Stockholms universitet

Spectral Distance between Complex Networks using Graph Laplacians

Marina Herrera Sarrias

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2021:9
Matematisk statistik
Juni 2021

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2021:9**
http://www.math.su.se

Spectral Distance between Complex Networks using Graph Laplacians

Marina Herrera Sarrias*

June 2021

Abstract

In nature, there exist many complex networks with structures representing real system interactions and dynamics. Over the last two decades, the study and use of complex networks have extensively increased in many fields, such as biology, meteorology, neuroscience, and social studies. In order to establish a comparison between complex networks, a distance measure capturing the network structural differences has to be established. Many different distance measures, such as, the *Hamming distance*, have been proposed, although it is worth noting that most of them focus on comparing the number of nodes and edges between graphs rather than the structure and dynamics of the network. Amongst the research already initiated in the context of spectral similarity measures, Shimada *et al.* (2016), introduced in their paper "Graph distance for complex networks" a new graph distance called the *spectral graph distance* (SGD), defined in terms of the unnormalized graph Laplacian along with its associated eigenvectors. This thesis will investigate the statistical properties related to the spectral graph distance, both theoretically and by using simulated graphs. An overview of related topics in graph theory and network theory will also be presented.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: marina.sarrias@gmail.com. Supervisor: Chun-Biu Li.

Acknowledgements

I wish to express my gratitude, first and foremost, to my supervisor, professor Chun-Biu Li, whose guidance and support helped me grow in so many ways. This academic journey was also the beginning of many great friendships - to all the friends I made along the way, thank you! With special considerations for my dear Nga and Bao. I also wish to thank my partner Arthur, for his unconditional love and support throughout the years. Last but not least, thanks to my sister Elora, for always being there for me.

Contents

1 Introduction

Network theory has gained enormous amount of attention in previous years, primarily as it has provided answers to practical questions in many fields concerned in the study of natural systems phenomena. Throughout the last decade, *complex networks*, i.e., graphs with non-trivial topological features, have been a central topic of the research effort. As a result, new concepts and measures to identify statistical properties on the topology of complex networks have been proposed. Structural properties of complex networks are commonly explained by the arrangement of its edges. A common approach to understand the evolution of the underlying mechanism responsible for the appearance of such properties, is to generate networks able to reproduce such structural properties in a strictly controlled environment. Besides the study of structural properties, one of the many questions that can arise in network theory is how to quantify similarities between networks as a mean to compare them. Extensive research within different frameworks has already been initiated on the subject. For instance, in the context of spectral graph theory, spectral similarity measures denote a family of graph distances based on the spectra of the graph's connectivity matrices. The motivation for using the spectra derives from previous literature [1, 2], where it has been shown that a link can be established between the spectra of connectivity matrices and the structure and dynamics of a network. Following on this framework of analysis, Shimada *et al.* [3] proposed a novel tool called the spectral graph distance, defined by the empirical distribution function of the eigenvectors associated with the unnormalized Laplacian matrix. Using this graph distance measure, they succeeded in quantifying the structural distance between complex networks and benchmarked the method using graphs sampled from the Watts-Strogatz model. The spectral graph distance will be the main focus of this thesis.

1.1 Thesis outline

The outline of this thesis is as follows. Following an introductory overview of basic concepts in graph theory and network theory Section 2 will introduce the subject of matter of the thesis: the spectral graph distance. Two important types of simulated networks will be presented - the random graphs and the small work networks - along with different models to generate them - the Erdős-Rényi graph model and the Watts-Strogatz model. Next, different approaches for constructing graphs from data will be considered, the discussion leading to a review of graph matrices and the graph Laplacians. To conclude Section 2, the theory behind the spectral graph distance, as well as the technical aspects of the method will be presented. In Section 3, experiments on simulated graphs are performed and discussed. Finally, Section 4 presents the conclusions and Section 4.1 provides a general discussion on the spectral graph distance (SGD) and an outlook for further studies.

1.2 Objectives

The present work intends to provide a thorough introduction to the spectral graph distance, a spectral similarity measure based on the graph Laplacian. The main objective of this thesis project is to investigate the statistical properties of the graph distance, theoretically and by performing experiments on simulated graphs, as well as evaluating its performance and possible limitations and artifacts. An overview of related topics is provided to the reader to motivate the method. Focus will be put on general properties related to the different forms of defining the graph Laplacian, as well as the connection between them. Different ways of imposing a graph structure on the data will also be covered, in relation to fundamental concepts of graph theory and network theory.

2 Theory

2.1 Graph and Graph Theory

Graph theory was first introduced in the 18th century, with Euler's negative resolution of the Seven Bridges of Königsberg problem. Since then, the study of graphs and their mathematical properties have extensively increased. Graph theory is now widely used in a great variety of areas, from computer science to linguistics. In epidemiology, for instance, it is used to study infection contact patterns in a population; in biology, to represent protein structures and model protein interaction; in linguistics, to analyse language structures. This section will introduce some basic notation for defining graphs in connection with some important graph properties.

2.1.1 Basic Notation of Graphs

A graph $G = (\mathcal{N}, \mathcal{L})$ is defined by a finite set of nodes – or vertices – $\mathcal{N}$, and links – or edges – $\mathcal{L}$, where the nodes represent the objects of the graph, and the links the pairwise connections between them. Two distinct nodes n_i and n_j in G, such that $n_i \in \mathcal{N}$ and $n_j \in \mathcal{N}$, are connected if there exists a link $l_{ij} = (n_i, n_j) \in \mathcal{L}$ connecting the two nodes. The link is said to be *directed* if the link (n_i, n_j) is *outgoing* from n_i and *coming* to n_j – meaning that the order of the nodes in the link is relevant, such that $l_{ij} \neq l_{ji}$. It is said to be *undirected* if the link can be traversed in both directions – meaning that the order of the nodes is not relevant, such that $l_{ij} = l_{ji}$. Links between nodes may sometimes carry a weight, which is a numerical value that measures the strength of the connection between each pair of nodes – such graphs are called *weighted graphs*.

The number of nodes $N = |\mathcal{N}|$ where $|\cdot|$ denotes the cardinality of a set, is referred to as the *size* of G and the number of edges $K = |\mathcal{L}|$ is referred to as the *order* of G. For determining an order in the set of nodes, these are usually identified by labeling them with an integer index going from 1 to N. Nodes that are connected by a link are said to be *neighboring* or *adjacent*.

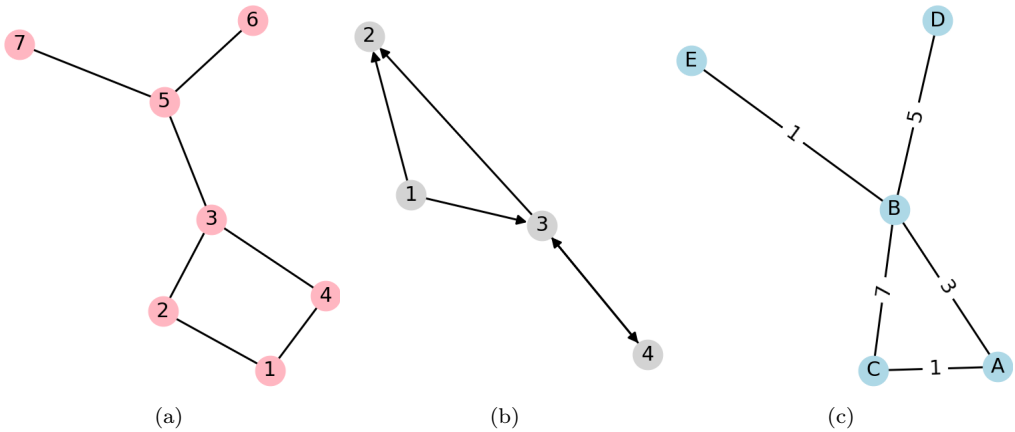
Some of the most common ways of defining graphs given in [4] are the following:

Definition 2.1.1 (**Undirected graph**). *An undirected graph $G = (\mathcal{N}, \mathcal{L})$ consists of two ordered sets $\mathcal{N} \neq \emptyset$ and $\mathcal{L}$. The elements of $\mathcal{N} = \{n_1, n_2, ..., n_N\}$ are distinct and are called the nodes, or vertices of the graph G. The elements of $\mathcal{L} = \{l_1, l_2, ..., l_K\}$ are distinct unordered pairs of distinct elements of $\mathcal{N}$ and are called links or edges.*

Definition 2.1.2 (**Directed graph**). *A directed graph $G = (\mathcal{N}, \mathcal{L})$ consists of an ordered pair of sets, $\mathcal{N} \neq \emptyset$ and $\mathcal{L}$. The elements of $\mathcal{N} = \{n_1, n_2, ..., n_N\}$ are the nodes of the graph G. The elements of $\mathcal{L} = \{l_1, l_2, ..., l_K\}$ are distinct ordered pairs of distinct elements of $\mathcal{N}$ and are called directed links.*

Definition 2.1.3 (**Weighted graph**). *A weighted graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ consists of a set of nodes $\mathcal{N} = \{n_1, n_2, ..., n_N\} \neq \emptyset$, a set of links $\mathcal{L} = \{l_1, l_2, ..., l_K\}$ and a weight function $\omega : \mathcal{N} \times \mathcal{N} \to \mathbb{R}$.*

Fig. 1 draws three basic ways of defining graphs: the graph in (a) refers to an undirected and unweighted graph; (b) to a directed and unweighted graph, where the arrows indicate the direction of the links; and (c) to an undirected and weighted graph, where the numerical values of the edges denote the weights.



(a) (b) (c)

Figure 1: Graphs examples: (a) undirected graph, (b) directed graph, and (c) weighted graph

The number of links each node has or the pairwise reachability of distinct nodes provides local and global information about the graph. Throughout this thesis, unless stated otherwise, only undirected and possibly weighted graphs will be considered. Some basic notations in graph theory given in [4] are:

Definition 2.1.4 (**Node degree**). *Given a graph $G = (\mathcal{N}, \mathcal{L})$ with $|\mathcal{N}| = N$, the node degree $deg(n_i) = |\{l_{ij} : l_{ij} \in \mathcal{L},\ 1 \leq j \leq N\}|$ of a node $n_i \in \mathcal{N}$, $1 \leq i \leq N$ is the number of links connected to node n_i. In the case of a weighted graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ the degree of a node n_i is the sum of the edges weights between n_i and the rest of*

nodes in the graph, written as:

$$deg(n_i) = \sum_{j=1}^{N} \omega(n_i, n_j)$$

.

Definition 2.1.5 (**Subgraph**). *A subgraph of $G = (\mathcal{N}, \mathcal{L})$ is a graph $G' = (\mathcal{N}', \mathcal{L}')$ such that $\mathcal{N}' \subseteq \mathcal{N}$ and $\mathcal{L}' \subseteq \mathcal{L}$. G' is said to be a graph generated by $\mathcal{N}'$ if G' contains all links of G that join two nodes in $\mathcal{N}'$.*

Definition 2.1.6 (**Walk**). *Given a graph $G = (\mathcal{N}, \mathcal{L})$. A walk $W(n_i, n_j)$ from node n_i to node n_j is a sequence alternating nodes and edges, that begins with n_i and ends with n_j. Let $(i_0, i_1, ..., i_k) \in [1, |\mathcal{N}|]$, be a sequence of indices such that $i_0 = i$ and $i_k = j$. Then let $l_p = (n_{i_{p-1}}, n_{i_p})$, and $W(n_i, n_j) = (n_i = n_{i_0},\ l_1,\ n_{i_1}, ..., l_k,\ n_{i_k} = n_j)$ A walk is usually indicated only by the sequence of traversed nodes: $W(n_i, n_j) = (n_i = n_{i_0},\ n_{i_1}, ...,\ n_{i_k} = n_j)$.*

Definition 2.1.7 (**Path**). *A path is a walk in which there are no repeated nodes, meaning that all nodes in the walk are different.*

Graphs can also be classified depending on their structural properties, their order, and their size. Let $G = (\mathcal{N}, \mathcal{L})$ be an undirected graph of size N and order K. Some examples are listed below:

- A *regular graph* is an undirected graph in which all nodes have the same degree.
- A *complete unweighted graph* is a special case of a regular graph, in which all nodes in the graph have a degree of $N - 1$, meaning that every pair of distinct nodes is connected by an edge.
- A *connected graph* - or single component graph - is a graph that for any pair of nodes n_i and n_j such that $n_i \in \mathcal{N}$ and $n_j \in \mathcal{N}$, there is a *path* from n_i to n_j.
- A *sparse graph* refers to a type of graph with low density, meaning that the graph order K is small in comparison to the maximum order $N(N-1)/2$ the graph could possibly have.

2.1.2 Graph drawing

As introduced in [4], *graph drawing* is an important area in mathematics and computer science concerned with the study of methods for graph visualization. The *drawing of a graph* is a geometric representation of the graph in a plane. It is important to know that the drawing of a graph is not the graph itself, as there can be many different ways of drawing the same graph, with each nodes' possible arrangement - also known as *layout* - highlighting different aspects of the graph. A *spring-based layout*, for instance, treats the graph as a physical system where the nodes act as repelling objects and the edges as springs, attracting neighboring nodes in space, thereby facilitating clustering. In a *tree-like layout*, nodes are arranged following a hierarchical structure given a root node. In a *circular layout*, nodes are evenly spaced on a circle, and the edges are usually drawn across the circle where nodes are ordered in such a way that they are closer to their neighbors to minimize edges crossing. The readability of the drawing of a graph tends to worsen as the graph size increases, making it more pertinent for graphs of smaller size.

2.2 Preliminary Network Theory

An important area of graph theory is *network theory*, which centers on the study of networks. Networks are graphs that are used to depict and understand real-world systems. A network can also be understood as an "instantiated" graph, with the graph being a purely abstract object with a set of properties, whereas the network defines a real setting where the graph object can be used. In a network, the mapping of the real system is done by assigning attributes to the edges and nodes of the graph. Graphs that represent real systems are called *complex networks*. Complex networks are characterized by being objects where order coexists with disorder, meaning that they are neither regular nor random. Complex networks share many particular features, some of which describe the network locally, i.e., by the connectivity between nodes or globally, i.e., by the global topology of the network. A network is said to have *small world* properties when the nodes are highly clustered, and when the average distance between the nodes is small. It is also said to have a *scale-free distribution* when the nodes' degree distribution follows a power law, meaning that, in the network, most of the nodes have a small degree and very few nodes have a high degree, also known as *hubs*. Another important feature of complex networks is that they tend to be large, in which case providing a global description of the network can become an unfeasible task. As a result, describing the network by its local properties is a more common approach. An illustrative example of a complex network is shown in Fig. 2, where the authors of [5] studied a population of 62 bottlenose dolphins living in the largest fjord in New Zealand to gain some understanding of dolphins' complex social behavior. The nodes in the network represent dolphins, and the edges, individuals pair-wise association occurring more often than expected by chance. Data were collected over a period of 7 years by direct observation of interactions between dolphins. The drawing in Fig. 2 was produced using a spring layout, where a visual inspection is enough to identify two sub-communities of dolphins among the sample population, where most of the links happen between dolphins of the same sex.

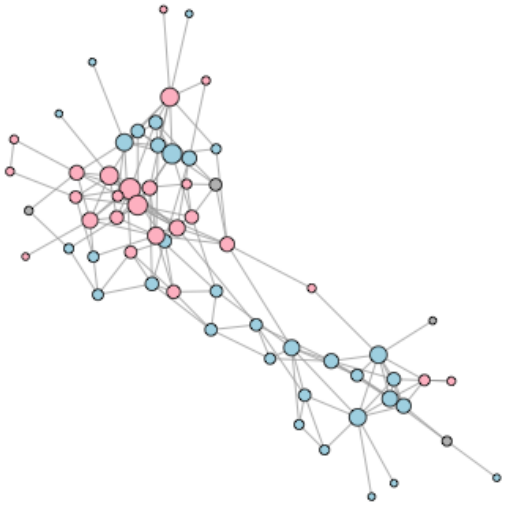


Figure 2: Bottlenose dolphins social network drawn using a spring-layout, nodes color indicates sex: the female (pink), male (blue) and unknown (grey) dolphins. The size of the nodes is proportional to each node degree.

Some classical properties of real complex networks are the following:

Definition 2.2.1 (**Degree distribution**). *Given a graph $G = (\mathcal{N}, \mathcal{L})$ with number of vertices $|\mathcal{N}| = N$, the degree distribution is defined as:*

$$p_k = \frac{N_k}{N}$$

Where N_k denotes the exact number of nodes with exactly k edges.

Definition 2.2.2 (**Geodesic path**). *A geodesic path -or shortest path- from node n_i to node n_j is the path with minimal number of edges. In the case of a weighted graph, it refers to the path with minimal sum of weights assigned to the edges connecting the nodes in the path.*

Definition 2.2.3 (**Geodesic distance**). *The geodesic distance between node n_i and node n_j refers to the length of the geodesic path. In the case of a weighted graph, it refers to the sum of weights assigned to the edges connecting the nodes in the geodesic path.*

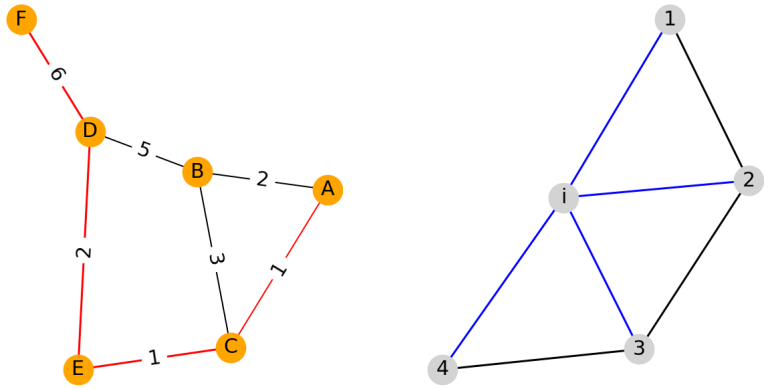
Definition 2.2.4 (**Clustering Coefficient**). *Given a graph $G = (\mathcal{N}, \mathcal{L})$ of size $N = |\mathcal{N}|$. The clustering coefficient, or local clustering coefficient c_i of a node $n_i \in \mathcal{N}$ is defined as:*

$$c_i = \frac{|\{l_{jk} : n_j,\ n_k \in M_i,\ l_{jk} \in \mathcal{L}\}|}{d_i(d_i - 1)/2}$$

Where $M_i = \{v_j : l_{ij} \in \mathcal{L} \vee l_{ji} \in \mathcal{L}\}$ refers to the set of neighbors, and $d_i = |M_i|$ to the number of neighbors of node n_i. The graph clustering coefficient C is the average of c_i over all nodes in the graph:

$$C = \frac{1}{N} \sum_{i=1}^{N} c_i$$

Fig 3 (a) illustrates the concept of geodesic distance and geodesic path in a weighted an undirected graph. The geodesic distance between vertices A and F is 10, and is simply obtained by adding the weights of the edges connecting the nodes of the geodesic path (F, D, E, C, A). On the other hand, the node clustering coefficient measures how the neighbors of a given node are connected to one another. A common example for understanding the concept of clustering coefficient is that of mutual friends: the clustering coefficient captures the social phenomena that friends of an individual are also likely to be friends. To further illustrate this notion, an unweighted graph of 5 nodes is presented in Fig. 3 (b); the blue lines represent the edges connecting the node i to its neighbors, and the black lines, the edges connecting the node i's neighbors. Since the maximum number of mutual connections the neighbors of the node i can have is 6, its clustering coefficient is $3/6 = 1/2$.



(a) (b)

Figure 3: Examples of: (a) geodesic path in a weighted graph, and (b) local clustering coefficient on an undirected graph.

Different types of simulated graphs are commonly used for studying and understanding properties of real-world systems in a controlled environment, mainly as they have the ability to mimic properties found in such real networks. This section will cover some aspects related to two important types of networks in network theory, namely *random graphs* and the so-called *small world networks*. Explicit connections with the main subject of this work will be made in Section 3.

2.2.1 Random graphs

A *random graph* is a type of graph whose structure is determined by a probability distribution, generated according to a graph model. Typically, this type of graph is constructed by randomly assigning, edges between pair of nodes to a set of isolated nodes. The theory of random graphs was first introduced in the late 50s by Hungarian mathematicians Paul Erdős and Alfréd Rényi in [6]. Their main study involved using probability theory and graph theory for identifying the appearance of certain properties that characterize such graphs. As of today, the *Erdős-Rényi graph models* remain the most popular way for generating random graphs. These graph models refer to two different procedures for generating random graphs, namely the *uniform random graphs* and the *binomial random graphs*. The main idea of these models is to consider not a single graph, but a collection of all possible graphs generated by fixing a particular parameter (e.g. the number of nodes, or the node degree), with both models sharing the possibility to be studied analytically. This thesis will cover the *binomial random graphs* only.

2.2.1.1 Erdős-Rényi model: Binomial random graph

Define $G^{ER}_{N,p}$ to be a random graph model generated by a collection of undirected and loop-less graphs with N labeled vertices, where each pair of nodes is connected independently, by an edge with a probability $p \in [0, 1]$. The binomial random graph is constructed by initially fixing a value of N and p. The pseudo-code of the algorithm for sampling this type of random graph can be found in the Algorithms section B.1. Some interesting properties of binomial random graphs are the following:

- Given a particular graph G of order K, the sampling probability P_G denotes the probability of finding the graph G in the collection $G_{N,p}^{ER}$ such that:

$$P_G(p) = p^K(1-p)^{M-K} \tag{1}$$

Where M refers to the maximum number of edges a graph of size N can have, i.e., $M = \binom{N}{2}$ and $0 \leq K \leq M$.

- The probability of finding graphs with K edges in $G_{N,p}^{ER}$ is the following:

$$P(K) = \binom{M}{K} p^K (1-p)^{M-K}$$

Note that the probability that a graph sampled from $G_{N,p}^{ER}$ has K links, follows a binomial distribution with parameters M and p, which is just the probability of sampling a graph with exactly K edges in M independent Bernoulli trials.

- A complex network of order K and size N can be approximated by a binomial random graph constructed with the value of p that maximizes the likelihood of including such network. That is, the value of p that maximizes the log-likelihood function of Eq. (1) with respect to p:

$$\mathcal{L}(G_p) = \log P_G(p) = K \log p + (M-K)\log(1-p)$$

The maximum likelihood estimator for p is:

$$\begin{aligned} \frac{\partial \mathcal{L}(G_p)}{\partial p} &= \frac{K}{p} - \frac{M-K}{1-p} \\ \text{set} \quad \frac{\partial \mathcal{L}(G_p)}{\partial p} &= 0 \\ \hat{p} = \frac{K}{M} &= \frac{2K}{N(N-1)} \end{aligned}$$

Meaning that for a model constructed with $\hat{p}$, all graphs will have on average K edges.

- The degree distribution of a random graph G in $G_{N,p}^{ER}$ follows a binomial distribution with parameters $N-1$ and p, such that:

$$\begin{aligned} p(k) = \frac{\bar{N}_k}{N} &= \frac{\sum_{i=1}^{N} P(\deg(n_i) = k)}{N} = P(\deg(n_i) = k) \\ &= \binom{N-1}{k} p^k (1-p)^{N-1-k} \end{aligned} \tag{2}$$

Where n_i refers to the i:th node in G and $k = 0, 1, 2, ..., N-1$. This equality holds since Definition 2.2.1 can be extended to the case of a collection of graphs, in which $\bar{N}_k$ denotes the average number of nodes with k edges in $G_{N,p}^{ER}$, and it can be written as the sum of the probability that each node in the graph has k links. Since $P(\deg(n_i) = k)$ is the same $\forall i = 1, 2, ..., N$ then $\bar{N}_k = NP(\deg(n_i) = k)$, which means that p_k has the same distribution as $P(\deg(n_i) = k)$. Note that for high values of N the binomial coefficient may

produce an overflow error, as $p^k(1-p)^{N-1-k}$ will be a small value, while the coefficient, a large one. In order to avoid this inconvenience, when $N \to \infty$ and $p \to 0$, Eq. (2) can be approximated by the *Poisson distribution*. Precise statements and proofs of this are omitted.

- As $N \to \infty$, the model $G_{N,p}^{ER}$ is almost surely connected if $p \geq \frac{\ln N}{N}$. To prove this, first let G be a graph in the model $G_{N,p}^{ER}$ and define the sequence of dependant random variable $\{Y_i\}_{1 \leq i \leq N}$:

$$Y_i = \begin{cases} 1 & \text{if } n_i \text{ is isolated} \\ 0 & \text{otherwise} \end{cases}$$

Where Y_i indicates if the i:th node is isolated. Let the random variable X_N be the number of isolated nodes in G, so that $X_N = Y_1 + Y_2 + ... + Y_N$. Then, the expected number of isolated nodes in G is:

$$E[X_N] = \sum_{i=1}^{N} E[Y_i] = \sum_{i=1}^{N} (1-p)^{N-1} = N(1-p)^{N-1}$$

Since the *critical probability* -or threshold- of having a connected component, is given by $\frac{\ln N}{N}$, consider $p = c\frac{\ln N}{N}$, where c is just a real-valued constant. Consider the ratio:

$$\frac{e^{(N-1)\ln(1-c\frac{\ln N}{N})}}{e^{N\ln(1-c\frac{\ln N}{N})}} = 1 - c\frac{\ln N}{N} \to 1 \text{ as } \; N \to \infty$$

Then,

$$\begin{aligned} \lim_{N\to\infty} E[X_N] &= \lim_{N\to\infty} N e^{N(-c\frac{\ln N}{N} - c^2\frac{\ln^2 N}{2N^2} + O(c^3\frac{\ln^3 N}{N^3}))} \\ &= \lim_{N\to\infty} N e^{-c\ln N} = \lim_{N\to\infty} N^{1-c} \end{aligned} \tag{3}$$

Note that when $c > 1$ the expected number of isolated nodes goes to zero while if $c < 1$ the expected number of isolated nodes goes to infinity.

Fig. 4 illustrates the critical probability for strong connectedness of simulated random graph models in comparison with the true critical probability. In general, the estimated critical probability has a good fit with respect to the theoretical true critical probability $c\ln N/N$, where in this case $c = 2$.

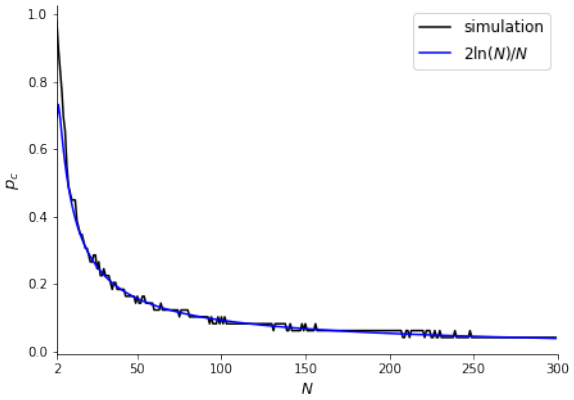


Figure 4: Critical probability p_c for which a random graph forms a connected component. The black line refers to the observed critical probability computed by simulating an ensemble of 100 random graphs per model. Each model was constructed for values of $0 \leq p \leq 1$ with a spacing of 0.02 and number of nodes $N \leq 300$. The blue line refers to the true critical probability $c\ln(N)/N$ where $c = 2$. The connectedness threshold for each model replicates is set to 0.95.

2.2.2 Small-World networks

Small world networks refers to a type of graphs characterized by having independent structural properties commonly observed in real-world systems, known as *small world effect* and high *clustering*. Small world networks were first introduced by Duncan J. Watts and Steven Strogatz in their paper [7] published in 1998. The authors presented the *Watts-Strogatz model* (also known as W-S model), based on the idea that by introducing randomness in a regular graph, small-world networks can be generated. As such, a small-world network can be defined as a junction between a regular graph with a high clustering, and a random graph with the small-world property. Ever since Watts and Strogatz laid the basis on small-world networks, many other methods that are able to generate networks with small-world properties have been proposed. In this section, the focus will remain on the classical model established by Watts and Strogatz.

2.2.2.1 Watts-Strogatz model

A *Watts-Strogatz graph* $G^{WS}_{N,k,p} = (\mathcal{N}, \mathcal{L})$ is an undirected graph that is generated by randomly rewiring all the edges of a regular ring lattice of size $|\mathcal{N}| = N$ with probability p. A network from the W-S model is generated in a two-step process. First, a regular network of N vertices and $Nk/2$ edges is constructed, where each node is connected by an undirected edge to the k nearest neighbors. That is, each node is connected to the $m = k/2$ right and left most near neighbors for a given value of k, where k is an even positive integer. In the second step, each edge in the graph is randomly rewired to a distinct node with probability p. In the case a node has degree $N - 1$, the rewiring of the edges is ignored. The rewiring process is done in a clockwise direction and is completed in m rounds, where each edge is only considered once and in an ordered manner, meaning that the edges are rewired from the nearest

to the farthest neighbor. In the first round for instance, only the edges connecting to the nodes' first right-most neighbors are rewired; in the second round, the second right-most neighbors. The same process is repeated until completing the m rounds. The pseudo-code of the algorithm for sampling a graph from the W-S model can be found in the Algorithms Section B.2.

Fig. 5 illustrates the rewiring process in sparse and connected graphs generated with the W-S model, where as the value of the rewiring probability p increases, the introduction of randomness in the graphs also increases. When $p = 0$ the regular ring lattice remains unchanged, while when $p = 1$ the graph reaches the limit of total randomness as all edges in the graph are rewired. When $p = 0.1$ given the low rewiring probability, the rewired regular graph remains almost unchanged, however, the few rewired edges are enough to set the transition from a regular network to a small-world network.

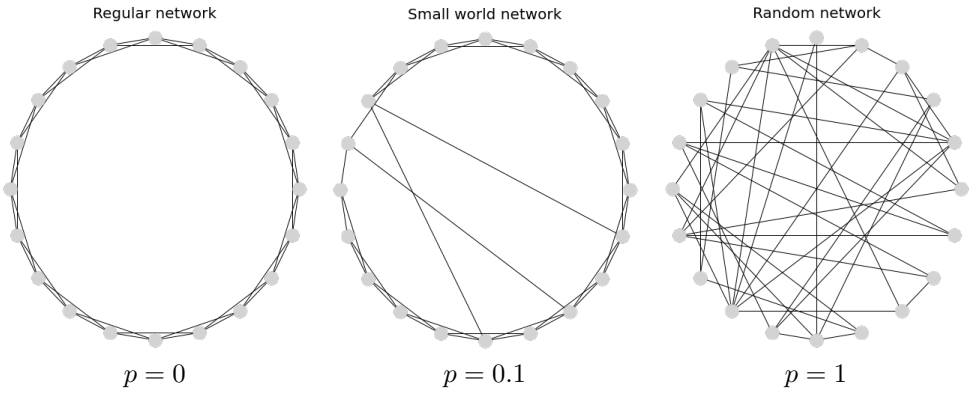


Figure 5: Graphs generated by the W-S model, constructed with fixed values of $N = 20$, $k = 4$ and increasing randomness.

The *small world effect* and *clustering* are two important qualitative properties that generate small world networks. *The small world effect* refers to the concept in which any node in a graph is connected to any other node in the same graph by a small path. This concept is captured by the *average path length* -or characteristic path length-, which for any network, represents the average of how far away any pair of nodes are in the network. The average path length is a global property of a graph and is obtained by taking the average of the geodesic distance (Definition 2.2.3) between all pairs of nodes in a graph. The average path is expressed as:

$$\frac{2}{N(N-1)} \sum_{i<j}^{N} \gamma(n_i, n_j)$$

where $\gamma(n_i, n_j)$ denotes the shortest geodesic distance between the nodes $n_i \in \mathcal{N}$ and $n_j \in \mathcal{N}$, for $i, j = 1, 2, ..., N$. The concept of *clustering* refers to the idea that nodes having common neighbors tend to be connected to each other rather than to distant nodes. This concept is captured by the clustering coefficient, a graph local property (Definition 2.2.4). Whenever the rewiring probability $p = 0$, the clustering coefficient and the average path length are both high. The first follows since in regular networks, nodes are only connected to neighboring nodes and not further away nodes,

while the second, since traversing the graph from one node to a more distant node will require a longer path distance. On the contrary, in the limit of total randomization, i.e., when $p = 1$, the random graph has a low clustering coefficient and low average path length, mainly as nodes in the graph are randomly connected to the rest of nodes in the graph. Intermediate values of p generate graphs of interest, as they are able to produce structural properties found in small-world networks – that is a high clustering coefficient and low average path length. The clustering coefficient and average shortest path length of a graph $G^{WS}_{N,k,p}$, are denoted as $C(p)$ and $L(p)$, respectively.

The transition from a regular ring lattice with high clustering coefficient and shortest average path length to a small-world network occurs with the introduction of *shortcuts* -or long-range edges- in the graph. For small values of p the clustering coefficient remains close to the one of a regular graph and only starts dropping for relatively high values of p, meaning that the rewired edges do not introduce important local changes on the graph's structure as only a small number of edges are actually rewired. On the contrary, the shortest path length drops rapidly starting at quite small values of p, making the intermediate region prior to the drop of the clustering coefficient a regime for small-world networks, this occurs mainly as it only takes a few number of rewired edges to connect distant neighborhoods.

Since Watts and Strogatz defined small-world networks qualitatively, in application the small-world network regime is found by comparing sampled graphs from the W-S model with a randomized version of the same graph. There is not a unique way to quantify the range of p for which the graph is a small-world network and certainly, this range is not defined by a sharp bound. One method for identifying the small world network regime was proposed in [8] where the authors introduced a small world metric called the *small world measurement* ω based on the idea of comparing the clustering coefficient C of a network to that of an equivalent ring lattice network C_{latt}, and comparing the path length L to that of the randomized version of the same network L_{rand}. The small world measurement is defined as the difference between the ratios:

$$\omega = \frac{L_{rand}}{L} - \frac{C}{C_{latt}}$$

Where $\omega \in [-1, 1]$. Within this context, a network is said to be a small world network, if ω is close to zero, i.e., whenever $C \approx C_{latt}$ and $L \approx L_{rand}$. Fig. 6 plots the clustering coefficient $C(p)$ and shortest average path length $L(p)$ for $0 < p < 1$. Both quantities are normalized by the clustering coefficient $C(0)$ and shortest average path length $L(0)$ of a regular ring lattice, respectively. For each rewiring probability, p and a fixed value of nodes N with degree k, the normalized clustering coefficient, and average path length were obtained as an average of an ensemble of 50 graph realizations. This figure demonstrates how randomly rewiring a very small percentage of edges in a regular ring lattice results in a rapid decrease of the average path length but marginal changes in the clustering coefficient.

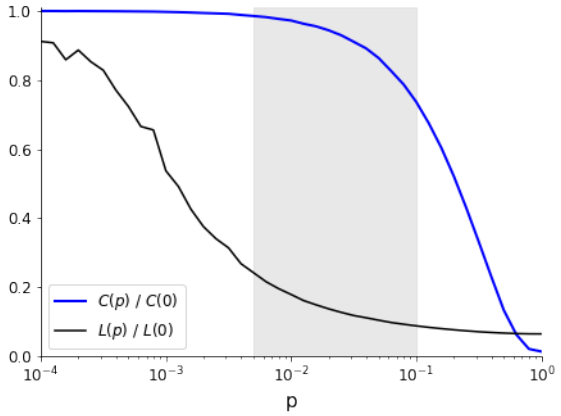


Figure 6: Watts and Strogatz's small-world model. A regular ring lattice with number of vertices $N = 1000$ and degree $k = 10$ was rewired at varying probabilities $0 \leq p \leq 1$. 50 realizations per rewiring probability were considered. $C(p)$ denotes the average clustering coefficient and $L(p)$ the average path length for a certain value of p. The x-axis represents p in log-scale. The shaded area denotes the small-world network regime.

Another important property of the W-S model is that the mean degree is always constant and the nodes in the graph always have at least $k/2$ edges. As a result, graphs sampled from the W-S model do not have isolated vertices and they usually form a connected component, meaning that in the limit of total randomness, the W-S model does not produce graphs with local similarities to a random graph. As introduced in [9], the degree of a node n_i in a W-S model graph can be decomposed into two parts: (i) the fixed number of untouched edges $m = k/2$, originally connected to the rightmost neighbors, and (ii) any additional edges connected to the node, denoted as u_i. These additional edges can also be divided in two quantities: (i) the number of connections left from the remaining m edges, denoted as $u_i^{(1)}$, where $u_i^{(1)} \leq m$, and (ii) all the additional edges in-going node n_i, denoted as $u_i^{(2)}$. That is: $\deg(n_i) = m + u_i$, where $u_i = u_i^{(1)} + u_i^{(2)}$.

The probability that the node n_i maintains $u_i^{(1)}$ connections from the remaining m original leftmost neighbor edges, follows a binomial distribution, expressed as:

$$P(u_i^{(1)}) = \binom{m}{u_i^{(1)}} (1-p)^{u_i^{(1)}} p^{m-u_i^{(1)}} \quad \text{where} \quad u_i^{(1)} \leq m \tag{4}$$

For graphs of high order, the probability of node n_i having $u_i^{(2)}$ additional coming edges follows a Poisson distribution expressed as:

$$P(u_i^{(2)}) = \frac{(pm)^{u_i^{(2)}}}{(u_i^{(2)})!} e^{-pm} \tag{5}$$

Hence, from Eq. (4) and Eq. (5), the degree distribution of the W-S model can be written as:

$$p(k) = \sum_{u^{(1)}=0}^{\min\{k-m,\ m\}} \binom{m}{u^{(1)}} (1-p)^{u^{(1)}} p^{m-u^{(1)}} \frac{(pm)^{k-m-u^{(1)}}}{(k-m-u^{(1)})!} e^{-pm} \tag{6}$$

Fig. 7 shows the degree distribution for $k_0 = 8$ and various values of p. Note that when $p = 0$ the degree distribution is a Kronecker delta function, and that for all values of p the degree distribution is centered at k_0. Also, as the value of p increases, the degree distribution of the graph becomes wider. One of the main limitations of the W-S model is that despite it manages to generate graphs with small-world properties, the degree distribution of these graphs tend to be homogeneous. This aspect is not commonly found in real-world systems, whose degree distribution commonly follows a power-law.

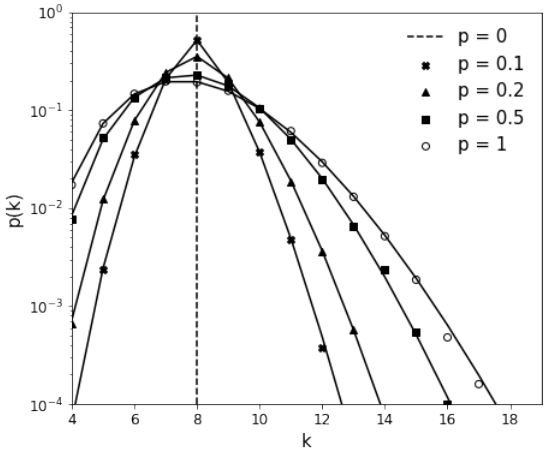


Figure 7: Probability distribution of the node degree k for W-S model graphs: the different markers correspond to the average observed degree distribution of 50 graphs, generated with fixed values of $N = 1000$ and $k_0 = 8$ and various values of p. The solid black line connecting the markers refer to the theoretical probability distribution from Eq. (6) computed with the same p of the overlying markers.

2.3 From data to Graph

Simulated graphs commonly used for studying real-world systems come in handy for benchmarking a distance measure such as the spectral graph distance. The reason behind this is that the structural properties of graphs generated from well-studied graph models provide prior knowledge of the structure of the sampled graph. Such is the case of the W-S model in [8]. Hence, having prior knowledge of the graph's structural properties serves as a mean to validate whether the spectral graph distance is able to uncover these structural differences or not. As an alternative, it is also possible to construct graphs directly from data without using a predefined graph model. This section will cover some of these approaches.

A *data set* is a collection of observations, where each observation (or data point) carries quantitative and (or) qualitative information, often referred to as *variables*. *Quantitative variables* refer to numerical variables that represent a measurement, while *qualitative variables* to categorical values that differ in quality and not in quantity. When mapping a dataset to a graph, each node in the graph represents a specific observation in the dataset, while the edges and weights in the graph are determined

by a *similarity measure* that measures the pairwise similarity between observations. A link between two nodes in a graph is assigned if the similarity measure between their corresponding observations in the dataset is positive, or if it meets a certain condition, such as exceeding a threshold.

The use of a similarity measure to assign edges to a graph is the reason why these graphs are often referred to as *similarity graphs*. There are different popular approaches, as well as similarity functions to construct graphs, all of them sharing the common purpose of capturing local neighborhood relationships between nodes. In the sense that when constructing a graph the main idea is not to provide information about the location of the nodes in the graph, if not, to understand how nodes are related between each other. Such construction will allow to identify strong connections between nodes. In order to introduce some of these approaches, certain notations need to be established. Let the matrix $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}) \in \mathbb{R}^{d \times n}$ represent a dataset of n observations, each having d variables, where each column corresponds to an observation: $\mathbf{x_i} = (x_{i1}, x_{i2}, ..., x_{id})$ for $i = 1, 2, ..., n$. Define an undirected and weighted graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ to be the graph representation of $\mathbf{X}$, where each observation $x_i \in \mathbf{X}$ corresponds to a node $n_i \in \mathcal{N}$. By convention, the set of links $\mathcal{L} \in G$ does not include self-loops – or self-connectivity links – and whenever G is a complete and weighted graph, the self-similarity measure is set to zero. There are many reasons why self-loops are avoided when constructing a similarity graph, the more intuitive follows from the main purpose of the similarity function, which is to measure pairwise similarities between observations and not similarities with themselves. Some available options to construct the graph G are introduced as follows.

2.3.1 Fully connected graph weighted by the Gaussian similarity

A *fully connected graph* is a complete graph with weighted edges connecting all pair of nodes. The weights of the graph are non-negative values defined by a similarity measure that models local neighborhoods. Among the many available similarity measures to construct graphs, the *Gaussian similarity function* is a popular and intuitive similarity measure able to capture the underlying structure of the data, as it has been previously shown in [1, 2, 10, 11]. The Gaussian similarity function is defined as:

$$\omega(x_i, x_j) = \begin{cases} \exp\left(- \frac{||x_i - x_j||^2}{2\sigma^2} \right) & x_i \neq x_j \\ 0 & x_i = x_j \end{cases}$$

Where $x_i, x_j \in \mathbf{X}$. Note that when x_i and x_j are a distinct pair of nodes, the similarity function is a non-linear function of the Euclidean distance, where the *width parameter* - or scale parameter - σ is instrumental for capturing the local structure of the data, as it somehow defines the extent of the observations' neighborhood. If the separation of two observations $x_i, x_j \in \mathbf{X}$ is smaller than σ, then their connection will have a significant weight $\omega(x_i, x_j)$, as both points will reside in the same neighborhood. However, as the distance between observations grows larger than σ, the weights rapidly decrease to a value close to zero.

Different choices of σ will describe the underlying structure of the data in different ways. A very large value of σ will entail strong links between nodes as the observations will lay in the same neighborhood. On the contrary, as the value of σ gets close to

zero, the connectivity links between nodes become very weak, possibly yielding to a fully disconnected graph as the neighborhood extend neglects each observation nearest neighbors. This means that whenever the value of σ is too large, or too small, the similarity measure fails to capture the local geometric structure of the data. Ideally, the parameter σ should be small enough to connect only neighboring points, in the sense that if by one σ a point is not able to reach the nearest neighbor, it is a sign that σ is too small.

The choice of σ will also depend on prior knowledge of the geometrical structure and density of the data. A dataset is said to be *inhomogeneous* when the density of the data changes from one region to another, and *homogeneous* when the density is constant in all regions. If the dataset has more than one well-defined component (or cluster) then a dataset can be characterized by being *imbalanced* or *balanced*. An imbalanced dataset refers to a classification problem in which the data are not uniformly distributed between components, while a balanced dataset, the opposite. In the case of a homogeneous dataset, a reasonable way of choosing σ is by computing all observations' k nearest neighbors' distances and selecting a value of σ that is at least the median (in order to avoid any outlier effect) of those distances multiplied by a constant factor. In the case of an imbalanced and inhomogeneous dataset, the more suitable option is to adapt σ according to the density in each region, i.e., using more than one value of σ. Finding a "good" choice of the scale parameter σ is one of the main challenges of using the Gaussian function, and even if there are other useful options, besides the ones already presented in this thesis, these methods tend to be computationally expensive. An example of such method is by choosing the value of σ that gives the largest spectral gap, i.e., the smallest nonzero eigenvalue. As it has been shown in [2], a large spectral gap can be associated to a pronounced separation of strongly connected components in the data. However, this method will not only require having prior information about the structure of the data, if not it will also require solving the eigenvalue problem of the graph's Laplacian matrix many times until finding a suitable choice.

Fig. 8 illustrates how the scale parameter $\sigma = 0.36$ delimits the neighborhood of two observations located in different regions of the two-dimensional dataset, where the center of the circles represents the observations, while the radius σ, the neighborhood extend. Each marked observation has a strong (weak) connective weight with the rest of the neighbors lying inside (outside) the dotted circle delimited by σ. The dataset shown in Fig. 8 is inhomogeneous and imbalanced in the number of observations; it is composed of two small and dense clusters and a noisy, larger and less dense cluster. Note that for the same σ the coverage is different in the two regions of the data. The similarity measure is able to capture the underlying structure of the larger cluster, but not for the small ones, as it creates strong connections between the observations of the two small clusters. In this case, using two different values of σ is a way to better capture the structure of the data.

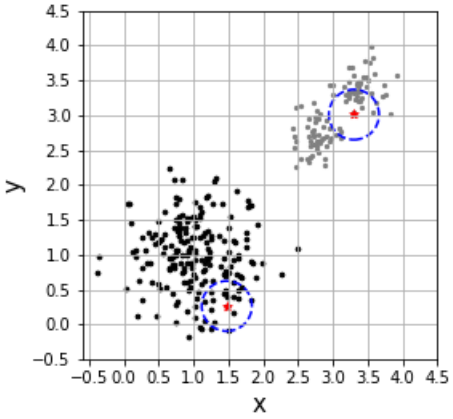


Figure 8: Example of the neighborhood coverage for two observations corresponding to a two-dimensional imbalanced and inhomogeneous spherical data set.

2.3.2 k-nearest neighbor graphs

A different approach to construct a graph is by using the *k-nearest neighbor method*. First, let G' be an unweighted and directed graph, whose set of nodes is represented by $\mathcal{N}' \in G'$. The k-nearest neighbor method tries to connect a *successor* node $n_i \in \mathcal{N}'$ to a distinct *predecessor* node $n_j \in \mathcal{N}'$ by an edge, if and only if the distance between n_i and n_j is among the k:th smallest distances between n_i and the rest of nodes in $\mathcal{N}'$. Generally, the edges in G' can only be traversed in a single direction, meaning that if n_j is among the k-nearest neighbors of n_i, the converse is not necessarily true. As a result of this non-symmetric nearest neighbor relation, two different weighted undirected graphs can be derived from G':

k-NNG: The *k-nearest neighbor graph* is a weighted and undirected graph generated by assigning a two-ways direction edge between each node in the graph and its corresponding k-nearest neighbors. In the case that the directed edge already had a two-way relationship, and given the symmetric nature of the edges in the graph, all duplicated edges are ignored.

Mutual k-NNG: In the *mutual k-nearest neighbor graph* two nodes n_i and n_j in $\mathcal{N}$ are connected if and only if the node n_j is among the k-nearest neighbors of n_i, and vice versa. The mutual k-NNG is the "mutually inclusive" version of the k-NNG.

After constructing either of the k-NNG's, weights are assigned to the edges of the graphs. This can be done by a similarity function such as the Gaussian similarity, or by simply weighing all edges with a value of 1. Regarding the parameter k, its role is to associate nodes in the graph that are considered to be "close" or "similar", without depending on a distance scale. The choice of the parameter k will determine the graph representation of the data. If the value of k is too small, so will the node's degree, meaning that the nodes in the graph will be connected to the few closest ones, resulting in an almost, if not totally disconnected graph. On the other hand, as the value of k increases, connections between distant nodes will be allowed, until coinciding with

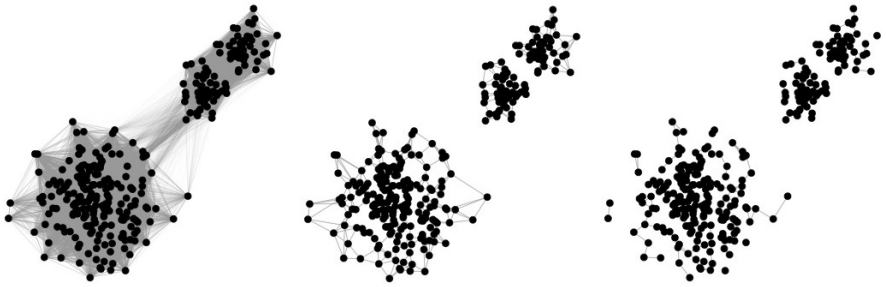
a complete graph – in which case if the edges are weighted by a similarity measure, the resulting graph could approach a fully connected graph. In the case where the edges have a unit weight instead, the parameter k will fail to capture the underlying structure of the data. A suitable value of k should be small enough that it allows the graph to connect nearby nodes without generating disconnected components.

The main structural differences between the two graphs rely on the nodes' degree distribution and the connectedness of the graph. In the case of the mutual k-NNG, the nodes' degree is at most k, while in the case of the k-NNG, it is at least k, meaning that the size of the k-NNG is generally larger than in the case of the mutual k-NNG. Another important difference between the two similarity graphs is that the mutual k-NNG does not produce nodes with high degree, and so the graph tends to have more disconnected components, some of them composed of very few observations of "noisy" or "isolated" observations. This is not usually the case with the k-NNG where nodes with a larger degree than the average node degree - also referred to as *hubs* - are more likely to appear.

2.3.3 Choosing a similarity graph

The matter of choosing a similarity graph strongly depends on the data, the purpose of the graph, and the users' preference, as on many occasions different similarity graphs can serve the same purpose. Such is the case with graph-based clustering methods as spectral clustering [2]. Note that both approaches, the fully connected graph, and the neighborhood graphs, use a distance metric. The choice of this metric should also be done according to the data representation. For instance, in the case of high-dimensional data, using the *Euclidean distance* is not the best option given that the pairwise distance between observations in the Cartesian coordinates becomes almost uniform, failing to distinguish close and distant observations. Choosing a suitable value of the parameters σ and k is not an easy task, and can lead to different representations of the data, where some of them manage to capture its local geometric structure, and others do not. Also, unlike the fully connected graph, the k-nearest neighbor graph with weights defined by a similarity measure might require defining an additional parameter σ, which could turn into a more complicated problem. In such cases, an alternative could be to assign weights using as similarity measure the reciprocal of the Euclidean distance.

Fig. 9 illustrates three different similarity graphs generated from the spherical dataset presented in Fig. 8, where (a) shows a fully connected graph with weights assigned by the Gaussian similarity function ($\sigma = 0.36$), (b) a k-NNG ($k = 4$), and (c) a mutual k-NNG ($k = 4$). In (a), given the choice of σ, the similarity measure could not distinguish the local properties of the more dense region in the data, as it was only able to generate two main strongly connected components. The mutual k-NNG in (c) formed many small disconnected components of "noisy" or more "distant" observations in all regions on the data, caused by the small choice of the parameter k. The k-NNG in (b) was the only similarity graph that succeeded in identifying the three components in the data, which indicates that the method managed to capture the underlying geometrical structure of the data in all regions.



(a) Fully Connected Graph (b) k-NNG (c) Mutual k-NNG

Figure 9: Example of three similarity graphs built from an imbalanced and inhomogeneous spherical dataset. The graph in (a) shows a fully connected graph with edges weighted by the Gaussian similarity function with width of $\sigma = 0.6$. The graphs in (b) and (c) show a k-NNG and a mutual k-NNG respectively, built with $k = 4$ and edges weighted with a value of 1.

2.4 The Graph Laplacian and other Graph matrices

The spectral graph distance is a method rooted in spectral-graph theory, thereby, with a close connection to the graph Laplacian. In this section, some basic matrices derived from graphs are introduced, followed by presenting the different ways of defining the graph Laplacian, namely, the *unnormalized* graph Laplacian and the *normalized* graph Laplacians, as well as discussing some important properties related to the spectrum of a graph in connection with each Laplacian form. Although the spectral graph distance is defined in terms of the unnormalized Laplacian only, the normalized Laplacian form is also presented in this section. The motivation is to provide the reader with a complete review of the different Laplacian forms, as a way to establish a comparison between their properties, and outline artifacts likely to arise in the unnormalized form and not in the normalized form. The material presented in this section is required for the main subject of this thesis.

Two important matrices that capture structural information of a graph are the *adjacency* and *degree* matrices. For an undirected unweighted graph of size N, the adjacency matrix $\mathbf{W}$ is a symmetric $N \times N$ matrix whose element $\mathbf{W}_{ij}$ equal one if there is an edge connecting vertices n_i and n_j, and zero otherwise. In the case where the graph is weighted, the elements of the adjacency matrix are the edges' weights, commonly assigned by a similarity measure. In both cases, the diagonal elements of the adjacency matrix is zero-valued, due to the absence of self-connectivity links. Moreover, the degree matrix $\mathbf{D}$ is a $N \times N$ diagonal matrix obtained from the adjacency matrix. Its diagonal elements are given by the nodes' degree, i.e., the row or column sum of the adjacency matrix, where the degree of node i, defined in Section 2.1.4 is given by the sum of the i:th row or i:th column of the matrix $\mathbf{W}$. The diagonal matrix provides information on how connected each node is to the rest of the nodes in the graph. The two matrices are formally defined as:

Definition 2.4.1 (Adjacency and Degree Matrix). *Given a graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ of size $|\mathcal{N}| = N$, nodes $n_i, n_j \in \mathcal{N}$, and node degree $deg(n_i)$:*

The adjacency matrix $\mathbf{W}$ of G is a $N \times N$ matrix with elements:

$$\mathbf{W}_{ij} = \begin{cases} \omega(n_i, n_j) & \text{if } (n_i, n_j) \subset \mathcal{L} \\ 0 & \text{otherwise} \end{cases}$$

The degree matrix $\mathbf{D}$ of G is a $N \times N$ diagonal matrix with diagonal elements:

$$d_i = deg(n_i)$$

The *graph Laplacian* or Laplacian matrix is the matrix representation of a graph. This matrix is one of the most studied matrices in graph theory, as well as one of the fundamental concepts of spectral graph theory, extensively discussed by Chung in [12]. Its use in different contexts and applications mainly originates from the ability to relate the structure of a graph with its Laplacian matrix spectrum, as done in [1, 2].

Any positive semi-definite $m \times m$ matrix $\mathbf{M}$ is characterized by having m non-negative real-valued eigenvalues associated with m real eigenvectors. The positive semi-definite property of a matrix provides that all its eigenvalues are non-negative, while the symmetric property allows to find orthonormal eigenvectors. The *algebraic multiplicity* of the eigenvalue $\lambda_i, i = 1, 2, ..., m$ refer to the number of times the eigenvalue appears as a root of the characteristic polynomial i.e., $\det(\lambda_i \mathbf{I} - \mathbf{M})$ where $\mathbf{I}$ is the identity matrix. The algebraic multiplicity of eigenvalues is an important property, commonly used in the graph partitioning method, namely, *spectral clustering* [2], as the algebraic multiplicity of the smallest eigenvalue of the Laplacian matrix is known for determining the number of connected components in the graph. The three types of graph Laplacian are now introduced:

2.4.1 The unnormalized Graph Laplacian

Given a graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ of size N, whose adjacency matrix and degree matrix (Definition 2.4.1) are denoted by $\mathbf{W}$ and $\mathbf{D}$, respectively. The *unnormalized graph Laplacian* of G is:

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{7}$$

With elements:

$$\mathbf{L}_{ij} = \begin{cases} d_i & i = j \\ -\omega(n_i, n_j) & i \neq j \end{cases}$$

$\mathbf{L}$ is a $N \times N$ *symmetric* and *positive semi-definite* matrix with precisely N non-negative real eigenvalues associated with N orthonormal eigenvectors. From Eq. (7), it is clear that the symmetry of $\mathbf{L}$ follows directly from the symmetry of $\mathbf{D}$ and $\mathbf{W}$, such that $\mathbf{L}^T = \mathbf{L}$. Moreover, the positive semi-definite property of $\mathbf{L}$ follows from the fact that for all vectors $f \in \mathbb{R}^N$ the quadratic form $f^T \mathbf{L} f$ is non-negative. The proof is shown below:

$$
\begin{aligned}
f^T\mathbf{L}f &= f^T(\mathbf{D}-\mathbf{W})f \\
&= f_1^2\sum_{j=1}^{n}\omega_{1j} + \ldots + f_n^2\sum_{j=1}^{n}\omega_{nj} - \Big(f_1\sum_{j=1}^{n}f_j\omega_{j1} + \ldots + f_n\sum_{j=1}^{n}f_j\omega_{jn}\Big) \\
&= \sum_{i=1}^{n}f_i^2\sum_{j=1}^{n}\omega_{ij} - \sum_{i=1}^{n}f_i\sum_{j=1}^{n}f_j\omega_{ji} = \frac{1}{2}\sum_{i,j=1}^{n}\omega_{ij}(f_i^2+f_j^2-2f_if_j) \\
&= \frac{1}{2}\sum_{i,j=1}^{n}\omega_{ij}(f_i-f_j)^2 \geq 0
\end{aligned}
\tag{8}
$$

Eq. (8) shows that $\mathbf{L}$ is a positive semi-definite matrix. Hence, it follows that $\mathbf{L}$ has N non-negative real-valued eigenvalues. From Eq. (7) it is clear that by construction the row or column sum of the elements of $\mathbf{L}$ is always zero. Therefore, $\mathbf{L}$ has at least one zero eigenvalue associated with the one-constant eigenvector $\mathbb{1} = (1, 1, 1, ..., 1, 1)$. The aforementioned follows since $\mathbb{1}\mathbf{L}\mathbb{1} = 0$.

2.4.2 The normalized Graph Laplacians

There are two main forms of defining the normalized graph Laplacians; that is, the *symmetric* and *random walk* graph Laplacian. Although both matrices are closely related, they offer a different perspective on the same problem. The matrices are defined as follows:

2.4.2.1 Symmetric Graph Laplacian

Given a graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ of size N, the *symmetric graph Laplacian* of such graph is:

$$
\begin{aligned}
\mathbf{L}_{sym} &= \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}(\mathbf{D}-\mathbf{W})\mathbf{D}^{-1/2} \\
&= \mathbf{D}^{-1/2}\mathbf{D}\mathbf{D}^{-1/2} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}
\end{aligned}
\tag{9}
$$

With elements:

$$
\mathbf{L}_{ij}^{sym} = \begin{cases} 1 & i = j \\ -\frac{\omega(n_i,n_j)}{\sqrt{d_i d_j}}, & i \neq j \text{ and } (n_i, n_j) \subset \mathcal{L} \end{cases}
$$

Where $\mathbf{I}$ denotes the identity matrix, $\mathbf{L}$ the unnormalized graph Laplacian, $\mathbf{W}$ the adjacency matrix with elements $\mathbf{W}_{ij} = \omega_{ij} = \omega(n_i, n_j);\ i, j = 1, ..., N$, and $\mathbf{D}$ the diagonal matrix with diagonal elements $d_1, .., d_N$, all of them derived from the graph G. $\mathbf{L}_{sym}$ is a $N \times N$ *symmetric, positive semi-definite matrix* with N non-negative real-valued eigenvalues, associated with N orthonormal eigenvectors. The positive semi-definite property of $\mathbf{L}_{sym}$, follows since the quadratic form $f'\mathbf{L}_{sym}f$ for all $f \in \mathbb{R}^N$ is non-negative. The proof is shown below:

$$
\begin{aligned}
f'\mathbf{L}_{sym}f &= f'\mathbf{I}f - f'\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}f \\
&= \sum_{i=1}^{n} f_i^2 - \sum_{i=1}^{n} \frac{f_i}{\sqrt{d_i}} \left(\sum_{j=1}^{n} \frac{f_j}{\sqrt{d_j}} \omega_{ji} \right) = \sum_{i=1}^{n} f_i^2 \frac{\sum_{j=1}^{n} \omega_{ij}}{d_i} - \sum_{i,j=1}^{n} f_i f_j \left(\frac{\omega_{ij}}{\sqrt{d_i}\sqrt{d_j}} \right) \\
&= \sum_{i,j=1}^{n} f_i^2 \frac{\omega_{ij}}{d_i} - f_i f_j \left(\frac{\omega_{ij}}{\sqrt{d_i}\sqrt{d_j}} \right) = \frac{1}{2} \sum_{i,j=1}^{n} \omega_{ij} \left(\frac{f_i^2}{d_i} + \frac{f_j^2}{d_j} - 2\frac{f_i f_j}{\sqrt{d_i}\sqrt{d_j}} \right) \\
&= \frac{1}{2} \sum_{i,j=1}^{n} \omega_{ij} \Big(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \Big)^2 \geq 0
\end{aligned}
$$

2.4.2.2 Random Walk Graph Laplacian

A discrete time *random walk* on a graph, is a finite Markov chain where each node in the graph represents a state in the Markov chain state space. Such Markov chain is described by an $N \times N$ *transition probability matrix*, or transition matrix $\mathbf{P}$, whose elements $\mathbf{P}_{ij} = p(j|i)$ denote the conditional probability of moving to state j from state i, for $i, j = 1, ..., N$.

Given a graph $G = (\mathcal{N}, \mathcal{L}, \omega)$ of size N, the *random walk graph Laplacian* is defined as:

$$
\begin{aligned}
\mathbf{L}_{rw} &= \mathbf{D}^{-1}\mathbf{L} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{W}) \\
&= \mathbf{I} - \mathbf{D}^{-1}\mathbf{W} = \mathbf{I} - \mathbf{P}
\end{aligned}
\tag{10}
$$

With elements:

$$
\mathbf{L}_{ij}^{rw} = \begin{cases} 1 & i = j \\ -\frac{\omega(n_i, n_j)}{d_i}, & i \neq j \text{ and } (n_i, n_j) \subset \mathcal{L} \end{cases}
$$

From Eq. (10), it is clear that the introduction of the transition matrix $\mathbf{P}$ in the definition of the random walk Laplacian $\mathbf{L}_{rw}$, provides a meaningful interpretation. Note that the rows in the transition matrix $\mathbf{P}$ describe a probability distribution where each element in the matrix is determined by the conditional probability $p(j|i) = \frac{\omega(n_i, n_j)}{d_i}$. This probability is proportional to the connectivity weight between node n_i and node n_j, meaning that the stronger (weaker) the connection between nodes, the higher (lower) the transition probability. Another connection between the two matrices is that the spectrum of $\mathbf{L}_{rw}$ and $\mathbf{P}$ will differ only by a constant factor of 1, meaning that the pair (λ, u) is an eigenvalue and eigenvector of $\mathbf{L}_{rw}$ if and only if the pair $(1 - \lambda, u)$ is an eigenvalue and eigenvector of $\mathbf{P}$.

Moreover, there exists a close relationship between the spectrum of $\mathbf{L}_{sym}$ and $\mathbf{L}_{rw}$, in the sense that λ is an eigenvalue of $\mathbf{L}_{rw}$ with eigenvector v if and only if λ is an eigenvalue of $\mathbf{L}_{sym}$ with eigenvector $w = \mathbf{D}^{1/2}v$. Eq. (11) shows this relationship.

$$
\begin{aligned}
\mathbf{L}_{sym}\ w &= \lambda w \\
(\mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})\mathbf{D}^{1/2}v &= \lambda \mathbf{D}^{1/2}v \\
(\mathbf{I} - \lambda)\mathbf{D}^{1/2}v &= \mathbf{D}^{-1/2}\mathbf{W}v \\
(\mathbf{I} - \mathbf{D}^{-1}\mathbf{W})v &= \lambda v \\
\mathbf{L}_{rw}\ v &= \lambda v
\end{aligned}
\tag{11}
$$

In general, L_{rw} is not a *symmetric* matrix, since $\mathbf{P}$ needs not be symmetric, given that $p(j|i) \neq p(i|j)$. From Eq. (11) it follows that since L_{sym} and L_{rw} share the same spectrum, L_{rw} has N non-negative eigenvalues associated with N non necessarily orthogonal eigenvectors, as L_{rw} is not symmetric. As a final remark, Eq. (12) verifies that solving the generalized eigenvalue problem of $\mathbf{L}$ provides the eigenvalues and associated eigenvectors of $\mathbf{L}_{rw}$.

$$
\begin{aligned}
\mathbf{L}_{rw}\ v &= \lambda v \\
(\mathbf{D}^{-1}\mathbf{L})v &= \lambda v \\
\mathbf{L}v &= \lambda \mathbf{D}v
\end{aligned}
\tag{12}
$$

2.4.3 Choosing the Graph Laplacian

The choice of the graph Laplacian is a fundamental question in related methods to spectral graph theory. The authors of [2] offer a discussion of different aspects to consider before choosing a Laplacian form. One of such considerations is that whenever the degree distribution of the graphs is regular the Laplacian choice is not very important, as all Laplacians will be very similar. If the degree distribution is broadly distributed, then the authors of [2] advocate for using the normalized Laplacian over the unnormalized Laplacian. A reason for this choice is that for high order eigenvalues of the unnormalized Laplacian, it can happen that some of the associated eigenvectors are almost constant vectors, such that, except for a single component, most elements of the eigenvectors take small values close to zero. While this issue only affects the unnormalized Laplacian, it can lead to undesirable effects, as detailed in [2] and Section 3. Besides, unlike the unnormalized Laplacian, the spectrum of the normalized Laplacian is not influenced by the graph size, which is more convenient when comparing graphs of different sizes. Finally, when it comes to choosing between the symmetric and random walk Laplacian, the symmetric Laplacian might be a better option, essentially by its symmetric nature and as it presents computational advantages when solving the eigenvalue problem.

2.5 Similarity between graphs: Matrix Distances and Spectral Distances

How to define similarities between graphs? Although it is a common problem in the field of graph theory, there is no unique definitive answer to the question, as the notion of similarity itself can be captured in different ways. A common approach for expressing differences between graphs is by means of a distance function. Formally, a distance function is defined as such:

Definition 2.5.1 (**Distance function**). *A distance function, or metric on a set $\mathcal{M}$ is a function $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\geq 0}$. For all $x, y, z \in \mathcal{M}$ the distance function satisfies the following axioms:*

- $d(x, y) = 0 \iff x = y$ *the distance from a point to itself is always zero*
- $d(x, y) = d(y, x)$ *is symmetric*
- $d(x, z) \leq d(x, y) + d(y, z)$ *satisfies the triangle inequality*

In graph theory, there are many definitions of graph distances [13], and while they are often referred to as distance functions, many of them do not satisfy all axioms in Definition 2.5.1. In [14], the authors propose to classify graph distances in two main categories: the *matrix distances* and the *spectral distances*. The matrix distances are a set of distances based on a direct comparison of graph matrices. These types of distances require node-to-node correspondence, and their main objective is to provide a global generalization of individual nodes' properties, as for instance by comparing edges perturbation between a pair of graphs. Due to this requirement, matrix distances can only detect changes at a local level, making them more suitable for comparing graphs differing in volume rather significantly. To address the problem of graph similarity, a different approach is to use spectral methods rooted in spectral graph theory. Spectral distances are a set of graph distances based on comparing the spectrum of graph matrices (see the ones introduced in Section 2.4). Recent studies presented in [15, 16] have shown that the spectrum of graph matrices manage to capture global structure of the graphs such as clustering structures, and thus provide new insights on the structure of complex networks. Nevertheless, as discussed in [17], there are reasons to advise against the choice of the spectrum as a graph representation, mainly as it can be too coarse. As a result, different graphs may end up with the same spectrum, also said to be *co-spectral* (with the exact same eigenvalues), raising doubts about the reliability of the method. In that respect, one of the novelties of the spectral graph distance is that, unlike many other spectral distances, it considers eigenvectors, thus mitigating the problem of co-spectrality.

This section will first introduce the *Hamming distance*, a matrix distance based on comparing the adjacency matrices of a pair of graphs, and then present the theory behind the main subject of this thesis: the *spectral graph distance* for complex networks. The motivation for introducing the Hamming distance is to then establish, in the Results 3 section, a comparison of the ability of both the spectral distance and the matrix distance to capture structural differences between graphs.

2.5.1 Hamming Distance

The *Hamming distance* between two graphs G and G' of equal size N is defined as:

$$H(G, G') = \frac{1}{N(N-1)} \sum_{i,j} |\mathbf{A}_{ij} - \mathbf{A}'_{ij}| \tag{13}$$

Where $\mathbf{A}$ and $\mathbf{A}'$ denotes the $N \times N$ adjacency matrix of each graph. The Hamming distance displays how distant two graphs are based on the differences between their edges, meaning that it only captures local information of edges perturbation. Although the Hamming distance can be adapted for comparing graphs of different sizes (by adding isolated nodes to the smaller graph), it is best suited for comparing graphs of the same size.

2.5.2 Spectral Graph Distance (SGD)

Shimada et *al.* first introduced the spectral graph distance in their 2016 publication [3]. Their main goal was to establish a graph distance that would be able to quantify dissimilarities between networks by comparing their Laplacian matrices, the Laplacian being chosen for the ability of its eigenvectors to describe aspects of the global structure of the network. Before diving into the complete definition of the SGD, some key concepts will be first introduced. Thus this section will start by commenting on the distribution of the elements of an eigenvector, and continue with the introduction of the *empirical cumulative distribution function*, and of the *Kruglov distance*, a distribution-distribution distance that is considered as it was used to define the SGD. Certain notations will also be established.

Define the network $G^{(i)}$ of size $N^{(i)}$ and order $K^{(i)}$, whose unnormalized Laplacian matrix eigendecomposition is $\mathbf{L}^{(i)} = \mathbf{V}^{(i)}\mathbf{\Lambda}^{(i)}(\mathbf{V}^{(i)})^T$, where $\mathbf{\Lambda}^{(i)}$ is a $N^{(i)} \times N^{(i)}$ diagonal matrix with diagonal elements given by the ordered eigenvalues $0 = \lambda_1^{(i)} \leq \lambda_2^{(i)} \leq ... \leq \lambda_{N^{(i)}}^{(i)}$. The matrix $\mathbf{V}^{(i)}$ is a $N^{(i)} \times N^{(i)}$ matrix whose columns are the eigenvectors associated with the eigenvalues in $\mathbf{\Lambda}^{(i)}$. That is, $\boldsymbol{v}_r^{(i)}$ is the r:th eigenvector located in the r:th column of $\mathbf{V}^{(i)}$ associated with the r:th eigenvalue $\mathbf{\Lambda}_{rr}^{(i)} = \lambda_r^{(i)}$ of $\mathbf{L}^{(i)}$; for $r = 1, 2, ..., N^{(i)}$. Similarly, define a second network $G^{(j)}$ of size $N^{(j)}$ and order $K^{(j)}$ with corresponding $N^{(j)} \times N^{(j)}$ unnormalized Laplacian matrix $\mathbf{L}^{(j)}$ with a diagonal $N^{(j)} \times N^{(j)}$ eigenvalues matrix $\mathbf{\Lambda}^{(j)}$, associated with a $N^{(j)} \times N^{(j)}$ eigenvectors matrix $\mathbf{V}^{(j)}$.

The central idea behind the SGD is to reflect how distant two graphs are based on the difference between the distribution of the elements of their eigenvectors (seen as different realizations of an underlying real-valued random variable). Let $\rho_r^{(i)}$ and $\rho_r^{(j)}$ be the probability distribution of the r:th eigenvector in $\mathbf{V}^{(i)}$ and $\mathbf{V}^{(j)}$, respectively. The distance between $\rho_r^{(i)}$ and $\rho_r^{(j)}$ can be obtained by directly comparing the distribution functions, or by comparing their *cumulative distribution function* (CDF), denoted as $\varrho_r^{(i)}$ and $\varrho_r^{(j)}$. In [3], the authors suggest comparing their *empirical cumulative distribution function* (eCDF). The reason behind this choice is that directly comparing distribution functions will require estimating the probability distributions analytically, which can be problematic. Mainly since it will require building a histogram and introduce an additional parameter: the bin size, while in the case of the eCDF, it can be directly calculated from the elements of the eigenvectors. The *em-*

pirical cumulative distribution function is defined as such:

Definition 2.5.2 (**Empirical cumulative distribution function**). *Let $\{X_i\}_{1\leq i\leq N}$ be a sequence of independent and identically distributed random variables with a common distribution function. The empirical cumulative distribution function (eCDF) is defined as:*

$$\varrho(x) = \frac{|\{i \leq N; X_i \leq x\}|}{N}$$

Note that comparing distribution functions allows comparing eigenvectors of different lengths, enabling the method to compare graphs of different sizes. It is also important to consider that, in order to compare the eCDFs of eigenvectors corresponding to a distinct pair of graphs, one needs to first ensure that the exact range of values is the same for all eigenvectors. Thus the authors of [3] suggest performing a *re-scaling* (or min-max normalization) on the eigenvectors, so that the minimum and maximum element value for all eigenvectors are respectively zero and one. The re-scaling function $z : \mathbb{R} \to [0, 1]$ is defined as:

$$z(\boldsymbol{v}) = \frac{\boldsymbol{v} - \min(\boldsymbol{v})}{\max(\boldsymbol{v}) - \min(\boldsymbol{v})} \tag{14}$$

For calculating the distance between the re-scaled eigenvectors distribution the authors of [3] suggest to use the *Kruglov distance*, defined as:

Definition 2.5.3 (**Kruglov distance**). *Let P be the set of all cumulative distribution functions. The Kruglov distance is a distance on P defined as:*

$$\int_{-\infty}^{\infty} f(\varrho(X \leq x) - \varrho'(X \leq x))\, dx$$

Where $\varrho, \varrho' \in P$, and the function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is any even strictly increasing function.

Given that the main purpose of the Kruglov distance is to capture the area between a pair of CDFs, a simple choice of the function f in Definition 2.5.1 is the *absolute value function*. The Kruglov distance between the distributions $\rho^{(i)}, \rho^{(j)}$ can be written as:

$$d(\rho^{(i)}, \rho^{(j)}) = \int_{-\infty}^{\infty} |\varrho^{(i)}(x) - \varrho^{(j)}(x)|\, dx, \tag{15}$$

Let P be the set of all CDFs, then $\forall \rho^{(i)}, \rho^{(j)}, \rho^{(k)} \in P$:

(i) $d(\rho^{(i)}, \rho^{(j)}) \geq 0$
(ii) $d(\rho^{(i)}, \rho^{(j)}) = 0 \iff \rho^{(i)} = \rho^{(j)}$
(iii) $d(\rho^{(i)}, \rho^{(j)}) = d(\rho^{(j)}, \rho^{(i)})$
(iv) $d(\rho^{(i)}, \rho^{(j)}) \leq d(\rho^{(i)}, \rho^{(k)}) + d(\rho^{(k)}, \rho^{(j)})$

Property (i) holds since the absolute value function is a non-negative function, (ii) holds since $d(\rho^{(i)}, \rho^{(j)}) = 0$ implies that the absolute value function is zero, and this

can only happen if and only if $\rho^{(i)} = \rho^{(j)}$, (iii) is trivial, finally, (iv) holds since $\forall \rho^{(i)}, \rho^{(j)}, \rho^{(k)} \in P$, and by using the triangle's inequality:

$$\begin{aligned} d(\rho^{(i)}, \rho^{(j)}) &= \int_{-\infty}^{+\infty} |\varrho^{(i)}(x) - \varrho^{(k)}(x) + \varrho^{(k)}(x) - \varrho^{(j)}(x)| \, dx \\ &\leq \int_{-\infty}^{+\infty} |\varrho^{(i)}(x) - \varrho^{(j)}(x)| + |\varrho^{(j)}(x) - \varrho^{(k)}(x)| \, dx \\ &= d(\rho^{(i)}, \rho^{(k)}) + d(\rho^{(k)}, \rho^{(j)}) \end{aligned} \tag{16}$$

The *spectral graph distance* between two networks $G^{(i)}$ and $G^{(j)}$ is finally defined as:

$$D(G^{(i)}, G^{(j)}) = \frac{1}{M_{ij} - 1} \sum_{r=2}^{M_{ij}} d(\rho_r^{(i)}, \rho_r^{(j)}) \tag{17}$$

Where M_{ij} is the cut-off value; $M_{ij} = \min(N^{(i)}, N^{(j)})$. The SGD in Eq. (17) is a weighted sum of the Kruglov distances between the distribution of the $M_{ij} - 1$ eigenvectors associated with the smallest non-zero eigenvalues. Since the eigenvector associated with the zero-value eigenvalue is always constant, it does not provide information about the structure of the graph. For this reason, it is not included in the sum in Eq. (17).

2.5.2.1 Properties of the SGD

The spectral graph distance is not a distance *per se*. Note that two eigenvectors having the same CDF does not necessarily result in them being equal i.e., whenever $D(G^{(i)}, G^{(j)}) = 0 \implies \forall r \in [2, M_{ij}] : \rho_r^{(i)} = \rho_r^{(i)} \not\implies G^{(i)} = G^{(j)}$. As mentioned in [14], this can be problematic if one aims to perform a rigorous analysis on the distance. However, in practice, finding two graphs where $D(G^{(i)}, G^{(j)}) = 0$ is unlikely to happen. Nevertheless, the SGD satisfies certain properties of a distance function (Definition 2.5.1). Let $\mathbb{G}$ be the set of all undirected and unweighted connected graphs and $\mathbb{P}$ the set of all distribution functions. Then, for all pairs of graphs $G^{(i)}, G^{(j)} \in \mathbb{G}$, some properties can be considered: The SGD is (i) **non-negative** $D(G^{(i)}, G^{(j)}) \geq 0$, since for all probability distributions $\rho^{(i)}, \rho^{(j)} \in \mathbb{P}$, the Kruglov distance $d(\rho^{(i)}, \rho^{(j)})$ is non-negative, then Eq. (17) must also be non-negative, and (ii) **symmetric** $D(G^{(i)}, G^{(j)}) = D(G^{(j)}, G^{(i)})$ since for any $\rho^{(i)}, \rho^{(j)} \in \mathbb{P}$ it holds that $d(\rho^{(i)}, \rho^{(j)}) = d(\rho^{(j)}, \rho^{(i)})$ which implies that $D(G^{(i)}, G^{(j)}) = D(G^{(j)}, G^{(i)})$.

2.5.2.2 The sign ambiguity of the eigenvectors of the graph Laplacian

Given a network G of size N, the eigendecomposition of its unnormalized Laplacian matrix $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, is given by the diagonal eigenvalue matrix $\mathbf{\Lambda}$ and the associated eigenvector matrix $\mathbf{V}$. The matrix $\mathbf{V}$ is an orthogonal matrix, whose eigenvectors are unique up to a constant factor ± 1, which means that there are 2^N possible forms of $\mathbf{V}$. Although the sign of the eigenvectors does not influence the information they convey, it does influence the Kruglov distance. Depending on which sign of the eigenvector is chosen, the result of the Kruglov distance might differ. In order to circumvent this artifact, the authors of [3] suggest rewriting the spectral graph distance in such a way that the evaluation of the Kruglov distance in terms of the eigenvectors is minimized

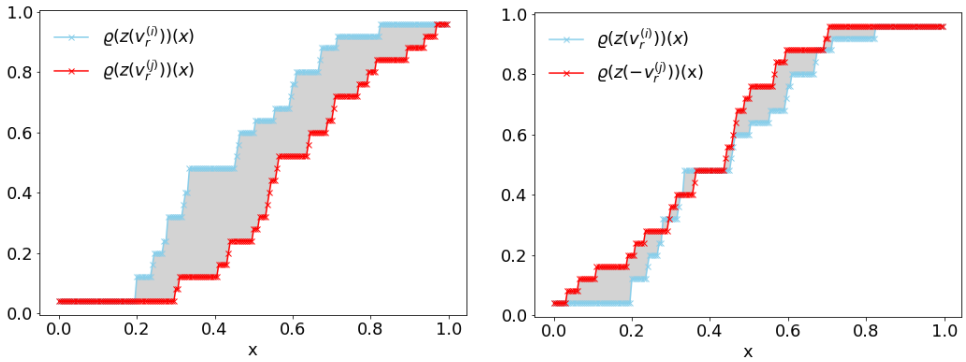
with respect to the sign of the eigenvectors. As a result, the spectral graph distance in Eq. (17) between two graphs $G^{(i)}$ and $G^{(j)}$ can be re-written as:

$$D(G^{(i)}, G^{(j)}) = \frac{1}{M_{ij}-1} \sum_{r=2}^{M_{ij}} \Big\{ \min_{s,l \in \{-1,1\}} d[\rho(z(s \cdot \boldsymbol{v}_r^{(i)})),\ \rho(z(l \cdot \boldsymbol{v}_r^{(j)}))] \Big\} \tag{18}$$

Where s, l determine the sign of the eigenvectors prior to the re-scaling. As shown in Appendix A.1, the SGD in Eq. (18) can be simplified to:

$$D(G^{(i)}, G^{(j)}) = \frac{1}{M_{ij}-1} \sum_{r=2}^{M_{ij}} \Big\{ \min\{d(\rho(\boldsymbol{v}_r^{(i)}), \rho(\boldsymbol{v}_r^{(j)})),\ d(\rho(\boldsymbol{v}_r^{(i)}), \rho(-\ \boldsymbol{v}_r^{(i)})\} \Big\} \tag{19}$$

Fig.10 illustrates how the chosen signs of the eigenvectors affect the evaluation of the Kruglov distance to the spectral graph distance. Note how the "flipping" the eigenvector $\boldsymbol{v}_r^{(j)}$ (prior to re-scaling) results in a reduction of the grey area between the two cumulative distributions. Hence, according to the minimization problem in Eq. (19), the solution to the distance between the r:th eigenvector of both graphs is given by the representation in Fig. 10 (b).



(a) (b)

Figure 10: Cumulative distribution distance for different choices of sign of the re-scaled r:th eigenvectors $\boldsymbol{v}_r^{(i)}, \boldsymbol{v}_r^{(j)}$ of a pair of graphs $G^{(i)}$ and $G^{(j)}$, where (a) illustrates the distance between $\varrho(z(\boldsymbol{v}_r^{(i)}))$ and $\varrho(z(\boldsymbol{v}_r^{(j)}))$, and (b) the distance between $\varrho(z(\boldsymbol{v}_r^{(i)}))$ and $\varrho(z(-\boldsymbol{v}_r^{(i)}))$. The grey areas of both plots denote the Kruglov distance of the r:th eigenvector $d(\rho_r^{(i)}, \rho_r^{(j)})$.

2.5.2.3 Limitations and artifacts of the SGD

Previous literature related to graph-based methods [1, 2] in spectral graph theory established that the eigenvectors and eigenvalues associated with the Laplacian matrix manage to capture meaningful information about the structure of the data. In that regard, and unlike the matrix distances, the spectral graph distance has the advantage of reflecting the information captured by the Laplacian matrix. However, certain aspects of the SGD still need to be taken into consideration; to begin with, the SGD does not consider the relative weights of the eigenvectors, as all eigenvectors in Eq. (17) contribute equally. Hence, the sum of the Kruglov distances is somewhat problematic: as previously discussed in [2], different eigenvectors carry a different amount of variability. Furthermore, the spectral graph distance assumes that the eigenvectors are uncorrelated, as the sum in Eq. (17) does not include cross terms.

The SGD also requires correspondence between the eigenvectors, so that comparison between eigenvectors is established in accordance to their associated eigenvalues. Another important aspect is related to the size difference between graphs: although the method is able to compare eigenvectors of different sizes, the risk of losing information from high order eigenvectors corresponding to a larger graph still exists. Generally, this issue will become a problem if one of the graphs is considerably smaller than the other; otherwise, it will not have an important effect, as high order eigenvectors, given their low variability, often do not provide much information about the structure of the graph; this is further explained in [2]. Finally, the use of the unnormalized Laplacian in defining the SGD is another aspect to consider: as discussed in Section 2.4.3 and [2, 18], it might happen that high order eigenvectors of the unnormalized Laplacian behave like a delta function, which can lead to an increase of the SGD. This may become a source of artifact in the computation of graph similarity through this measure. This last point will be discussed more in-depth in Section 3.

3 Results

In order to evaluate the performance of the SGD, this section will present the results obtained from different experiments on graphs sampled from the Watts-Strogatz (W-S) model introduced in Section 2.2.2. The experiments will be performed by comparing an ensemble of graphs with a reference graph generated by fixing certain parameters. In this context, the measure of performance is considered to be the ability of the graph distance to uncover the fixed parameter used to generate reference graphs, as a way of showcasing the power of the method to capture structural properties. The SGD between the reference graphs and the rest of graphs will also be compared to the *Hamming distance* introduced in Section 2.5.1, and the *degree distribution distance* (DDD). The degree distribution distance can be easily obtained by calculating the Kruglov distance in Eq. 15, between the CDF of the graph's degree distribution (Definition 2.2.1). The motivation for these comparisons is to (i) show if the SGD outperforms the matrix distance, and (ii) try to uncover if there exists any specific structural property (e.g. degree distribution) that could dominate the SGD.

The SGD was implemented in the Python Programming Language. Graphs were constructed using the `NetworkX` library, and the solution to the eigenvalue problem of the Laplacian matrix was obtained by using the `SciPy` library.

3.1 Watts and Strogatz model

This section will present the results of the following experiments:

(i) A reference graph $G(k_0)$ is generated from the W-S model with a fixed reference node degree k_0 and compared with a set of graphs, each generated for varying value of k, while fixing the number of nodes N and the rewiring probability p.

(ii) A reference graph $G(p_0)$ is generated from the W-S model with a fixed reference rewiring probability p_0, and then compared with a set of graphs, each generated for a value of the rewiring probability p, while fixing the number of nodes N and node degree k.

3.1.1 Experiments on networks with fixed rewiring probability p and varying node degree k

The different steps of this experiment were as follow:

- A reference model $G(k_0)$ was generated for each node degree $k_0 \in \{10, 100, 220, 280\}$.
- An ensemble of 50 graphs was generated for 22 values of k (50 graphs for each k), chosen between $[6, 294]$.
- All graphs are of size $N = 300$ and were constructed with a rewiring probability $p = 0.1$.
- For each value of k, the SGD between the reference graph and each graph of the ensemble was computed.

Fig. 11 shows the SGD average over the ensemble of graphs and each reference graph $G(k_0)$. Several important properties of the SGD are summarize as follows:

(i) All experiments show a rapid increase of the SGD around $k = 280$ and $k = 294$. This behavior is caused by the high Kruglov Distance contributions from the high order eigenvectors, whose elements are almost constant. To illustrate this effect, Fig. 12 plots the Kruglov distance contributions to the SGD between a graph $G(k); k \in \{10, 100, 220\}$ (sampled from the ensemble) and the reference graph $G(10)$. Note that for all cases the highest contributions come from the high order eigenvectors. In order to take a closer look into this issue, Fig. 13 plots the eigenvectors that contribute the most, for the case when $k_0 = 10$ and $k = 294$. The first plot in Fig. 13 highlights the three highest Kruglov distance contributions, in this case are attributed to the 300th, 294th and 292nd eigenvectors. Moreover, the eigenvectors of the reference and sampled graph associated with these high contributions are shown in the 3 left panels of Fig. 13. Note that the 300th eigenvector of the sampled graph is constant except for the minimum value component. For the rest of the eigenvectors, although they show more variability, (excluding the 292nd and 294th eigenvector of the reference graph), the elements are overall constant, except for the component with minimum value.

(ii) Another aspect to notice is that almost all experiments present a flat and constant phase around intermediate values of k. This is caused by the uniform contribution of the Kruglov distance from different eigenvectors. Fig. 12 illustrates this effect for the case when $k = 10$. Note that the transition from $k = 10$ to higher values of k is only affected by the contribution of eigenvectors associated with small and high order eigenvalues. For $k = 10$, this effect is most notable as all eigenvectors' distances contribute in a very similar manner. Although a change in the shape of the distribution can be noticed as k increases, the overall contribution of all eigenvectors remains very uniform.

(iii) Additionally, when $k_0 = k = 10$, the SGD in Fig. 11 manages to capture the difference between the networks, as the minimal SGD is reached when $k = k_0$. However, compared to the maximum SGD value (excluding the last experiment, when $k = 294$), the minimal SGD reached in all experiments is not close to zero, caused by the uniform contributions shown in Fig. 12. This means that, even though the parameters k and k_0 are the same, as long as the graphs are not identical, the SGD will not be close to zero.

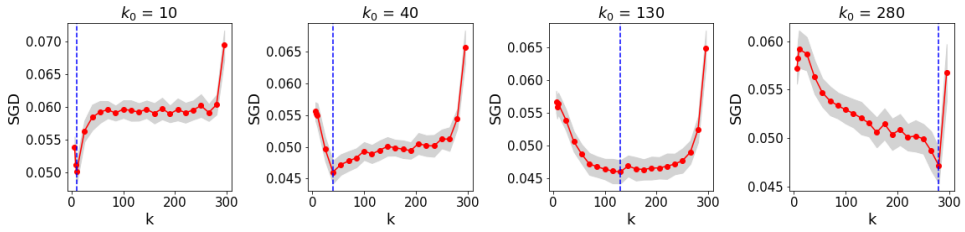


Figure 11: Average SGD between the reference graph $G(k_0)$ and 50 graph replicates $G(k)$ generated from the W-S model for $6 \leq k \leq 294$. Graphs were constructed with a rewiring probability of $p = 0.1$ and size $N = 300$. The dotted blue line corresponds to k_0. The grey area denotes $\pm$ standard deviation.

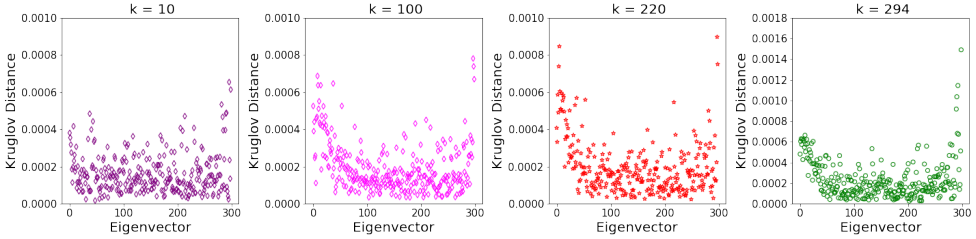


Figure 12: Kruglov distance contributions between eigenvectors of a reference graph $G(10)$ and a single graph $G(k)$ randomly chosen from the ensemble of 50 graphs generated for each $k \in \{10, 100, 220, 294\}$.

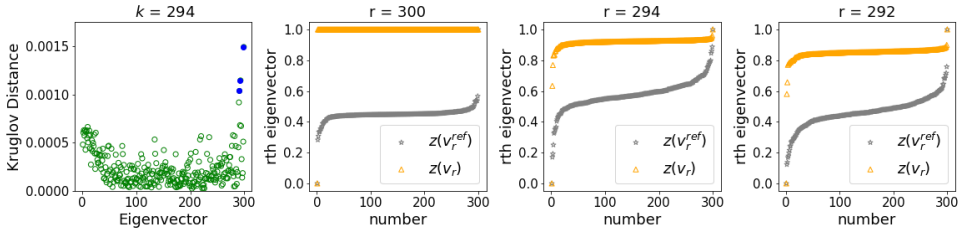


Figure 13: Left: Kruglov distance contributions between eigenvectors of the reference graph $G(10)$ and a graph $G(294)$. The highlighted markers refer to the 3 highest contributions. Last three plots refer to the r:th re-scaled eigenvectors of $G(10)$ and $G(294)$ in descending order of contribution to the SGD.

On the other hand, Fig. 14 shows the average over the (i) DDD, and (ii) Hamming distance between the ensemble of graphs and each reference graph. It is not surprising that both, the DDD in Fig. 14 (a) and the Hamming distance in Fig. 14 (b) (except when $k_0 = 280$) manage to distinguish the graphs and correctly identify the minimal distance when $k_0 = k$. In the context of this experiment, this results as the varying parameter k defines the main difference between the compared graphs.

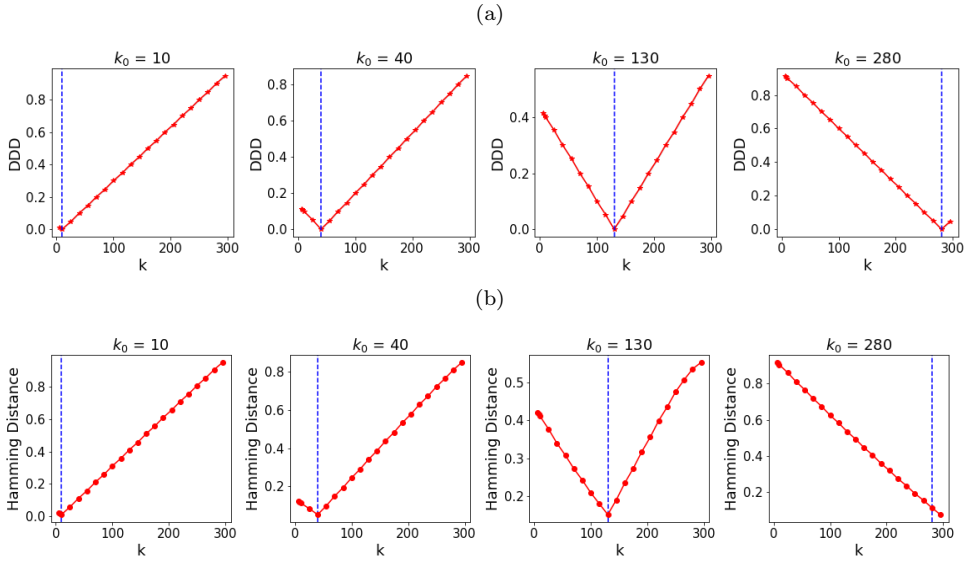


Figure 14: Average (a) DDD, and (b) Hamming distance between the reference graph $G(k_0)$ and 50 graph replicates $G(k)$ generated from the W-S model for $6 \leq k \leq 294$. Graphs were constructed with a rewiring probability of $p = 0.1$ and size $N = 300$. The dotted blue line corresponds to k_0.

3.1.2 Experiments on networks with fixed node degree k and varying rewiring probability p

The different steps of this experiment were as follow:

- A reference model $G(p_0)$ was generated for each rewiring probability $p_0 \in \{0.1, 0.2, 0.6, 0.9\}$
- An ensemble of 50 graphs was generated for 50 values of p (50 graphs for each p), chosen between $(0, 1]$.
- All graphs are of size $N = 300$ and have node degree $k = 10$.
- For each value of p, the SGD between the reference graph and each graph of the ensemble was computed.

Results of this experiment are shown in Fig. 15, where each row refers to the resulting average over the (i) SGD, (ii) DDD, and (iii) Hamming distance between the ensemble of graphs and each reference graph $G(p_0)$. Several important properties of the SGD are summarize as follows:

(i) For high values of p, the SGD in Fig. 15 (a) shows small or no changes, and fails to distinguish graphs. This behaviour is caused by the introduction of randomness in the reference graph and the rest of the graphs, as by increasing the values of p and p_0, the variance in the degree distribution of the graphs also increases.

(ii) Fig. 15 (b) shows that the degree distribution seems to explain the information captured by the SGD. Even if there is no clear connection between the SGD and the degree distribution, the degree distribution is solely based on the frequency of the nodes' degree. Therefore, it will fail to capture other structural properties of the network (e.g. clustering structures), that can be captured by the SGD. Hence, the similar behavior between the SGD and the DDD will not necessarily hold for other types of networks, as in the case of the experiment performed in Section 3.1.1, where the DDD showed a different behavior of the SGD.

(iii) For all p_0 the SGD outperforms the Hamming distance. Also note that when $p_0 = 0.1$, the minimal SGD is achieved when $p_0 = p$. Nevertheless, even if the SGD performs better, the DDD provides similar information as the SGD at a lower computational cost.

(iv) The Hamming distance in Fig. 15 (c) shows a linear behavior for all experiments, resulting from the introduction of randomness in the graphs. As the rewiring probability p increases, the graphs become less regular and the similarities of the node's connectivity more distant.

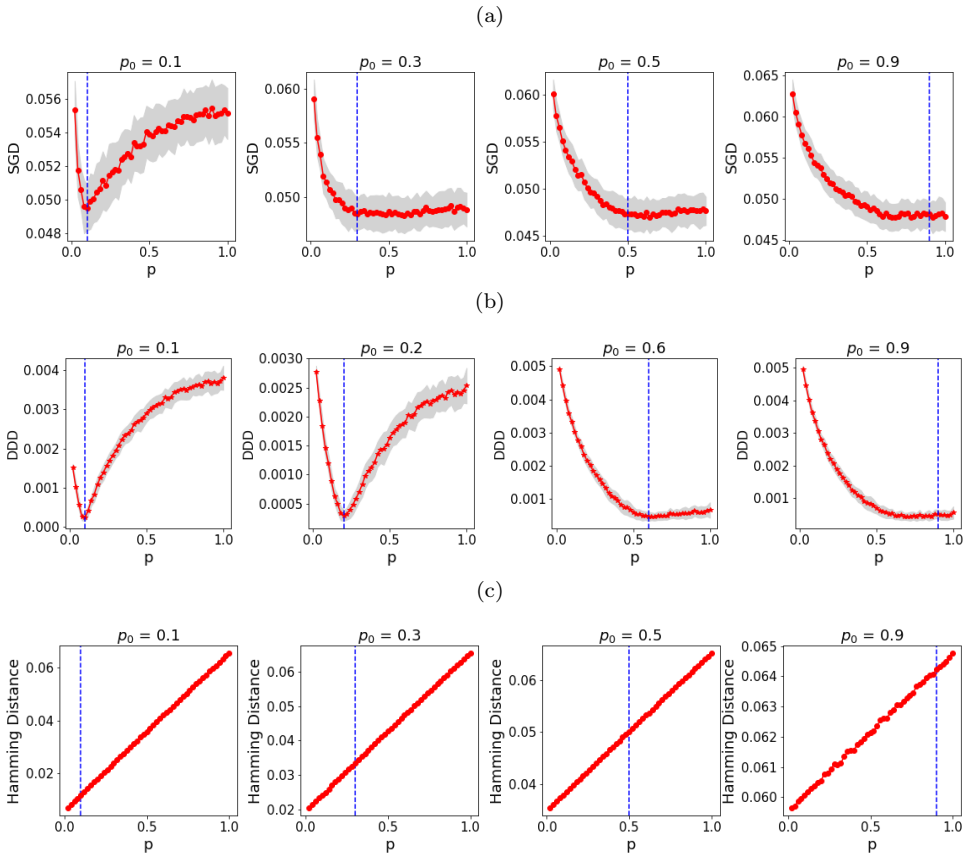


Figure 15: Average (a) SGD, (b) DDD, and (c) Hamming distance between the reference graph $G(p_0)$ and 50 replicates $G(p)$ generated from the W-S model for $0 \leq p \leq 1$. The number of nodes $N = 300$ with node degree $k = 10$. The dotted blue line corresponds to p_0. The grey area denotes $\pm$ standard deviation.

4 Discussion and Conclusions

The present work aimed to provide a thorough introduction to the SGD, as well as a review of related topics in graph theory and network theory. Section 2.1 presented different ways to define graphs and basic concepts in graph theory. A discussion regarding relevant aspects of complex networks followed in Section 2.2, focusing on two widely studied types of networks – the random graphs and the small-world networks. Due to their useful theoretical properties, random graphs are commonly used to compare properties of real-world networks. Similarly, properties of networks generated from the W-S model were also considered for their ability to generate graphs with structural properties commonly found in real systems. Section 2.3 introduced and compared different approaches to impose a graph structure on data. An overview of connectivity matrices derived from graphs was proposed in Section 2.4, which looked at the different ways of defining the graph Laplacian and their relationships, limitations, and artifacts. It was thus concluded that, in the context of the SGD, the main artifact introduced by the unnormalized Laplacian form is explained by the fact that high order eigenvectors could approximate a delta function. Section 2.5 provided with some background on matrix distances and spectral distances. The Hamming distance

was presented, followed by the theory supporting the SGD, as well as related technical aspects and practical considerations.

A simplified version of the original graph distance introduced by the authors in [3] was also offered to the reader. The main limitations and artifacts of the SGD were also identified and discussed. Concluding that the main limiting aspect of the SGD is that it considers the distances between eigenvectors to be equally important. As such, the SGD becomes affected by noisy high-order eigenvectors that could overshadow the contribution of the distribution distance of more relevant eigenvectors. Overall, the fact that high order eigenvectors have a significant contribution is a drawback of the SGD. Another important limitation of the SGD is that it requires correspondence between eigenvectors and it assumes that all eigenvectors are independent. In Section 3, two different experiments were performed on graphs sampled from the W-S model, where the SGD was able to correctly identify the reference models fixed parameters; the rewiring probability in the first case and the node degree in the second case (at least to some extent). As a result the SGD did not perform as well as the degree distribution distance (DDD) when comparing graphs sampled from the W-S model. An alternative proposal to evaluate the performance of the SGD could be to consider graphs that present a scale-free distribution (or scale-free networks), as they have different structural properties. In conclusion, even though the experiments presented in the Results Section 3 only covered graphs from the W-S model, the application of the SGD can be extended to all the possible ways of simulating graphs and constructing graphs from data that were introduced in this thesis.

4.1 Outlook on further studies

The authors of [3] established the graph distance by only considering unweighted graphs and the associated unnormalized graph Laplacian. Nevertheless, many aspects of the SGD can be generalized and further extended. For instance, the SGD could be extended to the general case of weighted graphs and it could also be reformulated in terms of the normalized Laplacian matrix representation. Another possible generalization could be to consider a different distribution-distribution distance such that the eigenvector's importance is reflected on each contribution.

References

[1] Belkin M and Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15:1373–1396, (2003).

[2] von Luxburg U. A tutorial on spectral clustering. *Statistics and computing*, 17:395–416, (2007).

[3] Shimada Y, Hirata Y, Ikeguchi Y, and Aihara K. Graph distance for complex networks. *Scientific Reports*, 6:1–6, (2016).

[4] Latora V, Nicosia V, and Russo G. *Complex Networks: Principles, Methods and Applications*. Cambridge University Press, 1st edition, (2017).

[5] Lusseau L and Newman M. Identifying the role that animals play in their social networks. *Biology Letters*, 271:S477–S481, (2004).

[6] Erdős P and Rényi A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–60, (1960).

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

[8] Telesford Q, Joyce K, Hayasaka S, Burdette J, and Laurienti P. The ubiquity of small-world networks. *Brain connectivity*, 1:367–375, (2011).

[9] Barrat A and Weigt M. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13:547–560, (2000).

[10] De la Porte J, Herbst BM, W. Hereman, and SJ. Van Der Walt. An introduction to diffusion maps. pages 15–25, (2008).

[11] Cortes C and Vapnik V. Support-vector networks. *Machine learning*, 20:273–297, (1995).

[12] Chung F and Graham R. *Spectral graph theory*. Number 92. American Mathematical Soc., (1997).

[13] Deza M and Deza E. *Encyclopedia of distances*. Springer, (2009).

[14] Wills P and Meyer F.G. Metrics for graph comparison: A practitioner's guide. *PloS ONE*, 15:e0228728, (2020).

[15] de Lange S, de Reus M, and Van Den Heuvel M. The Laplacian spectrum of neural networks. *Frontiers in computational neuroscience*, 7:189, (2014).

[16] Newman M. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103:8577–8582, (2006).

[17] Wilson R and Zhu P. A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41:2833–2841, (2008).

[18] von Luxburg U, Bousquet O, and Belkin M. *On the convergence of spectral clustering on random samples: the normalized case*. (2004).

Appendices

Appendix A Proofs

A.1 Equality of the Kruglov Distance of flipped eigenvectors.

Define G to be a graph of size N, whose Laplacian matrix $\mathbf{L}$ has N real eigenvalues associated with N real and linearly independent eigenvectors. Define also the $N \times N$ eigenvectors matrix $\mathbf{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N)$, such that $\forall r;\ \boldsymbol{v}_r \in \mathbf{V}$ it holds that:

$$\begin{aligned}
\max(-\boldsymbol{v}_r) = -\min(\boldsymbol{v}_r) &\iff \forall j\ \boldsymbol{v}_{rj} \leq \max(\boldsymbol{v}_r) \\
\min(-\ \boldsymbol{v}_r) = -\max(\boldsymbol{v}_r) &\iff \forall j\ -\boldsymbol{v}_{rj} \geq -\max(\boldsymbol{v}_r) = \min(\boldsymbol{v}_r)
\end{aligned}$$

The re-scaled form of $\pm\boldsymbol{v}_r$ is given by:

$$\begin{aligned}
z(\boldsymbol{v}_r) &= \frac{\boldsymbol{v} - \min(\boldsymbol{v}_r)}{\max(\boldsymbol{v}_r) - \min(\boldsymbol{v}_r)} \\
z(-\boldsymbol{v}_r) &= \frac{-\boldsymbol{v}_r + \max(\boldsymbol{v}_r)}{\max(\boldsymbol{v}_r) - \min(\boldsymbol{v}_r)} = \frac{-\boldsymbol{v}_r + \min(\boldsymbol{v}_i)}{\max(\boldsymbol{v}_r) - \min(\boldsymbol{v}_r)} + 1 = 1 - z(\boldsymbol{v}_r)
\end{aligned}$$

While the eCDF of the re-scaled form of $\pm\boldsymbol{v}_i$ is:

$$\begin{aligned}
\varrho(z(\boldsymbol{v}_r))(x) &= \frac{1}{N}\sum_{l=1}^{N} \mathbb{1}\{z_l(\boldsymbol{v}_r) \leq x\} \\
\varrho(z(-\boldsymbol{v}_r))(x) &= \frac{1}{N}\sum_{l=1}^{N} \mathbb{1}\{z_l(-\boldsymbol{v}_r) \leq x\} = \frac{1}{N}\sum_{l=1}^{N} \mathbb{1}\{1 - x \leq z_l(\boldsymbol{v}_r)\} = 1 - \varrho(z(\boldsymbol{v}_r))(1-x)
\end{aligned}$$

For a pair of graphs $G^{(i)}$ and $G^{(j)}$ of size $N^{(i)}$ and $N^{(j)}$ and unnormalized Laplacian matrices $\mathbf{L}^{(i)}$, $\mathbf{L}^{(j)}$ with respective eigenvector matrices $\mathbf{V}^{(i)}$ and $\mathbf{V}^{(j)}$. It can be shown that the Kruglov distance between the probability distributions $\rho(z(-\boldsymbol{v}_r^{(i)}))$ and $\rho(z(\boldsymbol{v}_r^{(j)}))$ is the same as the Kruglov distance between the probability distributions $\rho(z(\boldsymbol{v}_r^{(i)}))$ and $\rho(z(-\boldsymbol{v}_r^{(j)}))$. That is:

$$\begin{aligned}
d(\rho(z(-\boldsymbol{v}_r^{(i)})),\ \rho(z(\boldsymbol{v}_r^{(j)}))) &= \int_0^1 |1 - \varrho(z(\boldsymbol{v}_r^{(i)}))(1-x) - \varrho(z(\boldsymbol{v}_r^{(j)}))(x)|\ dx \\
&= \int_1^0 |1 - \varrho(z(\boldsymbol{v}_r^{(i)}))(u) - \varrho(z(\boldsymbol{v}_r^{(j)}))(1-u)|\ (-du) \\
&= \int_0^1 |1 - \varrho(z(\boldsymbol{v}_r^{(j)}))(1-u) - \varrho(z(\boldsymbol{v}_r^{(i)}))(u)|\ du \\
&= \int_0^1 |\varrho(z(-\boldsymbol{v}_r^{(j)}))(u) - \varrho(z(\boldsymbol{v}_r^{(i)}))(u)|\ du \\
&= d(\rho(z(\boldsymbol{v}_r^{(i)})),\ \rho(-z(\boldsymbol{v}_r^{(j)})))
\end{aligned} \tag{20}$$

Similarly, it can also be shown that the Kruglov distance between $\rho(z(-\boldsymbol{v}_r^{(i)}))$ and $\rho(z(-\boldsymbol{v}_r^{(j)}))$ is the same as the Kruglov distance between the probability distributions $\rho(z(\boldsymbol{v}_r^{(i)}))$ and $\rho(z(\boldsymbol{v}_r^{(j)}))$. That is:

$$
\begin{aligned}
d(\rho(z(-\boldsymbol{v}_r^{(i)})),\ \rho(z(-\boldsymbol{v}_r^{(j)}))) &= \int_0^1 |(1-\varrho(z(\boldsymbol{v}_r^{(i)}))(1-x)) - (1-\varrho(z(\boldsymbol{v}_r^{(j)}))(1-x))|\ dx \\
&= \int_0^1 |\varrho(z(\boldsymbol{v}_r^{(j)}))(1-x) - \varrho(z(\boldsymbol{v}_r^{(i)}))(1-y)|\ dx \\
&= \int_0^1 |\varrho(z(\boldsymbol{v}_r^{(j)}))(u) - \varrho(z(\boldsymbol{v}_r^{(i)}))(u)|\ du \\
&= d(\rho(z(\boldsymbol{v}_r^{(i)})),\ \rho(z(\boldsymbol{v}_r^{(j)})))
\end{aligned}
\tag{21}
$$

Appendix B Algorithms

B.1 ER model: binomial random graph

Algorithm 1: ER: Binomial Random Graph

Input: Number of nodes $N \in \mathbb{Z}^+$; probability $p \in [0,1]$

Output: A graph $G = (\mathcal{N}, \mathcal{L})$, where $G \in G_{N,p}^{ER}$.

$\mathcal{N} \leftarrow \{1, 2, ..., N\}$;
$\mathcal{L} \leftarrow \emptyset$;
for $i = 1, ..., N$ **do**
    for $j = i+1, ..., N$ **do**
        $\epsilon \leftarrow random(0,1)$;
        if $\epsilon < p$ **then**
            $\mathcal{L} \leftarrow \mathcal{L} \cup \{(i,j)\}$;
return $G = (\mathcal{N}, \mathcal{L})$

B.2 Watts-Strogatz model

Algorithm 2: Watts-Strogatz model

Input: Number of nodes $N \in \mathbb{Z}^+$; m; $p \in [0, 1]$

Output: A graph $G = (\mathcal{N}, \mathcal{L})$

$\mathcal{N} \leftarrow \{1, 2, ..., N\}$;
$k \leftarrow \lfloor \frac{m}{2} \rfloor$
for $i = 1, ..., N$ **do**
    for $j = 1, ..., k + 1$ **do**
        $l \leftarrow i + j \mod N$;
        $\mathcal{L} \leftarrow \mathcal{L} \cup \{(i,\ l)\}$;
for $j = 1, ..., k + 1$ **do**
    $i \leftarrow 1$
    for $l \in \{j + 1, ..., N\} \cup \{1, ..., j + 1\}$ **do**
        $\epsilon \leftarrow random(0, 1)$
        if $\epsilon < p$ **then**
            $\epsilon' \leftarrow random(0, 1)$
            $l' \leftarrow \lfloor N\epsilon' \rfloor$
            while $l' = i$ *or* $(i, l') \subset \mathcal{L}$ **do**
                $\epsilon' \leftarrow random(0, 1)$
                $l' \leftarrow \lfloor N\epsilon' \rfloor$
                if $deg(i) = N - 1$ **then**
                    break
            else
                replace (i, l) with (i, l')
        $i \leftarrow i + 1$
return $G = (\mathcal{N}, \mathcal{L})$