# Quantum Processes
# for Structural Analysis

Tesi di dottorato di Giorgia Minello
Matr. 797636

October, 2018

Author's Web Page: `http://www.dais.unive.it/~minello`


Author's e-mail:  giorgia.minello@unive.it


Author's address:

Dipartimento di Scienze Ambientali, Informatica e Statistica
Università Ca' Foscari Venezia
Via Torino, 155
30172 Venezia Mestre – Italia
tel. +39 041 2348465
fax. +39 041 2348419
web: `http://www.dais.unive.it`

# Abstract

Many real systems can be modelled as networks, being characterized by a set of items and links between them. Systems taking the form of networks, also called graphs, appear in a wide range of scenarios, varying from biological to technological domains. Illustrative examples abound and include neural networks, protein-protein interactions, metabolic reaction networks, social networks, coauthorship and citation relations, road maps, financial market stock correlations and the World Wide Web.

In the last decade network theory has proven to be a very useful instrument to model the structure of systems, albeit not sufficient to cover all issues in the scope of structural analysis. For this reason it has arisen the need of drawing on ideas from fields such as physics which actually helped in gaining new insight for a relevant class of problems.

In this thesis, we address matters encountering in graph structural analysis by exploiting new approaches based on quantum processes and the von Neumann entropy. In particular, we focus on the characterization aspects of graphs concerning structural properties, as well as on processes underlying network evolution. We commence by investigating spectral generative models for learning structural representations. Then we move on to quantum models, specifically quantum walks, and the von Neumann entropy characterization. Finally, we introduce a novel thermodynamic method to model time evolving networks.

# Contents

# List of Figures

# Published Papers

[1] GASPARETTO, A., MINELLO, G., AND TORSELLO, A. Non-parametric spectral model for shape retrieval. In *3D Vision (3DV), 2015 International Conference on* (2015), IEEE, pp. 344–352.

[2] LOCKHART, J., MINELLO, G., ROSSI, L., SEVERINI, S., AND TORSELLO, A. Edge centrality via the holevo quantity. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2016), Springer, pp. 143–152.

[3] MINELLO, G., ROSSI, L., AND TORSELLO, A. On the Von Neumann Entropy of Graphs. *ArXiv e-prints* (Sept. 2018).

[4] MINELLO, G., TORSELLO, A., AND HANCOCK, E. R. Quantum thermodynamics of time evolving networks. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (2016), IEEE, pp. 1536–1541.

[5] MINELLO, G., TORSELLO, A., AND HANCOCK, E. R. Thermodynamic characterization of temporal networks. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2016), Springer, pp. 49–59.

# 1

# Introduction

Networks are well-known frameworks that use the node-link paradigm to depict data relationships. Graphs, their mathematical counterpart, embody an optimal interface for dealing with objects and phenomena coming from a wide range of scenarios, regardless of each domain characteristics. Over the last decades, we have thus witnessed the growing adoption of the graph-based representation by researchers of diverse areas of study, to manage, organize and solve a vast variety of problems. Especially in the computer science field, graphs have been extendedly exploited, thanks to their representation power and flexibility. This kind of representation certainly carries considerable advantages when it comes to characterizing objects, at the same time it involves difficulties in terms of treatability though. In this sense, the main challenge lies in devising solutions to extract knowledge from graph structures, in the same way as standard analysis techniques capture salient properties from vectorial data.

In this thesis, we address different matters encountering in the structural analysis scope, by investigating a broad array of novel approaches for facing them. We focus both on the characterization aspects of graphs concerning structural properties and on processes underlying networks evolution. Chapter 3 is intended as an overture to the structural analysis study. Here we present a spectral generative model for learning structural representations. We then move on to the heart of our research starting from Chapter 4 which covers a wider topic, the structural analysis based on quantum concepts. More to the point, we explore the use of the von Neumann entropy as a means to characterize diverse structural aspects of a graph. In addition, through quantum walks, we probe graph structures with a view to measuring the similarity between graphs. Finally, in Chapter 5 we shift our attention to evolving structures, by introducing a novel thermodynamic method to model time-varying networks.

# Learning Graph Structure

Graph models are pervasive for describing information across many scientific areas where complex information is used. However, when abstracted in this way, structured data becomes more difficult to manipulate than pattern vectors. Despite considerable progress in characterizing data using graph structures, dealing with graphs is still a long-standing problem. Basically, graphs are non-Euclidean and this feature implies the impossibility of using well-established standard methods for analysing, comparing and matching graphs. This aspect affects many pattern recognition approaches including those about learning over graphs, such as generative models. Generative approaches are well known in the domain of statistical pattern recognition for the many attractive properties, first and foremost the ability to naturally establish explicit correspondences between model components and features. They also provide solutions for handling missing values (on the contrary of most of the discriminative approaches) and are robust to occlusion, allowing unsupervised learning in clutter [87]. Generative models typically describe the probability distribution of patterns in a vector space. However, without a straightforward vectorial representation, even the simplest statistical quantities, such as mean and variance, result difficult to define for a group of graphs. Two main reasons are behind the limited methodology available for learning structural representations. First, the lack of explicit correspondences between nodes of different graphs and second the variability of the graphs structure, *e.g.*, variations in edge connectivity, node composition, node attributes, *etc.* [197]. Therefore, even supposing to correctly encode the graph in a vectorial fashion, we may end up with vectors of different lengths.

In Chapter 3 we consider the problem of learning over graphs, by defining a novel model of structural representation based on the spectral decomposition of the Laplacian matrix which lifts the one-to-one node correspondence assumption. In particular, we follow White and Wilson [195] in defining separate models for eigenvalues and eigenvectors, but cast the eigenvector model in terms of observation over an implicit density function over the spectral embedding space, and we learn the model through non-parametric density estimation. The eigenvalue model, on the other hand, is assumed to be log-normal, due to consideration similar to [16]. In particular we show how the method is also suitable for non-rigid $3D$ shape retrieval application.

# Graph Structure Analysis

The task of analyzing graph structures has been an integral objective for a number of years in many research fields, ranging from chemistry to computer science. The leading challenge probably lies in capturing the various quantitative or qualitative aspects of a graph, through some sort of summarising descriptors, so that to handle problems involving structured data properly and possibly with standard techniques. Many basic parameters allow us to have a partial insight of the graph topology, such as the degree distribution, the clustering coefficient, the average path length, the diameter, the betweenness centrality, to

mention only a few. Later, by focusing on some parameters/measures, it has been understood there exists and it is possible to identify certain rather universal features shared by structures within a given class. One of the most famous examples concerns the power-law degree distribution for scale-free networks [22]. However, with the recent surge of interest in complex networks in different areas, many other methods have been proposed to characterize the structural properties of graphs, with attention towards single elements but at the same time oriented to some macroscopic level definition. In other words, instead of focusing on several aspects in detail, an alternative could be trying to obtain a concurrent representation of all its characteristics. Efforts to characterise complexity in networks in this direction have gathered pace since the dawn of the Internet. Quantum information and quantum computation met somehow the need of finding a new way to investigate structured data. Although they come with a plethora of inner issues, at the same time provide research topics and ideas suitable for tackling graphs [66]. For instance, quantum walks have been shown to be very sensitive to topological features, such as symmetries, and changes as well. This property has meant that many researchers started to further investigate, then making the walks play a fundamental role in the analysis of the structure and dynamical processes in networks.

Chapter 4 shows how the quantum approach to structural analysis can be an effective choice. Here, we present an edge centrality index rooted in quantum information as well as we propose an investigation over the von Neumann entropy in terms of structural patterns and variants formulations. Beside, we introduce a work in progress about graph similarity through quantum walks and divergences, which is an expansion of a previous study by Rossi *et al.* [154].

## Graph Evolution

As a framework for representation, networks help us to simplify and view the world as objects of pure structure whose properties are fixed in time, while the phenomena they are supposed to model are actually often changing [146]. Nevertheless, initially, the majority of effort in network theory just focused around a static view of the structure, thus aiming at developing only ways to describe networks [65] rather than studying the dynamics of these structures. However, recently, evolving network analysis has started to attract more attention. Examples of networks giving rise to structures that change with time are not so uncommon: neural networks, financial networks, citation networks, communications networks, to name but a few. Essentially, the main focus has shifted to how the structure affects the network performance and how the network evolves with time to respond to such needs [65]. In particular, the analysis of the processes underlying network evolution has acquired an increasingly crucial role in network science even if capturing the large-scale properties of a time evolving graph has proven to be a challenging problem. From the perspectives of both static and dynamic network analysis, many researchers have recognized that drawing on ideas from fields such as physics, may yield promising results. Especially, the use of analogies based on quantum thermodynamics, relating the

behaviour of microscopic particles to the macroscopic properties of a system [90, 95, 124], has proved singularly successful.

Chapter 5 introduces a novel method to characterize the behaviour of the evolving systems based on a thermodynamic framework for graphs. This framework aims at relating the major structural changes in time evolving networks to thermodynamic phase transitions. We start from a recent quantum-mechanical characterization of the structure of a network to derive the network entropy. Then we adopt a Schrödinger picture of the dynamics of the network to obtain a measure of energy exchange through the estimation of a hidden time-varying Hamiltonian from the data.

# 2

# Literature Review

In this Chapter we attempt to provide, as far as possible, a thorough overview of the current literature covering the principal subjects of this thesis. Note, nevertheless, that the Sections of this Chapter are not intended to give also a comprehensive survey of the state-of-art techniques in the corresponding topic. However, wherever necessary, references to more accurate reviews will be indicated or provided within the concerning Chapter. In Section 2.2 we introduce some basic notions of Spectral Theory while Section 2.3 will concern the generative models but for the structural representations niche. Section 2.5 is devoted to an introduction to quantum computation and in particular focused on quantum walks and von Neumann entropy. In fact, the latter will be the key concept of the entire Chapter 4. Finally, in Section 2.6 we present the main studies undertaken in graph analysis exploiting thermodynamic analogies and concepts, as a preamble for Chapter 5.

## 2.1   Graph Representation

Roughly speaking, a graph is a collection of points and lines, which connect some subset of points. Mathematically speaking, these points are called vertices while the lines linking the vertices are known as edges. Despite this very poor introduction, graphs are actually an excellent and powerful tool to model and characterize countless objects and phenomena. Generally nodes may represent an object, agents, financial entities or even a brain region. Instead edges tend to indicate causal relations or interactions between objects, *i.e.*, nodes. These links can be directed or undirected. In the last decades, the graph-based representation, which also includes sequences and trees as special cases, has increasingly become utilized by an extensive number of disciplines. This success stems from the fact the classical vectorial manner, amply exploited in the past, at a certain point has resulted no longer able to satisfy the increasing need of modelling relationships and dynamics exhaustively. In other words, those systems best defined in terms of interconnection or topological structure [93, 97, 100, 164] have found in the graph based representation the ultimate solution for the characterization problem. Data that can be regarded as structural graphs belongs to both nature and man-made sphere. Examples include RNA structures [59], natural language texts [121], biological sequences, semi structured data such as HTML and XML [7], phylogenetic trees, to name but a few.
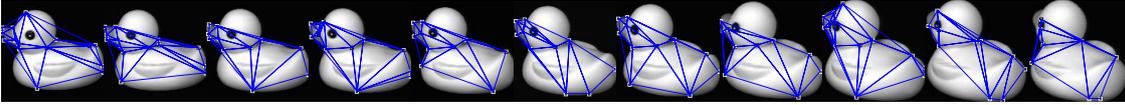
Figure 2.1: The Delaunay graph representation - a set of rotated objects. During the rotation, the set of interest points which are subjected to the Delaunay triangulation varies and as a result the structure of the graphs changes.

As regards the computer science field, in particular the areas of computer vision and pattern recognition, the graph-based representation has covered a particular role. In the literature there are many examples of application to different tasks, such as shape matching [164], object recognition [202], image segmentation [162], shape representation [17], just to cite the best-known ones.

The ability in concisely capturing the relational arrangement of object primitives, in a way that can be invariant to changes in object viewpoint [175], is probably the most alluring trait of this type of representation. In spite of the plenty possibilities of employment, graph-based representation presents some drawbacks in terms of handling data though. Especially, dealing with data not structured by default requires a preprocessing step where usually features are extracted from the object under analysis. For instance, converting an image into a graph involves first the feature points extraction and then their arrangement, in a structural way. That means the extracted features will become the nodes, while the edge structure will be placed so that to preserve the general layout. One of the first methods facing this problem was invented by Boris Delaunay [56] (1934), who attempted to turn feature points into graphs via the so called the Delaunay triangulation. An example is depicted in Fig. 2.1. The main property of this method is that it maximizes the minimum angle of all the angles of the triangles in the triangulation. Other famous graph representation methods include the Gabriel graph [75] and the K-nearest neighbour graphs [61]. In the Gabriel graph, two points are linked by an edge when there are no other points in the circle whose diameter is the line segment jointing the two points. The nearest neighbor graph representation, as its name obviously suggests, connects each node to its K-nearest neighbour nodes.

## 2.2 Spectral Graph Theory

A graph structure is defined by its adjacency relation, usually expressed in the form of a matrix. This standard encoding modality first allows handling any size structures but mainly may yield valuable information about graphs. Spectral graph theory studies the relationships between the eigenvalues of diverse matrices, *e.g.*, adjacency matrix, associated to a graph and the graph itself. It belongs to the branch of mathematics called algebraic graph theory, which investigates graphs by means of matrix algebraic properties [39]. Similarly, the leading goals in graph theory is to deduce the structure and key

properties of a graph from its graph spectrum.

Formally speaking, a graph is denoted by $G = (V, E)$, where $V$ is the set of nodes and $E \subseteq V \times V$ is the set of edges. $A = (a_{ij})$ defines its adjacency matrix where

$$a_{ij} = \begin{cases} 1, & v_i \sim v_j, \\ 0, & \text{otherwise.} \end{cases}$$

The degree $d$ of a node is the number of edges incident to the node and it can be represented through the degree matrix $D = (d_{ij})$ which is a diagonal matrix with $d_{ii} = \sum_j a_{ij}$. Starting from these two matrix representations of a graph, we derive the *Laplacian* matrix, which is defined as the difference between the degree matrix $D$ and the adjacency matrix $A$ 2.1. It can be also interpreted as a combinatorial analogue of the discrete Laplace-Beltrami operator [99].

$$L = D - A \tag{2.1}$$

Similarly the normalized Laplacian matrix $\tilde{L}$ is defined as

$$\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

The Laplacian has been extensively used to provide spectral representations of structures [113]. The spectral representation of the graph can be obtained from the Laplacian via eigendecomposition. Given a Laplacian $L$, its decomposition is

$$L = \Phi \Lambda \Phi^T$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_{|V|})$ is the matrix whose diagonal contains the ordered eigenvalues, while $\Phi = (\phi_1 | \phi_2 | ... | \phi_{|V|})$ is the matrix whose columns are the ordered eigenvectors. The Laplacian matrix is symmetric and positive semi-definite; its properties are quite interesting and convenient where appropriate. For instance, its eigenvalues are positive and the smallest eigenvalue is zero with multiplicity equal to the number of connected components in $G$.

$$0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{|V|}$$

Moreover, the zero eigenvalue is associated with all-ones vector. On the other hand, the eigenvector associated with the second smallest eigenvalues is the Fiedler vector.

Interestingly, the eigendecomposition is unique up to a permutation of the nodes of the graph, a change of sign of the eigenvectors, or a change of basis over the eigenspaces associated with a single eigenvalue, *i.e.*, the following properties hold:

$$\begin{aligned} L &\simeq PLP^T = P\Phi\Lambda(P\Phi)^T \\ L &= \Phi\Lambda\Phi^T = \Phi S\Lambda S\Phi^T \\ L &= \Phi\Lambda\Phi^T = \Phi B_\lambda \Lambda B_\lambda \Phi^T \end{aligned}$$

where $\simeq$ indicates isomorphism of the underlying graphs, $P$ is a permutation matrix, $S$ is a diagonal matrix with diagonal entries equal to $\pm 1$, and $B_\lambda$ is a block-diagonal matrix with

the block diagonal corresponding to the eigenvalues equal to $\lambda$ in $\Lambda$ and is orthogonal while all the remaining diagonal blocks are equal to the identity matrices.

There are many reasons, from several points of view, behind the intense study of the Laplacian spectra. For instance, the eigenvalues are a remarkable natural invariant which acts perfectly under operations such as disjoint union and Cartesian product. Being often related to other invariants, such a relationship can usually provide good approximations to less tractable computations. Moreover, there are a number of aspects in which the eigenvalues of a graph operate like the spectrum of a compact Riemannian manifold [48, 49].

From a practical perspective, there have been a number of useful applications of spectral approaches, for instance in image segmentation and graph matching. Remarkable the work of Shi and Malik [162], where they handled an image segmentation problem as a graph partitioning problem by introducing the normalized cut criterion to segment graphs. Through optimizing of such a formulation, they found out that the second smallest eigenvalue of the affinity matrix can be used to bipartition the graph. Another quite famous algorithm is the pioneer work for matching problem of Umeyama [183], aimed at finding a permutation matrix very close to the optimum one by means the outer product of the left eigenvector matrices of two graphs. However, the literature presents a myriad of other work in this area of interest. For instance Robles-Kelly and Hancock [147, 148] have used both a standard spectral method [147] and a more sophisticated one based on ideas from graph seriation [148] to covert graphs to strings to compares structures. Wilson and Zhu [198] by means of the Euclidean distance between spectra of graphs, measured the distance of graphs in classification and clustering tasks. Note that although graph spectra provide a convenient way of characterising graphs, their interpretation in terms of the underlying structure can prove to be opaque.

## 2.3   Graph Generative Models

Generative and discriminative approaches are two common but at the same time diverse techniques employed for classification tasks. Broadly speaking, generative approaches learn a model of the joint probability $p(x, y)$ of the input $x$ and the label $y$, make prediction by applying Bayes rules to obtain $p(x|y)$, and then pick the most likely label $y$. Instead, discriminative approaches model the posterior $p(x|y)$ directly from the data, or learn a direct map from input $x$ to the class labels.

Although there are many compelling reasons in favour of the use of discriminative classifiers, for instance superior performance, as far as graph domain is concerned, generative approaches seem to be more attractive. Probably, the crucial feature lies in their ability to capture the modes of structural variation of graphs, a process which has proved to be elusive and cumbersome in the past. However, even other properties are supportive to the generative choice. Indeed, having explicit hypotheses on the location and structure of the signal in the training set, they can also learn in presence of clutter, occlusion or missing data.

We may observe different types of structural variations, which can involve either the nodes or edges, or even both, other than attributes over them. In the contest of variations over node or edge attributes, Bayesian networks or general relational models [40, 73, 74], attempted to deal with the problem by associating random variables with the nodes of the structure and using a structural learning process to infer the stochastic dependency between these variables. In particular, Wong *et al.* [199], have introduced a first order random graphs for structural-based classification. In their model, edges and nodes are linked to discrete random variables, taking values over the attribute domain of the graphs. Unfortunately, discrete densities make difficult the learning process and in turn the practical application. The first order random graphs idea has been extended by Bagdanov and Worring in [18], where they opted for a continuous Gaussian distributions to model the densities of random variables in the graphs. The choice allowed to overcome many computational issues and speedup the classification process.

As regards the modelling of variations in node and edge composition, for the restricted class of trees, Torsello and Hancock [180] define a superstructure called tree-union that captures the relations and observation probabilities of all nodes of all the trees in the training set. The structure is obtained by merging the corresponding nodes and is critically dependent on the order in which trees are merged.

Still focusing on the class of trees Torsello and Hancock [181] investigated a method for estimating a structural archetype for a set of trees. They embedded the trees in a vector-space with fixed length, where the dimensions correspond to principal modes of structural variation. Then, from a set of trees a super-tree is built, from which each tree may be obtained by edit operations of node and edge removal. Although the initial promising results, the approach seems to fail in case of large databases containing several shape classes.

In [173] by Todorovic and Ahuja, we see how the tree-union approach can be also applied to object recognition based on a hierarchical segmentation of image patches and lifted the order dependence by repeating the merger procedure several times and picking the best model according to an entropic measure. While these approaches do capture the structural variation present in the data, the model structure and model parameter are tightly coupled, which forces the learning process to be approximated through a series of merges, and all the observed nodes must be explicitly represented in the model, which then must specify in the same way proper structural variations and random noise.

Similar to previous work, Torsello and Dowe [177] addressed the problem of learning archetypal structural models from examples. They proposed an approach where the initial assumption is simplified though. Indeed each observation over a node or edge is independent from all others and the existence of a vertex or an edge is modelled as a Bernoulli trial. They learned the model using an EM-like approach where they alternated the estimation of the node correspondences with the estimation of the model parameters, obtained within a Minimum Message Length framework. However the approach can be only used to choose from different learned structures since it has no way to change the complexity while learning the model.

In the scope of learning edge-connectivity, Xiao and Hancock [201] constructed a

generative model for graphs in a vectorial format by performing the Young-Householder decomposition on the heat kernel matrix, and describing the distribution of the coordinates of the nodes using a Gaussian distribution. The covariance matrix of the embedded node coordinates resulted to capture the variations in graph structure instead the reconstruction step from these representations turned out to be difficult. White and Wilson [195] have explored the use of a probabilistic model over the spectral decomposition of the graphs to produce a generative model of their structure. Similarly to [201], Luo *et al.* [117] build a generative model over a set of graphs by stacking the elements of the adjacency matrices of each sample graph. Once at hand vectorial data, statistical techniques are used. In order to align graphs they applied the algorithm of Luo and Hancock [116]. They also pad smaller graphs with dummy vertices to overcome graph size differences. Finally, variations in the resulting vector space is analysed by using principle component analysis (PCA), allowing to sample new graphs using their model. In more recent work [176, 182] Torsello and co-workers proposed a generalization for graphs which allowed to decouple structure and model parameters and used a stochastic process to marginalize the set of correspondences. The process however still requires a (stochastic) one-to-one relationship between model and observed nodes and could only deal with size differences in the graphs by explicitly adding a isotropic noise model for the nodes.

## 2.4   Graph Structure Analysis

In principle, structure and function of a network should be intrinsically correlated. However, learning about the function performed by a network from its structure or understanding if changes in structure reflect changes in function are both a demanding task. This is particularly hard above all when we do not properly know how to exploit the network characterization at hand with a view to analyse these relationships. Then, the quantification of the intrinsic complexity of graphs and networks has attracted significant attention due to its practical importance, not only in network analysis but also in other areas such as pattern recognition and control theory. Actually, when it comes to complexity over graphs, there emerges a dichotomous distinction. On the one hand randomness complexity aims to quantify the degree of randomness or disorganisation of a combinatorial structure. This approach addresses to characterise an observed graph structure probabilistically and compute its associated Shannon entropy. One of the earliest contributions was Mowshowitz complexity index [127] which links the complexity to the entropy of the distribution of orbits in the structure's symmetries. Recently, the quantum entropy [35, 140], applied to the graphs domain by mapping discrete Laplacians and quantum states, has been introduced as measure able to distinguish different structures. For instance it is maximal for random graphs, minimal for complete ones and takes on intermediate values for star graphs. The main drawback of randomness complexity is that it does not capture properly the correlations between vertices. On the other hand statistical complexity aims to overcome this problem by measuring regularities beyond randomness and by characterising a combinatorial structure using statistical features such as node degree statistics, edge density or

the Laplacian spectrum. For example Song *et al.* [168] have recently explored the use of the Laplacian energy [81], *i.e.*, the sum of absolute differences between the eigenvalues and the average vertex degree, as a complexity measure for graphs. In a regular graph, the Laplacian energy is equal to the energy of the graph. The Laplacian energy is also low for graphs associated to polygons. A relevant recent supplement to the graph-spectral literature is Estrada's network heterogeneity index [64]. The index gauges differences in degree for all pairs of connected vertices. It also can be expressed as a quadratic form of the Laplacian and it is equal to zero for regular graphs and random ones and equal to one for the star graph. Another measure taking on low values for both random and ordered systems is the thermodynamic depth, which relies on the causality of heat flow. In [63], Escolano and Hancock have devised a characterisation of network complexity using the phase transitions in thermodynamic depth. The obtained measure called flow complexity is being then utilized to analyse graphs and networks in terms of the thermodynamic depth.

It is clear that nowadays there is an increasingly voiced need to understand the behaviour of the system as a whole. Along this route, understanding the topology of the interactions between the components is unavoidable. Although physics has provided an arsenal of successful means to analysing the behaviour of a system as a whole, other paradigms anyhow attempted to reach the same goal. In particular, in the past, the definition of a model fitting the behaviour of real-world networks has prompted many studies towards that direction. In other words, scientists attempted to understand if the organizing principles displayed by networks were somehow encoded in their topology as well. Especially, they tried to provide at least a simpler representation. Simpler representations of possibly very complex structures have many advantages. They can gain insight into how networks form and how they grow, may allow mathematical derivation of certain properties, can serve to explain certain properties observed in real networks, can predict new properties or outcomes for networks that do not even exist and finally can serve as benchmarks for evaluating real networks.

The first work about large scale networks was proposed by Erdös and Rényi [62], who designed a model for networks with no apparent design principles, the so called random graphs. Even if it is very unlikely that real networks actually form like this, yet, the model can predict a surprising number of interesting properties. There are two definitions of a random network. In the $G(N, L)$ model $N$ nodes are connected with $L$ randomly placed links. Erdös and Rényi used this definition in their string of papers on random networks. On the other hand, in the $G(N, p)$ model, each pair of $N$ nodes is connected with probability $p$, a model introduced by Gilbert [79]. Hence, the $G(N, p)$ model fixes the probability $p$ that two nodes are connected whereas the $G(N, L)$ model fixes the total number of links $L$. This fundamental model then has lead to further investigate other kind of networks behaviour, resulting in the formulation of the scale–free model [22] and the small–world model [194].

A common feature of real world networks is the presence of hubs, which are the most striking difference between a random and a scale-free network. Hubs are those few nodes that are highly connected to other nodes in the network. The presence of hubs will give

the degree distribution a long tail, indicating the presence of nodes with a much higher degree than most other nodes. The recognition that growth and preferential attachment coexist in real networks has inspired a minimal model called the Barabási-Albert model (also known as the BA model or the scale-free model or preferential attachment) which can generate scale-free networks [22].

Watts and Strogatz in [194] propose a model for networks between order and chaos. The small world model describes the fact that in many networks there is a relatively short path between any two nodes but high clustering coefficients. The most popular manifestation of *small worlds* is the *six degrees of separation* concept, uncovered by the social psychologist Stanley Milgram (1967), who concluded that between most pairs of people in the United States there was a path of about six acquaintances. The small world model seems to characterize most complex networks. Even if the model is intriguing, it is not an index of a particular organizing principle. In fact, Erdös and Rényi have proved that random graphs are small worlds as well since the average distance between any two nodes in a random graph scales as the logarithm of the number of nodes.

A fundamental task in structural analysis is characterizing a graph locally, that means measuring the importance of any single element. To this end, various centrality indices have been introduced in the literature [65], since different aspects and different subjects imply different measures. Common examples usually refer to graph vertex, even if also edge centralities have been studied recently. The most intuitive notion of centrality focuses on degrees, the *degree centrality*. It can be interpreted as a measure of immediate influence, a measure of popularity or alternatively, as the risk of a node being infected in a disease spreading scenario.

The degree centrality [71] is defined as the number of edges incident upon a node. Given a graph $G$ with $n$ nodes and adjacency matrix $A$, the degree centrality of $u$ is

$$DC(u) = \sum_{v=1}^{n} A(u, v) \tag{2.2}$$

The closeness centrality links the importance of a vertex to its proximity to the remaining vertices of the graph. Then, if the farness of a vertex is the sum of the lengths of the geodesics to every other vertex, its reciprocal is the *closeness centrality* [72]. Thus a node is considered important if it is relatively close to all other nodes. More precisely, the closeness centrality is defined as the as the inverse of the sum of the distance of a vertex to all other nodes of the graph,

$$CC(u) = \frac{n-1}{s(u)} \tag{2.3}$$

where $s(u)$ denotes the sum of the distances from $u$ to all the other nodes of the graph, i.e.,

$$s(u) = \sum_{v=1}^{n} d(u, v) \tag{2.4}$$

where $d(u, v)$ denotes the distance between $u$ and $v$.

The betweenness centrality [134] tells us how influential a node is in communicating between node pairs.In practice it measures the number of times a shortest path between nodes $v_1$ and $v_2$ travels through a node $u$ whose centrality is being measured. It is a measure of the extent to which a vertex lies on the paths between others, where the path may be either the shortest path or a random walk between the nodes.

If $sp(v_1, v_2)$ denotes the number of shortest paths from node $v_1$ to node $v_2$, and $sp(v_1, u, v_2)$ denotes the number of shortest paths from $v_1$ to $v_2$ that go through $u$, the betweenness centrality of $u$ is

$$BC(u) = \sum_{v_1=1}^{n} \sum_{v_2=1}^{n} \frac{sp(v_1, u, v_2)}{sp(v_1, v_2)} \qquad (2.5)$$

Note that this definition assumes that the communication takes place along the shortest path between two vertices. A number of measures have been introduced to take into account alternative scenarios in which the information flows through different paths [65, 71, 72, 134].

The flow centrality is also known as random-walk betweenness centrality [134]. The flow centrality measures the importance of an edge proportionally to the expected number of times a random walk passes through the edge $e$ when going from $u$ to $v$. The name *flow centrality* comes from the analogy between the graph and an electrical circuit where one is interested in measuring the amount of current that flows through $e$, averaged over all source nodes $u$ and target nodes $v$.

## 2.5 Quantum Computation

Quantum mechanics is a mathematical framework or set of rules for the construction of physical theories. Ever since its discovery, quantum mechanics has been applied with enormous success to everything under and inside the Sun [135]. For instance, modern technology has exploited quantum effects both to our benefit, *e.g.*, laser technology and detriment, *e.g.*, the atomic bomb [102]. However, not only did mere applied sciences take advantage of this framework but also the pure ones. Indeed, the interest of obtaining complete control over single quantum systems allowed for the development of quantum computation and quantum information. In both cases, the basic objective is studying the information processing tasks through quantum mechanical systems. For example, quantum information theory has attracted intense interest because of the possibilities of employing its formalizations to devise tools of surprising power [135]. This fruitful axis of research has uncovered many new opportunities, such as quantum cryptography [92] or quantum computation. In the last years, quantum computing has especially merited consideration because of the potential for considerable speed-ups over classical algorithms. In that regard, it suffices to recall Peter Shor [163], in 1994, who discovered a quantum algorithm to factor numbers efficiently, namely polynomially with the length of the number to be factored. Needless to say, a wave of activity across a wide array of fields of study

has been unleashed. However, in the next Sections we will concentrate on the pertinent literature on quantum walks and the von Neumann entropy.

### 2.5.1 Quantum Walks

In the last decades, formulations based on quantum walks to model novel quantum algorithms [12, 45, 102, 157] on graph structures have become part of the mainstream.

A classical random walk on a graph corresponds to a quantum walk on a graph with the difference the latter is addressed by the quantum mechanical viewpoint. A continuous time classical random walk on a graph models a Markovian process over the nodes of the graph [23], where a step only occurs if two vertices are connected by an edge. Similarly, a quantum random walk is defined as a dynamical process over the vertices of a graph.

However a quantum random walk presents a substantial distinction in terms of characterization. Basically, the behaviour of quantum walks is governed by unitary matrices [203] rather than stochastic ones. In other words, the state-vector is composed of complex numbers rather than probabilities. Such an aspect cannot be overlooked, because it actually typifies the quantum walk. The unitary evolution also renders the quantum walk reversible, implying so the walks are is non-ergodic and do not have limiting distribution.

Nonetheless, many algorithms are based on classical random walks, so it is natural to ask which properties make the quantum walks more appealing.

In general, in the quantum case, interference and entanglement lead to interesting phenomena, not present in classical case. For instance, quantum walks have been shown to outperform their classical counterparts in a wide range of tasks, leading to speedups over classical computation [69, 159]. That seems to be due to the constructive and destructive interference effects of quantum processes. As far as the graph topology is concerned, their use has been broadly investigated over a wide variety of graphs [103, 128], such as the infinite line, cycles, regular lattices, star graphs and complete graphs. In addition, quantum walks have been shown to better discriminate between different structures than their classical counterparts [21, 154].

Quantum walks on graphs have both continuous time and discrete time and variants. The first type was initially investigated by Farhi and Gutmann [69] They devised an algorithm based on a continuous-time Hamiltonian to move through decision trees, building a correspondence between nodes and quantum states. More precisely, they took two co-joined $n$-level binary trees linked by their leaves and ran a quantum walk starting from the root of the first tree. They were among the first to prove the potential for speed-up over classical algorithms. Indeed, in their work, they have shown how the quantum walk can hit the root of the second tree exponentially faster than a similarly defined classical random walk, without relying on the quantum Fourier transform.

In the last decades, more and more researchers have started to exploit the continuous-time quantum walk as well as the discrete-time quantum walk approach. For the discrete-time quantum walk, an exponential speed-up for the hitting time was observed by Kempe [102]. Instead Aharonov *et al.* provided a lower bound on the possible speed up by quantum walks for general graphs [8] Similar results to [102] were presented by Childs *et*

*al.* [46] but for continuous-time quantum walks on graphs. By following Childs *et al.* [46] Kempe also took in consideration the walk on the *n*-dimensional hypercube, showing that the hitting time from one node to the one opposite is polynomial in *n* [101] whereas Ahmadi *et al.* [9] studied continuous time quantum walks in an *N*-cycle. Also Moore and Russell [126] worked on a *n*-dimensional hypercube, but they analyzed both the continuous-time and a discrete-time quantum walk.

Recently, various authors have investigated the evolution of continuous-time quantum walks over graphs along with the notion of the quantum Jensen-Shannon divergence [120], to diverse aims. In [155], Rossi *et al.* studied the connection between quantum walks and graph symmetries, showing that the quantum divergence between the evolution of two quantum walks with certain initial states is maximum when the graph presents symmetries. In the field of structural classification, Torsello *et al.* [178] generalized a structural kernel based on the Jensen Shannon divergence between quantum walks over structures, by introducing a novel alignment step. Instead Rossi *et al.* in [154], proposed a quantum inspired kernel for unattributed graphs where they gauge the similarity between graphs through continuous-time quantum walks. More specifically, they compute the divergence between two suitably defined quantum states. We extend upon the results of Rossi *et al.* [154] to include different Hamiltonians, kernel signature and directed graphs. More details about this work in Section 4.3.

## 2.5.2 Von Neumann entropy

Entropy is a key concept of quantum information theory and it is mainly known as a measure of uncertainty. For instance, the Shannon entropy measures the uncertainty associated with a classical probability distribution. However, if we generalize the definition of the Shannon entropy, by replacing probability distributions with density operators, we have to refer to a different type of entropy, the so called *quantum entropy*. That step is possible because quantum states are described in a similar fashion to probability distributions. John von Neumann defined the von Neumann entropy in his principal work in quantum mechanics, as far back as 1955 [130]. Nevertheless, the von Neumann entropy (or quantum entropy) is still an outstanding tool in quantum information theory [135,136]. In this section we attempt to review the leading approaches which have involved the von Neumann entropy as concerns the graph-based representation universe.

The ability to measure the complexity of a structure plays a central role in structural analysis. Intuitively, the complexity of a graph should capture the level of organization of its structural features, *e.g.*, the scaling behaviour of its degree distribution. To this end, a number of entropic complexity measures have been proposed in the past years [13, 14, 30, 55, 63, 140].

The von Neumann entropy of a network was introduced by Braunstein et al. [34] and then analyzed further in a number of later work [13, 14, 50, 51, 58, 140, 166]. The intuition behind this measure is that of associating graphs to density matrices and measuring the complexity of the graphs in terms of the von Neumman entropy of the corresponding density matrices. This in turn is based on the mapping between quantum states and the

combinatorial graph Laplacian proposed by Braunstein et al. [34]. In [140], Passerini and Severini briefly investigated the use of the normalized Laplacian, although their analysis mainly focused on the unnormalized version. In both cases, a necessary step is the computation of the eigenvalues of the (normalized) graph Laplacian. This has computational complexity quadratic in the number of nodes of the network, thus making the application to large networks unfeasible.

Han et al. [82] sought to overcome this by looking at the second order polynomial approximation of the Shannon entropy. They considered the von Neumann entropy obtained from the normalized graph Laplacian and they showed that its quadratic approximation can be computed in terms of degree statistics. A similar result was obtained by Lockhart et al. [114] for the graph Laplacian. With this approximation to hand, the von Neumann network entropy has found applications in the analysis of several real-world networks [82, 149, 206] as well as in pattern recognition [20, 196]. More recently, Simmons et al. showed that the von Neumann entropy can be used as a measure of graph centralization [166], *i.e.*, the extent to which a graph is organized around a number of central nodes. Unfortunately, due to the spectral nature of this measure, it remains unclear how different structural patterns influence its value. Despite several attempts, a general structural interpretation of the von Neumann entropy remains an open problem.

## 2.6 Graph Thermodynamics

From the perspectives of both static and dynamic network analysis, many researchers have recognized that drawing on ideas from fields such as physics, may yield promising results. For instance, the use of analogies based on statistical mechanisms [11, 67, 90, 95, 138], thermodynamics [63, 204, 205] as well as quantum information [14, 35, 140] provides an excellent framework for describing complex systems. Among these the approaches based on thermodynamic analogies, relating the behaviour of microscopic particles to the macroscopic properties of a system [90, 95, 124], have proved remarkably successful.

For example, the heat bath analogy from thermodynamics described in [68] allows us to define physical measures of communicability and balance in networks. When the network is described by a partition function, thermodynamic quantities, such as entropy, energy and temperature can be straightforwardly derived from it [57]. Usually the Hamiltonian is obtained from the adjacency matrix or Laplacian matrix of the network and their eigenvalues. Clearly the type of partition function not only governs how the different energy levels are populated and how thermodynamic quantities charactering a network are computed. For instance in [204] the partition function is computed from a characteristic polynomial. Here the authors use the Maxwell-Boltzmann partition function to describe a thermalized network. However, in [26], the Bose Einstein partition function models a Bose gas over a network. In this case the process of Bose-Einstein condensation provides some deep insights into low temperature network behaviour and the coalescences of particles into the lowest energy states.

In network analysis and especially for those methods based on statistical physics

analogies, most approaches adopted to study complex structures are based on concepts from spectral graph theory [49]. For instance, in the heat bath analogy utilized to characterize networks, the energy states of a network are captured using the eigenvalues of the matrix representation of the network [189]. Here, in particular, global network characteristics, including the entropy, can be computed by using the Maxwell-Boltzmann partition function. By contrast, in [190, 193] energy states are occupied according to either Bose-Einstein or Fermi-Dirac statistics. For time evolving structures, thermodynamic approaches and spectral analysis have also proved to be extremely useful as tools for characterizing dynamic networks. For instance the von Neumann entropy of a network was firstly introduced by Braunstein [35] and analysed further in a number of subsequent works [14, 140]. The idea underpinning this measure is that the combinatorial graph Laplacian can be interpreted as the density matrix of a quantum system. In [205] the spectrum of the normalized Laplacian are used to define the microstates of a complex system, and the associated microstate occupation probabilities are used to define the thermodynamic entropy of the network. Additionally, the authors show how to compute thermodynamic variables in terms of node degree statistics. Similarly, in [192] the authors investigate the variation of entropy in time evolving networks and show how global modifications of network structure are determined by correlations in the changes in degree statistics for nodes connected by edges.

Thus, thermodynamic approaches can not only be used to characterize networks utilizing classical measures, such as entropy or energy, but can also be used to develop new measures of network complexity. For instance in [63], for the case of graph representation of complex networks, a flow complexity measure is presented and a thermodynamic characterization is developed based on the variation of local histories over graphs.

# 3

# Generative Models

Graph-based representations have been used with considerable success in computer vision, for instance in the abstraction and recognition of object shape. Non-rigid 3D shape retrieval is a crucial research topic in content-based object retrieval. Often this problem is cast in terms of the shapes intrinsic geometry because of its invariance to a broad array of non-rigid deformations.

In this Chapter we present a generative model devised in collaboration with Andrea Gasparetto for shape retrieval. This study is thus intended as a specific application and extension of the method presented in a previous work [77], designed for structural representations and based on the spectral decomposition on the graph Laplacian. Despite this, even for the shapes domain, the generative model relies on a spectral approach but in this case takes into account the Laplacian of a mesh. The peculiarity of this approach is given by the possibility of avoiding the pervasive correspondence issue by turning the eigenvectors of the Laplacian into a density in the spectral-embedding space which is estimated non-parametrically. We show that this model can accurately be learned from a set of 3D meshes. The experimental outcomes on the SHREC'14 benchmark demonstrate the effectiveness of the approach compared to the state-of-the-art.

## 3.1   Literature Survey

Non-rigid 3D shape retrieval is an active and crucial research topic in content based object retrieval. 3D models are widely-used in many application areas, such as computer aided design, medical modelling, bioinformatics, and a large number of 3D models have become available on the web. A shape retrieval algorithm can be interpreted as a query executor where, given a shape, the result set is the collection of shapes which belong to the same class. Moreover, the shapes returned should be ordered by decreasing similarity to the query shape. Various approaches have been introduced to address the non-rigid shape retrieval problem. Depending on the paradigm adopted to represent the objects, there are view based methods and model based methods.

The view based methods attempt to exploit the fact that similar 3D shapes look akin from the same perspective. Hence, many 2D projections have been employed to represent the shape, *e.g.*, silhouettes [44]. On the contrary, model based techniques make use of the

3D shape directly. Both structural and topological techniques and geometric techniques belong to this class of approaches.

Structural and topological techniques take into account structural properties like connected components or the holes in the shape. An example of such approach is the Surface Penetration Map by Yu *et al.* [208], where topological information is extracted from a model by morphing it into a sphere.

Finally, geometric techniques utilize the quantitative properties of the shapes that could be used to characterize the shape either globally or locally. Global methods aim at capturing the features of the shape as a whole and tend to be more computationally efficient. As an example, Zhang and Chen [209] propose several methods to compute efficiently global features and use them for 3D shape retrieval purpose.

Features are also employed indirectly in the methods which exploit the distributions of those measurements in place of the value of the feature itself [137]. With respect to global methods, local approaches can be used for partial matching, but they are usually less computationally efficient. These methods consider local properties around the neighbourhoods of points on the surface, such as curvature, volume and area [98].

One of the most popular approach within the geometric techniques involves the definition of an invariant representation of a shape capable of capturing its geometrical and topological properties, but at the same time being insensitive to transformations like bending and stretching, and robust to acquisition resolution or noise. On top of that, storage and computational costs of the representation should be taken into account.

For these reasons, a widely used approach to define shape descriptors is through spectral shape analysis. Research efforts have recently resulted in many spectral descriptors [16, 144, 170] usually based on the spectral decomposition of the Laplace-Beltrami operator, due to its invariance to isometries. In particular, the mesh-Laplacian, the discrete counterpart of the Laplace-Beltrami operator applied to surfaces, has been extensively used to provide spectral representations of structures [113]. Reuter *et al.* [145] suggest to use the sequence of eigenvalues (spectrum) of the Laplace-Beltrami operator of a surface as fingerprints, while Jain and Zhang [94] propose to use the eigenvalues of the geodesic distance matrix of a 3D object in order to build the associated shape descriptor. Huang *et al.* [91], on the other hand, build the signature directly over local features, selecting discriminative volumetric features over pre-aligned shapes.

The aggregation of local descriptors to build a global descriptor is a general thread in the literature. For this purpose, the *Bag-of-Features* (BoF) paradigm is quite popular and has been successfully applied to 3D shape description [37, 109, 115, 174]. Li and Hamza [109] used the BoF paradigm combining the exploitation of hierarchical structures of the shape, such as pyramid matching [80] and spatial relationship [37]. They proposed to adopt the eigenfunction associated with the second-smallest eigenvector of the Laplace-Beltrami operator in order to build a global surface coordinate system which is insensitive to shape deformation, showing that the introduction of global spatial context could improve the effectiveness of their descriptor in 3D shape recognition. Spatial pyramid [109, 115], is the term used to identify this approach. Other approaches inspired by text-analysis have been proposed. For instance, in [27, 86] the authors adopt higher-

order models defining relations between 'geometric words'. Within the bag of features model, features quantization is generally performed through unsupervised learning. Departing from this approach, Litman *et al.* [112] recently proposed a new supervised BoF framework mapping the discriminative training directly into the dictionary construction step. Finally, methods which aim at finding the correspondences between two or more shapes have been utilized for shape retrieval, employinf the correspondences found as a similarity measure between shapes and classifying the query shape accordingly [25, 42].

## 3.2 Background

A surface is a smooth compact 2-manifold $\mathcal{S}$ without boundary isometrically embedded in the Euclidean space $\mathbb{R}^3$ with geometry induced by the embedding. With shape, we denote the quotient group of surfaces modulo isometries. Finally, we define a mesh as a discrete representation of a surface embedded in $\mathbb{R}^3$. In order to address the shape retrieval task, we define an invariant representation of a shape obtained as the result of a data-driven process. To this end, let us define the discrete Laplacian as the discretization of the continuous Laplace-Beltrami operator on the mesh. The Laplacian operator is a differential operator which stores intrinsic geometry information (vertex-wise) about a mesh, allowing to preserve the relationship among vertices when isometric transformations are applied. Several approaches have been proposed to compute the Laplacian matrix from a mesh. In this work we adopt the algorithm proposed by Belkin *et al.* [24] which offers point-wise convergence guarantees and was experimentally shown to be quite robust. In particular, it approximates the Laplace operator of a surface from a mesh with point-wise convergence and they show that it converges (for fine meshes) to the Laplace-Beltrami operator. The Laplacian matrix is computed as $\mathcal{L} = A^{-1} * W$, with $A$ the area elements on the diagonal and $W$ a symmetric weight matrix. $\mathcal{L}$ is not symmetric but satisfies $\Phi^T A \Phi = I$ ($I$ the identity matrix). By setting $\Psi = A^{\frac{1}{2}}\Phi$ and $\mathcal{L}_N = A^{1/2}\mathcal{L}A^{-\frac{1}{2}} = A^{-\frac{1}{2}} * WA^{-1/2}$, the normalised Laplacian matrix $L_N$ has the same eigenvalues as $\mathcal{L}$ and its eigenvectors $\Psi$ are a discrete approximation of the eigenfunctions by taking integrals over the area elements rather than point samples over the vertices as in $\Phi$. Indeed $\mathcal{L}_N$ is symmetric, positive definite and its eigenvectors are robust with respect to changes in sampling densities. The spectral representation of a mesh can be computed from the Laplacian via spectral decomposition. In particular, given a discrete Laplacian $\mathcal{L}$, its eigendecomposition is $\mathcal{L} = \Phi\Lambda\Phi^T$, where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_{|V|})$ is the matrix whose diagonal contains the ordered eigenvalues, while $\Phi = (\phi_1|\phi_2|...|\phi_{|V|})$ is the matrix whose columns are the ordered eigenvectors.

## 3.3    Spectral Generative Model

The leading contribution of the proposed work consists in the definition of a generative model based on the spectrum of the Laplacian matrices associated to a set of meshes representing the same shape. The defined model is able to capture the structural variations between the meshes that belong to the same class, and it takes into account all the invariances of the spectral representation. Let $M$ be a mesh, and let $\mathscr{L}_M = \Phi_M \Lambda_M \Phi_M^T$ be the eigendecomposition of the associated Laplacian matrix. In the definition of our model, we follow [195], by building two separate and independent models for the eigenvalues and eigenvectors of the Laplacian[1]

$$P(M|\Theta) = P(\Lambda^M|\Theta^\Lambda)P(\Phi^M|\Theta^\Phi) \tag{3.1}$$

where $\Theta$ is the shape model divided into its eigenvalue-model component $\Theta^\Lambda$ and eigenvector-model component $\Theta^\Phi$.

As regards to the eigenvalue model $\Theta^\Lambda$, we follow Aubry *et al.* [16] for the choice of the distribution to be used. In their work, they showed empirically that the eigenenergies of an articulated shape are log-normally distributed random variables due to stability considerations derived from matrix perturbation theory. As a result, we model the set of eigenvalues as a series of independent log-normal distributions, one for each of the $d$ eigenvalues involved in the construction of the model. More formally, we define the posterior probability (relative to the eigenvalue component) of a mesh $M$ to belong to the class modelled by $\Theta^\Lambda$ as

$$P(\Lambda^M|\Theta^\Lambda) = (2\pi)^{\frac{d}{2}} \prod_{i=1}^{d} \frac{1}{\lambda_i \sigma_i} e^{\frac{-(\ln\lambda_i - \mu_i)^2}{2\sigma_i^2}} \tag{3.2}$$

where $\mu_i$ and $\sigma_i$ are model parameters to be learned from data and $d$ is the number of eigenvalues/eigenvectors used in the model, *i.e.*, the embedding dimension. Note that the log-normal distribution concerns corresponding eigenvalues among the different meshes of the training set.

On the other hand, the eigenvector component is modelled as an unknown distribution $\mathscr{F}$ on the $d$-dimensional spectral embedding space $\Omega_d \subseteq \mathbb{R}^d$. The $d$-dimensional embedding of the eigenvector matrix is obtained from a simple sub-sampling operation over the matrix $\Phi^M$, taking the first $d$ columns (first $d$ non-constant eigenvectors) of the matrix or, in other words, taking the first $d$ non-constant eigenvectors which are associated with the $d$ smallest non-trivial eigenvalues. The resulting $n \times d$ matrix is assumed to be a collection of $n$ $d$-dimensional points that belong to the embedding space $\Omega_d$. Eigenvectors are often assumed to be of unit Euclidean norm. Unfortunately, this results in a compression of the value of the components of the spectral embedded points as the number of the vertices of the mesh (or alternatively, the $n$ points of our representation) increases. To overcome

---

[1]The lack of eigenvalue modulation in the spectral embedding is compensated by a data driven modelling of the variation in each spectral dimension through non-parametric density estimation.

this issue, we scale the $d$ eigenvectors which belong to the embedding by multiplying for the number of vertices of the mesh $n$. So, let $\hat{\Phi}^M = (\bar{\Phi}^M) \cdot n$, with $\bar{\Phi}^M \in \mathscr{R}^{n \times d}$, be the resulting embedded eigenvector matrix. Note that we are not assuming nor requiring the meshes to have the same number of vertices. Indeed, the number of vertices of the meshes belonging to the datasets used in the experimental section are different.

With this model we cast the learning phase into a non-parametric density estimates of the distribution of the spectral embedding points $\phi_1^M, \ldots, \phi_n^M$. Under these assumptions, the eigenvector model parameter $\Theta^\Phi$ is constituted of a collection of $N$ $d$-dimensional vectors $\theta_1^\Phi, \ldots, \theta_N^\Phi$ corresponding to samples from the unknown density function $\mathscr{F}$. In the learning phase, the parameter $\Theta^\Phi$ is obtained by aligning and stacking (and eventually sub-sampling) all those spectral embedding points together from the sample meshes which belong to the same class, or in other words, all the meshes that belong to the training set and represent the same shape. The density model allows us to avoid the vertex-alignment problem allowing many-to-many relations among vertices from different meshes and the separation of eigenvector and eigenvalue constrains the set of isometries of the embedding space to the discrete set of sign changes. However, sampling noise and approximate isometries of the surfaces result in mixing of the eigenspaces, especially between those with similar eigenvalues. We address this by adopting a kernel Procrustes alignment approach, locally optimizing over the orthogonal group $\mathbb{O}(d)$. Then, the posterior probability $P(\Phi^M | \Theta^\Phi)$ can be computed solving the problem:

$$\max_{\mathscr{O} \in \mathbb{O}(d)} \max_{S \in \{\pm 1\}^d} (Nh^d)^{-n} \prod_{i=1}^{n} \sum_{j=1}^{N} e^{-\frac{\|\mathscr{O}S\phi_i^M - \theta_j^\Phi\|^2}{2h^2}} \tag{3.3}$$

expressing the optimization problem in terms of the product of Parzen-Rosenblatt kernel density estimators. In particular, $\phi_i^M$ is one of the sample points obtained taking the first $d$ components of the $i$-th row of the eigenvector matrix associated to the mesh $M$, while $\theta_j^\Phi$ is the $j$-th component (still a $d$-dimensional row vector) of the eigenvector model $\Theta^\Phi$. Here the eigenvector model is assumed to be a collection (an array) of samples taken from the training set during the learning phase. Furthermore, in order to minimize the variance between all the eigenvector matrices of the meshes of the training set (and simultaneously increasing the variance with the eigenvector matrices which does not represent the same shape), we introduce two alignment steps. In particular, the matrix $\mathscr{O}$ that appears in the optimization problem 3.3 represents an orthogonal transformation used to align as much as possible all the points of our representation, while $S$ is a matrix containing just $\pm 1$ used to solve the sign ambiguity problem.

In the computation of the kernel density estimator, a rule of thumb must be chosen in order to estimate the bandwidth $h$ of the kernel. In this work, we employ the Silverman's rule of thumb for the multivariate case [165] and we estimate the bandwidth as

$$h = \left( N \frac{d+2}{4} \right)^{-\frac{1}{d+4}} \sigma \tag{3.4}$$

where $d$ is the embedding dimension, $N$ is the number of meshes employed in the con-

struction of the model $\Theta^\Phi$ (*i.e.,* the dimension of the training set) and $\sigma$ is the standard deviation computed as the squared root of the mean variance (the trace of the covariance matrix $\Sigma$ divided by the $n$ nodes of the model) of the eigenvector model

$$\sigma = \sqrt{\frac{1}{n} Tr(\Sigma)} \tag{3.5}$$

Note that we are not assuming that the eigenfunctions are independent. Indeed, we build our eigenvector model as a point cloud. Given an eigenvector matrix (as columns), each point is given by a row of this eigenvector matrix. We assume those points to be independent observations of an unknown underlying distribution which we estimate non parametrically. For this reason we state that our model is independent of the sampling, since even if we prune out part of those points, the underlying distribution remains the same. Robustness analysis presented in section 3.3.3 confirms the assumption. A more practical analysis about the construction of such model is demanded to section 3.3.1.

### 3.3.1   Model Learning

In order to address the retrieval task, we define two separate phases. The first one is the learning phase. The main goal of the learning process is to estimate the model parameters for both the eigenvector model and the eigenvalue model. Before estimating the parameters, the training dataset must undergo a pre-processing phase. Given a set of meshes $S = \{M_1, M_2, \ldots, M_N\}$ representing the same shape $\mathscr{S}$ in different poses (training set), the first step involves the computation of the Laplacian matrix associated to each mesh $M_i$, obtaining the set $\{\mathscr{L}_1, \mathscr{L}_2, \ldots, \mathscr{L}_N\}$. Applying a singular valued decomposition to each Laplacian matrix, we obtain the eigendecomposition $\mathscr{L}_i = \Phi_i \Lambda_i, \Phi_i^T$ where the diagonal elements of $\Lambda_i$ are in ascending order and we remove the trivial eigenvalue and the corresponding eigenvector (which results to be constant). Finally, we are not interested in the whole Laplacian eigendecomposition, but just an embedding of it. In other words, given $d$ as the embedding dimension, we keep the first $d$ eigenvalues and the first $d$ eigenvectors of the eigendecomposition of the Laplacians. As a result of this pre-processing phase, we obtain the set $\{(\Lambda_i, \Phi_i)\}_{i=1,\ldots,N}$, where $\Lambda_i \in \mathscr{R}^{d \times d}$ and $\Phi_i \in \mathscr{R}^{n_i \times d}$ ($n_i$ the number of vertices of the mesh $i$).

Then, the eigenvector model of the class $\mathscr{S}$, denoted as $\Phi^{\mathscr{S}}$, is defined as

$$\Phi^{\mathscr{C}} = \begin{bmatrix} \phi_1^1 & \phi_2^1 & \cdots & \phi_d^1 \\ \phi_1^2 & \phi_2^2 & \cdots & \phi_d^2 \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1^m & \phi_2^m & \cdots & \phi_d^m \end{bmatrix}$$

where $\phi_j^i$ denotes the $j$-th non-trivial eigenvector (still a column vector) of the $i$-th mesh of the training set. In other word, we perform a vertical concatenation of all the eigenvectors matrices of the meshes that belong to class $\mathscr{S}$. Therefore, the final dimension of the

model $\Phi^{\mathscr{S}}$ can be expressed as $\left(\sum_{i=1}^{N} ||M_i||\right) \times d$, with $||M_i||$ the number of vertices of the $i$-th mesh.

### 3.3.1.1 Estimating the Eigenvector Sign-Flips
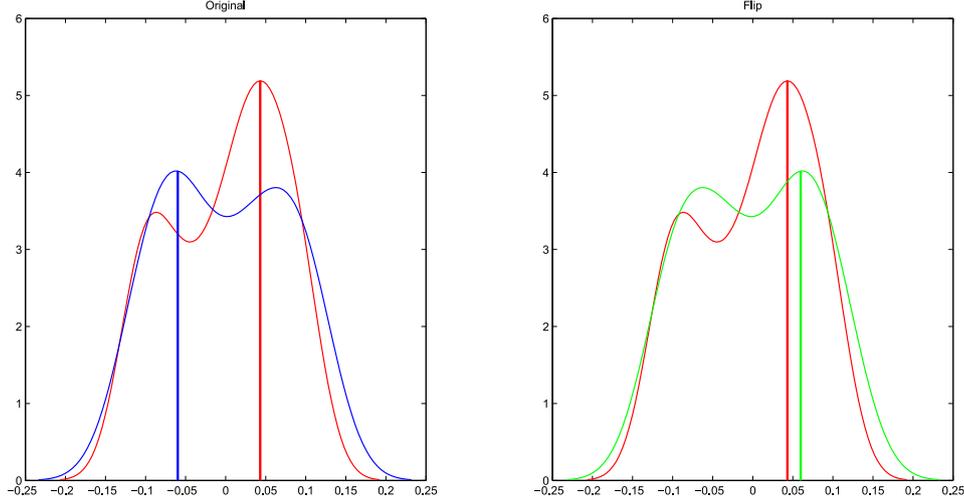


Figure 3.1: The red curve is the pdf of the $i$-th eigenvector of matrix A (reference mesh), calculated with the kernel density estimation, and the blue curve is the pdf of the $i$-th eigenvector of matrix B. On the right, the green curve represents the *pdf* after the sign flip.

The eigenvector matrix is unique up to a sign factor. Since our method characterizes every vertex of a mesh with a feature vector, a sign disambiguation is mandatory. There are diverse ways that allow to detect and solve this ambiguity, like using the correlation between two functions (*i.e.*, probability density functions). If the correlation grows after a flip, then the eigenvector sign should be flipped. Unfortunately, with increasing size, this method becomes computationally heavy.

For such a reason, we have to employ an heuristic-based method in order to solve the sign-ambiguity problem. However, being heuristic, it does not guarantee the detection of all the correct signs.

Given the eigenvector matrices $\Phi^A$ and $\Phi^B$, computed via the eigendecomposition of the Laplacian of two meshes $M_A$ and $M_B$ (representing the identical shape but in different poses), let $\phi_i^A$ and $\phi_i^B$ be the $i$-th eigenvector. We assume the eigenvectors to be random variables whose probability density functions are unknown. In addition, we assume all $i$-th eigenvectors associated to each mesh representing the same shape to share a similar probability density function (up to a linear transformation). Such linear transformation does not influence the shape of the *pdf*, but it influences the sign of the peak of such function. In other words, if the sign of the peaks disagree, a sign flip occurs. More
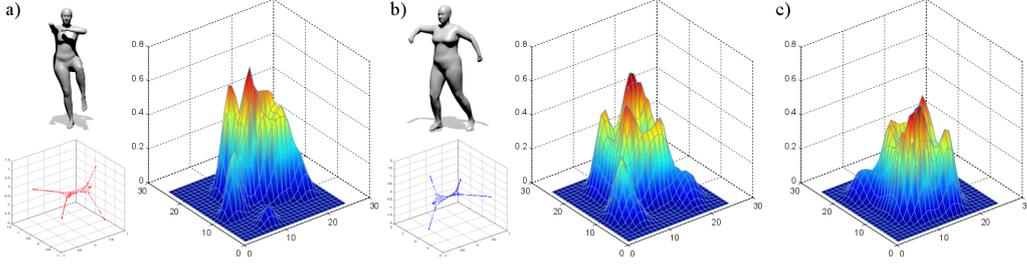
Figure 3.2: An example of the effect of Orthogonal transformation over the eigenvector matrix. In the left-hand side panel *a)* the reference mesh - the graph below the mesh shows the plot of the first three non trivial eigenvectors of the reference mesh whereas on the right of the mesh there is the kernel density estimation computed on the eigenvector matrix. In the middle panel *b)* the mesh to be aligned according to *a)*. Finally, in the right-hand side panel *c)* the result of the optimization process.

formally, after the selection of a mesh as reference mesh (for example *A*), we define our heuristic approach to the sign ambiguity solution as

$$
\phi_j^B = \begin{cases} \phi_j^B(-1) & \text{if } x_j^{A*} < 0 \text{ and } x_j^{B*} \geq 0 , \\ \phi_j^B(-1) & \text{if } x_j^{A*} \geq 0 \text{ and } x_j^{B*} < 0 , \\ \phi_j^B & \text{otherwise.} \end{cases} \tag{3.6}
$$

The *pdf*s of each eigenvector are estimated using kernel density estimation. The density estimates are evaluated at 100 points covering the range of the eigenvectors. Those evaluations are then used to find the peaks, more precisely the related independent variables $x_j^{A*}$ and $x_j^{B*}$ of the functions.

Hence, before the construction of the eigenvector model $\Phi^{\mathscr{S}}$, we randomly select a mesh $M_j$ as the reference mesh. Then we flip the eigenvector matrices associated to the other meshes according to 3.6. The main drawback of the proposed solution is its non-robustness to multiplicity of eigenvalues. Furthermore, similar eigenvalues could result in mixing of the eigenspaces. The problem is addressed through an additional alignment step, which is explained in detail in the next section.

### 3.3.1.2   Estimating the Eigenvector Orthogonal Transformation

The alignment step introduced in the previous section acts as an orthonormal transformation and thus it does not violate the orthogonality of the basis. Nevertheless, an isometric difference between the eigenvector matrix of the reference mesh and the eigenvector matrix of an another mesh is expected. In order to overcome the misalignment and consequently to minimize distances between feature vectors of the two matrices, a further alignment step is employed. From this viewpoint, the formulation of such optimization problem is similar to the Orthogonal Procrustes Problem (OPP), that is solving for the optimal orthogonal matrix that maps the first set of points into the second one. However,

our algorithm differs from the standard OPP since we seek for the orthonormal transformation which maximizes a certain probability. In particular, we define the probability density in terms of Parzen-Rosenblatt kernel density estimator (Gaussian kernel)

$$P_h(x) = \frac{1}{n \cdot h} \sum_{j=1}^{n} e^{-\frac{1}{2}\frac{\|x - y_j\|^2}{h^2}} \tag{3.7}$$

By integrating the constraints of the optimization problem into our definition of probability density, we get

$$\underset{\mathscr{O} \in \mathbb{O}(d)}{\arg\max} \prod_{i}^{m} \sum_{j}^{n} e^{-\frac{1}{2}\frac{\|\mathscr{O}x_i - y_j\|^2}{h^2}} \tag{3.8}$$

where $m$ is the number of vertices of the mesh we want to align to the reference mesh $y$ and the parameter $h$ is the bandwidth computed according to equation 3.4.

In order to solve 3.8, we firstly have to calculate the gradient with respect to an incremental variation and later compute iteratively the orthonormal transformation. The log-likelihood of the problem, after the introduction of the additive rotation $\mathscr{T}$, is given by

$$\ell_{\mathscr{L}}(y|x) = \sum_{i}^{m} \log\left(\sum_{j}^{n} e^{-\frac{1}{2}\frac{\|\mathscr{T}\mathscr{O}x_i - y_j\|^2}{h^2}}\right) \tag{3.9}$$

Let $\alpha_{i,j}$ be defined as $\alpha_{i,j} = e^{-\frac{1}{2}\frac{\|\mathscr{O}x_i - y_j\|^2}{h^2}}$. Then, by deriving the function in 3.9, with respect to the additional rotation $\mathscr{T}$, we get

$$\frac{\partial \ell_{\mathscr{L}}}{\partial \mathscr{T}} = \sum_{i} \frac{\sum_{j} \alpha_{ij}\left(-\frac{1}{2}\frac{\frac{\partial}{\partial \mathscr{T}}\|\mathscr{T}\mathscr{O}x_i - y_j\|^2}{h^2}\right)}{\sum_{j} \alpha_{ij}} \tag{3.10}$$

In particular, the partial derivative obtained by deriving with respect to the identity $\mathscr{I}$ ($\mathscr{T} = \mathscr{I}$), results

$$\frac{\partial}{\partial \mathscr{T}}\|\mathscr{T}\mathscr{O}x_i - y_j\|^2 = -2y_j(\mathscr{O}x_i)^T \tag{3.11}$$

$$= -2y_j x_i^T \mathscr{O}^T \tag{3.12}$$

We can rewrite 3.10 as

$$\frac{\partial \ell_{\mathscr{L}}}{\partial \mathscr{T}} = \underbrace{\left(\sum_{i} \frac{\sum_{j} \alpha_{ij} h^{-2} y_j x_i^T}{\sum_{j} \alpha_{ij}}\right)}_{A} \mathscr{O}^T = A\mathscr{O}^T \tag{3.13}$$

In order to find the maximum, we want to project the gradient to the null space. $\mathcal{T}$ is an orthogonal rotation matrix, hence it belongs to the Lie group $\mathbb{O}(d)$ (with $d$ the embedding dimension). The tangent space (which we compute through the gradient of the log-likelihood 3.13) at the identity element of a Lie group is its Lie algebra, which represents the skew-symmetric matrices space. Since the skew-symmetric component of a matrix $B$ is given by $\frac{B-B^T}{2}$, in order to nullify such component, we want to make $B$ symmetric. More precisely, in order to find the maximum we want to make $A\mathcal{O}^T$ symmetric (which means $A\mathcal{O}^T = (A\mathcal{O}^T)^T$ . Indeed

$$\frac{A\mathcal{O}^T - (A\mathcal{O}^T)^T}{2} = 0 \tag{3.14}$$

Treating the problem as an Orthogonal Procrustes Problem, the rotation matrix $\mathcal{O}$ which symmetrizes $A\mathcal{O}^T$ is computed through singular value decomposition of the matrix $A$. The decomposition yields $svd(A) = ULV^T$, while we can compute the rotation matrix as $\mathcal{O} = UV^T$. It is easy to see that $\mathcal{O}$ symmetrizes the gradient defined above, indeed

$$A\mathcal{O}^T = (ULV^T)(VU^T) = ULU^T \tag{3.15}$$

To summarize, the computation of the rotation matrix $\mathcal{O}$ is achieved using the following algorithm.

1. Initialize $\mathcal{O} = I$

2. Compute $\alpha_{ij}$ (**??**) for each $i = 1,\ldots,n$ ($n$ the number of vertex of a mesh) and $j = 1,\ldots,N$ ($N$ the number of points of the model)

3. Compute the matrix $A$ (according to 3.13)

4. $svd(A) = ULV^T$

5. $\mathcal{O} = UV^T$

6. If the convergence is achieved, *i.e.,* $A \approx A^T$, or the maximum number of iterations allowed is reached, end the algorithm, otherwise repeat from 2

Once both the alignment steps are computed on the spectral decomposition of each mesh belonging to the training set, we are able to build the matrix $\Phi^{\mathscr{S}}$ through the stacking operation mentioned at the beginning of this section. We decide to not combine the two alignment steps (the sign disambiguation and the orthonormal transformation) since the former represents just a coarse alignment, while the latter represents a fine alignment step which is effective only for smaller and local alignments.

### 3.3.1.3 Estimating the Eigenvalue Model

The second independent model involved in the proposed method regards the eigenvalues computed on the Laplacian of each mesh of the training set $S$ (see section **??**). The eigendecomposition produces the set of pairs $\{(\Lambda_i, \Phi_i)\}_{i=1,\dots,N}$. Especially, $\Lambda_i$ contains the first $d$ non-trivial eigenvalues of the $i$-th mesh of the training set as its diagonal elements. Let $\Lambda^{\mathscr{S}}$ be a $N \times d$ matrix whose rows are the eigenvalues extracted from the $\Lambda_i$s.

$$
\Lambda^{\mathscr{S}} = \begin{bmatrix} diag(\lambda_1^{\mathscr{S}}) \\ diag(\lambda_2^{\mathscr{S}}) \\ \vdots \\ diag(\lambda_N^{\mathscr{S}}) \end{bmatrix} = \begin{bmatrix} \lambda_{1_{1,1}}^{\mathscr{S}} & \cdots & \lambda_{1_{d,d}}^{\mathscr{S}} \\ \lambda_{2_{1,1}}^{\mathscr{S}} & \cdots & \lambda_{2_{d,d}}^{\mathscr{S}} \\ \vdots & \ddots & \vdots \\ \lambda_{N_{1,1}}^{\mathscr{S}} & \cdots & \lambda_{N_{d,d}}^{\mathscr{S}} \end{bmatrix}
$$

We assume that all the $j$-th eigenvalues of $\Lambda^{\mathscr{S}}$, with $j = 1,\dots,d$, are distributed as a log-normal distribution (see equation 3.2). In order to learn the model parameters $\mu$ (the mean) and $\sigma^2$ (the variance), we do a maximum likelihood estimation using

$$
\hat{\mu} = \frac{\sum_i \ln x_i}{N}, \quad \hat{\sigma}^2 = \frac{\sum_i (\ln x_i - \hat{\mu})^2}{N} \tag{3.16}
$$

As a result of this learning phase, we compute a pair of parameter $(\mu, \sigma^2)$ for each of the $d$ eigenvalues of the model. Hence, the eigenvalue model for a particular shape $\mathscr{S}$ is represented by the set $\Theta^{\Lambda} = \left\{(\mu_i, \sigma_i^2)\right\}_{i=1,\dots,d}$.

## 3.3.2 Prediction

The learning phase of the proposed method produces two separate models, both referring to a shape $\mathscr{S}$. The eigenvector model $\Theta^{\Phi}$ contains the embedded eigenvector matrices stacked together after the two alignment steps introduced in section 3.3.1.1 and 3.3.1.2, while the eigenvalue model $\Theta^{\Lambda}$ contains the parameters set learned on the eigenvalues of the training set. Once both models are computed, we can combine them to compute the posterior probability of a new mesh $M^*$ (whose spectral decomposition is $(\Phi^*, \Lambda^*)$) with respect to the model representing a mesh $\mathscr{S}$. Assuming the independence of the two models, we can define the conditional probability as

$$
P(\mathscr{S} \mid M^*) = P(\Phi^* \mid \Theta^{\Phi})P(\Lambda^* \mid \Theta^{\Lambda}) \tag{3.17}
$$

Since both $P(\Phi^* \mid \Theta^{\Phi})$ and $P(\Lambda^* \mid \Theta^{\Lambda})$ come from a log-derivation, equation 3.17 can be rewritten as

$$
\log \quad P(\mathscr{S} \mid M^*) = \ell_{\mathscr{L}}(\Phi^* \mid \Theta^{\Phi}) + \ell_{\mathscr{L}}(\Lambda^* \mid \Theta^{\Lambda}) \tag{3.18}
$$

where the eigenvector model log-likelihood is defined as
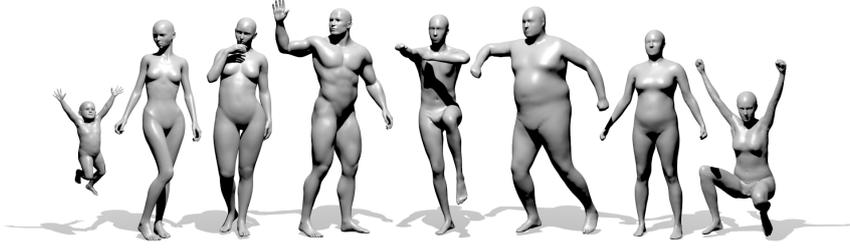
Figure 3.3: Examples of Shapes in SHREC'14 Humans dataset

$$\ell_{\mathscr{L}}(\Phi^*|\Theta^\Phi) = \prod_{i=1}^{n} P(x_i) = \sum_{i=1}^{n} \log P(\bar{x}_i|\Theta^\Phi) \tag{3.19}$$

Here, $n$ is the number of vertices of the mesh $M^*$, while $\bar{x}_i$ is a $d$-dimensional row vector representing the $i$-th feature vector of the mesh. Note that the eigenvector matrix associated to $M^*$ undergoes the same alignment steps defined in section 3.3.1.1 and 3.3.1.2. On the other hand, the eigenvalue model log-likelihood is defined as

$$\ell_{\mathscr{L}}(\Lambda^*|\mu_i^\Theta, \sigma_i^\Theta) = \prod_{i=1}^{d} P(\lambda_i) = \sum_{i=1}^{d} \log P(\lambda_i) \tag{3.20}$$

where $\mu_i^\Theta$ and $\sigma_i^\Theta$ are the mean and the variance parameters learned according to 3.16. The posterior probability is computed against each model of each shape contained in the dataset. As a decision rule, we classify a certain mesh as representing the shape whose model yields the highest probability. Even though we employ the proposed method as a classifier, the output of the approach is still a probability value which can be employed as a similarity score.

### 3.3.3   Experiments

In this section we show the performance achieved by the proposed method while addressing a classification task. The evaluation is based on the comparison of the results obtained by our method with respect to the current state-of-the-art on a very popular shape retrieval benchmark,*i.e.*, SHREC'14 benchmark. The efficacy of the method is evaluated in terms of mean classification accuracy. Besides, we test the robustness and its sensitiveness to sub-sampling, showing the retrieval performance with different amount of data employed in the learning process and with different embedding dimensions. Finally, a short performance analysis (in terms of execution time) is proposed at the end of the section.
**SHREC'14 Humans** [142] is a dataset containing a large number of meshes which represent 55 different shapes. It consists of two different sub-sets. The first one (synthetic) contains 15 different human models each in 20 different poses, while the second one

| Method | Synthetic | Scanned |
|---|---|---|
| ISPM [108] | 90.2 | 25.8 |
| DBN [142] | 84.2 | 30.4 |
| R-BiHDM [207] | 64.2 | 64.0 |
| HAPT [78] | 81.7 | 63.7 |
| ShapeGoogle (VQ) [37] | 81.3 | 51.4 |
| Unsupervised DL [112] | 84.2 | 52.3 |
| Supervised DL [112] | 95.4 | 79.1 |
| NPSR | 95.0 | 79.0 |

Table 3.1: Comparison of different retrieval methods in terms of average retrieval precision on the SHREC'14 Humans datasets.

contains scans of 40 human subjects, each in 10 different poses. All shapes were down-sampled to have about $6 \times 10^3$ triangles for efficiency purpose. Both datasets are extremely challenging, as they contain geometrically similar human shapes. We employed a 10-fold test over the datasets resulting respectively in 2 positives and 28 negatives per query for the synthetic dataset and 1 positive and 39 negatives for the scanned dataset. The process was repeated for each possible subset of the datasets and the whole test was averaged over 100 iterations. We compare our approach with the most accurate methods that participate to the SHREC'14 benchmark: Histograms of Area Projection Transform (HAPT) [78], Deep Belief Network (DBN) [142], Intrinsic Spatial Pyramid Matching (ISPM) [108], Reduced Bi-harmonic Distance Matrix (R-BiHDM) [207] and to a group of approaches based on the Bag-of-Features paradigm ( [37, 112]). Table 3.1 shows the results yielded by those methods. The proposed method performs well and is competitive with the current state-of-the-art. We are able to achieve high retrieval accuracies in both the synthetic and scanned datasets, while the differences with respect to the current state-of-the-art are not statistically significant.

Finally, we present a robustness analysis of the proposed method. The random sub-sampling has been performed directly on the Laplacian matrices of the meshes which belong to the same training set. Indeed, performing the sub-sampling after the model construction through the elimination of some feature vectors (which means, taking out some rows from the model $\Theta^\Phi$) could lead to biased results since both the alignment steps employed in our learning phase would have benefited from the contribution of data that would no longer be available in the next steps. Figure 3.4 shows the average precision while both the embedding dimension and the sampling percentage vary. As you can see, for the lowest level of sub-sampling (respectively, 75% and 100% of data used in the learning process) we were not able to produce the results with respect to all the embedding dimensions due to the dimension of the model. Even so, it is easy to see that the performance achieved are consistent at every sub-sampling level. In particular, the results show that the defined model is robust with respect to the sub-sampling of points. This
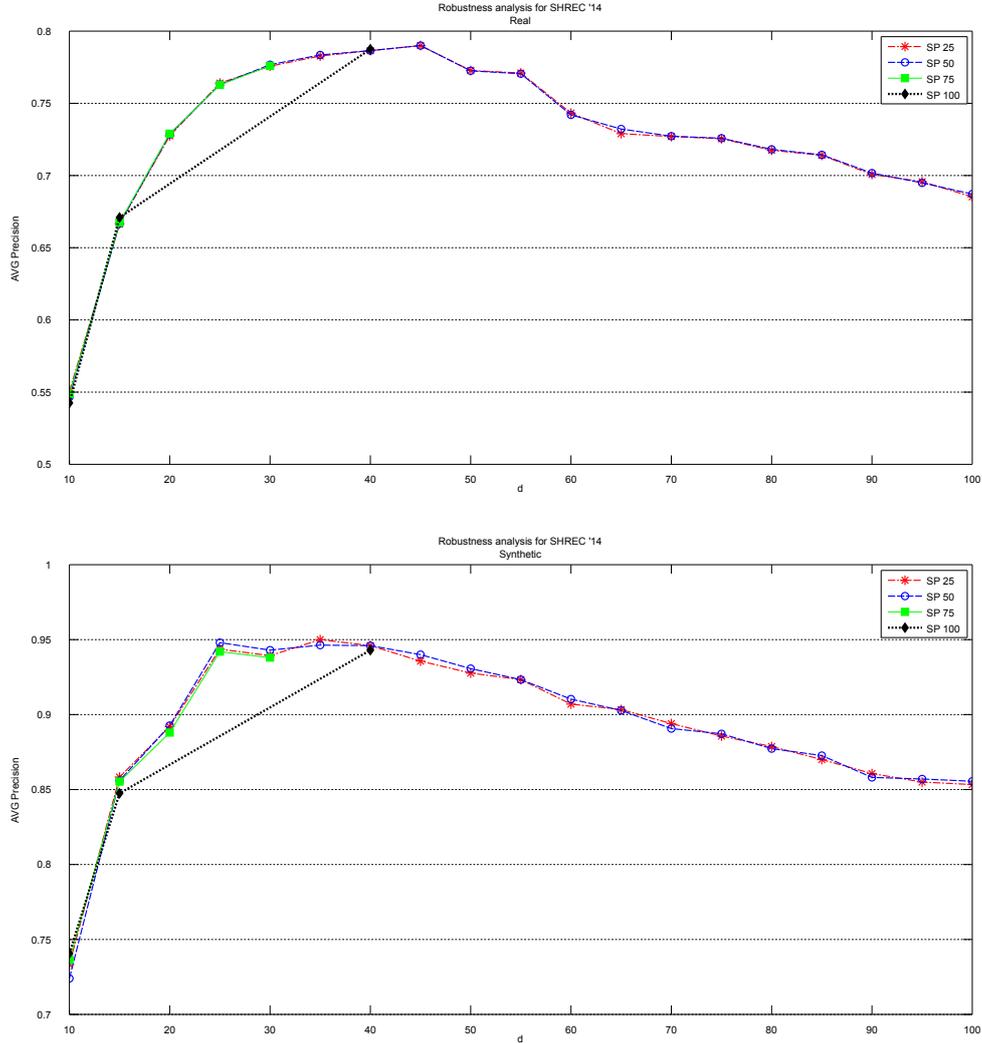
Figure 3.4: Robustness analysis of the proposed method in both Real (top) and Synthetic (bottom) datasets. The marks represent the average precision obtained with a certain embedding dimension. Different lines represent different sub-sampling percentage, *e.g.*, the red line represents the average accuracy with respect to different embedding dimensions after removing 75% of the data on the model (*i.e.,* keeping only 25% of the data).

allows to greatly reduce the dimension of the model itself and consequently increasing the performance of the whole pipeline.

## 3.4   Conclusion

The work here introduced, investigates over the application of a generative model for shapes retrieval task. The model, originally devised for graph structures, is grounded on the spectral decomposition of the Laplace-Beltrami operator. In particular, it is based

on the definition of two separated models, one for the eigenvalues and the other for the eigenvectors. The former ones are assumed to be log-normal distributed whereas there is no assumption for the eigenfunctions. Indeed, in this case, a kernel density estimation is employed.

Outcomes prove that the approach is easily adaptable to any graph based representation and actually, for the $3D$ shape retrieval task, the method is able to compete with the current state-of-the-art algorithms. Besides, by robustness analysis viewpoint, the approach also seems to ensure sampling, that means a better computation efficiency is achievable by removing data from the models without affecting the accuracy.

# 4

# Structural Analysis via Quantum Processes

In this Chapter, we address some of the main issues encountering in the wide subject of structural analysis. Despite the differences in contents between the various matters taken into account, a common denominator, or better say a fil rouge, actually binds all these studies together. Specifically, the quantum approach adopted in each work allowed us to undertake a path of analysis rich in novel tools but at the same time rooted in quantum information theory framework.

We commence in 4.1 by attempting to shed light on the meaning of the von Neumann entropy by the structural patterns interpretation viewpoint. We also investigate the potential relationships between the two variants, one based on the Laplacian matrix and the other on the normalized Laplacian matrix. Still concerning the two variants, we seek to estimate the quadratic approximations quality of both.

In 4.2 we turn our attention to a novel centrality index for edges. Even in this case we refer to a well known quantum information theoretical measure, the Holevo quantity. More to the point, we define the influence of an edge in terms of the contribution to the Von Neumann entropy of the graph.

Finally, in 4.3 we extend a previous work measuring the similarity between two structures. This algorithm makes use of a kernel on undirected graphs and evaluates the similarity through continuous-time quantum walks and the quantum Jensen-Shannon divergence. We contribute by investigating different Hamiltonian choices and the application to directed graphs.

## 4.1   On the von Neumann Entropy of Graphs

The von Neumann entropy of a graph is a spectral complexity measure that has recently found applications in complex networks analysis and pattern recognition. Two variants of the von Neumann entropy exist based on the graph Laplacian and normalized graph Laplacian, respectively. Due to its computational complexity, previous works have proposed to approximate the von Neumann entropy, effectively reducing it to the computation of simple node degree statistics. Unfortunately, a number of issues surrounding the von

Neumann entropy remain unsolved to date, including the interpretation of this spectral measure in terms of structural patterns, understanding the relation between its two variants, and evaluating the quality of the corresponding approximations.

In this Section we aim to answer these questions by first analysing and comparing the quadratic approximations of the two variants and then performing an extensive set of experiments on both synthetic and real-world graphs. In particular we intend to: 1) shed light on the relation between the structure of a network and its von Neumann entropy, both for the version based on the graph Laplacian and the normalized Laplacian, thus also 2) deepening our understanding of the difference between these two entropies; 3) evaluate the quality of the quadratic approximation. Han et al. [82] also briefly analysed the accuracy of the quadratic approximation, but only for the version of the von Neumann entropy based on the normalized Laplacian. As explained in 4.2.4, their analysis is also strongly influenced by the use of datasets with graphs of varying size, whereas our experimental evaluation is on datasets of fixed graph size. We are also particularly interested in looking at how different edges contribute to the overall graph entropy, revealing additional inaccuracies introduced by the quadratic approximation.

The rest of this work is organized as follows: in 4.1.1 the necessary mathematical and physical background is being introduced. Then in 4.1.2 the quadratic approximation of the two variants of the von Neumann entropy considered in this paper are presented. In 4.2.4 we empirically compare the exact and approximated entropies. Finally, we discuss in 4.2.5 the results of our investigation.

## 4.1.1 Background

In quantum mechanics, a system can be either in a pure state or a mixed state. Using the Dirac notation, a pure state is represented as a complex-valued column vector $|\psi_i\rangle$. A mixed state, on the other hand, is a statistical ensemble of pure states $|\psi_i\rangle$, each with probability $p_i$. Density matrices are trace-one positive semidefinite matrices introduced to describe mixed state systems [131]. For such a system, $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle$ is a pure state and $p_i$ is the probability associated to it. Density matrices play a pivotal role in quantum mechanics and are linked with the observables of quantum systems, e.g., the expectation value of the measurement of an observable $O$ is $\langle O \rangle = \text{Tr}(\rho O)$.

### 4.1.1.1 The von Neumann entropy

Given a quantum mechanical system described by a density matrix $\rho$, its von Neumann entropy [131] is defined as

$$S(\rho) = -\text{Tr}(\rho \ln \rho), \tag{4.1}$$

where Tr denotes the trace operator and ln denotes the matrix logarithm. The von Neumann entropy of $\rho$ can also be computed as the Shannon entropy of the spectrum of $\rho$, i.e.,

$$S(\rho) = -\sum_{i=1}^{n} \lambda_i \ln \lambda_i \tag{4.2}$$

where $\lambda_i$ denotes the $i$-th eigenvalue of $\rho$, with the convention $0 \ln 0 = 0$.

The von Neumann entropy measures the maximum amount of classical information that we can extract from a mixture of pure states [186]. It has also been extensively used in the literature to study correlated systems and to define entanglement and distinguishability measures [120, 135, 136]. Finally note that the von Neumann entropy of a pure state $\rho = |\psi_i\rangle\langle\psi_i|$ is always zero. On other hand, a mixed state always has non-zero entropy. Therefore, the von Neumann entropy $S(\rho)$ can also be seen as a measure of how close $\rho$ is to being a pure state.

### 4.1.1.2  Graph density matrices

Let $G$ be an undirected graph with vertex set $V$ and edge set $E \subseteq V \times V$. Recall that the adjacency matrix of the graph $G$ is the symmetric matrix with elements

$$A_{uv} = \begin{cases} 1 \text{ if } (u,v) \in E \\ 0 \text{ otherwise} \end{cases} \tag{4.3}$$

Let $D$ be the diagonal matrix with elements $d_u = \sum_{v=1}^{n} A(u,v)$, where $d_u$ is the degree of the node $u$. Then $L = D - A$ is the graph Laplacian, the combinatorial analogue of the Laplace-Beltrami operator [99].

Braunstein et al. [34] proposed to use the graph Laplacian to map graphs to quantum states. More specifically, let $G$ be a graph with Laplacian $L$, then its density matrix is defined as $\rho(L) = \frac{L}{\text{Tr}(L)} = \frac{L}{2m}$, where $m$ denotes the number of edges of $G$. Passerini and Severini [140] proposed an alternative version of the von Neumann entropy for graphs based on the normalized Laplacian $\mathscr{L} = D^{-1/2}LD^{-1/2}$. Given a graph $G$ with $n$ nodes and normalized Laplacian $\mathscr{L}$, they define the density matrix of $G$ as $\rho(\mathscr{L}) = \frac{\mathscr{L}}{\text{Tr}(\mathscr{L})} = \frac{\mathscr{L}}{n}$.

## 4.1.2  The von Neumann entropy of a Graph

With the density matrix of a graph to hand, one can compute its von Neumann entropy using either Eq. 4.1 or Eq. 4.2. In the remainder of this paper, we refer to the von Neumann entropies computed on $\rho(L)$ and $\rho(\mathscr{L})$ as the *Laplacian entropy* and *normalized Laplacian entropy*, respectively.

A number of previous works have made steps toward a general interpretation of the Laplacian entropy, although this remains an open problem [13, 14, 50, 51, 58, 140]. Passerini and Severini [140] have observed that the Laplacian entropy of a graph tends to grow with the number of connected components, long paths and nontrivial symmetries. They have also shown that the Laplacian entropy of a graph $G$ is upper bounded by $\ln(n-1)$, where $n$ denotes the number of nodes of $G$, and that this upper bound is saturated by both complete graphs and regular graphs (for large $n$), suggesting that the Laplacian entropy can be interpreted as a measure of regularity. Du et al. [58] proved that the same bound holds also for Erdös-Rényi random graphs, highlighting a connection between randomness and regularity. In [13], the authors showed that for scale free networks the Laplacian entropy of a graph is linearly related to the Shannon entropy of the

graph ensemble [13]. More in general, Anand et al. observed in [14] that for graphs with heterogeneous degree distributions there exists a correlation between these entropies.

De Beaudrap et al. have shown that the Laplacian entropy of a graph can be interpreted as a measure of the amount of entanglement between a system corresponding to the vertices and a system corresponding to the edges of the graph [51]. This in turn allows them to identify cospectral graphs (i.e., graphs having the same graph spectrum) as graphs with local unitarily equivalent pure states [51]. Finally, Dairyko et al. [50] show that adding an edge to a graph can result in a decrease of its Laplacian entropy, i.e., the Laplacian entropy does not satisfy the subadditivity property [53]. More recently, Simmons et al. [166] have proved that the Laplacian entropy of a graph is related to both the graph Theil index and the graph Jain fairness index, highlighting an interesting connection between the Laplacian entropy and the level of centralization across a graph.

### 4.1.2.1 Quadratic approximation of the von Neumann entropy

While the von Neumann entropy of a graph has found many applications in the analysis of real-world networks [20, 82, 196, 206], a major drawback of this entropic measure is the fact that it requires the computation of the eigenvalues of the (normalized) graph Laplacian. This has computational complexity which is cubic in the number of nodes of the network, thus making the application to large networks unfeasible.

For this reason, a number of researchers resorted to a quadratic approximation of the entropy [82, 114, 206]. Although this only captures simple degree statistics of the graph, Han et al. [82] show that for Erdös-Rényi, scale-free, and Delaunay graphs this is a sufficiently good approximation. However their analysis is limited to the normalized Laplacian entropy, and does not consider the unnormalized version. In fact, to the best of our knowledge, no previous study has investigated the difference between the Laplacian and the normalized Laplacian entropies. Interestingly, despite a lack of evidence suggesting that one formulation should be preferred to the other, most works in the literature make use of the normalized version [20, 82, 196, 206].

One of the main aims of this paper is indeed that of shedding light on the differences between these two formulations. To this end, we rewrite the Shannon entropy $-\sum_i \lambda_i \ln(\lambda_i)$ using the second order polynomial approximation $k\sum_i \lambda_i(1 - \lambda_i)$, where the value of $k$ depends on the dimension of the simplex. Given a graph $G$, let $\rho(G)$ denote its associated density matrix, i.e., $\rho(G)$ is either $\rho(L)$ or $\rho(\mathscr{L})$. We obtain

$$S(\rho) = -\mathrm{Tr}\left(\rho \ln \rho\right) \approx \mathrm{Tr}\left(\rho(I_n - \rho)\right), \tag{4.4}$$

where $n$ is the number of nodes of $G$, $I_n$ is the $n \times n$ identity matrix and we ignored the node set size-dependent factor $\frac{|V|\ln(|V|)}{|V|-1}$.

In the next subsections we look at the specific form of these approximations in the case of the Laplacian and normalize Laplacian entropies. We also derive the expressions for the change in approximated entropy when a single edge is added to the graph, which in turn allows us to shed light on the type of structures that lead to maximal entropy changes.

**4.1.2.1.1 Laplacian** We start by considering the Laplacian entropy. Recall that in this case $\rho(L) = \frac{L}{2m}$, where $m$ denotes the number of edges of $G$. Using simple algebra, we can rewrite Eq. 4.4 as

$$S(\rho(L)) \approx 1 - \frac{1}{2m} - \frac{1}{4m^2} \sum_{v \in V} d_v^2 \tag{4.5}$$

In other words, the quadratic approximation of the Laplacian entropy can be expressed in terms of simple degree statistics. More interestingly, this allows us to probe into the behaviour of the (approximated) Laplacian entropy as the edge set of the graph grows. This was already investigated numerically in Passerini and Severini [140], but the quadratic approximation allows us to get a deeper analytical insight, although dependent on the approximation.

Let $\Delta(\rho(L)) = S(\rho(L_{\cup(x,y)})) - S(\rho(L))$ be the increment in entropy when a new edge is added to a graph $G$. From Eq. 4.5, we see that

$$\Delta(\rho(L)) \approx \frac{1}{2m} + \frac{1}{4m^2} \sum_{v \in V} d_v^2 - \frac{1}{2(m+1)} - \frac{1}{4(m+1)^2} \left( \sum_{v \neq x,y} d_v^2 + (d_x + 1)^2 + (d_y + 1)^2 \right)$$

$$= \frac{1}{2m(m+1)} - \frac{1}{4m^2(m+1)^2} \left( (m+1)^2 \sum_{v \in V} d_v^2 - m^2 \left( \sum_{v \neq x,y} d_v^2 + (d_x + 1)^2 + (d_y + 1)^2 \right) \right)$$

$$= \frac{1}{2m(m+1)} - \frac{1}{4m^2(m+1)^2} \left( (2m+1) \sum_{v \in V} d_v^2 - 2m^2(d_x + d_y + 1) \right)$$

$$= \frac{1}{2m(m+1)} - \frac{2m+1}{(m+1)^2} \left( 1 - \frac{1}{2m} - \tilde{S}(\rho(L)) \right) - \frac{d_x + d_y + 1}{2(m+1)^2}$$

$$= - \frac{d_x + d_y}{2(m+1)^2} - \frac{1 + (2m+1)(1 - \tilde{S}(\rho(L)))}{(m+1)^2}, \tag{4.6}$$

where $\tilde{S}(\rho(L))$ denotes the approximated Laplacian entropy. Eq. 4.6 indicates that edges connecting low degree nodes produce the maximum increment in the graph entropy, while connecting high degree nodes has the opposite effect. This in turn suggests that highly regular graphs with low average degree will be assigned higher values of the approximated Laplacian entropy. This is also the case for the exact version of the Laplacian entropy, as shown in Section 4.2.4.

Note, however, that this does not explain the emergence of structures such as long paths, connected components, and non-trivial symmetries observed by Passerini and Severini in the Laplacian entropy [140] and confirmed in our experimental evaluation. Indeed, the quadratic approximation provides an interesting but incomplete picture of the structural patterns captured by the Laplacian entropy.

**4.1.2.1.2 Normalized Laplacian** We now consider the normalized Laplacian entropy. In this case $\rho(\mathscr{L}) = \frac{\mathscr{L}}{n}$, where $n$ denotes the number of nodes of $G$. We can rewrite Eq. 4.4 as

$$S(\rho(\mathscr{L})) \approx 1 - \frac{1}{n} - \frac{1}{n^2} \sum_{(u,v) \in E} \frac{1}{d_u d_v}, \tag{4.7}$$

as previously observed by Han et al. [82].  As for the Laplacian entropy, the quadratic approximation is based on simple degree statistics.  Unlike the approximated Laplacian entropy, however, Eq. 4.7 shows that the approximated normalized Laplacian entropy is defined in terms of degree statistics for pairs of nodes that are connected by edges.

As in the previous subsection, we now turn our attention to the increment in entropy when the edge set of $G$ grows.  Let $\Delta(\rho(\mathscr{L})) = S(\rho(\mathscr{L}_{\cup(x,y)})) - S(\rho(\mathscr{L}))$ denote this increment.  Then let $N_x$ and $N_y$ denote set of vertices connected to $x$ and $y$ in $G$ (before introducing the edge $(x, y)$), respectively.  We have that

$$
\begin{aligned}
\Delta(\rho(\mathscr{L})) &\approx -\frac{1}{n^2}\Bigg( \sum_{v \in N_x} \frac{1}{d_v(d_x + 1)} + \sum_{v \in N_y} \frac{1}{d_v(d_y + 1)} + \frac{1}{(d_x + 1)(d_y + 1)} \\
&\qquad - \sum_{v \in N_x} \frac{1}{d_v d_x} - \sum_{v \in N_y} \frac{1}{d_v d_y} \Bigg) \\
&= -\frac{1}{n^2}\Bigg( \frac{d_x - (d_x + 1)}{d_x(d_x + 1)} \sum_{v \in N_x} \frac{1}{d_v} + \frac{d_y - (d_y + 1)}{d_y(d_y + 1)} \sum_{v \in N_y} \frac{1}{d_v} + \frac{1}{(d_x + 1)(d_y + 1)} \Bigg) \\
&= \frac{1}{n^2}\Bigg( \frac{1}{d_x(d_x + 1)} \sum_{v \in N_x} \frac{1}{d_v} + \frac{1}{d_y(d_y + 1)} \sum_{v \in N_y} \frac{1}{d_v} - \frac{1}{(d_x + 1)(d_y + 1)} \Bigg) \\
&= \frac{1}{n^2}\Bigg( \frac{1}{(d_x + 1)H(d_{N_x})} + \frac{1}{(d_y + 1)H(d_{N_y})} - \frac{1}{(d_x + 1)(d_y + 1)} \Bigg),
\end{aligned} \tag{4.8}
$$

where $H(d_{N_x})$ and $H(d_{N_y})$ denote the harmonic means of the degrees of the vertices in $N_x$ and $N_y$, respectively.  Compared to Eq. 4.6, Eq. 4.8 shows a more complex relation between the node degrees and the graph entropy. The third term of the last line of Eq. 4.8 drives the entropy change in the opposite direction of Eq. 4.6, as maximizing (minimizing) the entropy requires establishing connections between high (low) degree nodes. The first two terms, on the other hand, highlight the importance of the neighbourhood of the nodes being connected, with the connection of pairs of low degree nodes with low average degree neighbourhoods yielding the maximum increment in the entropy of the graph.


### 4.1.2.2  Discussion

The analysis of the quadratic approximations of the two entropies suggests that these may be only weakly correlated, if not perhaps negatively correlated, depending on the topology of the underlying graphs. Note that simply looking at Eq. 4.5 and Eq. 4.7 one may conclude that the correlation between the quadratic approximations of the Laplacian and normalized Laplacian entropies should be negative. However the actual relation is more subtle, and it is better understood through Eqs. 4.6 and 4.8. In fact, note that while Eq. 4.5 involves a summation over the nodes of the graph, Eq. 4.7 involves a summation over its edges, thus making the relation between the two quantities more complex. Indeed, the negative correlation suggested by Eqs. 4.5 and 4.7 is also observed when examining Eqs. 4.6 and 4.8, though the second pair of equations reveals a more subtle relation be-
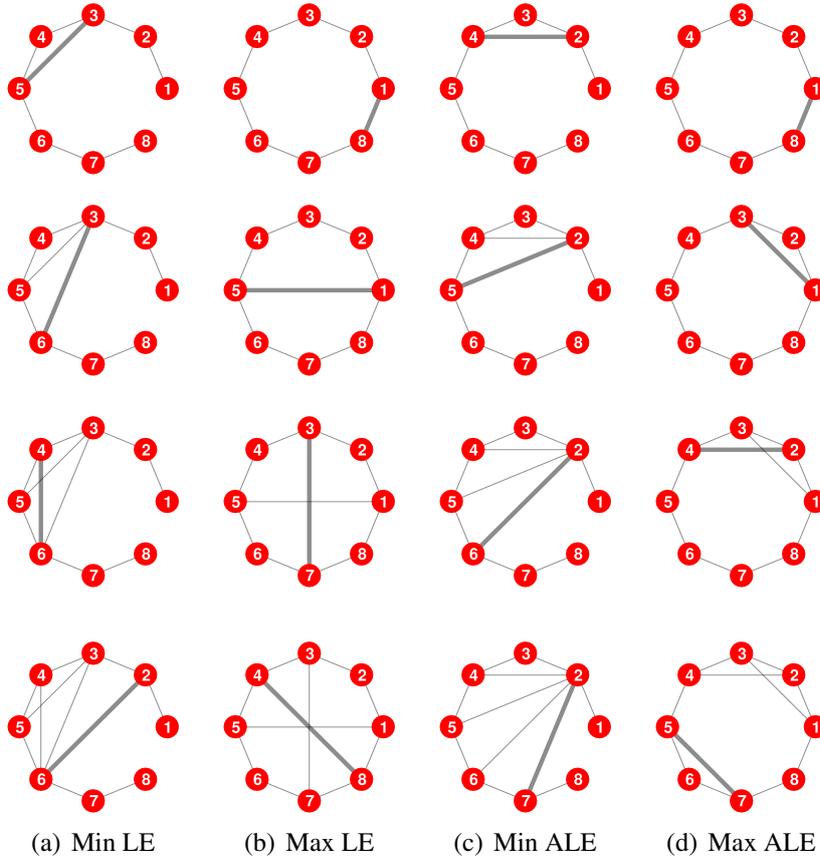
Figure 4.1: Evolution of the edge structure of a path graph over 8 nodes when we iteratively add edges that (a) minimize and (b) maximize the Laplacian entropy (LE). (c) and (d) show similar results for the approximate Laplacian entropy (ALE).

tween the two entropies, with the degree distribution of the nodes neighbourhoods playing an important role.

As for the exact version of the entropies, it is harder to draw any conclusion on their relation as we do not know what type of structural information (beyond simple degree statistics) is being lost in the approximation. In the next section we aim to answer the following questions: 1) are the Laplacian and normalized Laplacian capturing similar structural patterns? and 2) can we rely on the quality of their quadratic approximations when the high computational complexity of the exact version becomes an issue? To answer this questions, in the next section we run an extensive set of numerical experiments on both synthetic and real-world graphs.

### 4.1.3 Experiments

In the previous sections we have introduced the concepts of (normalized) Laplacian entropy of a graph and its quadratic approximation. This in turn provided us with a partial

(a) Min NLE  (b) Max NLE  (c) Min ANLE  (d) Max ANLE
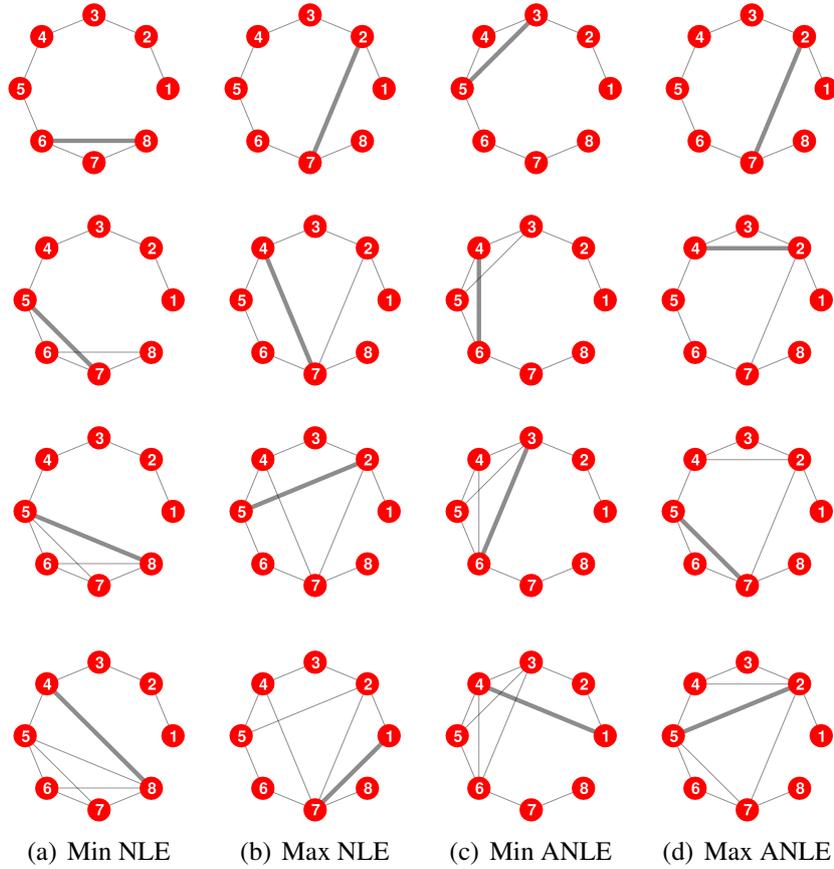
Figure 4.2: Evolution of the edge structure of a path graph over 8 nodes when we iteratively add edges that (a) minimize and (b) maximize the normalized Laplacian entropy (NLE). (c) and (d) show similar results for the approximate normalized Laplacian entropy (ANLE).

intuition of the relation between graph structure and entropy. In this section we aim to validate these initial intuitions with an extensive set of experiments and to investigate further the relation between the normalized and unnormalized Laplacian entropies, as well as the quality of their quadratic approximation.

### 4.1.3.1 Entropy-driven Graph Evolution

We commence by investigating how the structure of a graph changes as we add new nodes and edges to it. To this end, we introduce a simple growth model where new connections are established if they maximise (minimise) the graph entropy. We perform the same experiment for both the Laplacian and the normalized Laplacian entropy, as well as their quadratic approximations.

(a) Erdös-Rényi  (b) Scale-free  (c) Small world
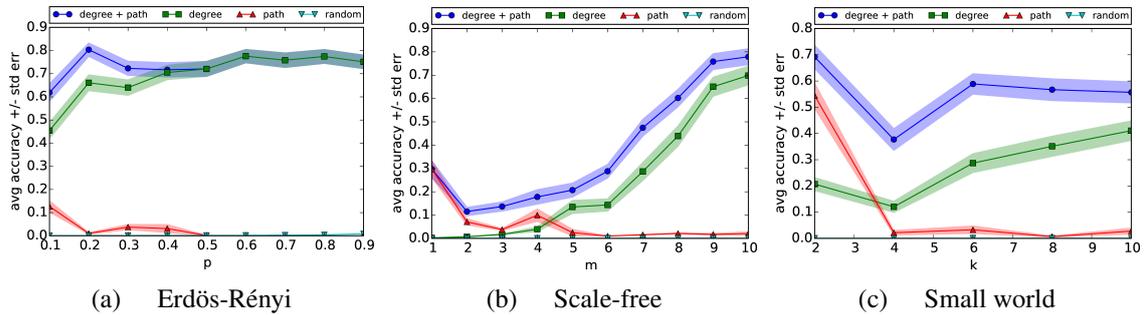
Figure 4.3: Average accuracy when predicting the edge whose addition maximizes the Laplacian entropy when using four different heuristics, i.e., choose (1) the pair of nodes with minimum degree sum; (2) the pair of nodes with maximum geodesic distance; (3) the pair of nodes with minimum degree sum and maximum geodesic distance; (4) a pair of nodes $u$ and $v$ picked at random, with $(u, v) \notin E$.

### 4.1.3.2 Edge Growth Model

We first consider the case where the number of nodes is fixed and new edges are iteratively added to the graph. Fig. 4.1 shows the first four stages of the evolution of a graph with eight nodes where the growth process is driven by the Laplacian entropy. Each column of Fig. 4.1 corresponds to a different choice of the process (maximization or minimization) and entropy (exact or approximated). Similarly, Fig. 4.2 shows the results for the normalized Laplacian entropy.

Figs. 4.1(c) and (d) confirm what already observed when looking at the quadratic approximation of the Laplacian entropy. Indeed, edges that maximize the approximate Laplacian entropy are edges that connect low degree nodes, as shown in Fig. 4.1(d). In contrast to Fig. 4.1(b), where we maximize the exact entropy, in Fig. 4.1(d) all pairs of nodes with minimum degree sum have the same probability of being connected. This is not the case in Fig. 4.1(b), where, given two pairs of nodes with equal sum of their degrees, the pair of nodes with the highest geodesic distance[1] leads to a higher increment in the entropy. As a result of this, each new edge added in the Fig. 4.1(b) seems to act as an axis of symmetry. While it would be tempting to argue that the latter is evidence of structural symmetries being picked up by the exact Laplacian entropy as opposed to its approximated version, a quick numerical investigation proves that this hypothesis is incorrect.

It is also not true, for a general graph, that the pair of nodes being connected is always one with minimum degree sum and maximum distance. However the distance between the nodes being connected clearly plays a role, together with a number of yet unspecified structural properties. To show this, we perform the following experiment. Starting from a random graph $G$, we use four different heuristics to predict what edge will lead to the

---

[1] Recall that the geodesic distance between two nodes $u$ and $v$ is the number of edges in the shortest path connecting $u$ and $v$.

maximum increment of the Laplacian entropy. Each of the four heuristics selects the pairs of nodes that optimize the following measures, respectively: (1) the pair of nodes with minimum degree sum (which corresponds to the structural information contained in the approximated Laplacian entropy); (2) the pair of nodes with maximum geodesic distance; (3) the pair of nodes with minimum degree sum and maximum geodesic distance; (4) a pair of nodes $u$ and $v$ picked at random, with $(u, v) \notin E$. The prediction accuracy of a heuristic is computed as the fraction of edges it identified correctly. Fig. 4.3 shows the results for the different random graph models, the Erdös-Rényi model, the Watts–Strogatz model, and the Preferential Attachment model [22] (see Section 4.1.3.4 for a detailed description of the models and their parameters). In all cases, the addition of the path length information leads to a significant increment in the accuracy of the heuristic which solely looks for the pair of nodes with minimum degree sum. In other words, both degree statistics (captured by the quadratic approximations) and path length information are important structural patterns captured by the exact version of the Laplacian entropy. Note also that as the graphs becomes denser (which, given a fixed number of nodes, for the three random models considered correspond to increasing values of $p$, $m$, and $k$, respectively) the path length information loses importance. This is due to the fact that for sufficient higher densities all pairs of nodes lie at the same distance from each other.

We also compute a number of statistics that capture different structural properties of the graph during its evolution, namely the average shortest path length, the index of dispersion of the degree distribution, and the average clustering coefficient, as shown in Fig. 4.4. Recall that the index of dispersion of a distribution measures the ratio of its variance to its mean, and the clustering coefficient quantifies the degree to which nodes in a graph tend to cluster together. Fig. 4.4 highlights once again the differences between the structural information captured by the Laplacian and the normalized Laplacian entropy, as well as their quadratic approximations. The tendency of the process which maximizes the exact Laplacian entropy to connect low degree nodes is particularly evident in the plots of the index of dispersion. As explained above, maximizing the (approximated) Laplacian entropy tends to create connections between low degree nodes. This in turn tends to create a regular structure where each node has the same degree, thus keeping the index of dispersion of the degree distribution low throughout the graph evolution. Note that this does not happen when maximizing or minimizing the (approximated) normalized Laplacian entropy. The difference between the exact Laplacian entropy and its approximated version is instead clear by looking at the average clustering coefficient. By connecting nodes that have both low degree sum and high distance, maximizing the exact Laplacian entropy keep the clustering coefficient as it effectively attempts to avoid creating triangles, at least in the first stages of the evolution. This however does not happen for the approximated Laplacian entropy, where the connection of two nodes with a common neighbour introduces a new triangle and thus increases the value of the average clustering coefficient.
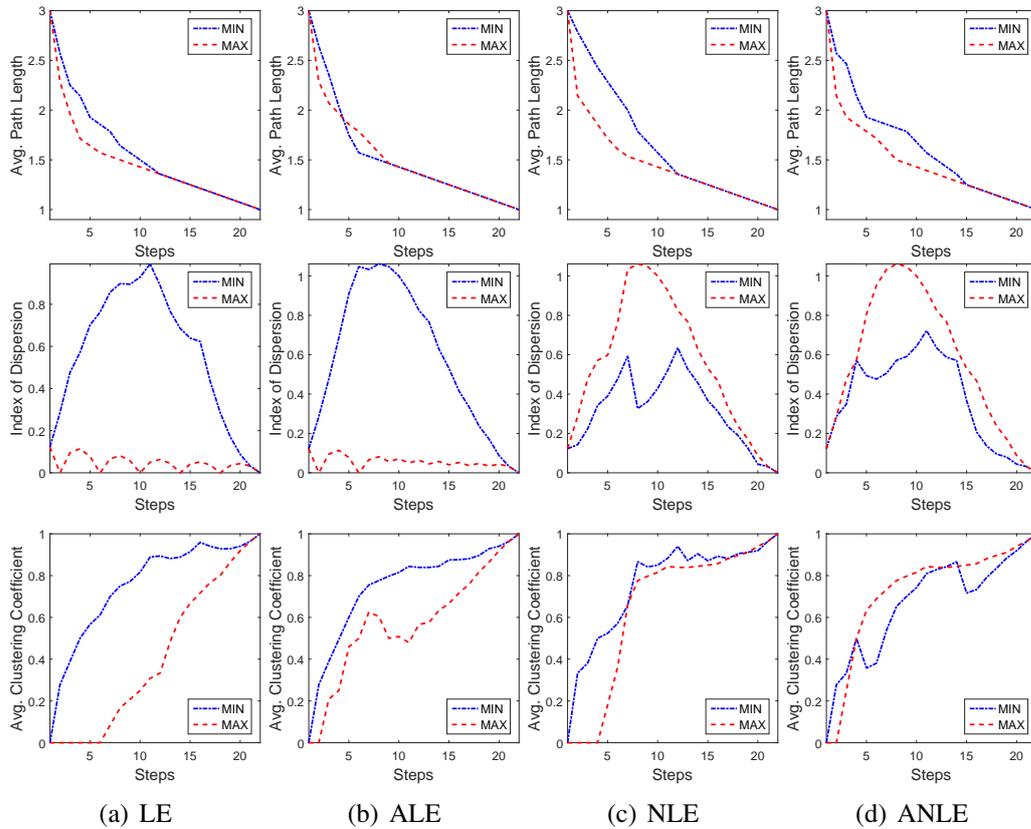
Figure 4.4: Top-to-bottom: Average path length, index of dispersion, and average clustering coefficient at different steps of the temporal evolution of the graphs in Fig. 4.1.

### 4.1.3.3 Node Growth Model

We also consider the case of graph where the both the number of nodes is not fixed over time. Instead, at each time step we add a new node and we connect it to the $m$ nodes that lead to a maximal increment of the entropy. Each column in Fig. 4.5 corresponds to a different choice of the process (maximization or minimization) and entropy (exact or approximated). Similarly, we show the results of the same experiment for the normalized Laplacian entropy in Fig. 4.6. In both cases we start from a clique over three nodes. We only show the results for $m = 1$, as we observe the same behaviour for larger values of $m$. In contrast to the edge growth model, here maximizing (minimizing) the exact and approximated entropies yields the same structural evolution. Interestingly, while minimizing the (approximated) Laplacian entropy yields the formation of hubs, minimizing the (approximated) normalized Laplacian entropy leads to the formation of a long tail of low degree nodes. We observe opposite behaviour when maximizing the (approximated) Laplacian and normalized Laplacian entropy, respectively.
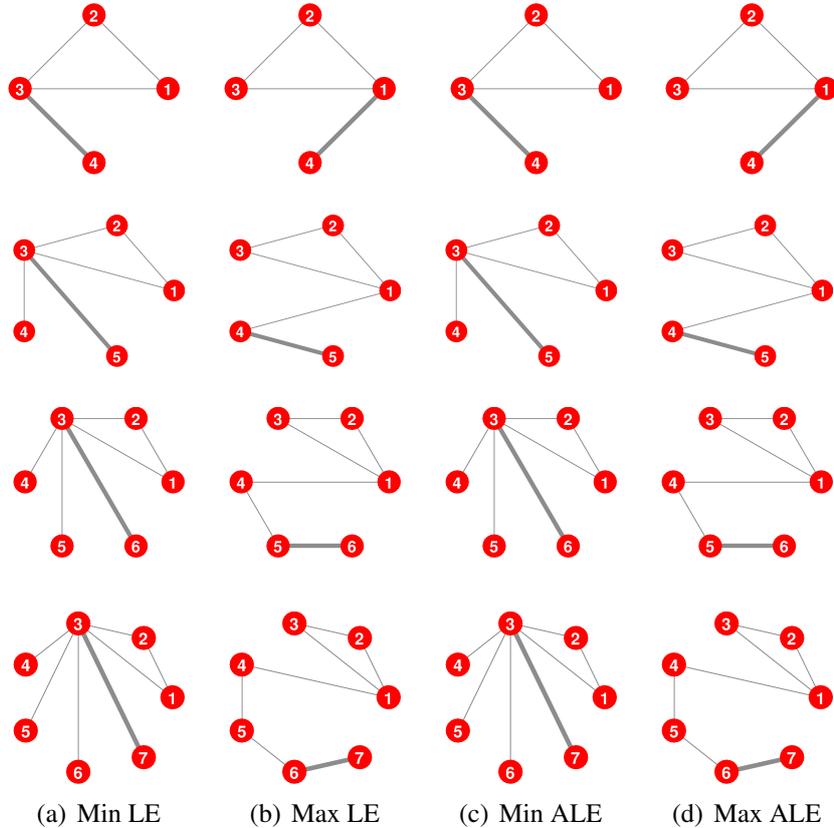
Figure 4.5: Evolution of the graph structure when we iteratively add a new node and we connect it to the graph so as to (a) minimize and (b) maximize the Laplacian entropy (LE). (c) and (d) show similar results for the approximate Laplacian entropy (ALE). Here the seed graph is a clique over 3 nodes.

**4.1.3.3.1   Experiments on Random Graphs**   While the previous experiments gave us some first interesting insights in the nature of the structural pattern captured by the (approximated) Laplacian and normalized Laplacian entropies, in this section we aim to perform a more thorough analysis of the two entropies and their approximated versions on a large set of synthetically generated graphs.

### 4.1.3.4   Datasets

We perform our experiments on synthetic networks generated by three well-known random graph models: 1) the Erdös-Rényi model, 2) the Watts–Strogatz model and 3) the Preferential Attachment model [22]. For each model we vary its parameters as explained below, except from the number of nodes, which is fixed to $n = 100$ in all three cases. This is to control for the well-known dependency between the value of the von Neumann entropy of a graph and its vertex set size, which would otherwise skew the results of our correlation study. This is particularly evident when our results are compared to those of

(a) Min NLE      (b) Max NLE      (c) Min ANLE      (d) Max ANLE
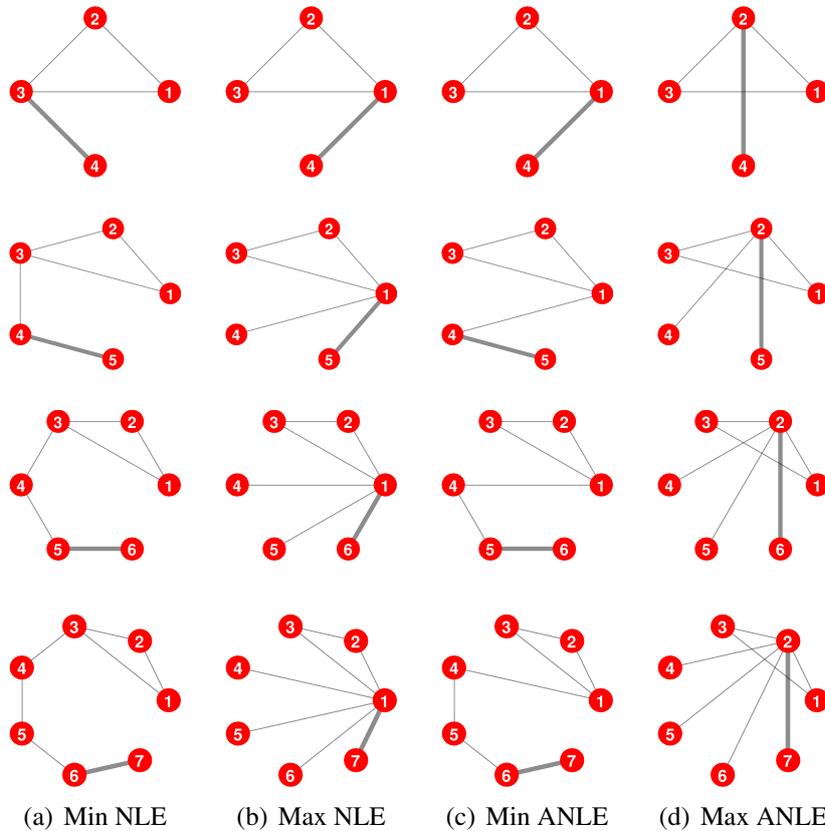
Figure 4.6: Evolution of the graph structure when we iteratively add a new node and we connect it to the graph so as to (a) minimize and (b) maximize the normalized Laplacian entropy (NLE). (c) and (d) show similar results for the approximate normalized Laplacian entropy (ANLE). Here the seed graph is a clique over 3 nodes.
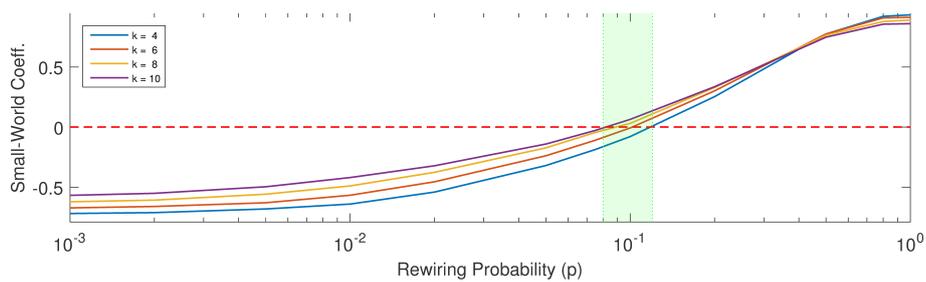


Figure 4.7: We choose the optimal value of $p$ in order to generate graphs displaying the small-world property [172].

Han et al. [82], where no such control was introduced.

**Erdös-Rényi model**: the graphs in this dataset are generated by varying the parameter $p$, namely the probability of connecting two nodes, between 0.1 and 0.9. Unless other-

wise stated, for each choice of $p$ we generate 100 instances, for a total of 900 graphs.

**Preferential Attachment model**: the parameter of this model is $m$, i.e. the number of edges to add from a new node to the existing nodes, at each temporal iteration. We let $m$ vary from 2 to 10, and, unless otherwise stated, we generate 100 instances for each choice of $m$, for a total of 900 graphs.

**Watts–Strogatz model**: here the model parameters are $k$ and $p$. Starting from a ring graph where each node is connected to its $k$ nearest neighbours, we rewire each edge with probability $p$. When $p = 0$, the graph is regular. As $p$ increases the graph structure becomes more random. We follow the quantitative metric presented in [172] to measure the small-worldness of a graph and we select the value of $p = 0.1$, as shown in Fig. 4.7. More precisely, in [172] the authors propose a way to measure the small-worldness of a network based on the original model described by Watts and Strogatz, comparing the network clustering coefficient to an equivalent lattice network and the path length to a random network. This in turn ensure that the generated graphs display the small-worldness property [172], i.e., they simultaneously have high clustering coefficient and low path length. As for the parameter $k$, we let it vary from 2 to 10. Unless otherwise stated, we generate 100 instances for each choice of $k$, for a total 900 graphs.

### 4.1.3.5   Correlation Analysis

With the synthetic graph datasets to hand, we perform a correlation study between the various version of the von Neumann entropy considered so far. More specifically, we measure the Pearson correlation coefficient (denoted as $\rho$ in Figs. 4.8-4.11) between 1) the approximate and exact Laplacian entropy, 2) the approximate and exact normalized Laplacian entropy, 3) the exact Laplacian entropy and the exact normalized Laplacian entropy, and 4) the approximate Laplacian entropy and the approximate normalized Laplacian entropy. Note that, for each model and each choice of the model parameters (see Section 4.1.3.4), we generate 1000 graphs.

   Fig. 4.8 shows the results of the correlation analysis on the Erdös-Rényi graphs. The first column refers to the Laplacian entropy, the second one to the normalized Laplacian entropy, whereas the third and fourth columns concern the approximate and exact formulation. Here, we consider three choices of $p$. We first observe that there exists a strong correlation between the exact and approximate versions of each entropy. On the other hand, when we compare the normalized Laplacian entropy and unnormalized Laplacian entropy, both in their exact and approximate forms, the correlation becomes weaker. Indeed, as observed in the previous section, we expect the quadratic approximations of the two entropies to show a weak positive, or potentially negative, correlation. We posit that the weak positive correlation observed for Erdös-Rényi graphs is a consequence of the degree distribution of the nodes neighbourhoods being close to uniform. Interestingly, here we observe that this result holds also for the exact versions of the entropies. This in
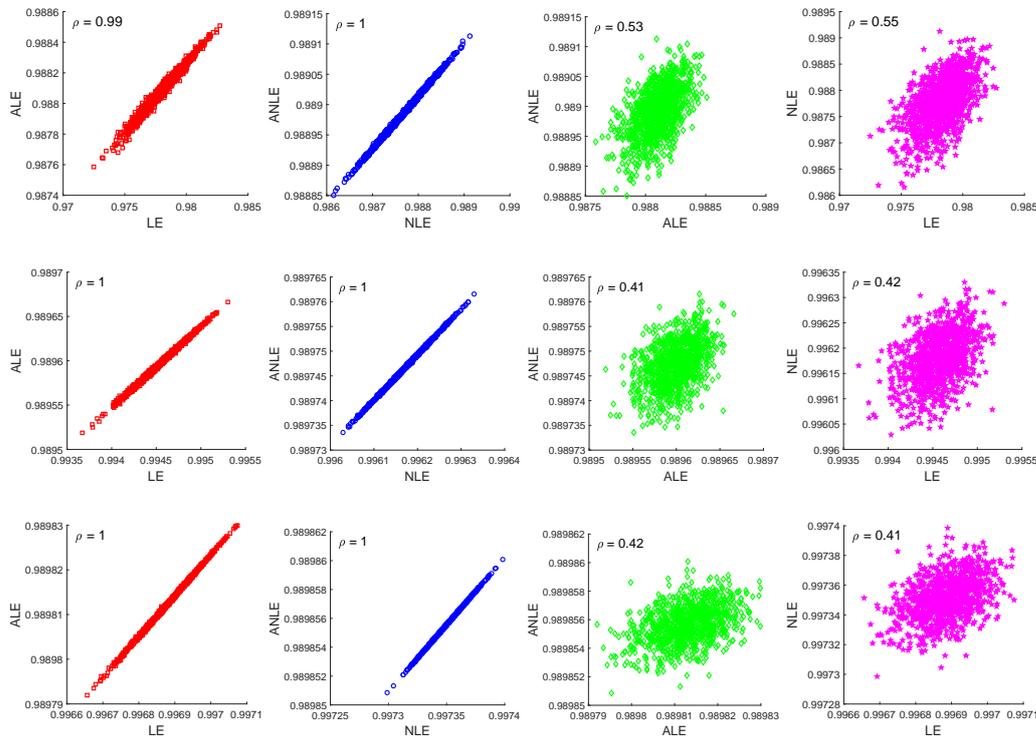
Figure 4.8: Entropies correlations on the Erdös-Rényi graphs for $p = 0.1$ (top), $p = 0.4$ (middle), and $p = 0.7$ (bottom).
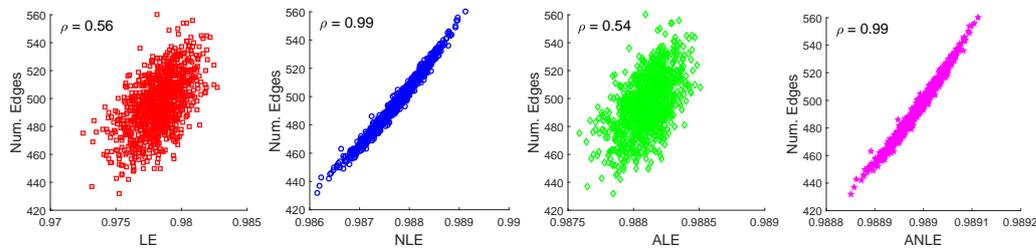


Figure 4.9: Correlation between entropy and number of edges for Erdös-Rényi graphs with $p = 0.1$.

turn suggests that the structural patterns captured by the two entropies are not necessarily the same. This is an important observation, as it implies, for example, that when using the von Neumann entropy in pattern recognition applications swapping one entropy for the other is likely not to give the same result.

As for the strong correlation observed between the exact and approximate version of the normalized Laplacian entropy, this is likely to be due to the tight relationship with the number of edges of a graph and its normalized Laplacian entropy, as shown in Fig. 4.9.
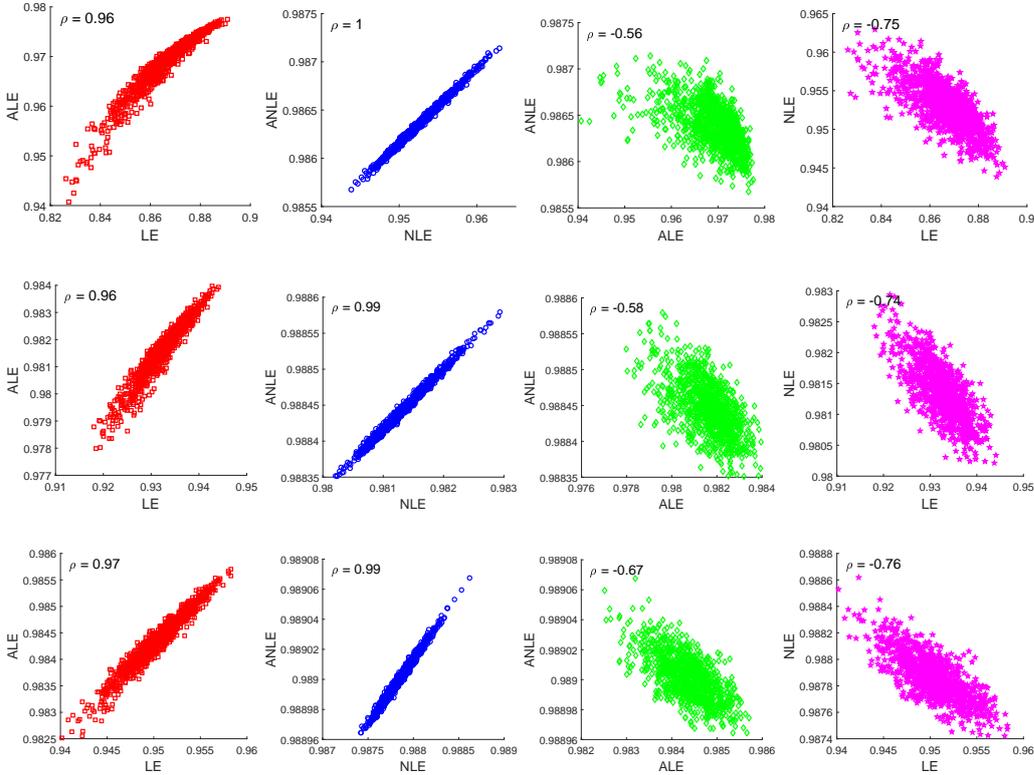
Figure 4.10: Entropies correlation on the scale-free graphs for $m = 1$ (top), $m = 3$ (middle), and $m = 5$ (bottom).

More precisely, in Fig. 4.9 we show the correlation between the number of edges of a graph and its entropy for Erdös-Rényi graphs with $p = 0.1$. Indeed, Eq. 4.7 suggests a strong dependency between the number of edges of a graph and the quadratic approximation of the normalized Laplacian entropy. Note however that we do not observe a strong correlation between the (approximate) Laplacian entropy and the number of edges.

We then continue the correlation study on the set of scale-free graphs generated by the Preferential Attachment model. The results are shown in Fig. 4.10. On the one hand, when we consider the relationship between the approximate and the exact versions of the entropies, we observe a similar behaviour to that seen for the Erdös-Rényi graphs, with a strong correlation for both the Laplacian and the normalized Laplacian entropy with their quadratic approximations. Note that in this case, given a pair of values for $n$ and $m$, the number of edges of the generated graphs does not vary, so the observed effect cannot be explained by a varying edge set size.

On the other hand, in this case we observe a negative correlation between the two entropies, both in their exact and approximated versions. The correlation is stronger between the approximated entropies. Indeed, in Eq. 4.5 the term $-\sum_{u \in V} d_u^2$ prevails, whereas in Eq. 4.7 the leading term is $-\sum_{(u,v) \in E} \frac{1}{d_u d_v}$. In other words, when a graph contains very
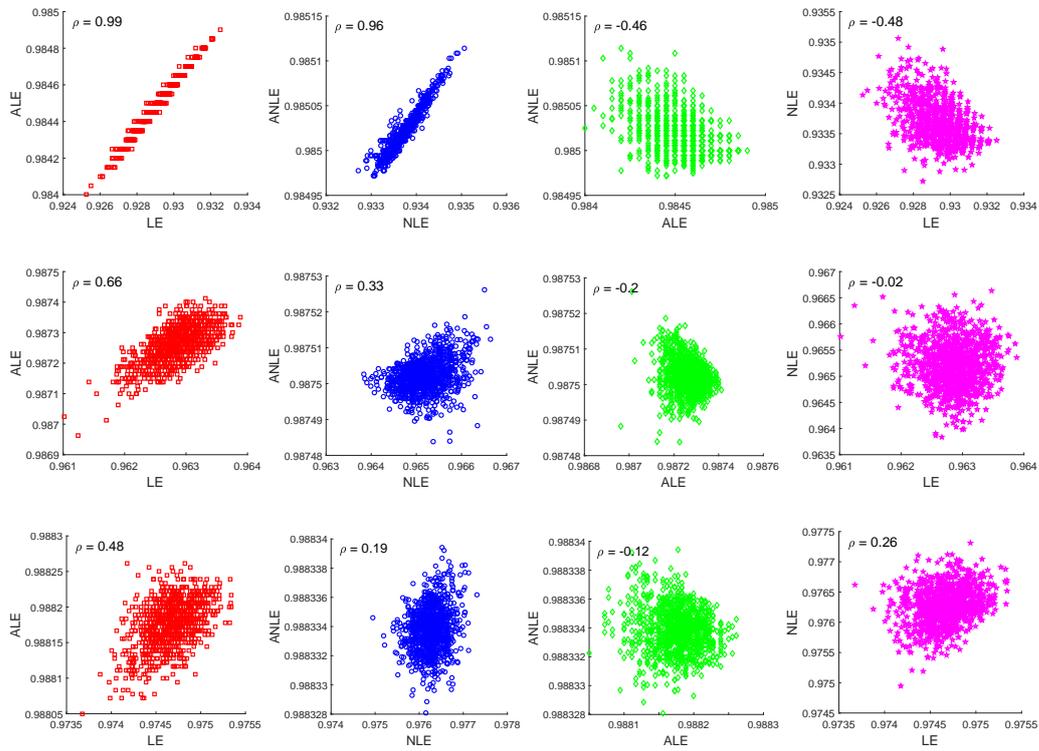
Figure 4.11: Entropies correlation on the small-world graphs for $k = 2$ (top), $k = 4$ (middle), and $k = 6$ (bottom).
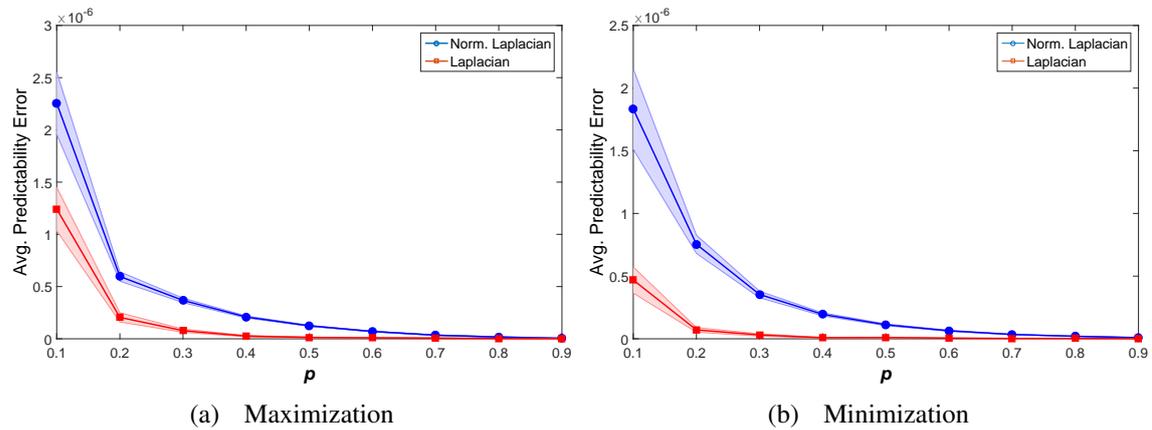


Figure 4.12: Average predictability error ($\pm$ standard error) on Erdös-Rényi graphs.

high degree nodes, the Laplacian entropy becomes very small while the normalized counterpart tends to increase.

We conclude this correlation study on the small-world graphs generated by the Watts–Strogatz model. As for the scale-free graphs, note for a choice of the parameters $n$ and

$k$ the number of edges in the generate graphs does not vary. Fig. 4.11 shows a stark contrast between the results obtained for $k > 2$ and those obtained for $k = 2$. To understand why this happens, recall that $k$ controls the number of neighbours for each node in the initial ring graph. The higher the value of $k$, the more robust the graph structure is to the edge flips that turn the ring into a small-world graph by reducing the average path length. The result is a quasi-regular ring lattice structure with relatively uniform degree where the approximate entropy. Since the approximated entropies only capture simple degree statistics, they are unable to capture the structural differences observed by the exact entropies, which go beyond structural information at degree level. As a consequence, the correlation between the exact entropies and the approximated ones decreases as $k$ increases, until the two are practically uncorrelated. However this does not happen when $k = 2$. In fact, in this case the regularity of the initial ring graph is easily disrupted by the noise addition process, with the removal and addition of a few edges causing significant deviations from the initial lattice structure.

### 4.1.3.6　Edge Predictability

From the previous analyses it is clear that the quality of the quadratic approximations of exact entropies depend on the topology of the underlying graphs. We have also seen that the Laplacian and the normalized Laplacian entropies generally capture different types of structural information. We now take our investigation one step further and we look at entropic contribution of a single edge, when either the approximate entropy or the exact entropy ares used. Previous works [114] have looked at the entropic content of an edge as a way to measure its centrality. More in general, our interest is again to understand how well the quadratic approximations of the Laplacian and normalized Laplacian entropies are able to capture the contributions of single edges to the overall graph entropy.

　　We generate three sets of synthetic graphs as described in 4.1.3.4. Let $G = (V, E)$ be one such graph with edge set $E \subseteq V \times V$, where $V$ denotes the node set. For each edge not in $E$, we calculate the increase in entropy obtained by adding that connection to the graph, both for the exact and the approximate form of the entropy. Let $list^{\mathscr{E}}$ and $list^{\mathscr{A}}$ be two edge-indexed lists containing the values of the exact and the approximate entropic increases, respectively. Given $list^{\mathscr{A}}$, we choose the index $iM$ of the edge that leads to the maximum entropic increment, and the index $im$ of the edge that leads to the minimum entropic increment. With these indices to hand, we select the corresponding value of the entropic increment for the same edges in the exact case, i.e., $list^{\mathscr{E}}_{iM}$ and $list^{\mathscr{E}}_{im}$. Then, we define the predictability error for the edge that leads to the maximum entropic increment as

$$Prd_M = 1 - \frac{list^{\mathscr{E}}_{iM}}{\max(list^{\mathscr{E}})}. \qquad (4.9)$$

Similarly, we define the predictability error for the edge that leads to the minimum entropic increment as

$$Prd_m = 1 - \frac{\min(list^{\mathscr{E}})}{list^{\mathscr{E}}_{im}} \qquad (4.10)$$

(a)   Maximization

(b)   Minimization
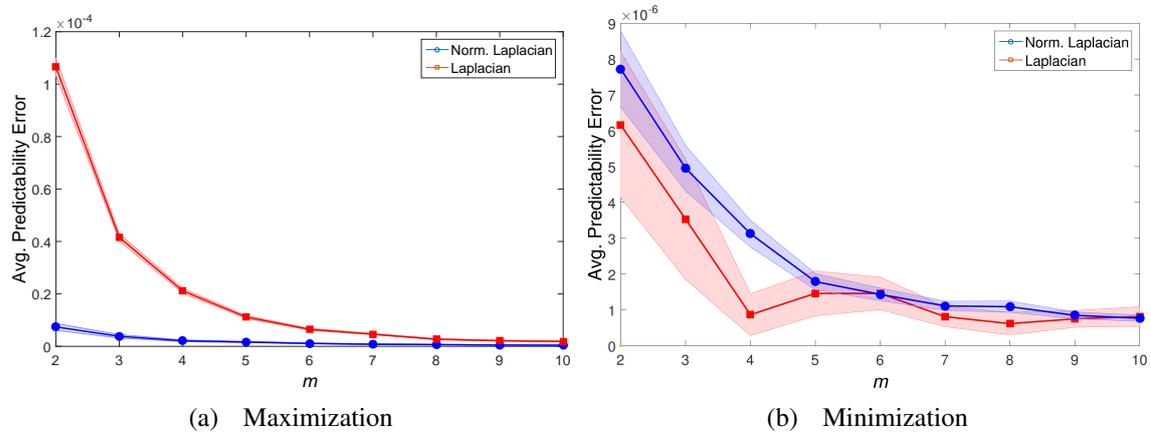
Figure 4.13: Average predictability error ($\pm$ standard error) on scale-free graphs.

In both cases, the smaller the predictability error the better we are able to approximate the exact maximum using the quadratic entropy.

Figs. 4.12, 4.13, and 4.14 show how the error changes as we vary the model parameters of the Erdös-Rényi, Preferential Attachment Watts-Strogatz model, respectively. We observe that in general, regardless of the model, the error tends to decrease. In other words, as graphs become denser the number of non-existing edges decreases and thus it becomes easier to correctly identify the edges associated to the maximum (minimum) entropic increment. The only exception is that of the Watts-Strogatz model, where the error first increases and then decreases, as shown in Fig. 4.14(a). Note that this fits with our previous observation of a higher correlation between the approximate and exact entropies for this type of graphs when $k = 2$. However while in the correlation study we observe that for $k > 2$ the correlation decreases, in this case the graph densification appears to dominate and to lead to a decrease of the observed predictability error.

Interestingly, we see that the predictability error is significantly lower when we are trying to approximate the Laplacian entropy as opposed to the normalized Laplacian entropy, with the only exception being that of scale free graphs. This is probably due to an ambiguity created by the approximate version. Indeed, according to Eq. 4.6, in order to maximize the approximate Laplacian entropy, two nodes with low degree should be linked. However, when we take into account networks whose degree distribution follows a power law, the choice gets nearly random. More specifically, the formula does not consider the node neighbourhood and thus two nodes belonging to the same hub or two nodes belonging to different hubs may be indistinctly connected. However, the exact version could be making a distinction and favouring the connection between nodes belonging to different hubs. This in turn could be explained as an effort to connect distant nodes, already observed in Fig. 4.3. Such an ambivalence between the approximate and exact Laplacian entropy eventually leads to a poor predictability. On the other hand, Fig. 4.13(b) shows no substantial difference between the two entropies because there is

(a) Maximization              (b) Minimization
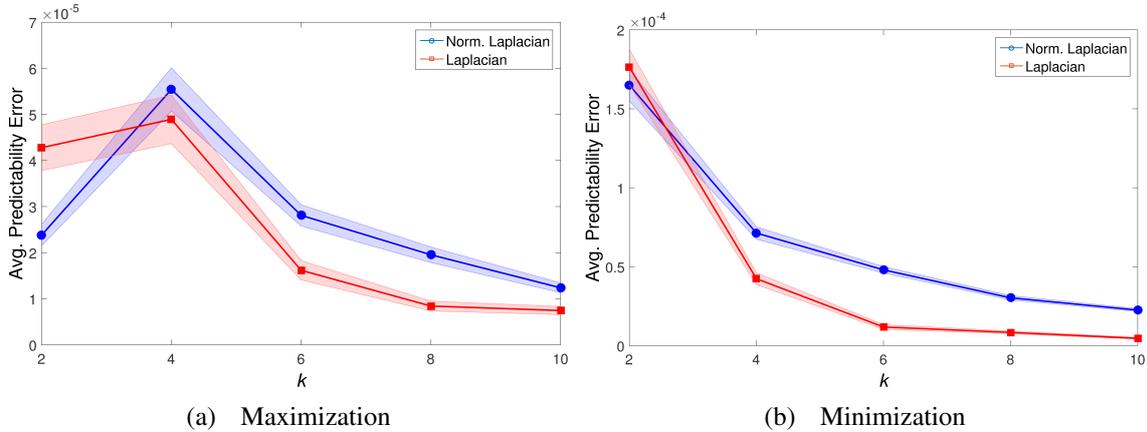
Figure 4.14: Average predictability error ($\pm$ standard error) on small-world graphs.

less uncertainty in choosing which pair of nodes (with high degree) to connected.

### 4.1.3.7   Experiments on Real-world Networks

We conclude our analysis by considering networks extracted from real-world complex systems.

### 4.1.3.8   Datasets

*Dataset* **1:** the USSM dataset is extracted from a database consisting of the daily prices of 431 companies in 8 different sectors from the New York Stock Exchange (NYSE) and the Nasdaq Stock Market (NASDAQ). To construct the dynamic network, 431 stocks with historical data from January 1995 to December 2016 are selected. The dataset is arranged to be around 5500 trading days. In order to build an evolving network, a time window of 28 days is used and it is moved along time to obtain a sequence (from day 29 to day 5500); in this way, each temporal window contains a time-series of the daily return stock values over a 28 day period. Afterward, trades among the different stocks are set as a network. For each time window, we compute the cross correlation coefficients between the time-series for each pair of stocks and create connections between them if the absolute value of the correlation coefficient exceeds a threshold. The result is a stock market network which changes over the time, with a fixed number of 431 nodes and varying edge structure for each of trading days.

*Dataset* **2:** this dataset collects protein-protein interaction PPI networks related to histidine kinase [2]. Histidine kinase is a key protein in the development of signal transduction. The graphs describe the interaction relationships between histidine kinase in different

---

[2]Lars J Jensen, Michael Kuhn, Manuel Stark, Samuel Chaffron, Chris Creevey, Jean Muller, Tobias Doerks, Philippe Julien, Alexander Roth, Milan Simonovic, et al. String global view on proteins and their functional interactions in 630 organisms. Nucleic acids research 2009
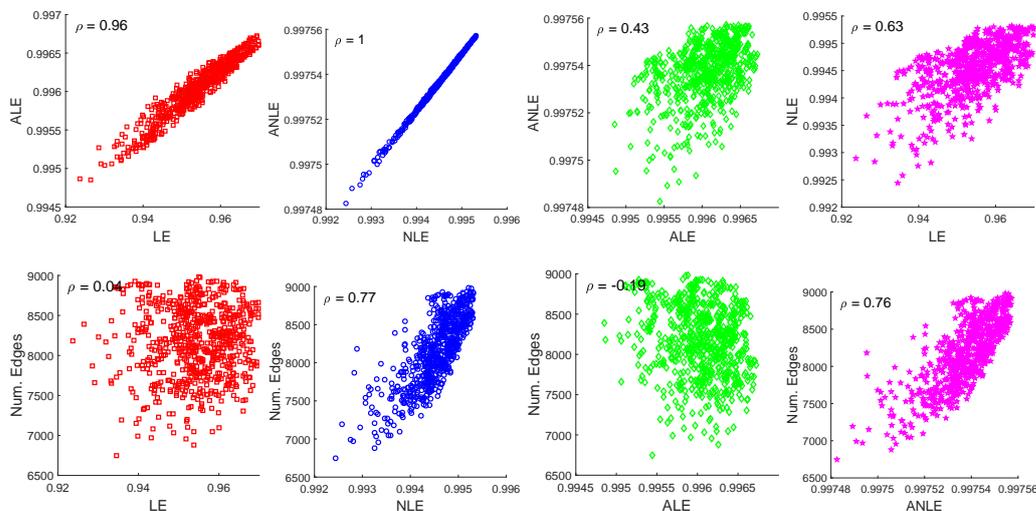
Figure 4.15: Correlation between the entropies (top) and entropy and number of edges (bottom) on the USSM dataset.

species of bacteria. If two proteins (graph nodes) have direct (physical) or indirect (functional) association, they are connected by an edge. PPIs are collected from 5 different kinds of bacteria with the following evolution order (from older to more recent): 4 PPIs from *Aquifex aelicus* and 4 PPIs from *Thermotoga maritima*, 52 PPIs from Gram-Positive *Staphylococcus aureus*, 73 PPIs from Cyanobacteria *Anabaena variabilis* and 40 PPIs from Proteobacteria *Acidovorax avenae*.

### 4.1.3.9   Correlation Analysis

The USSM dataset contains a time-evolving complex network consisting of graphs having components of different sizes. Thus, we selected only some defined instances among the available ones. Specifically, we chose 707 samples. Each sample has to satisfy two requirements: a) being a connected graph, b) having maximum size (431 nodes). The correlation plots between the entropies are shown in Fig. 4.15 (top), where $\rho$ denotes the Pearson correlation coefficient. We first observe that the correlation is always strong. This is presumably due to the fact all instances belong to the same time-varying process, making them intrinsically correlated to each other. However, it is worth recalling that another factor may be causing this dependence. We have already stressed that the entropy can be influenced by the volume of a graph (i.e., the number of node it contains) as well as by its density (i.e., the number of edges it contains). While in this case the volume is fixed, the density changes over time. Indeed, in Fig. 4.15 (bottom) we see that the normalized Laplacian entropy of is highly correlated with the graphs density, in accordance to what already observed in Fig. 4.8.

The PPI dataset consists of connected graphs with varying number of nodes. Due to the limited number of graphs in the PPI dataset, we prefer not to restrict our analysis to
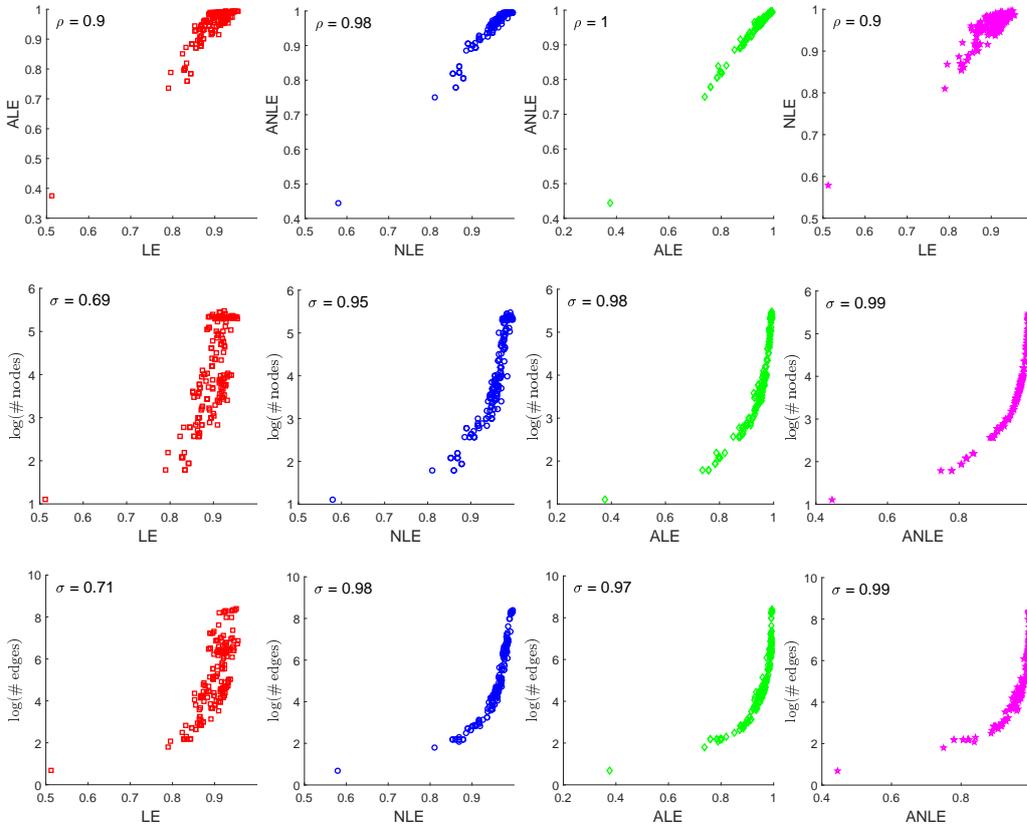
Figure 4.16: Correlation between entropies (top), entropy and number of nodes (middle) and edges (bottom) on the PPI dataset.

same size graphs. Fig. 4.16 (top) shows the correlation plots for the PPI dataset. Once again, we observe a high (Pearson) correlation between all pairs of entropies. With the exception of the Laplacian entropy, this appears to be largely due to the correlation between the entropy of a graph and its number of edges and nodes (Fig. 4.16, middle and bottom). Finally, note that in Fig. 4.16 $\sigma$ denotes the Spearman rank correlation coefficient. Indeed, we observed the existence of a non-linear relation between the entropy and the number of edges (nodes).

### 4.1.4   Conclusion

In this Section we have investigated two variants of the von Neumann entropy of a graph, based on the normalized and unnormalized Laplacian, respectively. With their quadratic approximations to hand, we have studied the entropic change as the new edges are added to the graph, giving new insight in the type of structural patterns that influence the value of the (approximated) entropy.

We performed an extensive set of experiments which showed that 1) the Laplacian and the normalized Laplacian entropies capture the presence of related yet different struc-

tural patterns, 2) the quadratic approximation fail to explain the emergence of non-trivial structures, in particular for the case of the Laplacian entropy, and that in general 3) the quality of the quadratic approximation, as well as which variant of the von Neumann entropy is better approximated, depends on the topology of the underlying graph. Our results suggest that the quadratic approximation of the von Neumann entropy can be an efficient way to measure the complexity of large networks, however the quality of this approximation depends on the topology of the network being studied. In particular, with the exception of small world networks, we find that the Laplacian entropy is easier to approximate. The normalized Laplacian entropy, on the other hand, can be approximated better for Erdös-Rényi and scale-free networks with low edge density.

Our analysis shows that:

- the two versions of the von Neumann entropy based on the Laplacian and normalized Laplacian (respectively) are connected to the presence of similar structural patterns, although with some significant differences;

- the correlation between these two entropic measures ranges from weakly positive to strongly negative, depending on the underlying graph structure;

- the quadratic approximations fail to explain the presence of non-trivial structures observed when the growth is driven by the exact entropies;

- the quality of the approximations, as well as which variant of the von Neummann entropy is better approximated, depends on the topology of the underlying graph;

## 4.2    Edge Centrality via the Holevo Quantity

In the study of complex networks, vertex centrality measures are used to identify the most important vertices within a graph. A related problem is that of measuring the centrality of an edge. In this section, we propose a novel edge centrality index rooted in quantum information. More specifically, we measure the importance of an edge in terms of the contribution that it gives to the Von Neumann entropy of the graph. We show that this can be computed in terms of the Holevo quantity, a well known quantum information theoretical measure. While computing the Von Neumann entropy and hence the Holevo quantity requires computing the spectrum of the graph Laplacian, we show how to obtain a simplified measure through a quadratic approximation of the Shannon entropy. This in turns shows that the proposed centrality measure is strongly correlated with the negative degree centrality on the line graph. We evaluate our centrality measure through an extensive set of experiments on real-world as well as synthetic networks, and we compare it against commonly used alternative measures.

### 4.2.1    Introduction

The study of complex networks has recently attracted increasing interest in the scientific community, as it allows to model and understand a large number of real-world systems [65]. This is particularly relevant given the growing amount of available data describing the interactions and dynamics of real-world systems. Typical examples of complex networks include metabolic networks [97], protein interactions [93], brain networks [169] and scientific collaboration networks [132].

One of the key problems in network science is that of identifying the most relevant nodes in a network. This importance measure is usually called the *centrality* of a vertex [105]. A number of centrality indices have been introduced in the literature [29, 65, 71, 72, 134, 153], each of them capturing different but equally significant aspects of vertex importance. Commonly encountered examples are the degree, closeness and betweenness centrality [71, 72, 134]. A closely related problem is that of measuring the centrality of an edge [33, 105]. Most edge centrality indices are developed as a variant of vertex centrality ones. A common way to define an edge centrality index is to apply the corresponding vertex centrality to the line graph of the network being studied. Recall that, given a graph $G = (V, E)$, the line graph $\mathscr{L}(G) = (V', E')$ is a dual representation of $G$ where each node $uv \in V'$ corresponds to an edge $(u, v) \in E$, and there exists and edge between two nodes of $\mathscr{L}(G)$ if and only if the corresponding edges of $G$ share a vertex. By measuring the vertex centrality on $\mathscr{L}(G)$, one can map it back to the edges of $G$ to obtain a measure of edge centrality. However, as observed by Koschützki et al. [105], this approach does not yield the same result as the direct definition of the edge centrality on $G$. Moreover, the size of the line graph is quadratic in the size of the original graph, thus making it hard to scale to large networks when the chosen centrality measure is computationally demanding.

Here, we introduce a novel edge centrality measure rooted in quantum information

theory. More specifically, we propose to measure the importance of an edge in terms of its contribution to the Von Neumann entropy of the network [140]. This can be measured in terms of the Holevo quantity, a well known quantum information theoretical measure that has recently been applied to the analysis of graph structure [154, 155]. We also show how to approximate this quantity in the case of large networks, where computing the exact value of the Von Neumann entropy is not feasible. This in turns highlights a strong connection between the Holevo edge centrality and the negative degree centrality on the line graph. Finally, we perform a series of experiments to evaluate the proposed edge centrality measure on real-world as well as synthetic graphs, and we compare it against a number of widely used alternative measures.

## 4.2.2 Quantum Information Theoretical Background

### 4.2.2.1 Quantum States and Von Neumann Entropy

In quantum mechanics, a system can be either in a pure state or a mixed state. Using the Dirac notation, a *pure state* is represented as a column vector $|\psi_i\rangle$. A *mixed state*, on the other hand, is an ensemble of pure quantum states $|\psi_i\rangle$, each with probability $p_i$. The density operator of such a system is a positive unit-trace matrix defined as

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| . \tag{4.11}$$

The *von Neumann entropy* [135] $S$ of a mixed state is defined in terms of the trace and logarithm of the density operator $\rho$

$$S(\rho) = -\text{Tr}(\rho \ln \rho) = -\sum_i \lambda_i \ln(\lambda_i) \tag{4.12}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\rho$. If $\langle \psi_i| \rho |\psi_i\rangle = 1$, i.e., the quantum system is a pure state $|\psi_i\rangle$ with probability $p_i = 1$, then the Von Neumann entropy $S(\rho) = -\text{Tr}(\rho \ln \rho)$ is zero. On other hand, a mixed state always has a non-zero Von Neumann entropy associated with it.

### 4.2.2.2 A Mixed State from the Graph Laplacian

Let $G = (V, E)$ be a simple graph with $n$ vertices and $m$ edges. We assign the vertices of $G$ to the elements of the standard basis of an Hilbert space $\mathcal{H}_G$, $\{|1\rangle, |2\rangle, ..., |n\rangle\}$. Here $|i\rangle$ denotes a column vector where 1 is at the $i$-th position. The *graph Laplacian* of $G$ is the matrix $L = D - A$, where $A$ is the adjacency matrix of $G$ and $D$ is the diagonal matrix with elements $d(u) = \sum_{v=1}^n A(u, v)$. For each edge $e_{i,j}$, we define a pure state

$$|e_{i,j}\rangle := \frac{1}{\sqrt{2}}(|i\rangle - |j\rangle). \tag{4.13}$$

Then we can define the mixed state $\{\frac{1}{m}, |e_{i,j}\rangle\}$ with density matrix

$$\rho(G) := \frac{1}{m} \sum_{\{i,j\} \in E} |e_{i,j}\rangle\langle e_{i,j}| = \frac{1}{2m} L(G). \tag{4.14}$$

Let us define the Hilbert spaces $\mathscr{H}_V \cong \mathbb{C}^V$, with orthonormal basis $\mathbf{a}_v$, where $v \in V$, and $\mathscr{H}_E \cong \mathbb{C}^E$, with orthonormal basis $\mathbf{b}_{u,v}$, where $\{u, v\} \in E$. It can be shown that the graph Laplacian corresponds to the partial trace of a rank-1 operator on $\mathscr{H}_V \otimes \mathscr{H}_E$ which is determined by the graph structure [51]. As a consequence, the Von Neumann entropy of $\rho(G)$ can be interpreted as a measure of the amount of entanglement between a system corresponding to the vertices and a system corresponding to the edges of the graph [51].

### 4.2.2.3  Holevo Quantity of a Graph Decomposition

Given a graph $G$, we can define an ensemble in terms of its subgraphs. Recall that a *decomposition* of a graph $G$ is a set of subgraphs $H_1, H_2, ..., H_k$ that partition the edges of $G$, i.e., for all $i, j$, $\bigcup_{i=1}^{k} H_i = G$ and $E(H_i) \cap E(H_j) = \emptyset$, where $E(G)$ denotes the edge set of $G$. Notice that isolated vertices do not contribute to a decomposition, so each $H_i$ can always be seen a subgraph that contains all the vertices. If we let $\rho(H_1), \rho(H_2), ..., \rho(H_k)$ be the mixed states of the subgraphs, the probability of $H_i$ in the mixture $\rho(G)$ is given by $|E(H_i)|/|E(G)|$. Thus, we can generalise Eq. 4.14 and write

$$\rho(G) = \sum_{i=1}^{k} \frac{|E(H_i)|}{|E(G)|} \rho(H_i). \tag{4.15}$$

Consider a graph $G$ and its decomposition $H_1, H_2, ..., H_k$ with corresponding states $\rho(H_1), \rho(H_2), ..., \rho(H_k)$. Let us assign $\rho(H_1), \rho(H_2), ..., \rho(H_k)$ to the elements of an alphabet $\{a_1, a_2, ..., a_k\}$. In quantum information theory, the classical concepts of uncertainty and entropy are extended to deal with quantum states, where uncertainty about the state of a quantum system can be expressed using the density matrix formalism. Assume a source emits letters from the alphabet and that the letter $a_i$ is emitted with probability $p_i = |E(H_i)|/|E(G)|$. An upper bound to the accessible information is given by the *Holevo quantity* of the ensemble $\{p_i, \rho(H_i)\}$:

$$\chi(\{p_i, \rho(H_i)\}) = S\left(\sum_{i=1}^{k} p_i \rho(H_i)\right) - \sum_{i=1}^{k} p_i S(\rho(H_i)) \tag{4.16}$$

## 4.2.3  Holevo Edge Centrality

We propose to measure the centrality of an edge as follows. Let $G = (V, E)$ be a graph with $|E| = m$, and let $H_e$ and $H_{\bar{e}}$ denote the subgraphs over edge sets $\{e\}$ and $E \setminus \{e\}$, respectively. Note that $S(\rho(H_e)) = 0$ and

$$\frac{m-1}{m} \rho(H_{\bar{e}}) + \frac{1}{m} \rho(H_e) = \rho(G). \tag{4.17}$$

Then the Holevo quantity of the ensemble $\{(m-1/m, H_{\bar{e}}), (1/m, H_e)\}$ is

$$\chi\left(\left\{\left(\frac{m-1}{m}, H_{\bar{e}}\right), \left(\frac{1}{m}, H_e\right)\right\}\right) = S\left(\rho(G)\right) - \frac{m-1}{m} S\left(\rho(H_{\bar{e}})\right) \tag{4.18}$$

**Definition 1.** *For a graph $G = (V, E)$, the* Holevo edge centrality *of $e \in E$ is*

$$HC(e) = \chi\left(\left\{\left(\frac{m-1}{m}, H_{\bar{e}}\right), \left(\frac{1}{m}, H_e\right)\right\}\right) \tag{4.19}$$

When ranking the edges of a graph $G$, the scaling factor $(m-1)/m$ is constant for all the edges and thus can be safely ignored. The Holevo edge centrality of an edge $e$ is then a measure of the difference in Von Neumann entropy between the original graph and the graph where $e$ has been removed. In other words, it can be seen as a measure of the contribution of $e$ to the Von Neumann entropy of $G$. From a physical perspective, this can also be interpreted as the variation of the entanglement between between a system corresponding to the vertices and a system corresponding to the edges of the graph (see the interpretation of the graph Laplacian in Section 4.2.2).

### 4.2.3.1 Relation with Degree Centrality

In this subsection we investigate the nature of the structural characteristics encapsulated by the Holevo edge centrality. Let $G = (V, E)$ be a graph with $n$ nodes, and let $I_n$ be the identity matrix of size $n$. We rewrite the Shannon entropy $-\sum_i \lambda_i \ln(\lambda_i)$ using the second order polynomial approximation $k \sum_i \lambda_i(1 - \lambda_i)$, where the value of $k$ depends on the dimension of the simplex. We obtain

$$S(\rho(G)) = -\text{Tr}\left(\rho(G) \ln \rho(G)\right) \approx \frac{|V| \ln(|V|)}{|V| - 1} \text{Tr}\left(\rho(G)(I_n - \rho(G))\right) \tag{4.20}$$

By noting that $\rho(G) = L(G)/(2m)$ and using some simple algebra, we can rewrite Eq. 4.20 as

$$S(\rho(G)) \approx \frac{|V| \ln(|V|)}{|V| - 1}\left(1 - \frac{1}{4m^2} \sum_{v \in V} \left(d^2(v) + d(v)\right)\right) \tag{4.21}$$

where $d(v)$ denotes the degree of the vertex $v$. This in turn allows us to approximate Eq.4.19 as

$$HC(e) = S(\rho(G)) - S(\rho(H_{\bar{e}})) \approx -\frac{|V| \ln(|V|)}{|V| - 1} \frac{d(u) + d(w)}{2m^2} \tag{4.22}$$

where $e = (u, w)$, we omitted the scaling factor $(m-1)/m$ and we made use of the fact that $1/(4m^2) \approx 1/(4(m-1)^2)$.

Eq. 4.22 shows that the quadratic approximation of the Holevo centrality is (almost) linearly correlated with the negative edge degree centrality (see Section 4.2.4). This in turn gives us an important insight into the nature of the Holevo edge centrality. However, the quadratic approximation captures only part of the structural information encapsulated by the exact centrality measure. In particular, Passerini and Severini [140] suggested that

(a) Exact  (b) Approximated
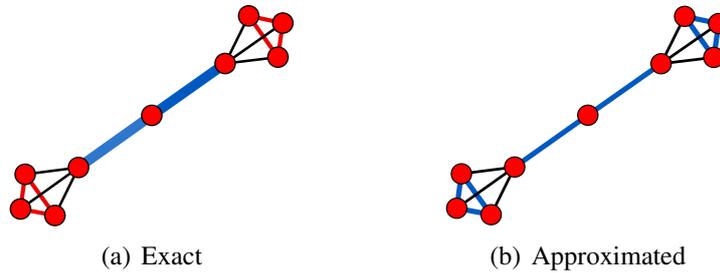
Figure 4.17: The Holevo edge centrality and its quadratic approximation on a barbell graph. Here the edge thickness is proportional to the value of the centrality. In (a) the blue edges have a higher centrality than the red edges, but in (b) all these edges (blue) have the same degree centrality.

those edges that create longer paths, nontrivial symmetries and connected components result in a larger increase of the Von Neumann entropy. Therefore, such edges should have a high centrality value, higher than what the degree information alone would suggest.

Fig. 4.17 shows an example of such a graph, where the central bridge has a high value of the exact Holevo edge centrality, but a relatively low value of the approximated edge centrality. In Fig. 4.17(b), the blue edges have all the same degree centrality, i.e., they are all adjacent to four other edges. However, from a structural point of view, the removal of the edges connecting the two cliques at the ends of the barbell graph would have a higher impact, as it would disconnect the graph. As shown in Fig. 4.17(a), the Holevo centrality captures this structural difference, i.e., the weight assigned to the two bridges (blue) is higher than that assigned to the edges in the cliques (red).

### 4.2.4 Experimental Evaluation

In the previous sections we have derived an expression for the Holevo edge centrality, both exact and approximated. Here, we first evaluate this measure on a number of standard networks, and we compare it against other well known edge centralities. We also analyse the behaviour of the proposed centrality measure when graphs endure structural changes.

#### 4.2.4.1 Experimental Setup

We perform our experiments on two well known real-world networks, the Florentine families graph and the Karate club network. We consider as edge centrality measures the Degree Centrality, the Betweenness Centrality and the Flow Centrality (more details in 2.4).

**Degree Centrality:** the centrality of an edge $e$ is computed as the degree of the corresponding vertex in the line graph. The idea underpinning the vertex degree centrality is that the importance of a node is proportional to the number of connections it has to

(a) Degree     (b) Betweenness     (c) Flow     (d) Holevo

(e) Degree     (f) Betweenness     (g) Flow     (h) Holevo
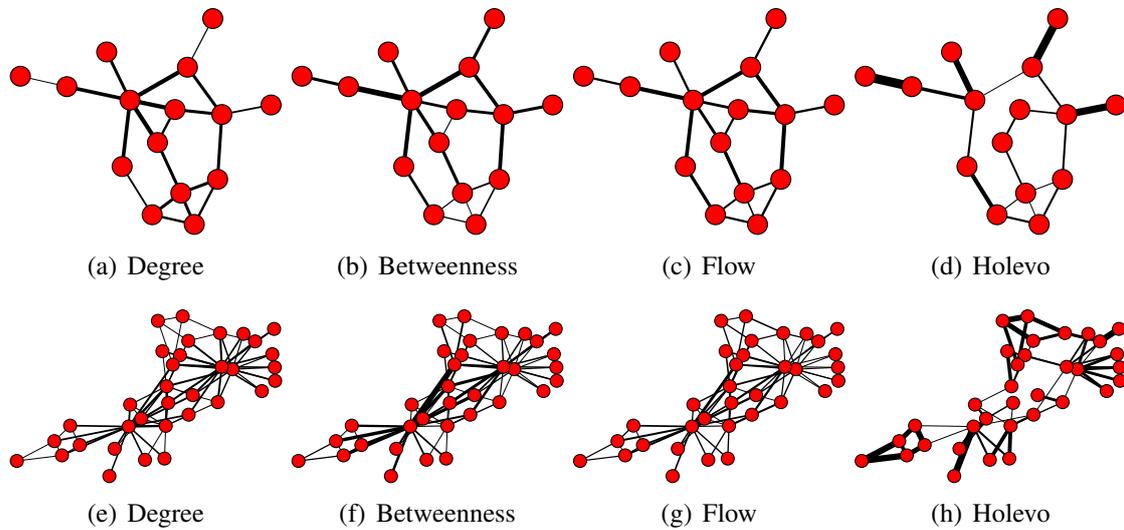
Figure 4.18: Edge centralities on the Florentine families network (a-d) and the Karate club network (e-h). A thicker edge indicates a higher value of the centrality.



(a) Florentine Families        (b) Karate Club
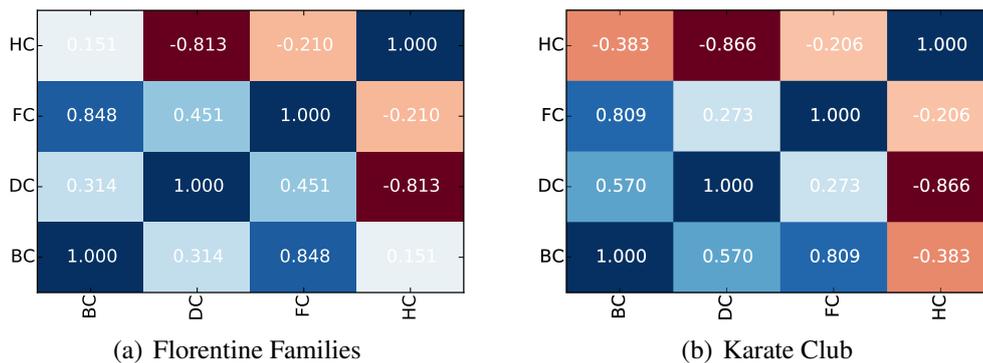
Figure 4.19: Correlation matrices for the centrality measure on the Florentine family network and the karate club network. DC, BC, FC, and HC denote the degree, betweenness, flow and Holevo centralities, respectively.

other nodes. This is the simplest edge centrality measure, but also the one with the lowest computational complexity.

**Betweenness Centrality:** the centrality of an edge $e$ is the sum of the fraction of all-pairs shortest paths that pass through $e$, i.e., $EBC(e) = \sum_{u,v \in V} \frac{\sigma(u,v|e)}{\sigma(u,v)}$ where $V$ is the set of nodes, $\sigma(u,v)$ and $\sigma(u,v|e)$ denote the number of shortest paths between $u$ and $v$ and the number of shortest paths between $u$ and $v$ that pass through $e$, respectively [33]. An edge with a high betweenness centrality has a large influence on the transfer of information through the network and thus it can be seen as an important bridge-like connector between two parts of a network. Note that the implementation we use does not rely on the

(a) Degree                              (b) Holevo
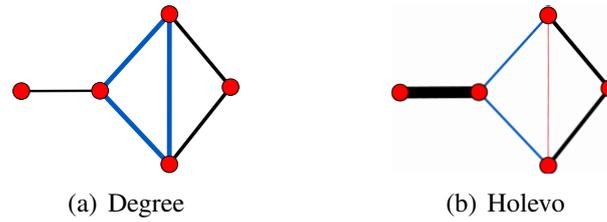
Figure 4.20: Toy example showing the difference in the structural information captured by the degree and Holevo centralities.

line graph, but measure the centrality of an edge directly on the original graph.

**Flow Centrality:** this centrality measure is also known as random-walk betweenness centrality [134]. While the betweenness centrality measures the importance of an edge $e$ in terms of shortest-paths between pairs of nodes that pass through $e$, the flow centrality is proportional to the expected number of times a random walk passes through the edge $e$ when going from $u$ to $v$. Similarly to the betweenness centrality, here we measure the flow centrality directly on the original graph.

### 4.2.4.2   Edge Centrality in Real-world Networks

In order to compare the Holevo edge centrality with the measures described in the previous subsection, we compute, for each network, the correlation between the Holevo quantity and the alternative measures. Fig. 4.18 shows the value of these centralities on the Florentine families graph and the Karate club network. In these plots, the thickness of an edge is proportional to the magnitude of the centrality index. Fig. 4.19, on the other hand, shows the correlation matrix between the different centralities. Here DC, BC, FC and HC denote the degree, betweenness, flow and Holevo centrality, respectively.

The Holevo centrality is always strongly negatively correlated with the degree centrality. This is in accordance with the properties discussed in 4.2.3. However, there are some significant differences. In general, the Holevo centrality is higher on edges that connect low degree nodes. In this sense, it can be seen as a measure of *peripherality*, rather than centrality. However, when two edges have the same degree centrality, edges that would disconnect the network or break structural symmetries are assigned a higher weight, as Fig. 4.17 shows. Similarly, in Fig. 4.20(a) the three edges highlighted in blue have the same degree centrality, but the same edges in Fig. 4.20(b) have different Holevo centralities. In fact, the removal of the red edge does not result in significant structural changes, while the removal of one of the blue edges increases the length of the tail.

### 4.2.4.3   Robustness Analysis

We then investigate the behaviour of the Holevo edge centrality when the graph undergoes structural perturbations. To this end, given an initial graph, we gradually add or delete
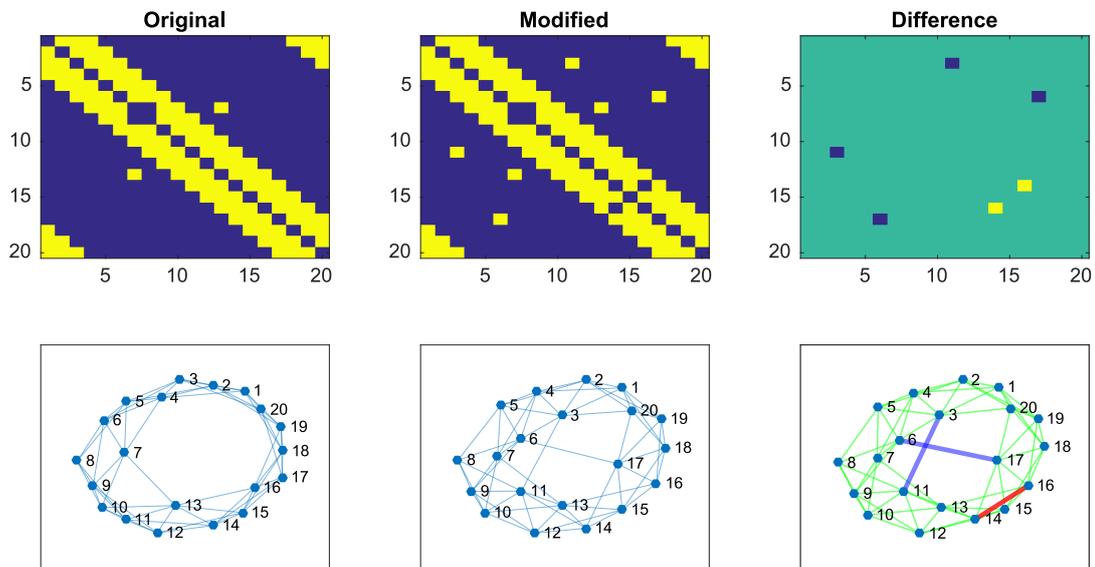
Figure 4.21: Perturbation process: on the left, adjacency matrix and plot of the starting graph; in the middle, the edited graph; on the right, the differences between initial and modified graph are highlighted.

edges according to an increasing probability $p$. Fig. 4.21 shows an instance of the noise addition process. Starting from a randomly generated graph, we compute the Holevo edge centrality for all its edges. Then, we perturb the graph structure with a given probability $p$ and again we recompute the Holevo edge centrality for all the graph edges. We compute the correlation between the Holevo centrality of the edges of the original graph and its noisy counterpart. More specifically, we measure the correlation between the centralities of the edges that belong to the intersection of their edge sets. In other words, we analyse how the centrality changes during the perturbation process, with respect to the starting state.

Since we are interested in the variation of the Holevo centrality as the graph structure changes, we use three different random graph models to generate the initial graph: 1) the Erdös-Rényi model, 2) the Watts–Strogatz model and 3) the Preferential Attachment model. For each model, we generate a starting graph with the same number of nodes $n$ and we create 100 noisy instances as $p$ varies from 0.01 to 0.3. We perform the same experiment for both the Holevo centrality and the betweenness centrality.

Fig. 4.22 shows the average correlation as we perturb the graph structure, for both the Holevo and betweenness centrality. As expected, in both cases the correlation decreases as the similarity between the original graph and the edited one decreases. However, while the correlation for centrality measures decreases rapidly in the case of Erdös and Rényi graphs, on scale-free graphs our centrality measure decreases linearly with the value of $p$, while the betweenness centrality drops significantly more quickly. On the other hand, we observe the opposite behaviour on small-world graphs. This can be explained by noting that in small-world graphs there exist multiple alternative paths between every pair of

(a) Erdös-Rényi                 (b) Watts-Strogatz                 (c) Preferential Attachment
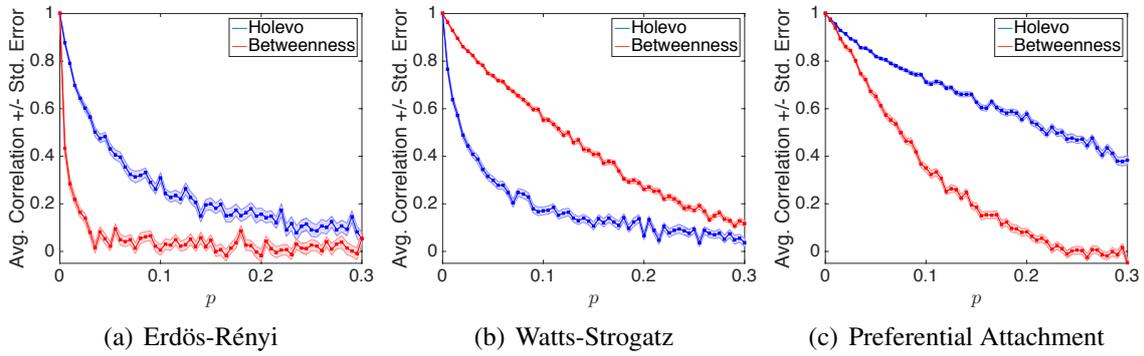
Figure 4.22: Average correlation between the centrality of the edges of the original graph and those of increasingly noisy version of it. The different columns refer to different starting graphs: (a) Erdös-Rényi, (b) Watts-Strogatz and (c) Preferential Attachment.

nodes, and thus the betweenness centrality is less affected by structural modifications. On the other hand, in scale-free graphs most shortest-paths pass through a hub, and thus adding a random edge can create shortcuts that greatly affect the value of the betweenness centrality. The Holevo centrality, however, assigns large weights to long tails and leaves, which are less affected by the structural noise.

## 4.2.5   Conclusion

In this section we have introduced a novel edge centrality measure based on the quantum information theoretical concept of Holevo quantity. We measured the importance of an edge in terms of the difference in Von Neumann entropy between the original graph and the graph where that edge has been remove. We showed that by taking a quadratic approximation of the Von Neumann entropy we obtain an approximated value of the Holevo centrality that is proportional to the negative degree centrality. We performed a series of experiments on both real-world and synthetic networks and we compared the proposed centrality measure to widely used alternatives. Future work will investigate higher order approximations of this centrality measure as well as the possibility of defining network growth models based on the Holevo quantity.

# 4.3 Graph Similarity through CTQW and QJSD

In this Section, we extend upon the results of Rossi *et al*. [154] by further investigating the Hamiltonian choice and applying the algorithm to a different graphs typology as well. Note that it is still work in progress but with the prospect of publishing the outcomes shortly.

The original method aims at measuring the similarity between two structures. For this purpose, the authors introduce a novel kernel on unattributed and undirected graphs which evaluates the similarity by way of the evolution of a continuous-time quantum walk on their structures. More to the point, two structures are merged by creating a complete set of connections between their nodes. Then, they define two continuous-time quantum walks over this resulting structure such that the density operators of the quantum states are orthogonal whenever the two original graphs are isomorphic. Finally they measure the quantum Jensen-Shannon divergence between these states. Clearly, the algorithm is not intended to provide any speedup with respect to classical counterparts, rather to pinpoint how quantum walks are excellent tools to capture information about the structural similarities between two graphs.

Our contribution concerns different facets related to the original algorithm. In particular, we investigate whether the method may benefit from different choices of Hamiltonian not taken into consideration yet. To this end, we take into account the normalized Laplacian matrix option. As the original method has been intended for undirected and unattributed graphs, first we are interested in probing its efficacy over attributed graphs. For that purpose, we enrich the graph information through feature descriptors. More specifically, we make use of the heat kernel signature [170] and the wave kernel signature [16] to extract node signatures and thus enhance the topology information. Finally, we also look at directed graphs. In this case, we exploit the (normalized) Laplacian counterpart for directed graphs.

The remainder of the Section is organized as follows: in 4.3.1 we provide a brief literature survey and some mathematical background about specific topics taken into account in this work while 4.3.2 is intended as an introduction to continuous-time quantum walks and the quantum Jensen-Shannon divergence. In 4.3.4 we introduce the quantum method. Finally in 4.3.5 we illustrates the experimental results, while the conclusions are presented in 4.3.6.

## 4.3.1 Literature Survey

The full expressivity of graphs usually comes at the cost of an increased trouble in applying standard machine learning and pattern recognition techniques to them. The lack of a canonical ordering for the nodes in a graph or even the varying dimension of the embedding space are well-known classical issues.

The so called kernel trick allows to transform the problem of finding an embedding of

the set of data entities into that of defining a positive semidefinite kernel between them. Kernel methods [158] are based on this elegant transformation.

The most famous example of this approach is the support vector machines (SVMs) [185]. The data under consideration may be of different types: entire graphs, vectors or just nodes. Given a positive semidefinite kernel $k : X \times X \to \mathbb{R}$ on a set $X$, we know that there exists a map $\phi : X \to H$ into a Hilbert space $H$, such that $k(x, y) = < \phi(x), \phi(y) >$ for all $x, y \in X$, where $< \cdot, \cdot >$ denotes the scalar product in $H$.

Therefore, any algorithm that can be defined as scalar products of the $\phi(x)$'s can be applied to a set of data on which a kernel is defined. Haussler [83] was one of the pioneer of such a method and its R-convolution kernel inspired many other studies. In fact, over the years in the literature, a great number of graph kernels have been introduced [31, 76, 161] . The common denominator behind these kernels is that of defining the similarity between two graphs by breaking them down and then contrasting the resulting more elementary substructures. This prompted to the development of of the random walk kernel by Gärtner et al. [76], based on the list of common random walks between two graphs. In a similar way, Borgwardt *et al.* [31] estimate the similarity by comparing the shortest paths in the structures, whereas Shervashidze *et al.* [161] by evaluating the presence of small subgraphs. Another interesting method is the one proposed by Bai and Hancock in [19], where they study the possibility of defining a graph kernel based on the Jensen-Shannon kernel. The Jensen-Shannon kernel is a non-extensive information theoretic kernel defined in terms of the entropy of probability distributions over the structures being compared [123]. Bai and Hancock also extend this concept to the graph domain by linking with each graph either the steady state distribution of a random walk on the graph or its Von Neumann entropy [139]. However, it is worth noting that the problem of defining a *complete* kernel, *i.e.*, a kernel whose implicit map $\phi$ is injective, is at least as hard as the graph isomorphism problem [76].

A large amount of work has been done on designing local point signatures especially in the context of shape analysis. A local point signature is a data set that represents specific features of a given point. Strictly speaking a descriptor or signature over a graph is a means to capture the unique properties of the structure (either locally or globally) basically to the aim of distinguishing it from others belonging to other classes. At the same time it should be invariant to a certain class of transformations a graph can undergo [38, 171]. A a considerable number of methods has been proposed by various scientist to deal with this problem during the last years. One of the initial point signature was introduced by Chua and Jarvi [47], in 1997. Almost one decade later Rustamov [156] proposed an approach for a global point signature based on the eigenfunction and eigenvalues of the Laplace Beltrami operator. Recently, two other spectral node signatures, the heat kernel signature (HKS) [170] and the wave kernel signature (WKS) [16], have been drawing significant attention. These constructions are inspired by physical processes (*e.g.*, heat propagation) on graphs, and are expected to inherit the physical processes' stability to perturbations of the underlying graph. However, the analysis of stability of spectral signatures in general is lacking, which hinders the ability to design spectral node signatures

that are not derived from physical processes [89]. Sun *et al.* [170] propose a point signature based on the heat kernel, indeed taking the name of Heat Kernel Signature (HKS). It is obtained by restricting the heat kernel to the temporal domain where the kernel is actually the fundamental solution to the heat equation. More to the point, the heat kernel $h_t(x, y)$ relates the amount of heat transferred from $x$ to $y$ after time $t$. The eigen decomposition of the heat kernel is expressed as

$$h_t(x, y) = \sum_{i=0}^{\infty} \exp\left(-\lambda_i t\right)\phi_i(x)\phi_i(y) \tag{4.23}$$

where $\lambda_i$ and $\phi_i$ are the $i^{th}$ eigenvalue and eigenfunction of $\Delta$, the Laplace-Beltrami operator. This implies that the Heat Kernel Signature is being defined as

$$HKS(x, t) = h_t(x, x) \tag{4.24}$$

The HKS is intrinsic and isometry-invariant (*i.e.*, two isometric graphs have equal HKS), multi-scale and captures both local features and global structure. In addition, it is informative: under mild conditions, if two graphs have equal heat kernel signatures, they are isometric [170]. Of course, the HKS has been amply and effectively used for characterizing objects. Especially, the study of this signature has been applied to graphs domain by employing the eigen-system of the Laplacian matrix to derive the heat kernel.

The Wave kernel signature (WKS) [16] follows a similar idea to the HKS's, but replacing the heat equation with the Schrödinger wave equation. It is defined to be the time-averaged probability of detecting a particle of a certain energy distribution at the point $x$, formulated as

$$WKS(x, e) = \sum_k \exp\left(-\frac{(e - log(\lambda_k))^2}{2\sigma^2}\right)\phi_k(x)^2 \tag{4.25}$$

Quantum walks have recently emerged as a primitive for designing novel quantum algorithms [12, 45, 102, 157] on graph structures. Indubitably, the most renowned property of quantum walks is that of providing polynomial or even exponential speedups over classical computation in many problems [69, 159]. It turned out that such a speedup is often caused by the highly symmetrical structure of the graphs. Recall that the automorphism group of a graph is the set of all permutation on vertices that preserve the adjacency. In this sense also a symmetry of a graph is an automorphism. Krovi and Brun [106] have demonstrated that the occurrence of infinite hitting times and of exponential speedup are generally an effect of the decays within the eigenspace of the evolution operator. In turn, these are linked with the symmetry group of the graph. The intuition behind is that, in case of symmetry, a quantum walker will hit a node by multiple paths with an identical phase. in that event, the interference is constructive and the node has a strong chance of being hit. This ends up in a quicker hitting time for node. However, based on how the initial state of the walk are set, the quantum walker may also hit the vertex on paths with

phases that coincide to destructive interference. In fact, in some cases the likelihood of the walker to hit a node can be zero, *i.e.*, the hitting time is infinite. Also Emms *et al.* [60] investigated over the connection between symmetries and quantum walks. They showed how degenerate directions in the quantum commute time embedding space match with the symmetries of the graph. Recently, a way to find approximate axial symmetries in graphs by measuring the interference patterns of continuous-time quantum walks has been introduced by Rossi *et al.* [150] Unfortunately, the analyses of Rossi et al. [150] as well as of Emms et al. [60] are not based on a principled observable and are thus semi-classical.

Rossi *et al.* in [154] have grounded the analysis of the behaviour of the walks on the recently introduced quantum Jensen-Shannon divergence. The quantum Jensen-Shannon divergence is employed as a way to compute the distance between quantum states [107, 120]. The computation of the kernel is naturally managed by means of the interference effects of quantum walks, hence avoiding the use of either the development of a product union graph [19] in the classical case or a rotation in Hilbert space [178]. The Quantum Jensen-Shannon Divergence (QJSD) has recently been developed as a generalization of the classical Jensen-Shannon divergence to quantum states by Majtey, Lamberti and Prato [107,119,120]. It is symmetric, bounded and always defined, as the classical Jensen-Shannon divergence [111]. Contrary to its classical analogue, it has been demonstrated to be the square of a metric only for pure states [107], while only empirical clues indicate that it is for mixed states [107]. Besides, it has been shown that for mixed quantum states the quantum Jensen-Shannon divergence has a good capacity of distinguish. Since the QJSD is determined with respect to the Von Neumann entropy, that implies it is not straight a quantum-mechanical observable, *i.e.*, there exists no operator whose expected value is the QJSD. However, it can be calculate over density matrices whose entries are actually observables. The drawback affecting the analysis not based on principled observable has been overcome by Rossi *et al.* [155], who resorting to using the quantum Jensen-Shannon divergence, rendered the analysis fully based on observable properties effectively. In another studies, based on the classical Jensen-Shannon kernel, Bai and Hancock [19], build a product graph from the two input graphs in order to measure the composite entropy. Note that a number of alternative graph kernels based on the classical Jensen-Shannon divergence and its quantum counterpart have been recently introduced in the literature [151, 152].

**Directed Graphs Laplacian** One of the aims in this work is to explore whether the method proposed by Rossi *et al.* [154] can be extended from undirected to directed graphs. In a directed graph edges are directed from one vertex to another, then the adjacency matrix is non-symmetric (presents in-degree and out-degree) and consequently the spectrum is complex. For this reason directed graphs pose some interesting questions, above all for the spectra viewpoint.

Suppose $G = (V, E)$ is a directed graph with node set $V$ and link set $E$, whose adjacency matrix is defined as for the undirected version. However, for a directed edge $(u, v)$ in $E$ we say that there is an edge $(u, v)$ from $u$ to $v$, namely $u$ has an out-neighbor $v$ while $v$

has an in-neighbor $u$ [48]. The in-degree and out-degree of node $u$ is thus

$$d_u^{in} = \sum_{v \in V} A_{vu}, \qquad d_u^{out} = \sum_{v \in V} A_{uv}. \qquad (4.26)$$

With a view to provide a definition of Laplacian matrix for such graph type, first we have to specify the random walk transition matrix P. For a given directed graph $G$, a typical random walk transition matrix is

$$P = D_{out}^{-1} A \qquad (4.27)$$

where $D_{out}$ is formed from row sums (out-degree), with the convention that when $D_{out}$ is not defined ($d_{out} = 0$), we modify $D_{out}^{-1}$ to ensure the entry is 0 in that specific case. Strongly connected directed graphs (*i.e.*, there exists a path between every pair of vertices and therefore there are no sinks) have P irreducible and an unique nonnegative left eigenvector $\pi$, which has eigenvalue 1 (by applying the Perron-Frobenius theorem [88]).

$$\pi = \pi P \qquad (4.28)$$

Once determined the random walk transition matrix, we can define the Laplacian on a directed graph[3]. Let $\Omega$ be a diagonal matrix with the elements of $\pi$ on the diagonal. Then the Laplacian matrix is defined as

$$L = \Omega - \frac{1}{2} \left( \Omega P + P^\top \Omega \right) \qquad (4.29)$$

whereas the normalized Laplacian will be

$$\tilde{L} = \boldsymbol{I} - \frac{1}{2} \left( \Omega^{\frac{1}{2}} P \Phi^{-\frac{1}{2}} + \Omega^{-\frac{1}{2}} P^\top \Omega^{\frac{1}{2}} \right) \qquad (4.30)$$

The above matrices are symmetric and coincide with the undirected analogues.

**Similarity Matrices** Let $G(V, E, u)$ be an attributed graph, with node set $V$, edge set $E$ while $u$ is the function assigning attributes to the nodes. That means at each vertex is attached a descriptor, usually taking the form of a $n$-dimensional vector. We determine the similarity for vertex pairs as $w(v_j, v_k) = e^{||v_j(u) - v_k(u)||_2^2}$ where $||v_j(u) - v_k(u)||_2$ is the Euclidean distance between feature points. A matrix of similarities $W$ is computed by using the above measure between each pair of nodes, in the following way

$$W_{jk} = w(v_j, v_k) \qquad \forall j, k \in V \qquad (4.31)$$

Feature points over graphs can be obtained via Heat Kernel Signature (HKS) or Wave Kernel Signature (WKS), for the sake of clarity. In the section 4.3.4 we will see when e where apply a similarity matrix.

---

[3]for further technical details see [48]

### 4.3.2   Continuous-Time Quantum Walks

A *continuous-time random walk* on the graph $G$ is generally described by a Markovian diffusion process over its vertices, and transitions are permitted only along the edges connecting adjacent nodes.

Let $\mathbf{p}_t \in \mathbb{R}^n$ be a vector denoting the state of the walk at time $t$, such that its $u$-th entry gives the probability of the walk being at vertex $u$ at time $t$. Then the state vector evolves according to the equation

$$\mathbf{p}_t = e^{-Lt}\mathbf{p}_0 \tag{4.32}$$

where $L$ is the generator matrix of the underlying continuous-time Markov process.

The *continuous-time quantum walk* is the quantum counterpart of the continuous-time random walk, and it is similarly defined as a dynamical process over the vertices of the graph [69]. Here the classical state vector is replaced by a vector of complex amplitudes over $V$ whose squared norm sums to unity, and as such the state of the system is not constrained to lie in a probability space. In fact, the lack of restrictions on the sign and complex phase allows for interference effects to take place. Let us denote, using Dirac notation, the basis state corresponding to the walk being at vertex $u \in V$ as $|u\rangle$. A general state of the walk is a complex linear combination of the basis states, such that the state of the walk at time $t$ is defined as

$$\left|\psi_t\right\rangle = \sum_{u \in V} \alpha_u(t)\left|u\right\rangle \tag{4.33}$$

where the amplitude $\alpha_u(t) \in \mathbb{C}$ and $\left|\psi_t\right\rangle \in \mathbb{C}^{|V|}$ are both complex. Moreover, we have that $\alpha_u(t)\alpha_u^*(t)$ gives the probability that at time $t$ the walker is at the vertex $u$, and thus $\sum_{u \in V} \alpha_u(t)\alpha_u^*(t) = 1$ and $\alpha_u(t)\alpha_u^*(t) \in [0,1]$, for all $u \in V$, $t \in \mathbb{R}^+$.

The evolution of the walk is then given by the Schrödinger equation, where we denote the time-independent Hamiltonian as $\mathcal{H}$.

$$\frac{\partial}{\partial t}\left|\psi_t\right\rangle = -i\mathcal{H}\left|\psi_t\right\rangle. \tag{4.34}$$

Given an initial state $\left|\psi_0\right\rangle$, we can solve Equation (4.34) to determine the state vector at time $t$

$$\left|\psi_t\right\rangle = e^{-i\mathcal{H}t}\left|\psi_0\right\rangle. \tag{4.35}$$

Usually the Laplacian matrix is chosen as the system Hamiltonian, *i.e.*, $\mathcal{H} = L$. Nonetheless, any Hermitian operator encoding the structure of the graph can be adopted as alternative.

We conclude by remark that we can rewrite Eq. 5.6 in the following way. Given the spectral decomposition of the Hamiltonian $\mathcal{H} = \Phi\Lambda\Phi^\top$ [4] and the fact that $\exp[-i\mathcal{H}t] = \Phi\exp[-i\Lambda t]\Phi^\top$ we can then write

$$\left|\psi_t\right\rangle = \Phi e^{-i\Lambda t}\Phi^\top\left|\psi_0\right\rangle. \tag{4.36}$$

---

[4] where $\Phi$ is the $n \times n$ matrix $\Phi = (\phi_1|\phi_2|...|\phi_j|...|\phi_n)$ with the ordered eigenvectors $\phi_j$s of $\mathcal{H}$ as columns and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_j, ..., \lambda_n)$ is the $n \times n$ diagonal matrix with the ordered eigenvalues $\lambda_j$ of $\mathcal{H}$ as elements

### 4.3.3 Quantum Jensen-Shannon Divergence

The observation process for a quantum system is defined in terms of projections onto orthogonal subspaces associated with operators on the quantum state-space called *observables*. Let $O$ be an observable of the system, with spectral decomposition

$$O = \sum_i a_i P_i \tag{4.37}$$

where the $a_i$ are the (distinct) eigenvalues of $O$ and the $P_i$ the orthogonal projectors onto the corresponding eigenspaces. The outcome of an observation, or projective measurement, of a quantum state $|\psi\rangle$ is one of the eigenvalues $a_i$ of $O$, with probability

$$P(a_i) = \langle\psi| P_i |\psi\rangle \tag{4.38}$$

After the measurement, the state of the quantum systems becomes

$$|\bar\psi\rangle = \frac{P_i |\psi\rangle}{||P_i |\psi\rangle||}, \tag{4.39}$$

where $|| |\psi\rangle || = \sqrt{\langle\psi|\psi\rangle}$ is the norm of the vector $|\psi\rangle$.

Density operators play an important role in the quantum observation process. The observation probability of $a_i$ is $P(a_i) = \text{Tr}(\rho P_i)$, with the mixed state being projected by the observation process onto the state represented by the modified density matrix $\rho' = \sum_i P_i \rho P_i$. The expectation of the measurement is $\langle O \rangle = \text{Tr}(\rho O)$. The projective properties of quantum observation means that an observation actively modifies the system, both by altering its entropy and forcing an energy exchange between quantum system and observer.

The *von Neumann entropy* [135] $H_N$ of a mixture is defined in terms of the trace and logarithm of the density operator $\rho$

$$H_N = -\text{Tr}(\rho \log \rho) = -\sum_i \xi_i \ln \xi_i \tag{4.40}$$

where $\xi_1,\ldots,\xi_n$ are the eigenvalues of $\rho$. If $\langle\psi_i| \rho |\psi_i\rangle = 1$, i.e., the quantum system is a pure state $|\psi_i\rangle$ with probability $p_i = 1$, then the Von Neumann entropy $H_N(\rho) = -\text{Tr}(\rho \log \rho)$ is zero. On other hand, for a mixed state described by the density operator $\sigma$ we have a non zero Von Neumann entropy associated with it. With the Von Neumann entropy to hand, the quantum Jensen-Shannon divergence between two density operators $\rho$ and $\sigma$ is defined as

$$D_{JS}(\rho,\sigma) = H_N\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2} H_N(\rho) - \frac{1}{2} H_N(\sigma) \tag{4.41}$$

This quantity is always well defined, symmetric and positive definite.
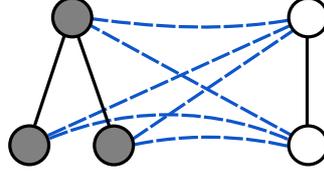
Figure 4.23: Given two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ we construct a new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = V_1 \cup V_2$, $\mathcal{E} = E_1 \cup E_2$ and we add a new edge $(u, v)$ between each pair of nodes $u \in V_1$ and $v \in V_2$.

It can also be shown that $D_{JS}(\rho, \sigma)$ is bounded, i.e., $0 \le D_{JS}(\rho, \sigma) \le 1$. Let $\rho = \sum_i p_i \rho_i$ be a mixture of quantum states $\rho_i$, with $p_i \in \mathbb{R}^+$ such that $\sum_i p_i = 1$, then one can prove that

$$H_N(\sum_i p_i \rho_i) \le H_S(p_i) + \sum_i p_i H_N(\rho_i) \tag{4.42}$$

where $H_S$ indicates the Shannon entropy and the equality is attained if and only if the states $\rho_i$ have support on orthogonal subspaces. By setting $p_1 = p_2 = 0.5$, we see that

$$D_{JS}(\rho, \sigma) = H_N\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \le 1 \tag{4.43}$$

Hence $D_{JS}$ is always less than or equal to 1, and the equality is attained only if $\rho$ and $\sigma$ have support on orthogonal subspaces.

The interest in the quantum Jensen-Shannon divergence stems from the fact that it verifies several interesting properties which are necessary to be a measure of distinguishability between quantum states [107, 120]. The distinguishability problem is of pivotal importance in quantum mechanics, and strictly connected to the concept of distance between states. For instance, Wootters [200] measures the distance between two states $|\phi\rangle$ and $|\psi\rangle$ of the same physical system by enumerating the distinguishable states between $|\phi\rangle$ and $|\psi\rangle$. Wootters' work is basically grounded on the extension of a distance over the space of probability distributions to the Hilbert space of pure quantum states. In the literature many other metrics have been introduced, such as the Bures distance [41], which is quite comparable to the QJSD since both need the same number of observations. Nevertheless, the QJSD seems to be preferable as it is faster to compute.

### 4.3.4  The QJSD Kernel

Let $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ be two unattributed graphs, *i.e.*, graphs with no attributes or features attached to their nodes and edges. Given $G_1$ and $G_2$, we build a new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (4.44) where $\mathcal{V} = V_1 \cup V_2$, $\mathcal{E} = E_1 \cup E_2 \cup E_{12}$, and $(u, v) \in E_{12}$ only if $u \in V_1$ and $v \in V_2$ (see Fig. 4.23 for an example).

On the contrary, in the case of attributed graphs, $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ are replaced by $G_1^*(V_1, E_1, u_1)$ and $G_2^*(V_2, E_2, u_2)$. The new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is no longer composed

by edges and nodes but replaced by a different structure, made up of similarity matrices. Given $G_1^*$ and $G_2^*$, first we compute four similarity matrices, namely $Gs_{11}$, $Gs_{22}$, $Gs_{12}$ and $Gs_{21}$, as expressed in Eq. 4.31. Here $Gs_{ij}$ means nodes and attributes from graph $G_i^*$ and nodes and attributes from graph $G_j^*$ are taken on board for the computation. Note that $Gs_{11}$ and $Gs_{22}$ have zero diagonal though. Then, a block matrix $\boldsymbol{G}$, in place of $\mathcal{G}$ 4.44, is defined as

$$\boldsymbol{G} = \left( \begin{array}{c|c} Gs_{11} & Gs_{12} \\ \hline Gs_{21} & Gs_{22} \end{array} \right)$$

Nevertheless, the initial states remain the degree distributions of $G_1$ and $G_2$.

With this new structure to hand (either $\mathcal{G}$ or $\boldsymbol{G}$), we define two independent continuous-time quantum walks with starting states

$$\left| \psi_0^- \right\rangle = \frac{\sum_{u \in V_1} d_u \left| u \right\rangle - \sum_{v \in V_2} d_v \left| v \right\rangle}{C} \qquad \left| \psi_0^+ \right\rangle = \frac{\sum_{u \in \mathcal{V}} d_u \left| u \right\rangle}{C} \qquad (4.45)$$

where the basis state corresponding to the walk being at vertex $v \in \mathcal{V}$ is denoted as $\left| v \right\rangle$, $d_v$ denotes the degree of vertex $v$, and $C$ is the normalisation constant such that the probabilities sum to one. We set the initial amplitude on the nodes of $G_1$ and $G_2$ to be respectively in antiphase and in phase. In other words, we form the initial states of the walks in this way in order to stress the presence of destructive and constructive interference schemes.

We let the two quantum walks evolve under Eq. 5.6 until a time $T$ and we define the average density operators $\rho_T^-$ and $\rho_T^+$ as

$$\rho_T^- = \frac{1}{T} \int_0^T \left| \psi_t^- \right\rangle \left\langle \psi_t^- \right| \mathrm{d}t \qquad \rho_T^+ = \frac{1}{T} \int_0^T \left| \psi_t^+ \right\rangle \left\langle \psi_t^+ \right| \mathrm{d}t \qquad (4.46)$$

Strictly speaking, we define two mixed systems with equivalent likelihood of being in any of the pure states determined by the quantum walks evolutions.

According to the designers of the original method [154], the justification for the proposed approach is that, whenever $G_1$ and $G_2$ are isomorphic, the distinguishability between the two states $\rho_T^-$ and $\rho_T^+$, which accentuates respectively destructive and constructive interference, will be maximal. Indeed, experiments are conceived so that the starting states are orthogonal and remain orthogonal during the quantum walk evolution, provided that $G_1$ and $G_2$ are isomorphic. Then, given two graphs $G_1$ and $G_2$, the quantum Jensen-Shannon kernel $k_T(G_1, G_2)$ between them is

$$k_T(G_1, G_2) = D_{JS}(\rho_T^-, \rho_T^+) \qquad (4.47)$$

where $\rho_T^-$ and $\rho_T^+$ are the density operators defined as in Eq. 4.46. Note that this kernel may be parametrised by the time $T$ and the choice of the time parameter can affect the computational complexity of the kernel. Finally, recall that in Eq. 4.45 we defined the initial state to be proportional to the nodes degree in the original graphs. As a consequence, the kernel is not defined on graphs $G = (V, E)$ with $E = \emptyset$, i.e., completely disconnected graphs.

#### 4.3.4.1   Kernel Computation

Here we evaluate the computational complexity of the kernel. In particular, we show that the solution to Eq. 4.46 can be computed analytically. Define $P_\lambda = \sum_{k=1}^{\mu(\lambda)} \phi_{\lambda,k} \phi_{\lambda,k}^\top$ to be the projection operator on the subspace spanned by the $\mu(\lambda)$ eigenvectors $\phi_{\lambda,k}$ associated with the eigenvalue $\lambda \in \Lambda$, where $\Lambda$ is the set of eigenvalues of the Hamiltonian. Given this set of projectors, the unitary operator inducing the quantum walk can be rewritten as

$$U^t = \sum_\lambda e^{-i\lambda t} P_\lambda \tag{4.48}$$

Recall that $|\psi_t\rangle = U^t |\psi_0\rangle$. Given Eq. 4.48 we can express the density matrix at time $t$ in terms of the projectors $P_\lambda$, *i.e.*,

$$\rho_t = U^t \rho_0 (U^t)^\dagger = \sum_{\lambda_1 \in \Lambda} \sum_{\lambda_2 \in \Lambda} e^{-i(\lambda_1 - \lambda_2)t} P_{\lambda_1} \rho_0 P_{\lambda_2}^\top \tag{4.49}$$

As a consequence, we can reformulate Eq. 4.46 as

$$\rho_T = \sum_{\lambda_1 \in \Lambda} \sum_{\lambda_2 \in \Lambda} P_{\lambda_1} \rho_0 P_{\lambda_2}^\top \frac{1}{T} \int_0^T e^{-i(\lambda_1 - \lambda_2)t} \, \mathrm{d}t \tag{4.50}$$

The integral in Eq. 4.50 can be solved yielding

$$\rho_T = \sum_{\lambda_1 \in \Lambda} \sum_{\lambda_2 \in \Lambda} P_{\lambda_1} \rho_0 P_{\lambda_2}^\top \frac{i(1 - e^{iT(\lambda_2 - \lambda_1)})}{T(\lambda_2 - \lambda_1)} \tag{4.51}$$

Letting $T \to \infty$, the integral in Eq. 4.50 reduces to the Dirac delta function $\delta(\lambda_1 - \lambda_2)$. Hence, Eq. 4.50 simplifies to

$$\rho_\infty = \sum_{\lambda \in \tilde{\Lambda}} P_\lambda \rho_0 P_\lambda^\top \tag{4.52}$$

where $\tilde{\Lambda}$ is the set of distinct eigenvalues of the Hamiltonian, *i.e.*, the eigenvalues $\lambda$ with multiplicity $\mu(\lambda) = 1$. Finally, along the same lines of Rossi et al. [155], one can show that as a consequence of Eq. 4.52 the infinite-time limit of the average density matrix commutes with the Hamiltonian $\mathcal{H}$, and thus the complexity of computing the Von Neumann entropy of $\rho_\infty$, *i.e.*, the Shannon entropy of its eigenvalues, is $O\left(\sum_{\lambda \in \tilde{\Lambda}} \mu(\lambda)^2\right)$, where $\mu(\lambda)$ is the multiplicity of the eigenvalue $\lambda$. As a consequence, we have that the complexity of computing the QJSD kernel with $T \to \infty$ is upper bounded by that of computing the eigendecomposition of $\mathcal{H}$, *i.e.*, $O(|\mathcal{V}|^3)$.

Note that the authors in [154] also provided and proved some interesting properties of the QJSD kernel. They proved that if $G_1$ and $G_2$ are two isomorphic graphs, then $\rho_T^-$ and $\rho_T^+$ have support on orthogonal subspaces. As corollary they also showed that given a pair of graphs $G_1$ and $G_2$, the kernel satisfies the following properties: 1) $0 \le k(G_1, G_2) \le 1$ and 2) if $G_1$ and $G_2$ are isomorphic, then $k(G_1, G_2) = 1$. Unfortunately, a formal proof of the positive semi definiteness has not been produced yet.

| Datasets | MUTAG | PPI | PTC | COIL | NCI1 |
|---|---|---|---|---|---|
| Max # vertices | 28 | 232 | 109 | 241 | 111 |
| Min # vertices | 10 | 3 | 2 | 72 | 3 |
| Avg # vertices | 17.93 | 109.60 | 25.56 | 144.97 | 29.87 |
| # graphs | 188 | 86 | 344 | 360 | 4110 |
| # classes | 2 | 2 | 2 | 5 | 37 |

Table 4.1: Information on the undirected graph datasets

## 4.3.5 Experimental Results

In this Section we evaluate the accuracy of the QJSD kernel in a classification task. In particular, for the undirected version of the algorithm, we use the following standard graph datasets:

**MUTAG** [54] is a dataset consisting originally of 230 chemical compounds assayed for mutagenicity in Salmonella typhimurium [143]. Among the 230 compounds, however, only 188 (125 positive, 63 negative) are considered to be learnable and thus are used in our simulations. The 188 chemical compounds are obviously represented by graphs. The aim is predicting whether each compound possesses mutagenicity.

**PPIs** (Protein-Protein Interaction) is a dataset collecting protein-protein interaction networks related to histidine kinase [96] (40 PPIs from Acidovorax avenae and 46 PPIs from Acidobacteria) [63]. The graphs describe the interaction relationships between histidine kinase in different species of bacteria. 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. The original dataset comprises 219 PPIs from 5 different kinds of bacteria with the following evolution order (from older to more recent): Aquifex 4 and Thermotoga 4 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. There is an additional class (Acidobacteria 46 PPIs) which is more controversial in terms of the bacterial evolution since they were discovered.

**PTC** (Predictive Toxicology Challenge) dataset records the carcinogenicity of several hundred chemical compounds for Male Rats (MR), Female Rats (FR), Male Mice (MM) and Female Mice (FM) [110]. These graphs are very small and sparse. We select the graphs of Male Rats (MR) for evaluation. There are 344 test graphs in the MR class.

**COIL** Columbia Object Image Library consists of 3D objects images of 100 objects [129]. There are 72 images per object taken in order to obtain 72 views from equally spaced viewing directions. For each view a graph was built by triangulating the extracted Harris corner points. In our experiments, we use the gray-scale images of only five objects, referring to it as COIL5.

**NCI1** The Anti-cancer activity prediction dataset represents a balanced subset of data sets of chemical compounds screened for activity against non-small cell lung cancer lines. [187].

As regards directed graphs we utilized the following datasets:

**Shock**  The Shock dataset consists of graphs from a database of 2D shapes [179]. Each graph is a medial axis-based representation of the differential structure of the boundary of a 2D shape. There are 150 graphs divided into 10 classes, each containing 15 graphs. The original version contains directed trees each with a root node, the undirected version has been created by removing the directionality.

**Alzheimer**  The dataset is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [6] and concerns interregional connectivity structure for fMRI (functional magnetic resonance imaging) activation networks for normal and Alzheimer subjects. Each image volume is acquired every two seconds with BOLD signals (Blood Oxygenation Level Dependent). 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 [188]. Subjects fall into four categories according to their degree of disease severity: AD - full Alzheimer's (30 subjects), LMCI - Late Mild Cognitive Impairment (34 subjects), EMCI - Early Mild Cognitive Impairment (47 subjects), HC - Normal Healthy Controls (38 subjects). The LMCI subjects are more severely affected and close to full Alzheimerś, while the EMCI subjects are closer to the healthy control group (Normal). 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. In order to fairly evaluate the influence caused by edges directionality, an undirected copy has been created as well. In particular, let $A_d$ be the adjacency matrix of a directed graph. Then its projection over the symmetric matrices space will be given by $(A_d + A_d^\top)/2$.

We use a binary C-Support Vector Machine (C-SVM) to test the efficacy of the QJSD kernel [43]. More specifically, we perform 10-fold cross validation, where for each sample we independently tune the value of C, the SVM regularizer constant, by considering the training data from that sample. The process is averaged over 100 random partitions of the data, and the results are reported in terms of average accuracy ± standard error. We recall that, in classification tasks, the accuracy is intended as the fraction of the data occurrences that are assigned the correct class label. In our case, an occurrence is a graph. Moreover, we contrast the performance of the kernel with that of other well-

established alternative graph kernels, namely the shortest-path kernel [31], the classic random walk kernel [76] and the the Weisfeiler-Lehman subtree kernel [160]. Except for the Weisfeiler-Lehman kernel, all the kernels do not consider graph attributes. In the case of the Weisfeiler-Lehman, nevertheless, each node is labelled with its degree. As concerns the maximum subtree height $h$, in our experiments we let $h = \{1 \cdots 3\}$ and we opted for the best value by cross-validation [160].

When we evaluate the accuracy of the QJSD kernel, we assume that the kernel is computed for $T \to \infty$. In the previous section, we showed that in the large time limit the density matrix of the quantum walk commutes with the Hamiltonian, and thus the value of the QJSD kernel can be easily calculated by eigendecomposition of $\mathscr{H}$. This, in turn, reduces the computational complexity, motivating in part the choice of $T \to \infty$. However, also results from the original work [154] influenced the preference. Specifically, experiments carried out over the QJSD kernel as a function of the time parameter $T$ suggest that for $T \to \infty$ the algorithm is able anyhow to reach a classification accuracy close to the optimum.

| Kernel | $ALZ_d$ | $ALZ_u$ | $SHOCK_d$ | $SHOCK_u$ |
|---|---|---|---|---|
| $QJSD_A$ | - | **$65.87 \pm 0.25$** | - | $41.48 \pm 0.15$ |
| $QJSD_L$ | $79.26 \pm 0.24$ | $60.42 \pm 0.23$ | $45.89 \pm 0.23$ | $35.77 \pm 0.21$ |
| $QJSD_{NL}$ | **$82.07 \pm 0.17$** | $61.45 \pm 0.22$ | **$46.05 \pm 0.20$** | **$44.38 \pm 0.21$** |
| SP | $59.86 \pm 0.25$ | $59.00 \pm 0.29$ | $22.09 \pm 0.29$ | $40.16 \pm 0.24$ |
| RW | $79.06 \pm 0.21$ | $58.71 \pm 0.21$ | $8.48 \pm 0.17$ | $24.34 \pm 0.28$ |
| WL | $70.87 \pm 0.27$ | $59.46 \pm 0.35$ | $38.74 \pm 0.27$ | $35.78 \pm 0.26$ |

Table 4.2: Classification accuracy ($\pm$ standard error) on directed and undirected counterpart graph datasets. $QJSD_{\mathscr{H}}$ is the QJSD kernel, where $\mathscr{H}$ denotes the Hamiltonian, *i.e.*, Adjacency matrix (A), Laplacian matrix (L) or normalized Laplacian matrix (NL), SP is the shortest-path kernel [31], RW is the random walk kernel [76] while WL is the Weisfeiler-Lehman subtree kernel [160].

We commence by commenting results reported in Table 4.2, regarding the directed graphs datasets. Here we evaluate the accuracy ($\pm$ standard error) of the QJSD kernel for different choices of the Hamiltonian. More specifically, we let the Hamiltonian be either the adjacency (only for undirected graphs) or the Laplacian or the normalized Laplacian matrix. Interestingly, we achieve the best performance with the normalized Laplacian matrix in almost all cases, with the exception for the Alzheimer in the undirected version. Moreover, for both dataset the best accuracy is given by the directed version which hence seems to better capture information from the graph topology. However, to contrast our outcomes with other studies we carried out further experiments. In particular we refer to the work of Wang *et al*. [191] where they use entropic measurements to distinguish

| Kernel | AD/Norm$_d$ | AD/Norm$_u$ | EMCI/LMCI$_d$ | EMCI/LMCI$_u$ |
|--------|-------------|-------------|---------------|---------------|
| QJSD$_A$ | - | $100.0 \pm 0.00$ | - | $100.0 \pm 0.00$ |
| QJSD$_L$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $98.58 \pm 0.04$ |
| QJSD$_{NL}$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ |
| SP | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $96.60 \pm 0.11$ | $96.68 \pm 0.09$ |
| RW | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ | $100.0 \pm 0.00$ |
| WL | $97.63 \pm 0.22$ | $99.38 \pm 0.18$ | $98.02 \pm 0.22$ | $99.63 \pm 0.12$ |

Table 4.3: Classification accuracy (± standard error) on directed and undirected counterpart graph datasets. QJSD$_{\mathscr{H}}$ is the QJSD kernel, where $\mathscr{H}$ denotes the Hamiltonian, *i.e.*, Adjacency matrix (A), Laplacian matrix (L) or normalized Laplacian (NL), SP is the shortest-path kernel [31], RW is the random walk kernel [76] while WL is the Weisfeiler-Lehman subtree kernel [160].

subjects falling into different categories. These experiments differ from the previous because here the aim is distinguishing subjects falling into only two classes, namely AD vs Normal and EMCI vs LMCL (see Table 4.3). Unlike Wang *et al.* [191], whose approach proved to be well-performing already, achieving a peak of ~ 91% of accuracy, in our case, the worst value is ~ 98% - but in the undirected forms, derived by symmetrizing and thus by definition with a loss of topological information. Even if the results are excellent, it must be pinpointed, to be fair, that in this last case the classes to be distinguished are quite far away from each other.

As far the Hamiltonian choice is concerned, for undirected graphs exclusively, results are shown in Table 4.4. The best values for each dataset are in bold font style whereas the second best values are in italic font style. Differently from previous outcomes, there is not an apparent homogeneity. However we can observe that in case of datasets with average graph size > 100 nodes the wave kernel in the Laplacian declination tends to be the best performer, while for the other dataset this assumption does not hold. The signature choice anyhow appears to provide an improvement to the MUTAG dataset and is the second best value also for PTC. The fact that NCI1 has outstanding outcomes only with WL is not unexpected at all [160] even if the Laplacian matrix is the best among the worst ones. Generally, however, the QJSD kernel obtains a classification accuracy which is better than that of the other kernels SP, RW, and WL (or sometimes comparable), regardless of how we set $\mathscr{H}$.

## 4.3.6   Conclusions

Graph-based representations are undoubtedly a powerful tool for modelling real-world complex systems. However, in terms of tractability a number of problem are posed, especially when it comes to apply pattern recognition or machine learning techniques. Even if kernel methods give a way to shift this representational issue, designing novel graph

| Kernel | MUTAG | PPI | PTC | NCI1 | COIL |
|--------|-------|-----|-----|------|------|
| $\text{QJSD}_A$ | $86.72 \pm 0.14$ | $78.71 \pm 0.30$ | $56.09 \pm 0.15$ | $66.90 \pm 0.03$ | $69.90 \pm 0.08$ |
| $\text{QJSD}_L$ | $84.92 \pm 0.18$ | $73.79 \pm 0.42$ | $\mathbf{59.70 \pm 0.16}$ | $\mathit{69.48 \pm 0.03}$ | $\mathit{70.72 \pm 0.07}$ |
| $\text{QJSD}_{NL}$ | $87.10 \pm 0.14$ | $74.63 \pm 0.37$ | $55.16 \pm 0.18$ | $66.36 \pm 0.03$ | $69.61 \pm 0.10$ |
| $\text{QJSD}_A^{hk}$ | $\mathbf{88.51 \pm 0.13}$ | $\mathit{81.56 \pm 0.34}$ | $58.76 \pm 0.14$ | $63.88 \pm 0.03$ | $69.68 \pm 0.06$ |
| $\text{QJSD}_L^{hk}$ | $86.36 \pm 0.16$ | $77.08 \pm 0.28$ | $57.63 \pm 0.13$ | $64.96 \pm 0.02$ | $70.24 \pm 0.06$ |
| $\text{QJSD}_{NL}^{hk}$ | $\mathit{87.79 \pm 0.12}$ | $74.38 \pm 0.37$ | $58.45 \pm 0.16$ | $64.01 \pm 0.04$ | $70.55 \pm 0.09$ |
| $\text{QJSD}_A^{wk}$ | $85.97 \pm 0.14$ | $74.91 \pm 0.32$ | $\mathit{58.91 \pm 0.13}$ | $63.52 \pm 0.05$ | $70.48 \pm 0.06$ |
| $\text{QJSD}_L^{wk}$ | $85.81 \pm 0.16$ | $\mathbf{84.66 \pm 0.26}$ | $58.01 \pm 0.13$ | $64.45 \pm 0.03$ | $\mathbf{71.34 \pm 0.05}$ |
| $\text{QJSD}_{NL}^{wk}$ | $87.61 \pm 0.16$ | $74.74 \pm 0.30$ | $57.46 \pm 0.16$ | $63.34 \pm 0.04$ | $70.31 \pm 0.06$ |
| SP | $84.98 \pm 0.16$ | $66.40 \pm 0.31$ | $56.89 \pm 0.71$ | $65.44 \pm 0.04$ | $70.50 \pm 0.13$ |
| RW | $78.02 \pm 0.20$ | $69.94 \pm 0.27$ | $55.59 \pm 0.01$ | $58.80 \pm 0.04$ | $21.03 \pm 0.22$ |
| GR | $81.93 \pm 0.17$ | $52.34 \pm 0.42$ | $56.20 \pm 0.10$ | $62.28 \pm 0.02$ | $67.22 \pm 0.11$ |
| WL | $84.62 \pm 0.23$ | $79.93 \pm 0.35$ | $55.64 \pm 0.20$ | $\mathbf{78.55 \pm 0.04}$ | $31.33 \pm 0.21$ |

Table 4.4: Classification accuracy (± standard error) on undirected graph datasets. $\text{QJSD}_{\mathcal{H}}$ is the QJSD kernel, where $\mathcal{H}$ denotes the Hamiltonian, *i.e.*, Adjacency matrix (A), Laplacian matrix (L) or normalized Laplacian matrix (NL). QJSD*indicates the heat kernel signature ($hk$) or the wave kernel signature ($wk$), SP is the shortest-path kernel [31], RW is the random walk kernel [76] while WL is the Weisfeiler-Lehman subtree kernel [160].

kernels is still an open challenge. In this work, we proposed an extension of a quantum inspired kernel where the similarity between the input structures is gauged through continuous-time quantum walks and quantum Shannon-Jensen divergence. More specifically, we investigated over the Hamiltonian choices, especially we proposed to take into consideration even attributed graphs, where features descriptors are obtained via HKS and WKS. Besides, we tested the kernel over directed graphs as well.

Our experimental validation has shown that the QJSD kernel can outperform state-of-the-art kernels in a graph classification task. We noted that, for the directed version, the best choice is always the normalized Laplacian matrix. On the other hand, for undirected graphs, it seems that the graph dimensionality may somehow influence the performance of the Hamiltonian and the use of kernel signature. In any case, on average, all the new choices did not lead to poorer results but only improvements.

Future work should first include further detailed studies in order to figure out whether there exists any connection between the kernel and the size of the graphs. Also it should be analysed in detail how the similarity matrices actually affect the kernel behaviour. Moreover, an investigation over the normalized Laplacian matrix and its influence in the directed graphs case could lead to uncovering some interesting properties.

# 5

# Quantum Thermodynamics of Time-Varying Graphs

In this Chapter we present a novel analysis of time-evolving networks, based on a thermodynamic representation of graph structure. We show how to characterize the evolution of time-varying complex networks by relating major structural changes to thermodynamic phase transitions. In particular, we derive expressions for a number of different thermodynamic quantities (specifically energy, entropy and temperature), which we use to describe the evolutionary behaviour of the network system over time. Since in the real world no system is truly closed and interactions with the environment are usually strong, we assume an open nature of the system. We adopt the Schrödinger picture as the dynamical representation of the quantum system over time. First, we compute the network entropy using a recent quantum–mechanical representation of graph structure, connecting the graph Laplacian to a density operator. Then, we assume the system evolves according to the Schrödinger representation. This allows us to obtain a measure of energy exchange through the estimation of a hidden time-varying Hamiltonian from the available network data. Using the thermodynamic relationship between changes in energy, entropy, pressure and volume, we recover the thermodynamic temperature. We assess the utility of the method on real-world time-varying networks representing complex systems in the financial and biological domains. We also compare and contrast the different characterizations provided by the thermodynamic variables (energy, entropy, temperature, and pressure). The study shows that the estimation of the time-varying energy operator strongly characterizes different states of a time evolving system and successfully detects critical events occurring during network evolution.

The remainder of the Chapter is organized as follows. We commence in Section 5.1 by outlining the work and clarifying the exact advantage of the present approach. In Section 5.2 we review the thermodynamic background and introduce the novel analysis thermodynamic method while Section 5.3 shows results of the resulting thermodynamic representation to a number of real-world time-varying networks. In Section 5.4 we generalized the analysis to size-changing networks, and finally in Section 5.5 we conclude discussing outcomes and future work.

# 5.1   Overview

The problem of identifying an effective and succinct characterization of complex network structure has been the ongoing focus of interest in network analysis for several decades [10, 13, 184]. Initially efforts focused around a static view of network structure and on developing ways to describe networks through identifying salient substructures, such as communities, hubs or clusters [65, 70, 133]. In these cases the resulting representation is defined by the connectivity structure [14], and this approach has proved effective for problems of both clustering and classification [52, 122]. Recently, on the other hand, dynamical network analysis has attracted more attention. In particular the analysis of the processes underlying network evolution and the detailed mechanisms involved, have both acquired an increasingly crucial role in network science. However, capturing the large-scale properties of a time evolving structure has proven to be an extremely challenging problem. An excellent framework for the study of complex networks relies on statistical physics and thermodynamics, connecting the macroscopic properties of a system to the behavior of microscopic particles [90, 95, 124]. In particular, thermodynamics defines the macroscopic properties of a system through three variables, subject to constraints imposed by the four laws of thermodynamics.

During the last years, the literature has witnessed a lot of work seeking to understand the evolution of time-varying networks by means of analogies and formal similarities with thermodynamics and quantum information. However, most attempts built their approach around a given Hamiltonian, *de facto* assuming a complete representation of the system to be studied. In this work we look at a different and less studied scenario, where the network is only a partial component of a larger system and its evolution is affected by the interaction with the larger environment. Following a classic renormalization approach, we assume that these forces act at a different time-scale, so that they can be effectively approximated by a (time-varying) potential term in the Hamiltonian. The fact that the effects of the environment are unknown, results in an unknown potential and thus an unknown effective Hamiltonian, which we need to estimate from the evolution of the network. This results in an inverse-problem formulation from the observed dynamics to estimate the thermodynamic quantities. With this formulation to hand, we aim at analysing, comparing, and contrasting the characterization quality of the obtained thermodynamics measures. The goal is to see whether major variations in the thermodynamic-variables are linked with fundamental topological changes and key events in the evolution of the system. In our approach the graph Laplacian at each time epoch is viewed as a quantum mixed state undergoing free evolution through the Schrödinger equation under an unknown time-dependent Hamiltonian. In particular, the Hamiltonian represents the change in potential due to external factors. Entropy and energy change with time as a result of direct interaction with the environment. From the observed evolution of the network, we estimate the Hamiltonian together with the Energy-exchange at each time epoch, as well as the variation in entropy of the underlying structure. Finally, from these variations we derive the actual thermodynamic variables of the evolving network system, including the free energy and temperature.

# 5.2  Quantum Thermodynamics of the Network

Let $G(V, E)$ be an undirected graph with node set $V$ and edges set $E \subseteq V \times V$ and let $A = a_{ij}$ be the adjacency matrix, where

$$a_{ij} = \begin{cases} 1, & v_i \sim v_j, \\ 0, & \text{otherwise.} \end{cases}$$

The degree $d$ of a node is the number of edges incident to the node and it can be represented through the degree matrix $D = (d_{ij})$ which is a diagonal matrix with $d_{ii} = \sum_i a_{ij}$. The graph Laplacian is then defined as $L = D - A$, and it can be interpreted as a combinatorial analogue of the discrete Laplace-Beltrami operator. The normalized Laplacian matrix $\tilde{L}$ is defined as

$$\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \tag{5.1}$$

In standard quantum mechanics, the state of a quantum mechanical system associated to the $n$–dimensional Hilbert space $\mathbf{H} \cong \mathbb{C}^n$ is identified by a $n \times n$ positive semidefinite hermitian matrix, with $\text{Tr}(\rho) = 1$, the *density matrix*. The *density operator* is introduced in quantum mechanics to describe a system whose state is an ensemble of pure quantum states $|\psi_i\rangle$, each with probability $p_i$. In other words, every density matrix can be written as a weighted sum of pure states, with real non-negative weights summing up to 1.

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| . \tag{5.2}$$

There are different ways to associate graphs to specific states or dynamics, *e.g.*, graph-states [36, 84] or spin networks [28, 32, 104]. However here we opt for another approach, the mapping between discrete Laplacian and quantum states, firstly introduced by Braunstein et al. [35](see also [85]). In the first place, it is easy to observe how the normalized Laplacian matrix, scaled by the number of vertices in the graph, has the characteristic features of a quantum mechanical density matrix: it is a positive semidefinite unit trace matrix, which actually provides a link to quantum states. In [140], Passerini and Severini suggest how the normalized Laplacian can be also seen as a density matrix in a quantum system representing a quantum superposition of the transition steps of quantum walk over the graph. Furthermore, the normalized Laplacian may be interpreted as a mixture of pure states (*i.e.*, , a convex combination of rank-1 operators), more precisely a uniformly random mixture of pure states in the vertex-space, where each state in the mixture corresponds to a single edge of the graph [51]. In this way, we can associate with any graph $G$ a specific mixed quantum state [85].

The Shannon entropy measures the uncertainty associated with a classical probability distribution. However, if we replace probability distributions with density operators, we can generalize the definition of the Shannon entropy to quantum states (actually described in a similar fashion), obtaining in fact the entropy of a quantum state. More to the point,

the *von Neumann entropy* [135] or *quantum entropy* $H_N$ of a mixed state is defined in terms of the trace and logarithm of the density operator $\rho$

$$H_N = -\text{Tr}(\rho \log \rho) = -\sum_i \xi_i \ln \xi_i \tag{5.3}$$

where $\xi_1, \ldots, \xi_n$ are the eigenvalues of $\rho$. In this sense, the quantum entropy is related to the distinguishability of the states, *i.e.*, , the amount of information that can be extracted from an observation of the mixed state. In [35] Braunstein et al., introduce the von Neumann entropy as a quantitative measure of mixedness of the density matrix $\rho$. This definition in turn is based on the mapping between quantum states and the combinatorial graph Laplacian, as discussed above. Many detailed studies followed this work [13, 14, 50, 51, 58, 140], but one in particular deserves more attention. Precisely, Passerini and Severini in [140] investigated the use of the normalized Laplacian as a density operator, to define the current state of the network, and allowing so to derive the network entropy in terms of the von Neumann entropy

$$S_{VN} = -\sum_{i=1}^{|V|} \frac{\tilde{\lambda}_i}{|V|} \ln \frac{\tilde{\lambda}_i}{|V|} \tag{5.4}$$

where $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ are the eigenvalues of $\tilde{L}$ and $|V|$ defines the number of vertices. Still in their work, they provide an interpretation of this quantity as a measure of regularity for graphs. The resulting intuition is thus that we can somehow measure the complexity of a graph through the von Neumann entropy. Nevertheless, the interpretation of this spectral measure, in terms of structural patterns, is as yet unclear, as pointed-out in [125]. Specifically, the ability of distinguish certain structural patterns is strictly influenced by the topology of the underlying graph. In other words, these kind of measures are extremely susceptible to structural properties of graphs, such as graph symmetries [155]. Consequently, we adopt this quantum approach in the analysis of network, not so much because we believe the system or its evolution to be quantum in nature, but due to the observed sensitivity to structural changes of these quantum-probing mechanisms.

We have already recalled in the previous Chapters that the *continuous-time quantum walk* is the quantum analogue of the continuous-time random walk [69], where the classical state vector is replaced by a vector of complex amplitudes over $V$. The general state of the walk is a complex linear combination of the basis states $|v\rangle$, $v \in V$, then the state of the walk at time $t$ is defined as

$$\left| \psi_t \right\rangle = \sum_{u \in V} \alpha_u(t) |u\rangle \tag{5.5}$$

where the amplitude $\alpha_u(t) \in \mathbb{C}$ and $\left| \psi_t \right\rangle \in \mathbb{C}^{|V|}$ are both complex. Moreover, we have that $\alpha_u(t)\alpha_u^*(t)$ gives the probability that at time $t$ the walker is at the vertex $u$, and thus $\sum_{u \in V} \alpha_u(t)\alpha_u^*(t) = 1$ and $\alpha_u(t)\alpha_u^*(t) \in [0,1]$, for all $u \in V$, $t \in \mathbb{R}^+$.

The evolution of the walk is then given by the Schrödinger equation, where we denote the (possibly time-dependent) Hamiltonian as $\mathcal{H}$.

$$\frac{\partial}{\partial t}|\psi_t\rangle = -i\mathcal{H}|\psi_t\rangle. \tag{5.6}$$

Time-dependent Hamiltonian, within the context of the Schrödinger equation, means that there is an external interaction of the system which manifests itself by a time-dependent potential term. In general, the Schrödinger's equation must be solved for the joint state describing system and environment together. However, for certain Hamiltonians and initial quantum states of the environment, the open system's dynamics can be well described by an effective (possibly time-dependent) potential term in the Hamiltonian, to some good approximation.

Here, we assume that the dynamics of the network is governed by a free evolution following the Schrödinger equation under an unknown time-varying Hamiltonian $\mathcal{H}_t$, together with an interaction with the external world, which acts as an observer thus affecting the entropy and exchanging energy with the system. Indeed, the free evolution does not change the thermodynamic variables and the cause of any variation in entropy has to be understood in terms of the external interaction process, which also causes an energy exchange.

To measure the energy exchange we need to recover the potential term expressed by the unknown Hamiltonian. In fact, the Hamiltonian acts as an energy operator, resulting in the following expression for the change in energy between state $\rho_t$ and $\rho_{t+1}$

$$dU = \mathrm{Tr}(\mathcal{H}_{t+1}\rho_{t+1}) - \mathrm{Tr}(\mathcal{H}_t\rho_t) \tag{5.7}$$

We estimate the Hamiltonian $\mathcal{H}_t$ as the one that minimizes the exchange of energy through the interaction with the external environment. To this end we assume that the interaction intervenes at the end of the free evolution, where the density matrix $\rho_t$ is transformed by the Schrödinger equation into

$$\hat{\rho}_t \;=\; \mathcal{U}(t,t_0)\rho_{t_0}\mathcal{U}^\dagger(t,t_0) \tag{5.8}$$

Approximating the Hamiltonian $\hat{\mathcal{H}}_{t_0}$ with a piece-wise constant operator, we can use the time-invariant Schrödinger equation to estimate the constant value of the Hamiltonian between time epochs $t$ and $t+1$ through the use of the unitary transformation $\mathcal{U}(t,t_0) = \exp(-i\hat{\mathcal{H}}_{t_0}(t-t_0))$, resulting in

$$\hat{\rho}_{t+1} \;=\; \exp(-i\hat{\mathcal{H}}_{t_0})\rho_{t_0}\exp(i\hat{\mathcal{H}}_{t_0}) \tag{5.9}$$

The exchange of energy in the interaction is then

$$\begin{aligned}
\Delta E_t &= \mathrm{Tr}(\hat{\mathcal{H}}_{t_0}\rho_{t+1}) - \mathrm{Tr}(\hat{\mathcal{H}}_{t_0}\hat{\rho}_{t+1}) \\
&= \mathrm{Tr}\big(\hat{\mathcal{H}}_{t_0}\big(\rho_{t+1} - \exp(-i\hat{\mathcal{H}}_{t_0})\rho_{t_0}\exp(i\hat{\mathcal{H}}_{t_0})\big)\big)
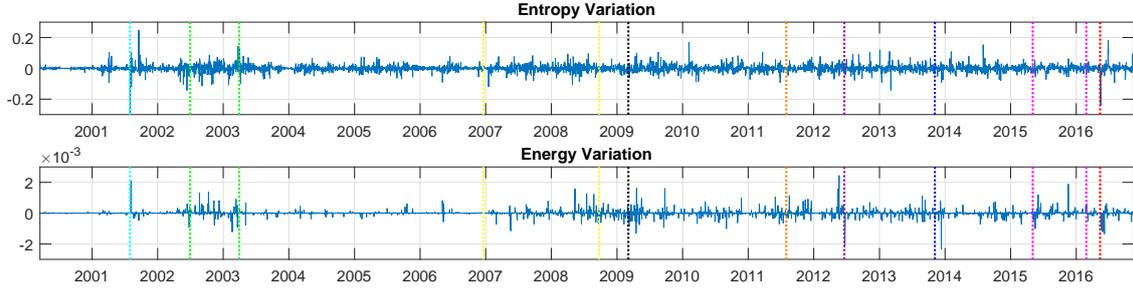\end{aligned} \tag{5.10}$$

Figure 5.1: USSM dataset - Entropy and Energy variation vs time, respectively top and bottom chart. The vertical coloured lines mark some important events for the trade market: September 11 attacks, Downturn of 2002-2003, Financial Crisis of 2007-2008, Dow Jones lowest point (March 2009), Stock Markets Fall (August 2011), Greek legislative election (June 2012), United States debt-ceiling crisis (October 2013), Chinese stock market turbulence 2015, Brexit Referendum (June 2016).
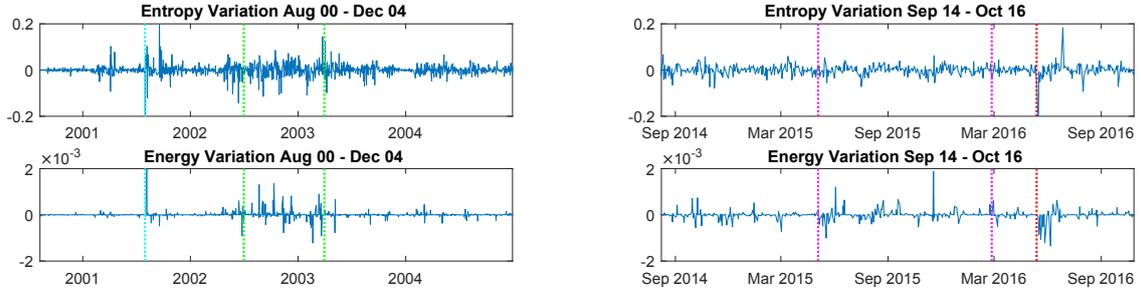


Figure 5.2: USSM dataset - Entropy and Energy variation vs time, respectively top and bottom, details. The vertical coloured lines refer to political and financial affecting the trade market. Left: September 11 attacks, Downturn of 2002-2003. Right: Chinese stock market turbulence 2015-16, Brexit Referendum (June 2016).

from which we can estimate the Hamiltonian as the operator that minimizes this quantity.

Let $\rho_t = \Phi_t \Lambda_t \Phi_t^T$ be the spectral decomposition of the state of the network at time $t$, equation (5.10) can be solved by noting that the minimum energy exchange intervenes when the interaction changes the eigenvalues of the density matrices, and with them the entropy, but does not change the corresponding eigenspaces. In other words, the Hamiltonian is the cause of the eigenvector rotation, while the interaction with the environment causes only changes in eigenvalues. Under this assumption, the Hamiltonian can be recovered from the rotation:

$$\hat{\mathcal{H}}_t \approx i \log(\Phi_{t+1}\Phi_t^T) \tag{5.11}$$

It is worth noting that we have computed a lower bound of the Hamiltonian, since we cannot observe components on the null spaces of $\rho$. Furthermore, we have

$$\underbrace{\Phi_{t+1}\Phi_t^T}_{\mathcal{U}} \rho_0 \underbrace{\Phi_t\Phi_{t+1}^T}_{\mathcal{U}} = \hat{\rho}_{t+1}, \tag{5.12}$$

where $\mathcal{U} = \Phi_{t+1}\Phi_t^T$ is the unitary evolution matrix. The final change in internal energy is

then

$$dU_t = Tr(\hat{\mathcal{H}}_{t+1}^- \rho_{t+1}) - Tr(\hat{\mathcal{H}}_t^+ \rho_t) = Tr(\hat{\mathcal{H}}_{t+\frac{1}{2}}(\rho_{t+1} - \rho_t)), \tag{5.13}$$

Where $\hat{\mathcal{H}}_{t+1}^-$ and $\hat{\mathcal{H}}_t^+$ are respectively the left and right limit of the top and bottom of the interval. Due to the piecewise-constant approximation we have

$$\hat{\mathcal{H}}_{t+1}^- = \hat{\mathcal{H}}_t^+ = \hat{\mathcal{H}}_{t+\frac{1}{2}}. \tag{5.14}$$

In the general time-varying case the evolution operator can be expressed as

$$\mathcal{U}(t, t_0) = \exp(-i\Omega(t, t_0)) \tag{5.15}$$

where $\Omega(t, t_0)$ can be obtained from the Hamiltonian through a Magnus series [118].

To estimate the exchange in energy, we compute the variation in energy from $\rho_t$ to $\hat{\rho}_{t+1}$ as

$$dU = \int_t^{t+1} \mathrm{Tr}\left(\mathcal{H}_s \rho_s\right) ds =$$

$$\int_t^{t+1} \mathrm{Tr}\left(\mathcal{H}_s \mathcal{U}(s,t) \rho_t \mathcal{U}^\dagger(s,t)\right) ds =$$

$$\int_t^{t+1} \mathrm{Tr}\left(i\frac{d\mathcal{U}(s,t)}{ds} \rho_t \mathcal{U}^\dagger(s,t)\right) ds =$$

$$\int_t^{t+1} \mathrm{Tr}\left(\mathcal{U}(s,t)\left(\frac{I - \exp(i\,\mathrm{ad}_\Omega)}{-i\,\mathrm{ad}_\Omega}\right)\Omega'(s,t)\rho_t \mathcal{U}^\dagger(s,t)\right) ds =$$

$$\mathrm{Tr}\left(\int_t^{t+1} \mathcal{U}(s,t)\left(\frac{I - \exp(i\,\mathrm{ad}_\Omega)}{-i\,\mathrm{ad}_\Omega}\right)\Omega'(s,t)\,ds\,\rho_t\right) =$$

$$\mathrm{Tr}\left(\Omega(t+1,t)\rho_t\right) \tag{5.16}$$

where the derivative of the exponential $\mathcal{U}(s,t) = \exp(-i\Omega(s,t))$ is

$$\frac{d\mathcal{U}(s,t)}{ds} = \mathcal{U}(s,t)\left(\frac{I - \exp(i\,\mathrm{ad}_\Omega)}{-i\,\mathrm{ad}_\Omega}\right)\Omega'(s,t) \tag{5.17}$$

with $\mathrm{ad}_X$ representing the adjoint action of a Lie algebra on itself [1]. As a consequence, the exchange in energy is

$$dU_t = Tr(\Omega(t+1,t)\rho_{t+1}) - Tr(\Omega(t+1,t)\rho_t) \tag{5.18}$$

This means that the unitary operator minimizing the exchange in energy is still

$$\mathcal{U} \approx \Phi_{t+1}\Phi_t^T \tag{5.19}$$

---

[1]The adjoint action is an endomorphism of Lie algebras. Let $G$ be a Lie group with Lie algebra $\mathfrak{g}$, then the operator is defined as $\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}$ with $\mathrm{ad}_X(Y) = [X, Y]$, that is a linear transformation of the Lie algebra where $[\,,\,]$ is the bracket operation in the algebra $\mathfrak{g}$.

and that logarithm of the unitary operator can be used as an effective "mean energy" operator to compute the energy differences.

With the energy operator to hand, the thermodynamic temperature $T$ can then be recovered through the fundamental thermodynamic relation $dU = TdS - PdV$ (5.20) but where we assume that the volume is constant, *i.e.*, , $dV = 0$ (isochoric process). As a result, the temperature $T$ is simply the rate of change of entropy with energy

$$T = \frac{dU}{dS}$$ (5.21)

This definition can be applied to evolving complex networks which do not change the number of nodes during their evolution.

## 5.3 Experiments and Evaluations

In this section, we explore the ability of the thermodynamic formulation to identify global topological changes in structure. In particular, we apply the analysis to three real-world time-evolving networks in order to assess the ability of the thermodynamic formulation to identify important topological transitions in network structure and characterize the overall dynamics of the system. First, we give a brief overview of the three datasets extracted from real-world complex systems, then we discuss the outcomes of the analysis. Wherever possible we make a comparison with similar approaches.

### 5.3.1 Datasets

Here, we use three data sets. The first and third sets are based on financial data, whereas the second comes from the biological domain.

*Dataset 1 – `NYSE`*: The dataset is extracted from a database consisting of the daily prices of 3799 stocks continuously traded on the New York Stock Exchange (`NYSE`). To construct the dynamic network, 347 stock with historical data from May 1987 to February 2011 are selected [141]. In order to construct an evolving network, a time window of 28 days is used and it is moved along time to obtain a sequence (from day 29 to day 6004); in this way, each temporal window contains a time-series of the daily return stock values over a 28 day period. Trades among the different stocks are represented as a network. For each time window, we compute the cross correlation coefficients between the time-series for each pair of stock and create connections between them if the absolute value of the correlation coefficient exceeds a threshold. The result is a stock market network which changes over the time, with a fixed number of 347 nodes and varying edge structure for each of trading days.

*Dataset 2 – Drosophila*: The dataset comes from the field of developmental biology, and concerns the interactions among genes of *Drosophila Melanogaster* - better known as the fruit fly - during its life cycle. The data is sampled at 66 sequential developmental time points. The fruit fly life cycle is divided into four stages, namely a) the embryonic
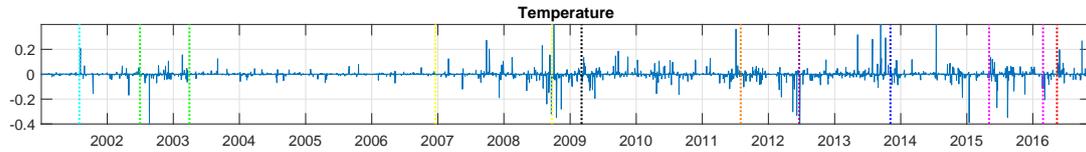
Figure 5.3: USSM dataset - Temperature vs time. The vertical coloured lines refer to some important events for the trade market: September 11 attacks, Downturn of 2002-2003, Financial Crisis of 2007-2008, Dow Jones lowest point March 2009, August 2011 stock markets fall, Greek legislative election June 2012, United States debt-ceiling crisis October 2013, Chinese stock market turbulence 2015-16, Brexit Referendum June 2016.

(samples 1-30),b) larval (samples 31-40) and c) pupal (samples 41-58) periods together with d) the first 30 days of adulthood (samples 59-66). Early embryos are sampled hourly and adults are sampled at multi-day intervals, according to the speed of the morphological changes. To represent this data using a time evolving network, the following steps are followed [167]. At each developmental point the 588 genes that are known to play an important role in the development of the *Drosophila* are selected. These genes are the nodes of the network, and edges are established based on the microarray gene expression measurements reported in [15]. To make the normalized Laplacian more tractable, any self-loop in the obtained undirected graph has been removed, at each time step. This dataset yields a time-evolving network with a fixed number of 588 nodes, sampled at 66 developmental time points.

*Dataset 3 – **USSM***: The dataset USSM is extracted from a database of the daily prices of 431 companies in 8 different sectors from the New York Stock Exchange and the NASDAQ Stock Market (NASDAQ). Data has been gathered from January 1995 to December 2016. The dataset is arranged to be around 5500 trading days and built similarly to the NYSE dataset, *i.e.*, a time window of 28 days is used and moved along time to obtain sequences of the daily return stock values. Then the cross correlation coefficients between each pair of stocks time-series are used to define links. The resulting network is an evolving stock market network where the number of nodes is fixed (equal to 431) but the edge structure is constantly changing.

## 5.3.2   Experiments

We aim at analysing how the thermodynamic variables describe network evolution and specifically if they can detect critical events as the system develops (*e.g.*, financial crises or crashes in the stock market and morphological changes during the fruit fly life cycle). To conduct our analysis, at each time interval, we computed the normalized Laplacian of the network which we used to estimate the Hamiltonian. Then entropy and energy state variables have been calculated using equations 5.4 and 5.7, respectively. However, it should be recalled we deal with open systems, which do not evolve towards an equilibrium. Therefore, we do not expect to observe any conservation in energy or a constant

increase in entropy either.

We commence by studying variations in thermodynamic quantities and consequent relationships between them. The first measures analysed are the variations in energy and entropy. We are particularly interested in these variations because if there is any pattern variation in these quantities, they are sufficient evidence of topological reactions in the networks. For instance, as mentioned in Section 5.2, the von Neumann entropy allows us to have an insight in terms of network complexity. Even if we do not have any direct link, a map, between entropy values and structural changes happening in the network, anyway we can see an effect of such alterations over the entropy trend.

In Fig. 5.1 we show their time series for the U.S. Stock Market data. Both quantities detect financial and political factors, which have influenced the structure of the trading network. However, the entropy variation sometimes appears somewhat noisier than the energy, probably because it is more susceptible even to slight structural mutations.

In Fig. 5.2 we study two financial crises in detail, namely the Downturn of 2002-2003 and Chinese stock market turbulence 2015-16. We observe how crisis are well-localised in energy dimension but less well so in entropy. Moreover, the entropy variation time series between 2004 and 2006 does not look so stable in comparison to the energy dimension. In fact, during this intermediate period, the energy exchange is small, and this corresponds to a period of prolonged network stability. In other words, energy-exchange proves to be effective in characterizing the network state as well as in the estimation of the hidden time-varying Hamiltonian which successfully extracts information concerning the change in network structure.

Next, we investigate the relationship between the previous thermodynamic quantities by studying the network temperature. This variable has been calculated similarly to Eq. 5.21. However, to reduce excessive surges or drops, we performed a Laplace smoothing of the series by adding a constant $\epsilon$ to the denominator, *i.e.*, to the entropy variation $dS$. The individual time-series for the thermodynamic variable shown in Fig. 5.3 clearly presents many significant fluctuations, most of them corresponding to some realistic major financial crises, *e.g.*, the Downturn of 2002-2003, as well as significant influential political events, *e.g.*, the Brexit referendum. Another interesting feature in the figure is that, although the network structure becomes particularly unstable after the Financial Crisis of 2007-2008, different crucial events are still detectable. This in turn tells us that this dimension provides the fundamental information to capture the market trend. Unlike the quantum entropy in graphs field, which is a well-established topic in the literature, the temperature, specifically its interpretation in terms of network analysis, has not reached a clear formalization yet. Nevertheless, Ye et al. in [205] have attempted to provide an insight about the connection between temperature and network behaviour, by defining the thermodynamic variables by means of node degree statistics for nodes connected by edges. In particular, according to their definitions, a low temperature is observed when there are large local variations in edge structure, whereas when there is a significant overall change in the number of edges, the temperature tends to become higher. Moreover, since this observation is based on the von Neumann entropy properties, such insights also apply in our formulation.

Figure 5.5 shows the scatter plot of energy versus entropy, for the U.S. Stock Market dataset. The main feature to note is that different epochs of network evolution form distinct clusters, occupying different regions of the plot. In fact, the plot reveals an interesting feature of the network whereby there exists a clustering tendency of the market when the system goes through strong modifications. Actually, each pattern presents a wide variation in entropy a low variation in energy. Thus, we conclude network states are better identified by energy, which effectively captures cluster compactness. Entropy on the other hand is more dispersive.

Interesting results can also be observed in the experiments on the Fruit Fly data too. In Fig. 5.4, we show an alternative representation of the network temperature, presented as the scatter plot of the difference in energy $dU$ over the difference in entropy $dS$. Most of the data clusters around a straight-line (shown in red), whose slope corresponds to the temperature of the network. However, there are also outliers exhibiting substantial departures from the main cluster.

Interestingly, these samples represent fundamental developmental landmarks in the evolution of the *Drosophila Melanogaster*. Specifically, they belong to the early embryonic stage, the mid larval transaction and final stage of the pupal step. According to [15],
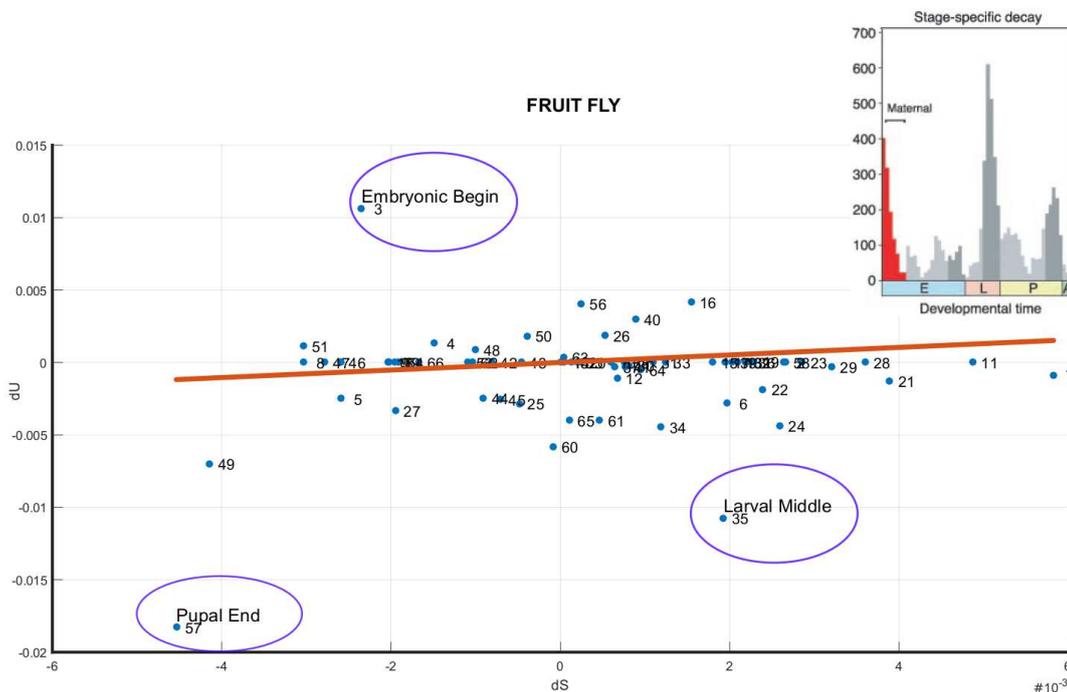


Figure 5.4: *Drosophila Melanogaster* dataset - Scatter plot of the difference of energy vs difference of entropy, as alternative representation of the network temperature. The red line fits the trend of the temperature. Outliers are epochs when the transcript levels of the genes has changed.
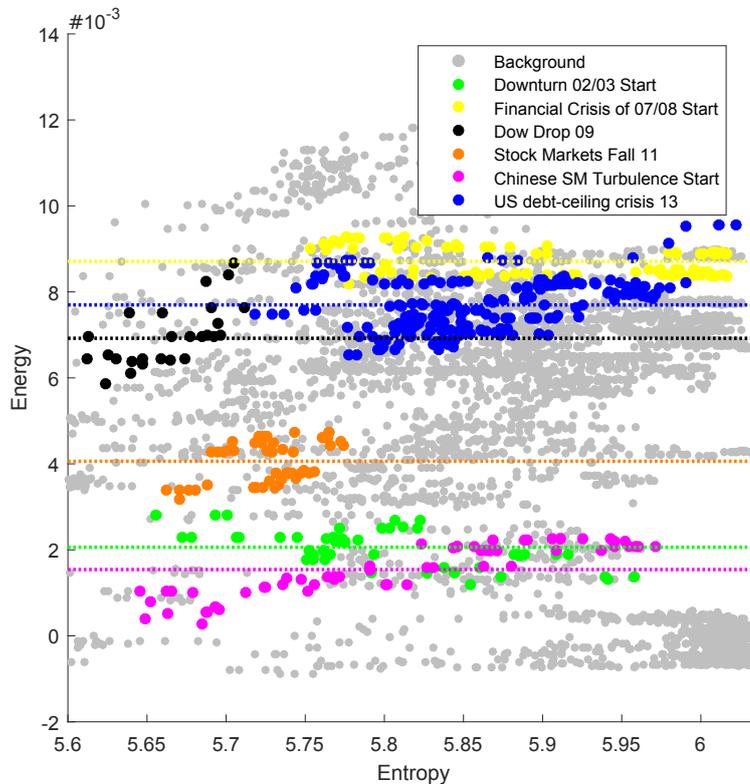
Figure 5.5: USSM dataset - Scatter plot of Energy ($U$) over Entropy ($S$). Each dot is a day and grey dots are the background. Dots of the same color belong to the same network phase. Horizontal lines represent cluster centroids for the energy dimension.

these are three of the four major morphological changes in terms of the transcript levels of genes. In fact, they mark the beginning and end of embryogenesis, the larval-pupal transition and the end of the pupal period, as the histogram in the upper right corner of Fig. 5.4 shows.

In addition, we analyse entropy and temperature. As before we performed Laplacian smoothing of the estimation of the temperature time series. In Fig. 5.6 the temperature (bottom time-series) clearly outperforms the entropy (top time-series), by detecting salient developmental changes during the fruit fly life cycle. Actually, entropy does not capture any information about the transcript level decline. Furthermore, by comparing the temperature dimension with the corresponding ones in [205] and [204], our method would seem to be able to capture genuine developmental stages effectively, rather than limiting the analysis to morphological changes alone.
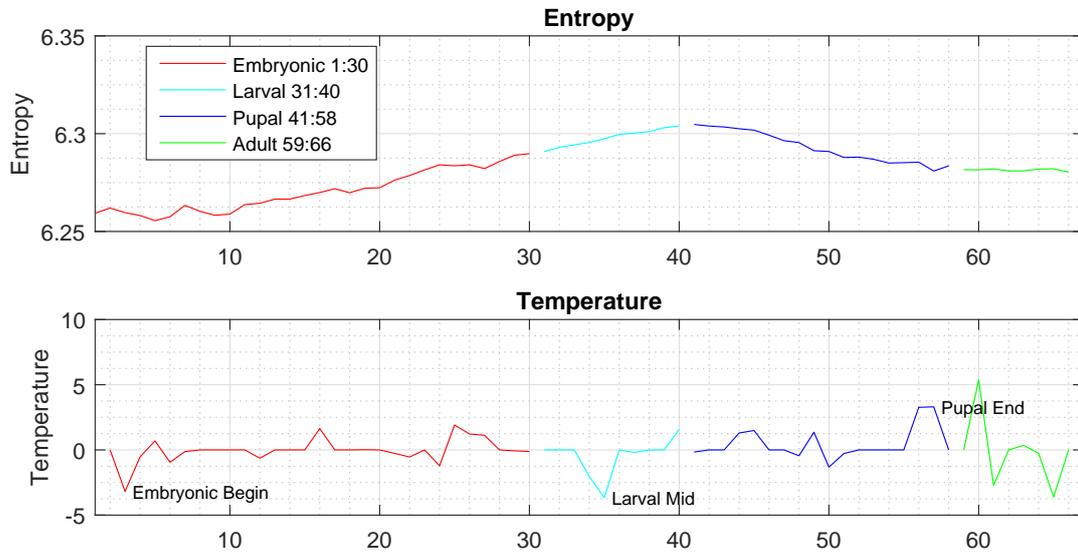
Figure 5.6: *Drosophila Melanogaster* dataset - Entropy and Temperature vs time, respectively top and bottom. The developmental periods of the fruit fly are coloured differently and the significant moments are highlighted.
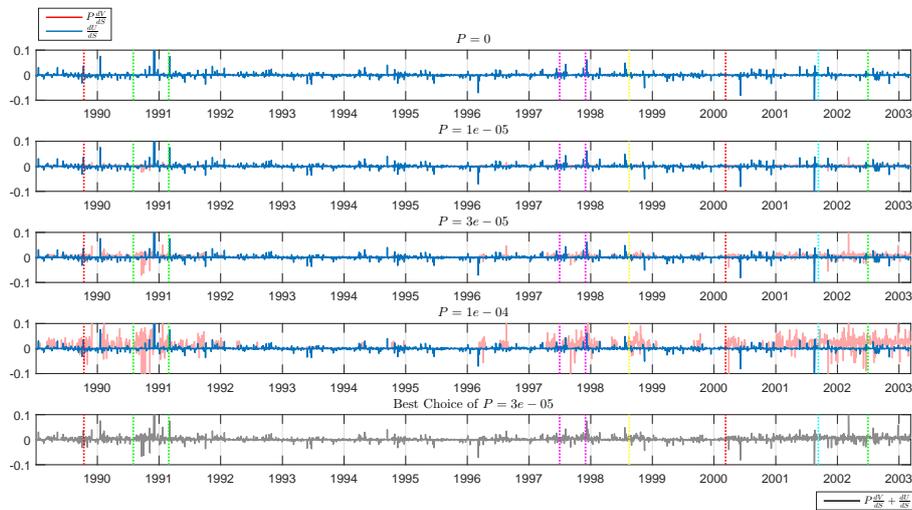


Figure 5.7: NYSE dataset - Temperature components vs time, as pressure changes. The vertical coloured lines indicate important events for the trade market: Friday the 13th Mini-Crash (October 1989), Persian Gulf War (August 1990 âĂŞ- January 1991), Asian Financial Crisis (July 1997 – October 1997), Russian Financial Crisis – Ruble devaluation (August 1998), Dot-com bubble - climax (March 2000), September 11 attacks, Downturn of 2002-2003.

## 5.4   Non-Isochoric processes

The method outlined so far was initially designed for analysing isochoric (constant volume) processes. Nevertheless, we can generalize the formulation to deal with variations in network size and, thus, in volume. We do this by assuming an isobaric process, *i.e.*, the pressure from the external environment is assumed independent of the system's evolution. By keeping a system at constant pressure, the change in volume becomes an integral part of the thermodynamic identity (see Eq. 5.20). Clearly, diverse pressure levels affect the temperature differently, as the volume variation effect changes. In fact, the temperature formula can be rewritten as

$$
\begin{aligned}
T \;&=\; \frac{dU + P\,dV}{dS} \\
&=\; \frac{dU}{dS} + P\,\frac{dV}{dS}
\end{aligned}
\tag{5.22}
$$

where $P\frac{dV}{dS}$ depends on the volume and pressure whereas $\frac{dU}{dS}$ depends on the energy variation.

     As an implementation note, to handle density matrices with different sizes. We overcome the dimension incongruity by padding with zeros the normalized Laplacian matrices. Then, once both the energy and the entropy variations have been recovered, we can compute the network temperature.

     In Fig. 5.7 we analyse the effect of the pressure parameter in the NYSE dataset. Here, the two temperature components are depicted simultaneously. When pressure is very small, the volume component does not contribute to the temperature. On the other hand, at high pressure values, the volume variation adds noise to the temperature. By balancing this term (in our case $P = 3 \cdot 10^{-5}$), the estimated temperature time-series accurately identifies occurrences of major events. In particular, the volume contribution often reinforces crucial parts of the signal, *e.g.*, during the Persian Gulf War or throughout the Russian Financial crisis, and only occasionally generates noise, for instance between the September 11 attacks and the beginning of the Downturn of 2002-2003. In fact, a comparison with [204] and [205] shows similar performance in terms of event-detection accuracy, proving that the generalization to size-varying graphs is still capable of modelling the underlying dynamical processes over the networks.

## 5.5   Discussion

In this work, we have introduced a novel thermodynamic framework to characterize network structure. Specifically, our driving goal was the visualization and understanding of the evolution of the time-varying network systems. This analysis is based on quantum thermodynamics and connects to recent work on the von Neumann entropy of networks. To this end, we provided expressions for a variety of thermodynamic variables, such as entropy, energy and temperature. In particular, energy is derived by estimating an unknown Hamiltonian operator governing the free evolution through the Schrödinger equation.

We have evaluated the method on real world complex systems, from the financial and biological domains. The experimental outcomes prove that the thermodynamic state variables are effective at characterizing the evolution properties of dynamic networks, including the detection of global topological changes, phase transitions in structure or distinctive periods in the evolution of time-varying complex networks. However, we observed that noise in the system's observables or overlap of consecutive significant events may affect the signals clarity, thus highlighting a certain level of sensitivity of the approach. Despite this, our objective of determining a correspondence between network evolution and related thermodynamic-variables has been mostly accomplished, although we are still far away from a truly and definitive mapping between thermodynamics measures and information provided by the system. Nevertheless, even without such a mapping, still we can indirectly get knowledge. For instance, by following Ye *et al.* [205] suggestions, where entropy is seen as a measure of degree change correlations on the edges, we can infer temperature significance. Indeed, a low temperature implies a large correlation. Conversely, the greater the disruption of the pattern of correlations, the higher is the temperature. Moreover, we were not able to identify clear precursors for the events, nor indicators of the scale of change; thus the approach as presented is mainly descriptive and, at this stage, not predictive.

There are many ways in which the work reported here can be improved. For instance, it would be interesting to study in greater depth the manifold structure of the thermodynamic space, and understand its links to phase transitions in the structure of networks.

# 6

# Conclusions and Future Work

In this thesis, we focused on problems involving the deeper questions posed by the structural analysis of graphs. These include developing generative models, characterizing graph structures by different perspectives and modelling of the dynamics defined over the networks. The leitmotiv behind our research has been to adopting analogies and formal similarities in order to relate one theoretical approach to another so that to acquire new intuitions and knowledge for a wide set of subjects. In fact, much of the work is rooted in a quantum information theory framework. This choice has been prompted by the recently uncovered connections between graphs and information theory. More to the point, we took advantage from new insights about the interpretation of the quantum entropy when applied to the domain of graphs, made possible by mapping between Laplacian matrix of a graph and quantum states.

We commenced our treatise with a brief synopsis of the more relevant literature, intended as a general outline for the reader. Then we moved on to the core of the thesis, where we presented our studies. Specifically, our contributions can be regarded as appertaining to three different themes, which share many common concepts though. In Chapter 3 we focused on generative models. Specifically, we used a generative approach initially designed for graphs in the context of $3D$ shape retrieval. The overarching aim in Chapter 4 has been to exploiting quantum processes to interpret and investigate problems concerning the structural analysis - an umbrella term referring to a wide-ranging variety of subjects though. Here, we dealt with network complexity, centrality indexes, quantum walks, graph similarities, but not only that. In the next section, we will anyhow detail each of them singularly. Finally, in Chapter 5 we proposed a novel thermodynamic framework for studying the properties of time-evolving complex networks.

## Contributions and Novelty

Generative models in graphs domain are important tools for learning the class-structure of data abstracted in terms of relational graphs. Indeed, in Chapter 3 we investigated if the application of a generative model originally designed for the general-purpose of graphs may lead to good outcomes also for non-rigid $3D$ shape retrieval task. Not so unexpectedly, the method worked fine and also allows to adopt a sampling scheme to

improve computation complexity.

In Chapter 4 quantum processes allowed us to undertake diverse analysis paths. Quantum information theory and in particular the well-known von Neumann entropy was the first topic treated in 4.1. Here we studied and compared the two variants of the von Neumann entropy along with the related quadratic approximations and we analysed how the graph topology can bias this measure as well. We found that 1) the two entropies lead to the emergence of similar structures, but with some significant differences; 2) the correlation between them ranges from weakly positive to strongly negative, depending on the topology of the underlying graph; 3) the quadratic approximations fail to capture the presence of non-trivial structural patterns that seem to influence the value of the exact entropies; 4) the quality of the approximations, as well as which variant of the von Neumann entropy is better approximated, depends on the topology of the underlying graph.

In 4.2 we proposed a novel edge centrality index rooted in quantum information. The importance of an edge is measured in terms of the difference in Von Neumann entropy between the original graph and the graph where that edge has been removed. Since this centrality appears higher on edges connecting low degree nodes we can deduce that it better gauges a structural component by the peripheral viewpoint. We also compared the proposed centrality measure to well-established alternatives, figuring out that the underlying topology of a graph influences the correlation with other centralities. Probably the reason stems from the fact that diverse random graph models differently behave at leaves level.

In the last part of Chapter 4, we extended the Rossi *et al.* work [154]. They devised a graph kernel aimed at probing the structure of graphs through continuous-time quantum walks and at defining similarity through the quantum Shannon-Jensen divergence. In our extension, we explored alternatives to the Laplacian and adjacency matrices, as Hamiltonian. We successfully applied the kernel to directed graphs and attributed graphs as well. In this latter case, we extracted feature points in terms of kernel signature and then worked with similarity matrices in place of the adjacency structures. Outcomes are promising, but as it is work in progress, these final considerations cannot be considered mature totally.

Finally, in Chapter 5 we adopted a thermodynamic representation of network structure in order to visualize and understand the evolution of time-varying networks. We provided expressions for thermodynamic variables on networks, including the entropy, energy and temperature. Even this analysis is connected to recent work on the von Neumann entropy of networks. Energy and temperature are derived by estimating an unknown Hamiltonian operator governing the free evolution through the Schrödinger equation. We have evaluated the method experimentally using data representing real world complex systems taken from the financial and biological domains. The experimental results proved that the thermodynamic variables are efficient in capturing abrupt changes and phase transitions in structure, as well as distinctive periods in the evolution of time-varying complex networks.

## 6.1   Future Work

There are several directions in which the work as a whole reported here can be extended. First, unless already applied, the work is expandable to directed graphs. For instance, as regards Chapter 5, in a recent paper [206] an approximation to the von Neumann entropy of directed graphs has been reported, which can be used to develop a similar framework for graph in which the edges posses directionality. This may provide particularly interesting in the context of characterising the thermodynamics of sources and sinks in a network. Also learning directed graphs could be a feasible continuation of the generative model in Chapter 3. As concerns the centrality index via Holevo quantity in Section 4.2, future work may include investigation of higher order approximations of this centrality measure as well as the possibility of defining network growth models based on the Holevo quantity. With respect to the analysis over the von Neumann entropy in Section 4.1 it is already under consideration a parallel work with directed graphs.

To conclude, it is worth mentioning that methods and studies here presented, like many others in this area of research, may open up fascinating possibilities of the analysis of data represented in terms of graphs. That means studying in detail the structure of real-world networks, leading so impact in a broad array of subject domains sampled from the physical, biological and social sciences. For instance, characterising the functionality of data furnished by brain imaging experiments, we may help in issues of major significance for the problem of brain-mapping, as well as new complexity indices might yield to surprising understanding of demographic networks.

# Bibliography

[6] Alzheimer's Disease Neuroimaging Initiative(adni). `http://adni.loni.usc.edu/`. Accessed: 2018-09-26.

[7] ABITEBOUL, S., BUNEMAN, P., AND SUCIU, D. *Data on the web: from relations to semistructured data and XML.* Morgan Kaufmann, San Francisco, 2000.

[8] AHARONOV, D., AMBAINIS, A., KEMPE, J., AND VAZIRANI, U. Quantum walks on graphs. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing* (2001), ACM, pp. 50–59.

[9] AHMADI, A., BELK, R., TAMON, C., AND WENDLER, C. On mixing in continuous-time quantum walks on some circulant graphs. *Quantum Information & Computation 3*, 6 (2003), 611–618.

[10] ALBERT, R., AND BARABÁSI, A.-L. Topology of evolving networks: local events and universality. *Physical review letters 85*, 24 (2000), 5234.

[11] ALBERT, R., AND BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics 74*, 1 (2002), 47.

[12] AMBAINIS, A. Quantum walks and their algorithmic applications. *International Journal of Quantum Information 1*, 04 (2003), 507–518.

[13] ANAND, K., AND BIANCONI, G. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E 80*, 4 (2009), 045102.

[14] ANAND, K., BIANCONI, G., AND SEVERINI, S. Shannon and von neumann entropy of random networks with heterogeneous expected degree. *Physical Review E 83*, 3 (2011), 036109.

[15] ARBEITMAN, M. N., FURLONG, E. E., IMAM, F., JOHNSON, E., NULL, B. H., BAKER, B. S., KRASNOW, M. A., SCOTT, M. P., DAVIS, R. W., AND WHITE, K. P. Gene expression during the life cycle of drosophila melanogaster. *Science 297*, 5590 (2002), 2270–2275.

[16] AUBRY, M., SCHLICKEWEI, U., AND CREMERS, D. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (2011), IEEE, pp. 1626–1633.

[17] BADAWY, O. E., AND KAMEL, M. Shape representation using concavity graphs. In *ICPR (3)* (2002), pp. 461–464.

[18] BAGDANOV, A. D., AND WORRING, M. First order gaussian graphs for efficient structure classification. *Pattern Recognition 36*, 6 (2003), 1311–1324.

[19] BAI, L., AND HANCOCK, E. Graph kernels from the Jensen-Shannon divergence. *Journal of Mathematical Imaging and Vision* (2012), 1–10.

[20] BAI, L., AND HANCOCK, E. R. Graph kernels from the jensen-shannon divergence. *Journal of mathematical imaging and vision 47*, 1-2 (2013), 60–69.

[21] BAI, L., ROSSI, L., TORSELLO, A., AND HANCOCK, E. R. A quantum jensen–shannon graph kernel for unattributed graphs. *Pattern Recognition 48*, 2 (2015), 344–355.

[22] BARABÁSI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *science 286*, 5439 (1999), 509–512.

[23] BASEI, G. Structural signatures through continuous-time quantum walks with decoherence. B.S. thesis, Università Ca'Foscari Venezia, 2015.

[24] BELKIN, M., SUN, J., AND WANG, Y. Discrete laplace operator on meshed surfaces. In *Proceedings of the Twenty-fourth Annual Symposium on Computational Geometry* (New York, NY, USA, 2008), SCG '08, ACM, pp. 278–287.

[25] BESL, P., AND MCKAY, N. D. A method for registration of 3-d shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 14*, 2 (Feb 1992), 239–256.

[26] BIANCONI, G., AND BARABÁSI, A.-L. Bose-einstein condensation in complex networks. *Physical review letters 86*, 24 (2001), 5632.

[27] BLEI, D. M., NG, A. Y., AND JORDAN, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res. 3* (Mar. 2003), 993–1022.

[28] BLINOV, B., MOEHRING, D., DUAN, L.-M., AND MONROE, C. Observation of entanglement between a single trapped atom and a single photon. *Nature 428*, 6979 (2004), 153.

[29] BONACICH, P. Power and centrality: A family of measures. *American journal of sociology 92*, 5 (1987), 1170–1182.

[30] BONCHEV, D., AND BUCK, G. A. Quantitative measures of network complexity. In *Complexity in chemistry, biology, and ecology*. Springer, 2005, pp. 191–235.

[31] BORGWARDT, K. M., AND KRIEGEL, H.-P. Shortest-path kernels on graphs. In *Data Mining, Fifth IEEE International Conference on* (2005), IEEE, pp. 8–pp.

[32] BOSE, S. Quantum communication through spin chain dynamics: an introductory overview. *Contemporary Physics 48*, 1 (2007), 13–30.

[33] BRANDES, U. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks 30*, 2 (2008), 136–145.

[34] BRAUNSTEIN, S. L., GHOSH, S., MANSOUR, T., SEVERINI, S., AND WILSON, R. C. Some families of density matrices for which separability is easily tested. *Physical Review A 73*, 1 (2006), 012320.

[35] BRAUNSTEIN, S. L., GHOSH, S., AND SEVERINI, S. The laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics 10*, 3 (2006), 291–317.

[36] BRIEGEL, H. J. Hj briegel and r. raussendorf, phys. rev. lett. 86, 910 (2001). *Phys. Rev. Lett. 86* (2001), 910.

[37] BRONSTEIN, A. M., BRONSTEIN, M. M., GUIBAS, L. J., AND OVSJANIKOV, M. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph. 30*, 1 (Feb. 2011), 1:1–1:20.

[38] BRONSTEIN, M. M., AND KOKKINOS, I. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (2010), IEEE, pp. 1704–1711.

[39] BROUWER, A. E., AND HAEMERS, W. H. *Spectra of graphs*. Springer Science & Business Media, 2011.

[40] BUNTINE, W. A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on knowledge and data engineering 8*, 2 (1996), 195–210.

[41] BURES, D. An extension of kakutani's theorem on infinite product measures to the tensor product of semifinite w*-algebras. *Transactions of the American Mathematical Society 135* (1969), 199–212.

[42] CARCASSONI, M., , CARCASSONI, M., AND HANCOCK, E. R. Alignment using spectral clusters. In *Proceedings of the 13th British Machine Vision Conference* (2002), pp. 213–222.

[43] CHANG, C., AND LIN, C. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST) 2*, 3 (2011), 27.

[44] CHEN, D.-Y., TIAN, X.-P., TE SHEN, Y., AND OUHYOUNG, M. On visual similarity based 3d model retrieval, 2003.

[45] CHILDS, A. M. Universal computation by quantum walk. *Physical review letters 102*, 18 (2009), 180501.

[46] CHILDS, A. M., CLEVE, R., DEOTTO, E., FARHI, E., GUTMANN, S., AND SPIELMAN, D. A. Exponential algorithmic speedup by a quantum walk. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing* (2003), ACM, pp. 59–68.

[47] CHUA, C. S., AND JARVIS, R. Point signatures: A new representation for 3d object recognition. *International Journal of Computer Vision 25*, 1 (1997), 63–85.

[48] CHUNG, F. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics 9*, 1 (2005), 1–19.

[49] CHUNG, F. R. *Spectral graph theory*. No. 92. American Mathematical Soc., 1997.

[50] DAIRYKO, M., HOGBEN, L., LIN, J. C.-H., LOCKHART, J., ROBERSON, D., SEVERINI, S., AND YOUNG, M. Note on von neumann and rényi entropies of a graph. *Linear Algebra and its Applications 521* (2017), 240–253.

[51] DE BEAUDRAP, N., GIOVANNETTI, V., SEVERINI, S., AND WILSON, R. Interpreting the von neumann entropy of graph laplacians, and coentropic graphs. *A Panorama of Mathematics: Pure and Applied 658* (2016), 227.

[52] DE DOMENICO, M., NICOSIA, V., ARENAS, A., AND LATORA, V. Structural reducibility of multilayer networks. *Nature communications 6* (2015), 6864.

[53] DE MONVEL, L. B., AND GUILLEMIN, V. *The Spectral Theory of Toeplitz Operators.(AM-99)*, vol. 99. Princeton university press, 2016.

[54] DEBNATH, A., LOPEZ DE COMPADRE, R., DEBNATH, G., SHUSTERMAN, A., AND HANSCH, C. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of medicinal chemistry 34*, 2 (1991), 786–797.

[55] DEHMER, M. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation 201*, 1-2 (2008), 82–94.

[56] DELAUNAY, B., ET AL. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk 7*, 793-800 (1934), 1–2.

[57] DELVENNE, J.-C., AND LIBERT, A.-S. Centrality measures and thermodynamic formalism for complex networks. *Physical Review E 83*, 4 (2011), 046117.

[58] DU, W., LI, X., LI, Y., AND SEVERINI, S. A note on the von neumann entropy of random graphs. *Linear Algebra and its Applications 433*, 11-12 (2010), 1722–1725.

[59] DURBIN, R., EDDY, S. R., KROGH, A., AND MITCHISON, G. J. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[60] EMMS, D., WILSON, R., AND HANCOCK, E. Graph embedding using a quasi-quantum analogue of the hitting times of continuous time quantum walks. *Quantum Information & Computation 9*, 3-4 (2009), 231–254.

[61] EPPSTEIN, D., PATERSON, M. S., AND YAO, F. F. On nearest-neighbor graphs. *Discrete & Computational Geometry 17*, 3 (1997), 263–282.

[62] ERDOS, P., AND RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci 5*, 1 (1960), 17–60.

[63] ESCOLANO, F., HANCOCK, E. R., AND LOZANO, M. A. Heat diffusion: Thermodynamic depth complexity of networks. *Physical Review E 85*, 3 (2012), 036206.

[64] ESTRADA, E. Quantifying network heterogeneity. *Physical Review E 82*, 6 (2010), 066102.

[65] ESTRADA, E. *The structure of complex networks: theory and applications*. OUP Oxford, 2011.

[66] ESTRADA, E., AND BENZI, M. Walk-based measure of balance in signed networks: Detecting lack of balance in social networks. *Physical Review E 90*, 4 (2014), 042802.

[67] ESTRADA, E., AND HATANO, N. Statistical-mechanical approach to subgraph centrality in complex networks. *Chemical Physics Letters 439*, 1 (2007), 247–251.

[68] ESTRADA, E., AND HATANO, N. Communicability in complex networks. *Physical Review E 77*, 3 (2008), 036111.

[69] FARHI, E., AND GUTMANN, S. Quantum computation and decision trees. *Physical Review A 58*, 2 (1998), 915.

[70] FELDMAN, D. P., AND CRUTCHFIELD, J. P. Measures of statistical complexity: Why? *Physics Letters A 238*, 4-5 (1998), 244–252.

[71] FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry* (1977), 35–41.

[72] FREEMAN, L. C. Centrality in social networks conceptual clarification. *Social networks 1*, 3 (1979), 215–239.

[73] FRIEDMAN, N., AND KOLLER, D. Being bayesian about network structure. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence* (2000), Morgan Kaufmann Publishers Inc., pp. 201–210.

[74] FRIEDMAN, N., LINIAL, M., NACHMAN, I., AND PE'ER, D. Using bayesian networks to analyze expression data. *Journal of computational biology 7*, 3-4 (2000), 601–620.

[75] GABRIEL, K. R., AND SOKAL, R. R. A new statistical approach to geographic variation analysis. *Systematic zoology 18*, 3 (1969), 259–278.

[76] GÄRTNER, T., FLACH, P., AND WROBEL, S. On graph kernels: Hardness results and efficient alternatives. In *Learning Theory and Kernel Machines*. Springer, 2003, pp. 129–143.

[77] GASPARETTO, A., MINELLO, G., AND TORSELLO, A. A non-parametric spectral model for graph classification. In *ICPRAM (1)* (2015), pp. 312–319.

[78] GIACHETTI, A., AND LOVATO, C. Radial symmetry detection and shape characterization with the multiscale area projection transform. *Computer Graphics Forum 31*, 5 (2012), 1669–1678.

[79] GILBERT, E. N. Random graphs. *The Annals of Mathematical Statistics 30*, 4 (1959), 1141–1144.

[80] GRAUMAN, K., AND DARRELL, T. The pyramid match kernel: discriminative classification with sets of image features. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (Oct 2005), vol. 2, pp. 1458–1465 Vol. 2.

[81] GUTMAN, I., AND ZHOU, B. Laplacian energy of a graph. *Linear Algebra and its applications 414*, 1 (2006), 29–37.

[82] HAN, L., ESCOLANO, F., HANCOCK, E. R., AND WILSON, R. C. Graph characterizations from von neumann entropy. *Pattern Recognition Letters 33*, 15 (2012), 1958–1967.

[83] HAUSSLER, D. Convolution kernels on discrete structures. Tech. rep., Technical report, UC Santa Cruz, 1999.

[84] HEIN, M. M. hein, j. eisert, and hj briegel, phys. rev. a 69, 062311 (2004). *Phys. Rev. A 69* (2004), 062311.

[85] HILDEBRAND, R., MANCINI, S., AND SEVERINI, S. Combinatorial laplacians and positivity under partial transpose. *Mathematical Structures in Computer Science 18*, 1 (2008), 205–219.

[86] HOFMANN, T. Unsupervised learning by probabilistic latent semantic analysis. *Mach. Learn. 42*, 1-2 (Jan. 2001), 177–196.

[87] HOLUB, A. D., WELLING, M., AND PERONA, P. Combining generative models and fisher kernels for object recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (2005), vol. 1, IEEE, pp. 136–143.

[88] HORN, R. A., HORN, R. A., AND JOHNSON, C. R. *Matrix analysis*. Cambridge university press, 1990.

[89] HU, N., RUSTAMOV, R. M., AND GUIBAS, L. Stable and informative spectral signatures for graph matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 2305–2312.

[90] HUANG, K. Statistical mechanics.

[91] HUANG, Q.-X., SU, H., AND GUIBAS, L. Fine-grained semi-supervised labeling of large shape collections. *ACM Trans. Graph. 32*, 6 (Nov. 2013), 190:1–190:10.

[92] HUGHES, R. J., ALDE, D. M., DYER, P., LUTHER, G. G., MORGAN, G. L., AND SCHAUER, M. Quantum cryptography. *Contemporary Physics 36*, 3 (1995), 149–163.

[93] ITO, T., CHIBA, T., OZAWA, R., YOSHIDA, M., HATTORI, M., AND SAKAKI, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences 98*, 8 (2001), 4569.

[94] JAIN, V., AND ZHANG, H. A spectral approach to shape-based retrieval of articulated 3d models. *Comput. Aided Des. 39*, 5 (May 2007), 398–407.

[95] JAVARONE, M. A., AND ARMANO, G. Quantum–classical transitions in complex networks. *Journal of Statistical Mechanics: Theory and Experiment 2013*, 04 (2013), P04019.

[96] JENSEN, L. J., KUHN, M., STARK, M., CHAFFRON, S., CREEVEY, C., MULLER, J., DOERKS, T., JULIEN, P., ROTH, A., SIMONOVIC, M., ET AL. String 8âĂŤa global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research 37*, suppl_1 (2008), D412–D416.

[97] JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z., AND BARABÁSI, A. The large-scale organization of metabolic networks. *Nature 407*, 6804 (2000), 651–654.

[98] JOHNSON, A. E., AND HEBERT, M. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1999), 433–449.

[99] JOST, J., AND JOST, J. *Riemannian geometry and geometric analysis*, vol. 42005. Springer, 2008.

[100] KALAPALA, V., SANWALANI, V., AND MOORE, C. The structure of the united states road network. *Preprint, University of New Mexico* (2003).

[101] KEMPE, J. Quantum random walks hit exponentially faster. *arXiv preprint quant-ph/0205083* (2002).

[102] KEMPE, J. Quantum random walks: an introductory overview. *Contemporary Physics 44*, 4 (2003), 307–327.

[103] KENDON, V. Decoherence in quantum walks–a review. *Mathematical Structures in Computer Science 17*, 6 (2007), 1169–1220.

[104] KIELPINSKI, D. D. kielpinski, c. monroe, and dj wineland, nature (london) 417, 709 (2002). *Nature (London) 417* (2002), 709.

[105] KOSCHÜTZKI, D., LEHMANN, K. A., PEETERS, L., RICHTER, S., TENFELDE-PODEHL, D., AND ZLOTOWSKI, O. Centrality indices. In *Network analysis*. Springer, 2005, pp. 16–61.

[106] KROVI, H., AND BRUN, T. Quantum walks with infinite hitting times. *Physical Review A 74*, 4 (2006), 042334.

[107] LAMBERTI, P., MAJTEY, A., BORRAS, A., CASAS, M., AND PLASTINO, A. Metric character of the quantum jensen-shannon divergence. *Physical Review A 77*, 5 (2008), 052311.

[108] LI, C., AND BEN HAMZA, A. A multiresolution descriptor for deformable 3d shape retrieval. *Vis. Comput. 29*, 6-8 (June 2013), 513–524.

[109] LI, C., AND HAMZA, A. Intrinsic spatial pyramid matching for deformable 3d shape retrieval. *International Journal of Multimedia Information Retrieval 2*, 4 (2013), 261–271.

[110] LI, G., SEMERCI, M., YENER, B., AND ZAKI, M. J. Effective graph classification based on topological and label attributes. *Statistical Analysis and Data Mining: The ASA Data Science Journal 5*, 4 (2012), 265–283.

[111] LIN, J. Divergence measures based on the shannon entropy. *Information Theory, IEEE Transactions on 37*, 1 (1991), 145–151.

[112] LITMAN, R., BRONSTEIN, A., BRONSTEIN, M., AND CASTELLANI, U. Supervised learning of bag-of-features shape descriptors using sparse coding. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 127–136.

[113] LITMAN, R., AND BRONSTEIN, A. M. Learning spectral descriptors for deformable shape correspondence. *IEEE transactions on pattern analysis and machine intelligence 36*, 1 (2014), 171–180.

[114] LOCKHART, J., MINELLO, G., ROSSI, L., SEVERINI, S., AND TORSELLO, A. Edge centrality via the holevo quantity. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2016), Springer, pp. 143–152.

[115] LÓPEZ-SASTRE, R., GARCÍA-FUERTES, A., REDONDO-CABRERA, C., ACEVEDO-RODRÍGUEZ, F., AND MALDONADO-BASCÓN, S. Evaluating 3d spatial pyramids for classifying 3d shapes. *Computers and Graphics 37*, 5 (2013), 473 – 483.

[116] LUO, B., AND HANCOCK, E. R. Structural graph matching using the em algorithm and singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 10 (2001), 1120–1136.

[117] LUO, B., WILSON, R. C., AND HANCOCK, E. R. A linear generative model for graph structure. In *International Workshop on Graph-Based Representations in Pattern Recognition* (2005), Springer, pp. 54–62.

[118] MAGNUS, W. On the exponential solution of differential equations for a linear operator. *Communications on pure and applied mathematics 7*, 4 (1954), 649–673.

[119] MAJTEY, A., LAMBERTI, P., MARTIN, M., AND PLASTINO, A. WoottersâĂŹ distance revisited: a new distinguishability criterium. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics 32*, 3 (2005), 413–419.

[120] MAJTEY, A., LAMBERTI, P., AND PRATO, D. Jensen-shannon divergence as a measure of distinguishability between mixed quantum states. *Physical Review A 72*, 5 (2005), 052310.

[121] MANNING, C. D., AND SCHÜTZE, H. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, Massachusetts, 1999.

[122] MARTÍN HERNÁNDEZ, J., LI, Z., AND VAN MIEGHEM, P. Weighted betweenness and algebraic connectivity. *Journal of Complex Networks 2*, 3 (2014), 272–287.

[123] MARTINS, A., SMITH, N., XING, E., AGUIAR, P., AND FIGUEIREDO, M. Nonextensive information theoretic kernels on measures. *The Journal of Machine Learning Research 10* (2009), 935–975.

[124] MIKULECKY, D. C. Network thermodynamics and complexity: a transition to relational systems theory. *Computers & chemistry 25*, 4 (2001), 369–391.

[125] MINELLO, G., ROSSI, L., AND TORSELLO, A. On the Von Neumann Entropy of Graphs. *ArXiv e-prints* (Sept. 2018).

[126] MOORE, C., AND RUSSELL, A. Quantum walks on the hypercube. In *International Workshop on Randomization and Approximation Techniques in Computer Science* (2002), Springer, pp. 164–178.

[127] MOWSHOWITZ, A. Entropy and the complexity of graphs: I. an index of the relative complexity of a graph. *The bulletin of mathematical biophysics 30*, 1 (1968), 175–204.

[128] MÜLKEN, O., AND BLUMEN, A. Continuous-time quantum walks: Models for coherent transport on complex networks. *Physics Reports 502*, 2-3 (2011), 37–87.

[129] NAYAR, S., NENE, S., AND MURASE, H. Columbia object image library (coil 100). *Department of Comp. Science, Columbia University, Tech. Rep. CUCS-006-96* (1996).

[130] NEUMANN, J. *Mathematical foundations of quantum mechanics*. Princeton university press, 1955.

[131] NEUMANN, J. V. *Mathematische grundlagen der quantenmechanik*, vol. 38. Springer-Verlag, 2013.

[132] NEWMAN, M. E. Scientific collaboration networks. i. network construction and fundamental results. *Physical review E 64*, 1 (2001), 016131.

[133] NEWMAN, M. E. The structure and function of complex networks. *SIAM review 45*, 2 (2003), 167–256.

[134] NEWMAN, M. E. A measure of betweenness centrality based on random walks. *Social networks 27*, 1 (2005), 39–54.

[135] NIELSEN, M., AND CHUANG, I. *Quantum computation and quantum information*. Cambridge university press, 2010.

[136] OHYA, M., AND PETZ, D. *Quantum entropy and its use*. Springer Science & Business Media, 2004.

[137] OSADA, R., FUNKHOUSER, T., CHAZELLE, B., AND DOBKIN, D. Shape distributions. *ACM Transactions on Graphics 21* (2002), 807–832.

[138] PARK, J., AND NEWMAN, M. E. Statistical mechanics of networks. *Physical Review E 70*, 6 (2004), 066117.

[139] PASSERINI, F., AND SEVERINI, S. The von neumann entropy of networks. *arXiv preprint arXiv:0812.2597* (2008).

[140] PASSERINI, F., AND SEVERINI, S.  Quantifying complexity in networks: the von neumann entropy. *International Journal of Agent Technologies and Systems (IJATS) 1*, 4 (2009), 58–67.

[141] PERON, T. D., AND RODRIGUES, F. A. Collective behavior in financial markets. *EPL (Europhysics Letters) 96*, 4 (2011), 48004.

[142] PICKUP, D., SUN, X., ROSIN, P. L., MARTIN, R. R., CHENG, Z., LIAN, Z., AONO, M., BEN HAMZA, A., BRONSTEIN, A., BRONSTEIN, M., BU, S., CASTELLANI, U., CHENG, S., GARRO, V., GIACHETTI, A., GODIL, A., HAN, J., JOHAN, H., LAI, L., LI, B., LI, C., LI, H., LITMAN, R., LIU, X., LIU, Z., LU, Y., TATSUMA, A., AND YE, J. SHREC'14 track: Shape retrieval of non-rigid 3d human models. In *Proceedings of the 7th Eurographics workshop on 3D Object Retrieval* (2014), EG 3DOR'14, Eurographics Association.

[143] RALAIVOLA, L., SWAMIDASS, S. J., SAIGO, H., AND BALDI, P. Graph kernels for chemical informatics. *Neural networks 18*, 8 (2005), 1093–1110.

[144] REUTER, M., ERICH WOLTER, F., AND PEINECKE, N. Abstract Laplace Beltrami spectra as Shape-DNA of surfaces and solids, 2005.

[145] REUTER, M., WOLTER, F.-E., AND PEINECKE, N. Laplace-spectra as fingerprints for shape matching. In *Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling* (New York, NY, USA, 2005), SPM '05, ACM, pp. 101–106.

[146] RIBEIRO, A. F. *Graph dynamics: learning and representation*. PhD thesis, Massachusetts Institute of Technology, 2006.

[147] ROBLES-KELLY, A., AND HANCOCK, E. R. Edit distance from graph spectra. In *ICCV* (2003), pp. 234–241.

[148] ROBLES-KELLY, A., AND HANCOCK, E. R. Graph edit distance from spectral seriation. *IEEE transactions on pattern analysis and machine intelligence 27*, 3 (2005), 365–378.

[149] ROSSI, L., AND TORSELLO, A. Measuring vertex centrality using the holevo quantity. In *International Workshop on Graph-Based Representations in Pattern Recognition* (2017), Springer, pp. 154–164.

[150] ROSSI, L., TORSELLO, A., AND HANCOCK, E. Approximate axial symmetries from continuous time quantum walks. *Structural, Syntactic, and Statistical Pattern Recognition* (2012), 144–152.

[151] ROSSI, L., TORSELLO, A., AND HANCOCK, E. Attributed graph similarity from the quantum jensen-shannon divergence. In *Similarity-Based Pattern Recognition.* Springer, 2013, pp. 204–218.

[152] ROSSI, L., TORSELLO, A., AND HANCOCK, E. A continuous-time quantum walk kernel for unattributed graphs. In *Graph-Based Representations in Pattern Recognition*. Springer, 2013, pp. 101–110.

[153] ROSSI, L., TORSELLO, A., AND HANCOCK, E. R. Node centrality for continuous-time quantum walks. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2014), Springer, pp. 103–112.

[154] ROSSI, L., TORSELLO, A., AND HANCOCK, E. R. Measuring graph similarity through continuous-time quantum walks and the quantum jensen-shannon divergence. *Physical Review E 91*, 2 (2015), 022815.

[155] ROSSI, L., TORSELLO, A., HANCOCK, E. R., AND WILSON, R. C. Characterizing graph symmetries through quantum jensen-shannon divergence. *Physical Review E 88*, 3 (2013), 032806.

[156] RUSTAMOV, R. M. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the fifth Eurographics symposium on Geometry processing* (2007), Eurographics Association, pp. 225–233.

[157] SANTHA, M. Quantum walk based search algorithms. In *International Conference on Theory and Applications of Models of Computation* (2008), Springer, pp. 31–46.

[158] SCHÖLKOPF, B., AND SMOLA, A. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[159] SHENVI, N., KEMPE, J., AND WHALEY, K. B. Quantum random-walk search algorithm. *Physical Review A 67*, 5 (2003), 052307.

[160] SHERVASHIDZE, N., SCHWEITZER, P., LEEUWEN, E. J. V., MEHLHORN, K., AND BORGWARDT, K. M. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research 12*, Sep (2011), 2539–2561.

[161] SHERVASHIDZE, N., VISHWANATHAN, S., PETRI, T., MEHLHORN, K., AND BORGWARDT, K. Efficient graphlet kernels for large graph comparison. In *Artificial Intelligence and Statistics* (2009), pp. 488–495.

[162] SHI, J., AND MALIK, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 8 (2000), 888–905.

[163] SHOR, P. W. Algorithms for quantum computation: Discrete logarithms and factoring. In *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on* (1994), Ieee, pp. 124–134.

[164] SIDDIQI, K., SHOKOUFANDEH, A., DICKINSON, S. J., AND ZUCKER, S. W. Shock graphs and shape matching. *International Journal of Computer Vision 35*, 1 (1999), 13–32.

[165] SILVERMAN, B. W. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.

[166] SIMMONS, D. E., COON, J. P., AND DATTA, A. The von neumann theil index: characterizing graph centralization using the von neumann index. *Journal of Complex Networks* (2018).

[167] SONG, L., KOLAR, M., AND XING, E. P. Keller: estimating time-varying interactions between genes. *Bioinformatics 25*, 12 (2009), i128–i136.

[168] SONG, Y.-Z., ARBELAEZ, P., HALL, P., LI, C., AND BALIKAI, A. Finding semantic structures in image hierarchies using laplacian graph energy. In *European Conference on Computer Vision* (2010), Springer, pp. 694–707.

[169] SPORNS, O. Network analysis, complexity, and brain function. *Complexity 8*, 1 (2002), 56–60.

[170] SUN, J., OVSJANIKOV, M., AND GUIBAS, L. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum* (2009), vol. 28, Wiley Online Library, pp. 1383–1392.

[171] TANGELDER, J. W., AND VELTKAMP, R. C. A survey of content based 3d shape retrieval methods. In *Shape Modeling Applications, 2004. Proceedings* (2004), IEEE, pp. 145–156.

[172] TELESFORD, Q. K., JOYCE, K. E., HAYASAKA, S., BURDETTE, J. H., AND LAURIENTI, P. J. The ubiquity of small-world networks. *Brain connectivity 1*, 5 (2011), 367–375.

[173] TODOROVIC, S., AND AHUJA, N. Extracting subimages of an unknown category from a set of images. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (2006), vol. 1, IEEE, pp. 927–934.

[174] TOLDO, R., CASTELLANI, U., AND FUSIELLO, A. The bag of words approach for retrieval and categorization of 3d objects. *The Visual Computer 26*, 10 (2010), 1257–1268.

[175] TORSELLO, A. *Matching hierarchical structures for shape recognition*. University of York, Department of Computer Science, 2004.

[176] TORSELLO, A. An importance sampling approach to learning structural representations of shape. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008), IEEE, pp. 1–7.

[177] TORSELLO, A., AND DOWE, D. L. Learning a generative model for structural representations. In *Australasian Joint Conference on Artificial Intelligence* (2008), Springer, pp. 573–583.

[178] TORSELLO, A., GASPARETTO, A., ROSSI, L., BAI, L., AND HANCOCK, E. R. Transitive state alignment for the quantum jensen-shannon kernel. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2014), Springer, pp. 22–31.

[179] TORSELLO, A., AND HANCOCK, E. Correcting curvature-density effects in the hamilton-jacobi skeleton. *Image Processing, IEEE Transactions on 15*, 4 (2006), 877–891.

[180] TORSELLO, A., AND HANCOCK, E. R. Learning shape-classes using a mixture of tree-unions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 6 (2006), 954–967.

[181] TORSELLO, A., AND HANCOCK, E. R. Graph embedding using tree edit-union. *Pattern recognition 40*, 5 (2007), 1393–1405.

[182] TORSELLO, A., AND ROSSI, L. Supervised learning of graph structure. In *International Workshop on Similarity-Based Pattern Recognition* (2011), Springer, pp. 117–132.

[183] UMEYAMA, S. An eigendecomposition approach to weighted graph matching problems. *IEEE transactions on pattern analysis and machine intelligence 10*, 5 (1988), 695–703.

[184] VAN DER HOFSTAD, R. Random graphs and complex networks. *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf* (2009), 11.

[185] VAPNIK, V. Statistical learning theory, 1998.

[186] VEDRAL, V. *Introduction to quantum information science*. Oxford University Press on Demand, 2006.

[187] WALE, N., AND KARYPIS, G. Acyclic subgraph based descriptor spaces for chemical compound retrieval and classification. Tech. rep., MINNESOTA UNIV MINNEAPOLIS DEPT OF COMPUTER SCIENCE, 2006.

[188] WANG, J., WILSON, R. C., AND HANCOCK, E. R. fmri activation network analysis using bose-einstein entropy. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2016), Springer, pp. 218–228.

[189] WANG, J., WILSON, R. C., AND HANCOCK, E. R. Network entropy analysis using the maxwell-boltzmann partition function. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (2016), IEEE, pp. 1321–1326.

[190] WANG, J., WILSON, R. C., AND HANCOCK, E. R. Thermodynamic network analysis with quantum spin statistics. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2016), Springer, pp. 153–162.

[191] WANG, J., WILSON, R. C., AND HANCOCK, E. R. Detecting alzheimer's disease using directed graphs. In *International Workshop on Graph-Based Representations in Pattern Recognition* (2017), Springer, pp. 94–104.

[192] WANG, J., WILSON, R. C., AND HANCOCK, E. R. Minimising entropy changes in dynamic network evolution. In *International Workshop on Graph-Based Representations in Pattern Recognition* (2017), Springer, pp. 255–265.

[193] WANG, J., WILSON, R. C., AND HANCOCK, E. R. Spin statistics, partition functions and network entropy. *Journal of Complex Networks* (2017).

[194] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of 'small-world' networks. *nature 393*, 6684 (1998), 440.

[195] WHITE, D., AND WILSON, R. C. Spectral generative models for graphs. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on* (2007), IEEE, pp. 35–42.

[196] WILSON, R., HANCOCK, E., ET AL. Generative graph prototypes from information theory. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1 (2015), 1–1.

[197] WILSON, R. C., HANCOCK, E. R., AND LUO, B. Pattern vectors from algebraic graph theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 7 (2005), 1112–1124.

[198] WILSON, R. C., AND ZHU, P. A study of graph spectra for comparing graphs and trees. *Pattern Recognition 41*, 9 (2008), 2833–2841.

[199] WONG, A., CONSTANT, J., AND YOU, M. Random graphs. In *Syntactic And Structural Pattern Recognition Theory And Applications*. World Scientific, 1990, pp. 197–234.

[200] WOOTTERS, W. Statistical distance and hilbert space. *Physical Review D 23*, 2 (1981), 357.

[201]  XIAO, B., AND HANCOCK, E. R.  A spectral generative model for graph structure.  In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (2006), Springer, pp. 173–181.

[202]  XIAO, B., WILSON, R., AND HANCOCK, E. R.  Object recognition using graph spectral invariants. In *ICPR* (2008), IEEE, pp. 1–4.

[203]  YADAV, M. J., AND JUNNARKAR, M. A.  A survey of graph matching techniques using quantum walks.

[204]  YE, C., COMIN, C. H., PERON, T. K. D., SILVA, F. N., RODRIGUES, F. A., COSTA, L. D. F., TORSELLO, A., AND HANCOCK, E. R.  Thermodynamic characterization of networks using graph polynomials. *Physical Review E 92*, 3 (2015), 032810.

[205]  YE, C., TORSELLO, A., WILSON, R. C., AND HANCOCK, E. R.  Thermodynamics of time evolving networks.  In *Graph-Based Representations in Pattern Recognition*. Springer, 2015, pp. 315–324.

[206]  YE, C., WILSON, R. C., COMIN, C. H., COSTA, L. D. F., AND HANCOCK, E. R. Approximate von neumann entropy for directed graphs. *Physical Review E 89*, 5 (2014), 052804.

[207]  YE, J., YAN, Z., AND YU, Y.  Fast nonrigid 3d retrieval using modal space transform.  In *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval* (New York, NY, USA, 2013), ICMR '13, ACM, pp. 121–126.

[208]  YU, M., ATMOSUKARTO, I., LEOW, W. K., HUANG, Z., AND XU, R.  3d model retrieval with morphing-based geometric and topological feature maps.  In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2003), pp. 656–661.

[209]  ZHANG, C., , ZHANG, C., AND CHEN, T.  Indexing and retrieval of 3d models aided by active learning, 2001.

# Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente:    GIORGIA MINELLO            matricola: 797636

Dottorato:    INFORMATICA

Ciclo:        31

Titolo della tesi[1] : QUANTUM PROCESSES FOR STRUCTURAL ANALYSIS

Abstract:

Systems modelled as networks, also called graphs, appear in many scenarios. Here we address matters concerning graph structural analysis by exploiting approaches based on quantum processes and the von Neumann entropy. We begin by studying spectral generative models for learning structural representations. Then we move on to quantum models and the von Neumann entropy characterization. Finally, we introduce a novel thermodynamic method to model time evolving networks.

Riassunto:

I sistemi modellati come reti, chiamati anche grafi, appaiono in molti scenari. Qui affrontiamo temi riguardanti l'analisi strutturale dei grafi sfruttando degli approcci basati sui processi quantistici e l'entropia di von Neumann. Iniziamo studiando dei modelli generativi spettrali per l'apprendimento di rappresentazioni strutturali. Poi passiamo ai modelli quantistici e alla caratterizzazione dell'entropia di von Neumann. Infine, introduciamo un nuovo metodo termodinamico per modellare le reti che evolvono nel tempo.

Firma dello studente

_____

---

[1] Il titolo deve essere quello definitivo, uguale a quello che risulta stampato sulla copertina dell'elaborato consegnato.