also go to all the colleagues at the Department of Computer Science. I have formed friendships with fellow members of the CVPR group from all over the world. They provide an excellent environment for both work and study. Particularly thanks to the Department Overseas Research Student Scholarship (DORS) for the funding of my studentship.

Besides, I am grateful to all my friends and flatmates, especially to Jinghua Huang, Jixun Wu, Shuyi Peng, Kaichi Chan and Ruiming Li who are my first undergraduate students when working as a Teaching Assistant. Their great help and supports made me have a wonderful and enjoyable time during the last four years at York.

Lastly and most importantly, I would like to dedicate this thesis to all my family members for their enormous and unstinting supports in my study and my life. My heartfelt gratitude goes to my grandparents, Yansi Li, Zhiling Xu, Junhua Ye and Zhonghao Wang. The deep love in my heart, for all of you, will go on forever. I also would like to say thank

you to my parents, Hengnian Li and Weigang Wang, for their remarkable love in my upbringing and education. Finally, thanks to my amazing girlfriend, Liuyi Xia, without whom it indeed would not have been possible.

Declaration

I declare that the research described in this thesis is original work, which I undertook at the University of York during 2014 - 2018. Except where stated, all of the work contained within this thesis represents the original contribution of the author. This work has not previously been presented for an award at this, or the other, university. All sources are acknowledged as References. Most of the material in this thesis have been published in conference proceedings and journals. Below is a complete list of publications can be found.

Journal Papers

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Spin Statistics, Partition Functions and Network Entropy. *Journal of Complex Networks*, Oxford University Press, Vol. 89, No. 6, pp. 858–883, 2017.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Directed and Undirected Network Evolution from Euler-Lagrange Dynamics. In *Pattern Recognition Letters*, Elsevier, 2018.

Conference Papers

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Directed Graph Evolution from Euler-Lagrange Dynamics. In Proceedings of *24rd International Conference on Pattern Recognition (ICPR)*, IEEE Computer Society Press, 2018.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Quantum Edge Entropy for Alzheimer's Disease Analysis. *Joint IAPR International Workshop in Structural*,

Syntactic, and Statistical Pattern Recognition (S+SSPR), Lecture Notes in Computer Science, Springer, vol.11004, p449-459, Springer, 2018.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Euler Lagrange Network Dynamics. *International Conference on Energy Minimisation Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, Lecture Notes in Computer Science, vol.10746, p399-413, Springer, 2018.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Network Edge Entropy from Maxwell-Boltzmann Statistics. *International Conference on Image Analysis and Processing (ICIAP)*, Lecture Notes in Computer Science, vol.10484, p254–264, Springer, 2017.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Minimising Entropy Changes in Dynamic Network Evolution. *Graph-Based Representations in Pattern Recognition (GbRPR)*, Lecture Notes in Computer Science, vol.10310, p255-265, Springer, 2017.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Detecting Alzheimer's Disease using Directed Graphs. *Graph-Based Representations in Pattern Recognition (GbRPR)*, Lecture Notes in Computer Science, vol.10310, p94-104, Springer, 2017.

- Jianjia Wang: Report on S+SSPR 2016, Mexico. *The Newsletter of the British Machine Vision Association and Society for Pattern Recognition*, vol. 27, No. 3, March, 2017.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Network Entropy Analysis using the Maxwell-Boltzmann Partition Function. In Proceedings of *23rd International Conference on Pattern Recognition (ICPR)*, IEEE Computer Society Press, Los Alamitos, CA, USA, p1321-1326, 2016.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Thermodynamic Network Analysis with Quantum Spin Statistics. *Joint IAPR International Workshop in*

Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR), Lecture Notes in Computer Science, vol.10029, p153-162, Springer, 2016..

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: fMRI Activation Network Analysis using Bose-Einstein Entropy. *Joint IAPR International Workshop in Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR)*, Lecture Notes in Computer Science, vol.10029, p49-59, Springer, 2016.

- Jianjia Wang, Richard C. Wilson, Edwin R. Hancock: Partition Function in Complex Networks. In Presentations of *International School and Conference on Network Science (NetSci)*, Zaragoza, Spain, 2015.

Chapter 1

Introduction

In this chapter, we provide a roadmap detailing the research for this thesis. Commencing with the background of network science, especially regarding complex networks and network entropy, we present findings of our study regarding network complexity with entropy and the structural evolution over time. Then, we outline the motivation behind the study, the state-of-the-art methods available to solve related problems, and briefly describe our novel methods with statistical characterisation and network evolution. We propose our research goals, accomplishments, and the novel contributions made in the thesis. Finally, an outline of the thesis is provided at the conclusion of the chapter.

1.1 Network Science

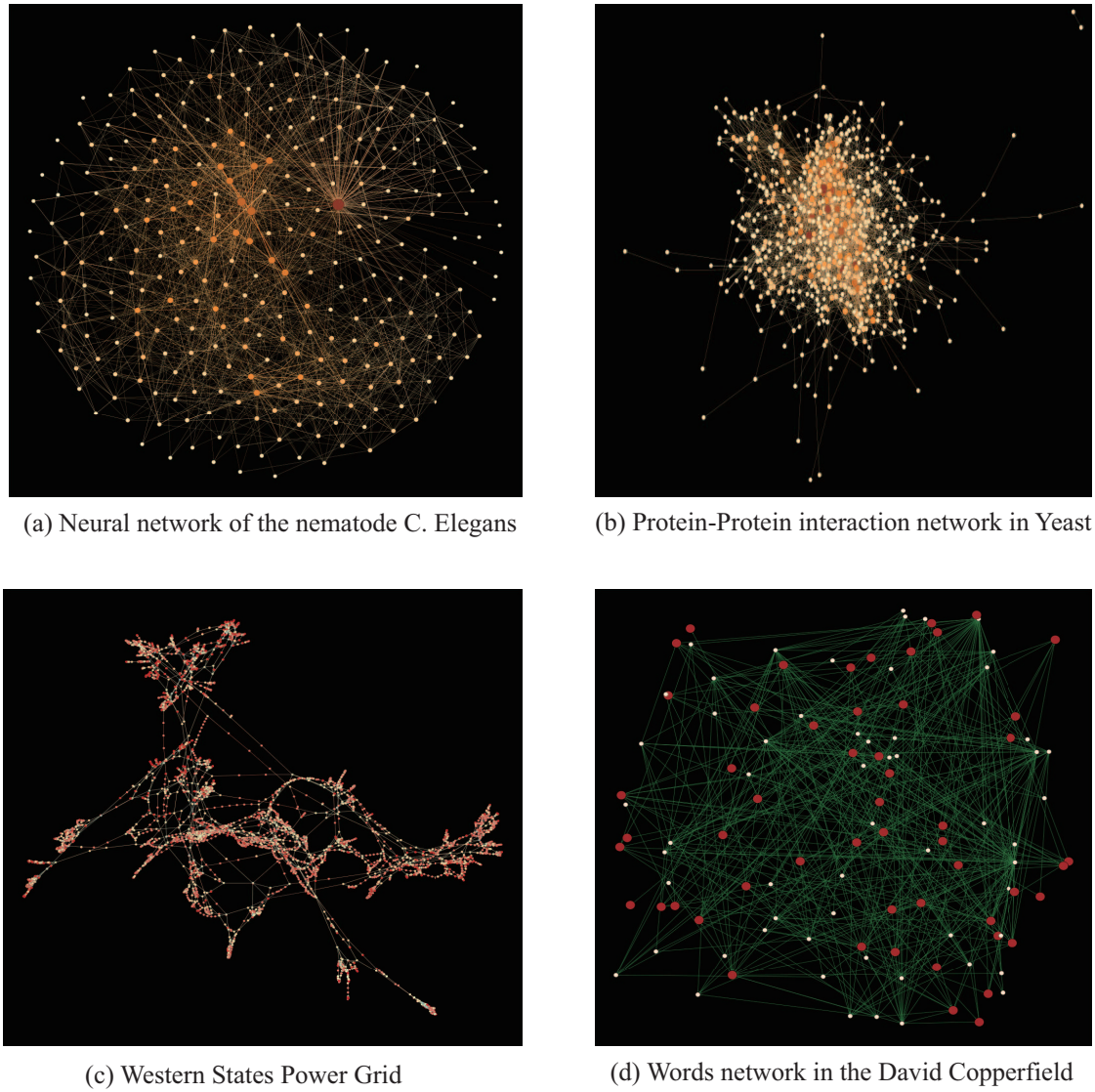
We are surrounded by a wide variety of systems in nature that can be represented as an abstract pattern of interactions or networks with vertices and edges [110, 62, 12]. Such network systems play a significant role in our daily life. For example, the Internet is comprised of enormous routers and computers in a kind of network connected by various physical or wireless links [3]; social relationships are another kind of network structure connected by human beings to spread ideas or knowledge [25, 102]. Trading markets maintain the financial networks for us to exchange goods and services for the economic prosperity [93, 83]. In nature, networks encode the interactions between genes, proteins, metabolites, and integrate chemical reactions into live cells [98, 62, 56]. The existence of a vast neural network in our brain, which includes the activity of billions of neurons,

holds the key for us to understand brain function and consciousness [48, 105, 37]. These are just a few of the many examples in the real world. Network science has proved to be an important innovation which allows us to investigate the mechanisms behind complex systems and the topological pattern in the network structure.

Network representation provides us with an abstract method to reduce complex systems to simple structural patterns of connection. Vertices and edges in a network can be labelled with additional information to capture more detail regarding their operation [98, 25]. In theory, tools developed to help us understand network characterisations can also be applied immediately to any other systems represented as a network [37]. If we want to understand complex systems thoroughly, we should first develop a deep understanding of the corresponding network structure behind them. In fact, most networks are driven by universal organising principles. Scientists have developed an extensive set of tools to analyse, model and predict them. These tools are complex and are comprised of developments from a wide variety of fields including mathematics, physics and computer science [45, 98, 33].

The study of networks can be traced back to the early 18th century which is known as the graph theory in the domain of discrete mathematics. The story begins with the work published by Leonhard Euler in 1736 with his historically notable solution to the problem of the Seven Bridges of Kőnigsberg [49]. Since the early days of the 19th century, the study of complex networks has been the territory of graph theory. A few historical remarks are established to provide the ideas of the interdisciplinary nature of topology. In 1960, a famous model was introduced by Paul Erdős and Alfréd Rényi which is known as the Erdős-Rényi random graph model [41, 42]. The traditional study of networks mainly focuses on regular graphs which provide a straightforward realization of complex networks. This model was relied on heavily in the past, however, it is not suitable for the realisation of the complex networks that are being studied today. Growing interest has prompted many scientists to review modelling paradigms without the fixed linking probabilities, and towards the end of the 20th century, we witnessed further movement in network research. Several novel concepts and measurement methods were proposed

and investigated in depth regarding large-scale complex networks [98]. Two well-known models which characterise the structural properties of complex networks are small-world networks [110, 75, 76] and scale-free networks [12, 13]. These both illustrate the specific statistical features of real-world network structure. The former specifies the short path lengths and high clustering in topology, while the later characterise the power-law degree distribution with preferential attachment [13].



(a) Neural network of the nematode C. Elegans

(b) Protein-Protein interaction network in Yeast

(c) Western States Power Grid

(d) Words network in the David Copperfield

Figure 1.1: Different kinds of Network. (a) A directed and weighted network representing the neural network of C. Elegans [110, 111]. (b) Protein-protein interaction network in budding yeast [29]. (c) An undirected and unweighted network representing the topology of the Western States Power Grid of the United States [110]. (d) Adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens [77].

The exploding interest in network science, particularly with regard to complex networks, reveals the fundamental laws and principles in the complex network domain [98]. Despite the apparent diversity in the real world, the structure and the evolution of the networks behind them is driven by a prevailing set of characterisations. If we disregard the nature of the components and the precise nature of the interactions, we find that these networks are more similar to one another than they are different from each other [56, 47]. Although there are hundreds of different technologies available which can analyse the universal principles of network structure, intellectual and scientific challenges are still encountered when we attempt to understand, describe, predict and eventually control the interwoven networks.

1.1.1 Complex Networks

Complex networks are large and varied networks which illustrate the interactions between the different parts of large complex systems [98, 25, 45, 109]. Well-known examples include social networks and power grids [45, 109]. These networks offer interesting and difficult challenges in terms of data analysis, as they are substantial and are characterised by relationships between objects rather than simple measurements. As a matter of fact, they are graphs [33, 46]. One of the most prominent challenges is to measure the complexity of the network. In essence, complexity is a measure of how much information exist within the network.

As the popularity of studying complex networks grows, many other inherent difficulties have attracted attention as we attempt to understand the possible complications in network structure [98, 25, 3]. Three main categories are usually cited as being useful when we illustrate network characterisation. The first is structural complexity [74, 44, 43]. Understanding the structural complexity is useful for identifying and classifying the network similarity as represented by the graphs. Although there is no widely accepted definition that can be used to define network complexity, there are still some parameters and structural features that people usually consider, such as the number of spanning trees [15], the length of pathways [48] and connectivity [47, 61], etc. The network complexity encodes structural and topological information to discriminate different kinds of networks [81, 59].

The second is to identify the connection diversity. The links between nodes could have distinct weights, directions and signs. Understanding of these so-called edge properties is still in its infancy for the rich characterisation of the practice and phenomenon [61, 4, 16]. The final consideration is to understand the network evolution. Since links and nodes can be created or could disintegrate at any moment, the network connection is not constant over time [42, 23, 70]. The evolution of networks usually occurs when connections are added or rewired from one component to another [78, 25]. Therefore, more effective and efficient methods of investigations are required if we want to understand the characterisation of complex networks.

Broadly speaking, many characterisations have been widely exploited, many different types of network structure have been classified, and networks have been analysed with regard to their evolution over time [46, 47, 2, 80]. Most of the available characterisations centre around ways of capturing network substructures using clusters, hubs and communities [47, 2, 80]. The underlying representations are based on simple degree statistics that capture the connectivity structures [110, 74]. These characterisations usually describe networks using macroscopic parameters [46, 103]. They delineate a particular network in terms of its structure, robustness, and performance or function through the statistics of linking and clustering [50, 37]. However, the structure of networks is not designed from the macroscopic perspective, and the connecting and relinking of individual nodes play a role in the microscopic structure [37]. The generalisation of these rules governs the network characterisations from both a purely deterministic domain to an entirely statistical domain [37, 2, 80].

1.1.2 Statistical Mechanics

Throughout scientific history one of the most influential innovations was the discovery of the laws of thermodynamics in the field of statistical physics [2, 80, 103, 86, 74, 114]. Statistical mechanics provides a framework based on which we can describe the macroscopic properties of matter from the microscopic points of view in the particles. This relationship provides a connection between the macro and the micro world in terms of thermodynamics [86]. For example, by using a heat bath analogy from thermodynamics,

principled physical measures of communicability and balance in networks can be defined [47]. Tools from statistical mechanics can also be used to characterise the degree distribution for different types of complex networks [2]. By maximising the ensemble entropy in exponential random graphs, the Boltzmann distribution developed in classical statistical mechanics can be used to predict the properties of time-evolving networks [80]. Furthermore, preferential attachment can lead to the phenomenon of condensation exhibited in growing networks [21, 23]. Both Bose-Einstein and Fermi-Dirac statistics have been used to account for the quantum geometry associated with different types of networks [20].

Another closely related approach is the heat bath analogy to graph spectrum from thermodynamics [114, 59, 108]. Classical statistical physics characterises the system with states to specify the energy dependence with the probability of finding certain states [37, 86]. This method can be extended in order to understand networks. A real-world network can also identify the possible energy states on the graph spectrum. This is known as the heat bath analogy, and it provides a convenient route to network characterisation. Here the energy states of a network are captured using the eigenvalues of a matrix representation of network structure [59, 6, 104]. The energy states are then populated by particles which are in thermal equilibrium with the heat bath. As a result of this thermalisation, the energy states are occupied according to the Boltzmann distribution [46, 114, 104]. Formally, this physical heat bath system can be described by a partition function with the energy micro-states of the network represented by a suitably chosen Hamiltonian. Usually, the Hamiltonian is computed from the adjacency or Laplacian matrix of the network, but recently, Ye *et al.* have shown that the partition function can be calculated based on a characteristic polynomial instead [114]. Moreover, thermodynamic and statistical analogy exploits various quantity in characterising the structural properties of the network [106]. From the commencement of mapping the network to a thermodynamic system, a partition function succinctly describes characterisations of the network such as entropy, total energy and temperature [46, 114]. For example, the Estrada index, as a measure of centrality and that is associated with the partition function of a network, describes the thermodynamic variables characterising network structures [46, 47]. These variables, such as

the entropy and internal energy, help us to interpret the structural properties within the graph spectrum [47]. Moreover, the micro-states of the network system can be explained by graph spectral theory [59, 117]. More investigations of network behaviour can be provided by this approach [33, 59].

Despite the interest in alternative models of the thermalised distribution of energy states, there has been no systematic study of the various thermodynamic characterisations resulting from different choices of statistical occupation, nor has there been a specific study of those associated with alternative assumptions concerning the graph spectrum with Laplacian matrix of the network [33, 119]. Here we attempt to understand the structural information by measuring the entropy of networks from thermodynamic characterisations [106, 104, 108]. The network complexity related to entropy cannot be specifically defined by different structural features when we take the topological complexity into account. Our goal is to find measurements which can encode unique structural information with a high degree of specificity. These parametric functions are based on metrical properties of graph spectra and present the notable features which allows for the detection of significant structural properties within networks and which can be used to identify network or graph similarities in pattern recognition problems and to predict the network evolution with appropriate data sets [106, 104, 108].

1.1.3 Network Entropy

Instead of describing the networks based on their structural characterisations, many types of research focus on quantifying network complexity using entropy as a measurement. The entropic measurements play an important role in understanding the structural and topological complexity of network systems. Borrowing from the ideas of information theory [5, 59], statistical mechanics [46, 2, 80, 103], and quantum information [84, 108, 23, 106], the measurement of entropy allows for a deep understanding of network evolution [107, 115] and an unveiling of the rich interplay between network topology and dynamics [93, 110]. More specifically, in the field of complex networks, the entropy of thermodynamic variables makes a great contribution to the information gathered on the network. The entropic quantities, such as Shannon entropy [23], Gibbs

entropy and von Neumann entropy [81, 59, 117], have proved to be useful for detecting significant structural characteristics of network ontologies [38].

The more complex, sophisticated, and unique the network structure, the more peculiar properties of entropy will be required to describe the networks. For example, collections of vertices usually share some similar properties that are known as communities [47]. The entropic measurements are useful for identifying communities with similar structural complexity and with shared attributes. The fact that they have the same entropy can be applied to explore symmetric and homogeneous networks [20, 94]. Furthermore, the concept of entropy can be used to inform inferences in the problems of complex networks [59]. The evaluation of the encoded information in network structures can allow for the challenging issues associated with inference to be addressed with network entropy. Usually, the statistical structure probability in the network is characterised as the likelihood using a generative model [60]. Some real networks present a high likelihood similarity during the evolution from which the phenomenon of preferential attachment associated with entropy emerges [119, 113].

Generally speaking, two kinds of entropy are well known in classical and quantum systems, i.e., the Shannon entropy [6, 10] and von Neumann entropy [6, 81, 59, 117], respectively. The Shannon entropy corresponds to the entropy for classical systems where statistical mechanics are used to interpret the various configurations within networks. It applies to the physical characteristics of states and specifies the energy dependence with regard to the probability of finding certain states [6]. For a network, it is sufficient and reasonable to describe the states of networks with all possible connection matrices. A network Hamiltonian introduces the energy of states [119]. This approach maps the network into the state of equilibrium where the specific Hamiltonian purportedly specifies the energy states associated with the topological properties. Formally, the network system can be described by a partition function with the energy micro-states of the network being represented by a suitably chosen Hamiltonian [104, 108]. Thermodynamic characteristics of the network, such as entropy, can then be derived from the partition functions

[74, 114]. By specifying the micro-states of the network system, statistical thermodynamics can provide deep insights into network behaviour.

On the other hand, von Neumann entropy is applied to describe the quantum statistics in a network system [81, 23, 63]. A mixed micro-state is a statistical mixture of pure quantum states which correspond to the Hilbert space with the maximum knowledge of the system [106]. The density matrix is used to describe the quantum state which represents the positive symmetric matrix with unitary trace [84]. The density matrix represents a convex combination of quantum states qualified by the von Neumann entropy [24]. It can be viewed as an extension of Shannon entropy that quantifies the incompressible information content of a quantum state [59].

The network entropy can be constructed from a density matrix as an operational meaning of connection between quantum mechanics and thermodynamics [106]. The network community has shown interest in discussing the challenging topics regarding the evaluation of the relationship between graph spectra and quantum states theoretically. Recently the spectra of the Laplacian or normalised matrix have provided a sophisticated way to define the von Neumann entropy and density matrix [81, 59]. Regarding the eigenvalues of Laplacian and quantum states, von Neumann entropy is derived from the spectrum of the Laplacian matrix [59, 117]. This efficiently and effectively distinguishes the different structures in extremal graph theory, where entropy is maximal for random graphs and minimal for regular graphs [59, 117]. Further extension approximates the calculation of von Neumann entropy in terms of simple degree statistics to reduce the quadratic complexity [106].

Overall, the entropic measurements play a crucial role in the quantification of the complexity of the network structure. It raises questions which merit investigation regarding the information that is encoded in the structural features of networks. Both Shannon entropy and von Neumann entropy offer novel methods to study the properties of pure states and mixed quantum states in network systems, which are central to the capture of differences and similarities between networks appearing in vastly different contexts [6].

1.2 Motivations and Goals

Although different physical analogies are useful for analysing the network characterisations, they are not always readily relatable to the spectral representation contained within the graphs presented in previous literature. The study of entropy in network science is an ongoing problem. The broad applicability of graph-based models offers a virtually limitless field for the use of entropy to measure structural differences [38]. Identification and classification of structural configurations in networks has shed light on the power of entropy measurements when compared to topological methods that were used in previous studies.

The goals of this thesis are to explore effective and efficient network characteristics and their evolution. We aim to exploit entropic quantities in characterising network structural properties in an effort to embark on thermodynamic and statistical analysis. We aim to find the structural information by comparing the entropy with network complexity and then utilising statistical models to derive entropic measurements with partition functions and the Hamiltonian operator. We aim to develop novel statistical models with network characterisations that allows us to address the issues that normally arise within complex networks. In this thesis, we have focussed on the problem of describing the network statistics and evolution with different partition functions. We apply the entropy to describe structural variations of edge-connectivity and time-varying evolution. Specifically,

- We explore the effects of occupation statistics on the populations of energy states when the Hamiltonian operator is the normalised network Laplacian, and the energy states are then given by its spectrum. Commencing from the heat bath analogy with the Laplacian matrix playing the role as the Hamiltonian, the energy states of the system are categorised according to a) Maxwell-Boltzmann, b) Bose-Einstein and c) Fermi-Dirac statistics respectively. Based on the relevant partition function, we use the statistical mechanical properties of the networks to calculate various thermodynamic quantities when the energy levels are occupied by particles in thermal equilibrium with the heat bath. We obtain different occupation statistics for the

energy levels by varying the partition function throughout the experiment. The network then can be characterised using thermodynamic quantities such as the entropy and energy derived from the relevant partition function [46, 114].

- We extend the entropy analysis of the heat bath analogy which provides a useful global characterisation of the network structure. We explore how to easily compute the entropy of edge or subnetwork structures, and attempt to provide a novel edge entropy projection which can be implemented at the global network entropy. We exploit this technique to analyse the distribution of edge entropy within a network and explore how this distribution reveals the intrinsic structural properties of different types of network.

- We explore whether the model of network entropy can be extended to detail the way in which the node degree distribution evolves with time, taking into account the effect of degree correlations caused by the degree structure of edges. We exploit this property by modelling the evolution of network structure using the Euler-Lagrange equations. Our variational principle is to minimise the changes in entropy during the evolution. By using our approximation of the von Neumann entropy, we are able to use update equations for the node degree which account for the effects of correlations induced by the edges of the network. It is effectively a type of diffusion process that models how the degree distribution propagates across the network.

- We explore whether the thermodynamic entropy can be used to construct an effective information theoretic graph-kernel for the purpose of classifying different types of graph or network structure. We construct a Jensen-Shannon kernel using the Bose-Einstein entropy for a sample of networks and then apply kernel principal components analysis (kPCA) to map graphs into low dimensional feature space. We apply the resulting method to classify fMRI activation networks from patients with suspected Alzheimer's disease. Furthermore, we are motivated to establish effective methods for measuring the structural properties of directed graphs representing inter-regional casual networks extracted from fMRI brain data. We aim to

use the directed network entropy to develop graph analytical methods to measure the degree of functional connectivity in brain networks.

1.3 Contributions

The significant contributions of this thesis are developing novel statistical models for network analysis with different partition functions, providing the study on edge entropy decomposition, minimising the entropy variance to model the network evolution, and applying the fMRI activation networks to distinguish Alzheimer's disease. These contributions are summarised as follows:

1.3.1 Partition Functions and Spin Statistics

The first contribution, outlined in Chapter 3, is to propose the thermodynamic characterisation of networks using the heat bath analogy when the energy states are occupied by different spin statistics, specified by a partition function. Applying the heat bath analogy and a matrix characterisation for the Hamiltonian operator, we consider the cases where the energy states are occupied according to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. We develop expressions for thermodynamic variables, such as entropy, for the system with particles occupying the energy states given by the normalised Laplacian eigenvalues. The chemical potential determines the number of particles at a given temperature. We provide a systematic study of the entropic measurements for network complexity resulting from the different partition functions and specifically those associated with alternative assumptions concerning the spin-statistics. Compared to the network von Neumann entropy to the corresponding normalised Laplacian matrix, these entropies are effective at characterising the significant structural configurations and distinguishing between the different types of network model (Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks). The effects of the spin-statistics are a) with regard to bosons - to allow the particles in the heat bath to congregate in the lower energy levels and b) with regard to fermions - to allow particles in the heat bath to populate higher energy levels. Bosons are more sensitive to the spectral gap in circumstances where a normalised Laplacian energy state exists and, hence, tend

to detect cluster or community structure, and fermions better sample the distribution of path lengths in a network. Numerical experiments for synthetic and real-world datasets are presented to evaluate the qualitative and quantitative differences that are present in the thermodynamic network characterisations derived from the different occupation statistics, which ultimately confirms these qualitative intuitions.

1.3.2 Edge Entropy Decomposition

The second substantial contribution, outlined in Chapter 4, is to propose a novel framework to show how to project edge-entropy components so that the detailed distribution of entropy across the edges of a network can be computed. This is particularly useful if the analysis of non-homogeneous networks with a strong community and hub structure is being attempted. To commence, we view the normalised Laplacian matrix as the network Hamiltonian operator which specifies a set of energy states with the Laplacian eigenvalues. The network is assumed to be in thermodynamic equilibrium with a heat bath. According to this heat bath analogy, particles can populate the energy levels according to the classical and quantum statistical distribution, and the distribution, together with the energy states, determines the thermodynamic variables of the network, such as entropy and average energy. We show how the entropy can decompose into components arising from individual edges using the eigenvectors of the normalised Laplacian. Compared to previous work based on the von Neumann entropy, this thermodynamic analysis is more effective in characterising changes in network structure since it better represents the edge entropy variance associated with edges connecting nodes of large degree. Numerical experiments on real-world datasets are presented to evaluate the qualitative and quantitative differences in performance.

1.3.3 Dynamic Network Evolution

The third contribution, outlined in Chapter 5, will investigate network evolution dynamics using the Euler-Lagrange equation. We use the Euler-Lagrange equation to develop a variational principle based on the von Neumann entropy for time-varying network structures. By utilising recent work to approximate the von Neumann entropy using simple degree

statistics, the changes in entropy between different time epochs are determined by correlations in the degree difference in the edge connection. Our Euler-Lagrange equation minimises the change in entropy and allows for the development of a dynamic model which predicts the changes of node degree with time. We first explore the effect of network dynamics on the three widely studied complex network models, namely a) Erdős-Rényi random graphs, b) Watts-Strogatz small-world networks, and c) Barabási-Albert scale-free networks. Our model effectively captures the structural transitions in the dynamic network models. We also apply our model to a time sequence of networks representing the evolution of stock prices on the New York Stock Exchange (NYSE). Here we use the model to differentiate between periods of stable and unstable stock price trading and to detect periods of anomalous network evolution. Our experiments demonstrate that the presented model not only provides an accurate simulation of the degree statistics in time-varying networks, but that is also captures the topological variations taking place when the structure of a network changes violently.

1.3.4 fMRI Network Application

The final contribution, outlined in Chapter 6, is to extend the theoretical approach to real-world networks and to discuss the application of fMRI brain network analysis. We present a novel method for characterising networks using the entropy associated with bosonic particles in thermal equilibrium with a heat bath. According to this analogy, the normalised Laplacian plays the role of the Hamiltonian operator, and the associated energy states are populated according to Bose-Einstein statistics. This model is subject to thermal agitation by the heat reservoir. The physics of the system can be captured using a partition function defined by the normalised Laplacian eigenvalues. Various global thermodynamic characterisations of the network including its entropy and energy then can be computed from the derivative of the corresponding partition function with respect to temperature. We explore whether the resulting entropy can be used to construct an effective information theoretic graph-kernel for the purpose of classifying different types of graph or network structure. To this end, we build a Jensen-Shannon kernel using the Bose-Einstein entropy for a sample of networks and then apply kernel principal components analysis (kPCA) to

map graphs into low dimensional feature space. We apply the resulting method to classify fMRI activation networks from patients with suspected Alzheimer's disease.

The neurobiology of Alzheimer's disease (AD) has been extensively studied by applying network analysis techniques to activation patterns in fMRI images. However, the structure of directed networks representing the activation patterns, and their differences in health and Alzheimer's people remain poorly understood. Here, we aim to identify the differences in fMRI activation network structure for patients with AD, late mild cognitive impairment (LMCI) and early mild cognitive impairment (EMCI). We use a directed graph theoretical approach combined with entropy measurements to distinguish subjects falling into these three categories from those within the normal healthy control (HC) group. We explore three methods. The first is based on applying linear discriminant analysis to vectors representing the in-degree and out-degree statistics of different anatomical regions. The second uses an entropic measure of node assortativity to gauge the asymmetries in the node with in-degree and out-degree. The final approach selects the most salient anatomical brain regions and utilizes the degree statistics of the connecting directed edges.

1.4 Thesis Outline

Having defined the problem in the domain of complex network and presented the overall goals of the thesis in Chapter 1, we propose a brief review of the relevant literature in Chapter 2, which includes a discussion about statistical mechanics in complex networks, network Hamiltonian, network entropy and dynamics.

Based on the heat bath analogy and the Hamiltonian operator, in Chapter 3, we detail a novel method that describes the thermodynamic characterisation of networks under different partition functions. We explore the case where the particle occupations correspond to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. From the related partition functions, we can compute the thermodynamic entropy and energy. Motivated by an interest in revealing the nontrivial properties of the network structure, we have compared the three resulting entropic characterisations with the von Neumann entropy. This study

investigates how the different entropies can be used to characterise changes in network structure, and how it can distinguish different types of network structure. Studies with synthetic data show that the entropies can distinguish Erdős-Rényi random graphs, Watts-Strogatz small-world networks, and Barabási-Albert scale-free networks. Experiments on real-world data, on the other hand, show that the thermodynamic variables can not only be used to detect both abrupt changes in network structure, but can also distinguish different classes of networks.

In Chapter 4, we combine the methods developed in Chapter 3 to explore the thermodynamic characterisations of networks, specifically those associated with the thermalisation effects of the heat bath on the occupation of the normalised Laplacian energy states. We extend the use of entropy as a tool to characterise network structures in both static and time series data. We conduct experiments which demonstrate that the thermodynamic edge entropy is better suited to represent the intrinsic structural properties associated with long-tailed degree distribution when compared with the extensively studied von Neumann entropy.

In Chapter 5, we apply the Euler-Lagrange equation to minimise the change associated with von Neumann entropy in the network structures. This treatment facilitates the prediction of the degree statistics varying with time and captures the effects of degree change correlations introduced by the edge-structure of the network. In other words, because of these correlations, the variance within one degree determines the translation of the connected nodes. We conduct numerical experiments using both synthetic and real-world network data in time evolution. Our model is capable of simulating the degree distribution and detecting significant variations in the network structure.

In Chapter 6, we demonstrate how to compute an information theoretic graph-kernel using Bose-Einstein entropy and the Jensen-Shannon divergence. This method is based on quantum statistics associated with the bosonic population of the normalised Laplacian eigenstates. By applying kernel PCA to the Jensen-Shannon kernel matrix, we are able to embed sets of graphs into a low dimensional space. We use discriminant classifier analysis to assign the graphs to different groups in order to evaluate the performance of thermal

entropies. The results of the experiment reveal that the method improves the classification performance for graphs extracted from fMRI data. The kernel method combined Bose-Einstein entropy and the Jensen-Shannon divergence provides an effective and efficient method for fMRI network analysis.

Furthermore, we are motivated to fill a gap in the literature regarding the analysis of fMRI regional brain interaction networks using directed graphs. We take advantage of the recently developed simplified approximations to the von Neumann entropy of directed graphs, which are dependent on the graph size and the in-degree and out-degree statistics of vertices. Assortativity of nodes in directed graphs provides insights into the neuropathology of Alzheimer's disease and allow us to characterise the functional organisation of the brain. Entropic measurements associated with node degree identify the edge connection features which offer high discrimination between subjects suffering from Alzheimer's disease and normal subjects.

Finally, in Chapter 7, we offer a brief conclusion regarding the advantages and shortcomings of the thesis. We summarise the contributions and make suggestions for future research. Overall, after developing the theoretical analysis, we present the experimental results on synthetic data and real-world networks throughout the thesis. It shows the potential applications of our theoretical methods.

Chapter 2

Literature Review

This chapter will review the existing literature about complex networks with consideration being given to statistical mechanics, entropic measures, quantum statistics and dynamic evolution. It starts with a discussion of network ensembles to introduce existing micro-canonical, canonical and grand-canonical ensembles and their associated complex networks. Then, the partition function is introduced to describe all possible configurations of network ensembles. We survey the thermodynamic concepts, such as entropy and temperature in networks, by quantifying the network with micro-states and heat bath analogy. Then, condensation phenomenon is observed in different network models and the quantum statistics are presented. Finally, we discuss structure, dynamics and other topics relevant to complex networks with statistical characterisations.

2.1 Network Characterisations in Statistical Mechanics

Statistical mechanics plays a vital role in helping us understand the important features of a network structure. They aim to develop effective characterisations of complex network structures and to interpret the process of network dynamics. These characterisations have been widely explored as the classification of different types of network structure and the methods of analysis pertaining to network evolution over time.

2.1.1 Network Ensembles

In 1878 in statistical mechanics, J. W. Gibbs introduced the concept of an ensemble to describe the microscopic properties of a thermal system [24]. Nowadays, people borrow

the idea of ensembles and apply it to thermal physics to analyse complex networks. Park and Newman [80] explored the properties of a graph ensemble with Boltzmann distribution. Bianconi quantified the complexity of networks with the concepts of ensembles in random networks [17, 18] and further extended our understanding to multiplex networks [19]. Waclaw described the problems associated with networks formulation with statistical ensembles by starting from the simplest construction of random graphs [103]. Garlaschelli and Loffredo proposed a grand-canonical ensemble model to construct networks with reciprocity [53]. All of these researchers regarded network ensembles as a fundamental tool in the analysis of complex systems.

The ensemble of networks can be used to construct networks with generalised hidden variables. Park and Newman assert that the ensemble models are not a single network, but rather a probability distribution over the whole set of possible networks [80]. Bianconi defines a network ensemble as a group of networks which satisfies certain structural constraints, such as degree distribution, community structure, etc [19]. Subsequently, certain undirected networks can be formulated in circumstances where we are given the number of links and nodes, the degree distribution or the community structure [18]. Further theoretical analysis presents the extensive constraints of network ensembles on the thermodynamic limits [17]. For example, given the total number of nodes and links, the degree of all nodes fix to the degree sequences $k_i, i = 1, \cdots, N$, and all networks with the same number of nodes form the ensemble with the uniform distribution $P(G) = 1/N$ [7]. The network ensembles follow the same structural constraints with the fixed distribution on average. They are under maximum-entropy constraints where the probability measurement in thermodynamic limits. The closer the graph G is to satisfying the constraints, the larger the probability value of certain structure $P(G)$ [7].

On the other hand, in terms of the statistical ensemble of networks, Garlaschelli *et al.* restrict the fixed number of vertices on unweighted networks without self-loops or multiple edges. Regarding each link as a "particle" between vertices, the constraint of "occupation number" for each pair of vertices follows certain statistics. Clearly, the adjacency matrix can be utilized to characterise this topology. Each possible adjacency

matrix corresponds to a configuration within a network structure, and the combination of these configurations defines the statistical ensemble of networks [52].

A similar description can be found in Waclaw's work [103]. Instead of defining the network ensembles as a statistical distribution for a network set, they refer to the network as the ideal gas containing the particles. Mapping the network with the same fixed number of nodes and links as the particles in a container, the canonical network ensemble emphasizes the conservation of the number of nodes and links, such as the thermal balance of ideal gas when it connects with a source of heat [103]. The partition function is quite useful to describe the configurations within the network ensemble. They define the partition function $Z(N,L)$ by summing over all combinations of nodes and links in adjacency matrices as [26]

$$Z(N,L) = \frac{1}{N!} \cdot C_L^{\frac{N}{2}} \tag{2.1}$$

where $C_L^{\frac{N}{2}}$ is the combination of the number of ways for choosing L links from $N/2$ edges.

Further considering the fluctuations of the number of edges, a new partition function was introduced with a chemical potential for links in the grand-canonical ensemble [26].

$$Z(N,\mu) = \sum_L exp(-\mu L) Z(N,L) \tag{2.2}$$

By analogy with classical physics, Waclaw defines a micro-canonical ensemble as a set of all equal probability graphs with prescribed sequences of degree $k_1, \cdots, k_N$. These degree sequences play a role in the micro-states. Then the micro-canonical ensemble is constructed by summing over all sequences obeying the conservation law $k_1 + \cdots + k_n = 2L$. With these definitions of the ensembles, Waclaw constructs complex networks with non-trivial statistical features, such as pow-law degree distribution, high clustering, degree-degree correlation etc [103, 26].

2.1.2 Micro-canonical and Canonical Ensembles

The micro-canonical ensemble is used to describe the possible states of a system constrained by total energy in statistical physics. And the canonical ensemble describes the system in terms of possible states within a heat bath at a certain temperature when it is in thermal equilibrium exchanging the total energy. Now, these two concepts have been extended to network systems in the literature.

A micro-canonical network ensemble means that structural constraints are strictly satisfied. A general framework of a random network ensemble can be built by using the micro-canonical ensemble and statistical mechanics [18]. A canonical network ensemble means these sets of structural constraints are satisfied on average [18]. This concept of network ensembles is consistent with the classical statistical description that considers the configurations of the system to be compatible with the fixed energy constraint, namely the micro-canonical ensembles, or the fixed average energy in the heat bath, which is the canonical ensembles [19]. By analogy, the random graph can be viewed as the micro-canonical ensemble which is formed by N nodes with a constant number of links L. Considering the Poisson distribution for the degree with an average $\langle L \rangle = p(N-1)$. It follows that the canonical ensemble of the random graph is formed by networks that satisfy the criteria as to the average number of links [5].

In addition, Anand and Bianconi use random graphs as an example in order to find the connection between micro-canonical and canonical network ensembles as the distribution reaches the thermodynamic limit [5]. The difference in entropy diverges between two cases when the imposed constraints are extensive. For a random graph with a fixed degree sequence, the entropy of the micro-canonical ensemble does not correlate to the entropy of the canonical ensemble case [5]. Recently, Bianconi *et al.* extend the micro-canonical and canonical ensembles to multiplex networks by implementing approaches from statistical mechanics [19]. They introduce the uncorrelated and correlated multiplex ensembles which related to the probability of the networks for every layer. They consider the multiplex ensembles with certain constraints on average, such as the fixed number of links and the degree sequence in each layer. They find the canonical ensembles are

more suitable to describe the overlap situation, while the micro-canonical ensembles can be extended to a number of other constraints [19].

2.2 Thermodynamic Variables in Networks

The analysis of statistical mechanics, in particular thermodynamic variables, within the network system allows us to gain a deep insight into network behaviour. The network can be succinctly described using a partition function when it is interpreted in accordance with the micro-states in the heat bath analogy. Additionally, thermodynamic variables, such as entropy, total energy, and temperature, can be derived from the partition function.

2.2.1 Network Entropy

The entropic measures provide a promising tool for understanding the structural and topological complexity of network systems. They may be able to resolve the issues encountered regarding evaluation of the network robustness and its ability to tolerate changes [90]. In 1978, E.T Jaynes first described the Maximum Entropy Principle in statistical mechanics. Since then, Strauss developed a class probability model for configurations of interacting points in 1986, which introduced this idea to graphs and lattices [97]. Park and Newman analysed exponential random graphs with Boltzmann distribution and Gibbs entropy. This means that in circumstance where we are given a set of networks, the expected properties to measure real-world networks can be derived from the data [80]. The network properties can be predicted by maximizing the Gibbs entropy of the graph ensemble when it is subjected to the constraints imposed by a given set of observations. Similar to the ideas contained in statistical mechanics, Lagrange multipliers were introduced, and graph Hamiltonian and partition function were defined [80].

Bianconi applies entropy to characterise randomized network ensembles that develop a logarithm to generate the total number of networks. They propose that entropy can be viewed as an indicator to assess structural features in network models [17]. Furthermore, Bianconi defines and evaluates the structural entropy to characterise undirected simple networks with certain constraints, such as the degree distribution [17]. It also reflects the community structure, in which case the networks might have to link probability with the

distance between the nodes [17]. Moreover, Anand *et al.* map the information theory to network topologies in order to quantify the complexity of networks [5]. They explain the relationship between the Shannon entropy, the Gibbs entropy and von Neumann entropy for the network ensembles, which can be used to solve the inference problems through maximum-entropy principle [5]. Krioukov *et al.* further explore entropy distribution in random networks given degree distribution [7]. They find that the network entropy has the property of self-averaging. The relative entropic variance vanishes in thermodynamic limit [19]. The fluctuations of entropy are also related to the average degree in networks [7]. Moreover, Bianconi finds that entropy is useful to solve inferential problems in multiplex networks [19]. In particular, the smaller the entropy of the ensemble, the smaller the number of networks satisfying the corresponding constraints, which implies that these networks are optimized [19].

In addition, a large number of approaches demonstrate that the entropy in thermodynamics is practically advantageous when it comes to measuring the robustness and complexity of the network. Since Boltzmann defined the general concept of entropy to the system when it is associated with different states, Shannon has extended the entropy principle to characterise communication systems with information theory [67, 86]. Such an application of the entropy principle can be used to characterise a network in terms of its complexity, robustness and heterogeneity. It has been extended to the graph spectrum domain. The normalised Laplacian spectrum can be viewed as a complexity level characterisation [81]. With the definition of von Neumann entropy, the density matrix associated with the Laplacian spectrum provides a novel way to study thermodynamic entropy on networks [81]. Anand, Bianconi and Severini study the relation between the Shannon entropy and the von Neumann entropy within networks exhibiting a given expected degree sequence [6]. Ye *et al.* [117] extend the approximate von Neumann entropy for directed graphs to characterise the structural complexity of networks. Overall, thermodynamic entropy has generally become an important mechanism through which we can characterise complex networks.

2.2.2 Network Temperature

In order to complete thermodynamic characterisations in networks, one parameter in particular should be mentioned. That is temperature. Generally speaking, there is a limited amount of literature that explores the impact of temperature on complex networks. Temperature is usually considered to be a pseudo-parameter. However, in order to completely describe the statistical formalism in complex networks, Garlaschelli *et al.* clearly discuss the impact of temperature on networks with a degree of topological optimization. They also developed various temperature-dependent versions of network models [52].

Furthermore, when using temperature as a parameter within a network, thermodynamic characterisations provide a convenient way to represent the structure within an associated graph spectrum. The normalised Laplacian has been shown to be related to the continuous time random walk and the heat flow on a graph [44]. Escolano *et al.* explain the thermodynamic depth relies on the heat flow to share the characterisation of a graph in entropy with statistical complexity [43]. Escolano, Bonev and Hancock extend the thermodynamic depth with their findings regarding heat diffusion on undirected and directed networks to quantify the complexity of structural patterns [43].

2.2.3 Partition Function

The partition function can be succinctly used to describe the network characterisations and properties. It refers to the statistical properties in the thermodynamic equilibrium. Thermodynamic characteristics of the network, such as entropy, total energy, and temperature can be derived from the partition functions.

Garlaschelli and Waclaw *et al.* [52, 103] view the networks as being at an equilibrium with the micro-canonical ensemble. It satisfies the constraints that have a fixed number of vertices N and a varying number of links $L_A = \sum_{ij} a_{ij}$, and which are controlled by the chemical potential μ. Then the probability of a graph A is introduced with temperature T

$$P_A = \frac{1}{Z} exp\left[\frac{\mu L_A - E_A}{T}\right] \tag{2.3}$$

where the partition function is $Z = \sum_A exp(\mu L_A - E_A)/T$. Considering the instructive

case, the energy E_A is summary of each individual link energies ε_{ij}. That is $E_A = \sum_{ij} a_{ij}\varepsilon_{ij}$. Then the partition function can be written as

$$Z = \sum_A \prod_{ij} e^{(\mu-\varepsilon_{ij})a_{ij}/T} = \prod_{ij}\left[1+e^{(\mu-\varepsilon_{ij})/T}\right] \tag{2.4}$$

So the probability of graph A is

$$P_A = \prod_{ij} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \tag{2.5}$$

And he probability of a link between node i and node j is

$$p_{ij}(T) = \frac{1}{e^{(\varepsilon_{ij}-\mu)/T}+1} \tag{2.6}$$

This form is the constant with quantum statistics, i.e., Fermi-Dirac statistics, which we review in the next section. The energy E_A implies that each link is drawn with probability p_{ij} independent of each other. Further, in order to simplify the form of ε_{ij}, Garlaschelli *et al.* obtained many important network models, such as the hidden-variable models and the configuration models of random graphs [52]. The optimization of topology in the low temperature case provides a deep understanding of the temperature-dependent network models, which in the end aid the investigation of the structural properties within complex networks [52].

2.3 Networks with Quantum Statistics

Quantum statistics are usually used to describe the statistical properties of complex networks. Quantum network states are characterised by quantum occupation numbers which are mapped relative to the nodes, links, and triangles.

In particular, the quantum states are characterised by the Hamiltonian operator, which defines the energy spectrum of the network system. Günthard and Primas first realised that the matrix form of Hamiltonian can be related to the adjacency matrix of a certain graph [55]. Gutman introduced the definition of graph energy to the absolute

eigenvalues of the adjacency matrix [57]. Further extension of the energy spectrum is defined as the eigenvalues of the Laplacian matrix [58]. Similar variants of the Hamiltonian operator are developed for the signless Laplacian [95], the distance matrix [54], the incidence matrix [64] and, recently, the normalised Laplacian for the connection to the Randić index [81].

By defining the micro-states within the network system, the quantum statistical picture allows for Bose-Einstein and Fermi-Dirac statistics to be used in two different ways. For example, the fitness of the nodes in the scale-free network model can be an analogy to the Bose gas when following the Bose-Einstein statistical properties [16]. Similar mapping of the Fermi gas can describe the growing Cayley trees in Fermi-Dirac statistics [15].

2.3.1 Bose-Einstein Condensation

In terms of Bose-Einstein statistics, one of the most interesting phenomena is the Bose-Einstein condensation. Strauss first discovered the condensation transition of a network model in 1986 [97]. He described these networks as the framework of an equilibrium model. More recently, Bianconi mapped the nonequilibrium scale-free growing network model with a fitness parameter to a Bose gas. This is known as Bose-Einstein condensation [21]. Starting with the fitness model, every node i is assigned an energy ε_i relating to the fitness distribution $\rho(\eta)$ which describes the ability of a node to attract new links.

$$\varepsilon_i = -\frac{1}{\beta}\ln \eta_i \tag{2.7}$$

where $\beta = 1/T$ plays the role of inverse temperature. A link between two nodes i and j with energies ε_i and ε_j corresponds to two noninteracting particles on the energy levels ε_i and ε_j. Adding a new node to the network corresponds to adding a new energy level ε_i and $2m$ particles to the system. The probability of a particle lands on the energy level ε_i is given by

$$\Pi_i = \frac{e^{-\beta\varepsilon_i}k_i}{\sum e^{-\beta\varepsilon_i}k_i} \tag{2.8}$$

The occupation number $k_i(\varepsilon_i, t, t_i)$ denotes the meaning that the number of links (particles) on the energy level ε_i at time t, while adding nodes to the system at time t_i. The rate at which particles accumulate on the energy level is given by

$$\frac{\partial k_i(\varepsilon_i, t, t_i)}{\partial t} = m\frac{e^{-\beta\varepsilon_i}k_i(\varepsilon_i, t, t_i)}{Z_t} \tag{2.9}$$

where Z_t is the partition function, defined as

$$Z_t = \sum_{j=1}^{t} e^{-\beta\varepsilon_i}k_j(\varepsilon_j, t, t_j) \tag{2.10}$$

Based on the assumption that each node increases its connectivity following a power law, the occupation number can be given as

$$k_i(\varepsilon_i, t, t_i) = m\left(\frac{t}{t_i}\right)^{f(\varepsilon_i)} \tag{2.11}$$

where the dynamic exponent $f(\varepsilon)$ satisfies $f(\varepsilon) = e^{-\beta(\varepsilon-\mu)}$, μ plays role of the chemical potential, satisfying the equation

$$\int deg(\varepsilon)\frac{1}{e^{\beta(\varepsilon-\mu)}-1} = 1 \tag{2.12}$$

where $deg(\varepsilon)$ is the degeneracy of the energy level ε. This equation suggests that in the $t \to \infty$ limit, the occupation number, given the number of particles with energy ε, follows the familiar Bose statistics [2, 21]

$$n(\varepsilon) = \frac{1}{e^{\beta(\varepsilon-\mu)}-1} \tag{2.13}$$

Each time a new node is added to the network, the links of this node exhibit a higher probability that they will attach to others with high fitness. That shows that the high connectivity nodes follow a generalised preferential attachment rule [12]. This kind of network model exists scale-free property [22]. When the node fitness is significantly higher than the mean value, this node grabs a finite fraction of all the links in the network.

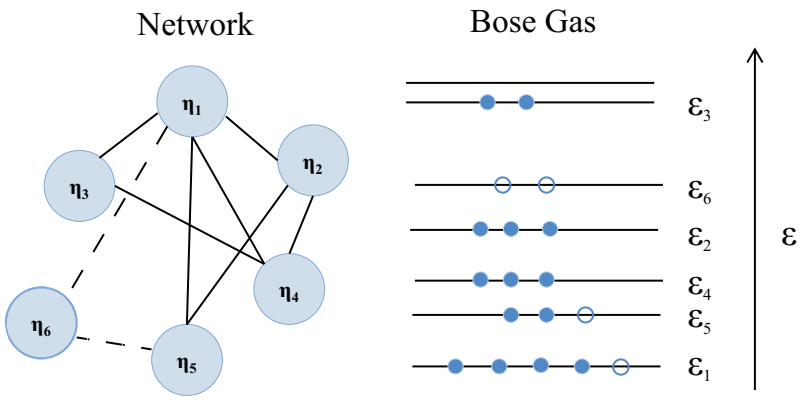


Figure 2.1: Schematic illustration of the mapping between the network model and the Bose gas. The fitness parameter η corresponds to the energy state ε, and a link between node i and node j corresponds to particles at the energy state ε_i and ε_j. The solid dots on the energy levels are the existing links. The cycle dots are the new links to be added in the network [21].

This phase transition can be mapped to the Bose-Einstein condensation in a Bose gas [21].

Furthermore, Ferretti, Mamino and Bianconi extend this work with their work pertaining to the rewiring of links [50]. The rewiring process tends to remove links from nodes that have a high negative fitness to optimize the network structure. Rewiring probabilities also satisfy the constraint of the network to be a simple graph. At low temperatures and high rewiring rates, this constraint induces a Bose-Einstein condensation, which, in turn, introduces the phase space with the connected component and the degeneracy of the networks [50]. They provide a mean-field solution to the model of the condensation phase transition. Via numerical simulation and analytical arguments below the phase transition, the structure is significantly different to the links near the neighbours of the condensed nodes [50].

Besides, Penrose also finds Bose-Einstein condensation in a solvable model of hard-core repulsive bosons on the complete graph [82]. In recent years, Bose-Einstein condensation has also been shown to take place in networks with complex topology, such as those with a set of infinite linear chains crossing at a single site [28], linear chains connected by a single line [30], an Apollonian scale-free network [36], and the infinitely ramified star and wheel graphs [101]. By analogy, the networks like the ideal Bose gas depict a spectral density with an anomalous behaviour at the bottom of energy states. They present a

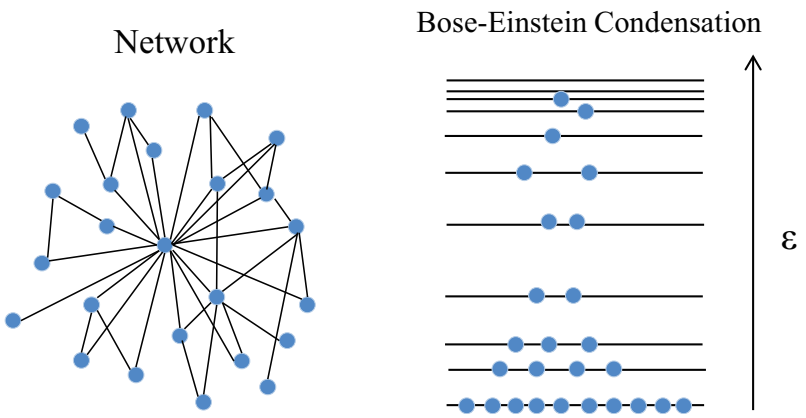


Figure 2.2: Bose-Einstein condensation in the network. The node fitness is significantly higher than the mean value, and this node grabs a finite fraction of all the links. A high density of particles occupies at the lowest energy state [21].

discontinuous jump of the specific heat at the transition temperature. Typically the critical temperature is proportional to the particle density. The ground state usually has a finite particle density located within a given region of the network, i.e. the condensation has a trapped fraction [101].

Interestingly, Anand *et al.* characterise the distribution of entropy of random networks with a given degree distribution [7]. They show that the condensation of the average degree is different from the Bose-Einstein condensation in complex networks. After defining the entropy of a random network ensemble, they find the fluctuations of entropy are mainly determined by the fluctuations of the average degree. Networks with an average degree exceeding a certain threshold exhibit large deviation or condensation effects, which means that a single node can attract $O(N)$ links. This is not the same as typical Bose-Einstein consideration in complex networks. Generally, they only correspond to some large deviation configurations in network ensembles [7].

2.3.2 Fermi-Dirac Statistics

Particles with half-integer spin are subject to Fermi-Dirac statistics and obey the Pauli exclusion principle. They give rise to models of network structure which are constrained by the occupancy of the nodes and edges. Examples include traffic flow and the modelling of certain types of geometric networks, such as the Cayley tree [15]. It can be used to

describe the form of super-symmetry multiplex networks [20]. Kriukov *et al.* detail a geometric framework to study the structure and functions of complex networks, interpreting edges as noninteracting fermions whose energies are hyperbolic distances between nodes [69]. Shen, Zhu and Liu discuss an inverse approach to network evolution defining a relative with an illness model and Fermi-Dirac statistics [92]. Baronchelli, Catanzaro and Romualdo define a framework using bosonic reaction-diffusion processes, with the aim of analysing dynamic systems on complex networks [14]. Javarone and Armano propose a theoretical model of network evolution inspired by fermions, which maps complex networks to Fermi gas [63]. They show that the emergence of different network structures can be represented in terms of quantum-classical transition [63].

2.4 Structural and Dynamic Networks

Except for thermal physics, many statistical methods are used to analyse complex networks based on their structural properties, dynamics, phase transition, modelling, etc. Martin, Zhang and Newman [72] use eigenvector centrality to measure the importance of nodes in the networks. This model relies on the small number of nodes which concentrate most of the weight of the centrality. Based on the non-backtracking matrix, an alternative method was established to avoid localisation and it is much more useful when it comes to fixing the problems in circumstances where the standard centrality fails. Zhang, Martin and Newman [118] use methods of statistical inference to detect the core-periphery structure within networks. They combine an expectation-maximization algorithm and a brief propagation algorithm to efficiently deconstruct networks into dense-core plus an outlying structure. Domenico *et al.* [39] introduce von Neumann entropy into multiplayer networks in order to distinguish different layers. They aggregate networks to minimise the number of layers and maximize the ability to distinguish between the multiplayer networks. Furthermore, they identify modular flows on multiplayer networks to reveal structures with a high level of overlap [40].

Zuev, Papadopoulos and Krioukov [119] describe the dynamics of complex networks

with Hamilton's equations. They derive the explicit form of the Hamiltonian in the canonical formalism to govern the growth of a network within the preferential attachment. Bouna, Kitsak and Krioukov [50] consider networks arising in cosmology. They show that networks grow following the power-law degree distribution with Lorentz-invariant. It encodes the maximum information about rewiring of the links which occurs according to a preferential attachment rule [50]. Ferretti, Mamino and Bianconi [50] consider growing networks with both heterogeneity of nodes and topological constraints. They found that at low temperature and high rewiring rates, a new phase transition was induced by an extended condensate of links. This transition further extends to the size of the connected component and the degeneracy of the networks [50].

Wu *et al.* [113] built a model to characterise the geometrical properties of dynamic networks. The growing geometrical networks follow the non-equilibrium rule which can generate scale-free networks with clustering and communities and planar random geometry with non-trivial modularity. Ostilli and Bianconi [79] detail a statistical mechanical approach to extract the coordinates of the nodes in random geometric graphs. They reveal the mechanism behind the typical configurations of the network model and explore the finding that the distribution of nodes is either uniform or condensed at the temperature limit [79]. The network structural transition is characterised by connectivity.

2.5 Summary

In this chapter, we briefly reviewed the existing literature about complex networks with regard to statistical mechanics, thermodynamic variables, quantum statistics and dynamic structure evolution. It started with a discussion about network ensembles and introduced existing micro-canonical and canonical ensembles within complex networks. Following the thermodynamic concept, literature about entropy and temperature in networks was then presented. Then, condensation phenomenon is observed in networks models and quantum statistics, i.e. Bose-Einstein statistics and Fermi-Dirac statistics, are employed in constructing different networks. Finally, we present some the state-of-art topics, namely structure, dynamics and others in complex networks with evolution.

Chapter 3

Partition Functions and Spin Statistics

In this chapter, we explore the thermodynamic characterisation of networks using the heat bath analogy when the energy states are occupied by different spin statistics, specified by a partition function. Utilising the heat bath analogy and a matrix characterisation for the Hamiltonian operator, we consider the cases where the energy states are occupied according to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. We derive expressions for thermodynamic variables, such as entropy, for the system with particles occupying the energy states given by the normalised Laplacian eigenvalues. The chemical potential determines the number of particles at a given temperature. We provide the systematic study of the entropic measurements for network complexity resulting from the different partition functions and specifically those associated with alternative assumptions concerning the spin statistics. Compared with the network von Neumann entropy corresponding to the normalised Laplacian matrix, these entropies are effective in characterising the significant structural configurations and distinguishing different types of network models (Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks). The effect of the spin statistics is a) in the case of bosons to allow the particles in the heat bath to congregate in the lower energy levels and b) in the case of fermions to populate higher energy levels. With normalised Laplacian energy states, this means that bosons are more sensitive to the spectral gap and hence to cluster or community structure, and fermions better sample the distribution of path lengths in a network. Numerical experiments for synthetic and real-world datasets are

presented to evaluate the qualitative and quantitative differences of the thermodynamic network characterisations derived from the different occupation statistics, and these confirm the qualitative intuitions.

3.1 Introduction

The literature contains many accounts of work aimed at developing effective characterisations of complex network structure. These characterisations have been widely exploited in both cluster and classify different types of network structure, and also to analyse how networks evolve with time [46, 47, 2, 80]. Broadly speaking, most of the available characterisations have centred around ways of capturing network substructure using clusters, hubs and communities [47, 2, 80]. The underlying representations are usually based on simple degree statistics that capture the connectivity structures [110, 74]. Although many of the methods available are goal-directed, most promising approaches are to draw on ideas from physics, using analogies based on statistical mechanics [2, 46, 80], thermodynamics [114] or quantum information [6].

One of the most powerful of these approaches is to use thermodynamics analogies suggested by statistical physics. For instance, by maximizing the ensemble entropy in exponential random graphs, the Boltzmann distribution from classical statistical mechanics can be used to predict the network properties of time-evolving networks [80]. Tools from statistical mechanics can also be used to characterise the degree distribution for different types of complex networks [2]. Furthermore, by using a heat bath analogy from thermodynamics, principled physical measures of communicability and balance in networks can be defined [47]. Ideas from quantum information theory are also useful in the understanding network structure. For instance, the preferential attachment can lead to the phenomenon of condensation exhibited in growing networks [21]. Both Bose-Einstein and Fermi-Dirac statistics have been used to account for the quantum geometries associated with different types of networks [20]. Although these different physical analogies are useful, they are not always easily related to the graph spectral representation.

Another closely related approach is heat bath analogy which provides a convenient

route to network characterisation. Here the energy states of a network are captured using the eigenvalues of a matrix representation of network structure. The energy states are then populated by particles which are in thermal equilibrium with the heat bath. As a result of this thermalisation, the energy states are occupied according to the Boltzmann distribution [46, 114]. Formally, this physical heat bath system can be described by a partition function with the energy micro-states of the network represented by a suitably chosen Hamiltonian. Usually, the Hamiltonian is computed from the adjacency or Laplacian matrix of the network [88], but recently, Ye *et al.* [114], have shown how the partition function can be computed from a characteristic polynomial instead.

To embark on this type of analysis, partition functions can be succinctly used to describe the network statistics and evolution. Thermodynamic characterisations of the network, such as entropy, total energy, and temperature then can be derived from the partition functions [74, 114]. By specifying the micro-states of the network system, statistical thermodynamics can provide deep insights into network behaviour. For example, by using the Maxwell-Boltzmann partition function to describe a thermalised network, the entropy, internal energy, and the Helmholtz free energy can be computed from the graph spectra, and this leads to natural definitions of notions such a centrality [46, 114].

However, the Boltzmann distribution does not take into account particle spin-statistics and their effects on the population of the thermalised energy levels. Unlike the classical case where particles are distinguishable, in quantum statistics particles are indistinguishable. Particles with integer spin are subject to Bose-Einstein statistics and do not obey the Pauli exclusion principle. As a result, they can aggregate in the same energy state. At low temperature, this leads to the phenomenon of Bose-Einstein condensation. There has been work aimed at extending the heat-bath model to take such effects into account. For instance, Bianconi and Barabási [21] have constructed a network model based on a Bose gas, and have studied the phase transitions in network structure associated with Bose-Einstein condensation [21]. This model has also been extended to understand processes such as supersymmetry in networks [47, 20]. On the other hand, particles with

half-integer spin are subject to Fermi-Dirac statistics and obey the Pauli exclusion principle. They thus give rise to very different models of network structure, and these have been exploited to model situations where there are constraints on the occupancy of the nodes and edges of a network. Examples include traffic flow and also the modelling of certain types of geometric networks such as the Cayley tree [15, 92].

Despite the interest in alternative models of the thermalised distribution of energy states under different particle spin statistics, there has been no systematic study of the various thermodynamic characterisations resulting from different choices of partition functions, and specifically those associated with alternative assumptions concerning the spin statistics. Here we consider the effects of occupation statistics on the populations of energy states when the Hamiltonian operator is the normalised network Laplacian, and the energy states are then given by its spectrum. Commencing from the heat bath analogy with the Laplacian matrix playing the role as the Hamiltonian, the energy states of the system are occupied according to a) Maxwell-Boltzmann, b) Bose-Einstein and c) Fermi-Dirac statistics respectively. From the relevant partition function, we use the statistical mechanical properties of the networks to compute various thermodynamic quantities when the energy levels are occupied by particles in thermal equilibrium with the heat bath. Making different choices for the partition function, we obtain different occupation statistics for the energy levels. The network then can be characterised using thermodynamic quantities such as the entropy and energy derived from the relevant partition function [46, 114]. In qualitative terms, the Pauli exclusion principle means that particles subject to Fermi-Dirac statistics are populated the energy states less densely than that in the classical Maxwell-Boltzmann case. On the other hand, since particles obeying Bose-Einstein are indistinguishable, they populate the energy states more densely.

The thermodynamic picture offered by quantum Bose-Einstein and Fermi-Dirac statistics differs from that offered by classical Maxwell-Boltzmann statistics in a number of important ways. Both quantum statistics additionally require a chemical potential to specify the distribution of states in the partition function. The chemical potential is determined by the heat reservoir, and modifies the occupation probability of the energy

levels [24]. In both cases of Bose-Einstein and Fermi-Dirac statistics, for energy levels greater than the chemical potential, the occupation probability increases. In other words, at a given temperature, the higher energy levels are, the more likely to be occupied in the quantum case than in the classical case. The difference between Fermionic and Bosonic statistics also manifests itself in important ways. For instance, at low temperatures where there is little thermal disruption of the occupation pattern dictated by the Pauli exclusion principle, Bosons tend to condense in the lowest energy states, while there is just one Fermion per energy state [24]. As a result, thermodynamic quantities such as the total energy or entropy of the system sample the spectrum of Laplacian energy states in different ways, and potentially convey different aspects of network structure. For instance, Bose-Einstein statistics are likely to respond more strongly to the cluster community structure since they are sensitive to the eigenvalue gap [20]. Fermi-Dirac statistics, on the other hand, are sensitive to a larger portion of the spectrum and are more sensitive to the density of energy states [15]. As a result, they are more sensitive to the details of the degree distribution and also to structural artefacts requiring more information concerning the Laplacian spectrum such as the path length and cycle length distributions [34].

The aim of this chapter is to explore the behaviour of the entropy and total energy of networks resulting from different choices of partition functions. We compare four different entropic network characterisations. The first three result from the partition functions for a) Maxwell-Boltzmann, b) Bose-Einstein and c) Fermi-Dirac occupation statistics, while the fourth is the von Neumann entropy associated with the normalised Laplacian matrix of the network [81, 59, 117]. We explore how these different entropies can be used to characterise the changes of network structure with time, and distinguish different types of network models (Erdős-Rényi random graphs, small-world networks [110], and scale-free networks [12]).

The remainder of the chapter is organised as follows. We first provide a review of the relationship between the partition function and the thermodynamic variables, i.e. the average energy, thermodynamic entropy, Helmholtz free energy, temperature and chemical potential. Then we provide a detailed analysis of the entropies resulting from the three

different choices of partition functions and explore their low and high-temperature limits. The numerical experiments on synthetic and real-world datasets are used to evaluate the effectiveness of the different thermodynamic network characterisations. Finally, we conclude the chapter and make suggestions for future work.

3.2 Thermodynamic Representation of Networks

Thermodynamic analogies provide powerful tools for analysing complex networks. The underpinning idea is that statistical thermodynamics can be combined with network theory to characterise both static and time-evolving networks [74].

A complex network can be viewed as a grand canonical ensemble, which not only exchanges energy but also exchanges particles with a heat reservoir. In general, the energy and entropy of the network depend on the assumptions concerning the Hamiltonian for the system and the corresponding partition function.

3.2.1 Preliminaries

Let $G(V,E)$ be an undirected graph with node set V and edge set $E \subseteq V \times V$, and let $|V|$ represent the total number of nodes on graph $G(V,E)$. The $|V| \times |V|$ adjacency matrix A of a graph is defined as

$$A = \begin{cases} 1 & \text{if } (u,v) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Then the degree of node u is $d_u = \sum_{v \in V} A_{uv}$.

The normalised Laplacian matrix $\tilde{L}$ of the graph G is defined as

$$\tilde{L} = D^{-\frac{1}{2}} L D^{\frac{1}{2}} = \Phi \tilde{\Lambda} \Phi^T \tag{3.2}$$

where $L = D - A$ is the Laplacian matrix and D denotes the degree diagonal matrix whose elements are given by $D(u,u) = d_u$ and zeros elsewhere. $\tilde{\Lambda} = diag(\lambda_1, \lambda_2, \ldots \lambda_{|V|})$ is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_{|V|})$ is the

matrix with the ordered eigenvectors as columns. The element-wise expression of $\tilde{L}$ is

$$\tilde{L}_{uv} = \begin{cases} 1 & \text{if } u = v \text{ and } d_u \neq 0 \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \neq v \text{ and } (u,v) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

3.2.2 Hamiltonian Operator

In quantum mechanics, the Hamiltonian operator is the sum of the kinetic energy and potential energy of all the particles in the system. It is the energy operator of the system and the standard formulation on a manifold is

$$\hat{H} = -\nabla^2 + U(r,t) \tag{3.4}$$

In our case, we assume the graph to be in contact with a heat reservoir. The eigenvalues of the Laplacian matrix can be viewed as the energy eigenstates, and these determine the Hamiltonian and hence the relevant Schrödinger equation which governs the particles in the system. The particles occupy the energy states of the Hamiltonian subject to thermal agitation by the heat bath. The number of particles in each energy state is determined by the temperature, the assumed model of occupation statistics and the relevant chemical potential.

If we take the kinetic energy operator $-\nabla^2$ to be the negative of the adjacency matrix, i.e. $-A$, and the potential energy $U(r,t)$ to be the degree matrix D, then the Hamiltonian operator is the Laplacian matrix on the graph. Similarly, the normalised form of the graph Laplacian can be viewed as the Hamiltonian operator

$$\hat{H} = \tilde{L} \tag{3.5}$$

In this case, the energy states of the network $\{\varepsilon_i\}$ are then the eigenvalues of the Hamiltonian $\hat{H}|\psi_i\rangle = \tilde{L}|\psi_i\rangle = E_i|\psi_i\rangle$.

The eigenvalues are all greater than or equal to zero, and the multiplicity of the zero

eigenvalues is the number of connected components in the network. Furthermore, the density matrix commutes with the Hamiltonian, i.e. the associated Poisson bracket is zero,

$$[\hat{H},\rho] = [\tilde{L},\frac{\tilde{L}}{|V|}] = 0 \tag{3.6}$$

which means that the network is in equilibrium when there are no changes in the density matrix which describes the system.

3.2.3 Thermodynamic Quantities

Here we consider the thermodynamic system specified by a system of N particles with energy states given by the network Hamiltonian and immersed in a heat bath with temperature T. The ensemble is represented by a partition function $Z(\beta,N)$, where $\beta = 1/k_BT$ is an inverse of temperature parameter [104].

When specified in this way, the various thermodynamic characterisations of the network can be computed. For instance, the average energy of the network can be expressed in terms of the density matrix and the Hamiltonian operator,

$$U = \left[-\frac{\partial}{\partial\beta}\log Z(\beta,N)\right]_N = \mathrm{Tr}\,(\rho H) = k_BT^2\left[\frac{\partial}{\partial T}\log Z\right]_N \tag{3.7}$$

the thermodynamic entropy by

$$S = k_B\left[\frac{\partial}{\partial T}T\log Z\right]_N \tag{3.8}$$

and the chemical potential by

$$\mu = -k_BT\left[\frac{\partial}{\partial N}\log Z\right]_\beta \tag{3.9}$$

The chemical potential is a measure of how resistive the system is to the addition of new particles. It acts to offset the energy levels of the Hamiltonian. In the case of Fermi-Dirac statistics, the chemical potential is equal to the Fermi level and at zero temperature

determines the highest occupied energy state. In Bose-Einstein statistics, the chemical potential tends to zero at zero temperature, and this leads to the formation of the Bose-Einstein condensate. In the remainder of the paper, we set the Boltzmann constant $k_B = 1$.

Both the energy and the entropy can be regarded as weighted functions of the Laplacian eigenvalues which characterise the network structure in different ways. In the following sections, we will explore these differences in more detail, and in particular to which parts of the Laplacian spectrum they are most sensitive to different choices of the partition function resulting from different occupation statistics.

3.3 Partition Functions and Occupation Statistics

According to the picture adopted in this chapter, the normalised Laplacian of the graph specifies a series of energy states that can be occupied by particles. At a given temperature, there are a number of alternative ways in which the energy levels can be occupied, depending on the spin-statistics of the particles.

Here we consider the different situations that arise when the occupation of the energy levels is governed by Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. The Maxwell-Boltzmann distribution applies when spin statistics are ignored and the population of the different energy levels is governed by thermalisation. Bose-Einstein statistics apply to bosons of integer spin, and which are indistinguishable. Finally, Fermi-Dirac statistics apply when the particles are fermions with half-integer spin and are subject to the Pauli exclusion principle.

For each distribution, we capture the statistical mechanical properties of particles in the system using the partition function associated with the different occupation statistics. The network can then be characterised using thermodynamic quantities computed from the partition function, and these include the entropy, energy, and temperature.

3.3.1 Maxwell-Boltzmann Statistics

In statistical mechanics, the Maxwell-Boltzmann distribution relates the microscopic properties of particles to the macroscopic thermodynamic properties of matter [114]. It

applies to systems consisting of a fixed number of weakly interacting distinguishable particles. These particles occupy the energy levels associated with a Hamiltonian and in our case the Hamiltonian of the network, which is in contact with a thermal bath [80].

Taking the Hamiltonian to be the normalised Laplacian of the network, the canonical partition function for Maxwell-Boltzmann occupation statistics of the energy levels is

$$Z_{MB} = \mathrm{Tr}\left[\exp(-\beta\tilde{L})^N\right] = \left(\sum_{i=1}^{|V|} e^{-\beta\varepsilon_i}\right)^N \tag{3.10}$$

where $\beta = 1/k_BT$ is the reciprocal of the temperature T with k_B as the Boltzmann constant; N is the total number of particles and ε_i denotes the microscopic energy of system at each microstate i. Furthermore, from Eq.(3.7), the average energy of the network is

$$\langle U\rangle_{MB} = -\frac{\partial \log Z}{\partial \beta} = N\frac{\mathrm{Tr}[\tilde{L}\exp(-\beta\tilde{L})]}{\mathrm{Tr}[\exp(-\beta\tilde{L})]} = N\frac{\sum_{i=1}^{|V|}\varepsilon_i e^{-\beta\varepsilon_i}}{\sum_{i=1}^{|V|} e^{-\beta\varepsilon_i}} \tag{3.11}$$

and similarly derived from Eq.(3.8), the entropy of the system with N particles is

$$\begin{aligned} S_{MB} &= \log Z - \beta\frac{\partial \log Z}{\partial \beta} = -N\mathrm{Tr}\left\{\frac{\exp(-\beta\tilde{L})}{\mathrm{Tr}[\exp(-\beta\tilde{L})]}\log\frac{\exp(-\beta\tilde{L})}{\mathrm{Tr}[\exp(-\beta\tilde{L})]}\right\} \\ &= -N\sum_{i=1}^{|V|}\frac{e^{-\beta\varepsilon_i}}{\sum_{i=1}^{|V|}e^{-\beta\varepsilon_i}}\log\frac{e^{-\beta\varepsilon_i}}{\sum_{i=1}^{|V|}e^{-\beta\varepsilon_i}} \end{aligned} \tag{3.12}$$

For a single particle, the density matrix is

$$\rho_{MB} = \frac{\exp(-\beta\tilde{L})}{\mathrm{Tr}[\exp(-\beta\tilde{L})]} \tag{3.13}$$

Thus, the entropy in the Maxwell-Boltzmann system is simply N times the von Neumann entropy of a single particle, as we might expect.

3.3.2 Bose-Einstein Statistics

The Bose-Einstein distribution applies to indistinguishable bosons. Each energy state can accommodate an unlimited number of particles specified by the network Hamiltonian.

Bosonic particles subject to Bose-Einstein statistics do not obey the Pauli exclusion principle and can aggregate in the same energy state. Complex networks have been successfully characterised using systems of bosons to capture network topology. For instance, Bianconi and Barabási [21] have constructed a network model based on a Bose gas, and have studied the phase transitions in network structure associated with the Bose-Einstein condensation of the gas. This model has also been extended to understand processes such as supersymmetry in networks [20].

For a grand-canonical ensemble with a varying number of particles and a chemical potential μ, the Bose-Einstein partition function is

$$Z_{BE} = \det\left(I - e^{\beta\mu}\exp[-\beta\tilde{L}]\right)^{-1} = \prod_{i=1}^{|V|}\left(\frac{1}{1-e^{\beta(\mu-\varepsilon_i)}}\right) \tag{3.14}$$

From Eq.(3.7) and Eq.(3.8), the average energy is

$$\begin{aligned}\langle U\rangle_{BE} &= -\frac{\partial \log Z}{\partial\beta} = -\mathrm{Tr}\left\{\left[I - e^{\beta\mu}\exp(-\beta\tilde{L})\right]^{-1}(\mu I - \tilde{L})e^{\beta\mu}\exp(-\beta\tilde{L})\right\} \\ &= -\sum_{i=1}^{|V|}\frac{(\mu-\varepsilon_i)e^{\beta(\mu-\varepsilon_i)}}{1-e^{\beta(\mu-\varepsilon_i)}}\end{aligned} \tag{3.15}$$

while the corresponding entropy is

$$\begin{aligned}S_{BE} &= \log Z + \beta\langle U\rangle \\ &= -\mathrm{Tr}\left\{\log[I - e^{\beta\mu}\exp(-\beta\tilde{L})]\right\} - \mathrm{Tr}\left\{\beta[I - e^{\beta\mu}\exp(-\beta\tilde{L})]^{-1}(\mu I - \tilde{L})e^{\beta\mu}\exp(-\beta\tilde{L})\right\} \\ &= -\sum_{i=1}^{|V|}\log\left(1 - e^{\beta(\mu-\varepsilon_i)}\right) - \beta\sum_{i=1}^{|V|}\frac{(\mu-\varepsilon_i)e^{\beta(\mu-\varepsilon_i)}}{1-e^{\beta(\mu-\varepsilon_i)}}\end{aligned} \tag{3.16}$$

As a result the average energy is the average difference between the Laplacian energy states and the chemical potential, weighted by the Bose-Einstein factor $\exp[-\beta(\varepsilon_i - \mu)]/(1-\exp[-\beta(\varepsilon_i-\mu)])$. The weighted energy difference therefore decreases with energy. The entropy also decreases with the energy of the states.

3.3.3 Fermi-Dirac Statistics

The Fermi-Dirac distribution applies to indistinguishable fermions with a maximum occupancy of one particle in each energy state. Particles cannot be added to states that are already occupied, and hence obey the Pauli exclusion principle. These particles behave like a set of free fermions in the complex network with energy states given by the network Hamiltonian. The statistical properties of the networks are thus given by the Fermi-Dirac distribution of the equivalent quantum system [86, 24]. The corresponding partition function is

$$Z_{FD} = \det\left(I + e^{\beta\mu}\exp[-\beta\tilde{L}]\right) = \prod_{i=1}^{|V|}\left(1 + e^{\beta(\mu-\varepsilon_i)}\right) \tag{3.17}$$

From Eq.(3.7) the average energy of the Fermi-Dirac system is

$$\begin{aligned}\langle U\rangle_{FD} &= -\frac{\partial \log Z}{\partial \beta} = -\mathrm{Tr}\left\{\left[I + e^{\beta\mu}\exp(-\beta\tilde{L})\right]^{-1}(\mu I - \tilde{L})e^{\beta\mu}\exp(-\beta\tilde{L})\right\} \\ &= -\sum_{i=1}^{|V|}\frac{(\mu-\varepsilon_i)e^{\beta(\mu-\varepsilon_i)}}{1+e^{\beta(\mu-\varepsilon_i)}}\end{aligned} \tag{3.18}$$

And the entropy is

$$\begin{aligned}S_{FD} &= \log Z + \beta\langle U\rangle \\ &= \mathrm{Tr}\left\{\log[I + e^{\beta\mu}\exp(-\beta\tilde{L})]\right\} - \mathrm{Tr}\left\{\beta[I + e^{\beta\mu}\exp(-\beta\tilde{L})]^{-1}(\mu I - \tilde{L})e^{\beta\mu}\exp(-\beta\tilde{L})\right\} \\ &= \sum_{i=1}^{|V|}\log\left(1 + e^{\beta(\mu-\varepsilon_i)}\right) - \beta\sum_{i=1}^{|V|}\frac{(\mu-\varepsilon_i)e^{\beta(\mu-\varepsilon_i)}}{1+e^{\beta(\mu-\varepsilon_i)}}\end{aligned} \tag{3.19}$$

As the result, the average energy is the average difference between the Laplacian energy states and the chemical potential, this time weighted by the Fermi-Dirac factor $\exp[-\beta(\varepsilon_i-\mu)]/(1+\exp[-\beta(\varepsilon_i-\mu)])$. For a given chemical potential, the higher energy levels receive more weight than in the case of the Bose-Einstein statistics. Moreover, the entropy associated with the states peaks at the chemical potential.

3.3.4 Particle Population and Chemical Potential

We need to specify how the system associated with different partition functions is populated at various temperatures, and how we set the chemical potential in the case of Bose-Einstein and Fermi-Dirac statistics. Our approach is to compute the number of particles occupying each energy state, and sum over the different energy states.

In the case of Maxwell-Boltzmann statistics, the number of particles in the state with energy ε is

$$n_i = N\frac{e^{-\beta\varepsilon_i}}{Z_{MB}} = N\frac{\exp(-\beta\tilde{L})}{\mathrm{Tr}[\exp(-\beta\tilde{L})]} \tag{3.20}$$

and so the total number of particles is

$$N = \sum_{i=1}^{|V|} n_i \tag{3.21}$$

In both cases of Bose-Einstein and Fermi-Dirac occupation statistics, the partition function and hence both the average energy and entropy, depending on the chemical potential. This parameter is determined by the number of particles in the system and the temperature.

For Bose-Einstein statistics at the temperature corresponding to β, in order for the number of particles in each energy state to be non-negative, the chemical potential must be less than the minimum energy level, i.e. $\mu < \min\varepsilon_i$. Under Fermi-Dirac statistics, on the other hand, with a single particle per energy state, the chemical potential is hence just the nth energy level, and so $\mu = \varepsilon_n$.

3.3.5 High and Low Temperature Limits

High Temperature Limits $(\beta \to 0)$

At high temperature, i.e. when β approaches zero, thermalisation disrupts the effects of the occupation statistics captured by different partition functions, and both the Bose-Einstein and Fermi-Dirac models are equivalent to the Maxwell-Boltzmann case. For the Maxwell-Boltzmann distribution, the high-temperature limit of the average energy is

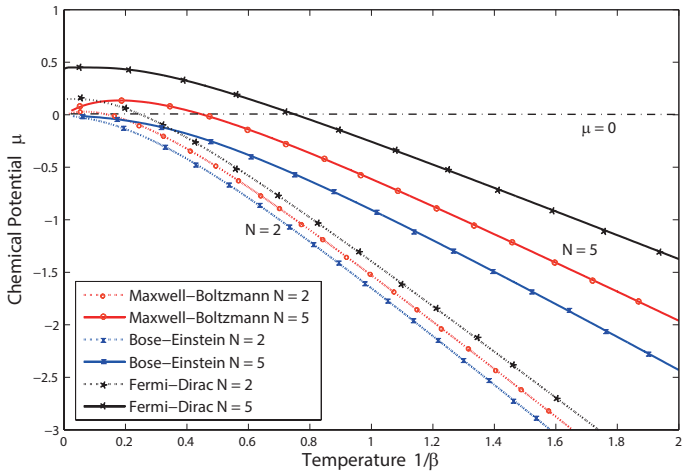


Figure 3.1: Plot of the chemical potential μ versus temperature T for Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. In the high-temperature region, the three chemical potentials exhibit similar behaviour. In the low-temperature region, the chemical potential for Bose-Einstein statistics is always less than 0. However, with Fermi-Dirac statistics, it is larger than 0 and increases with the number of particles N.

$\lim_{\beta\to 0}\langle U\rangle_{MB} = \frac{N}{|V|}\mathrm{Tr}[\tilde{L}]$, which is as expected proportional to the trace of the normalised Laplacian, giving an average energy per particle of $\frac{1}{|V|}Tr[\tilde{L}]$. The corresponding high-temperature limit of the entropy is

$$\lim_{\beta\to 0} S_{MB} = N\log|V| + \frac{N\beta^2}{2|V|}\left\{\frac{1}{|V|}\mathrm{Tr}[\tilde{L}] - \mathrm{Tr}[\tilde{L}^2]\right\} \tag{3.22}$$

This is similar to the result obtained by Han et al. [59] for the von Neumann entropy. As a result, the entropy at high temperature is a constant for all three models of the occupation statistics.

Low Temperature Limits $(\beta \to \infty)$

The low temperature limits of the energy and entropy under Maxwell-Boltzmann statistics satisfy when $\beta \to \infty$ are $\lim_{\beta\to+\infty}\langle U\rangle_{MB} = 0$ and $\lim_{\beta\to+\infty} S_{MB} = N\log c$, where c is the number of connected components in the network. We usually deal with graphs having a single connected component and as a result we have that the limit of entropy in Maxwell-Boltzmann case at the low temperature tends to zero.

In the case of the Bose-Einstein and Fermi-Dirac partition functions, the chemical

potential plays a pivotal role in determining the low-temperature limit.

For Fermi-Dirac statistics, the constant chemical potential μ is equal to the energy of highest state occupied by one of the N particles at zero temperature. With a single particle per energy state, this is hence just the Nth energy level, namely ε_N. As the temperature approaches zero, the chemical potential μ approaches the Fermi energy ε_N, so that $\mu = \varepsilon_N$. There is only one configuration for each identical particle occupies at each energy state, and the corresponding entropy is $\lim_{\beta\to+\infty} S_{FD} = 0$.

For Bose-Einstein occupation statistics, at $T = 0$ all particles are in the ground state and it is straightforward to show that $\lim_{\beta\to+\infty} S_{BE} = (N+1)\ln(N+1) - N\ln N$. As N goes to infinity, the limits of entropy tends to $\lim_{N\to+\infty}\lim_{\beta\to+\infty} S_{BE} = \ln N$. The main difference between the thermal quantities of the classical statistical system and that of the quantum spin systems is that the partition function results in different occupation of the energy levels according to the relevant population statistics.

In the Maxwell-Boltzmann case, without thermalisation of the levels at zero temperature, all particles occupy the zero energy ground state. But in the case of Bose-Einstein and Fermi-Dirac statistics, this pattern is modified by the chemical potential, and this modified the way which the higher energy levels are populated. For Bose-Einstein statistics, the effect is to shift the occupation number from the zero energy Maxwell-Boltzmann ground-state by an amount proportional to the chemical potential. In other words, the particles are found with higher probabilities at lower energy levels. In the case of the Fermi-Dirac statistics, the effect is exaggerated since the chemical potential is the energy of the state corresponding to the number of particles in the system.

3.4 Physical Intuitions

The network Laplacian defines a set of energy levels for a system which is in thermodynamic equilibrium with a heat bath of known temperature. The different partition functions govern how a system of non-interacting particles populate these energy levels at a particular temperature. From the partition functions, we can calculate the energy and

entropy associated with the system of particles, at a particular temperature and for different numbers of particles. Our idea is to use these two thermodynamic quantities to characterise the network from which the Laplacian was computed.

Of course, different networks will have different graph spectra (i.e. distributions of Laplacian eigenvalues or energy levels of our thermodynamic system), and this, in turn, will give rise to a different population of energy levels with temperature. More importantly, in this study, the choice of partition function will also control how the different energy levels can be populated depending on the spin statistics of the particles, and the number of particles added to the system. When we work with the classical Maxwell-Boltzmann distribution, then the temperature is the only controlling parameter. By increasing temperature, we simply thermalise the population of the energy levels. On the other hand, when we envoke non-classical spin statistics, quantum effects become evident. In the case of Fermi-Dirac statistics, only one particle can occupy each energy state. For Bose-Einstein statistics, on the other hand, particles can condense in the lower energy states, particularly at low temperatures, but these particles are indistinguishable, leading to different statistics. In the quantum cases, the effect of changing the number of particles can be modelled by adding a chemical potential which effectively shifts the energy levels.

We use the entropy and energy associated with the distribution of energy levels and their different occupation probabilities to explore whether the different partition functions allow us to probe differences in network structure in different and hopefully more useful ways. The main interest here lies in the low-temperature behaviour since at high temperature the effects of the quantum statistics are disturbed by thermal effects all three partition functions give identical results. At low temperature, we are more likely to find bosonic particles in the low energy states when compared to the Maxwell-Boltzmann distribution. On the other hand, because of the Pauli exclusion principle, we are more likely to find fermions at higher energies. Hence by populating the energy states in different ways, the particles respond to the Laplacian spectrum in different ways depending on which of the three partition functions governs their behaviour. The question we seek to answer is when measured in terms of their entropy or energy to the different partition functions allow us

to probe network structure in different ways.

It is well known that different types of network have different degree distributions, and this is reflected in their Laplacian spectra. For instance, Erdős-Rényi random graphs the eigenvalues follow a semi-circular (or Wigner) distribution, with mean controlled by the connection probability. Scale-free networks have a triangular distribution and networks of the Watts-Strogatz type have a more complex spectrum which depends on the parameters and may contain sharp peaks. For cluster-structure, the distribution of the lowest eigenvalues and the spectral gap are most important. Hence, the choice of how the eigenvalues are sampled, or choice of the partition function, can be sensitive to the type of structure. One might, for instance, expect Bose-Einstein statistics to be better suited to detecting networks with strong community structure because they preferentially sample the lower energy levels. Fermi-Dirac statistics, on the other hand, may be better for distinguishing different network models because they probe a wider range of energy levels, and are hence more sensitive to the mean and variance of the eigenvalue distribution.

3.5 Experiments and Evaluations

We explore whether the thermodynamic characterisations resulting from the three alternative models for the energy level occupation statistics can be employed as a useful tool for better understanding the structural properties and the evolution of networks. Specifically, we numerically simulate the effects of the three different models and examine whether the resulting entropies can distinguish different structures, and compare their relative performance. Furthermore, we compute the thermodynamic characterisations for a number of real-world time-evolving networks in order to investigate whether they can be used to detect abrupt changes in network structure at different time epochs. Finally, we use the different entropies to classify tumour mutation networks and protein to protein interaction networks resulting from different groups. To simplify the calculation, we set the Boltzmann constant to unity throughout our experiments.

This section is structured as follows. We commence by describing the data-sets used

in our experiments and the kernel principal components method used to visualise the entropy differences between different networks. We then explore how the different entropies depend on their free parameters, name temperature of a number of particles for network drawn from different models (random graph, small world and scale-free networks). Using kernel PCA we visualise how the networks from different models distribute themselves in three dimensions with the different partition functions, and comment on which gives the best separation. Finally, we report results on real-world data-sets.

3.5.1 Data Sets

Here, we use four different datasets. The first contains synthetically generated artificial networks, while the remaining three are extracted from real-world complex systems.

Synthetic Networks Data-set: Contains a large number of graphs which are randomly generated according to one of three different complex network models, namely, a) the classical Erdős-Rényi random graph model, b) the small-world model introduced by Watts and Strogatz [110], and c) the scale-free model, developed by Barabási-Albert model [12, 13]. These are created using a variety of model parameters, e.g., the graph size and the connection probability (randomly generated between 0.1 to 0.9) in the random graph model, the link rewiring probability (randomly generated between 0.2 to 0.8) in the small-world model [110] and the number of added connections (set to 1) at each time step [12] in the scale-free model. The networks are randomly generated by a normal distribution with the number of node between 100 and 1,000.

NYSE Stock Market Networks Data-set: The New York Stock Exchange dataset consists of the daily prices of 3,799 stocks traded continuously on the New York Stock Exchange over 6000 trading days. The stock prices were obtained from the Yahoo! financial database (http://finance.yahoo.com) [93]. A total of 347 stock were selected from this set, for which historical stock prices from January 1986 to February 2011 are available. In our network representation, the nodes correspond to stock and the edges indicate that there is a statistical similarity between the time series associated with the stock closing prices [93]. To determine the edge structure of the network, we use a time window of 20 days is to compute the cross-correlation coefficients between the time-series for each

pair of stock. Connections are created between a pair of stock if the cross-correlation exceeds an empirically determined threshold. In our experiments, we set the correlation coefficient threshold to the value to $\xi = 0.85$. This yields a time-varying stock market network with a fixed number of 347 nodes and varying edge structure for each of 6,000 trading days. The edges of the network, therefore, represent how the closing prices of the stock follow each other.

Tumour Mutation Networks Data-set: Contains tumour mutation data for three major cancers taken from the Cancer Genome Atlas (TCGA). These are a) ovarian cancer b) uterine cancer and c) lung adenocarcinoma [35]. There are 356 patients with mutations in 9,850 genes in the ovarian cancer cohort, 248 patients with mutations in 17,968 genes in the uterine endometrial cancer cohort and 381 patients with mutations in 15,967 genes in the lung adenocarcinoma cohort [62]. The raw patient mutation data are binary vectors, with elements corresponding to different genes. The binary numbers indicate if the relevant gene is mutated or not (1 indicates the presence of a mutation, 0 that a mutation is absent). So each individual is characterised by a 0-1 binary gene sequence of mutation indicators. Patient mutation networks were mapped onto gene interaction networks by aggregating information from several pathways and interaction databases, describing physical protein-protein interactions (PPIs) and functional relationships between genes in both regulatory, signalling and metabolic pathways [48].

Protein-Protein Interaction Networks Data-set: The PPIs dataset extracted from STRING-8.2 [99] consisting of networks which describe the interaction relationships between *histidine kinase* and other proteins. Histidine kinase is a key protein in the development of signal transduction. If two proteins have direct (physical) or indirect (functional) association, they are connected by an edge. There are 173 PPIs in this dataset and they are collected from 4 different kinds of bacteria with the following evolution order (from older to more recent). Aquifex and Thermotoga-8 PPIs from Aquifex aelicus and Thermotoga maritima, Gram-Positive-52 PPIs from Staphylococcus aureus, Cyanobacteria-73 PPIs from Anabaena variabilis and Proteobacteria-40 PPIs from Acidovorax avenae [44].

3.5.2 Visualising the Distribution of Networks using Jensen-Shannon Divergence

We require a tool for visualising the similarity of sets of graphs measured by the entropies computed from the different partition functions. To this end, we measure similarity using the Jensen-Shannon divergence [73], which is asymmetric information theoretic divergence measure computed from the entropies of pairs of graphs. We characterise the similarities of a set of graphs using a kernel matrix and then embed the graphs into a vector space using kernel-embedding for the purposes of visualisation.

Here we deal with the case where the nodes in the graphs are labelled, and at each time step, the node-sets are identical. Only the edge-set varies between time-steps. Moreover, since the nodes are labelled it is straightforward to determine which edges have been added, removed or remained unchanged between different time steps. Suppose that G_i and G_j are two graphs, and that $G_i \oplus G_j$ is the union graph with the set of edges formed from those edges that are present at either time step i or time step j. With the union graph to hand, the Jensen-Shannon divergence for the pair of graphs G_i and G_j is

$$D_{JS}(G_i, G_j) = S(G_i \oplus G_j) - \frac{S(G_i) + S(G_j)}{2} \tag{3.23}$$

where $S(G_i)$ is the entropy associated with the graph G_i, and $S(G_i \oplus G_j)$ is the entropy associated with the corresponding union graph G_U. Then the Jensen-Shannon kernel[10] is given by

$$k_{JS}(G_i, G_j) = \log 2 - D_{JS}(G_i, G_j) \tag{3.24}$$

With the graph kernel to hand, we embed the graphs into a vector space. To compute the embedding, we commence by computing the eigendecomposition of the kernel matrix, which will reproduce the Hilbert space with a non-linear mapping. In such a case, graph features can be mapped to low dimensional feature space with linear separation. The graph kernel decomposition is

$$k_{JS} = \Phi \Lambda \Phi^T \tag{3.25}$$

where Λ is the diagonal eigenvalue matrix and Φ is the matrix with eigenvectors as columns. To recover the matrix X with embedding coordinate vectors as columns, we write the kernel matrix in Gram-form, where each element is an inner product of embedding coordinate vectors

$$k_{JS} = XX^T \tag{3.26}$$

and as a result $X = \sqrt{\Lambda}\Phi^T$. In practice, we embed the samples of graphs into a three-dimensional space and hence use just the three leading eigenvalues and corresponding eigenvectors of k_{JS} to compute the embedding.

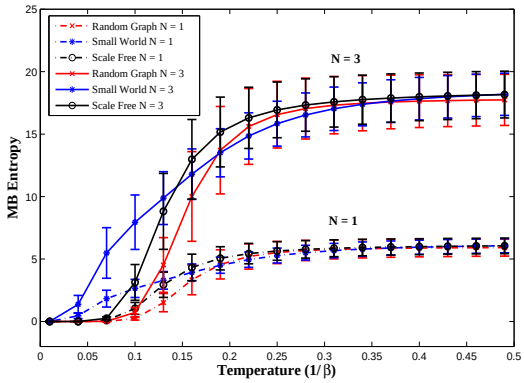
3.5.3 Parameter Dependence

In this section, we investigate how well the different models of the energy level occupation statistics can be used to distinguish synthetic networks generated using the Erdős-Rényi random graphs, Watts-Strogatz small-world models [110] and Barabási-Albert scale-free network models [12, 13]. We conduct numerical experiments to evaluate whether the thermodynamic variables, especially entropy, can represent differences in the structure and topology of networks.

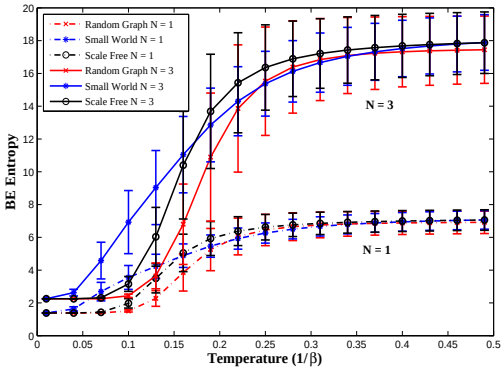
Fig.3.2(a) shows the behaviour of the entropies resulting from Maxwell-Boltzmann occupation statistics as a function of temperature $(1/\beta)$. We explore the effect of varying the number of particles occupying the system and explore the cases where $N = 1$ and $N = 3$. From Eq.(3.12), it is clear that the effect of varying N is simply to scale the entropy by a multiplicative factor.

For the three different graph models (Erdős-Rényi random graph model, Watts-Strogatz small-world model and Barabási-Albert scale-free model), there is different behaviour with temperature. For small-world networks, the entropy increases fastest at low values of temperature. But it is quickly overtaken by the scale-free networks at intermediate temperatures. The Erdős-Rényi random graph model shows the slowest rate of increasements. The common feature is that all three entropies increase monotonically with temperature. However, the detailed dependence on $1/\beta$ depends on the partition function and the underlying occupation statistics. Specifically, in the low-temperature region

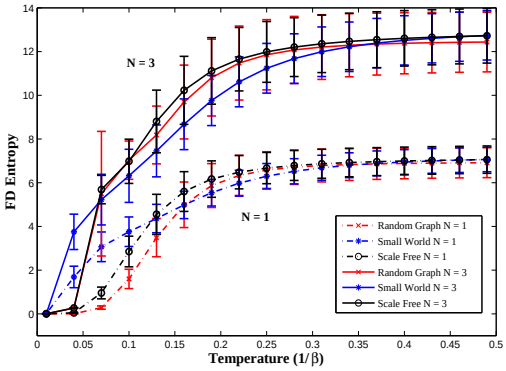
($0.07 \sim 0.12$), the entropy distinguishes strongly among the different types of network models.



(a) Maxwell-Boltzmann Statistics



(b) Bose-Einstein Statistics



(c) Fermi-Dirac Statistics

Figure 3.2: Mean and standard deviations of the entropies for three different network models versus temperature. Number of particles N = 1 and N = 3. Red cross line: Erdős-Rényi random graphs; blue star line: Watts-Strogatz small world networks; black circle line: Barabási-Albert scale free networks.

Fig.3.2(b) and Fig.3.2(c) respectively show similar plots for the entropies derived from the Bose-Einstein and Fermi-Dirac partition functions. In the case of the Bose-Einstein entropy, the curves for the three different graph-modes exhibit the same pattern as in the Maxwell-Boltzmann case. As a result, at low temperatures, the ordering of the Bose-Einstein entropy can be used to separate the different network models. In both the Bose-Einstein and Fermi-Dirac, the number of particles N affects the entropy via the chemical potential μ. Hence, the entropy is not simply scaled by changing N. In the case of the Fermi-Dirac partition function, the pattern of entropies for the different modes is more complex for the various network models. Firstly, for different values of N, the behaviour is very different with temperature. For $N = 1$, we see a similar pattern

to the Maxwell-Boltzmann and Bose-Einstein cases, but with $N = 3$ the behaviour is different with the scale-free and random graphs having small separation for all values of temperature. Additionally, the small-world model is overtaken by the random graphs and scale-free models at a lower value of temperature. This is a consequence of the exclusion principle manifesting itself at low temperature, and hence modifying the distribution of entropy for the different models.

Comparing the plots for the Bose-Einstein and Fermi-Dirac entropies, the following features should also be noted: a) in each case for the different models approach the same limiting value for a given value of N, b) in the case of the Fermi-Dirac partition function all networks have zero entropy as temperature approaches zero, c) in the case of the Bose-Einstein model the entropy approaches the finite value $\ln N$ at zero temperature determined by the number of particles in the system, d) the Fermi-Dirac entropy increases more rapidly with increasing temperature than the Bose-Einstein entropy. On the other hand, as the temperature increases, the occupation probability for the higher energy states increases and particles begin to occupy the higher energy states. Moreover, the occupation probabilities for the three different partition functions become identical.

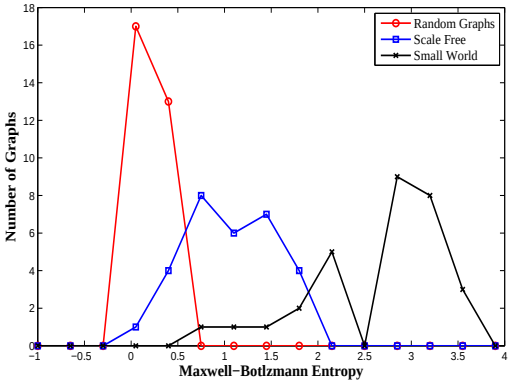
As expected, the differences between the different models are most evident at low temperature. These observations also fit with the intuitions outlined in Section 3.4. The faster rise of the Fermi-Dirac entropy with temperature is a consequence of the greater probability of finding fermions in the higher energy levels. For Bose-Einstein entropy, the greater separation between the different network models as low temperature is a consequence of the different shape of their degree distributions.

3.5.4 Distinguishing Different Network Models

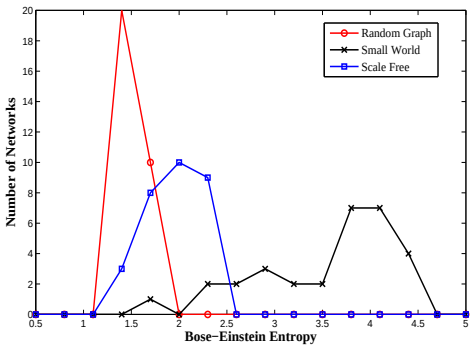
We now explore the ability of the different entropies, resulting from the three different partition functions (Maxwell-Boltzmann, Bose-Einstein, and Fermi-Dirac) to distinguish the three types of complex networks (random graphs, small-world networks and scale-free networks).

Fig.3.3 shows histograms of the entropy for data generated from the three network models in the synthetic dataset. Each figure shows the entropy computed using a different

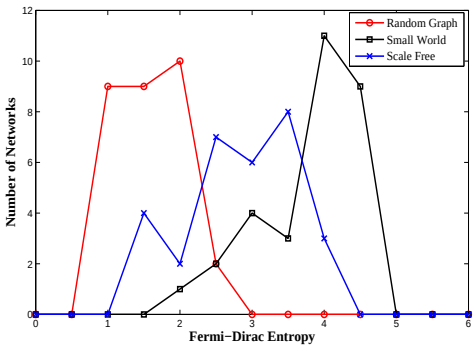
partition function. The differently coloured curves in the histograms correspond to the distribution from three network models. In each plot, the Erdős-Rényi random graphs occupy the low entropy region while the small-world networks stay at the high entropy-area. The distributions of random graphs and scale-free networks are closer in Maxwell-Boltzmann and Bose-Einstein cases when compared to the small-world networks. However, using entropy simply as a unary feature is insufficient to obtain good separation between the different network models (Erdős-Rényi random graphs, Watts-Strogatz small-world networks and Barabási-Albert scale-free networks). Better separation can though be obtained



(a) Maxwell-Boltzmann Statistics



(b) Bose-Einstein Statistics



(c) Fermi-Dirac Statistics

Figure 3.3: Histograms of entropy for Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. The networks are randomly generated with the number of nodes generated from a normal distribution with the number of nodes between 100 and 1,000. The red line represents Erdős-Rényi random graphs; the black line small-world networks and the blue line scale-free networks. Temperature $\beta = 10$ and the number of particles $N = 1$.

if we analyse the pattern of entropy differences between pairs of graphs. Fig.3.4 shows the results of applying the kernel embedding technique outlined in Section 3.5.2 to the entropies computed from the three different partition functions. The differently coloured

points correspond to the data generated from the three different network models (red - Erdős-Rényi random graphs, blue - small world networks, back - scale-free networks). In the case of Maxwell-Boltzmann and Bose-Einstein, the different models from non-overlapping subspaces and can be easily separated. In the case of Bose-Einstein statistics, the effect of changing the number of particles is negligible. In the case of Fermi-Dirac statistics, on the other hand, although more scattered when the number of particles is low, they form tightly clustered subspaces when a larger number of particles are used. This is in line with our physical intuition, since if the number of particles is increased then so the number of energy levels populated increases, even at low temperature. This is in contrast to the Bose-Einstein case, where particles congregate at low energy levels.

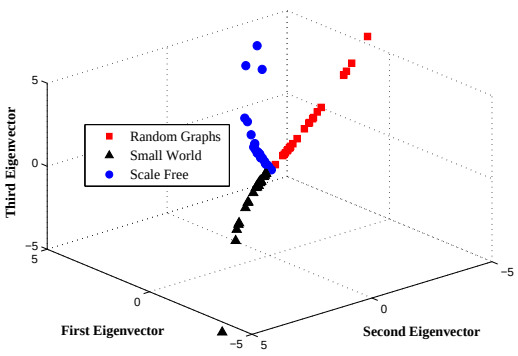
The results above are obtained, using entropies derived from Maxwell-Boltzmann and Bose-Einstein partition functions, the Jensen-Shannon divergence with kernel embedding provides a better visualisation of the separation of the different numerical network models.

3.5.5 Real World Data

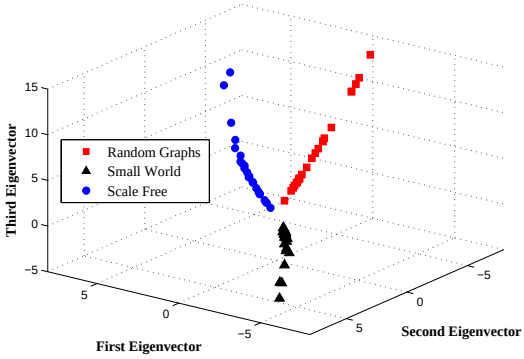
We now compute the entropy characterisations obtained from the three different partition functions on real-world data. Specifically, we explore whether the entropy can be used as an effective tool for better understanding the evolution of real-world complex networks. First, we focus on the detail of New York Stock Exchange in stock market dataset and then provide analysis of the tumour mutation networks and protein-protein interaction networks.

Stock Market Data

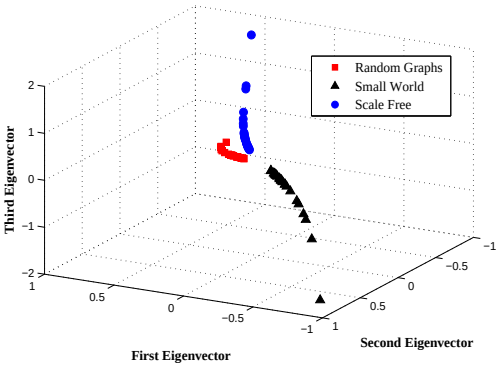
Fig.3.5 and Fig.3.6 show the entropy time-series for the NYSE data obtained from different partition functions. In Fig.3.5, the entropy is derived from the Maxwell-Boltzmann partition function. It is annotated to show the positions of significant financial events such as Black Monday, Friday the 13th mini-crash, Early 1990s Recession, 1997 Asian Crisis, 9.11 Attacks, Downturn of 2002-2003, 2007 Financial Crisis, the Bankruptcy of Lehman Brothers and the European Debt Crisis. In each case, the entropy undergoes significant



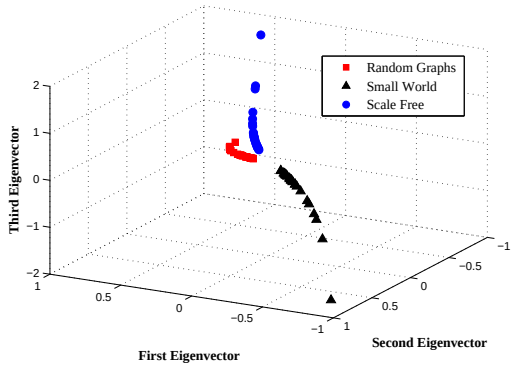
(a) Maxwell-Boltzmann Statistics $N = 5$



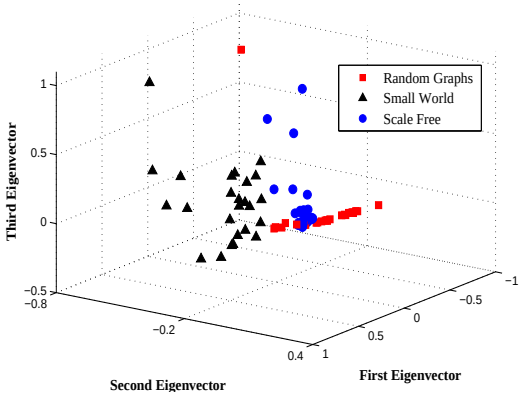
(b) Maxwell-Boltzmann Statistics $N = 10$



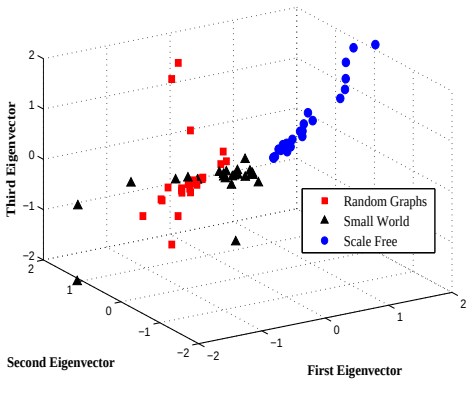
(c) Bose-Einstein Statistics $N = 5$



(d) Bose-Einstein Statistics $N = 10$



(e) Fermi-Dirac Statistics $N = 5$



(f) Fermi-Dirac Statistics $N = 10$

Figure 3.4: Kernel embedding from Jensen-Shannon divergence computed with Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac entropies. We compare the effect of different numbers of particles ($N = 5$ and $N = 10$) with fixed temperature $\beta = 10$.

fluctuations during the financial crises, associated with dramatic structural changes.

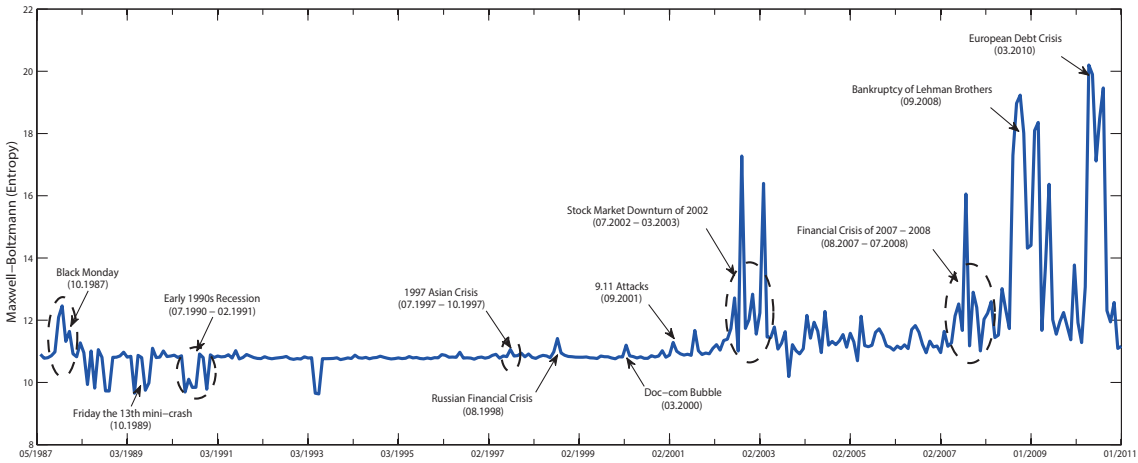


Figure 3.5: Entropy from Maxwell-Boltzmann occupation statistics for NYSE (1987-2011). Critical financial events, i.e., Black Monday, Friday the 13th mini-crash, Early 1990s Recession, 1997 Asian Crisis, 9.11 Attacks, Downturn of 2002-2003, 2007 Financial Crisis, the Bankruptcy of Lehman Brothers and the European Debt Crisis, all appear as distinct events. Particle number $N = 5$ and temperature $\beta = 7$.

A good example is the downturn of 2002-2003. After the 9.11 attacks, investors became unsure about the prospect of terrorism affecting the United States economy. Following the subsequent collapse of many internet companies, numerous large corporations were forced to restate earnings and investor confidence suffered. This considerably altered the inter-relationships among stocks and resulted in significant variance in the structure of the entire market.

Fig.3.6 compares the entropy derived from the three different partition functions with the von Neumann entropy. In the figure, entropies coming from the three partition functions perform better in evaluating the structural changes in the network time-series when compared to the von Neumann entropy. Further exploration shows that entropies, derived from Bose-Einstein and Fermi-Dirac partition functions, exhibit the similar behaviour in the evolution of stock markets. Compared to the Maxwell-Boltzmann case, the Bose-Einstein and Fermi-Dirac entropies are more sensitive to the critical events in the financial data, such as Black Monday in 1987 and the Asian Financial Crisis in 1997.

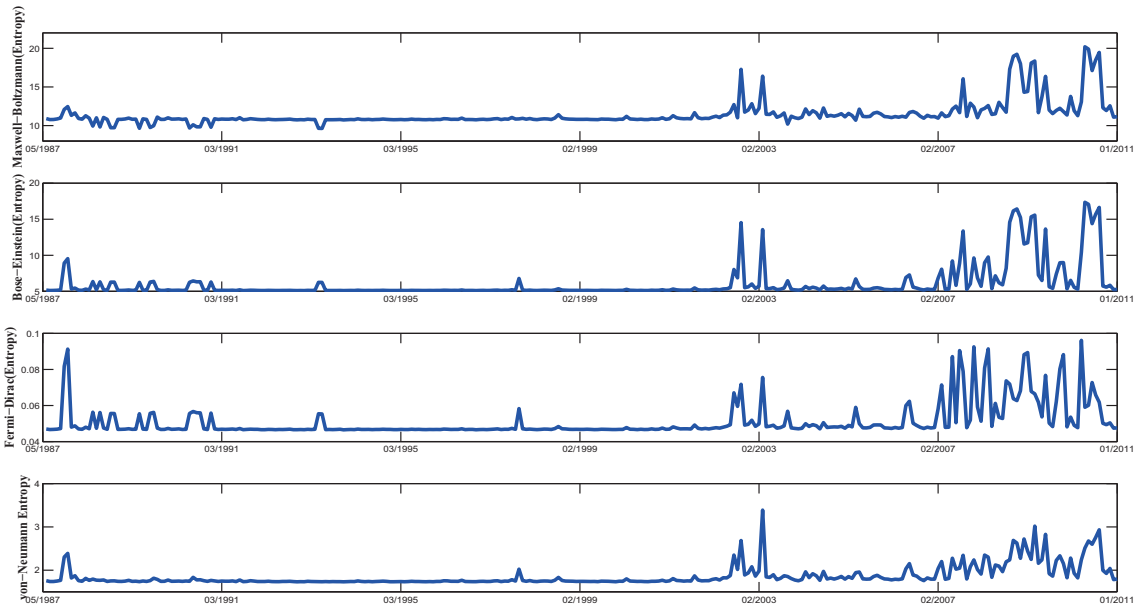


Figure 3.6: von Neumann Entropy and thermodynamic entropy compared for NYSE (1987-2011): (a) Maxwell-Boltzmann occupation statistics, (b) Bose-Einstein occupation statistics and (c) Fermi-Dirac occupation statistics. (d) von Neumann entropy.

Tumour Mutation Networks

Next, we turn our attention to the tumour mutation networks for the three different cancers, i.e. a) ovarian cancer, b) uterine cancer and c) lung adenocarcinoma.

In Fig.3.7(a), we provide the histogram of the entropy computed from the Maxwell-Boltzmann partition function. The different colour of curves represent the three types of cancers. The most striking feature of this plot is that the three kinds of tumour networks dominate different entropy intervals. By applying two separate thresholds to the entropy histogram, we can assign the patients to three classes. We have searched for the two thresholds which give the maximum pooled classification accuracy over the three cancer classes. We find that the best result is given when the uterine and ovarian classes are separated using an entropy threshold at $S_{MB} = 2.92$, and the ovarian and lung adenocarcinoma with a threshold at $S_{MB} = 4.38$. The resulting classification accuracies are 33.87% for uterine cancer, 83.71% for ovarian cancer and 78.48% for lung adenocarcinoma.

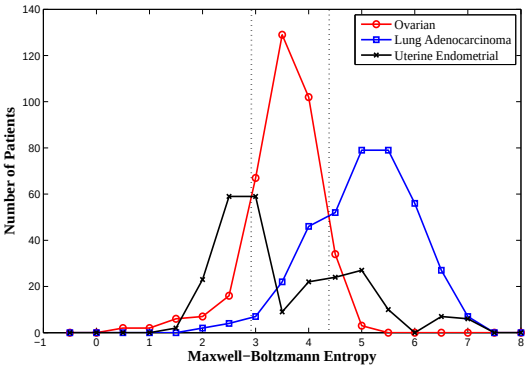
Fig.3.7(b) repeats the analysis using the entropy derived from the Bose-Einstein partition function. Here the corresponding thresholds are 2.49 and 4.52, giving correct classification rates of 75.00%, 93.54%, and 83.96% for the uterine, ovarian, and lung adenocarcinoma classes respectively. For the case of the Fermi-Dirac entropy, as shown in

Table 3.1: Classification accuracy of three different partition functions. The thermodynamic entropy thresholds for Maxwell-Boltzmann statistics are 2.92 and 4.38. The values of entropy separation for Bose-Einstein statistics are 2.49 and 4.52. And the corresponding thresholds of entropy for Fermi-Dirac statistics are 0.56 and 2.08.

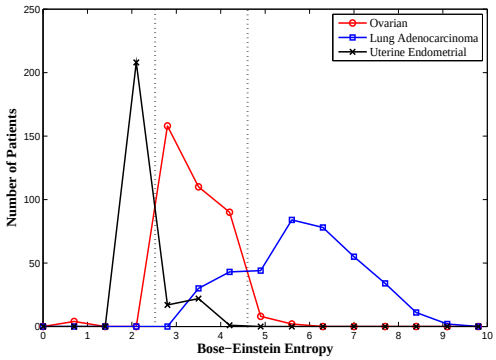
Accuracy	Uterine Cancer	Ovarian Cancer	Lung Adenocarcinoma	Total
Maxwell-Boltzmann	33.87% (84/248)	83.71% (312/356)	78.48% (300/381)	70.66% (696/985)
Bose-Einstein	75.00% (186/248)	93.54% (333/356)	80.84% (308/381)	83.96% (827/985)
Fermi-Dirac	63.71% (153/248)	74.16% (264/356)	79.53% (303/381)	73.10% (720/985)

Fig.3.7(c), the thresholds are $S_{FD} = 0.56$ and $S_{FD} = 2.08$ giving classification accuracies of 63.71%, 74.16%, 73.10% for the uterine, ovarian and lung adenocarcinoma groups.

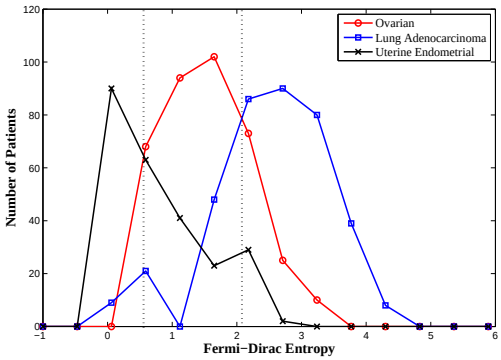
To improve the separation of the data, we use the kernel embedding based on the Jensen-Shannon divergence to measure network similarity, as outlined in Section 3.5.2. The results of the tumour networks, embedding into the three-dimensional space spanned by the first three leading eigenvectors of the kernel matrix, are shown in Fig.3.8. The plot



(a) Maxwell-Boltzmann Statistics



(b) Bose-Einstein Statistics



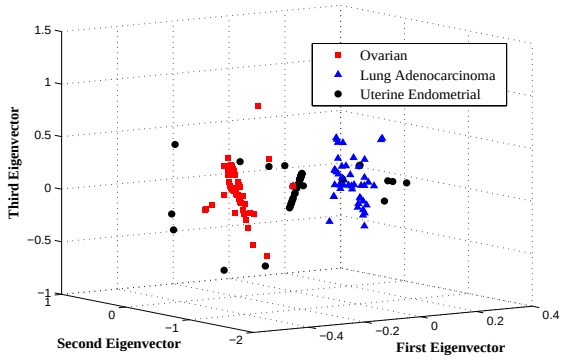
(c) Fermi-Dirac Statistics

Figure 3.7: Histograms of entropy from three statistics for tumour mutation networks (ovarian, uterine and lung adenocarcinoma). Particle number $N = 2$, temperature $\beta = 10$.

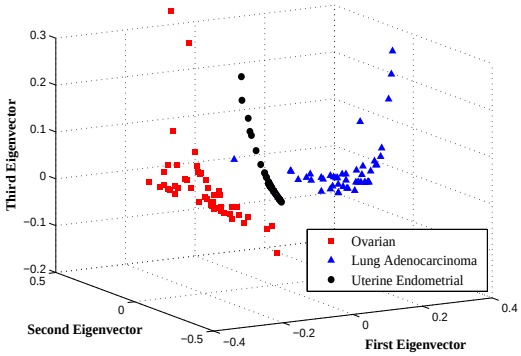
sheds light on the three different classes of data (shown in different colours) exhibiting a

compact manifold structure for three statistics. For each entropy, the different groups of tumour mutation networks are well separated in the embedding space, and this is especially so in the case of the Bose-Einstein entropy.

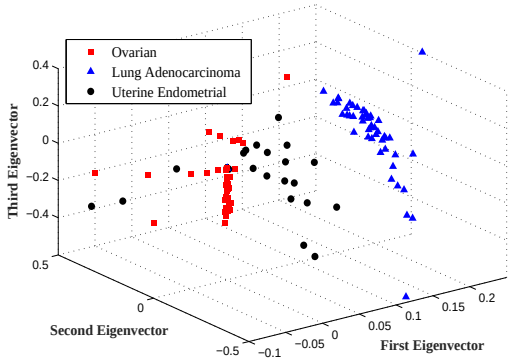
In the Maxwell-Boltzmann and Fermi-Dirac cases, although, the groups of lung adenocarcinoma and ovarian cancer are well separated, the outliers of uterine tumour networks are interspersed among remaining two classes. The best results are obtained in the Bose-Einstein case where the individual networks of the uterine group form the most compact cluster.



(a) Maxwell-Boltzmann Statistics



(b) Bose-Einstein Statistics



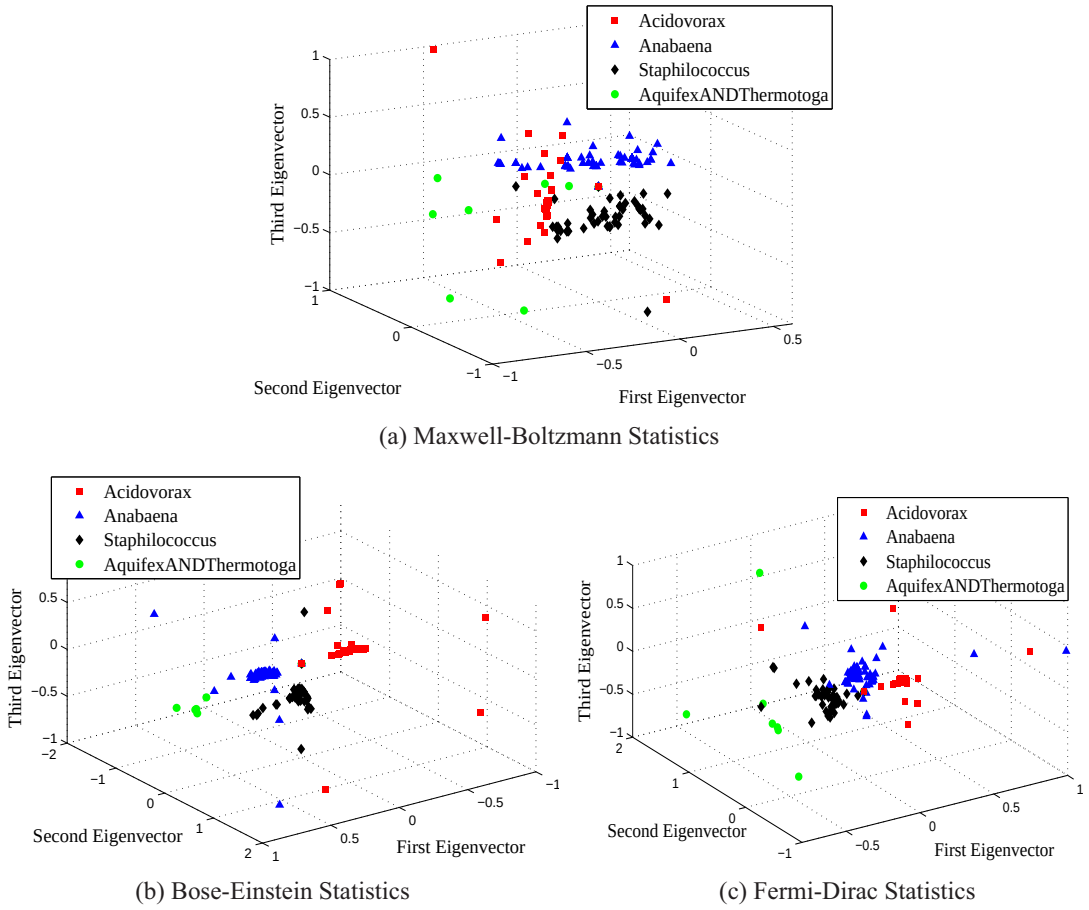
(c) Fermi-Dirac Statistics

Figure 3.8: Kernel embedding with the Jensen-Shannon divergence computed from tumour mutation network entropies (ovarian, uterine and lung adenocarcinoma) for different partition functions. Particle number $N = 3$, temperature $\beta = 10$.

Protein-Protein Interaction Networks

Our final example is based on Protein-protein interaction networks. We perform kernel embedding on the protein-protein interaction networks to visualise their distribution and

to provide a comparison between the entropic discrimination obtained with different partition functions. To this end, we show the distribution of the PPI's in the space spanned by the leading three kernel principal components in Fig.3.9.



(a) Maxwell-Boltzmann Statistics

(b) Bose-Einstein Statistics

(c) Fermi-Dirac Statistics

Figure 3.9: Kernel embedding with the Jensen-Shannon divergence computed from PPI network entropies (Acidovorax, Anabaena, Staphilococcus and Aquifex & Thermotoga) for different partition functions. Particle number $N = 5$, temperature $\beta = 10$.

In each case, the embedded data exhibits a manifold structure which results in good separation of the different classes of PPI. Moreover, the Bose-Einstein entropy provides a better separation of more tightly formed clusters and fewer outliers. The reason for this is that Bose-Einstein statistics encourage particles to aggregate in the lower energy states at low temperature. This amplifies the influence of the number of connected components and the spectral gap in determining the entropy. The former is reflected by the multiplicity of the zero eigenvalues and the latter relates to the degree of bi-partivity in the network. The particle occupation of the low energy states produces a stronger entropic separation

in the Bose-Einstein case. By contrast, neither the Maxwell-Boltzmann nor Fermi-Dirac statistics strongly reflect the lower part of Laplacian spectrum, since they do not give a similar particle concentration in the lower energy states. As a result, Bose-Einstein statistics are more sensitive to the cluster structure of networks, and in the case of PPI's where there is a strong inhomogeneity of node degree which leads to better separation of different classes.

Conclusions from the real world data study

In the case of the tumour mutation networks, overall, the best-pooled performance comes from the Bose-Einstein entropy. Compared with the Maxwell-Boltzmann case, the entropies derived from spin statistical partition functions appear to be more sensitive to differences in network structure and more accurately reflect the structural differences between distinct types of tumour mutation networks. The same pattern emerges with protein-protein interactions networks, This is not surprising since the PPI's have a strong cluster (community) structure. This again fits with the intuitions given in Section 3.4.

3.6 Summary

Our study uses the normalised Laplacian matrix as the Hamiltonian operator of the network, and the associated energy states are given by the eigenvalues of the normalised Laplacian. We explore the case where the particle occupations correspond to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. From the relevant partition functions, we can compute the thermodynamic entropy and energy. Motivated by an interest in revealing the nontrivial properties of the network structure, we have compared the three resulting entropic characterisations and with the von Neumann entropy. We provide a detailed analysis of the three different partition functions, expressed both in terms of the normalised Laplacian matrix and its eigenvalues.

We evaluate the network models resulting from the three different partition functions on both synthetic and real-world datasets. This study investigates how the different entropies can be used to characterise the changes in network structure, and distinguish different types of network structure. Studies with synthetic data show that the entropies can

distinguish Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks. Experiments with real-world data, on the other hand, show that the thermodynamic variables can not only be used to detect both abrupt changes in network structure, but also distinguish different classes of networks.

The main conclusion from this study is that for distinguishing different network models, the Fermi-Dirac entropy appears best. The reason for this is that it is most sensitive to the higher eigenvalues of the normalised Laplacian and this allows it to better probe differences in the degree distributions for different models. Our real-world data, on the other hand, comes mainly from problems where there is a strong community or cluster structure. Here the Bose-Einstein model performs best, and the reason for this is that it is most sensitive to the eigenvalue gap.

Chapter 4

Edge Entropy Decomposition

In prior chapter, we have shown how to compute global network entropy using a heat bath analogy and partition functions with three statistics. In this chapter, we show how to project out edge-entropy components so that the detailed distribution of entropy across the edges of a network can be computed. This is particularly useful if the analysis of non-homogeneous networks with a strong community as hub structure is being attempted. To commence, we view the normalised Laplacian matrix as the network Hamiltonian operator which specifies a set of energy states with the Laplacian eigenvalues. The network is assumed to be in thermodynamic equilibrium with a heat bath. According to this heat bath analogy, particles can populate the energy levels according to the classical Maxwell-Boltzmann distribution and the quantum spin statistics. These distributions together with the energy states determine thermodynamic variables of the network, such as entropy. We analyse the partition functions relevant to Bose-Einstein and Fermi-Dirac statistics in terms of temperature. At high temperatures, the effects of quantum spin statistics are disrupted by thermalisation and correspond to the classical Maxwell-Boltzmann case. However, at low temperatures, the Bose-Einstein system condenses into a state where the particles occupy the lowest energy state, while in the Fermi-Dirac system there is only one particle per energy state. These two models produce quite different entropic characterisations, which are appropriate to different types of network structure. We show how the entropy can be decomposed into components arising from individual edges using the eigenvectors of the normalised Laplacian. Compared to previous work based on the

von Neumann entropy, this thermodynamic analysis is more effective in characterising changes of network structure since it better represents the edge entropy variance associated with edges connecting nodes of large degree. Numerical experiments on real-world datasets are presented to evaluate the qualitative and quantitative differences in performance.

4.1 Introduction

There has been a considerable recent interest in computing the entropy associated with different types of network structure [59, 117, 104]. Network entropy has been extensively used to characterise the salient features of the structure in static and dynamic network systems arising in biology, physics, and the social sciences [81, 59, 117]. For example, the von Neumann entropy can be used as an effective characterisation of network structure, commencing from a quantum analogy in which the Laplacian matrix on graphs [81] plays the role of the density matrix. Further development of this idea has shown the link between the von Neumann entropy and the degree statistics of pairs of nodes forming edges in a network [59], which can be efficiently computed for both directed and undirected graphs [59, 117]. Since the eigenvalues of the density matrix reflect the energy states of a network, this approach is closely related to the heat bath analogy in statistical mechanics. This provides a convenient route to network characterisation [117, 104]. By populating the energy states with particles which are in thermal equilibrium with a heat bath, this thermalisation, of the occupation statistics for the energy states can be computed using the different distribution [114, 104]. The properties of this physical heat bath system are described by a partition function with the energy micro-states of the network represented by a suitably chosen Hamiltonian [114].

The Hamiltonian specify the energy states which are populated by particles in thermal equilibrium with a heat bath. A key element in this thermalisation approach is to model how the energy states are occupied at a particular temperature. Normally this is assumed to follow the classical Maxwell-Boltzmann distribution, where the particles are

distinguishable and weakly interacting. But in the quantum mechanic domain, these particles obey spin statistics. In other words, they are indistinguishable and are either fermions (half-integer spin) or bosons (integer spin). Particles with integer spin are subject to Bose-Einstein statistics and do not obey the Pauli exclusion principle. They can aggregate in the same energy state. At low temperature, this leads to the phenomenon of Bose-Einstein condensation. There has been work aimed at extending this model to networks. For instance, by mapping the network model to a Bose gas, phase transitions have been studied in network evolution associated with Bose-Einstein condensation [21]. This model has also been extended to understand processes such as supersymmetry in networks [47]. In the meanwhile, particles with half-integer spin are subject to Fermi-Dirac statistics and obey the Pauli exclusion principle. They give rise to models of network structures constrained by the occupancy of the nodes and edges. Examples include traffic flow and also the modelling of certain types of geometric networks such as the Cayley tree [15].

Although entropic analysis of the heat bath analogy provides a useful global characterisation of network structure, it does not allow the entropy of edge or subnetwork structure to be easily computed. There has been a little systematic study of the resulting thermodynamic characterisations of network entropy on edges. In this chapter, we explore a novel edge entropy projection which can be applied to the global network entropy computed from Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. We characterise the thermalised system of energy states using partition functions relevant to three occupation statistics. From the partition functions, we compute the entropy of the network with particles. Because Bose-Einstein particles coalescence in low energy states and Fermi-Dirac particles have a greater tendency to occupy high energy states for the Puli exclusion principle, these types of spin statistics lead to very different distributions of entropy for a network with a given structure (i.e. a set of normalised Laplacian eigenvalues). Moreover, at low temperature, the distributions are also different from the classical Maxwell-Boltzmann case. It is these low-temperature differences in energy and entropy that we wish to investigate as a means of characterising differences in the network structure. We use this technique to analyse the distribution of edge entropy within

a network and explore how this distribution encodes the intrinsic structural properties of different types of network.

The remainder of the chapter is organised as follows. In Sec.4.2, we briefly introduce the von Neumann entropy with its approximate degrees of nodes connected by an edge. In Sec.4.3, we develop an entropic network characterisation from the heat bath analogy and Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics, and then describe our edge entropy projection. In Sec.4.5, we undertake experiments to demonstrate the usefulness of this novel method. Finally, we conclude our chapter with a summary of our contribution and suggestions for future work.

4.2 Entropy Representation

4.2.1 Density Matrix

In quantum mechanics, the density matrix is used to describe a system whose state is an ensemble of pure quantum states $|\psi_i\rangle$, each with probability p_i . The density matrix is defined as

$$\rho = \sum_{i=1}^{|V|} p_i |\psi_i\rangle\langle\psi_i| \tag{4.1}$$

Severini et al. [81] have extended this idea to the graph domain. Specifically, they show that a density matrix for a graph or network can be obtained by scaling the combinatorial Laplacian matrix by the reciprocal of the number of nodes in the graph. With this notation, the specified density matrix is obtained by scaling the normalised Laplacian matrix by the number of nodes, i.e.

$$\rho = \frac{\tilde{L}}{|V|} \tag{4.2}$$

When defined in this way the density matrix is Hermitian i.e. $\rho = \rho\dagger$ and $\rho \geq 0, \mathrm{Tr}[\rho] = 1$. It plays an important role in the quantum observation process, which can be used to calculate the expectation value of the measurable quantity.

4.2.2 Von Neumann Entropy

The interpretation of the scaled normalised Laplacian as a density operator, opens up the possibility of characterising a graph using the von Neumann entropy from quantum information theory. The von Neumann entropy is defined as the entropy of the density matrix associated with the state vector of a system. As noted above, Severini et al. [81] suggest how the von Neumann entropy can be computed by scaling the normalised discrete Laplacian matrix for a network. As a result the von Neumann entropy is given in terms of the eigenvalues λ_1, ..., $\lambda_{|V|}$ of the density matrix ρ,

$$S_{VN} = -\mathrm{Tr}(\rho \log \rho) = -\sum_{i=1}^{|V|} \frac{\lambda_i}{|V|} \log \frac{\lambda_i}{|V|} \tag{4.3}$$

The von Neumann entropy [81] computed from the normalised Laplacian spectrum has been shown to be effective for network characterisation. In fact, Han et al.[59] have shown how to approximate the calculation of von Neumann entropy in terms of simple degree statistics. Their approximation allows the cubic complexity of computing the von Neumann entropy from the Laplacian spectrum, to be reduced to one of quadratic complexity using simple edge degree statistics, i.e.

$$S_{VN} = 1 - \frac{1}{|V|} - \frac{1}{|V|^2} \sum_{(u,v)\in E} \frac{1}{d_u d_v} \tag{4.4}$$

This expression for the von Neumann entropy allows the approximate entropy of the network to be efficiently computed and has been shown to be an effective tool for characterising structural property of networks, with extremal values for the cycle and fully connected graphs.

Therefore, the edge entropy decomposition is given as

$$S_{VN}^{edge}(u,v) = \frac{1}{|E|} - \frac{1}{|V||E|} - \frac{1}{|E||V|^2} \frac{1}{d_u d_v} \tag{4.5}$$

where $S_{VN} = \sum_{(u,v)\in E} S_{VN}^{edge}(u,v)$. This expression decomposes the global parameter of von

Neumann entropy on each edge with the relation to the degrees from the connection of two vertexes.

4.3 Thermodynamic Representation

4.3.1 Maxwell-Boltzmann Entropy

Taking the Hamiltonian to be the normalised Laplacian of the network, the canonical partition function for Maxwell-Boltzmann occupation statistics of the energy levels can be achieved as Eq.(3.10). The entropy of the system with N particles is given by Eq.(3.12). Since the density matrix commutes with the Hamiltonian operator, we have $\partial\rho/\partial t = 0$ and the system can be viewed as in equilibrium. So the entropy in the Maxwell-Boltzmann system is simply N times the von Neumann entropy of a single particle, as we might expect.

4.3.2 Bose-Einstein Entropy

For a system of the network, as the grand-canonical ensemble with a varying number of particles N and a chemical potential μ, the Bose-Einstein partition function is achieved as Eq.(3.14). The corresponding entropy is given by Eq.(3.16). This entropy depends on the chemical potential for the partition function and hence it is determined by the number of particles in the system. At the temperature β, the corresponding number of particles in the level i with energy λ_i is

$$n_i = \frac{1}{\exp[\beta(\lambda_i - \mu)] - 1} \tag{4.6}$$

As a result, the total number of particles in the system is

$$N = \sum_{i=1}^{|V|} n_i = \sum_{i=1}^{|V|} \frac{1}{\exp[\beta(\lambda_i - \mu)] - 1} = \mathrm{Tr}\left[\frac{1}{\exp(-\beta\mu)\exp[\beta\tilde{L}] - I}\right] \tag{4.7}$$

In order for the number of particles in each energy state to be non-negative, the chemical potential must be less than the minimum energy level, i.e. $\mu < \min\lambda_i$.

The equivalent function of density matrix in this case is given by

$$\rho_{BE} = \frac{1}{\mathrm{Tr}(\rho_1)+\mathrm{Tr}(\rho_2)} \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix} \tag{4.8}$$

where $\rho_1 = -\left(\exp[\beta(\tilde{L}-\mu I)]-I\right)^{-1}$ and $\rho_2 = \left(I-\exp[-\beta(\tilde{L}-\mu I)]\right)^{-1}$.

Since Bose-Einstein statistics allow particles to coalesce in the lower energy levels, the corresponding entropy reflects the smaller Laplacian eigenvalues most strongly. As a result, the number of connected components (the multiplicity of the zero eigenvalues), and spectral gap (the degree of bi-partiality in a graph) are most strongly reflected.

4.3.3 Fermi-Dirac Entropy

The Fermi-Dirac distribution applies to indistinguishable fermions with a maximum occupancy of one particle in each energy state. Particles cannot be added to states that are already occupied, and hence obey the Pauli exclusion principle. These particles behave like a set of free fermions in the complex network with energy states given by the network Hamiltonian.

The statistical properties of the networks are thus given by the Fermi-Dirac statistics of the equivalent quantum system, and the corresponding partition function is Eq.(3.17). From Eq.(3.8), the associated entropy of the Fermi-Dirac system is given by Eq.(3.19).

Under Fermi-Dirac statistics, on the other hand, the number of particles occupying the *i*th energy state is

$$n_i = \frac{1}{\exp[\beta(\lambda_i-\mu)]+1} \tag{4.9}$$

and the total number of particles in the network system is

$$N = \sum_{i=1}^{|V|} n_i = \sum_{i=1}^{|V|} \frac{1}{\exp[\beta(\lambda_i-\mu)]+1} = \mathrm{Tr}\left[\frac{1}{\exp(-\beta\mu)\exp[\beta\tilde{L}]+I}\right] \tag{4.10}$$

With a single particle per energy state, the chemical potential is hence just the *n*th energy level, and so $\mu = \lambda_n$.

Similarly, we find that the equivalent density matrix function

$$\rho_{FD} = \frac{1}{\mathrm{Tr}(\rho_3) + \mathrm{Tr}(\rho_4)} \begin{pmatrix} \rho_3 & 0 \\ 0 & \rho_4 \end{pmatrix} \tag{4.11}$$

where $\rho_3 = \left(I + e^{-\beta\mu}\exp[\beta\tilde{L}]\right)^{-1}$ and $\rho_4 = \left(I + e^{\beta\mu}\exp[-\beta\tilde{L}]\right)^{-1}$.

Since Fermi-Dirac statistics exclude multiple particles from the same energy level, the corresponding entropy does not just reflect the lower part of the Laplacian spectrum and are sensitive to a greater portion of the distribution of Laplacian eigenvalues. As a result, we might expect them to be more sensitive to subtle differences in the network structure.

4.4 Edge Entropy Analysis

The edge entropy decomposition is to project the global network entropy onto the edges by multiplying the eigenvector matrices with the entropic elements. The matrix form for entropy can be written as,

$$S = -\mathrm{Tr}\left[\rho\log\rho\right] = -\mathrm{Tr}[\Sigma] \tag{4.12}$$

Since the spectral decomposition of the normalised Laplacian matrix is

$$\tilde{L} = \Phi\tilde{\Lambda}\Phi^T \tag{4.13}$$

We can decompose the matrix Σ as follows

$$\Sigma = \Phi\sigma(\tilde{\Lambda})\Phi^T \tag{4.14}$$

As a result, we can perform edge entropy projection using the Laplacian eigenvectors, with the result that the entropy of edge (uv) is given as,

$$S_{edge}(u,v) = \sum_{i=1}^{|V|} \sigma(\lambda_i)\varphi_i\varphi_i^T \tag{4.15}$$

where, for Maxwell-Boltzmann statistics,

$$\sigma_{MB}(\lambda_i) = -N\frac{e^{-\beta\lambda_i}}{\sum_{i=1}^{|V|}e^{-\beta\lambda_i}}\log\frac{e^{-\beta\lambda_i}}{\sum_{i=1}^{|V|}e^{-\beta\lambda_i}}$$

and for Bose-Einstein statistics,

$$\sigma_{BE}(\lambda_i) = -\sum_{i=1}^{|V|}\log\left(1-e^{\beta(\mu-\lambda_i)}\right) - \beta\sum_{i=1}^{|V|}\frac{(\mu-\lambda_i)e^{\beta(\mu-\lambda_i)}}{1-e^{\beta(\mu-\lambda_i)}}$$

and for Fermi-Dirac statistics,

$$\sigma_{FD}(\lambda_i) = \sum_{i=1}^{|V|}\log\left(1+e^{\beta(\mu-\lambda_i)}\right) - \beta\sum_{i=1}^{|V|}\frac{(\mu-\lambda_i)e^{\beta(\mu-\lambda_i)}}{1+e^{\beta(\mu-\lambda_i)}}$$

Thus, the global entropy can be projected on the edges of the network system. This provides useful measures for local entropic characterisation of network structure in a relatively straightforward manner.

4.5 Experiments and Evaluations

In this section, we provide experiments to evaluate the proposed methods of edge entropy decomposition. We commence by assessing the performance on edge degrees by comparing the previous con Neumann entropy. Then we apply the real-world networks, i.e. PPIs, NYSE and fMRI, to distinguish significant structural variance.

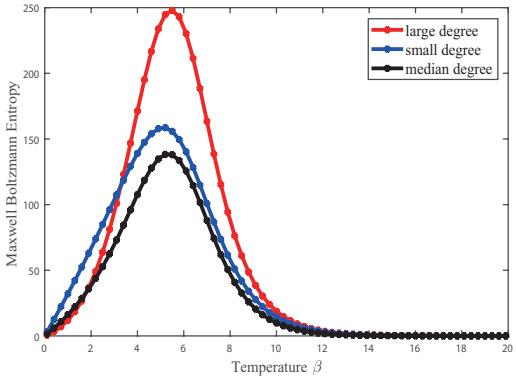
4.5.1 Data Sets

The first two datasets are Protein-Protein Interaction networks and New York Stock Exchange networks as introduced in chapter 3. The PPI dataset contains four groups of bacteria, i.e., Aquifex and Thermotoga-8, Gram-Positive-52, Cyanobacteria-73 and Proteobacteria-40 [44]. The financial networks are extracted from New York Stock Exchange over 6000 trading days with a total of 347 stocks.

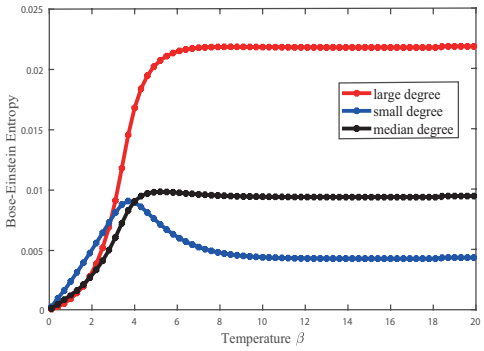
4.5.2 Experimental Results

We first investigate the temperature dependence of edge entropy with degree properties in three statistics. We select three different types of edges with different values of degrees at

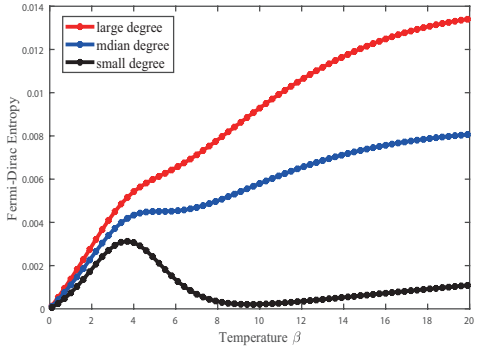
the vertices and explore how the entropy changes with temperature.



(a) Maxwell-Boltzmann Edge Entropy



(b) Bose-Einstein Edge Entropy

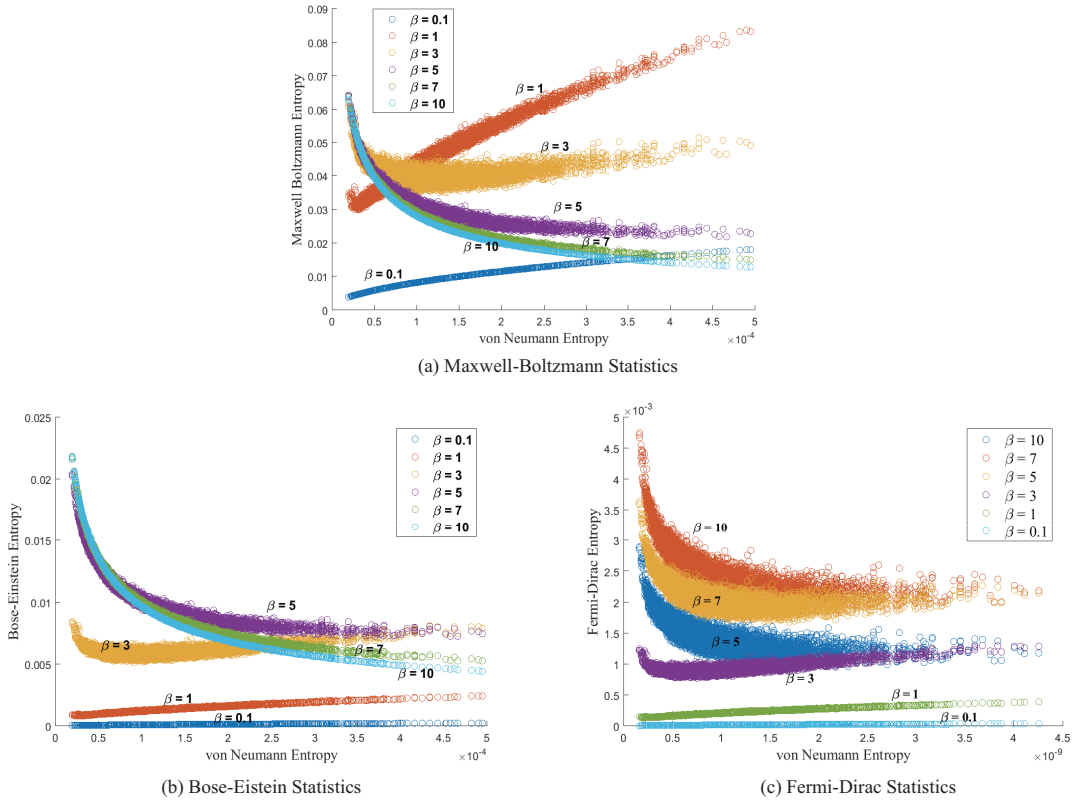


(c) Fermi-Dirac Edge Entropy

Figure 4.1: The temperature tendency of edge entropy with different degree on both ends in three statistics. The red line represents the high-degree edge; the blue line is the low-degree edge and the black line is the median value of degrees on the edge ends.

Fig.4.1(a) plots three selected edge entropies versus temperature with Maxwell-Boltzmann occupation statistics. The three edges show a similar dependence of entropy on the temperature. As the inverse of temperature (β) increases, the edge entropy reaches a maximum value. The edge entropy for vertices with the high degree increases faster than that for the low-degree in the high-temperature region. In the low-temperature limit, entropy approaches zero. This is because when the temperature decreases the configuration of particle occupation becomes identical as the particles always state at the low energy levels since the thermalisation effects vanish.

The quantum statistics, i.e., Bose-Einstein and Fermi-Dirac cases, exhibit the similar pattern like the Maxwell-Boltzmann statistics at the high temperature. As shown in Fig.4.1(b) and Fig.4.1(c), the edge entropy increases in both cases at the high-temperature



(a) Maxwell-Boltzmann Statistics

(b) Bose-Eistein Statistics

(c) Fermi-Dirac Statistics

Figure 4.2: Scatter plot of edge entropies compared to the von Neumann entropy with different value of temperatures.

region. However, as the temperature goes down (the inverse temperature β increases), the edge entropy in quantum statistics present a different tendency compared to the classical Maxwell-Boltzmann case. For Bose-Einstein statistics, the edge entropy reaches a constant value in the low-temperature limit. It is because the configuration of practices tend to condensate at the low energy state, which makes the edge entropy coalesce at a constant platform. On the other hand, particles in Fermi-Dirac statistics have a greater tendency to occupy high energy state with the Puli exclusion principle. This leads to edge entropy more distinguishable at the low-temperature limit with broad spread distribution of particles among the energy states. Two kinds of edge entropies in both quantum statistics cases are more sensitive to represent the degree structural difference in the low-temperature region compared to the Maxwell-Boltzmann statistics.

To better present the relationship between the edge entropies in three statistics and von Neumann entropy case, Fig.4.2 shows the edge entropy tendency with a different value of temperatures. All three statistical entropies exhibit a transition relationship with

von Neumann entropy in terms of changing temperature. For example, the Maxwell-

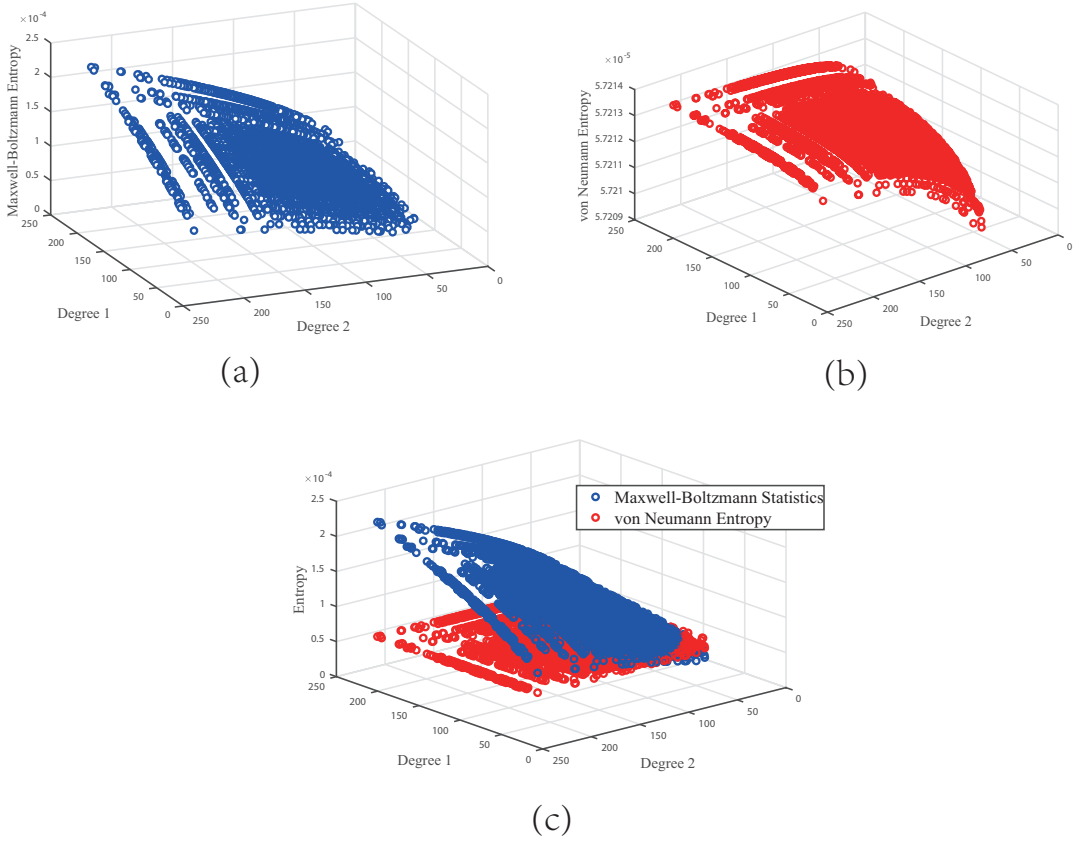


Figure 4.3: 3D scatter plot of edge entropy from Maxwell-Boltzmann statistics and von Neumann entropy. (a) Edge entropy in Maxwell-Boltzmann statistics. (b) Edge entropy from von Neumann formula. (c) Comparison of edge entropy between Maxwell-Boltzmann statistics and von Neumann entropy.

Boltzmann entropy is roughly in linear proportion to von Neumann entropy at the high temperature ($\beta = 0.1$). However, as the temperature reduces, it takes on an approximately exponential dependence. The Maxwell-Boltzmann edge entropy decreases monotonically with the von Neumann edge entropy in the low-temperature region ($\beta = 10$). The similar patterns can be observed in Bose-Einstein and Fermi-Dirac cases. The high temperature produces a proportional relationship to the von Neumann edge entropy, while the low temperature causes a transition to an inverse proportion between two entropies.

Furthermore, we explore the relationship between edge entropy in statistical methods and von Neumann case. Take the Maxwell-Boltzmann statistics as an example, we

show the 3D plots of edge entropy with the vertex degree in Fig.4.3 . The figure compares the edge entropy between Maxwell-Boltzmann statistics and von Neumann entropy with node degree connection for each edge in the network. The observation is that both entropies have a similar tendency with the degrees at the end. The Maxwell-Boltzmann edge entropy is more sensitive to the degree variance than the von Neumann entropy in the high degree region. The reason for this is the constant term in the von Neumann entropy formula dominates the value of edge entropy when the degrees are large. Thus, the Maxwell-Boltzmann edge entropy is better suited to represent the differences in graph structure associated with large degree nodes.

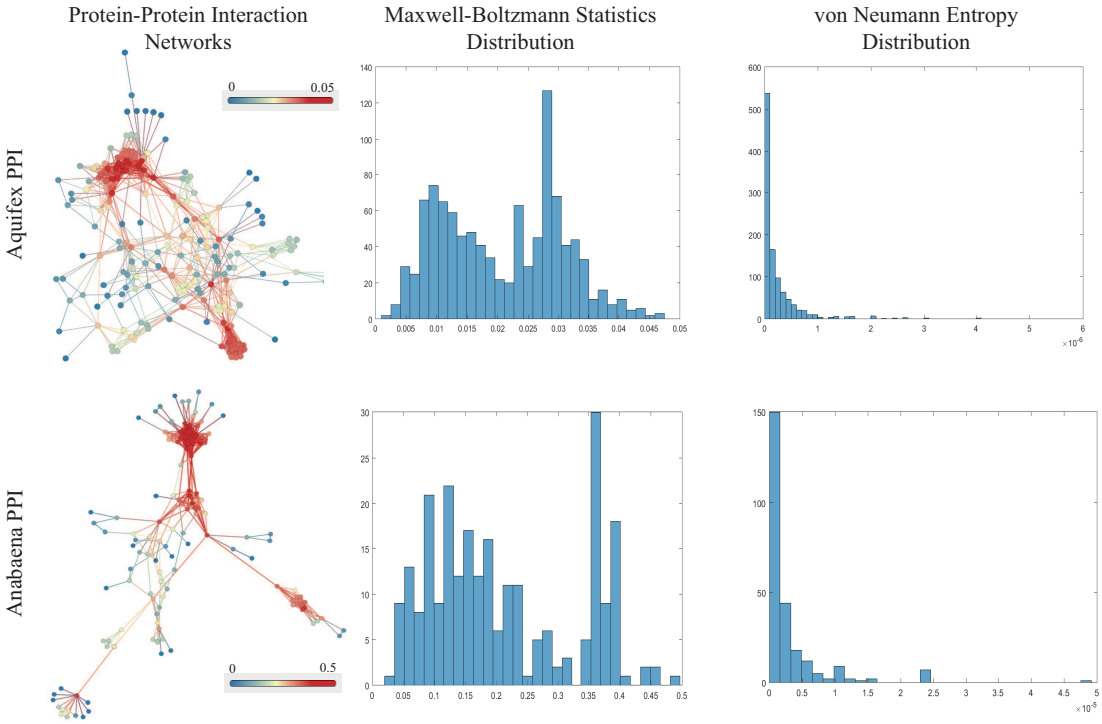


Figure 4.4: Examples of protein-protein interaction networks with edge entropy distribution of von Neumann entropy and Maxwell-Boltzmann statistics.

Now we apply the real-world PPI networks as an example to better illustrate the difference between edge entropy distributions in von Neumann entropy and the Maxwell-Boltzmann statistics. Fig.4.4 shows two examples of PPI networks, namely Anabaena variabilis and Aquifex aelicus together with their associated edge entropy histograms. The Maxwell-Boltzmann edge entropies are more sensitive to the presence of edges associated with high degree nodes, which provides better edge discrimination. This effect is

manifest in the differences of edge entropy histograms. In the Maxwell-Boltzmann case, the histogram shows two peaks in the edge entropy distribution, while the von Neumann edge entropy is concentrated at low values and has just a single peak. In other words, the von Neumann edge entropy offers less salient structure.

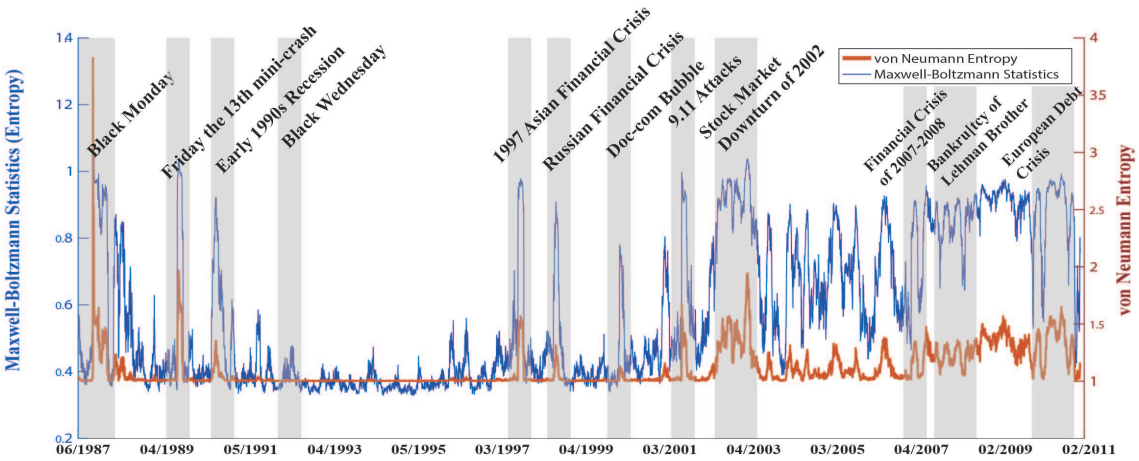


Figure 4.5: Entropy from Maxwell-Boltzmann statistics and von Neumann entropy for NYSE (1987-2011). Number of particles is $N = 1$ and temperature is $\beta = 10$.

Next, we turn our attention to the time evolution of networks. We take the NYSE network as an example to explore the entropic characterisation in the network structure. Fig.4.5 plots the total network for the Maxwell-Boltzmann and von Neumann cases. Both entropies reflect the positions of significant global financial events. In each case, the entropy undergoes significant fluctuations during the financial crises, associated with dramatic structural changes. Compared to the von Neumann entropy, the Maxwell-Boltzmann case is more sensitive to fluctuations in the network structure. A good example is Black Wednesday in 1992, which is obvious in the Maxwell-Boltzmann entropy but is not clear in the von Neumann case.

The similar entropic pattern can be observed in Fig.4.6, which shows both entropies from Bose-Einstein and Fermi-Dirac statistics with various financial events annotated, including Black Monday, Friday the 13th mini-crash, Early 1990s Recession, 1997 Asian Crisis, 9.11 Attacks, Downturn of 2002-2003, 2007 Financial Crisis, the Bankruptcy of Lehman Brothers and the European Debt Crisis. In each case, the entropy undergoes sharp increase corresponding to the financial crises, which are associated with dramatic

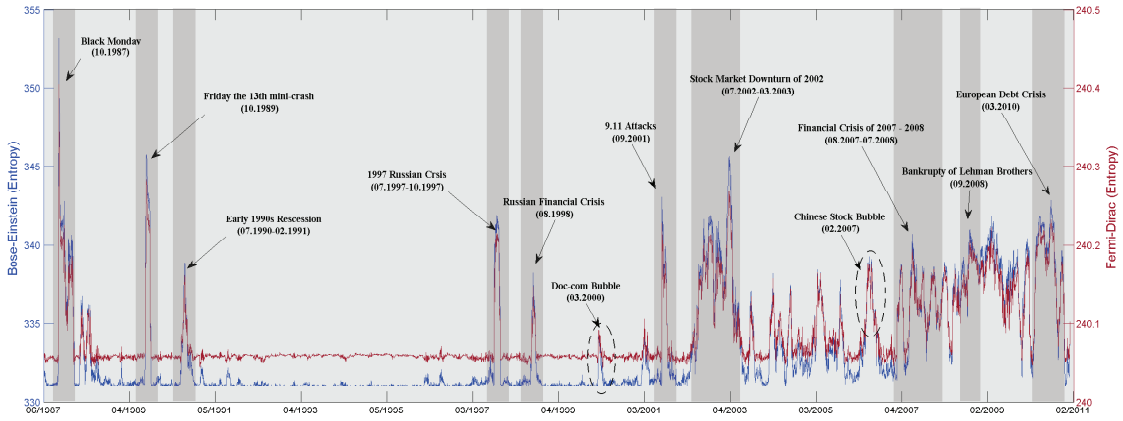


Figure 4.6: Entropy in NYSE (1987-2011) derived from Bose-Einstein and Fermi-Dirac statistics. Critical financial events, i.e., Black Monday, Friday the 13th mini-crash, Early 1990s Recession, 1997 Asian Crisis, 9.11 Attacks, Downturn of 2002-2003, 2007 Financial Crisis, the Bankruptcy of Lehman Brothers and the European Debt Crisis, can be represented in thermodynamic entropy with Maxwell-Boltzmann statistic. It is efficient to use the partition function associating with entropy to identify events in NYSE.

structural changes in the networks. Similarly the Maxwell-Boltzmann entropy in Fig.4.5, the quantum entropies are also effective in indicating the critical events. Moreover, the Bose-Einstein quantities show the greatest variation during the crises, suggesting that changes in cluster-structure (modularity) are important during these episodes.

We now focus in detail on one particular critical financial event, i.e., Black Monday in October 1987, to explore the dynamic structural difference with the entropic variance. We visualize the network structure at three-time epochs, i.e., before, during and after Black Monday, and compare the three statistical edge entropy distribution with von Neumann entropy case. Fig.4.7 shows the network structure and edge entropy distribution during the crisis. Before Black Monday, the stocks are highly connected with a large number of densely connected clusters of stocks following the same trading trends. This feature is reflected in edge entropy distribution of three statistics. However, during Black Monday, the number of connections between stock decrease significantly with large numbers of nodes becoming disconnected. Some stocks do though slightly increase their number of links with other stocks. This manifests itself as a shift of the peak to the high entropy region of the distribution. After Black Monday, the stocks begin to recover connections with another and a few stocks tend to form some clusters in the network structure. The node degree distribution also returns to its previous shape. In contrast, the von

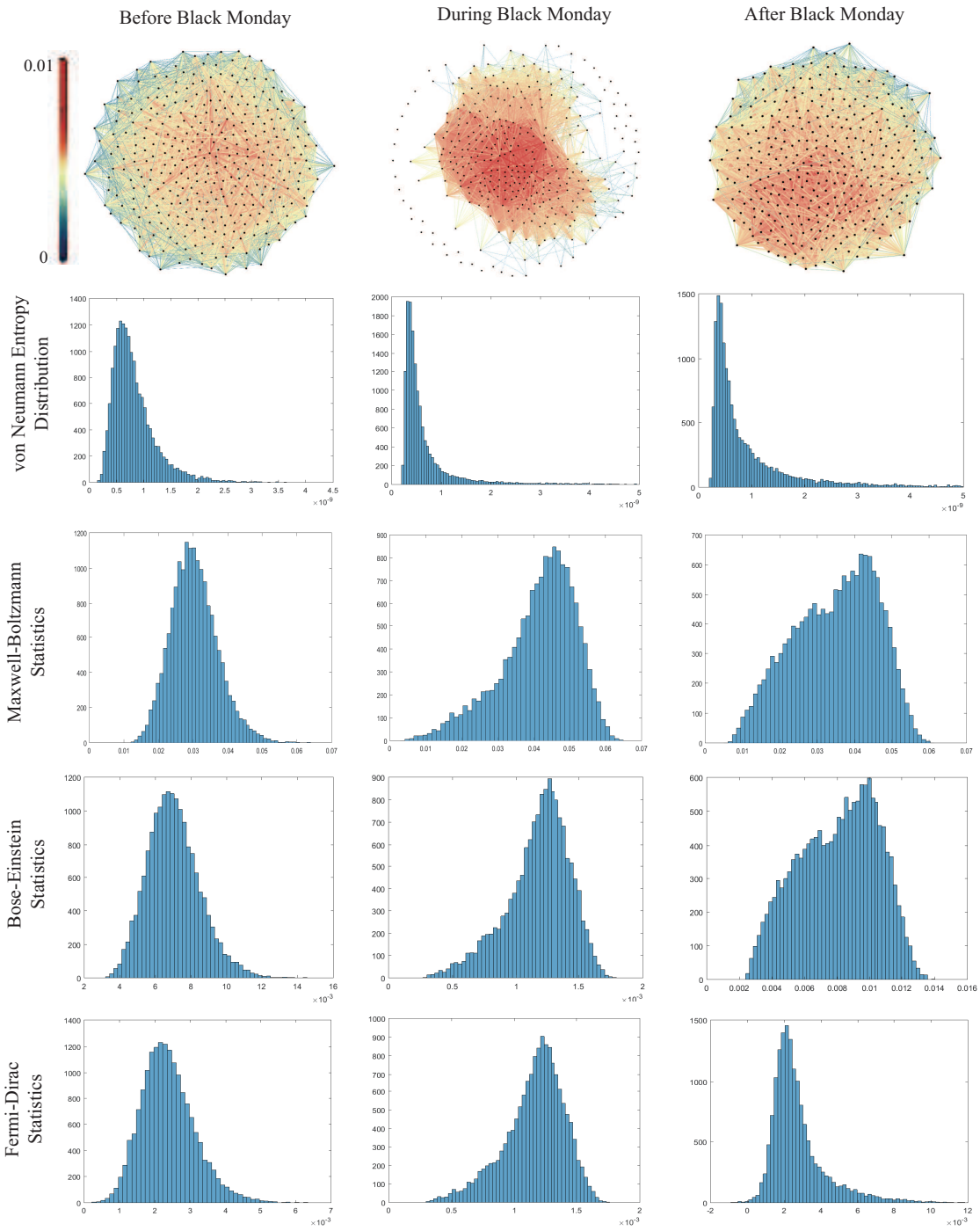


Figure 4.7: Visualisation of network structure before, during and after Black Monday. Edge entropy distribution is computed from von Neumann entropy, Maxwell-Boltzmann statistics, Bose-Einstein statistics and Fermi-Dirac statistics. The statistical model such as the Maxwell-Boltzmann case is more sensitive to represent the dynamic structure in the networks.

Neuman edge entropy distribution does not completely reflect the details of these critical structural changes. Compared to the three statistical edge entropy, the distribution of von Neumann edge entropy does not change significantly during Black Monday and hence does not effectively characterise the dynamic structure of the network. Moreover, an interesting observation is the difference of edge entropy distribution between Bose-Einstein and Fermi-Dirac statistics after Black Monday. This is because the networks exist some clusters with community structure. Since Bose-Einstein statistics preferentially sample the lower energy levels with the network eigenvalue spectrum, it is more suitable to detect networks with strong community edge connection. While Fermi-Dirac statistics may be more sensitive to the mean and variance of the eigenvalue distribution since they probe a wider range of energy levels.

In conclusion, all of the statistical methods and von Neumann edge entropies can be used to represent changes in network structure. Compared to the von Neumann edge entropy, the Maxwell-Boltzmann edge entropy is more sensitive to variance associated with the degree distribution. In the high-temperature region, the quantum statistics have similar degree sensitivity to the Maxwell-Boltzmann edge-entropy. However, in the low-temperature region, Bose-Einstein statistics is more sensitive to reflect strong community edge connection; while Fermi-Dirac edge entropy is more suitable to represent high degree variations.

4.6 Summary

This chapter has explored the thermodynamic characterisations of networks resulting from Maxwell-Boltzmann statistics, Bose-Einstein statistics and Fermi-Dirac statistics, and specifically those associated with the thermalisation effects of the heat bath on the occupation of the normalised Laplacian energy states. We view the normalised Laplacian matrix as the Hamiltonian operator of the network with associated energy states which can be occupied by classical distinguishable particles and quantum identical particles. This extends the use of entropy as a tool to characterise network structures in both static and

time series data. To compare with the extensively studied von Neuman entropy, we conduct the experiments which demonstrate that the thermodynamic edge entropy is better suited to represent the intrinsic structural properties associated to long-tailed degree distributions. The results reveal that all of the statistical entropies are effective in characterising dynamic network structure and distinguish different types of network models. Both quantum spin statistics present the similar effects correspond to the classical Maxwell-Boltzmann case since they are disrupted by thermalisation at the high temperatures. But, at low temperatures region, the phenomenon of Bose-Einstein and Fermi-Dirac statistics are significant different producing quite different entropic characterisations of network structure. Bose-Einstein system condenses into a state where the particles occupy the lowest energy state, which preferentially samples the lower energy levels with the network eigenvalue spectrum. The resulting entropy is more suitable to detect networks with strong community edge connection. Fermi-Dirac system, on the other hand, follows the Puli exclusion principle with only one particle per energy level. It probes a wider range of network spectrum which is more sensitive to the mean and variance of the eigenvalue distribution.

Chapter 5

Modelling Network Evolution

In this chapter, we investigate both undirected and directed network evolution using the Euler-Lagrange equation. We use the Euler-Lagrange equation to develop a variational principle based on the von Neumann entropy for time-varying network structure. Commencing from recent work to approximate the von Neumann entropy using simple degree statistics, the changes in entropy between different time epochs are determined by correlations in the degree difference of edge connections. Our Euler-Lagrange equation minimises the change in entropy and allows for the development of a dynamic model to simulate the changes of node degree with time. We first explore the effect of network dynamics on the three widely studied complex network models, namely a) Erdős-Rényi random graphs, b) Watts-Strogatz small-world networks, and c) Barabási-Albert scale-free networks. Our model effectively captures both undirected and directed structural transitions in the dynamic network models. We apply our model to a network time sequence representing the evolution of stock prices on the New York Stock Exchange (NYSE) and sequences of Drosophila gene regulatory networks containing different developmental phases of the organism from embryo to adult. Here we use the model to differentiate between periods of stable and unstable stock price trading and to detect periods of anomalous network evolution. Our experiments show that the presented model not only provides an accurate simulation of the degree statistics in time-varying networks, but that is also captures the topological variations taking place when the structure of a network changes violently.

5.1 Introduction

The study of network evolution plays an increasingly crucial role in modelling and predicting the structural variance of complex networks [112]. Previous studies have addressed this problem from the perspectives of both the local and the global characterisation of network structure. At the local level, the aim is to model how the detailed connectivity structure changes with time [115, 71]. Specifically, networks grow and evolve with the addition of new components and connections, or the rewiring of connections from one component to another [12, 47]. On the other hand, at the global level, the aim is to model the evolution of characteristics which capture the structure and hence the function of a network to allow different types of network function to be distinguished from one to another. Thermodynamic analysis of network structure allows the macroscopic properties of network structure to be described in terms of variables such as temperature, associated with the internal structure [115]. There are also models developed to learn the patterns of network evolution. Examples here include generative and autoregressive models which allow the detailed evolution of edge connectivity structure to be estimated from noisy or uncertain input data [60].

However, both the global and the local methods require to us to develop models that can be fitted to the available data by estimating their parameters, which describe how vertices interact through edges and how this interaction evolves with time. There are few methods that are both simple and effectively predict the evolution of network structure. 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, there have been a number of attempts to extend the scope of probabilistic generative models using various forms of regressive or autoregressive models [60, 8]. However, these essentially local models are parameter intensive and a simpler approach is to coach the model in terms of how different node degree configurations co-occur on the edges connecting them [115, 108].

In recent work we have addressed the problem by detailing a generative model of graph-structure [60] and have shown how it can be applied to network time series using an autoregressive model [8, 116]. One of the key elements of this model is a means of

approximating the von Neumann entropy for both directed and undirected graphs [59]. Von Neumann entropy is the extension of the Shannon entropy defined over the re-scaled eigenvalues of the normalised Laplacian matrix. A quadratic approximation of the von Neumann entropy gives a simple expression for the entropy associated with the degree combinations of nodes forming edges [107, 115]. In accordance with intuition, those edges that connect high degree vertices have the lowest entropy, while those connecting low degree vertices have the highest entropy [115, 108]. Making connections between low degree vertices is thus entropically unfavourable. Moreover, the fitting of the generative model to dynamic network structure involves a description length criterion which describes both the likelihood of the goodness of fit to the available network data together with the approximate von Neumann entropy of the fitted network. This latter term regulates the complexity of the fitted structure [112, 8], and mitigates against overfitting of the irrelevant or unlikely structure. Moreover, the change in entropy of the two vertices forming an edge between different epochs depends on the product of the degree of one vertex and the degree change of the second vertex. In other words, the change in entropy depends on the structure of the degree change correlations.

The aim of this chapter is to explore whether our model of network entropy can be extended to model the way in which the node degree distribution evolves with time, taking into account the effect of degree correlations caused by the degree structure of edges. We exploit this property by modelling the evolution of network structure using the Euler-Lagrange equations. Our variational principle is to minimise the changes in entropy during the evolution. Using our approximation of the von Neumann entropy, this leads to update equations for the node degree which include the effects of the node degree correlations induced by the edges of the network. It is effectively a type of diffusion process that models how the degree distribution propagates across the network. In fact, it has elements similar to preferential attachment [12], since it favours edges that connect high degree nodes [104, 108].

This model can also be extended to directed graphs. In prior work we have developed

approximate expressions for the von Neumann entropy of directed graphs [117], considering the cases where there is a) a mixture of unidirectional and bidirectional edges, b) where the unidirectional edges dominate (strongly directed graphs) and c) where the bidirectional edges outnumber the unidirectional edges (weakly directed graphs). Here we focus on the strongly directed graphs, where edges are purely unidirectional and there are no bi-directional edges. Our model distinguishes between the in-degree and out-degree of vertices, and we develop Euler-Lagrange equations for how the distributions quantities evolve with time.

The remainder of the chapter is organised as follows. In Sec.5.2, we provide a detailed analysis of entropy changes in dynamic networks and develop models for degree statistics by minimising the von Neumann entropy change using the Euler-Lagrange equations. We theoretically analyse both undirected and directed networks separately. In Sec.5.3, we conduct numerical experiments on the synthetic and real-world time-varying networks and apply the resulting characterisation of network evolution. Finally, we conclude the chapter and make suggestions for future work.

5.2 Variational Principle on Graphs

5.2.1 Directed Network Entropy

Severini *et al.* [81] exploit the concept of density matrix ρ from quantum mechanics in the network domain. They obtain the density matrix for a network by re-scaling the combinatorial Laplacian matrix by the reciprocal of the number of nodes in the graph. Han *et al.* render the computation of entropy more tractable by making a second order approximation to the Shannon entropy [59]. In so-doing they re-express the entropy it in terms of the traces of the normalised Laplacian and its square. The resulting approxminate von Neumann entropy depends on the degrees of pairs of nodes forming edges.

For directed graphs, the approximate von Neumann entropy is related to the in-degree and out-degree of the nodes [117]. First, the edge set E is divided into two subsets E_1 and E_2, where $E_1 = \{(u,v)|(u,v) \in E \text{ and } (v,u) \notin E\}$ is the set of unidirectional edges, $E_2 = \{(u,v)|(u,v) \in E \text{ and } (v,u) \in E\}$ is the set of bidirectional edges. The two edge-sets

satisfy the conditions $E_1 \cup E_2 = E, E_1 \cap E_2 = \emptyset$. With this distinction between unidirectional and bidirectional edges, the analogous approximation for the von Neumann entropy of a directed graph is,

$$S_d = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2} \left\{ \sum_{(u,v)\in E} \frac{d_u^{in}}{d_v^{in} d_u^{out^2}} + \sum_{(u,v)\in E_2} \frac{1}{d_u^{out} d_v^{out}} \right\} \tag{5.1}$$

To simplify the expression according to the relative importance of the sets of unidirectional and bidirectional edges E_1 and E_2, the von Neumann entropy can be further approximated to distinguish between weakly and strongly directed graphs. For weakly directed graphs, i.e., $|E_1| \ll |E_2|$ most of the edges are bidirectional, and we can ignore the summation over E_1 in Eq.(5.1), rewriting the remaining terms in curly brackets as

$$S_{wd} = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2} \left\{ \sum_{(u,v)\in E} \frac{\frac{d_u^{in}}{d_u^{out}} + \frac{d_v^{in}}{d_v^{out}}}{d_u^{out} d_v^{in}} \right\} \tag{5.2}$$

For the strongly directed graph the unidirectional edges dominate, i.e., $|E_1| \gg |E_2|$, there are few bidirectional edges, and we can ignore the summation over E_2 in Eq.(5.1), giving the approximate entropy as

$$S_{sd} = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2} \left\{ \sum_{(u,v)\in E} \frac{d_u^{in}}{d_v^{in} d_u^{out^2}} \right\} \tag{5.3}$$

Thus, both the strongly and weakly directed graph entropies depend on the graph size and the in-degree and out-degree statistics of edge connections [117].

5.2.2 Euler-Lagrange Equation

We would like to understand the dynamics of a network which evolves so as to minimise the entropy change between different sequential epochs. To do this we cast the evolution process into a variational setting of the Euler-Lagrange equation, and consider the system which optimises the functional

$$\mathcal{E}(q) = \int_{t_1}^{t_2} \mathcal{G}\left[t, q(t), \dot{q}(t)\right] dt \tag{5.4}$$

where t is time, $q(t)$ is the variable of the system as a function of time, and $\dot{q}(t)$ is the time derivative of $q(t)$. Then, the Euler-Lagrange equation is given by

$$\frac{\partial \mathcal{G}}{\partial q}[t, q(t), \dot{q}(t)] - \frac{d}{dt}\frac{\partial \mathcal{G}}{\partial \dot{q}}[t, q(t), \dot{q}(t)] = 0 \tag{5.5}$$

Here we consider an evolution which changes just the edge connectivity structure of the vertices and does not change the number of vertices in the graph. As a result, the factors $1 - \frac{1}{|V|}$ and $\frac{1}{|V|^2}$ are constants and do not affect the solution of the Euler-Lagrange equation.

5.2.3 Undirected Graphs

Suppose that two undirected graphs $G_t = (V_t, E_t)$ and $G_{t+\Delta t} = (V_{t+\Delta t}, E_{t+\Delta t})$ represent the structure of a time-varying complex network at two consecutive epochs t and $t + \Delta t$ respectively. Then the change of approximate von Neumann entropy between two sequential undirected graphs can be written a

$$\Delta S = S(G_{t+\Delta t}) - S(G_t) = \frac{1}{|V|^2} \sum_{(u,v)\in E,E'} \frac{d_u\Delta_v + d_v\Delta_u + \Delta_u\Delta_v}{d_u(d_u+\Delta_u)d_v(d_v+\Delta_v)} \tag{5.6}$$

where Δ_u is the change of degree for node u, i.e., $\Delta_u = d_u^{t+\Delta t} - d_u^t$; Δ_v is similarly defined as the change of degree for node v, i.e., $\Delta_v = d_v^{t+\Delta t} - d_v^t$. The entropy change is sensitive to degree correlations for pairs of nodes connected by an edge.

We aim to study evolutions that minimise the entropy change associated with the structure of the degree change correlations, i.e. minimise the entropy change between time intervals. In order to represent the change of entropy more accurately, here, we approximate the denominator in Eq.(5.6) to the quadratic term and apply the Euler-Lagrange equation $\mathcal{G} = \Delta S$ with the entropy change to obtain

$$\mathcal{G}[t, d_u(t), \Delta_u(t), d_v(t), \Delta_v(t)] = \frac{d_u\Delta_v + d_v\Delta_u + \Delta_u\Delta_v}{d_u^2 d_v^2} \tag{5.7}$$

For the vertex indexed u with degree d_u, the Euler-Lagrange equation in Eq.(5.5)

gives,

$$\frac{\partial \mathcal{G}}{\partial d_u} - \frac{d}{dt}\frac{\partial \mathcal{G}}{\partial \Delta_u} = 0 \tag{5.8}$$

First, solving for the partial derivative of the degree d_u, we find

$$\frac{\partial \mathcal{G}}{\partial d_u} = -\frac{d_u \Delta v + 2 d_v \Delta u + 2\Delta u \Delta v}{d_u^3 d_v^2} \tag{5.9}$$

The detailed analysis above not only involves the terms to first order in the node degree change but also those of second order, i.e. degree difference correlations of the form $\Delta u \Delta v$.

Then computing the partial time derivative to the first order degree difference Δ_u, we obtain

$$\frac{\partial \mathcal{G}}{\partial \Delta_u} = \frac{d_v + \Delta v}{d_u^2 d_v^2} \tag{5.10}$$

Substituting Eq.(5.9) and Eq.(5.10) into Eq.(5.8),

$$\frac{\partial \mathcal{G}}{\partial d_u} - \frac{d}{dt}\frac{\partial \mathcal{G}}{\partial \Delta_u} = \frac{2\Delta^2 u - d_u \dot{\Delta u}}{d_u^3 d_v^2} = 0 \tag{5.11}$$

The solution for Euler-Lagrange equation in terms of node degree difference is

$$\Delta_u = \left(\frac{d_u}{d_v}\right)^2 \Delta_v + C \tag{5.12}$$

where C is the constant term coming from the integral of the differential equation. This leads to a detailed degree update equation which involves a square term of d_u/d_v and plus a constant C. Since it considers the effects of second order terms in the change of von Neumann entropy, this solution is accurate in predicting the degree distribution

As a result, it gives a relationship between the degree changes of nodes connected by an edge when solving the Euler-Lagrange equation which minimises the change in entropy over time. Since we are concerned with understanding how network structure

changes with time, the solution of the Euler-Lagrange equation provides a way of modelling the effects of these structural changes on the degree distribution across nodes in the network. The update equation for the node degree is at time epochs t and $t+\Delta t$ is

$$d_u^{t+\Delta t} = d_u^t + \sum_{v\sim u} \dot{\Delta}_v \Delta_t = d_u^t + \sum_{v\sim u} \left(\frac{\Delta_u}{\Delta_t}\right)_v \Delta_t \tag{5.13}$$

In other words by summing over all edges connected to node u, we increment the degree at node u due to changes associated with the degree correlations on the set of connecting edges. We then leverage the solution of the Lagrange equation to simplify the degree update equation, to give the result

$$d_u^{t+\Delta t} = d_u^t + \sum_{v\sim u} \left(\frac{d_u}{d_v}\right)^2 \Delta_v + C \tag{5.14}$$

This can be viewed as a type of diffusion process, which updates edge degree so as to satisfy constraints on degree change correlation so as to minimise the entropy change between time epochs. Specifically, the update of degree reflects the effects of correlated degree changes between nodes connected by an edge.

5.2.4 Directed Graphs

Weakly Directed Graphs

In order to accommodate directed edges, we consider the node u and let d_u^{in} be the number of edges incident on vertex u or in-degree and d_u^{out} be the number of edges leaving vertex u or out-degree. The ratio of in-degree to out-degree is $r_u = \frac{d_u^{in}}{d_u^{out}}$ and $r_v = \frac{d_v^{in}}{d_v^{out}}$. We use this ratio to re-write the directed graph entropies in terms of in-degree and in/out degree ratio. As a result, the weakly directed graph entropy is

$$S_{wd} = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2}\left\{\sum_{(u,v)\in E} \frac{r_u(r_u+r_v)}{d_u^{in} d_v^{in}}\right\} \tag{5.15}$$

For two weakly directed graphs $G_{wd}^t = (V_t, E_t)$ and $G_{wd}^{t+\Delta t} = (V_{t+\Delta t}, E_{t+\Delta t})$, representing the structure of a time-varying complex network at two consecutive epochs t and

$t+\Delta t$ respectively, the change of von Neumann entropy is given by

$$\begin{aligned}\Delta S_{wd} &= S(G_{wd}^{t+\Delta t}) - S(G_{wd}^{t}) \\ &= -\frac{1}{2|V|^2}\sum_{(u,v)\in E,E'}\left\{\frac{(2r_u+r_v)\Delta r_u + r_u\Delta r_v}{d_u^{in}d_v^{in}} - \frac{r_u(r_u+r_v)(d_u^{in}\Delta_v^{in}+d_v^{in}\Delta_u^{in})}{(d_u^{in}d_v^{in})^2}\right\}\end{aligned} \tag{5.16}$$

where Δ_u^{in} is the change of in-degree for node u, i.e., $\Delta_u^{in} = d_u^{in}(t+\Delta t) - d_u^{in}(t)$; Δ_v^{in} is similarly defined as the change of in-degree for node v, i.e., $\Delta_v^{in} = d_v^{in}(t+\Delta t) - d_v^{in}(t)$. Δr_u and Δr_v are the change of degree ratio for the node u and node v respectively.

The Euler-Lagrange equation for r_u gives

$$\frac{\partial \Delta S_{wd}}{\partial r_u} - \frac{d}{dt}\frac{\partial \Delta S_{wd}}{\partial \Delta r_u} = -\frac{2(2r_u+r_v)(d_u^{in}\Delta_v^{in}+d_v^{in}\Delta_u^{in})}{(d_u^{in}d_v^{in})^2} = 0 \tag{5.17}$$

and similarly for r_v gives

$$\frac{\partial \Delta S_{wd}}{\partial r_v} - \frac{d}{dt}\frac{\partial \Delta S_{wd}}{\partial \Delta r_v} = -\frac{2r_u(d_u^{in}\Delta_v^{in}+d_v^{in}\Delta_u^{in})}{(d_u^{in}d_v^{in})^2} = 0 \tag{5.18}$$

Combining the Eq.(5.17) and Eq.(5.18), the relationship between d_u^{in} and d_v^{in} is

$$\frac{\Delta_u^{in}}{d_u^{in}} = -\frac{\Delta_v^{in}}{d_v^{in}} \tag{5.19}$$

Thus, for the weakly directed graph, there exists a linear correlation between Δ_u^{in}/d_u^{in} and Δ_v^{in}/d_v^{in}.

Strongly Directed Graphs

For a strongly directed graph, the von Neumann entropy in Eq.(5.2) can be expressed in terms of in-degree and in/out degree ratio as

$$S_{sd} = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2}\left\{\sum_{(u,v)\in E}\frac{r_u^2}{d_u^{in}d_v^{in}}\right\} \tag{5.20}$$

For two strongly directed graphs $G_{sd}^t = (V_t, E_t)$ and $G_{sd}^{t+\Delta t} = (V_{t+\Delta t}, E_{t+\Delta t})$, the

change of von Neumann entropy is

$$\begin{aligned}\Delta S_{sd} &= S(G_{sd}^{t+\Delta t}) - S(G_{sd}^{t}) \\ &= -\frac{1}{2|V|^2}\sum_{(u,v)\in E,E'}\frac{d_u^{in}d_v^{in}\Delta r_u - r_u(d_v^{in}\Delta_u^{in} + d_u^{in}\Delta_v^{in})}{(d_u^{in}d_v^{in})^2}\end{aligned} \tag{5.21}$$

where Δ_u^{in} is the change of in-degree for node u; Δ_v^{in} is similarly defined as the change of in-degree for node v.

Now we apply the Euler-Lagrange equation to the changes of entropy for strongly directed graph. The partial derivative of the ratio r_u is

$$\frac{\partial \Delta S_{sd}}{\partial r_u} = -\frac{d_u^{in}\Delta_v^{in} + d_v^{in}\Delta_u^{in}}{(d_u^{in}d_v^{in})^2} \tag{5.22}$$

And the partial time derivative to the first order ratio difference Δr_u is

$$\frac{\partial \Delta S_{sd}}{\partial \Delta r_u} = \frac{2}{d_u^{in}d_v^{in}} \tag{5.23}$$

Then, the solution of the Euler-Lagrange equation for r_u can be computed as

$$\frac{\partial \Delta S_{sd}}{\partial \Delta r_u} - \frac{d}{dt}\frac{\partial \Delta S_{sd}}{\partial \Delta r_u} = -\frac{2(d_u^{in}\Delta_v^{in} + d_v^{in}\Delta_u^{in})}{(d_u^{in}d_v^{in})^2} = 0 \tag{5.24}$$

Similarly, applying the Euler-Lagrange equation on the in-degree d_u^{in}, we get

$$\frac{\partial \Delta S_{sd}}{\partial d_u^{in}} - \frac{d}{dt}\frac{\partial \Delta S_{sd}}{\partial \Delta_u^{in}} = \frac{r_u(d_u^{in}\Delta_v^{in} + d_v^{in}\Delta_u^{in}) + d_v^{in}(r_u\Delta_u^{in} - 2d_u^{in}\Delta r_u)}{(d_u^{in})^3(d_v^{in})^2} = 0 \tag{5.25}$$

Substituting Eq.(5.24) into Eq.(5.25), the relationship between d_u and r_u can be obtained

$$\frac{\Delta_u^{in}}{d_u^{in}} = 2\frac{\Delta r_u}{r_u} \tag{5.26}$$

Therefore, the Euler Lagrange dynamics leads to a linear relationship between $\frac{\Delta_u^{in}}{d_u^{in}}$

and $\frac{\Delta r_u}{r_u}$ for strongly directed graphs. This should be compared to the analogous relationship which arises from the incremental analysis of the ratio $r_u = \frac{d_u^{in}}{d_u^{out}}$,

$$\Delta r_u = \frac{\Delta_u^{in}}{d_u^{out}} - \frac{d_u^{in}\Delta_u^{out}}{(d_u^{out})^2} \tag{5.27}$$

and as a result

$$\frac{\Delta r_u}{r_u} = \frac{\Delta_u^{in}}{d_u^{in}} - \frac{\Delta_u^{out}}{d_u^{out}} \tag{5.28}$$

Combining with Eq.(5.26) gives the growth equation

$$\frac{\Delta_u^{out}}{d_u^{out}} = \frac{1}{2}\frac{\Delta_u^{in}}{d_u^{in}} \tag{5.29}$$

which is the out-degree grows at half the rate of the in-degree. In the next section we explore empirically how well this relationship is observed.

5.3 Experimental Evaluation

5.3.1 Data Sets

Synthetic Time-evolving Networks: We generate three kinds of complex network models, namely, a) Erdős-Rényi random graph model, b) Watts-Strogatz small-world model [110], and c) Barabási-Albert scale-free model [12, 13]. These are created with the fixed number of vertices with changing the parameters with the network structure evolution. For the Erdős-Rényi random graph, the connection probability is monotonically increasing at the constant rate of 0.005. Similarly, the link rewiring probability in the small-world model [110] increases constantly between 0 to 1 as the network evolution. For the scale-free model [13], one vertex is added to the connection at each time step.

Drosophila Gene Regulatory Networks: The time-evolving network represents the DNA microarrays expressed at different developmental stages from fertilization to adulthood during the life cycle of Drosophila melanogaster. The developmental process has four stages, namely, the embryonic (1-30), larval (31-40), pupal (41-58) and adulthood (59-66). The vertices in the network are gene identities which vary in number from 588

to 4028 at different time epochs. This hence tests the ability of our method to deal with networks of variable size. The gene expression patterns are modelled as a binary Markov random field [96] which allow the edge connections to be determined.

Financial Directed & Undirected Networks: This dataset consists of the daily prices of 3,799 stocks traded continuously on the New York Stock Exchange over 6000 trading days. The stock prices were obtained from the Yahoo! financial database [93]. A total of 347 stock were selected from this set, for which historical stock prices from January 1986 to February 2011 are available. In our network representation, the nodes correspond to stock and the edges indicate that there is a statistical similarity between the time series associated with the stock closing prices [93].

To establish the edge-structure of the network we use a time window of 20 days is to compute the cross-correlation coefficients between the time-series for each pair of stock. Connections are created between a pair of stock if the cross-correlation exceeds an empirically determined threshold. In our experiments, we set the correlation coefficient threshold to the value to $\xi = 0.85$. This yields a time-varying stock market network with a fixed number of 347 nodes and varying edge structure for each of 6,000 trading days. The edges of the network, therefore, represent how the closing prices of the stock follow each other.

5.3.2 Synthetic Networks

We first conduct experiments on the synthetic networks. We generate three kinds of time-evolving network models from Erdős-Rényi random graphs, Watts-Strogatz small-world networks, and Barabási-Albert scale-free networks to evaluate our theoretical analysis.

Using the degree update equation derived from the principle of minimum entropy change and the Euler-Lagrange equation in Eq.(5.14), we turn our attention to synthetic network data to characterise the structural variance in network models. Fig.5.1 shows the visualisation of the time evolution for three complex networks. Since we fix the number of vertices to 200, for the random graphs, the networks evolve from an initially sparse set of edges with a low value of the connection probability. As the connection probability increases, the structure of the random graph exhibits a phase transition to a state with a

high density of connection and a giant connected component. A phase transition can also be observed for the Watts-Strogatz small-world model, as the rewiring probability evolves with time. Commencing from a regular ring lattice, the network structure evolves to a small-world network with high rewiring probability, and then to an Erdős-Rényi random graph structure with unit rewiring probability. For the scale-free network, the evolution takes place via preferential attachment. The nodes with the highest degree have the largest probability to receive new links. This process produces several high degree nodes or hubs in the network structure.

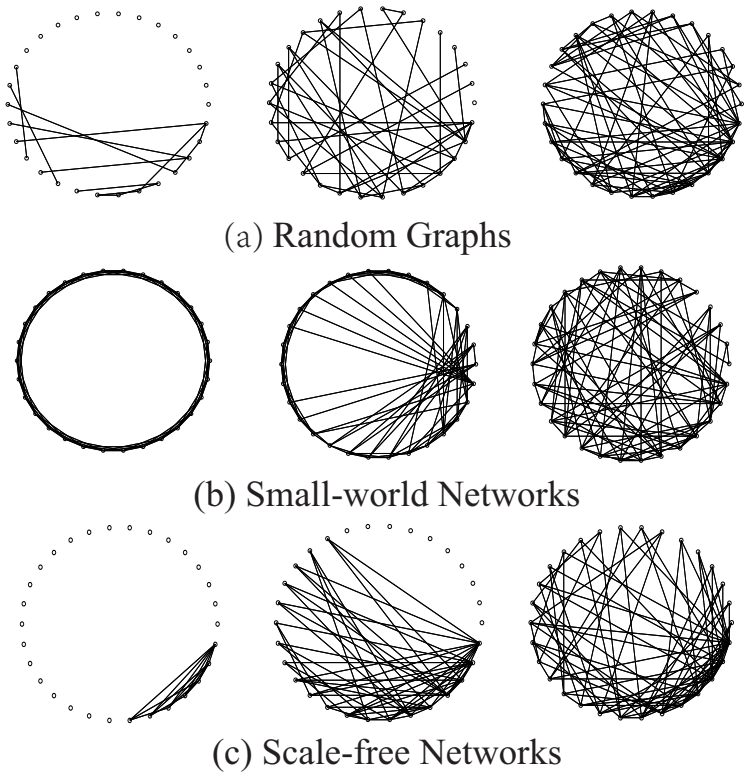
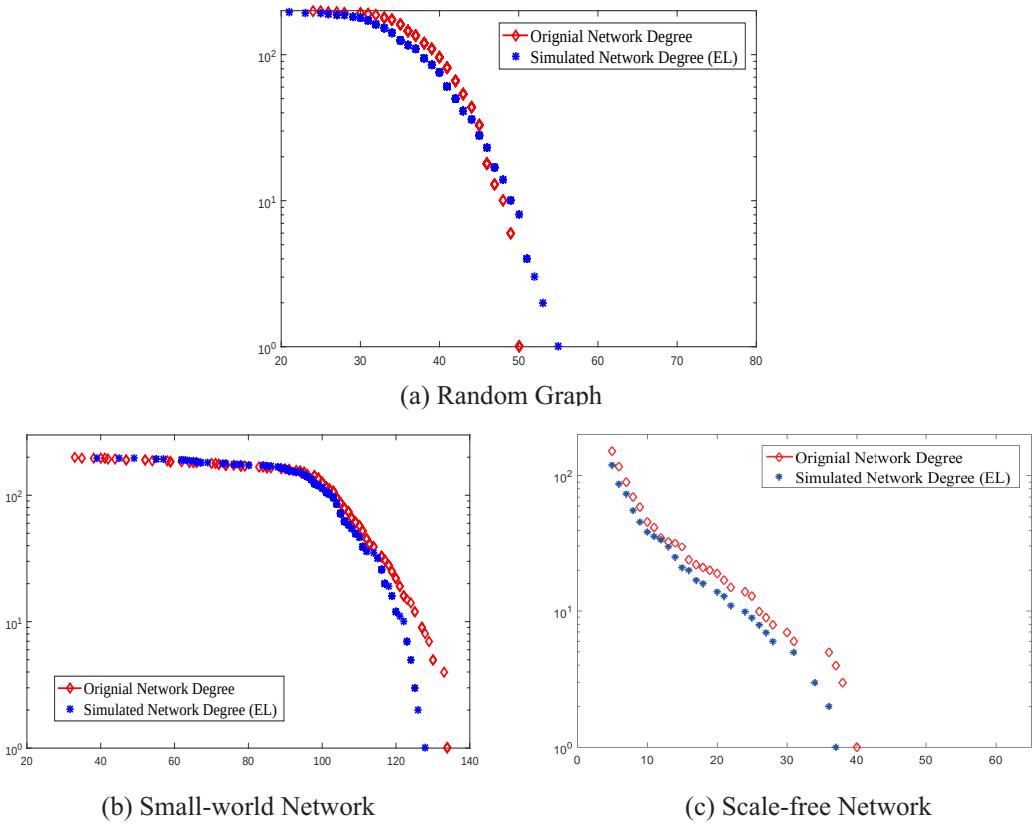


Figure 5.1: Visualisation of dynamic network structures in time evolution for three network models (Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks)

Now we explore whether the Euler-Lagrange equation can capture structural properties in the time evolution. We use our model to predict the network structure at subsequent time steps and simulate the degree distribution. We then to compare the predicted degree distribution with which from the original time series. Fig.5.2 shows the simulation results and degree distribution comparisons. The predicted degree distribution resulting from

Euler-Lagrange dynamics for the simulated networks fit quite well to the observed distributions. This provides empirical evidence that the Euler-Lagrange equation accurately predicts the short-term evolution of the different network models.



(a) Random Graph

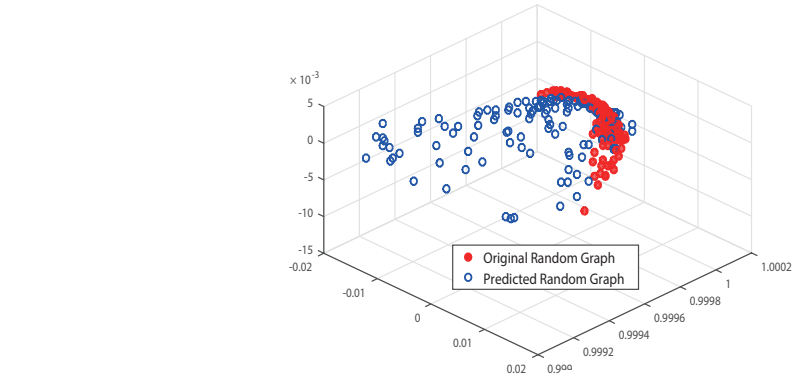
(b) Small-world Network

(c) Scale-free Network

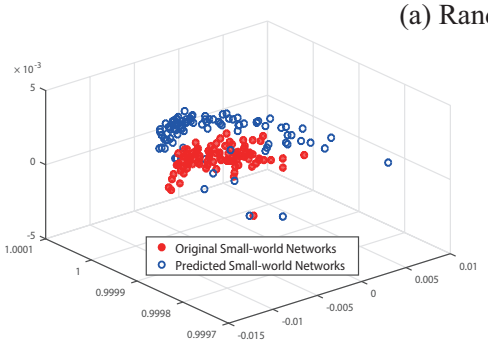
Figure 5.2: Degree distribution of original networks and simulated networks for three network models. The red line is for the originally observed networks and the blue line is for the results simulated with the second order Euler-Lagrange analysis. (Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks).

To visualise how the different networks evolve over extended time intervals, we apply the principal component analysis of the degree distribution to project the degree distribution sequences for the networks into a low dimensional space. To commence, we normalise the degree distributions so that the bin contents sum to unity, and then we construct a long vector from the normalised bin contents. We then construct the covariance matrix for the set of long vectors representing the observed degree distributions for the sample of networks. Finally, we apply principal component analysis to the sample covariance matrix for the sample of observed vectorised network degree distributions. We

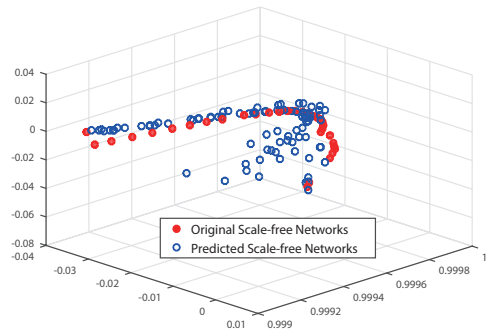
project both the observed and predicted distributions into the principal component space spanned by the leading three eigenvectors of the covariance matrix. In this way, we visualise the evolution of the observed and predicted degree distributions in the principal component space. The results are shown in Fig.5.3. The red points are the original network distributions and the blue ones are the predicted ones. Fig.5.3 clearly shows that for all three network models the predicted network degree distribution evolves in a similar manner to the observed network degree distribution.



(a) Random Graph



(b) Small-world Network



(c) Scale-free Network

Figure 5.3: Visualisation of degree distribution in network evolution with principal component analysis (Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks).

Then, we explore the effect of length of time step on the performance of the degree distribution prediction accuracy. Fig.5.4 shows the degree distribution error with a different value of the time step for the three different network models. The prediction error is the standard error over the normalised bin contents (the standard deviation of the difference in observed and predicted bin contents, divided by the square root of the number of

bins). The longer the time intervals Δt, the higher the prediction error in the degree distribution. For the random graph, the errors sharply increase around the step $\Delta t = 20$. This is because, during the evolution, the random graph undergoes a phase transition from being sparsely connected to containing a giant connected component. At large time intervals, the predictions fail for the reason of the presence of this giant component.

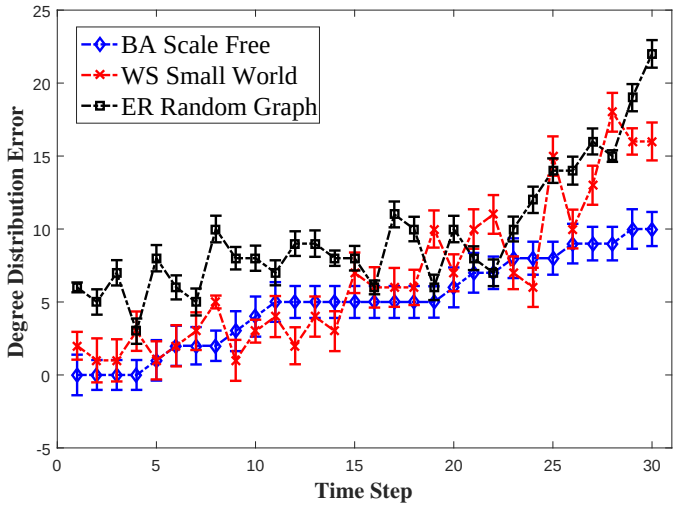


Figure 5.4: The degree distribution error with the different value of time steps for three network models (Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks). Degree prediction error increases quickly after time step $\Delta t = 20$.

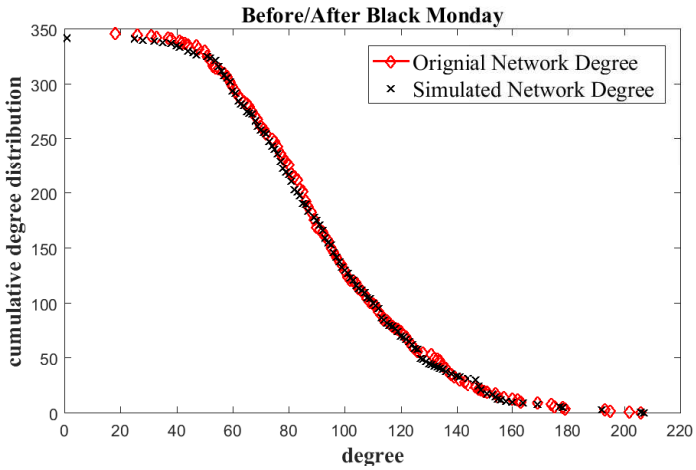


Figure 5.5: Degree distribution of originally observed networks and simulated networks before/after Black Monday.

A similar behaviour can be observed in the sample of small-world networks. As the time step interval increases, there are two instants in time separating three different

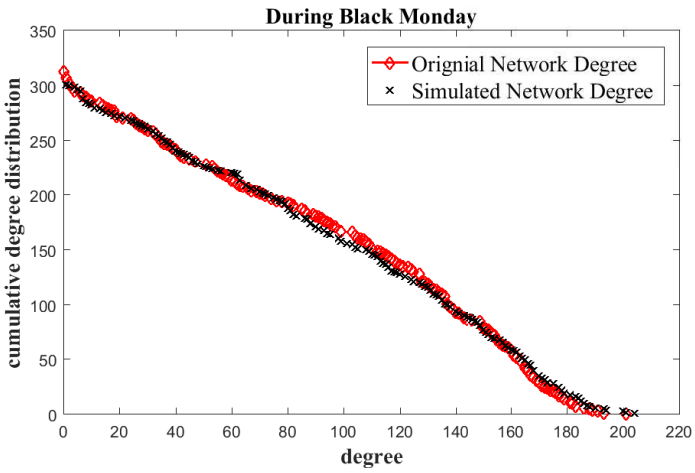


Figure 5.6: Degree distribution of originally observed networks and simulated networks during Black Monday. The network becomes disconnected and most vertices are disjoint, which results in the degree distribution following the power-law.

evolution models. The first event occurs around $\Delta t = 15$ and the second at $\Delta t = 25$. The reason is that, during the evolution, the structure of network changes from a regular lattice at the beginning to a small-world network, and then finally takes on a similar structure to a random graph. These three epochs and the associated with the impact of structural transition on the performance of degree distribution prediction. Finally, the degree prediction error for the scale-free network grows slowly and smoothly with the time step, since there are no significant structure transitions during the evolution. As a result, the topology of the scale-free network remains stable. Overall, increasing the value of the time interval results in a reduction of the prediction accuracy. Our new model is capable of capturing the local trends arising from the structural changes during the evolution.

5.3.3 Real-world Networks

For real-world network evaluation, we test our method on data provided by the Drosophila genes and the New York Stock Exchange. We first evaluate the undirected networks with the life cycle of Drosophila genes dataset. Then we construct the time sequential undirected and directed networks which consist of the daily prices of 3,799 stocks traded continuously on the New York Stock Exchange over 6000 trading days.

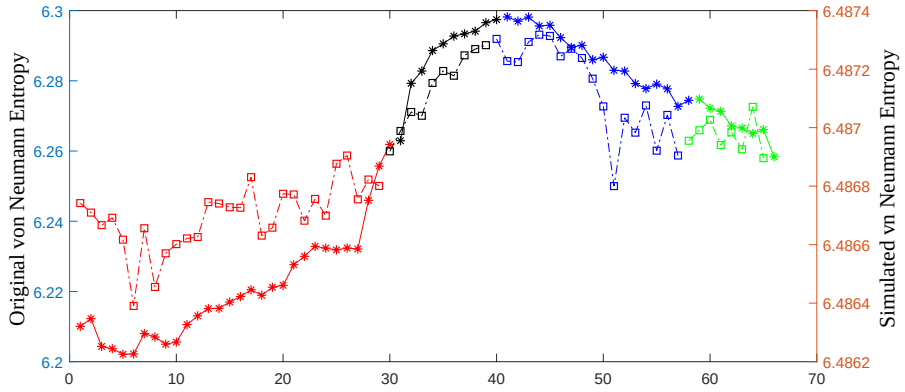


Figure 5.7: Comparison of entropy evolution in Drosophila gene regulatory networks using von Neumann entropy and the simulation with the Euler-Lagrange model. The four developmental phases are embryonic (red line), larval (black line), pupal (blue line), and adulthood (green line).

Undirected Drosophila Gene Regulatory Networks

To commence, we represent the Drosophila gene regulatory networks as undirected graphs evolving from the embryonic stage to the adulthood stage. The four phases of the Drosophila life cycle in genes represent the structural variations in the gene regulatory network connections.

We compare the computed von Neumann entropy of the network with that computed from the degree evolution predicted by the Euler-Lagrange model in Eq.(5.14). Fig.5.7 plots the two entropies for the entire life cycle of Drosophila development. The four developmental phases, namely, embryonic (red line), larval (black line), pupal (blue line), and adulthood (green line) are represented by different colours. The entropy predicted by the Euler-Lagrange model exhibits a similar time series compared to that obtained with the von Neumann entropy calculated from the observed degree distribution. In other words, the degree distribution predicted by the Euler-Lagrange equation effectively captures the changes in structure due to developmental changes in the gene regulatory networks.

Undirected Financial Networks

Now we simulate the behaviour of the financial market networks. Here we focus on how the degree distribution evolves with time. We compare the simulated structure and the observed network properties and provide a way to identify the consequence of structural

variations in time-evolving networks. Our procedure is as follows. We first select a network at a particular epoch from the time series and simulate its evolution using the degree update equation in Eq.(5.14). Then we compare the degree distributions for the real network sampled at a subsequent time and the simulation of the degree distribution after an identical elapsed time. One of the most salient events in the NYSE is Black Monday. This event occurred on October 19, 1987, during which the world stock markets crashed, dropping in value in a very short time.

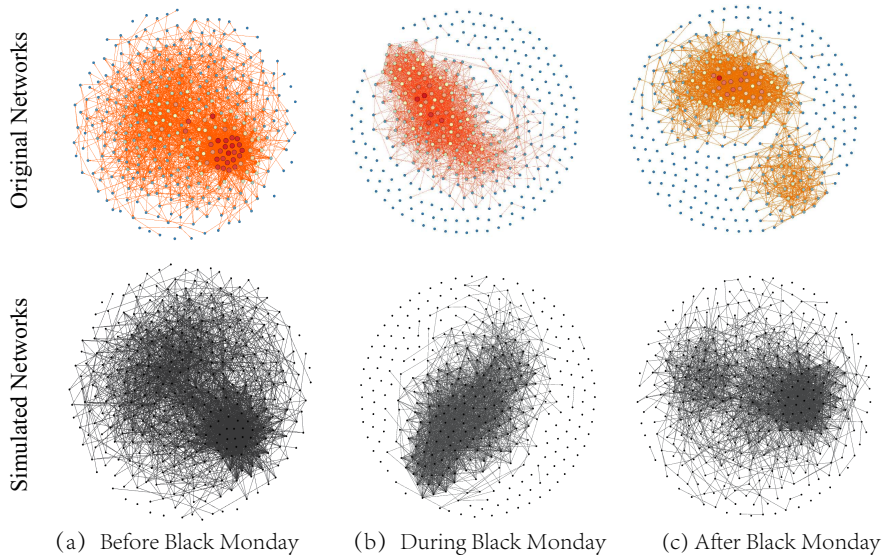


Figure 5.8: The visualisation of network structure for three specific days of Black Monday financial crisis. The red line corresponds to the entropy difference for the original networks and the gray line is the Euler-Lagrange model.

We compare the prediction of consecutive time steps at different epochs, before/after and during the Black Monday crisis. The results are shown in Fig.5.5 and Fig.5.6. The most obvious feature is that the degree distribution for the networks before and after Black Monday is quite different to that during the crisis period. During the Black Monday crisis, a large number of vertices in the network is disconnected. This results in a power-law degree distribution. However, for time epochs before and after Black Monday, the disconnected nodes recover their interactions to one another. This increases the number of connections among vertices and causes departures from the power law distribution. This phenomenon is also observed in the simulated networks using our degree update equation. This is an important result that empirically shows the simulated networks reflect the

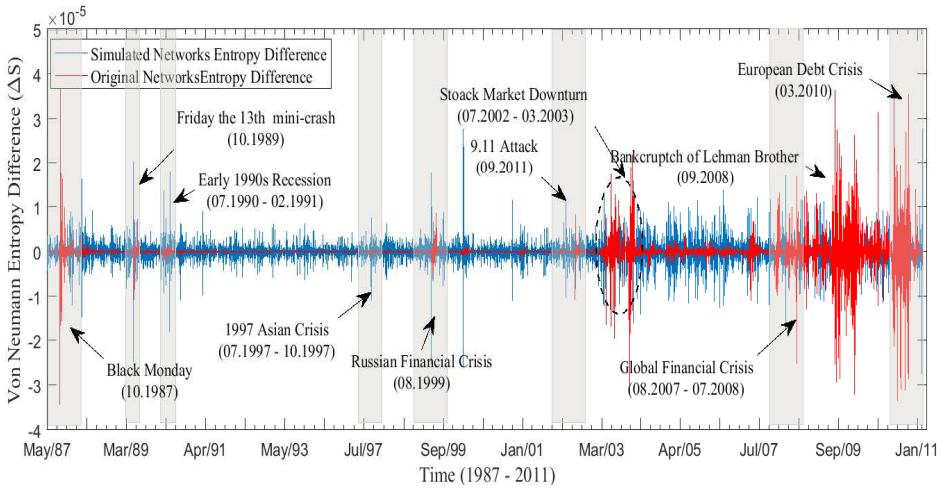


Figure 5.9: The von Neumann entropy difference in NYSE (1987-2011) for original financial networks and simulated networks. Critical financial events, i.e., Black Monday, Friday the 13th mini-crash, Early 1990s Recession, 1997 Asian Crisis, 9.11 Attacks, Downturn of 2002-2003, 2007 Financial Crisis, the Bankruptcy of Lehman Brothers and the European Debt Crisis, are associated with large entropy differences.

structural properties of the original networks from which they are generated. Moreover, our dynamic model can reproduce the topological changes that occur during the financial crisis.

In Fig.5.8, we show network visualisations corresponding to three different instants of time around the Black Monday crisis. In order to compare the simulated network structures resulting from the current model, we show the connected components (community structures) at three-time epochs. As the network approaches the crisis, the network structure changes violently, and the community structure substantially vanishes. Only a single highly connected cluster at the centre of the network persists. These features can be observed in both the simulations and original time evolution of the networks. At the crisis epoch, most stocks are disconnected, meaning that the prices evolve independently without strong correlations to the remaining stock. During the crisis, the persistent connected component exhibits a more homogeneous structure as shown in Fig.5.8. Our network prediction gives structures that more closely resemble the original network structure. After the crisis, the network preserves most of its existing community structure and begins to reconnect again. This result also agrees with findings in other literature concerning the structural organisation of financial market networks [93].

Finally, we explore the anomaly detection in dynamic networks. We validate our framework by analysing the entropy differences between simulated networks and actual stock market networks in the New York Stock Exchange (NYSE). In order to quantitatively investigate the relationship between a financial crisis and network entropy changes, we analyse a set of well-documented crisis periods. These periods are marked alongside the curve of the first order entropy difference in Fig.5.9, for all business days in our dataset.

The literature in the financial domain usually identifies the potential crashes using either a) the trading volumes [31], b) the variation of expected returns [11] or c) Spearman's rank correlation [1]. Recently, machine learning techniques, such as conditional random fields, support vector machines and artificial neural networks, have been used to identify trading patterns using various criteria on specific financial datasets [32]. Unfortunately, the complexity of these data-driven methods is generally high due to the combination of multiple techniques. By contrast, our entropy based analysis is easily effected using our dynamic model which clearly indicates the financial crises.

Directed Financial Networks

Next, we extend our study to directed graph representations of the New York Stock Exchange data. To extract directed graphs from the stock times series data, we compute the correlation with a time lag. We measure the correlation over 30-day windows separated by a time and then select the lag that results in the maximum correlation. As with undirected graphs we threshold the correlation to establish edges representing interactions between stock. We determine the directionality of the edges using the sign of the lag. All the resulting edges are unidirectional. We, therefore, explore how the time evolution follows our model for strongly directed graphs.

First, we investigate how the distribution of r_u evolves with the time. Fig.5.10 shows the distribution at three different time epochs, i.e., before, during and after Black Monday. Here, the parameter r_u reveals the relationship between in-degree and out-degree for each vertex. As shown in Fig.5.10, during the Black Monday, the cumulative distribution becomes concentrated over a small range of values around unity. This reflects the fact that

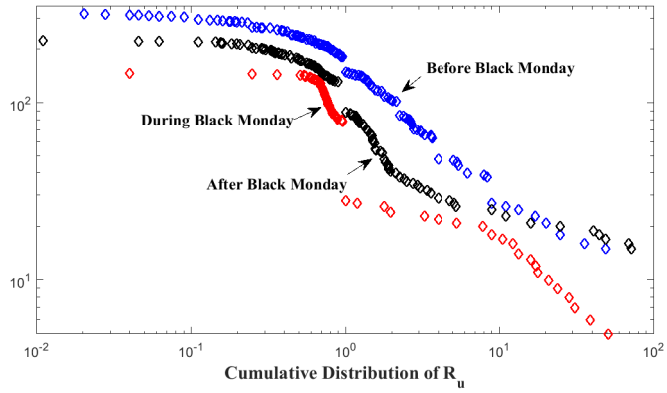


Figure 5.10: The cumulative distribution of parameter $r_u = d_u^{in}/d_u^{out}$ in directed financial networks before/during/after the Black Monday. The distribution shrinks during the Black Monday crisis.

a substantial fraction of vertices become isolated during the Black Monday, without the out-edges. The remaining connections exist with a balance between in-degree and out-degree. After Black Monday, the network structure begins to recover as the cumulative distribution widens to return to its previous shape.

From the analysis leading to Eq.(5.19), there is a linear relationship between the quantities $\frac{r_u}{\Delta r_u}$ and $\frac{d_u^{in}}{\Delta_u^{in}}$. In order to test whether this relationship holds in practice, Fig.5.12 shows scatter plots of $\frac{r_u}{\Delta r_u}$ versus $\frac{d_u^{in}}{\Delta_u^{in}}$ for epochs before, during and after the Black Monday crisis. This provides evidence that there exists a linear relationship between the fractional in-degree change and the degree ratio change. By fitting a linear regression to the sequence of scatter plots for the time series, we explore how the slope parameters of the regression line and the regression error evolve with time. Fig.5.13 shows the linear regression errors, as well as the fitted slope, during the period around Black Monday. Here we provide the regression error, for a) the flexible fitting of the slope and b) the regression for a fixed value of the slope. In the time interval around Black Monday, both the linear regression parameter and its error changes abruptly. This is because there are substantial structural differences in the network evolution. During the Black Monday, many nodes become disconnected and the connected components of vertices become small and fragmented. Only a small number of community structures remain highly inter-connected. During Black Monday itself, although the slope of the regression line is zero, the scatter about the line is relatively small.

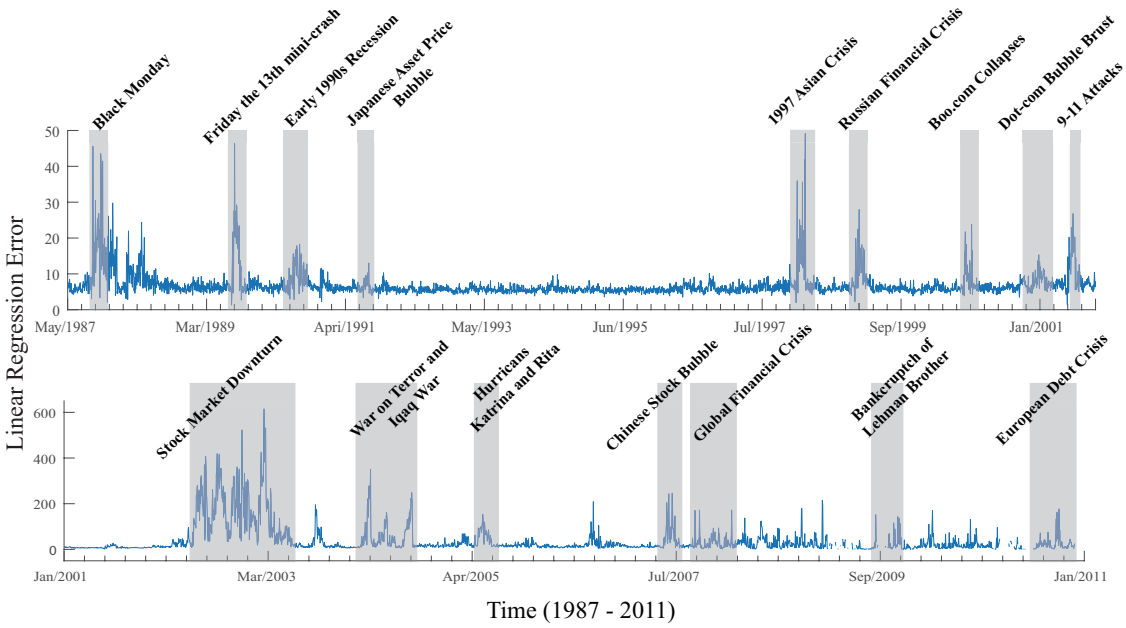


Figure 5.11: The linear regression error for the whole sequential financial data in NYSE (1987-2011). Critical financial events, i.e., Black Monday, Friday the 13th mini-crash, Early 1990s Recession, 1997 Asian Crisis, 9.11 Attacks, Downturn of 2002-2003, 2007 Financial Crisis, the Bankruptcy of Lehman Brothers and the European Debt Crisis, are associated with significant error peaks.

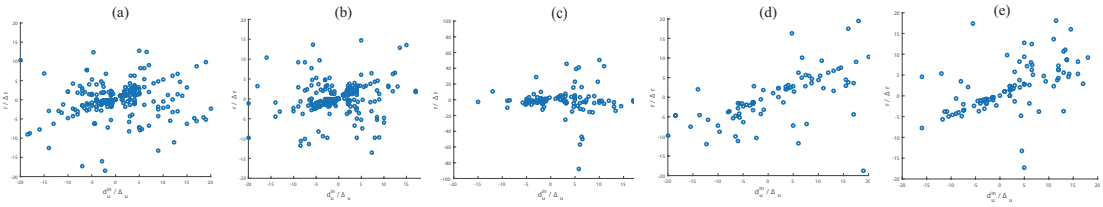


Figure 5.12: The scatter plots of d_u^{in}/Δ_u^{in} versus $r_u/\Delta r_u$ during the epoch of Black Monday (a)-(e). Before Black Monday: (a) October 1, 1987; (b) October 10, 1987. During Black Monday: (c) October 19, 1987. After Black Monday: (d) October 29, 1987; (e) November 10, 1987

Furthermore, the linear regression error sequence for the entire directed financial network time series is shown in Fig.5.11. The peaks in the regression error correspond closely to the occurrence of the financial crisis. Our analysis in the directed graph is effective and efficient to detect the abnormal structure in dynamic networks. The most striking observation is that the largest peaks of regression can be used to identify the corresponding financial crisis. This shows that the theoretical analysis of minimising the change of directed entropy is sensitive to significant structural changes in networks. The financial crises are characterised by significant entropy changes, whereas outside these critical periods remains stable.

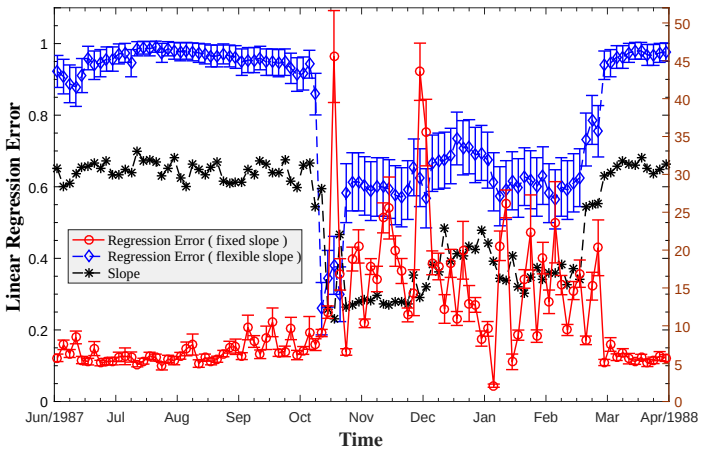


Figure 5.13: The linear regression error and standard deviation during Black Monday (June 1987 - April 1988). The blue diamond curve is the error bar with the flexible slope in the regression. Red circle line is the error bar with the fixed slope in the regression. Black star curve represents the value of the slope.

5.4 Summary

In this chapter, we explore how to model the time evolution of networks using a variational principle. We use the Euler-Lagrange equations to model the evolution of undirected and directed networks that undergo changes in structure by minimising the change in von Neumann entropy. This treatment leads to the model of how the node degree varies with time and captures the effects of degree change correlations introduced by the edge-structure of the network. In other words, because of these correlations, the variety of one degree determines the translation in connected nodes.

We conduct the experiments on a time-series of networks representing life cycle of Drosophila and the stock trades on the NYSE. Our model is capable of predicting how the degree distribution evolves with time. Moreover, it can also be used to detect abrupt changes in network structure.

Chapter 6

fMRI Network Application

The neurobiology of Alzheimer's disease (AD) has been extensively studied by applying network analysis techniques to activation patterns in fMRI images. However, the structure of the directed networks representing the activation patterns, and their differences in health and Alzheimer's people remain poorly understood. In this chapter, we aim to identify the differences in fMRI activation network structure for patients with AD, late mild cognitive impairment (LMCI) and early mild cognitive impairment (EMCI). We first use a directed graph theoretical approach combined with entropic measurements to distinguish subjects falling into these three categories and the normal healthy control (HC) group. Then we present a novel method for characterising networks using the entropy associated with bosonic particles in thermal equilibrium with a heat bath. To this end, we construct a Jensen-Shannon kernel using the Bose-Einstein entropy for a sample of networks and then apply kernel principal components analysis (kPCA) to map graphs into low dimensional feature space. We apply the resulting method to classify fMRI activation networks from patients with suspected Alzheimer's disease.

6.1 Introduction

Functional magnetic resonance imaging (fMRI) provides a sophisticated means of studying the neuropathophysiology associated with Alzheimer's disease (AD) [100]. Specifically, the blood oxygen level-dependent (BOLD) signal in fMRI indicates the activation

potential of different brain regions, and neuronal activity between the various brain regions can be determined by measuring the correlation between activation signals. The resulting network representation of region activity has proved useful in understanding the functional working of the brain [9]. Functional neuroimaging has also proved useful in understanding Alzheimer's disease (AD) via the analysis of intrinsic brain connectivity [89]. Abnormal brain function in AD is characterised by progressive impairment of episodic memory and other cognitive domains, resulting in dementia and, ultimately, death [87]. Although there is converging evidence about the identity of the affected regions in fMRI, it is not clear how this abnormality affects the functional organisation of the whole brain.

Tools from complex network analysis provide a convenient approach for understanding the functional association of different regions in the brain [89]. The approach is to characterise the topological structures present in the brain and to quantify the functional interaction between brain regions, using the mathematical study of networks and graph theory. Graph theory offers an attractive route since it provides effective tools for characterising network structures together with their intrinsic complexity. This approach has led to the design of several practical methods for characterising the global and local structure of undirected graphs [117]. Features based on the global and local measures of connectivity are widely used in functional brain analysis [66]. By comparing the structural and functional network topologies between different populations of subjects, graph theory provides meaningful and easily computable measurements to reveal connectivity abnormalities in both neurological and psychiatric disorders [87].

Furthermore, kernel-methods on graphs provide emerging and powerful set of tools to determine the class-structure of different graphs. There are many examples in the literature where graph kernels have successfully exploited topological information, and these include the heat diffusion kernel [68], the random walk kernel [65], and the shortest path kernel [27]. Once a graph kernel is to hand, it provides a convenient starting point from which machine learning techniques can be applied to learn potentially complex class-structure [10].

Unfortunately, there is relatively little literature aimed at studying structural network features using directed graphs and entropic kernel method. Although the success of existing graph kernels, one of the main challenges that remain open is to capture the variations present in different classes of graph in a probabilistic manner. The vast majority of techniques suggested by graph theory pertain to undirected rather than directed graphs. However, directed graphs are a more natural representation of brain structure, since they allow the temporal causality of activation signals for different anatomical structures in the brain. Moreover, Granger causality provides a powerful tool that can be used to investigate the direction of information flow between different brain regions [66]. When combined with machine learning algorithms, classification exhibited from directed graphs provides an effective way of detecting functional regions associated with Alzheimer's disease [66]. By explicitly defining anatomical and functional connections in a directed manner between brain regions, fMRI data may be analysed in a more detailed way and used to identify the different stages of neurodegenerative diseases [87, 66].

Recently, statistical mechanics and network entropy have been used to understand more deeply variations in network structure. One of the successes here has been to use quantum spin statistics to describe the geometries of complex networks [23]. For example, using a physical analogy based on a Bose gas, the phenomenon of Bose-Einstein condensation has been applied to study the salient aspects network structure [21]. This has been extended to understand processes such as supersymmetry in networks [20]. Although these types of analogy are useful and provide powerful tools for network analysis, they are not easily accommodated into the kernel-based approach to machine learning.

This chapter is motivated by the need to fill this important gap in the literature, and to establish effective methods for measuring the structural properties of directed graphs representing inter-regional casual networks extracted from fMRI brain data. In particular, we develop a link between statistical mechanics and kernel methods to define information theoretic kernels in terms of network entropy.

In order to characterise the functional organisation of the brain, firstly, our approach

uses as its starting point the von Neumann entropy for directed graphs. It provides a natural way of capturing the flow of information across a directed network, based on the asymmetry of edges entering and exiting its nodes. We aim to use the directed network entropy to develop graph analytical methods to measure the degree of functional connectivity in brain networks. Secondly, we explore whether the physical heat bath analogy and Bose-Einstein statistics can be used to furnish the required entropy, and implicitly the underlying probability distribution. We define information theoretic kernels in terms of network entropy to distinguish Alzheimer's disease subjects from normal healthy control population.

The heat bath analogy and Bose-Einstein statistics are proceed as follows. We commence from a physical analogy in which the normalised Laplacian plays the role as Hamiltonian (energy operator) and the normalised Laplacian eigenvalues are energy states. These states are occupied by bosonic (integer spin) particles and the resulting system is in thermodynamic equilibrium with a heat-bath, which is characterised by temperature. The bosons are indistinguishable, and each energy level can accommodate an unlimited number of particles. The effect of the heat bath is to thermalise or randomise the population of energy levels. The occupation of the energy states is therefore governed by Bose-Einstein statistics and can be characterised using an appropriate partition function. The partition function is the effective cumulative probability distribution function over the energy states in the network when the system of particles is in thermodynamic equilibrium with the heat bath. From the partition function, we can compute the entropy of the system of particles, and hence compute the Jensen-Shannon kernel. Once the kernel matrix is to hand, we use kernel principal components analysis (kPCA) [91] to embed the graphs into a low dimensional feature space where classification is performed.

We demonstrate that the resulting techniques can be used to distinguish the fMRI data from healthy controls and AD objects. The AD subjects exhibit significantly lower regional connectivity and exhibit disrupted the global functional organisation when compared to healthy controls. Moreover, the graph kernel in Bose-Einstein statistics combined with the linear discriminant analysis is applied to brain network data for two groups of

subjects with early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI). Our results indicate that in-degree and out-degree statistics for the nodes together with their associated entropy may be useful as a graph-based indicator to distinguish Alzheimer's disease subjects from normal healthy control population.

6.2 Entropy Analysis in fMRI Networks

In this section, we give the preliminaries on the directed graph representation in entropy analysis. We provide the concept of approximate von Neuman entropy for directed graphs. We then introduce the idea of edge entropy assortativity.

6.2.1 Approximate von Neumann Entropy for Directed Graphs

For an undirected graph, as shown in Chapter 5, the von Neumann entropy [81] computed from the normalised Laplacian spectrum has been proved to be effective for network characterisation. In fact, Han *et al.*[59] have shown how to approximate the calculation of von Neumann entropy in terms of simple degree statistics.

Their approximation allows the cubic complexity of computing the von Neumann entropy from the Laplacian spectrum, to be reduced to one of quadratic complexity using simple edge degree statistics in Eq.(4.4). This expression for the von Neumann entropy has been extended to characterise the structural properties of networks. It has extremal values for the cycles and fully connected graphs. Ye *et al.* [117] have extended this result to directed graphs by distinguishing between the in-degree and out-degree of nodes, giving the following expression for the entropy

$$S_d = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2} \sum_{(u,v)\in E_1} \frac{d_u^{in}}{d_v^{in} d_u^{out^2}} + \sum_{(u,v)\in E_2} \frac{1}{d_u^{out} d_v^{out}} \tag{6.1}$$

where the edge set E is partitioned into two disjoint subsets E_1 and E_2, which respectively contain the unidirectional and directional edges.

The two subsets E_1 and E_2 satisfy the conditions that $E_1 = \{(u,v)|(u,v) \in E \cap (v,u) \notin E\}$, $E_2 = \{(u,v)|(u,v) \in E \cap (v,u) \in E\}$. $E_1 \cup E_2 = E$, $E_1 \cap E_2 = \emptyset$. If most of the edges in the graph are unidirectional, i.e., $|E_1| \gg |E_2|$, then the graph is said to be strongly

directed. In this case we can ignore the entropy associated with the summation over E_2, giving the approximate entropy for strongly directed graphs as

$$S_{sd} = 1 - \frac{1}{|V|} - \frac{1}{2|V|^2} \sum_{(u,v)\in E} \frac{d_u^{in}}{d_u^{out}} \cdot \frac{1}{d_v^{in} d_u^{out}} \tag{6.2}$$

There are thus two factors determining the entropy. The first is the ratio of in-degree to out-degree for the node starts at u in the directed edge, i.e. $r_u = \frac{d_u^{in}}{d_u^{out}}$; while the second is the directed version of the edge entropy, i.e. $\frac{1}{d_u^{out} d_v^{in}}$. The former weights the contributions of the entropy associated with the directed edges exiting node u. The contributions to the entropy are thus large if the ratio r_u is small, and directed edge connects nodes with large both out-degree and in-degree.

6.3 Entropic Edge Assortativity for Directed Graphs

The assortativity is the tendency of nodes to connect to those of similar degree. This concept can be extended to directed graphs if we measure the tendency of nodes to connect to those nodes of similar in-degree and out-degree. Foster *et al.* [51] define the directed assortativity as

$$r(\alpha,\beta) = \frac{1}{|E|} \frac{\sum_{(u,v)\in E}[(d_u^{\alpha} - \bar{d_u^{\alpha}})(d_v^{\beta} - \bar{d_v^{\beta}})]}{\sigma^{\alpha}\sigma^{\beta}} \tag{6.3}$$

where $\alpha,\beta \in \{in, out\}$ is the incoming and outgoing direction for a directed edge. $\bar{d_u^{\alpha}} = |E|^{-1}\sum_{(u,v)\in E} d_u^{\alpha}$ and $\sigma^{\alpha} = \sqrt{|E|^{-1}\sum_{(u,v)\in E}(d_u^{\alpha} - \bar{d_u^{\alpha}})^2}$. The similar definitions are for $\bar{d_v^{\beta}}$ and σ^{β}.

Ye *et al.*[114] adopts a different approach to defining degree assortativity for directed graphs based on von Neumann entropy decomposition. The method is based on the observation that edges associated with high degree nodes have large entropy and preferentially attach to clusters in a graph. The entropic assortativity measurement provides a novel way to analyse the graph structure. For instance, based on the approximation for the von Neumann entropy for directed graph S_d, the coefficient of directed edge assortativity is given by [114]

$$R = \frac{\sum_{(u,v)\in E}[(S^u_{uv} - \bar{S^u_{uv}})(S^v_{uv} - \bar{S^v_{uv}})]}{\sigma^S_u \sigma^S_v} \tag{6.4}$$

where S^u_{uv} associate the entropy of all the outgoing edges from vertex u, and S^v_{uv} are all the incoming edges of vertex v.

6.4 Experiments and Evaluations

In this section, we describe the application of the above methods to the analysis of inter-regional connectivity structure for fMRI activation networks for normal and Alzheimer's patients. We first examine the differences in degree distribution for the four groups of subjects. Then we apply the entropy-based analysis to distinguish Early Mild Cognitive Impairment(EMCI) and Late Mild Cognitive Impairment (LMCI). Finally, we explore whether we can classify the subjects on the basis of similarity of the activation networks from the fMRI scans. To do this, we embed the network similarity data into a vector-space by applying kernel-PCA to the Jensen-Shannon kernel. To simplify the calculation, the Boltzmann constant is set to unity through the experiment.

6.4.1 Dataset

The fMRI data comes from the ADNI initiative [85]. Each image volume is acquired every two seconds with Blood-Oxygenation-Level-Dependent(BOLD) signals. The fMRI voxels here have been aggregated into larger regions of interest (ROIs). The different ROI's correspond to different anatomical regions of the brain and are assigned anatomical labels to distinguish them. There are 96 anatomical regions in each fMRI image. The correlation between the average time series in different ROIs represents the degree of functional connectivity between regions which are driven by neural activities [104].

A directed graph with 96 nodes is constructed for each patient based on the magnitude of the correlation and the sign of the time-lag between the time-series for different anatomical regions. To model causal interaction among ROIs, the directed graph uses the time-lagged cross-correlation coefficients for the average time series for pairs of ROIs. We detect directed edges by finding the time-lag that results in the maximum value of the

cross-correlation coefficient. The direction of the edge depends on whether the time lag is positive or negative. We then apply a threshold to the maximum values to retain directed edges with the top 40% of correlation coefficients. This yields a binary directed adjacency matrix for each subject, where the diagonal elements are set to zero. Those ROIs which have missing time series data are discarded.

Subjects fall into four categories according to their degree of disease severity. The classes are full Alzheimer's (AD), Late Mild Cognitive Impairment (LMCI), Early Mild Cognitive Impairment (EMCI) and Normal Healthy Controls (HC). The LMCI subjects are more severely affected and close to full Alzheimer's, while the EMCI subjects are closer to the healthy control group (Normal). We have fMRI data for 30 AD subjects, 34 LMCI subjects, 47 EMCI subjects, and 38 normal healthy control subjects.

6.4.2 Directed Degree Classification

We first investigate the in-degree and out-degree distribution of the data by showing a scatter plot with in-degree versus out-degree for each directed edge. In order to extract potential structural difference, the distribution of points in the scatter plot is analysed using a general linear model. Fig.6.1 and Fig.6.2 show the scatter plots of in-degree versus out-degree, comparing the first AD vs. Normal and secondly EMCI vs. LMCI respectively. The obvious difference is that normal subjects exhibit a high degree of interregional connection compared to Alzheimer's subjects. A similar effect is shown by Early and Late detection groups. Table 6.1 shows the coefficients of a linear model with 95% confidence bounds and root mean square error.

The results of fitting the linear model show that the in-degree and out-degree distributions for the nodes in the AD and LMCI groups of subjects have a greater slope than those of the Normal and Early groups. This implies that there is a greater imbalance in in-degree and out-degree in the Alzheimer's and late detection groups. In other words, the nodes in the fMRI inter-regional connectivity graphs for these two groups tend to have larger in-degree than out-degree. Moreover, the small value of RMSE in these two groups reveals that for Alzheimer's subjects the scatter about the regression lines is smallest. By contrast, for the normal and early control subjects the scatter is significantly higher. This

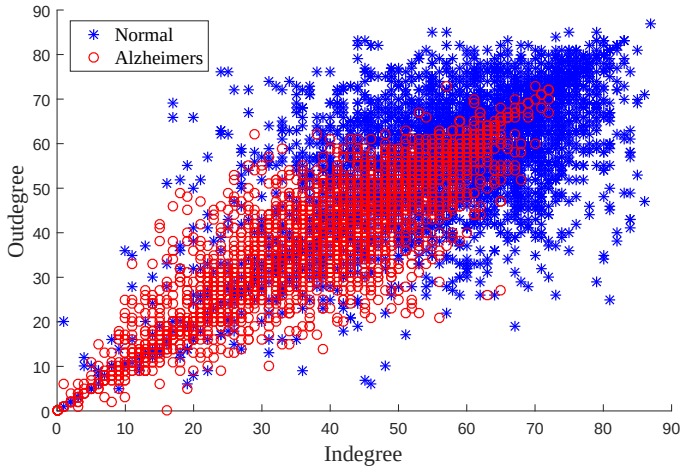


Figure 6.1: The in-degree distribution for edges in the directed graphs in Normal Healthy Control and Alzheimer's groups. The blue stars represent the edges of normal patients' graphs which occupy the high degree region with large variance. The red cycles indicate the AD patients' graphs with narrow and low degree occupation.

Table 6.1: Liner polynomial model to fit the edge in-degree/out-degree distribution

Groups	Coef (α)	CI (α)	Coef (β)	CI (β)	R^2	RMSE
AD	0.8582	[0.8406, 0.8758]	5.445	[4.719, 6.171]	0.7604	7.2444
Normal	0.6103	[0.5848, 0.6357]	22.45	[20.94, 23.96]	0.3771	11.3445
EMCI	0.7235	[0.7034, 0.7436]	14.6	[13.5, 15.7]	0.5253	10.3959
LMCI	0.9236	[0.9098, 0.9375]	2.933	[2.356, 3.509]	0.8395	6.4426

underlines the imbalance in in-degree for the subjects belonging to the diseased groups.

We can explore this asymmetry of in-degree and out-degree in more detail using Ye's entropy assortativity measure [114]. This gauges the extent to which nodes to connect to others with similar in-degree or out-degree [66]. To represent the structural difference regarding the entropy associated with degree of each node, we plot the histogram of edge entropy assortativity in Fig.6.3 and Fig.6.4. It shows the difference in entropy of the directed edges for subjects in AD vs. Normal, and EMCI vs. LMCI. By comparing the directed edges in the AD and normal groups, we conclude that the edges in the directed graphs for Alzheimer's subjects tend to have a higher value of entropy, and this reveals the structure is weakly connected with a lower average in out to in degree ratio. A similar result is shown in the EMCI and LMCI subject groups. For late Alzheimer's subjects, the shift in entropy to the right represents the weak degree connection in the nodes. This clearly reveals the loss of interregional connection for directed edges in Alzheimer's.

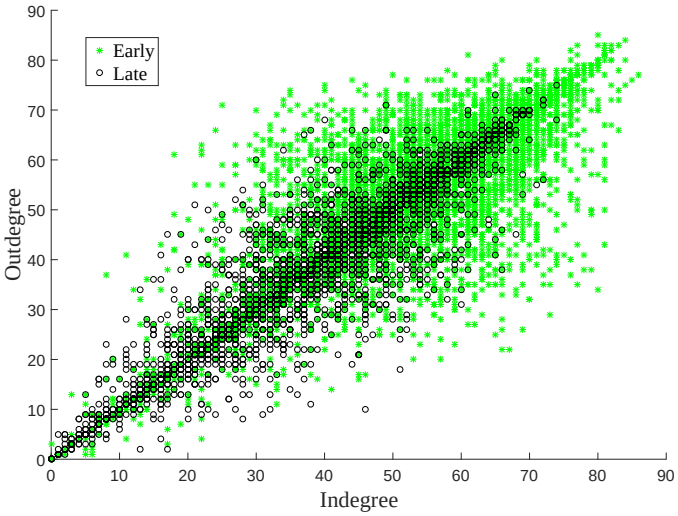


Figure 6.2: The in-degree/out-degree distribution for edges in the directed graphs in Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI).

Finally, the in-degree and out-degree of nodes are used as the features to distinguish the different group of subjects. For each edge, we construct four-dimensional feature vectors with two nodes and in and out degree measurements on each node. So the graph can be represented by these directed edges associated with four-dimensional feature vectors. We perform the linear discriminant analysis (LDA) on the Alzheimer's(AD) and Normal healthy control groups as the training process to find the decision boundary. Then the LDA model is applied to the EMCI and LMCI groups to classify patients. We compare the results and the labels to get classification accuracy.

Table 6.2 shows the classification accuracy of linear discriminant analysis (LDA). The directed graphs for the AD and Normal subjects are used as the training data to find the decision boundary. The performance of the resulting LDA classier is high with an accuracy of 87.87% when computed using 10-fold cross-validation. We randomly divide the AD and Normal subjects into 10 disjoint subsets of equal size. Remove one subset, train the LDA model using the other nine subsets. This process is repeated by removing each of the ten subsets once at a time and then average the classification accuracy. In order to evaluate the performance of classification, we provide results for sensitivity and specificity for LDA classifier. The sensitivity indicates the percentage of Alzheimer's people

who are correctly identified. It reaches 88.59% which represents the high percentage of correctly classified. In addition, the specificity shows the true negative that is the healthy people correctly identified as healthy. The accuracy of 87.10% reveals that most normal healthy people are correctly identified in the Normal group. Similarly to the LDA in AD and Normal classier, for the discrimination of subjects belonging to the EMCI and LMCI groups, we obtain a classification accuracy of 80.47%. Although this result is acceptable, the sensitivity is reduced to 75.85% indicating some percentage of patients are not correctly classified in LMCI groups.

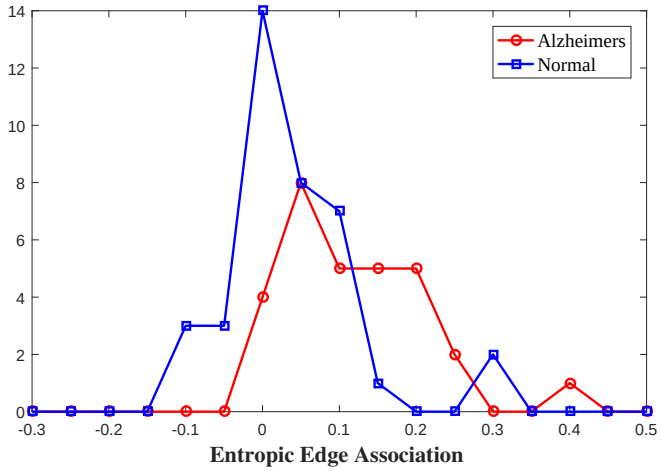


Figure 6.3: Histogram of directed edge entropy association. Normal exhibits low entropy association for each edge compared to the late and AD groups which the distributions shift to high entropy region.

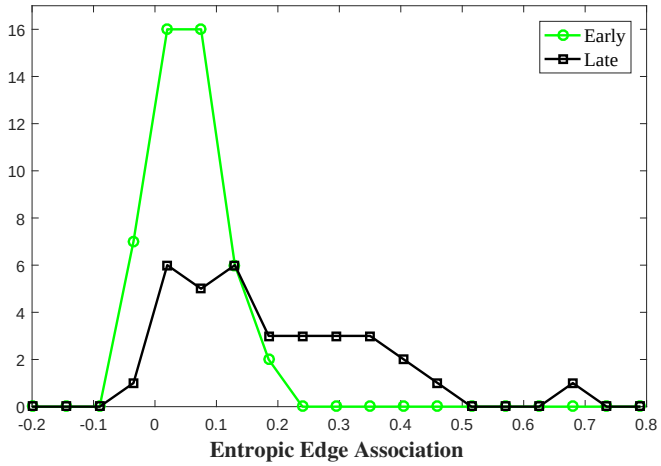


Figure 6.4: Histogram of directed edge entropy association between Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI).

Table 6.2: The classification accuracy with linear discriminant analysis (LDA) for training data (AD/Normal) and testing data (EMCI/LMCI) (in %)

LDA	Accuracy	Sensitivity	Specificity	Positive Predictivity
AD/Normal	87.87 ± 0.58	88.59	87.10	88.00
EMCI/LMCI	80.47 ± 0.41	75.85	86.18	87.14

6.4.3 Entropic Kernel Classification

Now we describe the application of the quantum statistical methods to investigate the structural dissimilarity of the fMRI activation networks, which is used to distinguish different groups of patients. We compute the Jensen-Shannon kernel matrix using the Bose-Einstein entropy and compare the performance obtained from von Neumann entropy. Given the spectra of a graph and the total number of particles, the chemical potential can be derived from Eq.(4.18), which is used to calculate the entropy. Fig.6.5 and

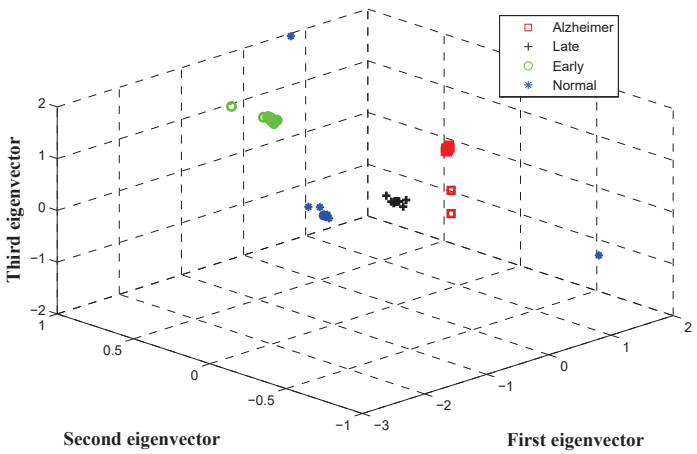


Figure 6.5: Kernel PCA performance of Jensen-Shannon Divergence in Bose-Einstein entropy. Temperature $\beta = 10$ and particle number $N = 1$.

Fig.6.6 show the results of mapping the graphs into a 3-dimensional feature space obtained by kernel principal components analysis (kPCA). We use first three eigenvectors to show the cluster of each group. The common feature is that both the Bose-Einstein and von Neumann entropies separate the four groups of subjects. In the case of Bose-Einstein statistics, the clusters are better separated than those obtained with von Neumann entropy.

To place our analysis on a more quantitative footing, we apply Fisher's linear discriminant analysis to classify graphs with the kernel features and compute the classification accuracy for the different groups of subjects. Table 6.3 summaries the results of classification accuracy obtained by Jensen-Shannon kernels computed from the two entropies. Compared to the accuracy with von Neumann entropy, that obtained with Bose-Einstein entropy exhibits a higher classification accuracy. The Bose-Einstein entropy outperforms the von Neumann entropy on three classes of data presented by a margin of about 10%. This reveals that the proposed graph kernel computed with Jensen-Shannon Divergence and Bose-Einstein entropy improve the classification performance for the fMRI data.

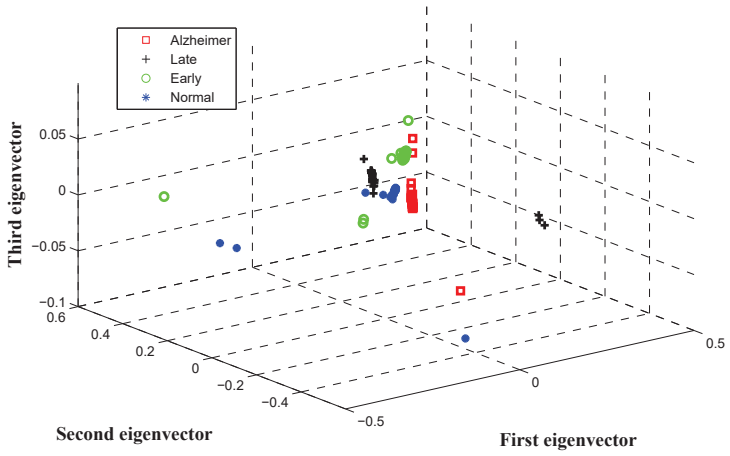


Figure 6.6: Kernel PCA performance of Jensen-Shannon Divergence in von Neumann entropy.

Table 6.3: Classification Accuracy for Entropy from Bose-Einstein Statistics and von Neumann Entropy

Classification Accuracy	Alzheimer	LMCI	EMCI	Normal
Bose-Einstein Statistics	93.33% (28/30)	100% (34/34)	89.36% (42/47)	92.11% (35/38)
von-Neumann Entropy	93.33% (28/30)	88.24% (30/34)	82.98% (39/47)	86.84% (33/38)

The main parameters of the Bose-Einstein entropy are the temperature and number of particles in the system. Here, we discuss the effects of the temperature on the energy level occupation statistics and hence upon the entropic kernel performance at low and high temperatures. We first focus on the average number of particles given the temperature β at each energy level ε_i from Eq.(4.17). In Fig.6.7, we plot the occupation number for the different normalised Laplacian energy states with different values of temperature.

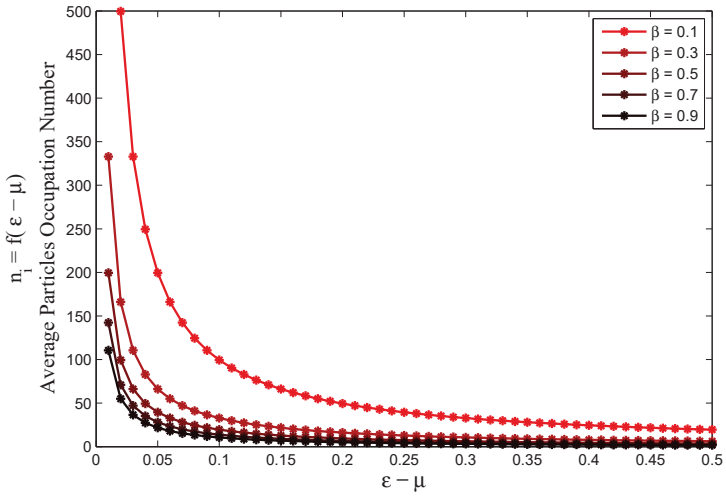


Figure 6.7: Average occupation number for energy state set different temperature for Bose-Einstein statistics.

As shown in this figure, with fixed temperature and increasing energy, the number of particles in each energy level decreases. As a result, the lower energy levels are occupied with the largest number of particles. Furthermore, as the temperature decreases, the number of particles in each energy state decreases. It should be noted that the number of particles in each state is determined by two factors, namely a) the Bose-Einstein occupation statistics, and b) the number of particles as determined by the chemical potential.

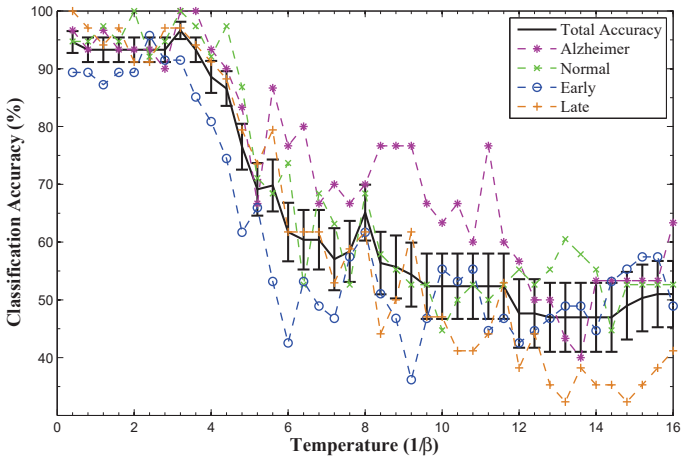


Figure 6.8: Classification accuracy changes with temperature in Jensen-Shannon Divergence with entropy from Bose-Einstein statistics.

In order to evaluate how temperature affects the performance of the Jensen-Shannon kernel, we compare its behaviour at low and high temperature. For the fMRI brain acti-

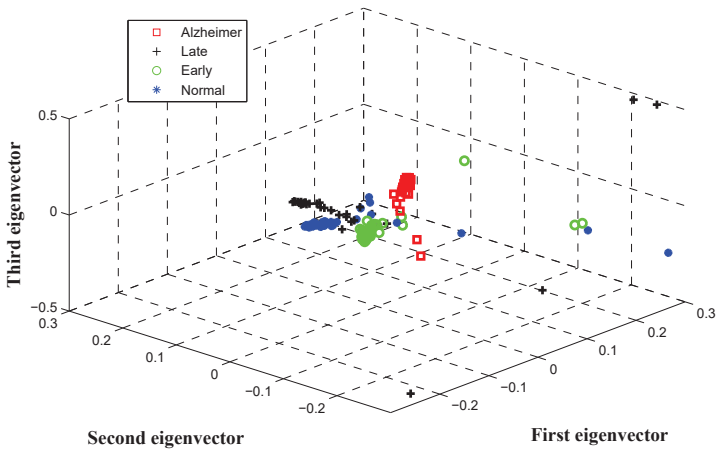


Figure 6.9: Kernel PCA performance of Jensen-Shannon Divergence with entropy from Bose-Einstein statistics at different values of temperature (β = 1, N = 5).

vation data, we set β = 1 and β = 0.1, leaving the total particle number N = 5 unchanged. Compared to the low temperature case ($\beta = 10$) in Fig.6.5, increasing temperature makes the four classes of graphs more densely clustered in feature space, shown in Fig.6.9 and Fig.6.10. This is term which reduces the performance of kernel PCA. Fig.6.8 shows the performance of classification changes with temperature. As the temperature increases, the occupation number at each energy level increases and particles become to propagate in the high energy states. This will rise up the entropy and its variance in each group, which reduces the performance of classification accuracy.

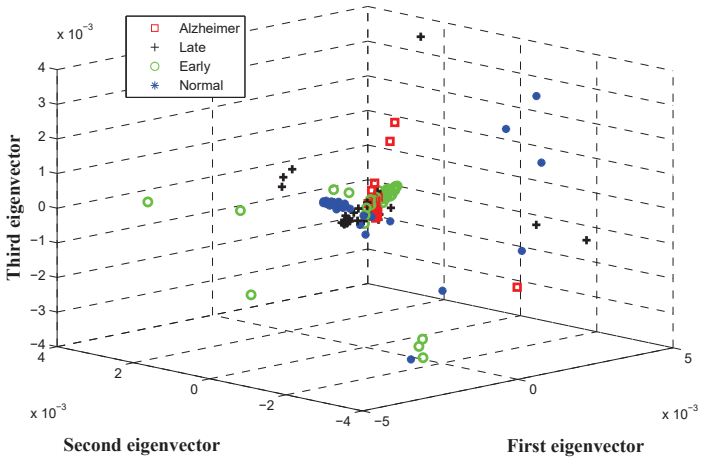


Figure 6.10: Kernel PCA performance of Jensen-Shannon Divergence with entropy from Bose-Einstein statistics at different values of temperature (β = 0.1, N = 5).

6.4.4 Identifying Salient Nodes for Disease Classification

Identifying diseased regions in the brain is also an important study in Alzheimer's analysis. Several studies have shown that in anatomical structures the corresponding ROIs are important for understanding brain disorders [100, 89]. Here, we compute the difference of out-degree and in-degree in our study and investigate the method for identification of the disease nodes in patients with Alzheimer's.

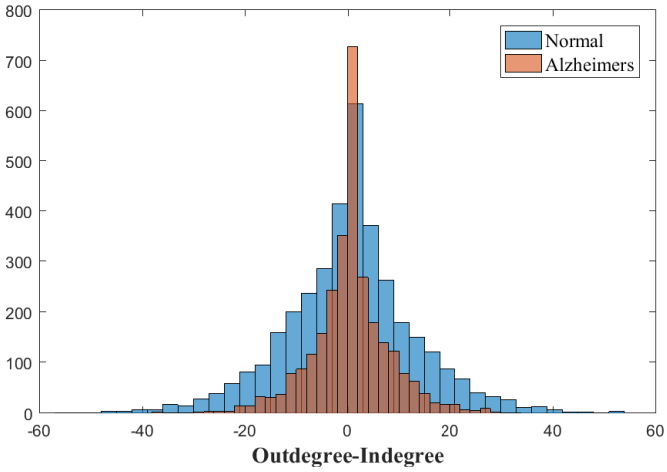


Figure 6.11: Histogram of degree difference between Alzheimer's (AD) and Normal Healthy Controls (HC) groups. The normal and early patients exhibit a wide bound range compared to the late and AD groups which the distribution narrows around zero.

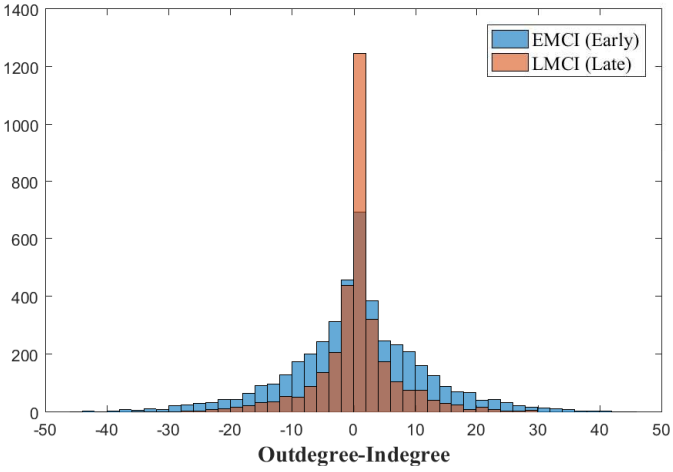


Figure 6.12: Histogram of degree difference between Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI).

We first compute the histogram of degree imbalance, i.e. out-degree minus in-degree for each node. Fig.6.11 and Fig.6.12 compare histograms obtained for AD and HC, and

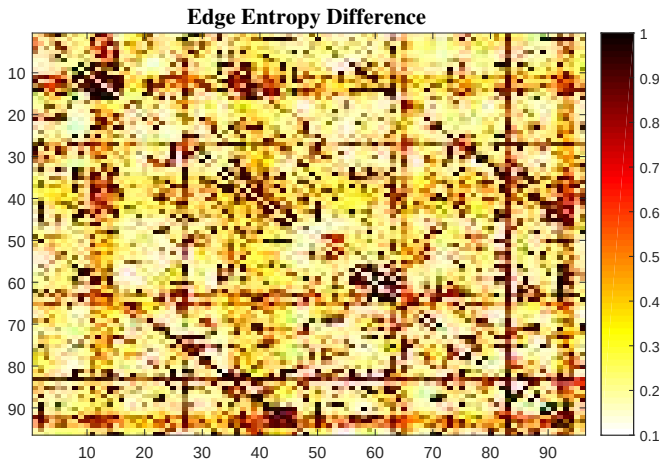


Figure 6.13: Directed edge entropy difference between Alzheimer's (AD) and Normal Healthy Controls (HC) groups. The significant changes of degree ratio in each nodes associate with the similar pattern in edge entropy plot, which illustrates the diseased area in the brain.

for EMCI and LMCI. The obvious feature is that the directed graphs for HC (normal) and EMCI (early development) groups give a much broader range of degree difference compared to that for the AD (fully developed disease) and LMCI (late development) groups. In other words, for subjects with fully developed AD, there is a loss of connection between brain regions and gives rise to a narrowing of the distribution of degree difference.

We now plot the difference in directed edge entropy between corresponding regions (nodes) in the directed graphs for the AD and HC groups. We find a similar feature pattern of the degree difference in both plots as shown in Fig.6.13 and Fig.6.14. The entropic measurements associated with degree difference in the brain areas, such as the Temporal Gyrus, Parahippocampal Gyrus, Operculum Cortex and Lingual Gyrus, suggest that subjects with AD experience loss of interconnection in their brain network during the progression of the disease.

As listed in Table 6.4, top ten anatomical regions with the largest entropy differences for subjects with full AD are right Parahippocampal Gyrus, left Inferior Temporal Gyrus, left Paracingulate Gyrus, right Temporal Fusiform Cortex, right Heschl's Gyrus, left Parietal Operculum Cortex, right Paracingulate Gyrus, left Temporal Fusiform Cortex, left Central Opercular Cortex and left Inferior Frontal Gyrus. This result is consistent with

the previous study [66, 87], which suggested that the middle temporal gyrus is an important region in AD pathology [89]. The parahippocampal gyrus has consistently been reported as being an affected region in EMCI and AD [51]. The loss of connection between these brain regions results in significant functional impairment between healthy subjects and patients with AD.

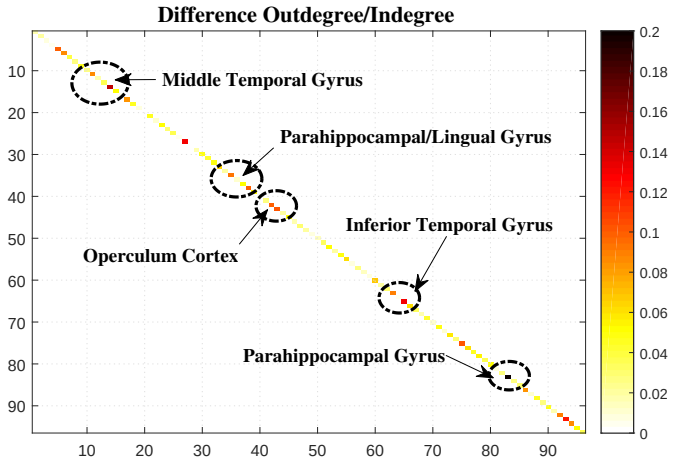


Figure 6.14: The ratio of out-degree and in-degree difference corresponding to each ROI in two groups of AD and Normal patients.

Table 6.4: Top 10 ROIs with the significant difference between groups of AD and Normal. These ROIs are extracted from the absolute value of out-degree to in-degree ratio.

Graph measure	ROI Number	Corresponding area in brain
Out-degree/In-degree Ratio Difference	83	Right Parahippocampal Gyrus
	14	Left Inferior Temporal Gyrus
	27	Left Paracingulate Gyrus
	65	Right Temporal Fusiform Cortex
	93	Right Heschl's Gyrus
	43	Left Parietal Operculum Cortex
	75	Right Paracingulate Gyrus
	38	Left Temporal Fusiform Cortex
	42	Left Central Opercular Cortex
	5	Left Inferior Frontal Gyrus

We now repeat our LDA analysis using just the salient regions listed in Table 6.4, since it is the impairment of connections to these anatomical structures that appear to determine the onset of AD. We perform LDA on the 4 vectors representing the pairs of listed anatomical regions. The classification accuracy is shown in Table 6.5. In comparison

Table 6.5: LDA classification accuracy with top 20 selected ROIs to distinguish AD/Normal and EMCI/LMCL (in %)

LDA	Accuracy	Sensitivity	Specificity	Positive Predictivity
AD/Normal	90.52 ± 0.67	91.36	89.61	91.20
EMCI/LMCI	86.20 ± 0.81	83.90	90.12	89.26

to the previous results in Table 6.3, the accuracy increases by about 3% in AD/Normal groups and 6% in the EMCI/LMCL groups. All other performances are also improved with these selected degree features.

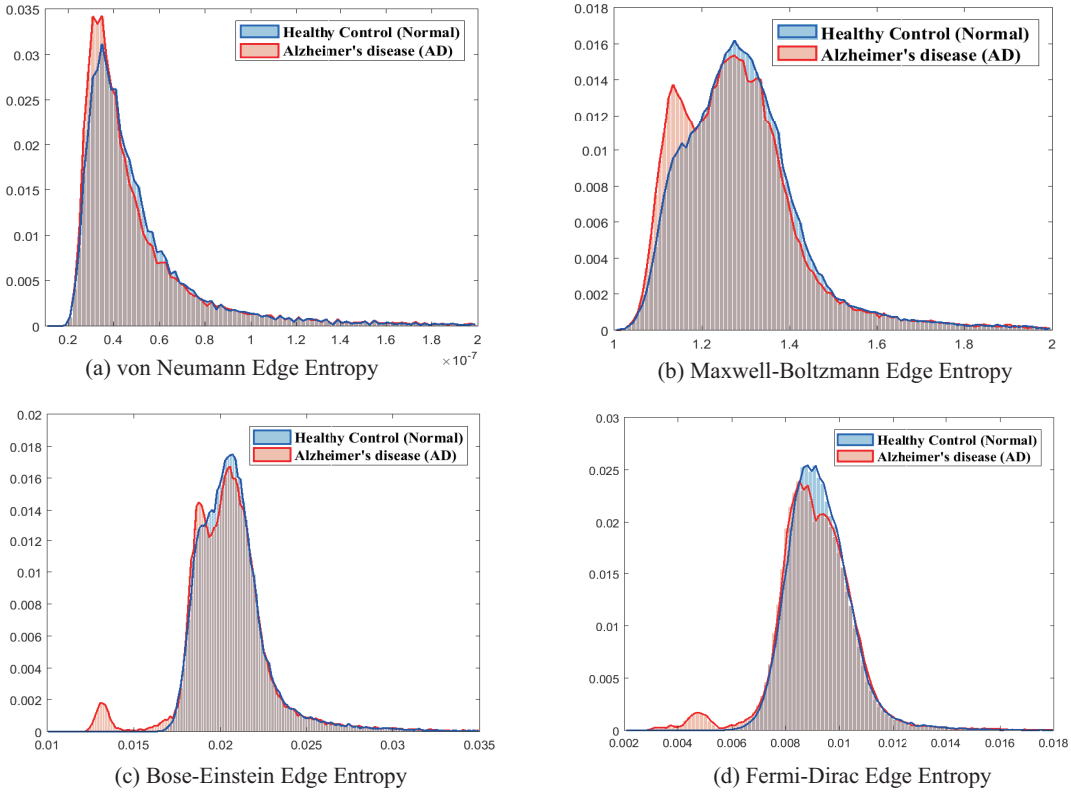


Figure 6.15: Edge entropy distribution of fMRI networks with (a) von Neumann entropy, (b) Maxwell-Boltzmann statistics, (c) Bose-Einstein statistics and (d) Fermi-Dirac statistics. Two groups of patients, Alzheimer's disease (AD) and healthy control (Normal).

Then, we apply the fMRI brain networks to further compare the edge entropy distribution with these statistical methods. Fig.6.15 shows the difference of edge entropy distribution with two groups of patients, i.e., Alzheimer's disease (AD) and healthy control (Normal). Compared to the von Neumann entropy which does not clearly represent

distributions difference between two groups, statistical methods are more robust to distinguish the detailed entropic edges. The Maxwell-Boltzmann distribution, as an example in Fig.6.15(b), illustrates that the edge entropy in Alzheimer's disease tends to present a low entropy value. This observation is more palpable in Bose-Einstein and Fermi-Dirac distributions, as shown in Fig.6.15(c) and Fig.6.15(d), with more edges tending to occupy the low entropy region. The Bose-Einstein edge entropy exhibits a more distinguishable property to separate two groups compared to the Fermi-Dirac distribution since the nonoverlapping area is much larger.

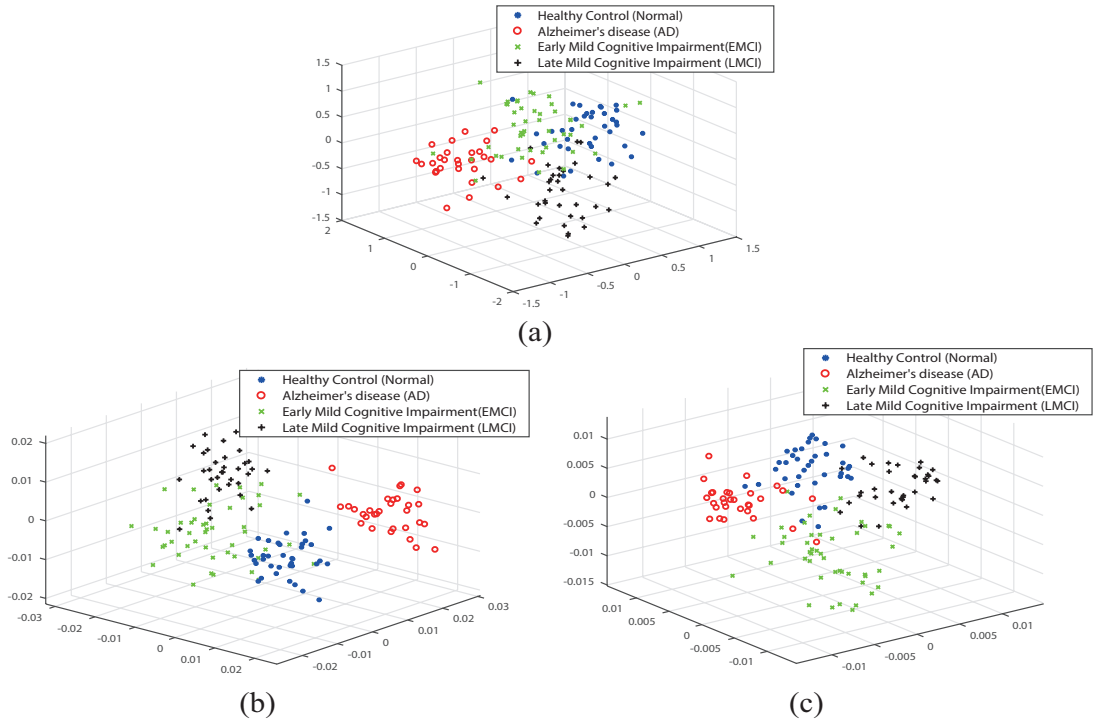


Figure 6.16: Visualisation of LDA performance with three dimensional principal components in four groups of Alzheimer's disease. (a) Maxwell-Boltzmann statistics, (b) Bose-Einstein statistics, (c) Fermi-Dirac statistics.

Finally, we select the edges with the largest 3% of entropy in the anatomical regions. This gives 278 significant edges as a feature vector. We explore whether these feature vectors can be used to classify normal healthy subjects and patients with the early Alzheimer's disease. Fig.6.16 is the visualisation of the three dimensional principal components for four groups using linear discriminant analysis (LDA). Three principal eigenvectors show the cluster of each group. The common feature is that each of the three statistical edge entropy (MB, BE, FD) can give the separation among the four

subject groups. Here, the Bose-Einstein edge entropy presents a better performance than Maxwell-Boltzmann and Fermi-Dirac cases.

6.5 Summary

In conclusion, motivated by filling the gap in the literature of analysing fMRI regional brain interaction networks using directed graphs. We commence from the recently developed simplified approximations to the von Neumann entropy of directed graphs, which are dependent on the graph size and the in-degree and out-degree statistics of vertices. In order to characterise the functional organisation of the brain, assortativity of nodes in directed graphs provides insights into the neuropathology of Alzheimer's disease. Entropic measurements associated with node degree identifies the edge connection features which offer high discrimination between subjects suffering from AD and normal subjects.

Furthermore, we show how to compute an information theoretic graph-kernel using Bose-Einstein entropy and the Jensen-Shannon divergence. We present a novel method for characterising networks using the entropy associated with bosonic particles in thermal equilibrium with a heat bath. According to this analogy, the normalised Laplacian plays the role of Hamiltonian operator, and the associated energy states are populated according to Bose-Einstein statistics. This model is subject to thermal agitation by the heat reservoir. The physics of the system can be captured by using a partition function defined over the normalised Laplacian eigenvalues. We explore whether the resulting entropy can be used to construct an effective information theoretic graph-kernel for the purposes of classifying different types of graph or network structure. To this end, we construct a Jensen-Shannon kernel using the Bose-Einstein entropy for a sample of networks and then apply kernel principal components analysis (kPCA) to map graphs into low dimensional feature space. We apply the resulting method to classify fMRI activation networks from patients with suspected Alzheimer's disease.

Chapter 7

Conclusions and Future Work

This chapter provides a summary of the main contributions in the thesis, which includes the novel methods in the network analysis with partition functions, spin statistics, edge entropy decomposition, dynamic network evolution and the application in fMRI networks. We analyse the limitations of these methods and provide the potential research in the future.

7.1 Summary of Contributions

The overall goal of this thesis is to apply statistical and entropic techniques to develop novel and effective methods for analysing the network structure and evolution, particularly paying attention to the application of fMRI data. To this end, we investigate thermodynamic characterisation of networks with different spin statistics specified by partition functions. We propose a novel framework to show how to project edge-entropy components. The detailed distribution of entropy across the edges are presented. We also develop an efficient method for simulating the dynamic network evolution using the Euler-Lagrange equation. Finally, we apply the fMRI brain networks to extend the theoretical approach to real-world datasets.

Our starting point is to develop statistical models with regard to partition functions and entropy to investigate the network structure. This is described in Chapter 3. The normalised Laplacian matrix is regarded as the Hamiltonian operator of the network, and the associated energy states are given by the eigenvalues of the normalised Laplacian.

We explore the classical and quantum statistical cases where the particle occupations correspond to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. From the relevant partition functions, we can calculate the thermodynamic entropy and energy. It provides the detailed analysis of three different partition functions deriving the entropic characterisations when compared to the extensively study of von Neumann entropy.

We conduct the experiments on both synthetic and real-world datasets to evaluate statistical properties. Entropies from three statistical models can be used to characterise changes in network structure, and distinguish different network structures. The synthetic data, generating from Erdős-Rényi random graphs, Watts-Strogatz small-world networks, Barabási-Albert scale-free networks can be distinguished by entropies. Experiments with real-world data show that the thermodynamic variables can not only be used to identify different classes of network, but can also to detect the abrupt changes in network structure.

In classical and quantum statistical models, i.e., Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics, the Fermi-Dirac entropy appears superior performance to distinguish different networks. This is because Fermi-Dirac statistics is more sensitive to the higher eigenvalues of the normalised Laplacian, which allows it to enlarge probe differences in the degree distributions for different models. Our real-world data, on the other hand, comes mainly from problems where a strong community or cluster structure exists in the networks. Thus, the Bose-Einstein model performs best for the reason of sensitivity to the eigenvalue gap.

Based on the thermodynamic entropy with spin statistics, we propose a novel framework in Chapter 4 to project out edge-entropy components so that the detailed distribution of entropy across the edges of a network can be computed. Combined the methods in Chapter 3 that Hamiltonian operator of the network associated with energy states derived from the eigenvalues of normalised Laplacian matrix, the particle occupation in the energy states result from three statistics. Then the corresponding thermodynamic entropy extends it as a tool to characterise network structures in both static and time serial data. Our results in the experiments demonstrate that the thermodynamic edge entropy is better

suited to represent the intrinsic structural properties associated to long-tailed degree distributions when compared with the von Neuman entropy. This is particularly valuable for the analysis of non-homogeneous networks with a hub structure.

The third contribution, as outlined in Chapter 5, is the development of a variational principle to investigate both undirected and directed network evolution. We apply the Euler-Lagrange equation based on the von Neumann entropy for time-varying network structure. Commencing from recent work to approximate the von Neumann entropy using simple degree statistics, the changes in entropy between different time epochs are determined by correlations in terms of the degree difference between edge connections. Our Euler-Lagrange equation minimises the change in entropy and develops a dynamic model to simulate the changes of node degree with time. We first explore the effect of network dynamics on the three widely studied complex network models, namely Erdős-Rényi random graphs, Watts-Strogatz small-world networks, and Barabási-Albert scale-free networks. Our model effectively captures both undirected and directed structural transitions in the dynamic network models. We also apply two real-world networks. One is the time sequential network representing the evolution of stock prices on the New York Stock Exchange (NYSE) from 1987 to 2011. The other is the sequences of Drosophila gene regulatory networks containing different developmental phases of the organism from embryo to adult. Our experiments show that the presented model not only provides an accurate simulation of the degree statistics in time-varying networks, but that is also captures the topological variations taking place when the structure of a network changes violently.

Finally, in order to fill the gap in the literature regarding to the analysis of fMRI regional brain interaction networks using directed graphs, in Chapter 6, we take advantages of the recently developed approximate von Neumann entropy for directed graphs, which are dependent on the graph size and the in-degree and out-degree statistics. Assortativity of vertices provides insights into the neuropathology of Alzheimer's disease to explore the functional organisation of the brain. Entropic measurements associated with node degree identifies the edge connection features which offer high discrimination between subjects suffering from the AD and normal subjects. Furthermore, we compute an

information theoretic graph-kernel using Bose-Einstein entropy and the Jensen-Shannon divergence. This method is based on the entropy associated with bosonic particles in thermal equilibrium with a heat bath. It is subjected to thermal agitation by the heat reservoir. The physics of the system can be captured by using a partition function defined over the normalised Laplacian eigenvalues. We explore whether the resulting entropy is useful to construct an effective information theoretic graph-kernel for the purposes of classifying different types of graph or network structure. To this end, we build a Jensen-Shannon kernel using the Bose-Einstein entropy for a sample of networks and then apply kernel principal components analysis (kPCA) to map graphs into low dimensional feature space. We apply the resulting method to classify fMRI activation networks from patients with suspected Alzheimer's disease.

7.2 Limitations

Although the novel methods provided in this thesis outperformed some of the state-of-the-art measures in network characterisations, there are still a number of limitations to be noticed as follows.

First, commencing with the definition of the Hamiltonian operator of the network, the associated energy states correspond to the eigenvalues of the normalised Laplacian. We specify the particle occupations correspond to Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics. This leads to an indeterminate definition of the meaning of particles on the network. In other words, it is difficult to present a clear physical meaning to the particles with regard to the structural characterisations in the network. This shortcoming clearly suppresses the utility of statistical structural applications in the network characteristics.

Another limitation is the Hamiltonian operator, which is regarded as the normalised Laplacian matrix, determines the energy of a given network. So far, there is comparatively little work to understand complex networks from a purely classical statistical mechanics point of view where the energy is not constrained. Classical thermodynamics describes systems in equilibrium. However, many networks which have emerged as a result of a

dynamical process are far from equilibrium. Real-world networks result from a combination of a growth process and some thermalisation processes. For example, the Internet grows, but at the same time, it continuously rearranges exhibiting a sort of thermalisation [2]. The introduction of a network Hamiltonian makes it disadvantage to study the evolutionary networks with thermalisation in a flexible way.

Moreover, in terms of dynamic network evolution, we apply the variational principle with Euler-Lagrange equation in the experiments. Although it effectively identifies abrupt changes and distinctive periods in time-varying financial networks, there are still some limitations with the large-scale networks evolving with inflexible vertices. Some unexpected random fluctuations may not associate with any identifiable events in the time-series. A few critical events do not give rise to unique patterns when it applies the change of approximate von Neumann entropy in variational principle.

Finally, when it comes to the application of fMRI networks, we generate the brain networks using the cross-correlation coefficients to measure the similarity between pairs of a time-serial signal. However, this method cannot adequately represent the functional structural activity in the brain. Actually, there does not exhibit a linear correlation between pairs of regions in the brain. The threshold for constructing the binary adjacency matrix can also lead to the lost of information in functional connectivity. In this case, more advantage technologies, such as mutual information and transfer entropy, should be investigated to compensate the drawbacks of cross-correlation.

7.3 Future Work

In this section, we provide the possible solutions to the limitations of this thesis, and discuss some approaches for the potential research.

First, in order to clearly explain the meaning of particles in the network system, we may introduce the network topologies with the thermodynamics of the ideal gas. By analogy with eigenvalues of graph spectrum, the one-particle energy spectrum can derive the thermodynamical properties in the scope of network characterisations. In this sense, the networks with connectivity can be viewed as the ideal gas in the study of general

physical systems. Moreover, a stochastic process can describe a gas of identical particles transporting on the networks. The transition rate for particles depends on the state of the node which corresponds to the quantum walks on the networks.

Moreover, in terms of network Hamiltonian, other graph matrices such as adjacency matrix, Laplacian matrix and singless Laplacian can be explored as the Hamiltonian operator to specify the energy spectrum. So far the equilibrium approaches have been proposed with specific partition function to study topological properties of networks, while the dynamic network evolution requires the network Hamiltonian to be more suitable to represent the non-equilibrium process. Perturbation theory may be further applied to the dynamic network described by Euler-Lagrange or Hamilton's equations. This is more interesting to derive the dynamic Hamiltonian that governs the network evolution in analogy to the physical intuition.

In addition, the entropic methods can further associate with network structure. The definition of entropy provides the concept of thermodynamics in networks, which establishes a link between microstates and macroscopic descriptions of networks. Therefore, the information of network topology can be computed from entropy to reflect the divergence between different structures. In fact, the von Neumann entropy with Jensen-Shannon divergence has been proved to be an efficient way to construct graph kernel to measure dissimilarity. This provides a new direction for the development of kernel methods. It allows us to further explore graph kernels and mutual information methods with thermodynamic entropies from different statistics. Particular information divergence from different network statistical entropy will provide a more powerful tool for characterising various structural patterns.

Finally, we acknowledge that we have explored a relatively limited quantity of real-world data. It would, for example, be interesting to see if the thermodynamic variables can be used to detect temporal anomalies and disturbances in the evolution of networks on a greater variety of data. Another interesting line of investigation would be to explore whether phase transitions can be detected with thermodynamic quantities to other network structures such as the multilayer networks and multiplex networks. We also plan

to extend the work in this thesis to the low-temperature limits for exploring the observed phenomenon of Bose-Einstein condensation in the networks.

List of Symbols

G	Graph
V	Vertex set of a graph
E	Edge set of a graph
u	Vertex Index
$\|V\|$	Number of Vertices in a graph
A	Adjacency Matrix of a graph
D	Degree Matrix of a graph
L	(Combinatorial) Laplacian matrix of a directed/undirected graph
$\tilde{L}$	Normalized Laplacian Matrix of a directed/undirected graph
$\tilde{\Lambda}$	Diagonal Eigenvalue Matrix of Normalized Laplacian
Φ	Eigenvector Matrix of Normalized Laplacian
I	Identity Matrix
η	Fitness Parameter for each node
d_u	Degree of Vertex u
d_u^{in}	In-degree of Vertex u
d_u^{out}	Out-degree of Vertex u
$\tilde{\lambda}$	Eigenvalue of Normalized Laplacian Matrix
$\|\psi\rangle$	Pure Quantum States
ρ	Density Matrix
$\hat{H}$	Hamiltonian Operator
∇^2	Kinetic Energy
$U(r,t)$	Potential Energy
N	Number of Particles
$\langle U \rangle$	Thermodynamic Average Energy
F	Helmholtz Free Energy
T	Thermodynamic Temperature
k_B	Boltzmann Constant
μ	Chemical Potential
ε	Energy State

Z	Partition Function
Z_{MB}	Partition Function in Maxwell-Boltzmann Statistics
Z_{BE}	Partition Function in Bose-Einstein Statistics
Z_{FD}	Partition Function in Fermi-Dirac Statistics
S	Thermodynamic Entropy
S_{MB}	Thermodynamic Entropy in Maxwell-Boltzmann Statistics
S_{BE}	Thermodynamic Entropy in Bose-Einstein Statistics
S_{FD}	Thermodynamic Entropy in Fermi-Dirac Statistics
S_{VN}	Von Neumann Entropy of a Graph
S_{edge}^{MB}	Edge Entropy Decomposition in Maxwell-Boltzmann Statistics
S_d	Von Neumann Entropy of a Directed Graph
S_{wd}	Von Neumann Entropy of a Weakly Directed Graph
S_{sd}	Von Neumann Entropy of a Strongly Directed Graph
$\mathcal{E}(q)$	Euler-Lagrange equation
$q(t)$	Variable in Euler-Lagrange Equation as a Function of Time
$\dot{q}(t)$	Time Derivative of $q(t)$
ΔS	Change of Approximate von Neumann Entropy
ΔS_{wd}	Change of Approximate von Neumann Entropy for Weakly Directed Graph
ΔS_{sd}	Change of Approximate von Neumann Entropy for Strongly Directed Graph
Δ_u	Degree change at vertex u
$\dot{\Delta}_u$	Time Differential to the Change of Degree at vertex u
$\mathcal{G}$	Graph Change in Euler-Lagrange Equation
d_u^t	Node Degree u at Time Epoch t
r_u	Ratio of In-degree to Out-degree for Node u
$r(\alpha, \beta)$	Directed Assortativity
R	Directed Edge Assortativity with von Neumann Entropy
D_{JS}	Jensen-Shannon Divergence
k_{JS}	Jensen-Shannon Divergence in Graph Kernel

Abbreviations

AD Alzheimer’s Disease

ADNI Alzheimer’s Disease Neuroimaging Initiative

BA Barabási-Albert Scale-free Networks

BE Bose-Einstein Statistics

BEC Bose Einstein Condensation

BOLD Blood Oxygen Level-Dependent

CI Confidence Interval

EMCI Early Mild Cognitive Impairment

ER Erdős-Rényi Random Graph

FD Fermi-Dirac Statistics

fMRI functional Magnetic Resonance Imaging

HC Healthy Control

JSD Jensen Shannon Divergence

kPCA kernel Principal Component Analysis

LDA Linear Discriminant Analysis

LMCI Late Mild Cognitive Impairment

MB Maxwell-Boltzmann Statistics

NYSE New York Stock Exchange

PCA Principal Component Analysis

PPI Protein-Protein Interaction

ROI Region of Interest

RMSE Root Mean Square Error

SD Strongly Directed

VNE von Neumann Entropy

WD Weakly Directed

WS Watts-Strogatz Small World Networks

Bibliography

[1] Alanyali, M., Moat, H. S., & Preis, T. (2013). Quantifying the relationship between financial news and the stock market. *Scientific reports*, 3, 3578.

[2] Albert, R. & Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Review Modern Physics*, 74(1), 47–97.

[3] Albert, R., Jeong, H., & Barabasi, A.-L. (2000). Error and attack tolerance of complex networks. *nature*, 406(6794), 378–382.

[4] Alstott, J., Pajevic, S., Bullmore, E., & Plenz, D. (2015). Opening bottlenecks on weighted networks by local adaptation to cascade failures. *Journal of Complex Networks*, 3, 552–565.

[5] Anand, K. & Bianconi, G. (2009). Entropy measures for networks: toward an information theory of complex topologies. *Physical Review E (Rapid communications) 045102(R)*, 80.

[6] Anand, K., Bianconi, G., & Severini, S. (2011). Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Physical Review E 036109*, 83(3).

[7] Anand, K., Krioukov, D., & Bianconi, G. (2014). Entropy distribution and condensation in random networks with a given degree distribution. *Physical Review E 062807*, 89.

[8] Andreas, L., Simonetto, A., & Leus, G. (2015). Distributed autoregressive moving average graph filters. *IEEE Signal Processing Letters*, 22.11, 1931–1935.

[9] Anwar, A. R., Hashmy, M. Y., Imran, B., Riaz, M. H., Mehdi, S. M. M., Muthalib, M., Perrey, S., Deuschl, G., Groppa, S., & Muthuraman, M. (2016). Complex network

analysis of resting-state fMRI of the brain. *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (pp. 3598–3601).

[10] Bai, L. & Hancock, E. (2012). Graph kernels from the Jensen-Shannon divergence. *Journal of Mathematical Imaging and Vision*, 47, 60–69.

[11] Bali, T. G. & Hovakimian, A. (2009). Volatility spreads and expected stock returns. *Management Science*, 55(11), 1797–1812.

[12] Barabasi, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512.

[13] Barabasi, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale free random networks. *Physics A*, 272, 173–187.

[14] Baronchelli, A., Catanzaro, M., & Pastor-Satorras, R. (2008). Bosonic reaction-diffusion processes on scale-free networks. *Physical Review E 016111*, 78.

[15] Bianconi, G. (2002). Growing cayley trees described by a Fermi distribution. *Physical Review E 036116*, 66.

[16] Bianconi, G. (2005). Emergence of weight-topology correlations in complex scale-free networks. *Europhysics Letters*, 71, 1029.

[17] Bianconi, G. (2008). Entropy of randomized network ensembles. *Europhysics Letters 28005*, 81.

[18] Bianconi, G. (2009). The entropy of network ensembles. *Physical Review E.*, 79, 036114.

[19] Bianconi, G. (2013). Statistical mechanics of multiplex networks: entropy and overlap. *Physical Review E. 062806*, 87.

[20] Bianconi, G. (2015). Supersymmetric multiplex networks described by coupled Bose and Fermi statistics. *Physical Review E*, 91.

[21] Bianconi, G. & Barabasi, A.-L. (2001a). Bose-Einstein condensation in complex networks. *Physical Review Letter*, 88, 5632.

[22] Bianconi, G. & Barabasi, A.-L. (2001b). Competition and multiscaling in evolving networks. *Europhysics Letters*, 54, 436.

[23] Bianconi, G., Rahmede, C., & Wu, Z. (2015). Complex quantum network geometries: Evolution and phase transitions. *arXiv:1503.04739v2*.

[24] Blundell, S. J. & Blundell, K. M. (2006). *Concepts in Thermal Physics*. Oxford University Press.

[25] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics reports*, 424(4), 175–308.

[26] Bogacz, L., Burda, Z., & Waclaw, B. (2006). Homogeneous complex networks. *Physica A: Statistical Mechanics and its Applications*, 366, 587–607.

[27] Borgwardt, K. M. & Kriegel, H.-P. (2005). Shortest-path kernels on graphs. In *Fifth IEEE International Conference on Data Mining* (pp. 74–81).

[28] Brunelli, I., Giusiano, G., Mancini, F. P., Sodano, P., & Trombettoni, A. (2004). Topology-induced spatial Bose-Einstein condensation for Bosons on star-shaped optical networks. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 37, S275.

[29] Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L., Zhang, N., et al. (2003). Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic acids research*, 31(9), 2443–2450.

[30] Buonsante, P., Burioni, R., Cassi, D., & Vezzani, A. (2002). Bose-Einstein condensation on inhomogeneous networks: Mesoscopic aspects versus thermodynamic limit. *Physical Review B 094207*, 66.

[31] Chesney, M., Crameri, R., & Mancini, L. (2015). Detecting abnormal trading activities in option markets. *Journal of Empirical Finance*, 33, 263–275.

[32] Choudhry, R. & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39(3), 315–318.

[33] Chung, F. (1997). Spectral graph theory. *CBMS Regional Conference Series in Mathematics*, 92.

[34] Chung, F. & Yau, S.-T. (1998). Coverings, heat kernels and spanning trees. *Journal of Combinatorics*, 6, 163–184.

[35] Consortium, I. C. G. et al. (2010). International network of cancer genome projects. *Nature*, 464(7291), 993.

[36] de Oliveira, I., FA de Moura, M. L., Andrade, J. J., & Albuquerque, E. (2010). Bose-Einstein condensation in the apollonian complex network. *Physical Review E 030104(R)*, 81.

[37] Dehmer, M. & Emmert-Streib, F. (2009). *Analysis of complex networks: from biology to linguistics*. John Wiley & Sons.

[38] Dehmer, M. & Mowshowitz, A. (2011). A history of graph entropy measures. *Information Sciences*, 181(1), 57–78.

[39] Domenico, D., Lancichinetti, A., Arenas, A., & Rosvall, M. (2015a). Structural reducibility of multilayer networks. *Nature Communications*, 6, 7864.

[40] Domenico, M. D., Nicosia, V., Arenas, A., & Latora, V. (2015b). Identifying modular flows on multilayer networks reveals highly overlapping organization in social systems. *Physical Review X 011027*, 5.

[41] Erdos, P. & Rényi, A. (1959). On random graphs. *Publ. Math. Debrecen*, 6, 290–297.

[42] Erdos, P. & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1), 17–60.

[43] Escolano, F., Hancock, E. R., & Lozano, M. A. (2012). Heat diffusion: Thermodynamic depth complexity of networks. *Physical Review E 036206*, 85, 190–198.

[44] Escolano, F., Lozano, M. A., Hancock, E. R., & Giorgi, D. (2010). What is the complexity of a network? The heat flow-thermodynamic depth approach. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, (pp. 286–295).

[45] Estrada, E. (2012). *The structure of complex networks: theory and applications*. Oxford University Press.

[46] Estrada, E. & Hatano, N. (2007). Statistical-mechanical approach to subgraph centrality in complex networks. *Chemical Physics Letters*, 439, 247–251.

[47] Estrada, E. & Hatano, N. (2008). Communicability in complex networks. *Physical Review E*, 77.

[48] Ethan, C., Benjamin, G., Emek, D., Igor, R., Ozgun, B., Nadia, A., Nikola, S., Gary,

B., & Chris., S. (2011). Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res*, 39, D695–D690.

[49] Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarium Imperialis Petropolitanae*, 8, 128–140.

[50] Ferretti, L., Mamino, M., & Bianconi, G. (2014). Condensation and topological phase transitions in a dynamical network model with rewiring of the links. *Physical Review E 042810*, 89.

[51] Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, 107(24), 10815–10820.

[52] Garlaschelli, D., Ahnert, S. E., Fink, T. M. A., & Caldarelli, G. (2008). Temperature in complex networks. *arXiv:cond-mat/0606805*.

[53] Garlaschelli, D. & Loffredo, M. I. (2006). Multispecies grand-canonical models for networks with reciprocity. *Physical Review E*, 73.

[54] Gopalapillai, I., Ivan, G., & Vijayakumar, A. (2007). On distance energy of graphs. *Match*.

[55] Günthard, H. H. & Primas, H. (1956). Zusammenhang von graphentheorie und motheorie von molekeln mit systemen konjugierter bindungen. *Helvetica Chimica Acta*, 39(6), 1645–1653.

[56] Gutfraind, A., Bradonjić, M., & Novikoff, T. (2014). Modelling the neighbour aid phenomenon for installing costly complex networks. *Journal of Complex Networks*, 3(2), 249–263.

[57] Gutman, I. (2001). The energy of a graph: old and new results. *Algebraic combinatorics and applications*, (pp. 196–211).

[58] Gutman, I. & Zhou, B. (2006). Laplacian energy of a graph. *Linear Algebra and its applications*, 414(1), 29–37.

[59] Han, L., Escolano, F., Hancock, E., & Wilson, R. (2012). Graph characterizations from von neumann entropy. *Pattern Recognition Letters*, 33, 1958–1967.

[60] Han, L., Wilson, R. C., & Hancock, E. R. (2015). Generative graph prototypes from

information theory. *IEEE transactions on pattern analysis and machine intelligence*, 37(10), 2013–2027.

[61] Hernández, J. M., Li, Z., & Mieghem, P. V. (2014). Weighted betweenness and algebraic connectivity. *Journal of Complex Networks*, 2, 272–287.

[62] Hofree, M., Shen, J. P., Carter, H., Gross, A., & Ideker, T. (2013). Network-based stratification of tumor mutations. *Nature Methods*, 10, 1108–1115.

[63] Javarone, M. A. & Armano, G. (2013). Quantum-classical transitions in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 04, P04019.

[64] Jooyandeh, M., Kiani, D., & Mirzakhah, M. (2009). Incidence energy of a graph. *Match*, 62(3), 561.

[65] Kashima, H., Tsuda, K., & Inokuchi, A. (2003). Marginalized kernels between labeled graphs. In *20th international conference on machine learning (ICML-03)* (pp. 321–328).

[66] Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behavioural Brain Research*, 322(Part B), 339 – 350.

[67] K.Huang (1987). *Statistical Mechanics*. New York: Wiley.

[68] Kondor, R. I. & Lafferty, J. (2002). Diffusion kernels on graphs and other discrete input spaces. *The Nineteenth International Conference on Machine Learning*, 2, 315–322.

[69] Krioukov, D., Papadopoulos, F., amd Amin Vahdat, M. K., & Boguna, M. (2010). Hyperbolic geometry of complex networks. *Physical Review E 036106*, 82.

[70] Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowledge Discovery from Data*, 1(1), 1–40.

[71] Lucas, L., Bartolo, L., Fernando, B., Jordi, L., Juan, C. N., & Author, A. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105.13, 4972–4975.

[72] Martin, T., Zhang, X., & Newman, M. (2014). Localization and centrality in networks. *Physical Review E 052808*, 90.

[73] Martins, A., Smith, N., Xing, E., Aguiar, P., & Figueiredo, M. (2009). Nonextensive theoretic kernels on measures. *Journal of Machine Learning Research*, 10, 935–975.

[74] Mikulecky (2001). Network thermodynamics and complexity: a transition to relational systems theory. *Computers & Chemistry*, 25, 369.

[75] Newman, M. & Watts, D. (1999a). Renormalization group analysis of the small-world network model. *Physics Letter A*, 263, 341.

[76] Newman, M. & Watts, D. (1999b). Scaling and percolation in the small-world network model. *Physical Review E*, 60, 7332.

[77] Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), 036104.

[78] Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E 056110*, 68.

[79] Ostilli, M. & Bianconi, G. (2015). Statistical mechanics of random geometric graphs: Geometry-induced first order phase transition. *Physical Review E 042136*, 91.

[80] Park, J. & Newman, M. (2004). Statistical mechanics of networks. *Physical Review E 066117*, 70(6).

[81] Passerini, F. & Severini, S. (2008). Quantifying complexity in networks: The von neumann entropy. *Inthernational Journal of Agent Technologies and Systems*, (pp. 58–67).

[82] Penrose, O. (1991). Bose-Einstein condensation in an exactly soluble system of interacting particles. *Journal of Statistical Physics*, 63, 761–781.

[83] Peron, T. & Rodrigues, F. (2011). Collective behavior in financial markets. *European Physical Letter*, 96(48004).

[84] Perseguers, S., Lewenstein, M., Acin, A., & Cirac, J. (2009). Quantum complex networks. *Nature Physics*, 6, 539 – 543.

[85] Petersen, R. C., Aisen, P., Beckett, L., Donohue, M., Gamst, A., Harvey, D., Jack, C., Jagust, W., Shaw, L., Toga, A., et al. (2010). Alzheimer's disease neuroimaging

initiative (ADNI) clinical characterization. *Neurology*, 74(3), 201–209.

[86] Rief, F. (1965). *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill Science, New York.

[87] Rombouts, S. A., Barkhof, F., Goekoop, R., Stam, C. J., & Scheltens, P. (2005). Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. *Human brain mapping*, 26(4), 231–239.

[88] Rovelli, C. & Vidotto, F. (2010). Single particle in quantum gravity and braunstein-ghosh-severini entropy of a spin network. *Physical Review D*, 81, 044038.

[89] Rubinov, M. & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3), 1059–1069.

[90] Safar, M. H., Sorkhoh, I. Y., Farahat, H. M., & Mahdi, K. A. (2011). On maximizing the entropy of complex networks. *Procedia Computer Science*, 5, 480–488.

[91] Schölkopf, B., Smola, A., & Müller, K.-R. (1997). Kernel principal component analysis. In *International Conference on Artificial Neural Networks* (pp. 583–588).: Springer.

[92] Shen, Y., Zhu, D., & Liu, W. (2004). Fermi-Dirac statistics of complex networks. *Chinese Phys. Lett.*, 22, 1281.

[93] Silva, F., Comin, C., Peron, T., Rodrigues, F., Ye, C., Wilson, R., Hancock, E., & Costai, L. (2015). Modular dynamics of financial market networks. *Physics and Society*, arXiv:1501.05040.

[94] Silva, F. N., Comin, C. H., Peron, T. K. D., Rodrigues, F. A., Ye, C., Wilson, R. C., Hancock, E. R., & Costa, L. d. F. (2016). Concentric network symmetry. *Information Sciences*, 333, 61–80.

[95] So, W., Robbiano, M., de Abreu, N. M. M., & Gutman, I. (2010). Applications of a theorem by ky fan in the theory of graph energy. *Linear Algebra and its Applications*, 432(9), 2163–2169.

[96] Song, L., Kolar, M., & Xing, E. P. (2009). Keller: estimating time-varying interactions between genes. *Bioinformatics*, 25(12), 128–136.

[97] Strauss, D. (1986). On a general class of models for interaction. *SIAM Review*,

28(4), 513–527.

[98] Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825), 268–276.

[99] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., et al. (2014). String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(D1), D447–D452.

[100] Van Den Heuvel, M. P. & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European neuropsychopharmacology*, 20(8), 519–534.

[101] Vidal, E., Lima, R., & Lyra, M. L. (2011). Bose-Einstein condensation in the infinitely ramified star and wheel graphs. *Physical Review E 061137*, 83.

[102] Viswanath, B., Mislove, A., Cha, M., & Gummadi, K. P. (2009). On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks* (pp. 37–42).: ACM.

[103] Waclaw, B. (2013). Statistical mechanics of complex networks. *arXiv:0704.3702*.

[104] Wang, J., Wilson, R., & Hancock, E. (2016a). Network entropy analysis using the maxwell-boltzmann partition function. *The 23rd International Conference on Pattern Recognition(ICPR)*, (pp. 1–6).

[105] Wang, J., Wilson, R. C., & Hancock, E. R. (2016b). fMRI activation network analysis using Bose-Einstein entropy. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, (pp. 218–228).

[106] Wang, J., Wilson, R. C., & Hancock, E. R. (2016c). Thermodynamic network analysis with quantum spin statistics. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, (pp. 153–162).

[107] Wang, J., Wilson, R. C., & Hancock, E. R. (2017a). Minimising entropy changes in dynamic network evolution. *International Workshop on Graph-Based Representations in Pattern Recognition*, (pp. 255–265).

[108] Wang, J., Wilson, R. C., & Hancock, E. R. (2017b). Spin statistics, partition functions and network entropy. *Journal of Complex Networks*, 5(6), 858–883.

[109] Wang, X. & Chen, G. (2003). Complex networks: Small-world, scale-free and beyond. *IEEE Cirsuits and System Magazine*, 3, 6–20.

[110] Watts, D. & Strogatz, S. (1998). Collective dynamics of 'small world' networks. *Nature*, 393, 440–442.

[111] White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode caenorhabditis elegans: the mind of a worm. *Phil. Trans. R. Soc. Lond*, 314, 1–340.

[112] Wolstenholme, R. J. & Walden, A. T. (2015). An efficient approach to graphical modeling of time series. *IEEE Transactions on Signal Processing*, 63, 3266–3276.

[113] Wu, Z., Menichetti, G., Rahmede, C., & Bianconi, G. (2015). Emergent complex network geometry. *Nature Scientific Reports*, 5, 10073.

[114] Ye, C., Comin, C. H., Peron, T. K. D., Silva, F. N., Rodrigues, F. A., Costa, L. d. F., Torsello, A., & Hancock, E. R. (2015a). Thermodynamic characterization of networks using graph polynomials. *Physical Review E*, 92(3), 032810.

[115] Ye, C., Torsello, A., Wilson, R. C., & Hancock, E. R. (2015b). Thermodynamics of time evolving networks. *International Workshop on Graph-Based Representations in Pattern Recognition*, (pp. 315–324).

[116] Ye, C., Wilson, R., & Hancock, E. (2016). Correlation network evolution using mean reversion autoregression. *S+SSPR 2016, Lecture Notes in Computer Science*, 10029, 163–173.

[117] Ye, C., Wilson, R. C., Comin, C. H., da F. Costa, L., & Hancock, E. R. (2014). Approximate von neumann entropy for directed graphs. *Physical Review E*, 89.

[118] Zhang, X., Martin, T., & Newman, M. (2015). Identification of core-periphery structure in networks. *Physical Review E 032803*, 91.

[119] Zuev, K., Papadopoulos, F., & Krioukov, D. (2016). Hamiltonian dynamics of preferential attachment. *Journal of Physics A: Mathematical and Theoretical*, 49(10), 105001.