

Politecnico di Milano

Doctoral Dissertation of:
Anna Calissano

Supervisor:
Prof. Simone Vantini

Co-Supervisor:
Prof. Valeria Fedeli

POPULATIONS
OF
GRAPHS

: STATISTICAL ANALYSIS
IN THE GRAPH SPACE
WITH APPLICATIONS
TO URBAN MOVEMENTS



POLITECNICO DI MILANO
DEPARTMENT OF MATHEMATICS
DOCTORAL PROGRAM IN
MATHEMATICAL MODELS AND METHODS IN ENGINEERING

POPULATIONS OF GRAPHS:
STATISTICAL ANALYSIS IN THE GRAPH SPACE
WITH APPLICATIONS TO URBAN MOVEMENTS.

Doctoral Dissertation of:
Anna Calissano

Supervisor:
Prof. Simone Vantini

Co-Supervisor:
Prof. Valeria Fedeli

The Chair of the Doctoral Program and Tutor:
Prof. Irene Maria Sabadini

2020 – XXXIII Cycle

Abstract

Populations of graphs are a complex and strongly non-Euclidean data type describing different relational phenomena in different fields. The aim of this PhD thesis is to develop statistical tools for the analysis of populations of unlabelled graphs, embedding them in the Graph Space, a quotient space of permuted adjacency matrices. We perform cluster analysis, we define Geodesic Principal Components, and Graph-valued regression model. We introduce an algorithm, namely Align All and Compute, to estimate the defined intrinsic statistic in the Graph Space. These original statistical tools are applied to quantify and analyse urban movements, in order to understand how people move within a square, a city, a region. We discuss if the description of a spatial occurrence in an abstract space can reveal interesting perspectives about the analysis of reality.

Sommario

Molti fenomeni appartenenti alle discipline più disparate possono essere quantificati grazie all'utilizzo dei grafi, ed in particolare all'uso di popolazioni di grafi. L'obiettivo di questa tesi è quello di sviluppare metodologie statistiche per l'analisi di popolazioni di grafi. In particolare, la tesi si focalizza su popolazioni di grafi senza corrispondenze di nodi tra un grafo e l'altro. Queste osservazioni sono rappresentate come classi di equivalenza di grafi con nodi permutati in uno spazio quoziente denominato Graph Space. All'interno di questo contesto geometrico, la tesi sviluppa metodologie statistiche quali le Componenti Principali, e la regressione per prevedere una popolazione di reti. Dal punto di vista applicativo, la tesi si concentra sull'utilizzo delle tecniche statistiche sviluppate allo studio della mobilità urbana su diverse scale: locale (piazza, parco), urbana (città, area metropolitana), e territoriale. L'applicazione pone quesiti interessanti relativi all'utilizzo di uno spazio geometrico astratto, quale il Graph Space per la descrizione di un fenomeno spaziale.

Acknowledgment

We thank PoliSocial Award to finance the "Safari Njema Project". We thank Trafikselskabet Movia and Cuebiq Inc. to share their precious data with us.

This thesis is the output of different encounters, with brilliant scientist and thinkers who guided me throughout this work. Among the others, I want to thank my supervisors to give me the freedom and the trust to explore this interesting and new topic; Aasa, to host and guide me, showing me a beautiful female leadership; Paola to discuss with me even the most bizarre ideas.

*To loneliness,
the best and worst partner
of a research journey.*

*To the boundaries of the mind,
when you discover how pleasant it is
to surmount them,
there is no way back from
the realm of freedom.*

A Note to the Reader

This PhD thesis is the output of a joint scholarship between the Department of Urban Studies and the Department of Mathematics of Politecnico di Milano. To help the reader navigates the thesis, we display at the beginning of each section a bar, summarizing the percentage of **Urban Studies** and the percentage of **Statistical Modelling** discussed in the Section. Reader interested in one or the other subject can focus only on the related sections.



At the beginning of each section, this bar is reported with the corresponding percentages describing the content of the section.

Contents

1	General Introduction	1
1.1	Populations of Graphs	2
1.2	Urban Movement	4
1.3	Populations of Graphs and Urban Movement	14
2	Embedding in the Graph Space	17
2.1	The Graph Space	19
2.2	Geometrical Characterization of the Graph Space	20
2.3	Discussion	25
3	Exploration in the Graph Space	27
3.1	Clustering in the Graph Space	29
3.2	The Fréchet Mean	34
3.3	Geodesic Principal Component Analysis	35
3.4	Discussion	46
4	Prediction in the Graph Space	53
4.1	Graph-Valued Regression	56
4.2	Conformal Prediction Intervals	68
4.3	Discussion	80
5	Conclusion	89
	Bibliography	91

CHAPTER *1*

General Introduction

How can we compare two graphs with different numbers of nodes? How can we describe the variability of a set of graphs? Is this variability defined in terms of topology, nodes, edges, or attributes? How can we predict a set of graphs from a set of regressors? All these questions can be framed within the analysis of populations of graphs. The aim of this PhD thesis is to answer some of these questions by developing statistical tools for the analysis of populations of graphs. Populations of graphs appear to be a powerful representation of reality in very different fields. We applied the developed statistical tools for the analysis of urban movements represented as a set of graphs. This Introduction 1 is divided in three main sections. In Section 1.1, we introduce the mathematical background of the analysis of population of graphs. In Section 1.2, we introduce the concept of urban movement and we discuss about the analysis of a spatial phenomena and the choice of a suitable geometrical embedding. The last Section 1.3 describes how a population of graphs can describe a urban movement phenomenon. The rest of the thesis is organized as follow:

Chapter 2 describes the chosen embedding space for the analysis of population of graphs, namely Graph Space. We detailed the geometrical properties of the space such as the metric, the geodesics, and the curvature required to define intrinsic statistical tools (Section 2.2).

Chapter 1. General Introduction

Chapter 3 focuses on the definition of new statistical models to explore population of graphs in the Graph Space. We start with a case study about cluster analysis of urban movements (Section 3.1), we discuss the analysis of the properties of the Fréchet Mean (Section 3.2), and we define the concept of Geodesic Principal Components (Section 3.3).

Chapter 4 is about prediction methods. In Section 4.1, we define a graph-valued regression model. In the last Section 4.2, we introduce a conformal prediction interval for graphs, applicable to different prediction models.

1.1 Populations of Graphs

A graph (or a network) is a mathematical structure used to study relational phenomena in different fields. A graph is characterized by a set of nodes (actors, vertices) and a set of relations (links, edges) between them. From a methodological perspective, the analysis of networks roots back in centuries and it has been challenging mathematicians due to its intrinsic complexity. The scientific literature has so far primarily focused on analysing graphs in a *first generation* setting (Wang and Marron, 2007): the analysis and modelling of a single network datum with a set of nodes and a set of edges. The *second generation* approach involves the analysis of a population of network-valued data: the analysis of not one but a set of networks. Network-valued data are an example of complex data and the statistical analysis of such data is generally referred to as *Object-Oriented Data Analysis* Wang and Marron (2007). Note that we will use the term graph and the term network interchangeably throughout the thesis.

The scientific literature regarding the first generation approach to networks is rich and lays at the intersection of different disciplines such as physics and applied science, e.g. (Barabási, 2016); sociology, e.g. (Wasserman and Faust, 1994); statistics, e.g. (Goldenberg et al., 2010; Kolaczyk and Csárdi, 2014); mathematics, e.g. (Chartrand, 1977); biology and medicine e.g. (Forster et al., 2020). Network Analysis studies the complexity of a network object and defines all sort of methods to find hubs and dependencies between nodes, to cluster and to classify nodes, to compute shortest paths, to predict edges, to compute graph density, etc. The second generation approach poses very different questions, aiming at defining statistical methods for the analysis of populations of network-valued data. A population of network-valued data is a sample of networks describing a phenomenon: a set of brain connectivity networks in different patients (Durante et al., 2017); a time series of mobility origin-destination matrices (Calissano et al., 2020b); students' social interactions at different grades at school, a social media network of different plat-

form users etc. These populations of complex data require statistical analysis and pose interesting research questions.

In the analysis of complex data like networks, the starting research question concerns the embedding space: the definition of a space where every complex datum is a point. If all the networks in the sample share the same nodes, the problem could be re-framed as an analysis of a set of matrices (i.e. adjacency matrices). This data type are called *labelled networks* (Jain and Obermayer, 2009). If the networks describe the same class of phenomena, but they have varying numbers of nodes or inconsistent node labelling, the mathematical challenges involved in the analysis become numerous. These data are called *unlabelled networks* and they are going to be the object of analysis in the current work. In Figure 1.1, two simple weighted undirected graphs are compared in a labelled or unlabelled setting, leading to very different results. Note that there are an infinite possible intermediaries between the fully unlabelled and the fully labeled networks. In Chapter 2, a detailed review of different embedding strategies for populations of unlabelled networks is offered, together with the description of the chosen embedding space of the current work: *Graph Space*. Graph Space is a quotient space obtained by applying node permutation to the space of adjacency matrices.

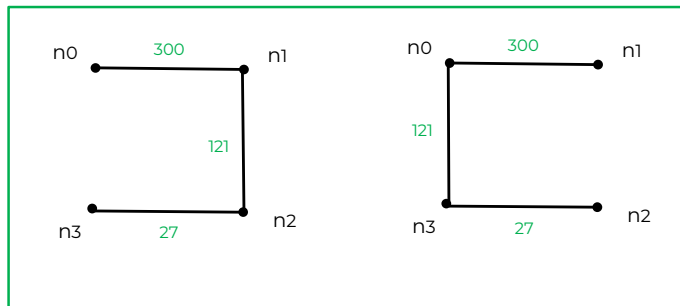


Figure 1.1: Two weighted undirected networks. They are different in a labelled node setting and the same in an unlabelled node setting if $n_0 \rightarrow n_1, n_2 \rightarrow n_3$ are matched.

Graph Space is an example of *natural* embedding space (Marron and Alonso, 2014), which is a space where each point is a potential meaningful network - in other words the embedding is bijective. It is not always necessary to study networks (or complex object in general) in a natural space. In fact, there is no requirement of interpolation between the data points for some statistical methods, such as the classification of population of networks, the clustering, and the regression of a scalar from a population of networks. Methods such as graph kernels (Shervashidze et al., 2011; Vishwanathan et al., 2010) or graph neural networks (Xu et al., 2018; Zhang et al., 2018) reduce each network to fixed-size feature vector and study the feature vectors in an Euclidean setting, recycling well-known

Chapter 1. General Introduction

multivariate statistical tools. When the aim of the analysis is to interpolate between networks (computing principal components to project the data on, estimating a regression line of networks), a natural embedding such as Graph Space is required (see Figure 1.2 for a conceptual representation). For different complex data rather than networks, building statistical models in a natural embedding space has been explored. Examples are studies in Functional Data Analysis (Ramsay and Silverman, 2007), Shape Analysis (Kendall, 1984), Population of Trees (Billera et al., 2001; Wang and Marron, 2007) - see Marron and Alonso (2014) for a review of different data type and methods. Networks are an example of strongly non-Euclidean data type. Developing statistical tools for the analysis of strongly non-Euclidean data is an essential research question, stimulating interdisciplinary research, breaking boundaries between statistics, geometry, and computer science.



Figure 1.2: *Predicting classes or real-valued regression outputs from graphs is "easy" – these problems can be handled by embedding network data into Euclidean feature space while only maintaining the structure of the output variable. Network-valued interpolation are far more challenging, as they require being able to interpolate between networks. Used by permission of A. Feragen*

To the best of our knowledge, very few researches have been conducted in the direction of building intrinsic methods for the analysis of population of unlabelled networks. This PhD thesis has the aim to start from the basis, describing a natural embedding for networks and building gradually statistical tools for the exploration and the prediction of networks within this embedding, proposing some solutions and stimulating new research questions related to this extremely complex and interesting topic.

1.2 Urban Movement

Human Individual Movement is defined as "how individual humans move within a network or a system" (Keyfitz, 1973). When the system or the network is a urban environment, we talk about *urban individual movement*. Understanding how individuals move in a specific environment is an intrinsically valuable research question, potentially revealing key information for planners, policy makers, landscape architects, urban designers, sociologists etc.

(Gonzalez et al., 2008; Toole et al., 2012; Wang et al., 2012). The interest is not only in urban studies, but also in other disciplines such as disease spread (Colizza et al., 2007), or socio-economic development (Pappalardo et al., 2015). Human individual movement can be studied at a different scale in space, according to the selected scale of the system. In the following, we focus on three possible level of analysis: micro-scale (or human scale), meso-scale (or urban scale), and macro-scale (or territorial scale). Scale is a continuous parameter, this interval separation is subjective and changes from case to case.

Focusing on the micro-scale (i.e. human-scale Giampieri et al. (2017)), urban movement can be framed in the concept of "Public Life", referring to citizens' "daily interactions with others within the built environment" (Gehl, 2011). Focusing on the urban movement in a square, a park, or a pedestrian cross, scholars can understand the quality of building environment, providing planners and policy makers an interesting insight on the effective usage of the public space (Gehl and Svarre, 2013; Giampieri et al., 2017; Riva et al., 2019; Whyte et al., 1980).



Figure 1.3: *A crossroad showing pedestrian interacting with the building environment and infrastructure (Gehl, 2011).*

Focusing on the meso-scale (i.e. a infra-urban area mobility), human movement can reveal mobility motifs and patterns, useful to transport modelling, sustainable urban studies, policy making (Lord et al., 2011; Ratti et al., 2006). The scale could even be a macro-scale (Brenner, 2000, 2014; Brunet, 1989). In the macro-scale system, individual mobility could reveal cosmopolitan and globalization behaviours (Beck, 2017), as well as new inter-regional patterns (Jensen and Richardson, 2004). Within this framework, studying movement on a different scale can help planners to interpret different characteristics of the urban environment, as for example the relationship between its layout and its usage, or

Chapter 1. General Introduction

the design features that influence human perception and decision making in such a realm. Note that, individual movement does not have to be limited to the analysis of the spatial movement of humans: it can refer to the whole related feature of these movement as well as its consequences. For example, one can be interested in the analysis of geo-localized opinions expressed by users on social media (Riva et al., 2019).

The analysis of the urban movement in the urban space can be conducted in a different methodological way. The problem can be tackle using a qualitative (e.g. Lord et al. (2011)), a quantitative (e.g. Gonzalez et al. (2008)), or mixed approach (e.g. Goetz et al. (2009)). It is not the aim of this section to open a discussion about whether it is better to use one approach or the other (we suggest Goetz et al. (2009); Røe (2000) for a discussions about this topic). In the past, many different quantitative tools have been developed to study urban movement - among all, we focus on statistics for two main reasons. The first one is the increasing interest in data analysis due to the augmented data gathering and storage. The second reason is related with the traditional role of statistics, as a mathematical field designed to study complex phenomena, not easily described by physics. Urban movement in fact is linked to the cognitive perception of space and the psychology of human beings, making this problem far from being a pure physical phenomenon.

Statistical Analysis of Urban Movement

In this section, we focus the statistical analysis of urban movement. A first part is dedicated to describe the increasing attention to data analysis and the increasing faith in the information carried by data. Then, we introduce the concept of "complex systems" and the related concept of "reductionism". Urban movement is an example of a complex phenomenon which can be analysed using data. To avoid reductionism in this analysis, we discuss the choice of geometry when we analyse a spatial problem. Arguing that the spatial analysis - such as the analysis of the movement - has mostly been conducted in the Euclidean setting, we discuss the reductionism of this choice. To introduce the reader to the choice of the geometrical framework, we briefly introduce Euclidean and non-Euclidean geometry and we summarize Nagel's discussion about geometry beyond classical space measurement tools. Even if apparently uncorrelated, the choice of a proper geometry is a first step to avoid the reductionism in measuring and analysing a complex phenomenon as urban movement.

Digital Positivism

Following the Google mantra "nobody is as smart as everybody" (Wyly, 2014), the last decade has been characterized by an increasing amount of stored data about many different aspects of life, including of course urban movement. Google is only one of the thousand of companies collecting everyday data. As stated by Wyly (2014), this data flow gives rise to a new era of positivism, also called digital positivism (Fuchs, 2017; Fuchs

and Mosco, 2015; Mosco, 2017). This era causes the growing of an unconditional faith in knowledge retrieval through data, with extremist statement as the one published on Wired by Chris Andersen: "Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves." (Anderson, 2008). A part from Anderson's extreme perspective on the topic, the usage of data has been becoming a paradigm in many different fields, including urban studies. These disciplines have been strongly influenced and reshaped by the digital positivism flow (Graham and Shelton, 2013). In particular, the analysis of urban movement has been studied under a data-driven perspective by enthusiastic scholars (e.g. Pucci et al. (2015); Ratti et al. (2006); Reades et al. (2007)), showing the power of data usage in the sector.

By contrast, a new form of post-positivism has been growing as a side-effects of the digital positivism (Goetz et al., 2009; Graham and Shelton, 2013; Wyly, 2014). For example, Graham and Shelton (2013) argue that "an increasing reliance on data in the political process opens up worries of ceding politics itself to data, tools, and machines. Allowing data and algorithms alone to determine policy can never be a value-neutral or depoliticized process...". The worrisome concerning political institutions and data driven policies is also expressed in Wyly (2014), under the criticism of the *Positivist City Hall*. Other critiques are related to the auto-referential loop and bias increase in the collection of data shadow (i.e. the data traces we unconsciously leave while doing our daily activities). "Not only there are always highly uneven data shadows, and not only do those data shadows have effects in the world, but the powerful links between the big data practices and the big data meme itself will only reinforce those very issues. Material and digital ghettoization and Balkanization are produced in a recursive relationship. Such out comes are only amplified by the popularity of the big data meme among technologists, journalists, and venture capitalists. As the data shadows of some people, places, and processes become increasingly dense and well defined, the utility of those data increases, reinforcing the alternatively virtuous or vicious cycles of data production, consumption, and usage." (Graham, 2014; Graham and Shelton, 2013). This last criticism is related to a more general political and philosophical critique to the positivism research paradigm, as a status-quo preserving approach to reality (Fuchs, 2017; Fuchs and Mosco, 2015). Note that the discussion about digital positivism expands far behind the academic circle, see for example O'neil (2016) and the 2020 documentary "The Social Dilemma" on Netflix.

Digital positivism in the urban context is related to *smart city* (Ratti et al., 2006). Smart city is a broad concept, focusing on many different aspect of the urban environment, such as data driven planning, transport optimization, and social welfare analysis. As to the analysis of urban movement and perception, the usage of digital traces has been boosted in the last decades. Digital traces offer a unique source of data to track perception and movement. However, digital positivism in the big data era has been strongly criticised.

Chapter 1. General Introduction

The critiques are both ontological and technical. Among the ontological critiques, digital positivism is related for example to a new form of "surveillance capitalism" (Zuboff, 2015). The technical critiques regard the effective quality of the data collected as well as the effective true knowledge gained from the analysis of this amount of data. Apart from all the possible critiques, it is a fact that these data are a unique occasion to understand urban movement. Urban movement is a perfect example of a problem hard to analyse with official and static data. However, to have all the information is only the first step of the analysis - and surely not the hardest. Complex phenomena generate complex data and require complex methodology.

Complexity and Reductionism in Statistics

As explained in Manson and O'Sullivan (2006), complexity theory can play an important role in "space-and-place-based studies". Complexity theory is a "complex" field, embracing different disciplines, different definitions, and different solutions. Defining complexity is not an easy task (see for an overview on the subject Richardson et al. (2001)). Here we refer to three type of complexity as described in Manson and O'Sullivan (2006):

- *aggregate complexity*, which is the study of how individual elements working in concert create complex systems which have internal structure relative to a surrounding environment.
- *algorithmic complexity*, from mathematical complexity theory and information theory, which contends that the complexity of a system resides in the difficulty of describing system characteristics.
- *deterministic complexity* attempts to simplify some classes of dynamic systems with the aid of chaos theory and, to a lesser extent, catastrophe theory.

Different complexity types affect the study of urban movement within space. In particular, scholars attempt to study the aggregation of individuals in urban landscape (aggregate complexity) by using mathematical models (algorithmic complexity). The quantitative description of this complex system can be conducted using different mathematical tools. Possible mathematical frameworks are partial differential equation (e.g. Colombo (2002)), stochastic equations (e.g. Boel and Mihaylova (2006)), and statistics (e.g. Secchi et al. (2015)). As stated in the previous paragraph, a mass of complex data concerning many different phenomena is sampled every second. Researchers on how to analyse in a meaningful way this source of data are required in every field. Among all the mathematical techniques, we are going to focus on the statistical approach to complex systems, where phenomena are studied via analysis of data collected from the system. The algorithmic complexity is declined in the statistical modelling of a complex system. In complex system analysis, a straightforward "sin" can be reductionism (Manson and O'Sullivan,

2006). There is a thin line between the holistic attempt to describe a complex system, the emergence phenomenon, and the reductionism of the analysis conducted. As well as complexity, reductionism is another complicated philosophical concept. Here we are referring to Honderich (2005) for a definition:

- Ontological reductionism: a belief that the whole of reality consists of a minimal number of parts.
- Methodological reductionism: the scientific attempt to provide explanation in terms of ever smaller entities.
- Theory reductionism: the suggestion that a newer theory does not replace or absorb an older one, but reduces it to more basic terms.

In this discussion, we are focusing on Methodological reductionism. Methodological reductionism has a deep root in the statistical field and it is strongly related with the concept of *sufficiency*. In the following, we decline both complexity and reductionism in the statistics.

Complexity theory has a long-time relationship with mathematics. One of the most important players of this discussion is the so-called "Chaos and Complexity theory" (see for example Cambel (1993)). We are interested in interpreting the complexity concept within the statistical framework. The statistical approach is dealing with phenomena that can not be physically described due to their intrinsic complexity. We report here a simple but enlightening example: "Probability [and Statistics] deals with the study and mathematical formalization of "random" phenomena, i.e. phenomena we cannot *a priori* predict the outcome of. The reasons why it is not possible to give a deterministic description of certain phenomenon are multiple. We could have incomplete information regarding the phenomenon, or there might be no theory to draw consequences, or the theory exists but it is difficult to apply, or it can simply happen that the phenomenon is really "random". Let's think about throwing a coin. The motion of a rigid body in space (e.g. the coin) is well described by the equations of Newtonian mechanics. Thus, we could theoretically calculate if the coin will fall with upper face as heads or tails, if we are able to take into account the initial speed, the air friction, and the inelastic collisions that the coin undergoes when it falls back to the ground. However, a real computation of these elements is impossible, both because it is not possible to experimentally measure the physical quantities involved, and because the system exhibits a sensitive dependence on the initial conditions. One small variation of the initial conditions (e.g. the force applied in the throwing or position from which it is thrown) leads to a remarkable macroscopic effect (e.g. the output head rather than cross). It is clear that if the currency is sufficiently symmetric we expect that the "possibility" of head is the same as cross. Hence the need to model this phenomenon through a theory different from Newtonian mechanics." (Epifani et al., 2005)

Chapter 1. General Introduction

A statistical model! In this very simple example, we can see how probability and statistics are the tools to describe complexity. When dealing with a complex system, the first basic question a statistician ask her/him self is: "What is a meaningful statistical unit to describe this complex system?". Historically, the statistical unit is a number (e.g. the height, the length, the weight). A set of indexes are built to describe the complex system and the effort is made in building a sufficiently complex statistical model, to describe complex research questions rising from a complex system. As already discussed, many scholars have started talking about *complex data* (Marron and Alonso, 2014; Secchi et al., 2015), trying to focus more on a complex representation of the system than on an hyper complicated modelling of "simple" data.

From 2011, when the word *big data* introduced by the McKinsey Global Institute has started spreading in both industry and research environment, some scholars start interpreting this stream of data not only in terms of velocity, variety, and volume, but also in terms of complexity (Marron and Alonso, 2014). Along with networks, complex data are for example texts, images, functions, videos. They are collected in different context, such as bio statistics (e.g. brain scan MRI (Wang et al., 2010)), social media analysis (e.g. images posted (Hu et al., 2014)), transportation (e.g. functional mobility data (Vantini et al., 2012)). The analysis of complex data poses serious methodological challenges. The complexity resides not only in the intrinsic nature of the data but also in methodological complexity required to analyse those data.

As previously said, methodological reductionism has a deep root in the statistical field and it is strongly related with the concept of *sufficiency*. Given a statistical model and its unknown parameter, a statistic is *sufficient* if "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter" (Fisher, 1922). In other terms, if we are interested in describing a specific phenomenon with certain summarising statistics, the estimation of these statistics is going to be exhaustive, if we represent the phenomenon with a sufficient amount of information. For example, if we are interested in how people moves around a square, sampling one position for each individual every day is probably an under-representation of the phenomenon, and measuring the microsecond movements with sensors on the legs of each individuals is an over-representation of the phenomenon. The representation choice of a problem is a key and a first step in a quantitative analysis, because it frames the problem and the analysis in the "space".

Both complexity and reductionism in statistics are related with geometry. Which is the correct geometry to use when studying a complex phenomenon? Which is the complex datum that represents a sufficient representation for the complexity of the phenomenon? Which is the geometry to study the complex data collected?



Figure 1.4: *Ernest Nagel*

Geometry, Complex Data, and Sufficient Representation

We are now going to introduce the concept of geometry, its history and its latest development mainly following Ernest Nagel perspective on the topic. Ernest Nagel (Nove Mesto, 1901 - New York, 1985) was a Bohemian philosopher. His work concerned the philosophy of mathematical fields such as geometry and probability, quantum mechanics, and the status of reductive and inductive theories of science. For an interesting overview and biography see Suppes (1994). In particular, Nagel's milestone book *The structure of science* (Nagel, 1961) offers an interesting argument about the usage of geometry in measuring space.

The history of measuring the space

The notion of how to measure space is thousands of years old. The first rudimentary ideas about the measurement of space dates back in Egypt, where ancient Egyptian built up a first geometrical framework to measure the dimension, the length, the area, the shape of a field. However, the theoretical development of this geometric theory as a set of axioms and theorems was formalized by Euclid in the IV century B.C.. The Euclidean geometry is based on a set of axioms:

- (I) To draw a straight line from any point to any point.
- (II) To produce (extend) a finite straight line continuously in a straight line.
- (III) To describe a circle with any centre and distance (radius).
- (IV) That all right angles are equal to one another.

Chapter 1. General Introduction

- (V) That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

Moving from these five axioms, the whole so called Euclidean geometry was built, studied by scholars, taught at school from first to last grade. In thousands of years, scholars used the Euclidean geometry as a "comfortable" mathematical framework to measure space and shape. Nagel started its discussion about geometry, reminding us that geometry was broadly applied as a base system, not only in the agro-measurement but also in astronomy, architecture, crafting, engineering, and fine arts. Within this context, scientist as Newton considered geometry as an applied sub-field of physics. Euclidean geometry was interpreted as an *a priori* system for measuring reality. Nagel defined two different geometry: *pure geometry* and *applied geometry*. Pure geometry studies the deductive results obtained by the axioms, while applied geometry tries to find a factual truth within the axioms and theorems. Focusing on applied geometry, there are two approaches: (1) objects such as surface, point, shapes are identified in reality and then prove that these configuration satisfied the Euclid's axioms within a certain measurement error bound; (2) Euclidean definitions and axioms are stated *a priori* and the real representation of the definition are searched in reality. Within the first approach, for a certain number of reality configuration, Euclidean geometry is very useful! In the second setting instead, it is highly probable that the observed figures are not so easily embedded within the Euclidean framework. This discussion is not proving the invalidity of the Euclidean geometry, but it shows the potential drawbacks in using the Euclidean framework.

Both the previous argument and the uncountable trials of proving the (V) axiom from the others were some of the reasons of the development of new form of geometry, often called non-Euclidean geometries. Here below, a short list and description of some of the most well known non-Euclidean geometries. For a detailed and technical introduction to non-Euclidean geometry we suggest: Coxeter (1998); Krause (1986); Lee (2013).

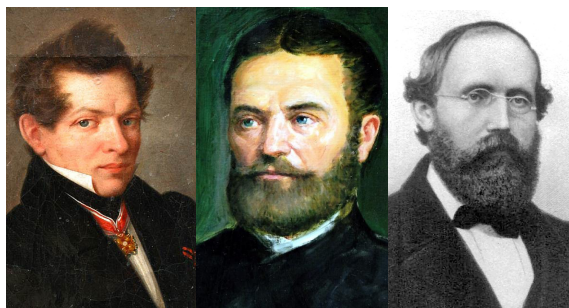


Figure 1.5: Lobachevsky, Bolyai, and Riemann

During the 19th century, two mathematicians, Lobačevskij (Novgorod, 1792 – Kazan’,

1856) and Bolyai (Kolozsvár, 1802 – Târgu Mureș, 1860), propose a new geometry asserting the negation of the (V) Euclidean's axiom (in other words, they state that given one line and one point, there is at least one line crossing the point and parallel to the line). From the negation of the axiom, a new geometry usually called *hyperbolic geometry* has been developed. In Figure 1.6, you can see parallel lines to a given line, where the space in exams

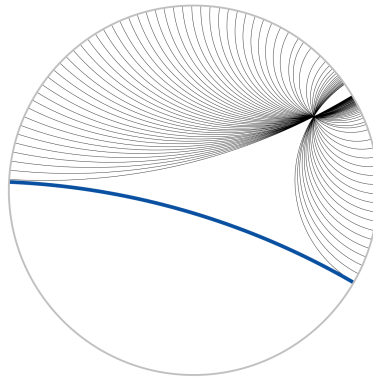


Figure 1.6: *Hyperbolic geometry. An example of parallel "lines" to a given point and a give "line".*

are all the inner point of a circle. Another geometry was introduced by the German mathematician Georg Riemann (Breselenz, 1826 – Selasca, 1866). Riemann substituted the (V) axiom with the following: "for a given point out of a given line, there is no parallel line crossing that point". This geometry called "riemannian" or elliptic geometry was developed after the curvature of the space concept. A last but not least non-Euclidean geometry we would like to introduce is the *projective geometry* (Casse and Casse Jr., 2006). This geometry does not focus on the concept of congruence, but on the concept of projection, defining invariant properties with respect to projection.

Which geometry when?

After this brief (and partial) discussion about new forms of geometry, Newton's (and many colleagues) approach to geometry as a practical tool to measure reality can not be taken for granted. Nagel noticed that the required geometry is a consequence of the rules adopted to measure space. This is clear for example in the way Riemann introduced his geometrical framework. In the nineteenth century, scholars started using the novel geometrical framework introduced to describe certain type of phenomena. A well known example is the relativity theory by Einstein. All his theory is based on the assumption that the space is an hyperbolic space, and all the computations are performed according to this assumption. As shown by this "successful" example, all the geometries can be adopted to measure space and phenomena with space dependency. The question is: given a problem and a set

Chapter 1. General Introduction

of properties and relations we want to analyse, which system is more convenient? In the analysis of urban movement, is the Euclidean setting always a good choice?

As described in Section 1.2, contemporary complex data have characteristics that are not always easily described with a simple Euclidean setting. For example, consider the subway system. The subway system can be described as the location of the stations. Analysis of the location of the station can be easily conducted by using Euclidean setting (i.e. the real space distance and operations). However, in the analysis of a subway system, someone could be interested in analysing also the topology of the network, the daily flow of passengers from one station to another, the prices of the tickets for each path on the network etc.. In this case, how can we mathematically describe all the information? The most common approaches summarize all the complexity with a list of numbers, trying to reduce the problem to the Euclidean setting. The following question arises: is it a sufficient representation? Is a vector of indexes enough to describe the topology and the dynamic of the subway? It depends of course on the research question. Shrinking such a complex system in an Euclidean setting is maybe useful, but sometimes unnatural and reductive. As already discussed in the Introduction to Population of Networks, OODA and statistical analysis of complex data in a natural embedding aim to describe the complexity, overcoming the reductionism, and searching for a meaningful sufficient representations of reality.

1.3 Populations of Graphs and Urban Movement

Urban movement has been successfully modelled within the framework of networks (or graphs). This because it is "characterized by a sequence of visited locations and the trips between them" Schneider et al. (2013), showing an intrinsic relational nature, which can be easily quantified using a network at every scale of the analysis. The same relational structure can be found in the public transport system, or in road infrastructure. Due to the appropriateness of this representation, the application of network analysis in transportation modelling has also been showing recent feasibility. Some examples of this are the analysis of public transport networks (von Ferber et al., 2009), urban structures (Zhong et al., 2014), and daily urban mobility motifs (Schneider et al., 2013). As in many other field, the application of the network analysis to urban movement has only followed a first generation approach. The second generation approach has been largely unexplored due to the lack of a robust methodological framework.

To describe the potential application of the second generation approach, consider the following example. Among the possible urban movement data type, we focus on mobility motif, which is the analysis of a single person mobility for a sampled day (Schneider

1.3. Populations of Graphs and Urban Movement



Figure 1.7: *Two schemes of two inhabitants' mobility motifs in Maputo, Mozambique. The two maps show the positions of the houses while the two graphs show the mobility motifs of all the barrios the inhabitants travel to in a certain day. The data are the two graphs.*

et al., 2013), moving between different places (office, shopping area, pub, cinema, ect.). The analysis of mobility motifs is complicated, because every inhabitant lives and works in different areas, and she/he crosses different neighbourhoods every day. However, a group of citizens could have a very similar mobility motif. Supposing we can study mobility motifs via GPS traces, we represent them as graphs, where the nodes are the places where the inhabitant is stopping for at least a certain amount of time and edges are the paths between them. How can you compare two mobility motifs that have different stopping locations (nodes)? How can we understand which mobility motifs are similar? To make these comparisons, we should be able to change from the geographical and physical embedding to a mathematical and relational embedding. We should change the geometry! This mathematical embedding should take into account the shape of the mobility motif and the attributes of the observation such as the speed, or the mobility mode used, while ignoring the actual geographical position of the nodes. With an object oriented data analysis embedding in a non Euclidean space, we are able to compare in a complete and meaningful way the mobility motifs. An example is the non Euclidean space to describe graphs that we are studying in this thesis. As shown in Figure 1.7, if we had maintained the geographical dependence of the mobility motif, we would have not been able to compare the two mobility motifs. Note that there are other mathematical representations of the mobility motifs, such as the trajectories representations. This simple example poses the basis (and hopefully stimulates the readers' interest) about which is the natural geometry to describe a spatial phenomenon. Following some of the concept introduced in this Section, we will show a possible path towards an original analysis of the urban movements.

CHAPTER 2

Embedding in the Graph Space

The aim of this chapter is to introduce a geometrical embedding for population of graphs, namely *Graph Space*, on top of which we are going to develop different statistical methods. Graph Space was firstly introduced in Jain and Obermayer (2009) in a more general framework. In this chapter, it is declined in the context of network analysis and studied in its geometrical properties.

As already discussed in Section 1.1, if the objective is a "simple" tasks such as the prediction of a class labels or a scalar values, or the cluster memberships, Euclidean graph embedding is often a scalable and powerful choice. See for example kernel methods (Sherwashidze et al., 2011), convolutional neural networks (Duvenaud et al., 2015), or feature selection algorithms (Bunke and Riesen, 2011). However, the Euclidean graph embedding methods do not ensure that every point in the embedding space is actually a network, and the probabilistic models therefore typically assign nonzero probability to points that do not represent networks, causing problems in extending statistical tools to these embedding spaces. As a response to these limitations, the natural embedding aims to model "objects" as residing in a space of precisely such objects - ensuring that every point in the embedding space is a meaningful object. However, most existing works in this category have so far focused on tree-valued data rather than graphs in general. Nonlinear data spaces such as

Chapter 2. Embedding in the Graph Space

the Billera-Holmes-Vogtmann space of leaf-labeled trees (Billera et al., 2001), or spaces of unlabeled trees (Feragen et al., 2010, 2013) are examples of non-Euclidean spaces in which tree-data can be analysed with geometric tools. The literature precisely concerning the natural embedding of networks is still very new and varied. For example, networks' Laplacian matrices are smoothly injected into a sub-manifold of a Euclidean space in Ginestet et al. (2017); a Bayesian generative model is defined for a set of network-valued data represented as adjacency matrices in Durante et al. (2017); or each network is represented as a metric space in Chowdhury and Mémoli (2018) and Lee et al. (2017). In the latter embedding context, unlabelled networks are also analysed, defining matching strategies as the one proposed in Chowdhury and Mémoli (2017). Even if the metric space representation is a very promising and flexible choice, there is still very little statistical tools developed for spaces of metric spaces (see Sturm (2003) for an introduction).

Introduced as a general space of "Structures" by Jain and Obermayer (2009), Graph Space is a quotient space of the Euclidean Space, obtained by applying permutation action to adjacency matrices. Graph Space is a natural space for graphs with different or equal number of nodes, and with labelled or unlabelled nodes, able to describe weighted or unweighted, uni- and multi-layer, directed and undirected networks. The same Graph Space appears in Kolaczyk et al. (2020), which studies the behavior of Fréchet mean, as well as in Guo et al. (2019), which proposes an extrinsic algorithm for principal components. The choice of focusing on Graph Space rather than other embedding strategies is guided by two main reasons. First, its relation with Euclidean space as total space makes Graph Space as an interesting starting point to extend the well known statistical tools available in the Euclidean setting. Second, its flexibility allows for description of different networks, covering many different application problem including the statistical analysis of human movement.

This chapter is organized as follows: After a general introduction to Graph Space and its notation, the nonlinear nature of the geometry of the space is studied, showing that it is a geodesic space, but not a manifold (Section 2.1). In Section 2.2, we prove that its curvature is unbounded from above, which renders standard approaches to nonlinear statistics unpractical, at least in terms of proving their convergence.

This chapter contains the contents developed in Calissano et al. (2020a)

2.1 The Graph Space

We consider graphs as triples $G = (V, E, a)$ where the node set V has at most n elements, and the edge set $E \subset V^2$ has maximal size n^2 . The nodes and edges are attributed with elements of an attribute space A , which in this thesis is assumed to be Euclidean, via an attribute map $a: E \rightarrow A$. Here, the map a allows us to describe attributes on both edges and nodes, as we use self loop edges (diagonal elements) to assign attributes to nodes. From here on, we represent networks mathematically as graphs, and consider these terms equivalent. In our modelling, we shall represent graphs with fewer nodes than n as having $n - |V|$ additional null nodes, allowing graphs to be represented via fixed-size adjacency matrices. More precisely, a graph with scalar attributes is completely specified by the adjacency matrix of dimension $n \times n$, residing in a space $X = \mathbb{R}^{n^2}$ of flattened adjacency matrices. If the attributes are vectors of dimension d , the graph is represented by a tensor of dimension $n \times n \times d$, residing in a space $X = \mathbb{R}^{n \times n \times d}$.

In many real world applications, populations of graphs describe the same type of relational phenomena in different contexts (e.g. routes of different airline companies, or brain connectivity networks of different patients). Differences in the labels or order of nodes make it challenging to investigate similarities between the topology and attributes of different graphs, and this is often alleviated by explicit or implicit matching of graph nodes. When the graphs are represented as $n \times n$ adjacency matrices, matching two graphs corresponds to finding optimal permutations of their nodes. The group T of node permutations can be represented via permutation matrices, acting on X through matrix multiplication. The binary operation:

$$\cdot: T \times X \rightarrow X, (T, x) \mapsto Tx$$

thus defines an action of the group T on X . We call the obtained quotient space $X_T = X/T$ *Graph Space*, and each element of X/T is a graph G , represented as an equivalence class $[x] = Tx$ which contains all the flattened adjacency matrices in X which can be obtained from x by permuting nodes. Note that this Graph Space is a special case of the A -attributed r -structures introduced by Jain and Obermayer (2009), which includes hyper-graphs and more general attribute types. The graphs considered in this thesis can be considered A -attributed 2-structures with A a Euclidean space. We illustrate the construction of Graph Space with a simple example:

Example. Consider the two weighted networks shown in Figure 2.1. To represent these as points in Graph Space: Add a fictional null node to the first graph; randomly enumerate the nodes; represent them in two weighted, symmetric adjacency matrices as shown in Figure 2.1. The adjacency matrices can be vectorized as a vector of dimension nine (e.g the

Chapter 2. Embedding in the Graph Space

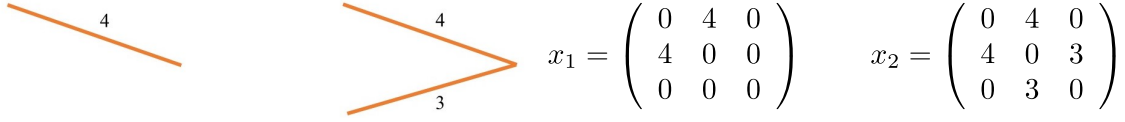


Figure 2.1: An example of two simple weighted undirected networks x_1 and x_2 and their matrix representations.

first network becomes $[0, 4, 0, 4, 0, 0, 0, 0, 0]$). The space of flattened adjacency matrices X is thus \mathbb{R}^9 , and the two networks are represented by two points in \mathbb{R}^9 . The permutation action can be represented as a binary 9×9 matrix: The permutation reorders the nodes and consequently the edges of the network by permuting the rows and columns of its adjacency matrix, and hence also the positions of the elements in the flattened matrix representation x_i . For example, if the nodes 2 and 3 are permuted in the first network, we obtain the new permuted vector $[0, 0, 4, 0, 0, 0, 0, 4, 0]$. Each point $[x_i]$ in the quotient space X/T consists of all possible permuted versions of x_i (i.e. permuting the rows and columns of the associated adjacency matrix). Now, the equivalence classes $[x_1]$ and $[x_2]$ are points in X/T , and the maximal size of an equivalence class in X/T is $3!$.

Remark. In practice, the maximum number of nodes n can limit the possible node matchings, and hence also the geometric relationship, between two graphs. To make sure no such limitations are present, set $n = 2 \max(|n_i|)$.

For symmetric adjacency matrices it would suffice to represent them with the upper triangular part. For the sake of generalization we keep the more complete representation.

2.2 Geometrical Characterization of the Graph Space

The aim of this section is to explore the geometrical properties of the Graph Space to understand how statistical estimators can be computed in Graph Space.

Graph Space is not a manifold

As Graph Space X/T is the quotient of the total space X of flattened adjacency matrices with respect to the node permutation group T , any metric d_X on X defines a quotient pseudo-metric

$$d_{X/T}([x_1], [x_2]) = \min_{t \in T} d_X(tx_1, x_2)$$

2.2. Geometrical Characterization of the Graph Space

on X/T which, since the permutation group T is finite, is indeed a metric. Examples of commonly used metrics on X include the l_p metrics for $p \in (0, \infty]$, where $p = 2$ gives the Euclidean distance used in this thesis, and $p = 1$ gives the Manhattan distance.

While Graph Space X/T is a metric space, it is not a manifold, even with the Euclidean distance on X . This follows from the fact that the structure of the isotropy subgroup $T_x = \{t \in T | tx = x\}$ varies for different points $x \in X$ (Bredon, 1972). One reason why this happens is that, as explained in the previous section, forcing the networks to all have the same number of nodes generates networks with a subset of null nodes. As a permutation $t \in T$ which acts only on the null nodes of a flattened adjacency matrix $x \in X$ does not have any effect on the adjacency matrix, such x will have a larger isotropy subgroup than generic points in X . As a consequence, many well-known tools from *manifold statistics* (Fletcher, 2013; Fletcher and Joshi, 2004; Huckemann et al., 2010; Kendall, 1984; Mallasto and Feragen, 2018; Pennec et al., 2006; Srivastava et al., 2005; Zhang and Fletcher, 2013) are unfortunately not directly applicable to Graph Space. However, as is commonly done in both general nonlinear statistics (Bačák, 2014; Duncan et al., 2018; Feragen and Nye, 2020; Feragen et al., 2011, 2013; Miller et al., 2015; Nye, 2011, 2014; Nye et al., 2017; Sturm, 2003; Turner et al., 2014) and in manifold statistics, we will utilize *geodesics*, or shortest paths, to define and compute statistical properties in Graph Space. In the absence of a manifold structure, we will define and understand geodesics, statistical properties built on geodesics, and their properties, by utilizing geometric constructions from *metric geometry* (Bridson and Haefliger, 1999). To that end, we dedicate this section to surveying necessary concepts from metric geometry and applying them to uncover geometric properties of Graph Space.

Graph Space is a geodesic space

Given a general metric space $(\mathfrak{X}, d_{\mathfrak{X}})$, the *length* of a path $\gamma: [0, 1] \rightarrow \mathfrak{X}$ is given by

$$l(\gamma) = \sup \left\{ \sum_{i=1}^m d_X(x_{i-1}, x_i) \mid x_0 = \gamma(0), x_1, \dots, x_m = \gamma(1) \text{ for some } m \in \mathbb{N} \right\} \quad (2.1)$$

where the supremum is taken over all approximations $x_0 = \gamma(0), x_1 = \gamma(t_1), \dots, x_{m-1} = \gamma(t_{m-1}), x_m = \gamma(1)$ of γ of arbitrary finite length m , where $0 < t_1 < \dots < t_{m-1} < 1$. Thus, the length of a path can be thought of as the supremum over lengths of all finite approximations of the path.

Given two points $a, b \in \mathfrak{X}$, a *geodesic* from a to b is a path $\gamma: [0, 1] \rightarrow \mathfrak{X}$ such that $\gamma(0) = a, \gamma(1) = b$ and $l(\gamma) = d_{\mathfrak{X}}(a, b)$. The metric space $(\mathfrak{X}, d_{\mathfrak{X}})$ is said to be a *length space* if, for every two points $a, b \in \mathfrak{X}$, we have

$$d_{\mathfrak{X}}(a, b) = \inf \{ l(\gamma) \mid \gamma: [0, 1] \rightarrow \mathfrak{X} \text{ s.t. } \gamma(0) = a, \gamma(1) = b \}.$$

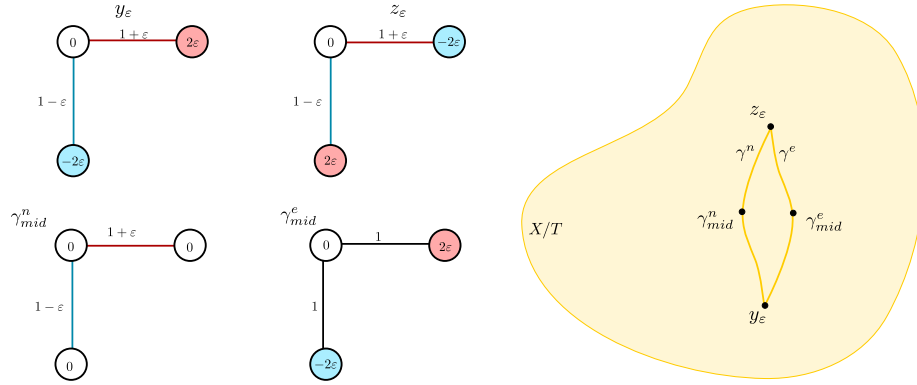


Figure 2.2: Depending on their node alignment, the two graphs y_ε and z_ε differ either in node attributes or in edge attributes, and there exist two different geodesics γ^n and γ^e connecting the two graphs: One which interpolates node attributes and one which interpolates edge attributes, respectively. These two geodesics are illustrated via their midpoints γ_{mid}^n and γ_{mid}^e .

That is, the distance between any two points a, b is the infimum over lengths of paths connecting them. Moreover, \mathfrak{X} is a *geodesic space* if every two points $a, b \in \mathfrak{X}$ are connected by a geodesic γ from a to b – that is, there actually exists a path attaining the infimum length.

Lemma 2.2.1. *Graph Space is a geodesic space.*

Proof. This result follows from standard properties of metric spaces. As our total space X is Euclidean, it is in particular a length space. Since Graph Space X/T is a metric space, X/T is a length space by (Bridson and Haefliger, 1999, Lemma I.5.20). Moreover, as X/T is the quotient with respect to a finite group, and X is locally compact, the quotient X/T is also locally compact (Bredon, 1972, Theorem I.3.1). Note that any Cauchy sequence $([x_i])_{i \in \mathbb{N}}$ in X/T is the image under π of a Cauchy sequence $(t_i x_i)_{i \in \mathbb{N}}$ in X such that for some $M \in \mathbb{N}$, and for $i, j \geq M$, we have $d_X(t_i x_i, t_j x_j) = d_{X/T}([x_i], [x_j])$. Since X is complete, the sequence $(t_i x_i)_{i \in \mathbb{N}}$ converges to some point $x \in X$, and hence the sequence $([x_i])_{i \in \mathbb{N}}$ converges to $[x] \in X/T$. In other words, X/T is complete. Thus, X/T is a geodesic space by the Hopf-Rinow theorem (Bridson and Haefliger, 1999, Proposition I.3.7). \square

Lemma 2.2.2. *There exist points $y_\varepsilon, z_\varepsilon \in X/T$ which are connected by more than one geodesic.*

Proof. We give an example in the case where node- and edge attributes are scalar. Note that the same example can be adapted to vector valued node- or edge attributes by appending this scalar value with zeros. Let $\varepsilon > 0$, and consider the two graphs y_ε and z_ε shown in

2.2. Geometrical Characterization of the Graph Space

the top row on the left hand side of Figure 2.2. There are two geodesic paths between these two graphs. The first consists of interpolating the node attributes with the node matching indicated by the planar embedding of the nodes. The midpoint of this geodesic is the graph γ_{mid}^n shown on the left hand side of the bottom row. The second geodesic between y_ε and z_ε consists of interpolating the edge attributes with the node matching indicated by the coloring of the nodes. The midpoint of this geodesic is the graph γ_{mid}^e shown on the right hand side of the bottom row. The two points connected by two geodesics are illustrated schematically on the right hand side. \square

Corollary 2.2.3. *Geodesics connecting pairs of points are not generally unique in X/T .* \square

Graph Space curvature

Curvature affects properties which are important for defining and computing statistical quantities. In particular, the convergence proofs for common statistical estimation algorithms (Chakraborty and Vemuri, 2015; Miller et al., 2015; Sturm, 2003) often rely on assumptions of bounded curvature. Below, we show that the curvature of Graph Space is unbounded from above, according to generalized concepts of curvature from metric geometry (Bridson and Haefliger, 1999). This will motivate our choice of statistics in the following section.

First consider the Graph Space where edges and nodes have real attributes $a: E \rightarrow \mathbb{R}$. We will show that the curvature of this Graph Space is not bounded from above. In metric geometry, curvature is approached through comparison with *model spaces* M_κ of curvature κ . Different model spaces are used for negative, zero and positive κ . When $\kappa < 0$, the model space is the hyperbolic space of negative curvature κ , namely $M_\kappa = \mathbb{H}_\kappa$. For $\kappa = 0$, the model space is $M_0 = \mathbb{R}^2$, namely the Euclidean plane. Finally, for $\kappa > 0$, the model space is the sphere of curvature κ , namely $M_\kappa = \mathbb{S}_\kappa^2$. An important property of the model spaces of curvature κ is that they are each accompanied by an *diameter* D_κ , such that any two points $a, b \in B(x, D_\kappa)$ for any $x \in M_\kappa$ can be joined by a unique geodesic.

We can compare any given geodesic space $(\mathfrak{X}, d_\mathfrak{X})$ to any one of the model spaces using *comparison triangles*, as follows: A geodesic triangle abc in \mathfrak{X} consists of vertices $a, b, c \in \mathfrak{X}$ joined by geodesic edges γ_{ab} , γ_{bc} and γ_{ac} . We assume that a, b and c are all contained in a ball of perimeter $< 2D_\kappa$. We can then construct a *comparison triangle* $\bar{a}\bar{b}\bar{c}$ in the model space M_κ with vertices \bar{a} , \bar{b} and \bar{c} joined by geodesic edges $\bar{\gamma}_{\bar{a}\bar{b}}$, $\bar{\gamma}_{\bar{b}\bar{c}}$ and $\bar{\gamma}_{\bar{a}\bar{c}}$, whose lengths are the same as the lengths of the edges γ_{ab} , γ_{bc} and γ_{ac} in abc . See Figure 2.3 for an illustration.

Definition 2.2.4 (*CAT*(κ) space, curvature in the sense of Alexandrov). *Let $(\mathfrak{X}, d_\mathfrak{X})$ be a geodesic metric space, and let abc be a geodesic triangle in \mathfrak{X} as described above. Note*

Chapter 2. Embedding in the Graph Space

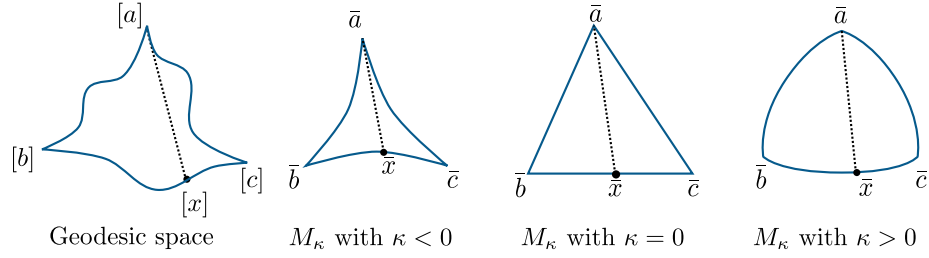


Figure 2.3: *Left:* A geodesic triangle in \mathfrak{X} . *Right:* the corresponding comparison triangles in hyperbolic space \mathbb{H}^2 , the plane \mathbb{R}^2 and the sphere \mathbb{S}^2 , respectively.

that any point x from the segment γ_{bc} has a corresponding point \bar{x} on the segment $\bar{\gamma}_{\bar{b}\bar{c}}$ in the comparison triangle, such that $d_{M_\kappa}(\bar{x}, \bar{b}) = d_{\mathfrak{X}}(x, b)$. If

$$d_{\mathfrak{X}}(x, a) \leq d_{M_\kappa}(\bar{x}, \bar{a}) \quad (2.2)$$

for every such x , and similarly for any x on γ_{ab} or γ_{ac} , then the geodesic triangle abc satisfies the $CAT(\kappa)$ condition. The metric space \mathfrak{X} is a $CAT(\kappa)$ space if any geodesic triangle abc in \mathfrak{X} of perimeter $< 2D_\kappa$ satisfies the $CAT(\kappa)$ condition given in eq. 2.2. Geometrically, this means that triangles in \mathfrak{X} are thinner than triangles in M_κ . The metric space \mathfrak{X} has curvature $\leq \kappa$ in the sense of Alexandrov if it is locally $CAT(\kappa)$.

Note that the properties of the model space in relation to its diameter D_κ also transfer to geodesic spaces \mathfrak{X} which are $CAT(\kappa)$ (Bridson and Haefliger, 1999, Proposition II.1.4): Given any $x \in \mathfrak{X}$, any two points $a, b \in B(x, D_\kappa) \subset \mathfrak{X}$ can be joined by a unique geodesic. In particular, $D_\kappa = \infty$ for $\kappa \leq 0$, meaning that in non-positively curved spaces, any two points can be joined by a unique geodesic, regardless of their distance. Moreover, $D_\kappa \geq \frac{\pi}{\sqrt{\kappa}}$ for $\kappa > 0$, meaning that the lower the bound on the (positive) curvature, the larger the radius within which all pairs of points have unique connecting geodesics.

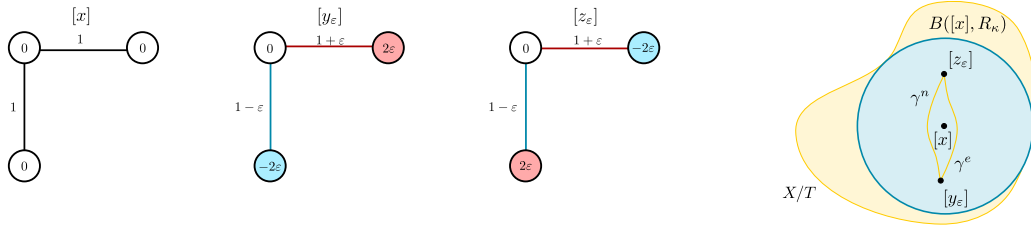


Figure 2.4: Consider the graph $[x] \in X/T$ shown on the left. A ball about $[x]$ of any radius $R_\kappa > 0$ will always contain two graphs $[y_\varepsilon]$ and $[z_\varepsilon]$, as shown second and third. As we saw in Figure 2.2, these two points are connected by two different geodesics.

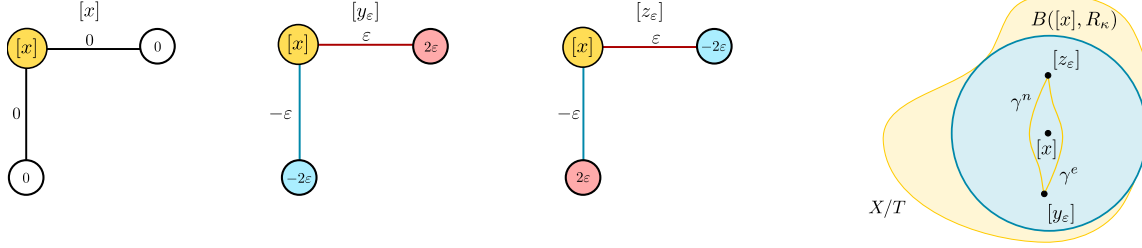


Figure 2.5: For any graph $[x]$, we can find two graphs $[y_\epsilon]$ and $[z_\epsilon]$ arbitrarily close to it, which are connected by more than one geodesic.

Theorem 2.2.5 (The curvature of Graph Space is unbounded from above). *Graph space does not have curvature $\leq \kappa$ in the sense of Alexandrov for any $\kappa \in \mathbb{R}$. In particular, for any graph $[x] \in X/T$ and any $R_\kappa > 0$, we can – by assuming a sufficiently high maximal number of nodes n – find two graphs $[y_\epsilon], [z_\epsilon] \in B([x], R_\kappa) \subset X/T$ which are connected by two geodesics.*

Proof. Let $[x], [y_\epsilon]$ and $[z_\epsilon]$ be as in Figure 2.4. Note that for any radius $R_\kappa > 0$, there will exist an $\epsilon > 0$ such that $[y_\epsilon], [z_\epsilon] \in B([x], R_\kappa)$. Moreover, as argued in Lemma 2.2.3, there will always be two equally long shortest paths connecting $[y_\epsilon]$ and $[z_\epsilon]$. Thus, by Bridson and Haefliger (1999) in Proposition II.1.4, Graph Space is not locally $CAT(\kappa)$ at $[x]$ for any κ , and thus cannot have curvature $\leq \kappa$ in the sense of Alexandrov for any κ . To prove the final statement, consider any graph $[x]$ and assume that the maximal number n of nodes considered in a graph is sufficiently high to construct the graphs $[y_\epsilon]$ and $[z_\epsilon]$ shown in Figure 2.5, which can be constructed to both be arbitrarily close to $[x]$. Now, again, there exist two geodesics connecting $[y_\epsilon]$ and $[z_\epsilon]$. \square

Remark. Note that while all of the results and examples above considered the case where node and edge attributes were real valued ($A = \mathbb{R}$), the proofs hold equally well for vector valued node and edge attributes ($A = \mathbb{R}^p$).

2.3 Discussion

The strong geometrical connotation of this chapter is required to pose the bases for the next chapters, and to acquire knowledge about which techniques are available or not in this context. Even if it is not a manifold and the geodesics are not unique, it is still a very convenient choice in terms of developing statistical methodology, because it is a quotient

Chapter 2. Embedding in the Graph Space

space of an Euclidean space. If a methodology to overcome the curvature issue is developed, statistical analysis can be conducted on the total space, inheriting all the well known statistical tools for multivariate statistics (for example Principal Components - see Section 3.3 in Chapter 3, and Regression - see Section 4.1 in Chapter 4).

A limitation of Graph Space embedding resides in the zero-adding values. In fact, to be able to analyse networks in a space of fixed dimension, all the graphs are grown to have the same number of nodes. This choice is causing several geometrical and applied problems. Adding zero rows and columns to the adjacency matrices is creating equivalence classes of different dimensions (if two null nodes are permuted the network and its representation remains unchanged). This is the main reason for the Graph Space to not be a manifold. Adding zero nodes is also misleading in terms of interpretation. In fact, a non existing node or edge and a node or edge with a zero attribute are considered equivalent in this framework. This choice does not represent a problem in a context of weighted networks, where the attributes describe the intensity of a link (e.g. a molecular bond, a flux, the intensity of a social interaction). However, it could cause problems in other settings, where the difference between an edge with a zero attribute and the absence of the edge itself is fundamental (e.g. the presence of a road on top of which the flux is null rather than the absence of the road itself). There are several approaches to solve this limitation. One is related to the concept of stratified spaces (Feragen et al., 2014). Instead of adding zeros and using Euclidean distance in the biggest space, a different distance can be used between graphs with different dimensions, imaging smaller graphs living in sub-spaces, stratified one into the other.

An interesting further development of Graph Space is related to the exploration of different matching strategies. In this work, the T action is the permutation of all the nodes $|T| = n!$. Different constraints to the permutation action can be required. An example is a permutation of a subset of nodes (i.e. nodes are divided in different groups among which the permutation is allowed), or a non-bijective permutation action (i.e. a node in a graph can be matched with a set of nodes in the other graph). These different permutation actions can also simplify the computational complexity of the problem, exploring only subsets of T .

CHAPTER 3

Exploration in the Graph Space

In this chapter, we introduce exploratory techniques such as Cluster Analysis, Fréchet Mean, and Geodesic Principal Components for the analysis of a population of graphs in the Graph Space.

Even if the Graph Space X/T is not a manifold and its curvature is unbounded from above, we can take advantages of its relation to the Euclidean total space X to define and compute *intrinsic* Graph Space statistics. Given the fact that Graph Space is a metric space, we start this chapter with a case study about clustering. Given the fact that Graph Space is a geodesic space, we study the properties of the Fréchet Mean and we extend tools such as Principal Components. However, we proved in the previous chapter that the geodesics in the Graph Space are non unique everywhere, causing problems in the computation of such statistics. These geometrical characteristics of the space make unavailable the usage of the so called *extrinsic* methods in this context (see Bhattacharya and Patrangenaru (2003) for more details about extrinsic methods, and Patrangenaru and Ellingson (2015) for a general overview). Extrinsic methods - as tangent space statistics (Fletcher and Joshi, 2004) - have been broadly used as a solution for the definition and the computation of statistics in a non-Euclidean settings. *Intrinsic* methods instead define the statistics on the space itself, regardless of a more easy-to-use tangent space (see Marron and Alonso (2014) for

Chapter 3. Exploration in the Graph Space

an overview of intrinsic methods). In this chapter, we show that intrinsic Graph Space statistics can be computed iteratively via a combination of choosing optimal graph representatives in X/T , and computing Euclidean statistics in X . We call this iterative strategy *Align All and Compute* and we will decline this general strategy to the Fréchet Mean, the Geodesic Principal Components, and the Regression in the next chapter.

The chapter is organized as follow. In Section 3.1, we describe a urban planning case study: how the clustering of a population of simple graph can have an impact in the analysis of human movement in a square. A step further is done in Section 3.2 and Section 3.3, where we discuss the property of the Fréchet Mean in Graph Space and we theoretically define an intrinsic concept of Geodesic Principal Component Analysis. We propose the general *Align All and Compute* (AAC) strategy for computing the above mentioned Graph Space statistics, proving both its convergence to the local minima and its convergence in finite time (Section 3.3).

3.1 Clustering in the Graph Space



The aim of this section is not the exploration of different clustering techniques, but the application of a well known clustering strategy, namely Hierarchical Clustering (see for example Chapter 14 pp. 520 – 528 Hastie et al. (2009)) to a population of unlabelled graphs embedded in Graph Space, using the distance defined in Section 2.1. The scientific relevance of this work resides in its interdisciplinary application and in the intent of describing a spacial phenomenon with a non physical space, such as Graph Space. The relational representation mathematically described by a graph structure is common in landscape analysis. Within the design disciplines, such kind of representation is called diagrammatic, and has been widely applied both in architecture, landscape architecture, and urban planning over the past 50 years (e. g. Batty (2004)). One example of such diagrammatic representation is the one obtained through the Space Syntax method (e. g. Hillier et al. (1987)). However, the relational representation of urban movement is still unexplored. In the following, we show the cluster analysis of a set of graphs describing the urban movements in Piazza Leonardo da Vinci (Milan), to find similar usages of the square.

The content of this section is published in Calissano et al. (2020e).

Case Study: Piazza Leonardo da Vinci

Leonardo da Vinci Square is a relevant testing ground for the scope of this research. The square, adjacent to the main entrance of Politecnico di Milano (PoliMi) was redesigned in 2018 to foster pedestrian circulation by eliminating parking lots and vehicular paths. The resulting public realm supports spontaneous gatherings, and pedestrian and bicycle crossings. Because of its geometry, it allows for installing fixed and temporary features like benches, kiosks, pavilions, and stages to support cultural events. The existing grasslands have been extended and reshaped to fit the spontaneous pedestrian desire lines defined by the students crossing them. Finally, the new lighting design expands the usability of the plaza by enabling evening activities and safety.

One year after completion, the new layout has shifted the practices of use of the plaza. The use of space varies throughout the day. In the early morning, students, professors and scholars from the two universities located in the area (Politecnico and Università Statale) cross the space from the bike-sharing stalls and public transportation stops (subway, tramways, and buses) to the universities main gates. At the same time, the users of the two primary public services on site, the BESTA neurology hospital and the Italian National Institute for Cancer hospital, cross the space in the same directions. Around 11 AM,

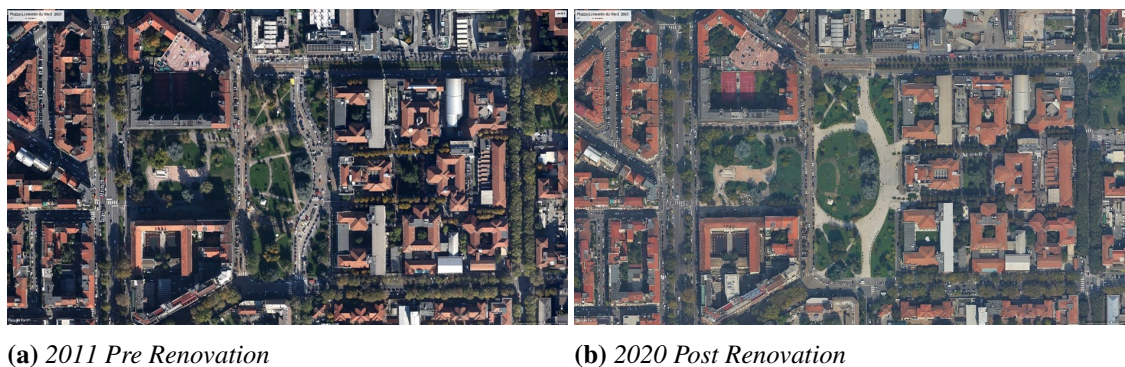


Figure 3.1: *Piazza Leonardo da Vinci, Milan, Italy. Note the pedestrian desire lines in the grasslands before the renovation. Image: Google Earth Pro*

residents walk around for leisure. During the lunch break, people buy street food at the numerous kiosks on site and eat seating on the benches or the grass. In the evening, space is quieter but still used because of a theater and the library of the PoliMi School of Architecture, which is open until midnight. Figure 3.1 show the view of the square before and after the redesign. The pedestrian-bikes shared surface, together with the lack of hard paths built in the grasslands, makes Leonardo da Vince Square a porous and flexible space, allowing for spontaneous use of the space, avoiding impositions by design. That is the reason why the case is relevant to this research. Following the square description usage previously described, three different types of square usage are simulated: (1) simple square crossing represented as a graph with two nodes (the entry point and the exit point of the square) and one link representing the trajectory; (2) seating in the square, represented as a graph with three nodes (the entry point, the seat, and the exit point) linked with three trajectories; (3) Kiosk, represented as a graph with four nodes (the entry point, the seat, the Kiosk, and the exit point) linked with four trajectories.

In Figure 3.2, we show the entrances and the kiosk. All the seats are parametrized according to both the real branches designed in the square and the steps in front of the university, where students use to sit. In Figure 3.2, all the generated trajectories are also shown on the map. To analyse them, we need to extrapolate the geo-referred data and embed them in Graph Space, where we compute the distance between graphs. Figure 3.3 shows three graphs dis-anchored from the geographical space. In real dataset, geo-localised data are sampled. There are several strategy to build graphs from trajectories. Here we interpret: (1) as graph nodes both the first and last measurement of every trajectory (i. e. the enter and the exit points) and all the points where the users stop for more than a certain time; (2) as graph edges the presence or absence of the trajectory between nodes. Different encoding are possible, as well as different attributes choice (e. g. the velocity of the trajectory,

3.1. Clustering in the Graph Space

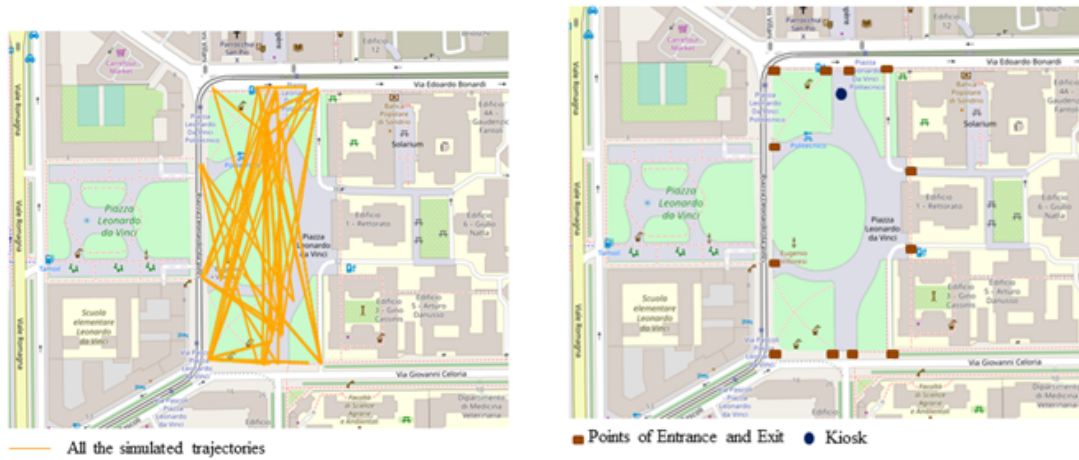


Figure 3.2: The simulated trajectories in Piazza Leonardo da Vinci. Each trajectory is simulated as a graph with an entry and an exit point, and eventually a seating point and a break at the Kiosk. Entry points, exit points, and kiosks are also shown. Image: OpenStreetMap

the stop-ping time). As shown in Figure 3.4, trajectories are correctly clustered using



Figure 3.3: Three graphs dis-anchored from the geographical space.

Hierarchical Clustering methodology (ward linkage). The Cluster 1 of single square cross, the Cluster 2 of people sitting in the square, and the Cluster 3 of people enjoying a break at the kiosk. The results show how the methodology can cluster the simulated data, thereby demonstrating the potential of the Graph Space approach. Also, the methodology can deal with unconventional paths. For example, in the Cluster 1 in Figure 3.4, we can see two paths that degenerate to a single point, where the pedestrian enters and exits the square at the same point without even crossing it.

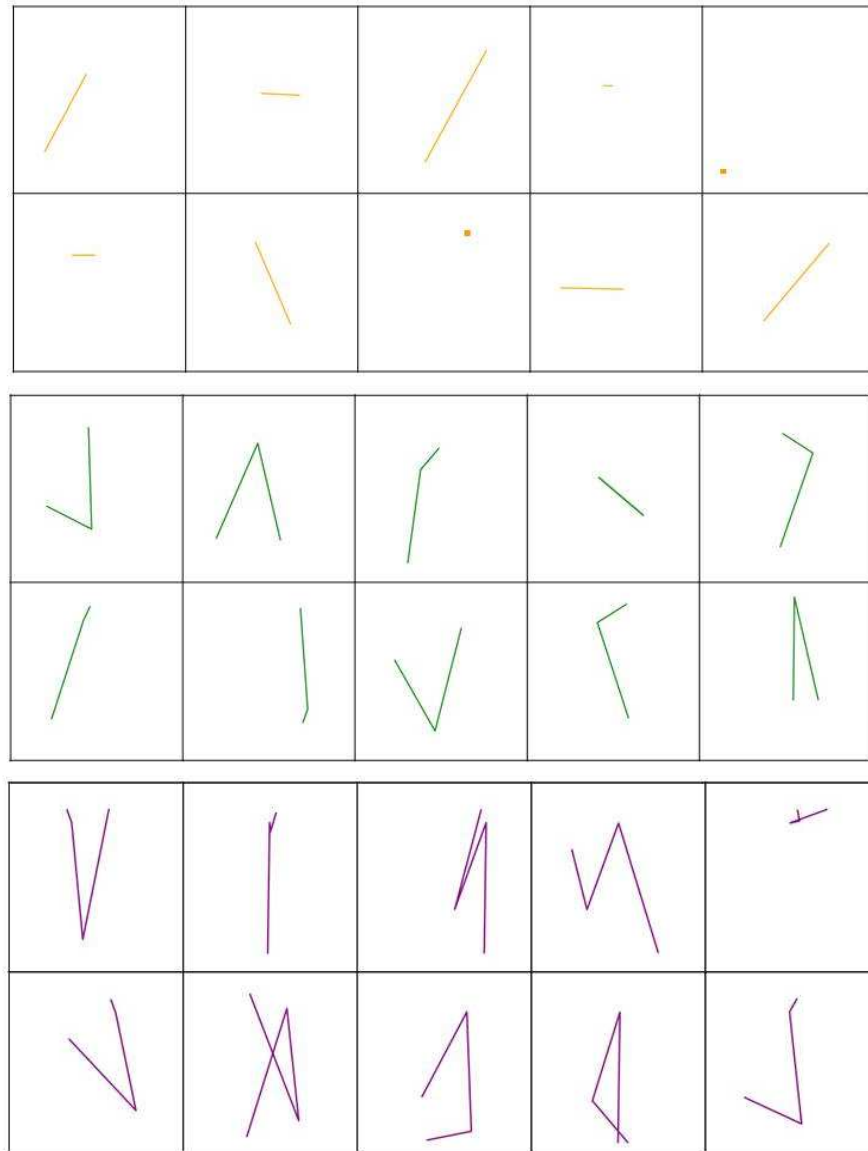


Figure 3.4: All the square graph trajectories are shown, divided in clusters. **Cluster 1** - top: trajectories of simple square cross **Cluster 2** - middle: trajectories of people crossing and seating in the square. **Cluster 3** -low: more complex trajectories of people seating in the square and going to the kiosk.

Case Study Discussion

This section aims to discuss if the measurement of the urban landscape and the urban movement should always be performed using a Euclidean setting or more complex embedding such as Graph Space. We focused on the graph representation of trajectories describing human movement in a square in Milan: Piazza Leonardo da Vinci. To show the potential of this framework, we simulated some simple trajectories of pedestrian crossing the square, seating in the square, and having a break at the kiosk in the square. Each trajectory is then mathematically represented as a graph and embedded into the Graph Space, to allow a relational more than a physical representation. Cluster analysis is performed, showing how this framework is able to cluster the simulated data, recognizing the three different types of usage. The simulation serves as a simple example of how the not Euclidean setting can show exciting results about a geographical dataset. From the point of view of the designer, the non-Euclidean graph could suggest unexpected, emergent uses of space, which can inform design in terms of paths, pedestrian shared surfaces and urban furniture location. The multidimensional nature of the Graph, not literally linked to a Euclidean 2D geometry, provides the designer with insights without forcing a layout, enabling interpretation and critical thinking in the workflow. The analysis in a more complex geometry produces a sort of diagram of the space, stressing on the relational aspects of the plaza inhabitation. The presented framework could be easily extended to other type of data generated by users interacting with the landscape (e.g. pictures, social media posts), creating original diagram about the space perception. It could be also easily enriched considering other variables such as the time evolution of the square usage.

3.2 The Fréchet Mean

The content of this section is part of the work Calissano et al. (2020a).

After discussing the usage of Graph Space distance to perform cluster analysis in a specific case study, we move back to the theoretical description of statistics on Graph Space. In this section we are going to discuss the property of the Fréchet Mean in Graph Space.

Consider the total space X , the Graph Space X/T obtained by quotienting out the action of the permutation group T on X , and a set of observations $\{[x_1], \dots, [x_k]\} \in X/T$. A basic quantity in nonlinear statistics is the *Fréchet mean*:

Definition 3.2.1. *The Fréchet mean of a sample $\{[x_1], \dots, [x_k]\} \in X/T$ is given*

$$[\bar{x}] = \arg \min_{[x] \in X/T} \sum_{i=1}^k d_{X/T}([x], [x_i])^2. \quad (3.1)$$

We note that for any geodesic space, the mean of two points is characterized as the midpoint of any geodesic connecting the two points. Thus, as illustrated in Figure 3.5, we obtain as a direct consequence of Corollary 2.2.3 and Theorem 2.2.5:

Corollary 3.2.2. *Fréchet means are not generally unique in Graph Space X/T . In fact, as shown by Figure 2.5, for any graph $[x] \in X/T$ and any radius $\varepsilon > 0$, there will be sets of points in $B([x], \varepsilon)$ (e.g. $\{[y_\varepsilon], [z_\varepsilon]\}$) whose Fréchet mean is not unique.*

Although Fréchet means in Graph Space are not necessarily unique, they are still explanatory and useful – just like on manifolds of positive curvature, where they are also not

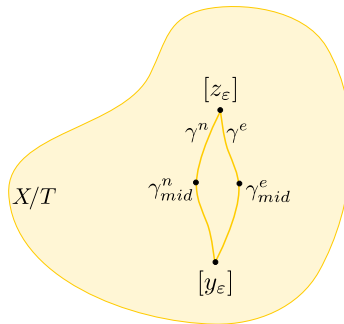


Figure 3.5: *Since geodesics are not necessarily unique in Graph Space X/T , Fréchet means are not necessarily unique either.*

unique. Examples of Fréchet Mean for real and simulated data are discussed in Section 3.3 together with Geodesic Principal Components.

3.3 Geodesic Principal Component Analysis

The content of this section is taken from Calissano et al. (2020a).

After the description of the properties of the Fréchet Mean in Graph Space, an exploration tool for the analysis of variance of a dataset is the Principal Component Analysis. One of the strategy in the literature to compute the principal components is to choose a node permutations that align each network to a fixed representative of a fixed network, such as the Fréchet mean μ , and performing principal components analysis in X (Guo et al., 2019)- this approach is analogous to tangent space methods from manifold statistics (Fletcher and Joshi, 2004). However, just like in manifold statistics, the projection to X based on optimal representatives x_i of individual graphs $[x_i]$ with respect to a single graph representative $\mu \in X$, leads to distortion of pairwise distances between observations $d_{X/T}([x_i], [x_j])$, which are not guaranteed to coincide with their distance $d_X(x_i, x_j)$ in the total space X . Noted this and considering the Graph Space environment properties, we introduce an intrinsic notion of generalized geodesic principal components (GGPCs), following Huckemann et al. (2010). To compute the GPCs we define an algorithm, namely Align All and Compute Algorithm, aimed at iteratively aligning the points to a new optimal estimation of the principal components. The algorithm is related to the Generalized Procruster Algorithm for shapes (Gower, 1975)).

Consider the *canonical projection* to the Graph Space X/T :

$$\pi: X \rightarrow X/T := \{[p] : p \in X\}$$

Definition 3.3.1. Denote by $\Gamma(X)$ the set of all straight lines (geodesics) in X . Following Huckemann et al. (2010), a curve δ is a *generalized geodesic on the Graph Space X/T* , if it is a projection of a straight line on X :

$$\Gamma(X/T) = \{\delta = \pi \circ \gamma : \gamma \in \Gamma(X)\}. \quad (3.2)$$

Since Graph Space is not an inner product space, we define orthogonality as:

Definition 3.3.2. Two *generalized geodesics* $\delta_1, \delta_2 \in \Gamma(X/T)$ are *orthogonal* if they have representatives in $\delta_1 = \pi \circ \gamma_1, \delta_2 = \pi \circ \gamma_2, \gamma_1, \gamma_2 \in \Gamma(X)$ which are orthogonal $\langle \gamma_1, \gamma_2 \rangle_X = 0$.

Chapter 3. Exploration in the Graph Space

In order to bridge computations in Graph Space X/T with computations in the total space X , we introduce a concept of alignment in X . We remind the reader that the metric on Graph Space is given by

$$d_{X/T}([x_1], [x_2]) = \min_{t \in T} d_X(tx_1, x_2),$$

which naturally leads to the concept of *optimal position* (Huckemann et al., 2010).

Definition 3.3.3 (Optimal position). *Given $\tilde{x} \in X$ and $t \in T$, the point $t\tilde{x}$ is in optimal position with respect to $x \in X$ if*

$$d_X(t\tilde{x}, x) = \min_{t' \in T} d_X(t't\tilde{x}, x) = d_{X/T}([\tilde{x}], [x]).$$

That is, the equivalence class $[\tilde{x}] \in X/T$ contains (at least) one point $t\tilde{x} \in [\tilde{x}]$ which has minimal distance to x , and this point is in optimal position with respect to x . Next, consider $[x] \in X/T$, $t \in T$ and δ a generalized geodesic in X/T with representative $\gamma \in \Gamma(X)$. The graph representative $tx \in X$ is in optimal position with respect to $\gamma \in \Gamma(X)$ if

$$d_X(tx, \gamma) = d_{X/T}([x], \delta).$$

The optimal position with respect to a generalized geodesic and its computation is detailed in the Algorithm 2. Having concepts of generalized geodesic, optimal position and orthogonality, we now define a set of geodesic principal components:

Definition 3.3.4. *Consider the canonical projection of the Graph Space $\pi: X \rightarrow X/T$ of X and consider a set $\{[x_1], \dots, [x_k]\} \subset X/T$ of graphs, $[x] \in X/T$, and $\delta \in \Gamma(X/T)$. The Generalized Geodesic Principal Components (GGPCs) for the set $\{[x_1], \dots, [x_k]\}$ are defined as:*

- *The first generalized geodesic principal component $\delta_1 \in \Gamma(X/T)$ is the generalized geodesic minimizing the sum of squared residuals:*

$$\delta_1 = \arg \min_{\delta \in \Gamma(X/T)} \sum_{i=1}^k (d_{X/T}^2([x_i], \delta)) \quad (3.3)$$

- *The second generalized geodesic principal component $\delta_2 \in \Gamma(X/T)$ minimizes (3.3) over all $\delta \in \Gamma(X/T)$, having at least one point in common with δ_1 and being orthogonal to δ_1 at all points in common with δ_1 .*
- *The point $\mu \in X/T$ is called Principal Component Mean if it minimizes*

$$\sum_{i=1}^k (d_{X/T}^2([x_i], [\mu])^2) \quad (3.4)$$

where $[\mu]$ only runs over points \tilde{x} in common with δ_1 and δ_2 .

3.3. Geodesic Principal Component Analysis

- The j^{th} generalized geodesic principal component is a $\delta_j \in \Gamma(X/T)$ if it minimizes (3.3) over all generalized geodesics that meet orthogonally $\delta_1, \dots, \delta_{j-1}$ and cross μ .

AAC: An algorithm to estimate statistics on the Graph Space

In this section, we show that intrinsic Graph Space statistics can be computed iteratively via a combination of choosing optimal graph representatives in X , not with respect to a single graph but with respect to the wanted statistic, and computing Euclidean statistics in X . Based on this derivation, we propose a general *Align All and Compute* (AAC) strategy for computing Graph Space statistics, which we use to compute both Fréchet means and, in particular, Generalized Geodesic Principal Components.

AAC for Fréchet Mean

The distance $d_{X/T}$ between two points in the Graph Space X/T corresponds exactly to the distance in X after posing one point in optimal position with the other. We now define the AAC algorithm, which is based on iteratively first posing observations in optimal position with respect to the current mean estimate, then re-estimating the mean based on the aligned observations, and repeating until convergence. In the case of computing Fréchet means of shapes, the AAC algorithm coincides with generalized Procrustes analysis (Gower, 1975).

Algorithm 1 AAC for the Fréchet Mean

Data: $\{[x_1], \dots, [x_k]\} \subset X/T$ observations in X ; a threshold $\varepsilon > 0$

Result: An estimate of the Fréchet Mean of $\{[x_1], \dots, [x_k]\}$.

Initialization: Select randomly an observed graph and randomly a representative: $\tilde{x} = \tilde{x}_i \in [x_i] \in \{[x_1], \dots, [x_k]\}$;

while $s > \varepsilon$ **do**

 Put every observation in optimal position with the representative \tilde{x} , obtaining an aligned set of representatives $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\} \subset X$;

 Compute the Fréchet Mean \bar{x} in X of $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$;

 Compute $s = d_X(\tilde{x}, \bar{x})$;

 Set $\tilde{x} = \bar{x}$

end

Return $[\bar{x}]$

This algorithm provides an estimate of a Fréchet Mean which is independent of the order of the data, and with improved convergence properties:

Chapter 3. Exploration in the Graph Space

Theorem 3.3.5. *AAC for Fréchet Mean (Algorithm 1) converges in finite time. Moreover, assume that Graph Space X/T is endowed with a probability measure η which is absolutely continuous with respect to the push-forward of the Lebesgue measure m on X . In particular, for $A \subset X/T$, we have $\eta(A) = 0$ if $m(\pi^{-1}(A)) = 0$. Let the dataset $[x_1], \dots, [x_k]$ is sampled from η ; now with probability 1, the estimator found by Algorithm 1 is a local minimum of the Fréchet function*

$$\sum_{i=1}^k d_{X/T}^2([x], [x_i]). \quad (3.5)$$

Proof. See Appendix. □

While we show that it theoretically converges in finite time, this might still be a long time, and we thus also add a convergence threshold ε in Algorithm 1.

Note also that our algorithm in practice relies on inexact graph matching, as graph matching is generally NP complete. We thus cannot be sure that our computed means were, in fact, based on completely correct graph matching. However, alternative mean algorithms also rely on graph matching, and are therefore similarly affected.

AAC for GGPCA

The AAC strategy, due to its iterative alignment with the chosen predictor, helps overcome the non-uniqueness of geodesics. It can readily be extended to other estimators that rely on geodesics such as Generalized Geodesic Principal Components. The alignment with a generalized geodesic, defined in Definition 3.3.3 is performed in a two step alignment procedure:

Algorithm 2 Algorithm for Optimal Position with respect to a generalized geodesic

Data: A point $x \in [x]$, a straight line $\gamma \in \Gamma(X)$, the domain $[s_{min}, s_{max}]$

Result: $t^* \in T$ such that t^*x is in optimal position wrt γ .

for $s \leftarrow s_{min}$ **to** s_{max} **do**

| Find $t(s) := \arg \min_{t \in T} d_X(tx, \gamma(s))$

end

Find $s^* = \arg \min_{s \in [s_{min}, s_{max}]} d_X(t(s)x, \gamma(s));$

Return $t^* = t(s^*)$

The obtained $t^* \in T$ is the permutation such that the point $t^*x \in [x]$ is the closest representative of $[x]$ to the geodesic γ in the interval selected.

Note that, due to the curvature of the space discussed in Section 2.2, the Fréchet mean is not ensured to be the same as the Principal Component Mean. A possible choice for the step distance function f is the proportion of variance explained by the first generalized geodesic principal component at the current and the previous step.

3.3. Geodesic Principal Component Analysis

Algorithm 3 AAC to compute the generalized geodesic principal components

Data: $\{[x_1], \dots, [x_k]\} \in X/T$ observations in X

Result: Geodesic principal components $\delta_1, \dots, \delta_k \in \Gamma(X/T)$

Initialize:

Select randomly $\tilde{x}_i \in [x_i] \{[x_1], \dots, [x_k]\}$;

Align all the observations to the representative \tilde{x}_i , obtaining a set of points $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\} \in X$ in optimal position with respect to \tilde{x}_i ;

Perform PCA on $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ in X obtaining $\gamma_1, \dots, \gamma_k \in \Gamma(X)$;

Project onto $\Gamma(X/T)$ as $\delta_i = \pi \circ \gamma_i$;

Set $\tilde{\delta}_1 = \delta_1, \dots, \tilde{\delta}_k = \delta_k$

while $s > \varepsilon$ **do**

Align all the points $\{[x_1], [x_2], \dots, [x_k]\}$ with respect to the generalized geodesic $\tilde{\delta}$, obtaining a new set of aligned points $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k \in X$ using Algorithm 2;

Perform PCA on $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ in X obtaining $\gamma_1, \dots, \gamma_k \in \Gamma(X)$;

Project onto $\Gamma(X/T)$ as $\delta_i = \pi \circ \gamma_i$;

Compute a step distance function $s = f(\tilde{\delta}_i, \delta_i)$;

Set $\tilde{\delta}_1 = \delta_1, \dots, \tilde{\delta}_k = \delta_k$.

end

Return $\delta_1, \dots, \delta_k \in \Gamma(X/T)$

Again, the AAC algorithm converges in finite time, and in the case of the first GGPC we can show that it converges to a local minimizer of the sum of squared residuals function:

Theorem 3.3.6. AAC for GGPCA (Algorithm 3) converges in finite time. Assume that Graph Space X/T is endowed with a probability measure η which is absolutely continuous with respect to the push-forward of the Lebesgue measure m on X , and let the dataset $[x_1], \dots, [x_k]$ be sampled from η . Now with probability 1, the estimator of the first GGPC found by Algorithm 1 is a local minimum of the sum of squared residuals function

$$\sum_{i=1}^k d_{X/T}^2(\delta, [x_i]). \quad (3.6)$$

where $\delta \in \Gamma(X/T)$.

Proof. See Appendix. □


For the case of the higher GGPCs, we do not have a proof of local minimization of (3.3) and Algorithm 3 should be considered a heuristic.

Remark. In the case of Fréchet means of shapes, the AAC coincides with Generalized Procrustes Analysis (Gower, 1975). In the statistical literature, a number of algorithms exist to compute means either approximately (Afsari et al., 2013; Arnaudon and Miclo,

Chapter 3. Exploration in the Graph Space

2014; Arnaudon et al., 2013; Bačák, 2014; Bonnabel, 2013; Hauberg et al., 2015; Miller et al., 2015; Sturm, 2003; Turner et al., 2014), or via heuristics (Billera et al., 2001; Feragen et al., 2011; Jain and Obermayer, 2008), whose applicability and efficiency vary with the complexity of the underlying nonlinear data space. A particularly popular strategy for computing Fréchet means in geodesic spaces is the iterative "midpoint" algorithm which obtains an updated mean estimate by stepping $1/k$ along the geodesic from a current mean estimate to a k^{th} random sample from the dataset. In Euclidean space, this computes the mean in finite time when samples are made without replacement. The same "without replacement" strategy is applied for trees in Feragen et al. (2011) and for graphs in Jain and Obermayer (2008), but these finite time algorithms do not generally return the mean in tree- or Graph Space, and should be considered heuristics. When running the algorithm *with* replacement, there are a number of scenarios in which it is known to converge towards the Fréchet mean, including non-positively curved spaces (Bačák, 2014; Miller et al., 2015; Sturm, 2003) and certain Riemannian manifolds of bounded curvature (Arnaudon and Miclo, 2014; Arnaudon et al., 2013; Chakraborty and Vemuri, 2015). Note, in particular, that for Riemannian manifolds, this algorithm is a special case of stochastic gradient descent (Bonnabel, 2013). While this algorithm is easy to generalize to Graph Space, its convergence proofs usually require bounded curvature to have some level of uniqueness for geodesics. Moreover, it does not come with an accompanying algorithmic strategy for computing more general statistics such as principal components. This motivates our choice to propose the AAC strategy. Other strategies have been proposed in the literature to compute geodesic principal components in nonlinear data space. For example Huckemann et al. (2010) proposes a general strategy with application to the Kendall Shape Spaces, and (Guo et al., 2019) proposes a simpler algorithm for principal components, analogous to the tangent space approaches known from manifold statistics.

Experiments on real and simulated data



In this section, we illustrate the introduced Graph Space statistics via three case studies emphasizing the framework's flexibility to model different graph features such as directed and undirected edges, or scalar and vector attributes, on both nodes and edges. Each computation of distances and geodesics require graph matching, which is an NP-complete problem with many existing heuristics for inexact matching (Conte et al., 2004; Emmert-Streib et al., 2016). In the simulations we use the graduate assignment algorithm (Gold and Rangarajan, 1996) for inexact matching. All experiments are conducted using our *GraphSpace* python package, which is available on github (Calissano et al., 2020c). Implementation details are listed in the package description. The convergence threshold ϵ used in the AAC

3.3. Geodesic Principal Component Analysis

algorithms is set to $\varepsilon = 0.001$ in Algorithm 1 and $\varepsilon = 0.01$ in Algorithm 3.



Figure 3.6: *The five networks used to simulate the dataset in Case study 1.*

Case study 1: Undirected Networks with Scalar Edge Attributes

We simulate networks with real-valued node- and edge attributes as follows. Figure 3.6 shows five networks with edge attributes $\{100, 80, 60, 40, 20\}$ as shown by the darkness of the color, whose equivalence classes in X/T we term $\{[x_1], [x_2], [x_3], [x_4], [x_5]\}$. We generate a synthetic dataset consisting of 50 observations randomly sampled in a stratified fashion by randomly picking one of the five equivalence classes $[x_i]$, $i = 1, 2, 3, 4, 5$, and then we randomly pick an element from this equivalence class by randomly permuting the network's nodes. The Fréchet mean and the GGPCs are computed using our AAC algorithm, which works on the quotient space X/T . These results are compared to Fréchet means and principal components computed on X , using the initial graph representations, as well as theoretically correct results which can be computed as for this particular synthetic dataset, the alignment can be made by hand. Figure 3.7 shows the three computed Fréchet means. The mean computed with the AAC algorithm is a good estimate of the theoretical one, both in terms of topological structure and edge weights. The mean obtained in the X space is a complete weighted network shown in Figure 3.7, capturing neither the topology nor the weights.

For GGPCA, Figure 3.8(a) shows the cumulative proportion of variance explained as a function of number of GGPCs, and Figure 3.8(b) visualizes the graph variation along the

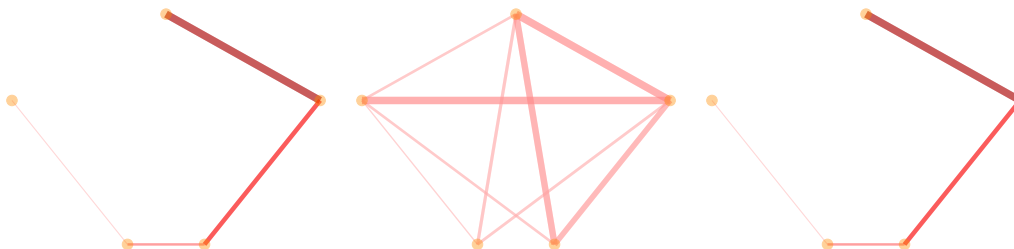


Figure 3.7: *Theoretical Fréchet mean; Fréchet mean in X ; Fréchet Mean computed via AAC. The scale is the same as in Figure 3.6.*

Chapter 3. Exploration in the Graph Space

first two principal components. Note that the Graph Space GGPCs obtained using AAC capture the same quantitative and visual level of variance as the theoretical GPCs, while this does not hold for the Euclidean version. In particular, the two GGPCs explain more than 90% of the original data variance, and the visualization of the 1st principal component shows how it runs from a single edge to the full structure. Note that the Geodesic Principal Components do not cross the Fréchet Mean, for the same reason this happens in Huckemann et al. (2010).

Case study 2: Undirected Networks with Vector Attributes

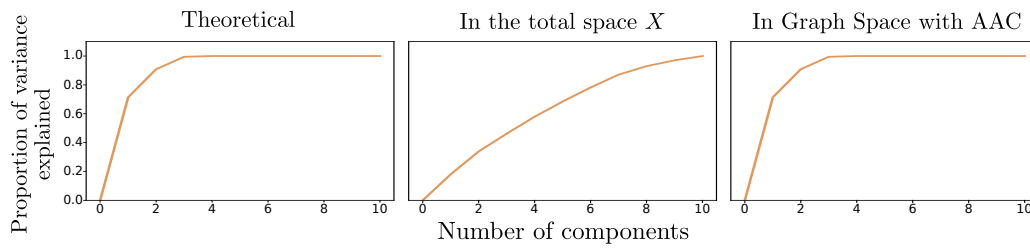
As an intuitive visual example with real data with vectors attributes, we sub-sample 20 cases of the letter "A" from the well known hand written letters dataset (Kersting et al., 2016; Riesen and Bunke, 2008). As shown in Figure 3.9(a), every network has node attributes consisting of the node's x - and y -coordinates, and binary (0/1) edge attributes indicating whether nodes are connected by lines. In Figure 3.9(a), the Fréchet Mean is shown, underling how the framework is capturing both the topology and the node coordinates. Figure 3.9(b) plots network variation along the three GGPCs. Note in particular how the principal components are capturing the variability in the way the letter A could be written: the variability of the length and the inclination of the horizontal bar, the angle between the vertical bars, and the reciprocal positions of the bars.

Case study 3: Directed Networks with Vector Attributes

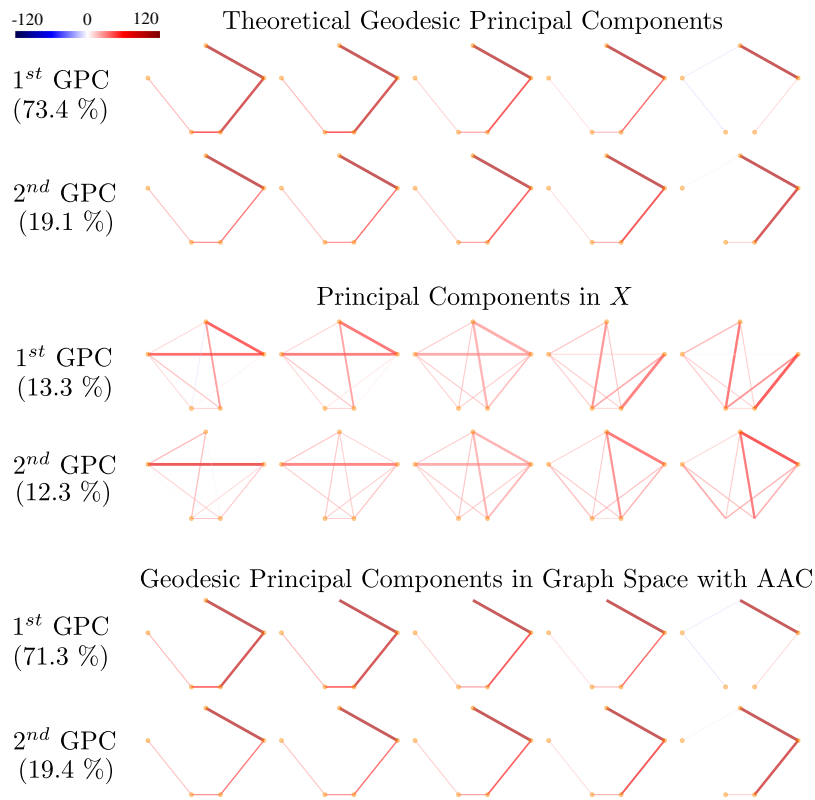
Our final example uses a mobility dataset from Open Data Regione Lombardia (Regione Lombardia, 2019). The dataset consists of origin-destination matrices of the commuting flux of people between the 11 provinces of the Lombardia Region in northern Italy. For every hour of a representative day in 2014, the fluxes were collected counting the number of people travelling by private mobility mode (car), railway system (train or metro), bus public transport system, or bike. This results in a set of 24 multi-layer networks, represented as graphs whose edge attributes are vectors in \mathbb{R}^4 . Each coordinate represents the flux associated to one of the mobility modes.

Figure 3.10, top, shows the multi-level networks associated with four different hours (left), as well as the Fréchet Mean on Graph space X/T computed with the AAC algorithm (right). Note that the density of the layers are well represented by the mean. By looking at the permutation of network nodes used to compute the mean and the generalized geodesics, we see that most of the time no permutation is performed (i.e. the node corresponding to a specific province at one hour is matched to the node representing the same province at another hour). This means that even if the province information is not stored in the graph, they are distinguished by their mobility properties. The only permutation happens at 5 p.m., when the town of Brescia is permuted with the town of Bergamo. These two towns

3.3. Geodesic Principal Component Analysis

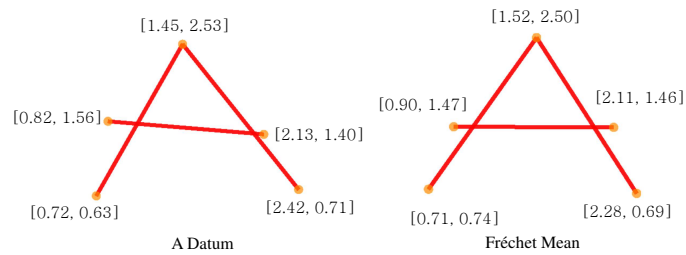


(a) Proportion of variance explained is shown by varying the number of the GGPCs. The proportion of variance explained correspond to the proportion of variance explained on the total space at the last step of the AAC.

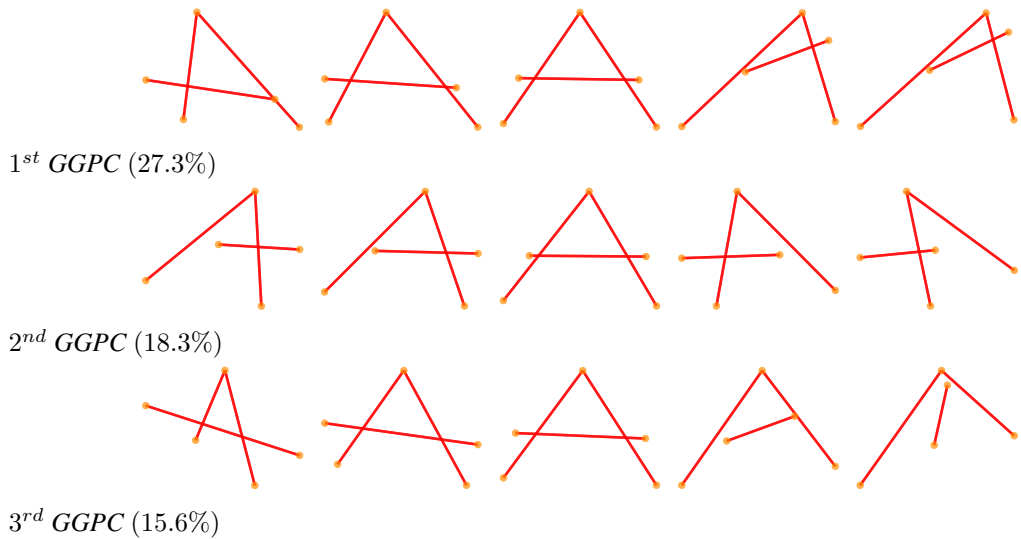


(b) The first two Generalized Geodesic Principal Components are shown by plotting the original data projected along the corresponding generalized geodesic (only the $q = 0.1, 0.25, 0.5, 0.75, 0.9$ quantiles are shown).

Figure 3.8: GGPCA analysis for Case study 1.



(a) **Left:** A datum extracted from the A dataset. Every unlabelled node has a bi-dimensional real valued attribute, while every edge has a 0, 1 attribute. **Right:** The Fréchet mean.



(b) Visualization of the GGPCs. 0.1, 0.25, 0.5, 0.75, 0.9 quantile of the projected scores are shown for the first three GGPCs.

Figure 3.9: GGPCA analysis for Case study 2.

3.3. Geodesic Principal Component Analysis

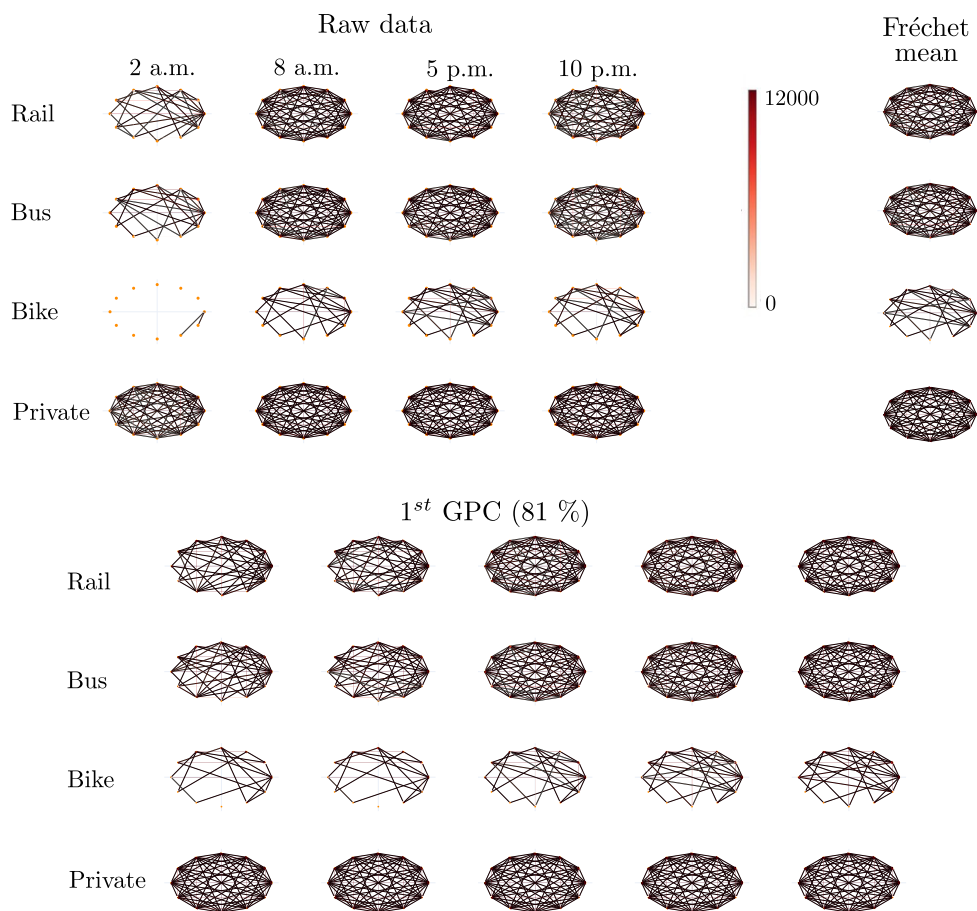



Figure 3.10: *Top left:* Network of the fluxes between the Lombardia Region provinces at 2 a.m., 8 a.m., 5 p.m. and 10 p.m.. *Top right:* Fréchet Mean in Graph Space X/T for the mobility modes bike, bus, rail and private. *Bottom:* GGPCA performed on scaled data. The first GGPC is shown, by plotting the quantiles of the original data projected along the corresponding generalized geodesic for each one of the levels ($q = 0.1, 0.25, 0.5, 0.75, 0.9$). The GGPC captures the density change along the day for all the transportation modes except from Private car mode.

Chapter 3. Exploration in the Graph Space

are both important commuting satellites of Milan, so their role is interchangeable with respect to the commuting flux in the afternoon. This application shows how this framework does not suffer the possible risk of “over-matching” also in case where the cardinality of the permutation group is pretty high (i.e., even though about 40 million possible node permutations are available, no artificial permutations are introduced by the algorithm). In Figure 3.10, bottom, we show that the majority of the variability is explained by the first GGPC. The first principal component captures 81% of the total variance, and we see that by moving along this component, the density of the layers changes, except from the private mode, which is a complete graph at at every hour of the day. As expected, the first GGPC captures the main variability of the dataset, which resides in the difference in number of travels along the day, from early-morning and late afternoon peaks to minima at night.

3.4 Discussion



In this chapter, we extended tools for intrinsic exploratory statistics to Graph Space. In terms of cluster analysis, we presented a case study about Piazza Leonardo da Vinci, Milan. The case study used the distance defined in Graph Space and the classical concept of Hierarchical Clustering to cluster citizens’ usage of a renewed square. This work offers potential extensions in terms of clustering techniques and applications. Cluster analysis of population of networks has been little explored in this thesis, and it is a broad research topic. In terms of application, the analysis of human movement as population of graphs provides interesting interdisciplinary results. The geometrical embedding of the human movement in an abstract space allows for location comparison (i.e. the alignment of networks -paths- and the matching of nodes -locations), which can reveal interesting insights about the space usage and the perception of the landscape. In the rest of the chapter, we discussed how to define intrinsic statistics. We described the properties of the Fréchet Mean and we defined an original concept of Generalized Geodesic Principal Components. Computing statistics on Graph Space is not straightforward in practice, due to the geometrical properties of Graph Space introduced in Chapter 2, especially because of the problems induced by its unbounded curvature. We proposed a general algorithm (Align All and Compute Algorithm - AAC), useful to compute intrinsic statistics on the Graph Space. First, we applied the AAC to estimate the the Fréchet mean, showing the convergence to a local minimum in finite time. Second, we applied the AAC to estimate the GGPCs. We demonstrated the practical utility of the introduced framework on three case studies: A simulated study, showing how AAC for Fréchet Mean and GGPCs estimates the expected theoretical results, and two real world examples including both directed and undirected networks, as well as both scalar and vector attributes. The examples emphasize

that GGPCA is an efficient tool to visualize the variability in a population of networks. As a further development, we would do a more detailed exploration of GPCA as a dimensional reduction tool for networks, understanding how nodes and edges can be described in a smaller subspace described by the geodesic principal components.

Note that Graph Space is analogous to the classical shape analysis (Dryden and Mardia, 1998; Kendall, 1984), with graph nodes analogous to landmarks. In this sense, Graph Space might also be a potential alternative to shape analysis when the indexing of landmarks is unknown or the number of landmarks varies. On a related note, as our computation of means and principal components is essentially a generalization of generalized Procrustes analysis, our estimators may suffer from similar biases as those shown to exist for shapes (Miolane et al., 2017); this would be an interesting problem for future research. As a further development, the analysis of networks with non-Euclidean attributes is a first interesting extension of the current framework. Moreover, the AAC algorithm is a general strategy to compute statistics on the Graph Space, which can naturally be extended to other statistical tools, such as regression and classification.

Appendix

Appendix A: Proof of the Theorem - AAC for Fréchet Mean (Algorithm 1)

Proof of Theorem 3.3.5. First, we prove convergence in finite time. Algorithm 1 consists of two steps repeated iteratively. Consider the function

$$\sum_{i=1}^k d_X^2(\mu^{cur}, x_i^{cur}), \quad (3.7)$$

where, at any point in time, μ^{cur} is the current representative in X of the current estimate of the Fréchet mean, and x_i^{cur} is the current representative (with current optimal node alignment to μ^{cur}) in X of the sample point $[x_i]$. Note that the first step, aligning data points to the current representative of the current mean estimate, cannot increase the value of (3.7) as an improved alignment would indeed lower the value of (3.7). Similarly, the second step, which is the re-estimation of the Fréchet mean given the new alignments, also cannot increase the value of (3.7) as, again, an improved estimate would lower its value. Moreover, if the value of (3.7) stays fixed two iterations in a row, the algorithm will terminate. Thus, the iterative algorithm will never see the same set of sample-wise alignments twice. As there are only finitely many such sets, the algorithm is forced to terminate in finite time.

Next, we move to convergence to a local minimum. Let $[\mu] \in X/T$ be the estimated mean, let $\mu \in X$ be a representative of it, and let $x_1, \dots, x_k \in X$ be optimally aligned representatives of the sampled graphs (as in the final step of the AAC algorithm). We will show below that with probability 1, there exists some $\varepsilon > 0$ such that for any $\mu' \in B_X(\mu, \varepsilon)$, the representatives x_1, \dots, x_k are also optimally aligned with μ' . In this case, since μ is a local minimizer of (3.7) within $B(\mu, \varepsilon)$, and $d_{X/T}([\mu'], [x_i]) = d_X(\mu', x_i)$ for all $\mu' \in B_X(\mu, \varepsilon)$, the estimated mean graph $[\mu]$ is a local minimizer of (3.5).

In order to prove the existence of such an $\varepsilon > 0$, we rely on the following lemma:

Lemma 3.4.1. *Given representatives x_1, \dots, x_k of $[x_1], \dots, [x_k]$ with mean μ in X , the following holds with probability 1:*

For all $i = 1, \dots, k$ and for all $t \in T \setminus T_{x_i}$,

$$d(\mu, x_i) \neq d(\mu, tx_i),$$

where T_{x_i} is the stabilizer $T_{x_i} = \{t \in T \mid tx_i = x_i\}$.

If the lemma holds, then we may define

$$\delta = \min\{d(\mu, tx_i) - d(\mu, x_i) \mid i = 1, \dots, k, t \in T \setminus T_{x_i}\} > 0.$$

We now set $\varepsilon = \frac{\delta}{2}$ and consider $\mu' \in B_X(\mu, \varepsilon)$. We wish to show that for all $i = 1, \dots, k$ and all $t \in T \setminus T_{x_i}$, we have $d(\mu', x_i) < d(\mu', tx_i)$, namely that the optimal representative of $[x_i]$ is left unchanged for all i .

Note that by the definition of δ , we have

$$d(\mu, x_i) \leq d(\mu, tx_i) - \delta,$$

and by the triangle inequality, we have

$$d(\mu', x_i) \leq \underbrace{d(\mu', \mu)}_{< \varepsilon = \frac{\delta}{2}} + d(\mu, x_i)$$

and

$$d(\mu, tx_i) \leq \underbrace{d(\mu, \mu')}_{< \varepsilon = \frac{\delta}{2}} + d(\mu', tx_i).$$

We compute

$$\begin{aligned} d(\mu', x_i) &\leq \underbrace{d(\mu', \mu)}_{< \frac{\delta}{2}} + d(\mu, x_i) \\ &< \frac{\delta}{2} + d(\mu, x_i) \leq \frac{\delta}{2} + d(\mu, tx_i) - \delta \\ &< -\frac{\delta}{2} + \frac{\delta}{2} + d(\mu', tx_i) = d(\mu', tx_i), \end{aligned}$$

which completes the proof of Theorem 3.3.5 under the assumption that Lemma 4.3.1 holds. \square

Proof of Lemma 4.3.1. In order to prove the lemma, we will show that the set

$$\mathcal{X}_T = \left\{ ([x_1], \dots, [x_k]) \in (X/T)^k \left| \begin{array}{l} d(\mu, x_i) = d(\mu, tx_i) \\ \text{for some representatives} \\ x_1, \dots, x_k, \\ i = 1, \dots, k \text{ and } t \in T \setminus T_{x_i} \end{array} \right. \right\}$$

has measure $\eta_k(\mathcal{X}_T) = 0$, where η_k is the product measure induced by η on $\underbrace{X/T \times \dots \times X/T}_k$.

For each element $t \in T$, denote by $X^t = \{x \in X \mid tx = x\}$ the fixed point set of t . Note that $\eta_k(\mathcal{X}_T) = m_k(\pi^{-1}(\mathcal{X}_T))$, and that

$$\pi^{-1}(\mathcal{X}_T) = \bigcup_{i=1}^k \bigcup_{t \in T} \mathcal{X}_{i,t},$$

Chapter 3. Exploration in the Graph Space

where

$$\begin{aligned} \mathcal{X}_{i,t} &= \{(x_1, \dots, x_k) \in X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X \mid d_X(\mu, x_i) = d_X \mu, tx_i)\} \\ &\subset \underbrace{X \times \dots \times X}_k. \end{aligned}$$

The preimage $f^{-1}(0)$ of the function

$$f: \underbrace{X \times \dots \times X}_k \rightarrow \mathbb{R}, \quad (x_1, \dots, x_k) \mapsto d_X^2(\mu, x_i) - d_X^2(\mu, tx_i)$$

satisfies

$$f^{-1}(0) \cap X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X = \mathcal{X}_{i,t}.$$

We show that f is a submersion on $X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X$ by showing that it has nonzero gradient. We can rewrite

$$\begin{aligned} f(x_1, \dots, x_k) &= \\ &= \left(\frac{1}{k} \sum_{j=1}^k x_j - x_i\right)^T \left(\frac{1}{k} \sum_{j=1}^k x_j - x_i\right) - \left(\frac{1}{k} \sum_{j=1}^k x_j - tx_i\right)^T \left(\frac{1}{k} \sum_{j=1}^k x_j - tx_i\right) = \\ &= \frac{2}{k} \sum_{j=1}^k (x_j^T tx_i - x_j^T x_i). \end{aligned}$$

For $j \neq i$ we obtain

$$\nabla_{x_j} f(x_1, \dots, x_k) = \frac{2}{k} (tx_i - x_i)$$

which is nonzero for $(x_1, \dots, x_k) \in X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X$. It follows that f is a submersion on $X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X$. As a result, the set

$$f^{-1}(0) \cap X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X = \mathcal{X}_{i,t}$$

has codimension 1 and, in particular,

$$m_k(\mathcal{X}_{i,t}) = m_k(f^{-1}(0) \cap X \times \dots \times \underbrace{X \setminus X^t}_{i^{th}} \times \dots \times X) = 0.$$

But then,

$$\eta_k(\mathcal{X}_k) = m_k(\pi^{-1}(\mathcal{X}_T)) = m_k\left(\bigcup_{i=1}^k \bigcup_{t \in T} \mathcal{X}_{i,t}\right) \leq \sum_{i=1}^k \sum_{t \in T} m_k(\mathcal{X}_{i,t}) = 0,$$

which proves the lemma. □

Appendix B : Proof of the Theorem - AAC for GPCA (Algorithm 3)

Proof of Theorem 3.3.6. The proof of convergence of AAC for GGPCA follows the exact same strategy as for the Fréchet mean in Theorem 3.3.5, noting:

- The algorithm converges in finite time because every step either decreases or leaves unchanged the value of the sum of squared residuals for representatives in X :

$$\sum_{i=1}^k d^2(x_i, \gamma)$$

with respect to the first GGPC $\delta \in \Gamma(X/T)$ and its representative $\gamma \in \Gamma(X)$.

- For the first GGPC, the algorithm converges to a local minimum of (3.3) following the same argument as above. Here, the ε -neighborhood of the obtained estimate δ has to be considered on the Grassmannian manifold of 1-dimensional sub-spaces of X .

□

CHAPTER 4

Prediction in the Graph Space

Prediction is the last main topic of this PhD thesis. We develop a regression model to predict a set of graphs as a function of a set of vectors and a prediction interval for a graph prediction. Defining a graph-valued regression model means to build a regression between a set of of real values (i.e. regressors) and a set of graphs (i.e. responses). This regression model is the first step to design a regression, an interpolation between graph points within the Graph Space.

Within the first generation approach to graphs, prediction embodies a well known scientific problem: the prediction of edges and nodes in a given graph. Statisticians and sociologists have been focusing on the analysis of random graphs since the 50s, starting from the seminal work by Erdős and Rényi (1960). From the Erdős-Rényi model, many different others have been proposed to describe the theoretical distribution behind the network datum and its variability. Exponential Random Graph Models (Lusher et al., 2013; Robins et al., 2007) and Stochastic Actor Oriented Models (Snijders, 2011, 2017) are some examples. Dynamic Network Analysis (DNA) is another stream of literature aiming to model the temporal evolution of a network (see Carley (2003) for an overview). Aside from the exploration of the generative models behind graphs, the effect of covariates should be taken into account in graph-on-variable regression model. Some examples are: a discrete partition of the space of the covariates to predict labelled networks (Liu et al.,

Chapter 4. Prediction in the Graph Space

2010); the definition of a regression model with continuous covariates (Ni et al., 2019); a regression model for graphs represented as Laplacian matrices (Severn et al., 2020); a Bayesian approach to the regression process of binary networks (Durante and Dunson, 2014); a multi-linear regression for a set of tensor data (Hoff, 2015). In machine learning, a frequently studied problem is the prediction of nodes and edges from scalars or vectors with Graph Neural Networks (Scarselli et al., 2008; Xu et al., 2018; Zhang et al., 2018) or Variational Autoencoder for Graphs (Kipf and Welling, 2016; Schlichtkrull et al., 2018). Within the second generation approach, a considerable amount of works considers population of graphs, where the graph plays the role of the independent, or input, variable. This includes problems such as graph classification, or regressing real-valued properties from graph-valued input. Such problems are often tackled by embedding the graphs, explicitly or implicitly, in a Euclidean feature- or embedding space (Kashima et al., 2003; Maron et al., 2018; Saigo et al., 2009; Vishwanathan et al., 2010; Xu et al., 2018), where much of the relational information is lost. This approach is fine when the information needed to make the prediction can be encoded in a Euclidean feature space. A more challenging problem is when the predicted dependent variable is a graph. Predicting an unlabelled network from a set of variables requires the definition of an interpolating regression function between graphs. We tackle this using interpolation in *Graph Space*. For other types of nonlinear data, such problems are frequently handled using tangent space methods, where the regression models are estimated in the Euclidean tangent space (Fletcher, 2013; Hinkle et al., 2014) of the embedding space (in our case Graph Space). However, these suffer from distorted residuals (Mallasto and Feragen, 2018), giving challenging conditions for model fitting. Another easily applicable approach is given by the non-parametric kernel smoothing approaches or the K -nearest neighbor regressions, which have appeared both in manifold statistics (Davis et al., 2010) and in the more general, stratified, tree-spaces (Skwerer, 2014). In the context of high dimensional data such as networks, these methods suffer from the curse of dimensionality, which could make them poorly performing in practice. Additionally, as these methods require computing local means or neighborhoods for every test point, they can also have significant computational cost. In terms of parametric models, linear regression models have been generalized to geodesic (Fletcher, 2013; Hong et al., 2014), polynomial (Hinkle et al., 2014) and more general parametric (Hong et al., 2016) regression models, defined exclusively on manifolds. Staying within the manifold-valued regression regime, more recent works also include manifold-valued models with uncertainty quantification (Hong et al., 2017; Mallasto and Feragen, 2018). However, as proven in Calissano et al. (2020a), Graph Space is not a manifold, so we cannot apply manifold methods directly.

In Section 4.1, we address this problem by designing an intrinsic, generalized linear regression model taking values in Graph Space. The resulting parametrized regression

models can be given a high level of flexibility via nonlinear basis functions supplied by the user. We provide an efficient estimate via a version of the AAC for regression, which combines statistical precision by using intrinsic, non-distorted residuals, with computational benefits as estimation is effectively made in a Euclidean "tangent space". We introduce an intrinsic generalized geodesic regression for a graph-on-vector regression problem. Inspired by the results obtained for Geodesic Principal Components in the previous chapter, the estimation of this regression function is done via Align All and Compute, which is declined hereby to minimize the prediction error.

Along with the prediction model, a technique to estimate the uncertainty of the predicted value is required. The strongly non-Euclidean nature of these data type and the difficulties of inferring a distribution of a population of networks (especially unlabelled) creates the requirement of the definition of a non-parametric strategy. The last Section 4.2 of this thesis focuses on the development of a conformal prediction interval (Vovk et al., 2005; Zeni et al., 2020) for every prediction model for a population of graphs. To the best of our knowledge, uncertainty quantification has been relatively overlooked in the context of strongly non-Euclidean data, despite its importance in the application tasks. While some attempts have been proposed in a functional data analysis setting (Antoniadis et al., 2016; Degras, 2011) and for the analysis of phylogenetic trees (Willis, 2019)), no uncertainty quantification techniques has been proposed for more general network data. Population of networks are very complex in terms of distribution. They show complex covariance structure due to their relational nature, and they can describe different phenomena on edges and nodes, making parametric inference hard to apply. These are the main reasons why we opt for a non parametric conformal prediction strategy.

In addition, this chapter contains two case studies about the effect of 2020 Covid-19 pandemic outbreak on the urban and the territorial mobility. In Section 4.1, the prediction of the public transport network usage in Copenhagen during Covid-19 is studied, analysing with an anova model the effect of the lockdown on the networks dynamics. In Section 4.2, we analyse the Origin Destination Matrices of Lombardy region before, during, and after the total lockdown, understanding the distribution of the fluxes between the different provinces.

This chapter is organized as follow. In Section 4.1, we introduce the Graph-Valued regression model, we propose a version of the Align All and Compute Algorithm to estimate the regression and we discuss the model with two case studies. In Section 4.2, we introduce the conformal prediction framework for labelled and unlabelled graphs. The conformal prediction strategy is applied to two simulated example and one case study.

4.1 Graph-Valued Regression

The content of this section is extracted from the work Calissano et al. (2020b).

Given a sample $(s_1, [x_1]), \dots, (s_k, [x_k])$, where $(s_i, [x_i]) \in \mathbb{R}^p \times X/T$, we aim to describe the relationship:

$$f: \mathbb{R}^p \rightarrow X/T$$

minimizing:

$$\sum_{i=1}^k d_{X/T}^2([x_i], f(s_i)) \quad (4.1)$$

over all the possible functions belonging to a prescribed family. In this section, we describe how such families of functions in Graph Space X/T can be defined and parametrized, and how to estimate the resulting regression model.

First, we recall the definition of *generalized geodesics*:

Definition 4.1.1 (Generalized Geodesics). *Denote by $\Gamma(X) := \{\gamma : \mathbb{R}^p \rightarrow X\}$ the set of all straight lines ($p = 1$), planes, or hyper-planes ($p > 1$) in the total space X . A curve, surface or hypersurface δ is a generalized geodesic, or generalized geodesic subspace, on the Graph Space X/T , if it is a projection of a straight line, plane, or hyper-plane on X :*

$$\Gamma(X/T) = \{\delta = \pi \circ \gamma : \gamma \in \Gamma(X)\}. \quad (4.2)$$

where $\pi : X \rightarrow X/T$ is the canonical projection from total to quotient space.

Next, we consider two potential classes of regression models.

Definition 4.1.2 (Generalized Linear Regression Models - Class I). *Consider the regression model*

$$f_\beta: \mathbb{R}^p \rightarrow X/T, \quad s \mapsto f_\beta(s) \in X/T$$

given by $f_\beta(s) = \pi \circ h_\beta(s)$, where $h_\beta: \mathbb{R}^p \rightarrow X = \mathbb{R}^J$ is a linear regression model on X of the form

$$h_\beta(s_i)_a = \sum_{j=0}^p \phi_j(s_i) \beta_j(a) \quad (4.3)$$

where the ϕ_j are continuous, possibly non-linear, basis functions $\phi_j: \mathbb{R} \rightarrow \mathbb{R}$, $j = 1, \dots, p$, for edge- and node-wise coefficients $\beta_j(a)$, where $a = 1, \dots, J$.

Definition 4.1.3 (Generalized Geodesic Regression Models - Class II). *Consider the regression model*

$$f_\beta: \mathbb{R}^p \rightarrow X/T, \quad s \mapsto f_\beta(s) \in X/T$$

where $f_\beta \in \Gamma(X/T)$ is a generalized geodesic. This can be written as $f_\beta(s) = \pi \circ h_\beta(s)$, where $h_\beta: \mathbb{R}^p \rightarrow X = \mathbb{R}^J$ is a linear regression on X of the form:

$$h_\beta(s_i)_a = \sum_{j=0}^p s_{ij} \beta_j(a) \tag{4.4}$$

again for edge- and node-wise coefficients $\beta_j(a)$ where $a = 1, \dots, J$.

See Appendix 4.3 for the matrix representation of the problem. Given these two classes, we define the family of generalized linear regression models by $\mathcal{F}(X/T) := \{f_\beta: \mathbb{R}^p \rightarrow X/T\}$, i.e. the family of models such that $f_\beta(s) = \pi \circ h_\beta(s)$, where $h_\beta(s)$ is defined as in Definition 4.1.3 or Definition 4.1.2.

Note that $\mathcal{F}(X/T)$ contains the family $\Gamma(X/T)$ of generalized geodesic regression models. To simplify the notation, we omit the β writing $h_\beta(s) = h(s)$ and $f_\beta(s) = f(s)$ in the following paragraphs.

By using the concept of generalized linear models and the concept of alignment with respect to a regression model, the Generalized Linear Regression Model is defined in the following way:

Definition 4.1.4 (Generalized Linear Regression). *Given a sample $\{(s_1, [x_1]), \dots, (s_k, [x_k])\}$ where $(s_i, [x_i]) \in \mathbb{R}^p \times X/T$, their Generalized Linear Regression $f \in \mathcal{F}(X/T)$ is the one that minimizes the residuals as specified by Equation (4.1).*

From intrinsic residuals on X/T to Euclidean residuals on X

Given a sample $\{(s_1, x_1), \dots, (s_k, x_k)\}$, $(s_i, x_i) \in \mathbb{R}^p \times X$ consisting of independent variables $s_i \in \mathbb{R}^p$ and dependent variables given by specific graph representatives $x_i \in X$, the modelling of a regression line $h: \mathbb{R}^p \rightarrow X$ is well known in statistics as a multiple output regression model. This regression line can be projected onto a generalized geodesic in the quotient space. However, the estimation of $h(s)$ entirely on how the representatives x_i for the graphs have been selected – since for any node permutation t_i , the representation $t_i x_i \in X$ is also a representative of the same graph. Here, we introduce the concept of optimal alignment with respect to a regression line in order to select the optimal representative $t_i x_i \in [x_i]$, $t_i \in T$, $i = 1, \dots, k$ for the graphs $[x_i]$. The original concept of optimal alignment was introduced to minimize the distance between a geodesic and an equivalence

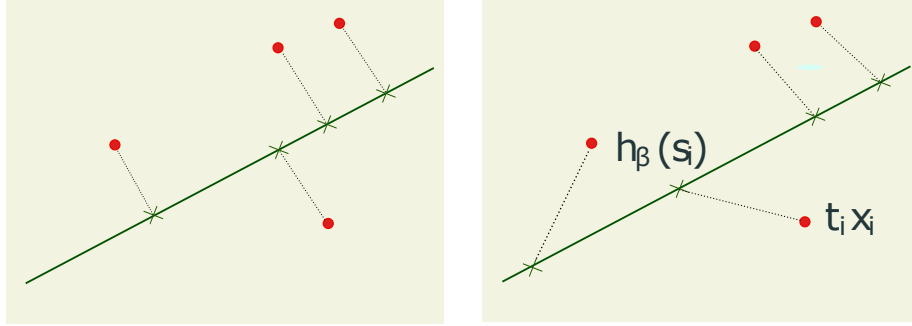


Figure 4.1: Alignment Differences between the Geodesic Principal Component Alignment and the Geodesic Regression Alignment

class in the context of Principal Components (Section 3.3). In a regression problem, we instead seek to minimize the prediction residual, which is the distance between the predicted points along the regression line $f(s_i)$ and the observed datum $[x_i]$. The optimal representative tx of the equivalent class $[x]$ is the point that minimizes the distance not with respect to the whole regression line (i.e. the projection along the line), but with respect to the predicted point $f(x)$ along the Generalized Linear Regression model:

Definition 4.1.5 (Alignment with respect to a regression model). Consider $(s_i, [x_i]) \in \mathbb{R}^p \times X/T$, $i = 1, \dots, k$, $t \in T$, and $f: \mathbb{R}^p \rightarrow X/T$ a generalized linear model in X/T with associated $h: \mathbb{R}^p \rightarrow X$. The graph representative $t_i x_i \in X$ is in optimal regression position with respect to the regression line f on X if

$$d_X(t_i x_i, h(s_i)) = d_{X/T}([x_i], f(s_i)). \quad (4.5)$$

Figure 4.1 illustrates the difference between alignment with respect to a Generalized Geodesic Principal Component (Section 3.3) and the alignment with respect to a regression model. In our case, the alignment distance is not the distance between an observation and its orthogonal projection onto the line, but the distance between the observation and the associated prediction along the regression line.

In Theorem 4.1.6 we show that this estimation strategy actually corresponds to regressing with intrinsic residuals from X/T .

Estimation: The Align All and Compute Algorithm for Regression

Inspired by the AAC algorithm for Generalized Geodesic Principal Components defined in the previous section, we next introduce an AAC algorithm for regression, where the alignment procedure is adapted to regression.

While the regression model of Definition 4.1.4 is framed intrinsically in the Graph Space X/T , we obtain a simple estimation procedure by the *Align All and Compute Algorithm for Regression* (AAC), which combines the Euclidean regression model estimation in the total space X with iterative alignment to the current model estimate. For completeness and notation, we recall the reader the analytical solution of the Multiple Output Least Square Linear Regression model in Appendix 4.3.

Reminding that we are minimizing the sum of squared residuals loss function

$$\sum_{i=1}^k d_{X/T}^2([x_i], f(s_i)).$$

The AAC algorithm optimizes the loss with respect to one argument at a time: first with respect to t_i , freezing $f(s_i)$ (i.e. aligning the points), and consequently optimizing with respect to $f(s_i)$ freezing the optimally aligned points $t_i x_i$, by minimizing the corresponding loss function on the total space X :

$$\sum_{i=1}^k d_X^2(t_i x_i, h(s_i)) \tag{4.6}$$

In Algorithm 4, the detailed steps of the implementation are shown. As a result (proven in Theorem 4.1.6 below), the loss decreases in every step, which is crucial for its convergence.

Chapter 4. Prediction in the Graph Space

Algorithm 4 AAC to compute the Generalized Geodesic Regression

Data: $\{(s_1, [x_1]), \dots, (s_k, [x_k])\} \in \mathbb{R}^p \times X/T$

Result: Generalized Geodesic Regression $f(s) \in \mathcal{F}(X/T)$

Initialize:

Select randomly $t_i x_i \in [x_i]$ among $\{[x_1], \dots, [x_k]\}$, $t_i \in T$;

Align all the observations to the representative $t_i x_i$, obtaining a set of points $\{t_1 x_1, t_2 x_2, \dots, t_k x_k\} \in X$ in optimal position with respect to $t_i x_i$;

Perform a Regression $\{(s_1, t_1 x_1), (s_2, t_2 x_2), \dots, (s_k, t_k x_k)\}$ in X obtaining $h(s) \in \Gamma(X)$ solving Equation 4.6;

Project as $f(s) = \pi \circ h(s)$;

Set $\tilde{f}(s) = f(s)$

while $\delta > \varepsilon$ **do**

 Align all the points $\{[x_1], [x_2], \dots, [x_k]\}$ with respect to the generalized geodesic regression $\tilde{f}(s)$, obtaining a new set of aligned points $t_1 x_1, t_2 x_2, \dots, t_k x_k \in X$;

 Perform GGR on $\{(s_1, t_1 x_1), (s_2, t_2 x_2), \dots, (s_k, t_k x_k)\}$ in X obtaining $h(s) \in \Gamma(X)$ by solving Equation 4.6;

 Project onto $\mathcal{F}(X/T)$ as $f(s) = \pi \circ h(s)$;

 Compute the step as the distance between the sum of square prediction errors $\delta = \mathcal{D}(\tilde{f}(s), f(s))$;

 Align all the observations wrt $f(s)$, obtaining a set of points $\{t_1 x_1, t_2 x_2, \dots, t_k x_k\} \in X$ as explained in Definition 4.5;

 Set $\tilde{f}(s) = f(s)$.

end

Return $f(s) \in \mathcal{F}(X/T)$

4.1. Graph-Valued Regression

The following Theorem 4.1.6 proves the convergence in finite time of the AAC to a local minimum.

Theorem 4.1.6. *Let Graph Space X/T be endowed with a probability measure η which is absolutely continuous with respect to the the push forward of the Lebesgue measure m on X , and let λ be a probability measure absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^p . Let the sample $\{(s_1, [x_1]), \dots, (s_k, [x_k])\}, (s_i, [x_i]) \in \mathbb{R}^p \times X/T$ be sampled from $\lambda \times \eta$.*

Assume that the AAC for Regression (Algorithm 4) fits the regression model f_β defined in Definition 4.1.3. Assume moreover that the basis functions $\phi_j: \mathbb{R}^p \rightarrow \mathbb{R}$ satisfy the following properties:

i) $\phi_0 := 1$

ii) Sample s from λ and let $(\beta_0, \dots, \beta_p) \neq (\tilde{\beta}_0, \dots, \tilde{\beta}_p)$. Then, with probability 1,

$$\sum_{j=0}^p \beta_j \phi_j(s) \neq \sum_{j=0}^p \tilde{\beta}_j \phi_j(s).$$

iii) The matrix

$$\Phi(S) = \begin{bmatrix} 1 & \phi_1(s_1) & \dots & \phi_p(s_1) \\ \vdots & \ddots & & \vdots \\ 1 & \phi_1(s_k) & \dots & \phi_p(s_k) \end{bmatrix}$$

has full rank.

Under these circumstances, we claim that

a) The AAC algorithm terminates in finite time, and

b) With probability 1, the estimated regression curve f_β returned by the AAC algorithm is a local minimum of the function

$$\beta \mapsto \sum_{i=1}^k d_{X/T}^2([x_i], f_\beta(s_i)). \quad (4.7)$$

Note that the assumptions *i)* and *ii)* are reasonable and hold both for the linear basis functions used in ordinary least square regression model, as well as for other type of basis function such as polynomial basis functions. In Appendix 4.3, the proof of the theorem is reported. In the Appendix 4.3, the above theorem is declined to the case where the $h(s)$ belongs to Class II as in Definition 4.1.2.

Case Studies

In this section, we show the potential of the model applied to two real datasets. To understand the AAC approach, we will compute at each iteration of Algorithm 4 two errors: the *Regression Error* and the *Post Alignment Error*. The *Regression Error* is the within sample prediction error at step m . It is computed as the distance between the prediction along the regression line at step m (i.e. $h_m(s_i)$) and the observations used to fit the current regression $t_{i(m)}x_i$:

$$\sum_{i=1}^k d_X^2(h_m(s_i), t_{i(m)}x_i) \quad (4.8)$$

The *Post Alignment Error* is the distance between the prediction along the line at step m , and the graph representative optimally aligned with respect to this prediction as defined in Definition 4.5. Note also, that this distance coincides with the distance in Graph Space X/T between the graphs represented by $h_m(s_i)$ and x_i , where the graph $[h_m(s_i)]$ coincides with $f_m(s_i) = \pi(h_m(s_i))$.

$$\sum_{i=1}^k d_X^2(h_m(s_i), t_{i(m+1)}x_i) = \sum_{i=1}^k d_{X/T}^2([h_m(s_i)], [x_i]) \quad (4.9)$$

In other words, this is an intrinsic residual between the Graph Space regression model f and the observation $[x_i] \in X/G$. Note also that the aligned points obtained at step m are the points used to fit the regression line at step $m + 1$. The AAC for Regression is implemented as a method in the *GraphSpace* python package (Calissano et al., 2020c). In the following experiments, the algorithm step is the difference between the coefficients of determination of the regression models, estimated at the current and previous iteration.

Crypto-currencies correlation networks

The analysis of how the stock market evolves in time is a broadly discussed and complex problem, and the correlation networks are commonly used to model currency interdependencies (Mizuno et al., 2006). We collect the prices in USD of the nine cryptocurrencies (Bitcoin, Dash, Digibyte, Dogecoin, Litecoin, Vertcoin, Stellar, Monero, Verge) from the first recorded price of bitcoin on July 18th 2010, until April 3rd 2020 (from the repository Coins). Based on the price data, we compute correlation networks describing how pairs of crypto-currencies varies in price over time by computing, for every 7 days, their correlation over a time period of the following 20 days. In this way, we obtain a set of 506 correlation matrices, split into a training set of 400 matrices and a test set of 106.

4.1. Graph-Valued Regression

The convergence of the algorithm is illustrated in Figure 4.2, where the plot shows that more than one step is required to converge to the optimal aligned final model. The necessity of using an iterative algorithm such as the AAC is clear from the decreasing tendency of the errors – it would not suffice to align all the graphs with a representative of the mean and carry out a single regression model in X . This would suffer from the same distorted residuals as tangent space regression for manifold data.

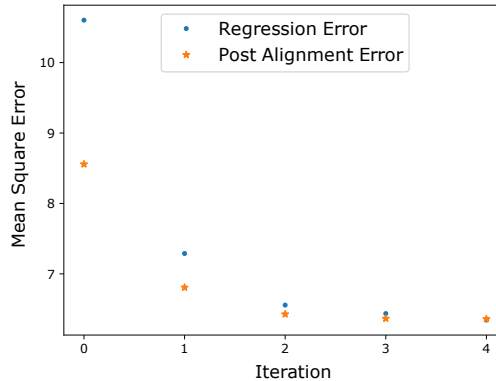


Figure 4.2: *The Regression Error and Post Alignment Error are computed at every iteration of the model estimation procedure. At the final step, $R^2 = 0.381$. The plot shows that the model converges within relatively few iterations, but also that a single iteration, analogous to the tangent space approach from manifold statistics, would not have been sufficient.*

In Figure 4.3, we show, for each crypto-currency, the frequency of permutation with other crypto-currencies in the analysis. This carries information on which crypto-currency has a more interchangeable or a unique role in the market. We see that Bitcoin, which appeared significantly earlier than the other crypto-currencies, is very rarely interchanged with the rest. Similarly, Dash, Vertcoin and Monero are interchanged at noticeable rates. In Figure 4.4, we see the predicted correlation networks at a number of different test time points, along with the ground truth network. The networks are illustrated both as adjacency matrices (heat-maps) and plotted as networks for intuitive comparison; note that while the nodes may be permuted compared to their original order, the prediction and ground truth are aligned and thus comparable.

Public Transport and Covid-19 in Copenhagen, Denmark

In this example, we analyse the public transport mobility networks in Copenhagen (Denmark) during the various phases of the Covid-19 epidemic in 2020. The mobility networks are derived from the *Rejsekort* (travel card) data provided by *Trafikselskabet Movia* - the

Chapter 4. Prediction in the Graph Space

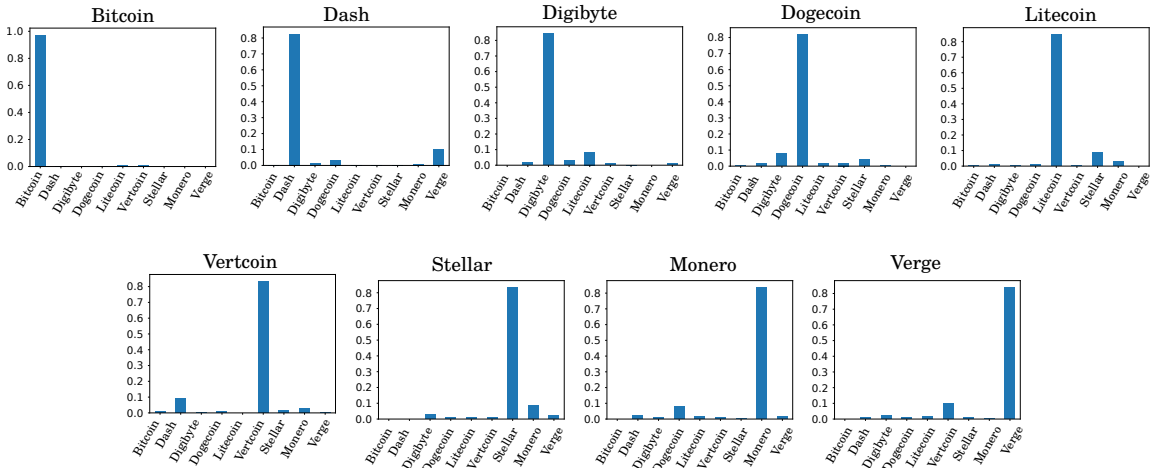


Figure 4.3: Permutation frequencies for the different cryptocurrencies.

bus company operating in the Copenhagen Region. The travel cards register the check in and check out on the buses of the travellers, along with the corresponding bus stops. Our mobility networks are based on trips between the 27th of February and the 13th of May, 2020. These transport data are modelled as the daily origin-destination matrices between the 10 different areas in the municipalities of Copenhagen and Frederiksberg. As shown in Figure 4.5, all the bus stops belonging to an area are aggregated into a single network node. The edges correspond to the number of people travelling between the areas during one day (00 : 01 to 23 : 59). The regression model describes the relationship between the origin destination networks and the categorical variable indicating the Covid-19 lock-down phases in Denmark. After the first registered case the 27th of February 2020, Denmark imposed a lock-down from the 13th of March to the 13th of April. During this month, the majority of the activities such as offices, gyms, and pubs were closed. During the following *Phase II*, a slow reopening has been taken place. The time regressors is modelled using three categorical variables describing the different phases. Note that this regression problem corresponds to an Anova problem, where the analysis of variance is conducted on the set of origin destination networks as a function of a three level categorical variable describing the lockdown phases. The regression is conducted both in the X space (i.e. without node permutation) and in the X/T space (i.e. with node permutation). The two analyses address two different research questions. In the X space regression, every neighbourhood maintains its own label. It is clear from the predicted trips in Figure 4.6, that the bus usage during lock-down almost disappeared without a full recover during the *Phase II*. If the regression is conducted on the X/T space, the neighbourhoods become interchangeable by allowing node permutation. In Figure 4.7, the network after the lock-down shows the same structure as the one of the pre-lock-down, but with an overall lower

4.1. Graph-Valued Regression

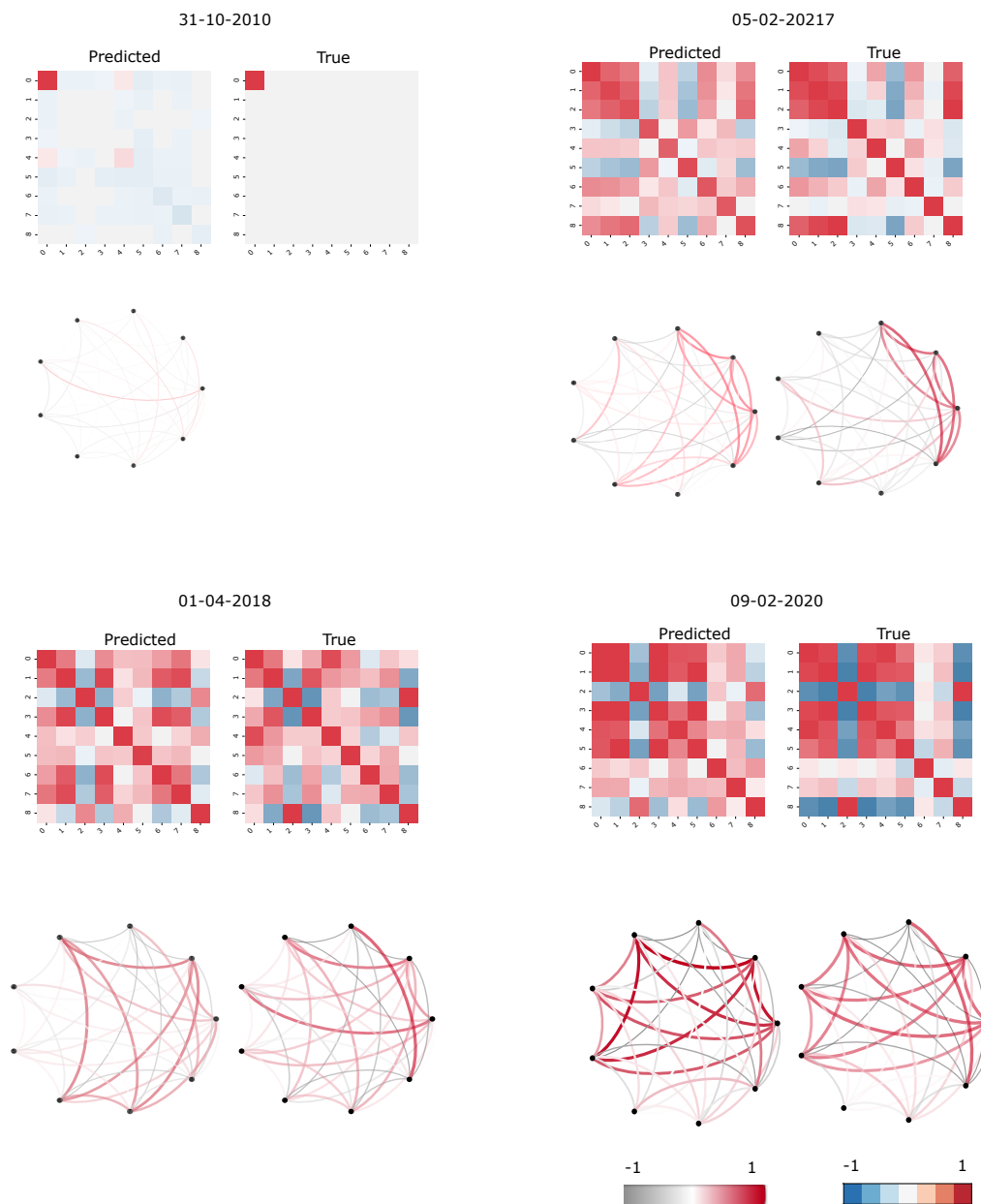


Figure 4.4: True and Predicted crypto-currencies represented as heat-maps and networks

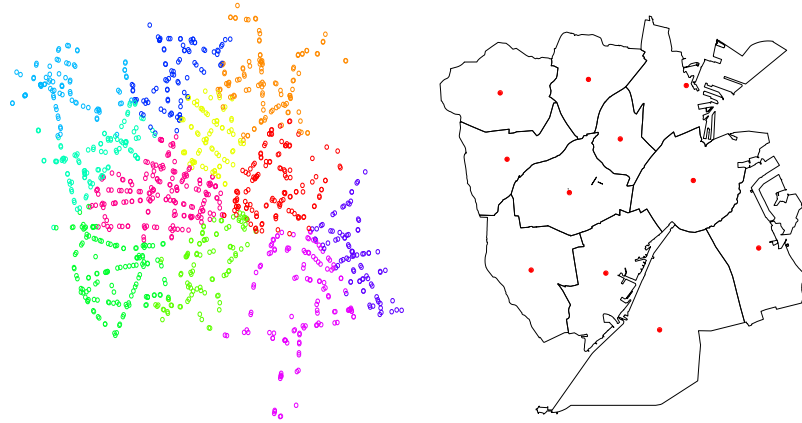


Figure 4.5: Bus stops in the different areas of Copenhagen and Frederiksberg.

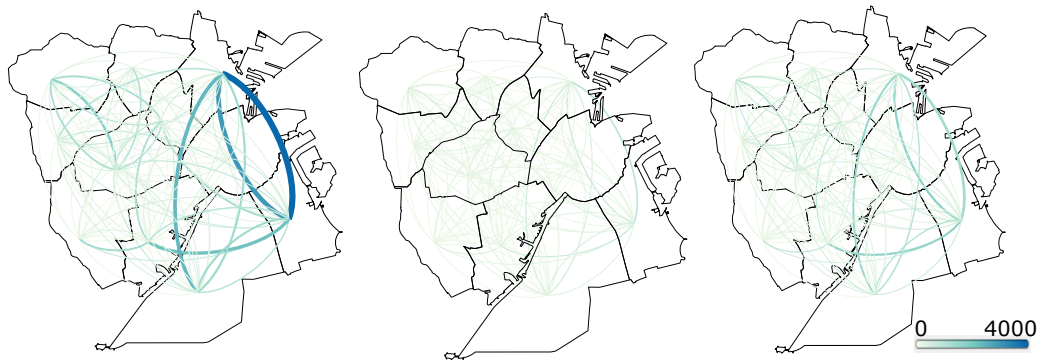


Figure 4.6: Prediction of the labelled networks (i.e. the nodes attributes) of three days randomly sampled from the three periods: 03/03/2020, 12/04/2020, and 22/04/2020

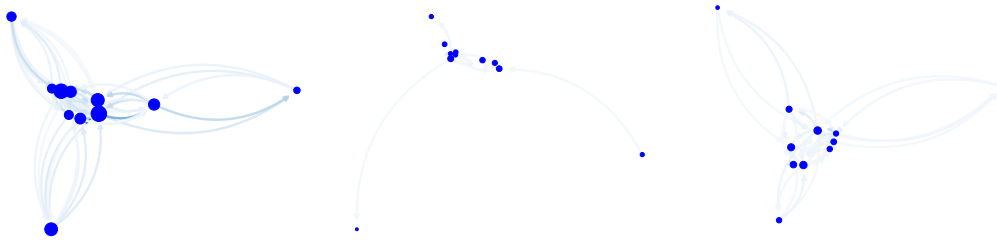


Figure 4.7: Prediction of the unlabelled networks of three days randomly sampled from the three periods: 03/03/2020, 12/04/2020, and 22/04/2020. The dimension of the node is proportional to the within area flux. The position of the nodes is computed using Spectral Layout of the Networkx python package.

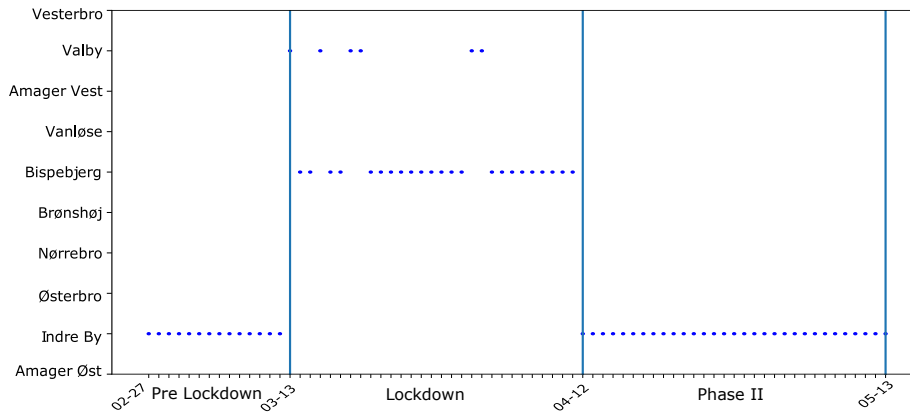


Figure 4.8: Most popular optimal matching with the node *Indre By* as a function time. Vertical Lines describe transitions between different phases of the lockdown.

intensity of trips. While the regression in the X space allows for the interpretation of each neighbourhood role in the network over time, the regression in the X/T space offers an unlabeled network vision, focusing on which role each area is playing in the whole system at every time step. To better understand this idea, in Figure 4.8 we represent the permutations along time of *Indre By*. *Indre By* is the shopping central area in Copenhagen. While before and after the lock-down, its role in the network is unique, during the lock-down it becomes interchangeable with Valby and Bispebjerg. These last two neighbourhoods are mostly residential areas. The permutations show how the city usage drastically changed during the emergency. As soon as the Phase II started, the unique role of *Indre By* in the city network is immediately reestablished.

4.2 Conformal Prediction Intervals

After the definition of a regression model for graphs, we move to the construction of a methodology for the uncertainty quantification. First, we introduce the concepts of conformal prediction in a labelled setting (Subsection 4.2). Second, we extend it in an unlabelled setting (Subsection 4.2).

The content of this section is extracted from the work Calissano et al. (2020d).

Conformal Prediction Parallelotopes for Labelled Graphs

Consider an i.i.d. population of graphs $X_1, \dots, X_k, x_i \in X$, sampled from a distribution \mathbb{P} . The problem we want to tackle is define an interval for an estimator. Formally, we define a prediction set $\mathcal{C}_{k,1-\alpha} := \mathcal{C}_{k,1-\alpha}(X_1, \dots, X_k)$ as

$$\mathbb{P}(X_{k+1} \in \mathcal{C}_{k,1-\alpha}) \geq 1 - \alpha \quad (4.10)$$

where $\alpha \in [0, 1]$. Let $X(j), j = 1, \dots, J^1$ be a generic element of the flattened adjacency matrix X , where $J = n^2$ since the attributes are scalars, and we are working with weighted adjacency matrices. As we said previously, $X(j) \in \mathbb{R}$, and the measure on X is the Lebesgue measure defined on the Borel σ -algebra. With respect to a univariate setting, the case of formulating prediction sets for complex data poses a serious question in terms of interpretability and practical usefulness of the obtained intervals. It is intuitive to understand that the best case in terms of interpretability for a prediction set is a region in space that allows a component-wise identification of an element that is inside or outside the prediction region. In more mathematical terms, we are interested in a set defined as:

$$\mathcal{C} := \{X \in X : X(j) \in \mathcal{C}(j), \quad \forall j \in 1, \dots, J\}, \quad (4.11)$$

where $\mathcal{C}(j) \subseteq \mathbb{R}$. The sets described in Equation 4.11 are the Cartesian product of J intervals of the real line. A set like this forms a parallelotope in \mathbb{R}^J

The prediction set in the shape of a parallelotope allows a practitioner to project the multivariate prediction region which is valid at a level α , in intervals for each element of X without changing the coverage level. Our applied goal is thus to identify parallelotope-shaped sets with a given unconditional coverage level, namely: $\mathbb{P}(X_{k+1} \in \mathcal{C}_{k,1-\alpha}) \geq 1 - \alpha$. A method which has the explicit aim to identify prediction sets of the type described in Equation 4.11 is the Conformal Prediction Method (Vovk et al., 2005; Zeni et al., 2020).

¹The index $j = 1, \dots, J$ was used to refer to $j = 1, \dots, p$ - regressors - in Section 4.1, to be coherent with the regression literature.

The key quantity around which a Conformal Prediction framework revolves is a so called conformity (or non-conformity) measure, on which a very weak “quasi-model” (in the sense of Cella and Martin, 2020) is imposed, and it allows to obtain prediction sets. By aptly choosing a non-conformity measure with the desired iso-contours, one is able to obtain prediction sets with the minimal size and/or with the desired shape.

Let us put ourselves in a Split/Inductive Conformal Framework (see Lei and Wasserman (2014); Papadopoulos et al. (2002) for an introduction). We start by splitting our X_1, \dots, X_k in a proper *training set* \mathcal{I}_1 and a *calibration set* \mathcal{I}_2 , where $|\mathcal{I}_1| + |\mathcal{I}_2| = k$, indexing the sets as: $l \in \mathcal{I}_1$ and $m \in \mathcal{I}_2$.

$\forall X_m, m \in \mathcal{I}_2$, one can compute an empirical P-value:

$$p_{X_m} := \frac{|\{i \in \mathcal{I}_2 : R_i \geq R_m\}|}{|\mathcal{I}_2| + 1}$$

where $R : X^k \rightarrow \mathbb{R}$ is a non-conformity measure as defined in Vovk et al. (2005). The conformal prediction set defined using the above definition of P-value is:

$$\mathcal{C}_{k,1-\alpha} := \{X \in X : p_X > \alpha\} \quad (4.12)$$

where α is the desired coverage level. To identify a prediction set that is also a set in the sense of Equation 4.11, we can use the L_{inf} metric. Namely, we define our non-conformity measure (NCM) $R : X^k \rightarrow \mathbb{R}$ to be

$$R_m = \max_{j \in \{1, \dots, J\}} |X_m(j) - \hat{\mu}(j)|, \quad m \in \mathcal{I}_2 \quad (4.13)$$

where $\hat{\mu} = \mathcal{A}(\{X_l, l \in \mathcal{I}_1\})$ is an estimator of central tendency based on a algorithm \mathcal{A} , trained on $X_l, l \in \mathcal{I}_1$. We can also note that, having defined a prediction set as in Equation 4.18, and the NCM as in Equation 4.13, one can say that $X_{k+1} \in \mathcal{C}_{k,1-\alpha} \iff R_{k+1} \leq h$, with h the $[(|\mathcal{I}_2| + 1)(1 - \alpha)]$ -th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$. Then

$$\begin{aligned} & \max_{j \in \{1, \dots, J\}} |X_{k+1}(j) - \hat{\mu}(j)| \leq h \\ \Rightarrow & |X_{k+1}(j) - \hat{\mu}(j)| \leq h \quad \forall (j) \\ \Rightarrow & X_{k+1}(j) \in [\hat{\mu}(j) - h, \hat{\mu}(j) + h] \quad \forall j = 1, \dots, J \end{aligned}$$

Therefore, the split conformal prediction set induced by the nonconformity measure (4.13) is

$$\mathcal{C}_{k,1-\alpha} := \{X \in X : X(j) \in [\hat{\mu}(j) - h, \hat{\mu}(j) + h] \quad \forall j = 1, \dots, J\}. \quad (4.14)$$

The described procedure is summarised in Algorithm 5. The calculation of these sets is very convenient: we require to train the central tendency estimation algorithm \mathcal{A} only once, and we have a closed form for the calculation of the semi-amplitude of the set.

Chapter 4. Prediction in the Graph Space

Algorithm 5 Split Conformal Prediction Parallelotopes for Populations of Graphs

- 1: **Require:** Data $X_i, i = 1, \dots, k,$, type-1 error level $\alpha \in (0, 1)$, central tendency estimation algorithm \mathcal{A}
 - 2: split randomly $\{1, \dots, k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
 - 3: $\hat{\mu} = \mathcal{A}(\{X_l, l \in \mathcal{I}_1\})$
 - 4: $R_m = \max_{j=1, \dots, J} (X_m(j) - \hat{\mu}(j)), m \in \mathcal{I}_2$
 - 5: h is the $[(|\mathcal{I}_2| + 1)(1 - \alpha)]$ -th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
 - 6: **Ensure:** $\mathcal{C}_{k, 1-\alpha} := \{X \in X : X(j) \in [\hat{\mu}(j) - h, \hat{\mu}(j) + h] \quad \forall j\}$
-

Amplitude Modulation

The main shortcoming of the approach proposed in Section 4.2 is that the identified parallelotope has constant amplitude across all $j = 1, \dots, J$. While there may be situations in which such feature is desirable, practitioners usually face cases in which edge attributes have different variability, and may want to take into account such variability when making a global prediction (for instance, with wider or narrower sets). The limit situation in this case would be the one where a vertex is completely absent from the population of graph analysed, meaning that $X(j) = 0$ for the position j assigned to the null node. Any amplitude different from 0 for the interval in $B(j)$ is not desirable, since we do not want to “add” a node that never appears in the observed data.

To do so, following Lei et al. (2018), we condition the amplitude of Equation 4.14 across $j \in 1, \dots, J$ using a local notion of variability. Namely, we modify Equation 4.13 in the following fashion:

$$R_m = \max_{j \in 1, \dots, J} \left[\frac{|X_m(j) - \hat{\mu}(j)|}{\hat{s}(j)} \right] \quad (4.15)$$

Where $\hat{s} = \mathcal{S}(\{X_l, l \in \mathcal{I}_1\})$ is an estimator of local variability, trained on the set $\{X_l, l \in \mathcal{I}_1\}$ using the symmetric algorithm \mathcal{S} . Please note that the modulation function is computed on the training set only. We summarise the Split Conformal Procedure with Modulation in Algorithm 6

According to the choice of the algorithm to compute \mathcal{S} , the modulating behaviour dramatically changes. We mention two notable cases: (i) When $\hat{s}(j) = 1 \forall j$ and no modulation is taking place, Algorithm 6 will yield results equal to Algorithm 5; (ii) If $\hat{s}(j) = \sqrt{\text{Var}(X(j))}$ the resulting set amplitude will be modulated according to the local variability of the attributes of the graph.

Remark. Consider the case when $\hat{s}(j) = \sqrt{\text{Var}(X(j))}$, if an attribute of an element of the graph is deterministic, the attentive reader can immediately understand that both the numerator and the denominator of the non-conformity function will be equal to zero. This yields to an indeterminate form. To solve this computational problem, we will use, as a

4.2. Conformal Prediction Intervals

Algorithm 6 Split Conformal Prediction Parallelotopes for Populations of Graphs with Amplitude Modulation

- 1: **Require:** Data $X_i, i = 1, \dots, k,$, type-1 error level $\alpha \in (0, 1)$, regression algorithm \mathcal{A} , amplitude modulation algorithm \mathcal{S}
- 2: split randomly $\{1, \dots, k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
- 3: $\hat{\mu} = \mathcal{A}(\{X_l, l \in \mathcal{I}_1\}), \hat{s} = \mathcal{S}(\{X_l, l \in \mathcal{I}_1\})$
- 4: $R_m = \max_{j=1, \dots, J} \left(\frac{|X_m(j) - \hat{\mu}(j)|}{\hat{s}(j)} \right), m \in \mathcal{I}_2$
- 5: $h =$ is the $\lceil (|\mathcal{I}_2| + 1)(1 - \alpha) \rceil$ -th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
- 6: **Ensure:** $\mathcal{C}_{k, 1-\alpha} := \{X \in X : X(j) \in [\hat{\mu}(j) - h\hat{s}(j), \hat{\mu}(j) + h\hat{s}(j)] \quad \forall j\}$

modulation function $\hat{s}(j) = \sqrt{\text{Var}(X(j))} + \epsilon$, where ϵ is a very small constant. This yields to a prediction interval for a deterministic values centered in the actual value of the constant, and of negligible length, while allowing the computation of the global set to be performed.

Prediction Class of Parallelotopes for Unlabelled Graphs

In this section, we extend the concept previously introduced for $\{X_1, \dots, X_k\} \in X$ to $\{[X_1], \dots, [X_k]\} \in X/T$. We recall the definition of probability measure on Graph Space, already introduced in Theorem 3.3.5.

Definition 4.2.1. *The Graph Space X/T is endowed with a probability measure η which is absolutely continuous with respect to the push-forward of the Lebesgue measure m on X . In particular, for $A \subset X/T$, we have $\eta(A) = 0$ if $m(\pi^{-1}(A)) = 0$.*

An interval in the quotient space becomes:

Definition 4.2.2. *A set $\mathcal{C} \in X$ as defined in Equation 4.11 can be projected on the Graph Space as:*

$$[\mathcal{C}] = \bigcup_{t=1}^{|\mathcal{T}|} \bigtimes_{j=1}^J \mathcal{C}(\sigma_t(j)), \quad [\mathcal{C}] \subseteq X/T \quad (4.16)$$

where $\sigma_t : \{1, \dots, J\} \rightarrow \{1, \dots, J\}$ is the relabelling function associated to the permutation $t \in T$

The idea is to define a set of intervals that follow the index permutation of the elements in the Graph Space. For the sake of simplicity we define:

$$\mathcal{C}^t = \bigtimes_{j=1}^J \mathcal{C}(\sigma_t(j))$$

Chapter 4. Prediction in the Graph Space

where $\mathcal{C}(\sigma_t(j)) \subseteq \mathbb{R}$, $\mathcal{C}^t \subseteq \mathbb{R}^J$.

The probability of this interval in the Graph Space can be computed using the projection on the total space X :

$$\mathbb{P}_\eta \left(\bigcup_{t=1}^{|\mathcal{I}|} \mathcal{C}^t \right) = \sum_{t=1}^{|\mathcal{I}|} \left((-1)^{t-1} \sum_{\substack{I \subseteq \{1, \dots, |\mathcal{I}|\} \\ |I|=t}} \mathbb{P}(A_I) \right)$$

where $A_I := \bigcap_{t \in I} \mathcal{C}^t$

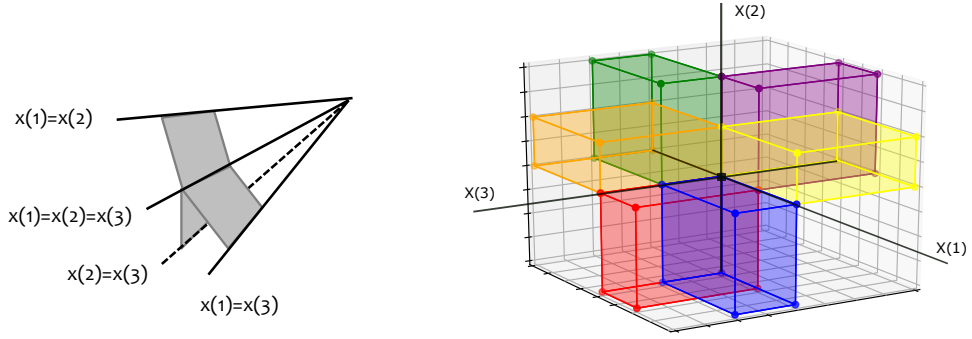


Figure 4.9: Conceptual visualization of the shape of an interval in the Graph Space and its back-projection π^{-1} in the total space for an undirected graph with three nodes and no attributes on the nodes.

Example. Consider an undirected graph with three nodes $n = 3$, real attributes on edges and no attributes on nodes. This graph can be described as a point in \mathbb{R}^3 . The number of permutation is $3! = 6$. The interval \mathcal{C} is the Cartesian product of three intervals on the real line $\mathcal{C} = \mathcal{C}(1) \times \mathcal{C}(2) \times \mathcal{C}(3)$. If we permute this shape with the following permutation $t = \{2, 1, 3\}$, we obtain a new set $\mathcal{C}^t = \mathcal{C}(\sigma_t(1) = 2) \times \mathcal{C}(\sigma_t(2) = 1) \times \mathcal{C}(\sigma_t(3) = 3)$. The union of all their possible permutations is $[\mathcal{C}]$, shown in Figure 4.9

As in the labelled case, we can define the interval with a given coverage level:

$$\mathbb{P}_\eta ([X]_{k+1} \in [\mathcal{C}]_{k,1-\alpha}) \geq 1 - \alpha \quad (4.17)$$

We start by splitting our set of unlabelled graphs $\{[X_1], \dots, [X_k]\}$ in a training set \mathcal{I}_1 and a calibration set \mathcal{I}_2 , where $|\mathcal{I}_1| + |\mathcal{I}_2| = k$.

$\forall [X_m], m \in \mathcal{I}_2$, one can compute an empirical P-value define exactly as in the labelled case:

$$p_{[X_m]} := \frac{|\{i \in \mathcal{I}_2 : R_i \geq R_m\}|}{|\mathcal{I}_2| + 1}$$

4.2. Conformal Prediction Intervals

where R is a non-conformity measure as defined by Vovk et al. (2005). The conformal prediction set defined using the above definition of P-value can be identified as

$$[\mathcal{C}]_{k,1-\alpha} := \{[X] \in X/T : p_{[X]} > \alpha\} \quad (4.18)$$

We define our non-conformity measure R_m to be

$$R_m = \max_{j=1,\dots,J} |t_m X_m(j) - \mathcal{A}(j)| \quad (4.19)$$

where:

$$t_m = \arg \min_{t \in T} (d_X(tX_m, \mathcal{A})) \quad (4.20)$$

where \mathcal{A} is a symmetric function of the data. This non-conformity measure selects the permutation that optimally aligns the two graphs and consequently selects the edge or node that are mostly far apart from each other.

In the case of amplitude modulation, the Equation 4.19 becomes:

$$R_m = \max_{j=1,\dots,J} \frac{|t_m X_m(j) - \mathcal{A}(j)|}{\hat{s}(j)} \quad (4.21)$$

Where \hat{s} is an estimator of the variability of the edge or node j after the alignment with respect to the central estimator \mathcal{A} . The whole procedure is summarised in Algorithm 7.

Algorithm 7 Split Conformal Prediction Parallelotopes for Populations of Unlabelled Graphs with Amplitude Modulation

- 1: **Require:** Data $[X_i]$, $i = 1, \dots, k$, type-1 error level $\alpha \in (0, 1)$, Predictive algorithm \mathcal{A} , amplitude modulation algorithm s
 - 2: split randomly $\{1, \dots, k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
 - 3: $\hat{\mu} = \mathcal{A}(\{[X_l], l \in \mathcal{I}_1\})$, $\hat{s} = \mathcal{S}(\{[X_l], l \in \mathcal{I}_1\})$
 - 4: Find $\{t_1, \dots, t_{|\mathcal{I}_2|}\}$ s.t. $t_m = \arg \min_{t \in T} (d_X(tX_m, \hat{\mu}))$
 - 5: $R_m = \max_{j=1,\dots,J} \left(\frac{|(t_m X_m)(j) - \hat{\mu}(j)|}{\hat{s}(j)} \right) = \max_{j=1,\dots,J} \left(\frac{|X_m(\sigma_{t_m}(j)) - \hat{\mu}(j)|}{\hat{s}(j)} \right)$, $m \in \mathcal{I}_2$
 - 6: h is equal to $\lceil (|\mathcal{I}_2| + 1)(1 - \alpha) \rceil$ -th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
 - 7: **Ensure:** $\mathcal{C}_{k,1-\alpha} := \{[X] \in X/T : (tX)(j) \in [\hat{\mu}(j) - h\hat{s}(j), \hat{\mu}(j) + h\hat{s}(j)] \quad \forall j \quad \forall t\}$
-

Remark. The reader should note how it is not required to specify anything about \mathcal{A} . This generality, which is one of the main interesting features of the conformal prediction framework, allows for the use of any predictive algorithm, either statistical inspired, machine-learning inspired, or a combination of the two. However, due to the geometrical complexity of the Graph Space, extending regression strategies as well as neural network strategy to this framework is not straightforward. In this section, we are going to use the Fréchet Mean as the \mathcal{A} (see (Calissano et al., 2020a) for definitions and details).

Simulation and Case Study



In this section, we illustrate the theoretical results described in the previous section on two simulated dataset and one case studies. In all these examples, the \mathcal{A} function is going to be a Fréchet Mean estimator. In the unlabelled case, the Fréchet mean corresponds to the Sample Mean, while in the unlabelled case, the Fréchet Mean is computed with the Align All and Compute Procedure (Calissano et al., 2020a). The Conformal Prediction Parallelotopes is implemented as a function in the Python Package *Graph Space* (Calissano et al., 2020c).

Simulation: Labelled Case

In this simulation, we compute the Empirical Coverage of different parametric intervals and the conformal prediction intervals. We generated a set of graphs 130 graphs ($|\mathcal{I}_1| + |\mathcal{I}_2| = 30$ is the training set - eventually divided in training and calibration for the split conformal method- and 100 is the test set). Every directed graph has 5 nodes with Gaussian attributes $N(0, 1)$ and 20 edges following four different distributions:

1. Gaussian attributes $N(0, 1)$,
2. Uniform attributes $U(-1.7, 1.7)$,
3. t-Student attributes with 4 degrees of freedom,
4. t-Student attributes with 1 degree of freedom,

Having two different distributions on nodes and edges attributes is very common in the applications, because nodes and edges usually describe two different phenomena. For every generated model, we compute the sample mean and three different prediction intervals, in a labelled fashion:

1. Univariate Gaussian Intervals:

$$\bar{x}_i \pm t_{k-1}(\alpha/2) \sqrt{1 + \frac{1}{k} \hat{s}_i}$$

2. Univariate Gaussian Intervals with Bonferroni Correction:

$$\bar{x}_i \pm t_{k-1}((\alpha/p)/2) \sqrt{1 + \frac{1}{k} \hat{s}_i}$$

3. Simultaneous Gaussian Intervals:

$$\bar{x}_i \pm \hat{s}_i \sqrt{\left(1 + \frac{1}{k}\right) \frac{(k-1)p}{(k-p)} F_{(p,k-p)}(\alpha)}$$

4. Gaussian Ellipse

$$(x - \bar{x})' \hat{S}^{-1} (x - \bar{x}) \leq \left(1 + \frac{1}{k}\right) \frac{(k-1)p}{(k-p)} F_{(p,k-p)}(\alpha)$$

where $t_{k-1}(\alpha/2)$ and $F_{(p,k-p)}(\alpha)$ denotes the upper quantile. \hat{S} denotes the sample covariance matrix and \hat{s}_i the estimated sample variance; $\bar{x} = [\bar{x}_1, \dots, \bar{x}_p]$ the sample vector mean.

Given a set of 100 different different α and 100 rounds of simulations, we compute the average empirical coverage on the test set, defined as:

$$\hat{E}(1 - \alpha) = \frac{1}{100} \sum_{k=1}^K \sum_{i=1}^{100} \frac{1_{x_i \in IC_k(\alpha)}}{100} \tag{4.22}$$

In the Figures 4.10, we show the calibration curves α for the the different generative models and the different intervals. As expected, and coherently with the theory, Univariate

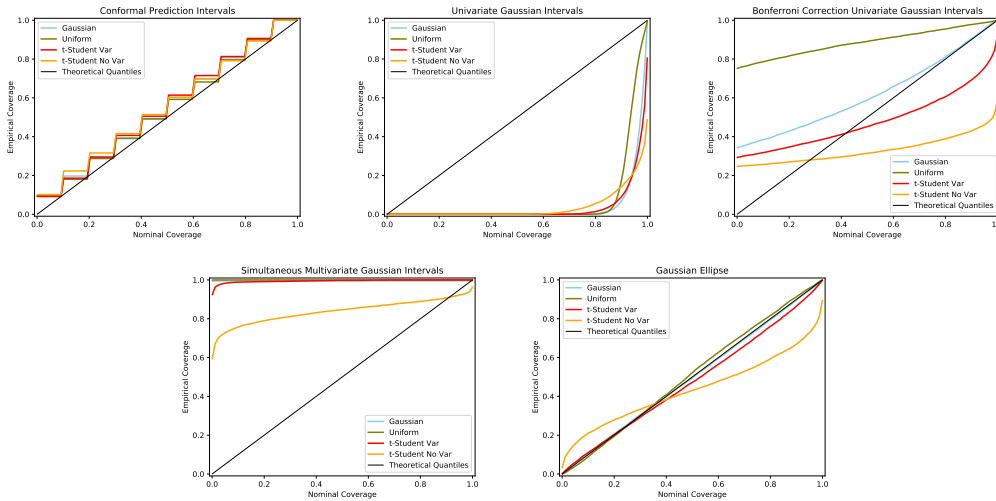


Figure 4.10: Empirical Coverage as a function of the Nominal Coverage for the different models. Each one of the plot visualizes a different interval.

gaussian intervals are wildly under-covering in a global sense, given the absence of any

Chapter 4. Prediction in the Graph Space

multiplicity correction. Bonferroni-corrected intervals are quite conservative in the gaussian case, and very conservative in leptocurtic cases such as the uniform one. With respect to the two platycurtic cases (t-student with 1 degree of freedom and t-student with 4 degrees of freedom), we observe a generic conservativeness for low levels of nominal coverage, generated by the Bonferroni correction: this effects tends to disappear for higher coverage levels. We see that Bonferroni-corrected intervals are under-covering for nominal levels that are commonly used in the practice. The projection over the components of multivariate gaussian intervals generates, similarly to our method, prediction sets with the shape of a parallelotope: It appears evident how they are grossly conservative. Multivariate gaussian prediction ellipses are exact in the gaussian case, and provide a conservative approximation in the uniform case, while they fail to cover for platycurtic distributions. In any case a high-dimensional prediction ellipse such as this one is of very relative practical value. The only method providing a properly calibrated prediction intervals - regardless of the distribution and interpretable (thanks to their parallelotopic shape)- is the proposed conformal one. In the Conformal Prediction Intervals, if the size of the calibration and training sets increases, the number of possible estimated values of the Empirical Coverage also increases, thus the Empirical Coverage converges to the Nominal Coverage. This is clear from Figure 4.11, where the CP intervals for different numbers of observations are estimated: $|\mathcal{I}_1| + |\mathcal{I}_2| = 30, 130, 230$. The lines of the four different distributions tent to the quadrant bisector, when the number of observations increases.

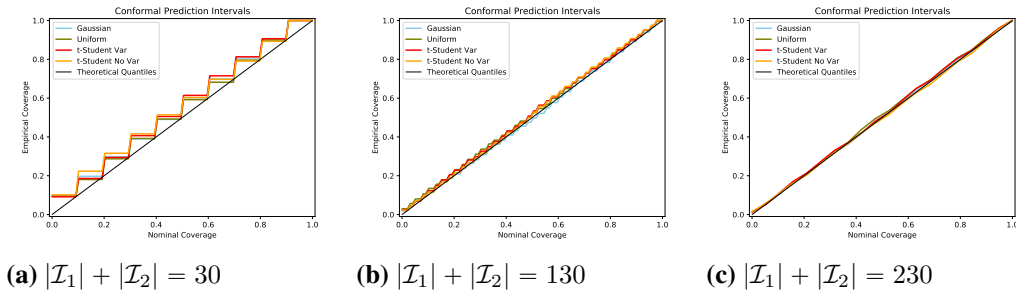


Figure 4.11: Estimations of the Empirical Coverage for the conformal prediction intervals with different calibration and training size. The size of the test set is 100 in all three simulations.

Simulation: Unlabelled Case

In this example, we simulated 500 pentagons from the equivalence classes shown in Figure 4.12. First, we randomly select one of the equivalent class and then we randomly pick one element from the class (i.e. a random permutation of the graphs shown in the figure). These pentagons have constant attribute on nodes (10) and decreasing attributes on edges (100, 80, 60, 40, 20). The interesting part of this example reside on the simplicity of the

graph topology and the possibility of visualizing and understanding the correct alignment as shown in Figure 4.12. Note that this dataset is the same as the one used in Section 3.3.

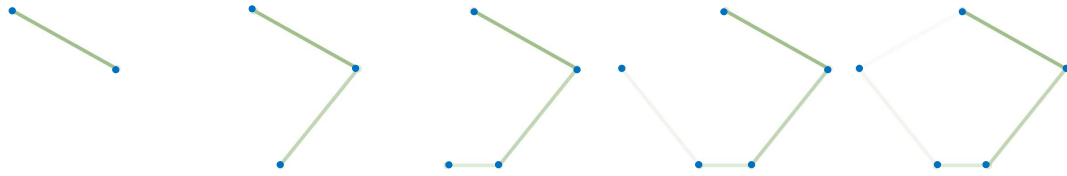


Figure 4.12: A candidate for each one of the equivalence classes of the unlabelled dataset.

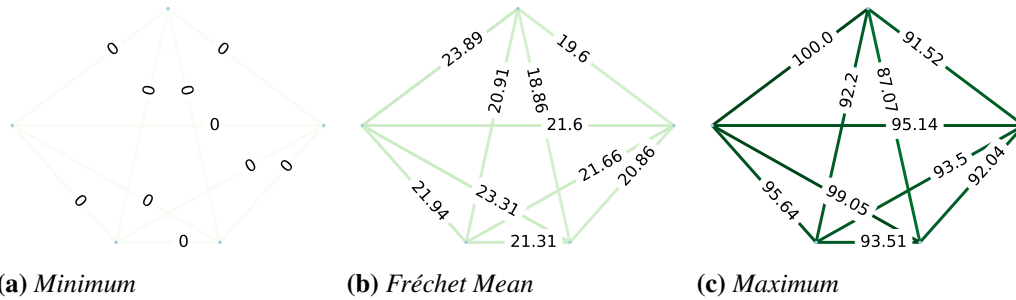


Figure 4.13: Labelled Setting: Conformal Prediction interval of level 95%.

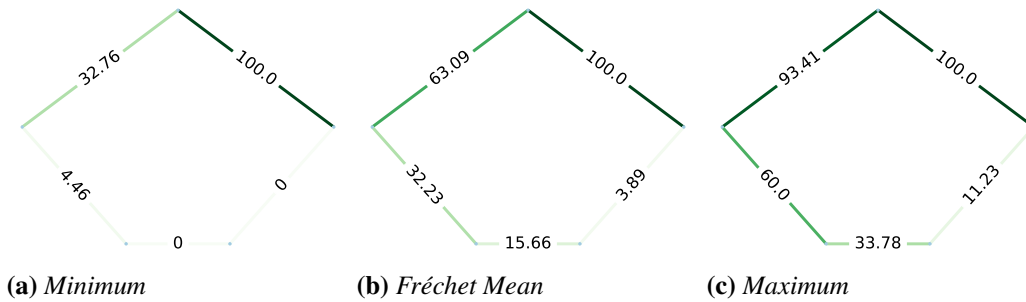


Figure 4.14: Unlabelled Setting Conformal Prediction interval of level 95%.

In Figures 4.13 and 4.14, we show the Fréchet Mean and the corresponding 95% intervals for both the labelled (no alignment procedure is applied to the data) and the unlabelled setting. The exact values of the intervals are visualized on top of each edge. The example shows the capability of the conformal prediction intervals to capture the topology and the attributes, when working in an unlabelled setting.

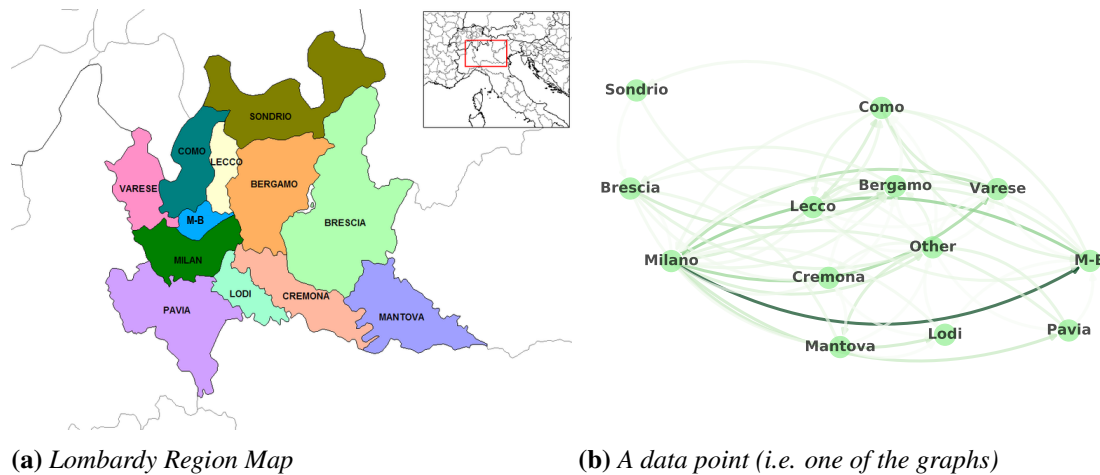


Figure 4.15: A map of the Lombardy Region with the 11 provinces (M-B stands for Monza-Brianza) and an example of a data point randomly sampled from the Origin-Destination Population.

Application: Mobility and Covid-19 in Lombardy Region

In this section, we apply the methodology to a case study concerning the mobility during the outbreak of the COVID-19 pandemic in the Italian region of Lombardy. Covid-19 hit with particular violence the northern part of Italy. The Italian government decided for a complete lockdown from the 21st of March 2020. The so-called *phase II* started the 4th of May 2020, with a slow reopening of the commercial activities. The data are provided by a location marketing company Cuebiq and consists of GPS locations gathered via smartphone of anonymous users in the Lombardy Region from the 17th of February to the 17th of May 2020. Among the anonymous users, we randomly sampled 50000. This anonymous data are collected from users who opt-in to share their data for research purposes, through a GDPR-compliant framework. Cuebiq then applies additional privacy preservation techniques to remove sensitive locations from the dataset, and to obfuscate personal areas such as home locations by “up-leveling” them to 600m x 600m geo-hash tiles. Data have been aggregated in Origin-Destination Matrices (ODM). ODMs - a standard data type commonly used in transport and mobility modelling - are graphs where the nodes are geographical locations and the edges are the flows of people between the locations. In this case study, we focus on the peoples’ trips arriving before 7 p.m. of the working days of the given period. The result is a dataset of 65 labelled graphs with 11 nodes. In Figure 4.15, the map of the 11 provinces of Lombardy and an example of a ODM is reported.

The conformal prediction intervals for the Fréchet Mean are computed at three different levels of $\alpha = \{0.25, 0.5, 0.75\}$, to shows how this conformal prediction strategy can be used to understand the distribution of a complex phenomena. In Figure 4.16, we report

4.2. Conformal Prediction Intervals

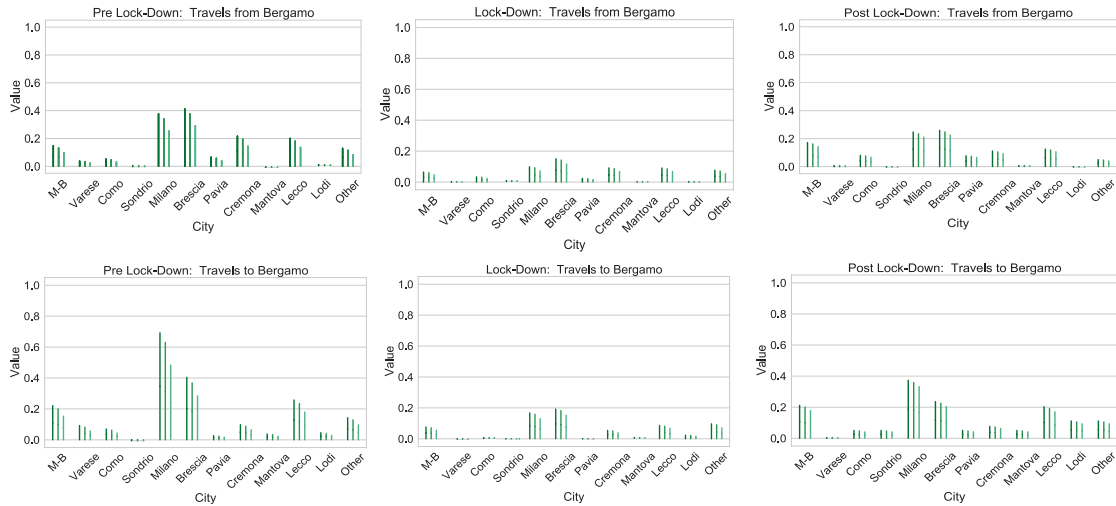


Figure 4.16: Intervals with $\alpha = 0.25, 0.5, 0.75$ for the outgoing and incoming edges of the province of Bergamo. Data has been standardized and negative values are set to zero.

the intervals of the incoming and outgoing edges from the province of Bergamo (cut for negative values and standardized due to privacy reasons). Bergamo was the province most hit by the COVID-19 epidemic in Lombardy and it is an important regional and national economic hub, being at the centre of a very industrialized area. As it is clear from the prediction intervals, the COVID lockdown decreased the mobility to and from Bergamo province and the mobility activities have not recovered in the *Phase II*. From a modelling perspective, the plots shows how the intervals sizes increase for higher values of $1 - \alpha$. Note that the prediction intervals are in dimension 144 and in the Figure 4.16 we are only showing the components along 11 axes, even if the coverage is in the higher dimension.

4.3 Discussion



In the first section of this chapter, we developed a graph-valued regression model, including scalar, basis functions, or vectors - as independent variables - and graph-valued datum - as dependent variable. The application of the model broads to every problem where a set of graphs varies according to an external value. To define an intrinsic regression model, we implemented the Align All and Compute Algorithm for Regression in the Graph Space by iteratively aligning points to the regression line and estimating the Multiple Output Least Square Regression as a regression model on the total space. The novelty of the AAC for Regression with respect to the AAC for GPCA (see Section 3.3) resides in making compatible the alignment procedure and the regression procedure. As a further development, this algorithmic framework can be extended to different regression strategies rather than the Multiple Output OLS regression, such as Gaussian Processes or Neural Networks. The requirement for these extensions is to check that the regression loss function and the optimization of the alignment procedure lead to the same minimum - i.e. getting the points closer to the geodesic and the geodesic closer to the points. Note that one of the main application of this framework is time series: a network evolving over time. Thus, a more tailored time series regression model can be defined for this specific application. The time series approach to networks could also be interpret as a network-on-network regression problem, where the inputs are the networks measured at $t - 1, \dots, t - k$ and the outputs are the networks measured at t . Defining a network-on-network regression problem is an interesting research question: this original regression should be equivariant with respect to the permutation action.

In the second section, we addressed the uncertainty quantification problem, proposing a model-free, computationally efficient forecasting method, based on Conformal Prediction. Due to the novelty of prediction intervals for graphs, we introduced the strategy for both labelled and unlabelled graphs. The issue of predicting with uncertainty a complex statistical units is a key research topic in modern statistics, from both a theoretical and an applied perspective. Only few works have been proposed on the topic, which focus on set forecasting techniques based on either distributional assumptions - that are hard to justify - or on heavy computational methods. In addition to that, they concentrate on producing forecasts for well-behaved objects, for which an embedding in an Euclidean or mildly non-Euclidean space is possible. The conformal prediction strategy instead creates intervals with the correct expected coverage in the quotient space and with a possible visualization of the intervals along the edges and the nodes, allowing for a possible interpretation of the results. A natural extension of this chapter is the application of conformal prediction intervals to the regression model, substituting the \mathcal{A} function with the devel-

oped regression model.

Two interesting case studies about the Covid-19 outbreak and the effect on urban movement are illustrated in this chapter. The first case study regards the public transport system in Copenhagen (Denmark). Thanks to the graph-valued regression, we showed how the system demand changed during the lockdown. The unlabelled perspective on the public transport system revealed the relational roles of the different areas and the evolution of the roles over time within the network. The second case study regards the analysis of GPS traces in the Lombardy Region (Italy), focusing on general urban movement (not only related to public transport). We showed how the distribution of the trips between the different provinces in the region changed over times, using conformal prediction as an exploration tools to visualize the distribution of a complex data. These two examples show the high potential impact of the developed statistical theory on the analysis of urban movements.

Appendix

Appendix 1: Proof of Theorem 4.1.6

Remark. Definition (4.1.3) describes a length $J = n^2$ vector (corresponding to a flattened $n \times n$ matrix) of linear regression models whose a^{th} entry is given by

$$h(s_i)_a = \sum_{j=0}^p \phi_j(s_i) \beta_j(a)$$

Note that the index a corresponds to either a node or an edge, and we thus have one regression model for each edge and each node, Generalized geodesic regression on X/T fits linear regression to aligned data point representatives in X , and corresponds to basis functions $\phi_0(s) = 1$ and $\phi_i(s) = s_{(i)}$ where $s_{(i)}$ denotes the i^{th} coordinate of s . However, the model also enables using more general nonlinear basis functions $\phi_i(s)$ as known from Euclidean statistics, leading to linear regression models with potentially nonlinear regressors, such as e.g. polynomial regression.

Proof. First, we prove convergence in finite time. Algorithm 4 consists of two steps repeated iteratively, fitting the generalised regression model 4.1.3 to the observations $\{(s_1, [x_1]), \dots, (s_k, [x_k])\} \in \mathbb{R} \times X/T$.

Consider the squared error loss function

$$\sum_{i=1}^k d_X^2(h_m(s_i), t_{i(m)}x_i), \tag{4.23}$$

where h_m is our current estimate of the regression model at step m in X , $h_m(s_i) \in X$ is the corresponding regression estimate corresponding to input s_i , and $t_{i(m)}x_i$ is the current representative in X of the sample network $[x_i]$. Note that the first step of Algorithm 4, which aligns output representatives of $[x_i]$ to the corresponding predicted value $h(s_i)$ along the current estimation of the regression line, cannot increase the value of (4.23) as an improved alignment would indeed lower it by definition. Similarly, the second step of Algorithm 4, which is the re-estimation of the generalized geodesic regression given the new alignments, also cannot increase the value of (4.23) as, again, an improved estimate would lower its value. Moreover, if the value of (4.23) stays fixed two iterations in a row, the algorithm will terminate. Thus, the iterative algorithm will never see the same set of sample-wise alignments twice without terminating. As there are only finitely many such alignments, the algorithm is forced to terminate in finite time.

Next, we turn to proving that the estimated regression model f_β is, indeed, a local minimum. We need to show that for some $\epsilon > 0$, $\|\tilde{\beta} - \beta\| < \epsilon$ implies that, with probability 1,

$$\sum_{i=1}^k d_{X/T}^2([x_i], f_\beta(s_i)) \leq \sum_{i=1}^k d_{X/T}^2([x_i], f_{\tilde{\beta}}(s_i))$$

We shall rely on the following lemma:

Lemma 4.3.1. *Given representatives $\{x_1, \dots, x_k\}$ of $\{[x_1], \dots, [x_k]\}$ with generalized linear regression model $f(s) = \pi \circ h(s)$ obtained minimizing (4.23), $h(s): \mathbb{R}^p \rightarrow X$, the following holds with probability 1:*

For all $i = 1, \dots, k$ and for all $t \in T \setminus T_{x_i}$,

$$d_X(h(s_i), x_i) \neq d_X(h(s_i), tx_i),$$

where T_{x_i} is the stabilizer $T_{x_i} = \{t \in T | tx_i = x_i\}$.

If the lemma holds, then we may define

$$\nu = \min\{d_X(h(s_i), tx_i) - d_X(h(s_i), x_i) \mid i = 1, \dots, k, t \in T \setminus T_{x_i}\} > 0.$$

Since the map $\beta \mapsto h_\beta(s)$ is continuous for any fixed $s \in \mathbb{R}^p$, where $h = h_\beta$ depends on the weights β as in Eq. 4.4, we can find some $\epsilon > 0$ such that $\|\beta - \tilde{\beta}\| < \epsilon$ indicates $d_X(h(s_i), \tilde{h}(s_i)) < \frac{\nu}{2}$ for all observed independent variables $s_i, i = 1, \dots, k$.

We now consider $\tilde{\beta} \in B(\beta, \epsilon)$; we wish to show that for all $i = 1, \dots, k$ and all $t \in T \setminus T_{x_i}$, we have $d(\tilde{h}(s_i), x_i) < d(\tilde{h}(s_i), tx_i)$, namely that the optimal representative of $[x_i]$ is left unchanged for all $i = 1, \dots, k$, even if we perturb the regression model. This would complete the proof.

Note that by the definition of ν , we have for any $i = 1, \dots, k$ and $t \in T \setminus T_{x_i}$

$$d(h(s_i), x_i) \leq d(h(s_i), tx_i) - \nu.$$

We compute

$$\begin{aligned} d(\tilde{h}(s_i), x_i) &\leq \underbrace{d(\tilde{h}(s_i), h(s_i))}_{< \frac{\nu}{2}} + d(h(s_i), x_i) < \frac{\nu}{2} + d(h(s_i), x_i) \leq \frac{\nu}{2} + d(h(s_i), tx_i) - \nu \\ &< -\frac{\nu}{2} + \underbrace{d(h(s_i), \tilde{h}(s_i))}_{< \frac{\nu}{2}} + d(\tilde{h}(s_i), tx_i) < -\frac{\nu}{2} + \frac{\nu}{2} + d(\tilde{h}(s_i), tx_i) = d(\tilde{h}(s_i), tx_i), \end{aligned}$$

where the second and fourth inequalities follow from the triangle inequality. This completes the proof of Theorem 4.1.6 under the assumption that Lemma 4.3.1 holds. \square

Chapter 4. Prediction in the Graph Space

The proof of Lemma 4.3.1 relies on the following :

Lemma 4.3.2. *Let $\beta \in \mathbb{R}^{(p+1) \times J}$ be the parameters of the output of AAC as stated in Theorem 4.1.6. This β encodes $p + 1$ flattened matrices (of dimension J) of coefficients, where β_j is the j^{th} flattened matrix, and $\beta_j(a)$ is the j^{th} coordinate corresponding to the a^{th} node or edge, and we denote by $\beta(a)$ the $(p + 1)$ -dimensional vector of coefficients for the node or edge a .*

Then, with probability 1, $\beta(a_1) \neq \beta(a_2) \in \mathbb{R}^{(p+1)}$ for all $a_1 \neq a_2 \in \{1, \dots, J\}$, giving $t\beta \neq \beta$ for all $t \in T \setminus \{Id\}$.

Proof. From the analytical solution of a linear regression model in X (see Eq. (4.27)), we recall that

$$\hat{\beta} = (\Phi(S)^T \Phi(S))^{-1} \Phi(S)^T X.$$

Since, by the assumptions of the theorem, $\Phi(S)$ has full rank, so does $(\Phi(S)^T \Phi(S))^{-1} \Phi(S)^T$. Thus, if $\beta(a_1) = \beta(a_2)$ for some $a_1 \neq a_2$, then the corresponding elements of X belong to the same fiber of $(\Phi(S)^T \Phi(S))^{-1} \Phi(S)^T$, which happens with probability 0. \square

Now, we are ready to prove the final Lemma 4.3.1. We prove the Lemma in the case where $s \in \mathbb{R}$, for simplicity of notation. The extension to the vector case $s \in \mathbb{R}^p$ easily follows.

Proof of Lemma 4.3.1. In order to prove the lemma, we will show that the set

$$\mathcal{X}_T = \left\{ \left((s_1, [x_1]), \dots, (s_k, [x_k]) \right) \in (\mathbb{R} \times X/T)^k \left| \begin{array}{l} d(h(s_i), x_i) = d(h(s_i), tx_i) \\ \text{for some repr. } x_1, \dots, x_k, \\ i = 1, \dots, k \text{ and } t \in T \setminus T_{x_i} \end{array} \right. \right\}$$

has measure $(\lambda \times \eta)_k(\mathcal{X}_T) = 0$, where $(\lambda \times \eta)_k$ is the product measure induced by $(\lambda \times \eta)$ on $\underbrace{(\mathbb{R} \times X/T) \times \dots \times (\mathbb{R} \times X/T)}_k$.

For each element $t \in T$, denote by $X^t = \{x \in X \mid tx = x\}$ the fixed point set of t . Note that $(\lambda \times \eta)_k(\mathcal{X}_T) = (\lambda \times m)_k((Id_{\mathbb{R}} \times \pi)^{-1}(\mathcal{X}_T))$, and that

$$(Id_{\mathbb{R}} \times \pi)^{-1}(\mathcal{X}_T) = \bigcup_{i=1}^k \bigcup_{t \in T} \mathcal{X}_{i,t},$$

where

$$\mathcal{X}_{i,t} = \left\{ (s_1, x_1, \dots, s_k, x_k) \in (\mathbb{R} \times X) \times \dots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{\text{th}}} \times \dots \times (\mathbb{R} \times X) \left| \begin{array}{l} d_X(h(s_i), x_i) \\ = d_X(h(s_i), tx_i) \end{array} \right. \right\}$$

and

$$\mathcal{X}_{i,t} \subset \underbrace{(\mathbb{R} \times X) \times \cdots \times (\mathbb{R} \times X)}_k.$$

The preimage $\mathcal{F}_i^{-1}(0)$ of the function

$$\mathcal{F}_i: \underbrace{X \times \cdots \times X}_k \rightarrow \mathbb{R}, \quad (s_1, x_1, \dots, s_k, x_k) \mapsto d_X^2(h(s_i), x_i) - d_X^2(h(s_i), tx_i)$$

satisfies

$$\mathcal{F}_i^{-1}(0) \cap (\mathbb{R} \times X) \times \cdots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \cdots \times (\mathbb{R} \times X) = \mathcal{X}_{i,t}.$$

We show that \mathcal{F}_i is a submersion on $(\mathbb{R} \times X) \times \cdots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \cdots \times (\mathbb{R} \times X)$ by showing that it has nonzero gradient.

Note that

$$\begin{aligned} \mathcal{F}_i(s_1, x_1, \dots, s_k, x_k) &= d_X^2(h(s_i), x_i) - d_X^2(h(s_i), tx_i) \\ &= \|h(s_i) - x_i\|^2 - \|h(s_i) - tx_i\|^2 \\ &= (h(s_i) - x_i)^T(h(s_i) - x_i) - (h(s_i) - tx_i)^T(h(s_i) - tx_i) \\ &= 2h(s_i)^T(tx_i - x_i). \end{aligned}$$

It follows that

$$\nabla_{x_i} \mathcal{F}_i(s_1, x_1, \dots, s_k, x_k) = 2h(s_i)^T(t - I).$$

We would like to show that $2h(s_i)^T(t - I)$ is nonzero with probability 1. Note that $2h(s_i)^T(t - I) = 0$ if and only if $t^T h(s_i) = h(s_i)$, which also indicates that $th(s_i) = tt^T h(s_i) = h(s_i)$. By Lemma 4.3.2, we know that $t\beta \neq \beta$ with probability 1, and by the assumptions of Theorem 4.1.6, we then have $th(s_i) \neq h(s_i)$ with probability 1. Hence, we may conclude that with probability 1, $2h(s_i)^T(t - I) \neq 0$, giving

$$\nabla_{x_i} \mathcal{F}_i(s_1, x_1, \dots, s_k, x_k) \neq 0.$$

It follows that \mathcal{F}_i is a submersion on

$$(\mathbb{R} \times X) \times \cdots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \cdots \times (\mathbb{R} \times X).$$

As a result, the set

$$\mathcal{F}_i^{-1}(0) \cap (\mathbb{R} \times X) \times \cdots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \cdots \times (\mathbb{R} \times X) = \mathcal{X}_{i,t}$$

Chapter 4. Prediction in the Graph Space

has co-dimension 1 and, in particular,

$$m_k(\mathcal{X}_{i,t}) = m_k(\mathcal{F}_i^{-1}(0) \cap (\mathbb{R} \times X) \times \dots \times \underbrace{(\mathbb{R} \times X \setminus X^t)}_{i^{th}} \times \dots \times (\mathbb{R} \times X) = 0.$$

But then,

$$\begin{aligned} (\lambda \times \eta)_k(\mathcal{X}_k) &= (\lambda \times m)_k((Id_{\mathbb{R}} \times \pi)^{-1}(\mathcal{X}_T)) = (\lambda \times m)_k\left(\bigcup_{i=1}^k \bigcup_{t \in T} \mathcal{X}_{i,t}\right) \\ &\leq \sum_{i=1}^k \sum_{t \in T} (\lambda \times m)_k(\mathcal{X}_{i,t}) = 0, \end{aligned}$$

which proves the lemma. \square

Appendix 2: Multiple Output Least Square

We recall the Ordinary Least Square regression model with multiple output (see Chapter 3, Hastie et al. (2009)), declined in the context where the output is a set of flattened adjacency matrices. Having a set of regressors and graphs $(s_1, [x_1]), \dots, (s_k, [x_k])$ where $(s_i, [x_i]) \in \mathbb{R}^p \times X/T$. Every graph can be represented as an equivalence class of flat adjacency matrices of dimension $J = n^2$, where n is the number of nodes. Suppose we have, for each equivalence class $[x] \in X/T$, chosen an optimal representative $x \in X$ using the criterion defined in 4.5. Then, the regression problem can be reformulated as follows:

$$\mathcal{X} = \mathcal{S} \times \mathcal{B} + \mathcal{E} \tag{4.24}$$

and in matrix form:

$$\begin{bmatrix} x_{1,1} & \dots & x_{1,J} \\ \vdots & \ddots & \vdots \\ x_{k,1} & \dots & x_{k,J} \end{bmatrix} = \begin{bmatrix} 1 & s_{1,1} & \dots & s_{1,p} \\ \vdots & \ddots & & \vdots \\ 1 & s_{k,1} & \dots & s_{k,p} \end{bmatrix} \begin{bmatrix} \beta_{1,1} & \dots & \beta_{1,J} \\ \vdots & \ddots & \vdots \\ \beta_{p+1,1} & \dots & \beta_{p+1,J} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,1} & \dots & \epsilon_{1,J} \\ \vdots & \ddots & \vdots \\ \epsilon_{k,1} & \dots & \epsilon_{k,J} \end{bmatrix}$$

where X represents the output matrix, S the matrix of regressors with the intercept added as a first column, B the matrix of coefficients, and E the error matrix. In terms of single elements of the graph - using the above index notation - the regression model describe the element (i, a) in the adjacency matrix:

$$x_{ia} = \sum_{j=0}^p s_{ij} \beta_j(a) + \epsilon_{ia}$$

Note that if $i = j$ the regression is describing a node, if $i \neq j$ the regression is describing an edge. As in Multiple Linear Regression, the estimated matrix of coefficients \hat{B} can

be computed by minimizing the least squares error. The analytical solution (under the hypothesis of full rank of the matrix S) is:

$$\hat{B} = (S^T S)^{-1} S^T X. \quad (4.25)$$

The model can be re-framed by using basis functions $\phi(s): \mathbb{R} \rightarrow \mathbb{R}$, rewriting the model as:

$$X = \Phi(S)B + E.$$

where $\Phi(S)$ is defined as:

$$\begin{bmatrix} 1 & \phi_1(s_1) & \dots & \phi_p(s_1) \\ \vdots & \ddots & & \vdots \\ 1 & \phi_1(s_k) & \dots & \phi_p(s_k) \end{bmatrix} \quad (4.26)$$

In this case, the analytical solution is given by

$$\hat{B} = (\Phi(S)^T \Phi(S))^{-1} \Phi(S)^T X \quad (4.27)$$

Appendix 3: Corollary of Theorem 4.1.6

Corollary 4.3.3. *Let Graph Space X/T be endowed with a probability measure η which is absolutely continuous with respect to the the push forward of the Lebesgue measure m on X , and let λ be a probability measure absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^p . Let the sample $\{(s_1, [x_1]), \dots, (s_k, [x_k])\}, (s_i, [x_i]) \in \mathbb{R}^p \times X/T$ be sampled from $\lambda \times \eta$.*

Assume that the AAC for Regression (Algorithm 4) fits the regression model f_β in Definition 4.1.2. Assume moreover that the design matrix $S \in \mathbb{R}^{k \times p}$ of the regressors satisfies the following properties:

- i) $s_0 = [1, \dots, 1]^T$ - the first column models the intercept;*
- ii) Sample s from λ and let $(\beta_0, \dots, \beta_p) \neq (\tilde{\beta}_0, \dots, \tilde{\beta}_p)$. Then, with probability 1,*

$$\sum_{j=0}^p \beta_j s_j \neq \sum_{j=0}^p \tilde{\beta}_j s_j$$

- iii) The matrix*

$$S = \begin{bmatrix} 1 & s_{11} & \dots & s_{1p} \\ \vdots & \ddots & & \vdots \\ 1 & s_{k1} & \dots & s_{kp} \end{bmatrix}$$

has full rank.

Chapter 4. Prediction in the Graph Space

Under these circumstances, we claim that

- a) The AAC algorithm terminates in finite time, and*
- b) With probability 1, the estimated regression curve f_β returned by the AAC algorithm is a local minimum of the function*

$$\beta \mapsto \sum_{i=1}^k d_{X/T}^2([x_i], f_\beta(s_i)). \quad (4.28)$$

CHAPTER 5

Conclusion

In this thesis, we focused on the statistical analysis of populations of graphs and the application to urban movements. We started from the description of the chosen embedding space, the Graph Space, and the study of its geometrical properties (Chapter 2). We showed an example of cluster analysis, we discussed the property of the Fréchet Mean, and we defined the Generalized Geodesic Principal Components (Chapter 3). We focused on the prediction of graphs, proposing a graph-valued regression model and a non-parametric conformal interval for the uncertainty quantification of a prediction (Chapter 4). Along these chapters, we introduced a general methodology for the estimation of intrinsic statistics on the Graph Space: the Align All and Compute algorithm (AAC). The AAC is a flexible strategy, which can be extended to the estimation of different other statistical methods. All these methodologies are applicable to both labelled and unlabelled, direct and un-direct, weighted and unweighted, one and multi-layered graphs. The mathematical innovation of the thesis lays at the intersection of geometry and statistics. Note that the Graph Space is a quotient space and it is not a manifold. This poses modelling questions intractable from the manifold statistics perspective. Defining statistical methods on quotient spaces (both intrinsic and extrinsic) is an urgent research field, which has been mostly explored in the shape analysis literature. Quotient spaces are a very general concept, potentially describing all sorts of phenomena, and they are yet very little explored in the

Chapter 5. Conclusion

statistical literature. In addition, this thesis offers a compendium of challenges and possible solutions encountered in the analysis of complex non-Euclidean data, starting from the embedding of the data, moving to the definition of original statistical methods. As for all complex problems, the thesis is a partial exploration of the phenomenon and every section arouse many new research questions discussed at the end of the chapters.

In terms of applications, the current work focused on the analysis of urban movement, both by discussing the importance of the usage of complex data analysis and by applying the developed tools to several case studies. In the analysis of movements, the intuitive representation has always been the spatial one, the Euclidean one. In this thesis, we lay the foundations to look at a spatial occurrence in a relational perspective, where movements become objects in an abstract space. If a spatial phenomenon is interpreted as a relational phenomenon, a different geometry is considered to analyse the physical space, and geometry is no more an auxiliary tool for measuring reality, taken for granted *a priori*. The choice of geometry becomes a first and important step of the analysis. To understand if the relational representation is a useful description of the reality, we proposed different case studies throughout the chapters: the micro scale analysis of citizens' usage of a square (Section 3.1); the mobility at urban level in the city of Copenhagen (Section 4.1); the origin destination matrices within the Lombardy region using census data (Section 3.3) and gps data (Section 4.2). Looking at these results, there are several gains in using the Graph Space framework. First, the Graph Space approach is trying to avoid reducing the complexity of the datum, by analysis it in its natural complexity and natural embedding. Second, the relational approach allows to perform comparison of spacial phenomena that would be otherwise incomparable (e.g. two equal square usage that differs only form the entry point). On the other hand, scholars can argue that the main point of the spatial analysis is the spatial dependency itself. The complex data analysis aims at giving insights about a phenomenon that could be *a posteriori* grounded again in the physical space. As clearly shown by the example of Piazza Leonardo (Section 3.1) and the example of Copenhagen bus system data (Section 4.1), analysing movement in a non-Euclidean space can reveal interesting insights about the actual usage of a location, dis-anchoring it from its pure geographical position and embedding it in a interconnected network perspective.

This thesis consists of two small steps: one in the direction of the statistical analysis of complex non Euclidean data and the other in the dialogue between statistics and qualitative disciplines, trying to define a trustful quantitative paradigm for the analysis of complex data and complex phenomena. The collaboration with urban planners and landscape designers revealed to me the importance of statistics: the power of embracing the complexity of reality, admiring its variability, and quantifying its blurriness.

Bibliography

- B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control and Optimization*, 51(3):2230–2260, 2013.
- C. Anderson. The end of theory: the data deluge makes the scientific methods obsolete. *Wired*, 2008.
- A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi. A prediction interval for a function-valued forecast model: Application to load forecasting. *International Journal of Forecasting*, 32(3):939–947, 2016.
- M. Arnaudon and L. Miclo. A stochastic algorithm finding generalized means on compact manifolds. *Stochastic Processes and their Applications*, 124(10):3463–3479, 2014.
- M. Arnaudon, F. Barbaresco, and L. Yang. Medians and means in Riemannian geometry: Existence, uniqueness and computation. In *Matrix Information Geometry*, pages 169–197. Springer, 2013.
- A. L. Barabási. *Network science*. Cambridge university press, 2016.
- M. Batty. *A new theory of space syntax*. Centre for Advanced Spatial Analysis (UCL), 2004.
- M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal of Optimization*, 24(3):1542–1566, 2014.
- U. Beck. Mobility and the cosmopolitan perspective. In *Exploring Networked Urban Mobilities*, pages 156–168. Routledge, 2017.

Bibliography

- R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *Annals of Statistics*, 31(1):1–29, 2003.
- L. J. Billera, S. P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.
- R. Boel and L. Mihaylova. A compositional stochastic model for real time freeway traffic simulation. *Transportation Research Part B: Methodological*, 40(4):319–334, 2006.
- S. Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.
- G. E. Bredon. *Introduction to compact transformation groups*. Academic press, 1972.
- N. Brenner. The urban question: reflections on henri lefebvre, urban theory and the politics of scale. *International journal of urban and regional research*, 24(2):361–378, 2000.
- N. Brenner. *Implosions/explosions*. Jovis, Berlin, 2014.
- M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer, 1999.
- R. Brunet. *Les villes "européennes": rapport pour la DATAR, Délégation à l'Aménagement du Territoire et à l'Action Régionale*. La Documentation Française, 1989.
- H. Bunke and K. Riesen. Improving vector space embedding of graphs through feature selection algorithms. *Pattern Recognition*, 44(9):1928–1940, 2011.
- A. Calissano, A. Feragen, and S. Vantini. Populations of unlabeled networks: Graph space geometry and geodesic principal components. *MOX Report*, 2020a.
- A. Calissano, A. Feragen, and S. Vantini. Graph-valued regression: Predicting network from vectors. *MOX Report*, 2020b.
- A. Calissano, A. Feragen, and S. Vantini. Graphspace python package. <https://github.com/annacalissano/GraphSpace.git>, 2020c.
- A. Calissano, M. Fontana, G. Zeni, and S. Vantini. Conformal prediction sets for populations of networks. *MOX Report*, 2020d.
- A. Calissano, P. Sturla, P. Pucci, V. Fedeli, and S. Vantini. Going beyond the euclidean setting in the statistical analysis of human movement in urban landscape. *Journal of Digital Landscape Architecture*, 1:150–158, 2020e.
- A. B. Cambel. *Applied chaos theory: A paradigm for complexity*. Elsevier, 1993.

- K. Carley. Dynamic network analysis. In *Summary in NRC Workshop on Dynamic social network modeling and analysis*, pages 133–145. National Academies Press, 2003.
- L. R. Casse and R. Casse Jr. *Projective geometry: an introduction*. Oxford University Press, 2006.
- L. Cella and R. Martin. Valid distribution-free inferential models for prediction. *arXiv:2001.09225*, 2020. arXiv: 2001.09225.
- R. Chakraborty and B. C. Vemuri. Recursive Fréchet mean computation on the Grassmannian and its applications to computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4229–4237. IEEE, 2015.
- G. Chartrand. *Introductory graph theory*. Courier Corporation, 1977.
- S. Chowdhury and F. Mémoli. Distances and isomorphism between networks and the stability of network invariants. *arXiv preprint arXiv:1708.04727*, 2017.
- S. Chowdhury and F. Mémoli. The metric space of networks. *arXiv preprint arXiv:1804.02820*, 2018.
- C. Coins. Cryptocurrency database. Accessed: 2020-09-04.
- V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine*, 4(1):e13, 2007.
- R. M. Colombo. A 2×2 hyperbolic traffic flow model. *Mathematical and computer modelling*, 35(5-6):683–688, 2002.
- D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.
- H. S. M. Coxeter. *Non-euclidean geometry*. Cambridge University Press, 1998.
- B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi. Population shape regression from random design data. *International journal of computer vision*, 90(2):255–266, 2010.
- D. A. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765, 2011.
- I. Dryden and K. Mardia. *Statistical analysis of shape*. Wiley, 1998.
- A. Duncan, E. Klassen, and A. Srivastava. Statistical shape analysis of simplified neuronal trees. *The Annals of Applied Statistics*, 12(3):1385–1421, 2018.

Bibliography

- D. Durante and D. B. Dunson. Nonparametric bayes dynamic modelling of relational data. *Biometrika*, 101(4):883–898, 2014.
- D. Durante, D. B. Dunson, and J. T. Vogelstein. Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1516–1530, 2017.
- D. K. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, pages 2224–2232, 2015.
- F. Emmert-Streib, M. Dehmer, and Y. Shi. Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, 346:180–197, 2016.
- I. Epifani, L. Ladelli, and G. Posta. *Appunti di calcolo delle probabilita*. Dispense di Probabilità e Statistica per l’Informatica, 2005.
- P. Erdős and A. Rényi. On the evolution of random graphs. *Mathematical Institute Hungarian Academy of Science*, 5(1):17–60, 1960.
- A. Feragen and T. M. W. Nye. Statistics on stratified spaces. In *Riemannian Geometric Statistics in Medical Image Analysis*, pages 299–342. Elsevier, 2020.
- A. Feragen, F. Lauze, P. Lo, M. de Bruijne, and M. Nielsen. Geometries on spaces of treelike shapes. In *Asian Conference on Computer Vision*, pages 160–173. Springer, 2010.
- A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. In *IEEE International Conference on Computer Vision (ICCV)*, pages 736–746, 2011.
- A. Feragen, M. Owen, J. Petersen, M. Wille, L. Thomsen, A. Dirksen, and M. de Bruijne. Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *Information Processing in Medical Imaging*, volume 7917 of *Lecture Notes in Computer Science*, pages 74–85. Springer, 2013.
- A. Feragen, M. Nielsen, E. B. Vedel Jensen, A. Plessis, and F. Lauze. Geometry and statistics: Manifolds and stratified spaces. *Journal of Mathematical Imaging and Vision*, 50(1-2):1–4, 2014.
- R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922.

- P. T. Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, 105(2):171–185, 2013.
- P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer vision and mathematical methods in medical and biomedical image analysis*, pages 87–98. Springer, 2004.
- P. Forster, L. Forster, C. Renfrew, and M. Forster. Phylogenetic network analysis of sars-cov-2 genomes. *Proceedings of the National Academy of Sciences*, 117(17):9241–9243, 2020.
- C. Fuchs. From digital positivism and administrative big data analytic towards critical digital and social media research. *European Journal of Communication*, 32(1):37–49, 2017.
- C. Fuchs and V. Mosco. *Marx in the age of digital capitalism*. Brill, 2015.
- J. Gehl. *Life between buildings: using public space*. Island press, 2011.
- J. Gehl and B. Svarre. *How to study public life*. Island press, 2013.
- M. A. Giampieri, S. Williams, W. Xu, H. Gunc, C. M. Langston, and D. C. Harvey. Benchmark: using sensors to study public space. In *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*, volume 17, 2017.
- C. E. Ginestet, J. Li, P. Balachandran, S. Rosenberg, and E. D. Kolaczyk. Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, 11(2):725–750, 06 2017.
- A. R. Goetz, T. M. Vowles, and S. Tierney. Bridging the qualitative–quantitative divide in transport geography. *The Professional Geographer*, 61(3):323–335, 2009.
- S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 18(4):377–388, 1996.
- A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. *A survey of statistical network models*. Now Publishers Inc, 2010.
- M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- J. C. Gower. Generalized Procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.
- M. Graham. The knowledge based economy and digital divisions of labour. pages 189–199. Routledge, 2014.

Bibliography

- M. Graham and T. Shelton. Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3):255–261, 2013.
- X. Guo, A. Srivastava, and S. Sarkar. A quotient space formulation for statistical analysis of graphical data. *arXiv preprint arXiv:1909.12907*, 2019.
- T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.
- S. Hauberg, A. Feragen, R. Enficiaud, and M. J. Black. Scalable robust principal component analysis using Grassmann averages. *IEEE transactions on pattern analysis and machine intelligence*, 38(11):2298–2311, 2015.
- B. Hillier, J. Hanson, and H. Graham. Ideas are in things: an application of the space syntax method to discovering house genotypes. *Environment and Planning B: planning and design*, 14(4):363–385, 1987.
- J. D. Hinkle, P. T. Fletcher, and S. C. Joshi. Intrinsic polynomials for regression on riemannian manifolds. *Journal of Mathematical Imaging and Vision*, 50(1-2):32–52, 2014.
- P. D. Hoff. Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics*, 9(3):1169–1193, 2015.
- T. Honderich. *The Oxford companion to philosophy*. OUP Oxford, 2005.
- Y. Hong, R. Kwitt, N. Singh, B. Davis, N. Vasconcelos, and M. Niethammer. Geodesic regression on the grassmannian. In *European Conference on Computer Vision*, pages 632–646. Springer, 2014.
- Y. Hong, R. Kwitt, N. Singh, N. Vasconcelos, and M. Niethammer. Parametric regression on the grassmannian. *IEEE transactions on pattern analysis and machine intelligence*, 38(11):2284–2297, 2016.
- Y. Hong, X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Regression uncertainty on the grassmannian. In *Artificial Intelligence and Statistics*, pages 785–793, 2017.
- Y. Hu, L. Manikonda, and S. Kambhampati. What we instagram: A first analysis of instagram photo content and user types. In *Eighth International AAAI conference on weblogs and social media*, pages 595–598. The AAAI Press, 2014.
- S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica*, 20(1):1–58, 2010.

- B. Jain and K. Obermayer. On the sample mean of graphs. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 993–1000. IEEE, 2008.
- B. J. Jain and K. Obermayer. Structure spaces. *Journal of Machine Learning Research*, 10(11):2667–2714, 2009.
- O. B. Jensen and T. Richardson. *Making European space: mobility, power and territorial identity*. Routledge, 2004.
- H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 321–328, 2003.
- D. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.
- K. Kersting, N. M. Kriege, C. Morris, P. Mutzel, and M. Neumann. Benchmark data sets for graph kernels, 2016. URL <http://graphkernels.cs.tu-dortmund.de>.
- N. Keyfitz. Individual mobility in a stationary population. *Population Studies*, 27(2): 335–352, 1973.
- T. N. Kipf and M. Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
- E. D. Kolaczyk and G. Csárdi. *Statistical analysis of network data with R*, volume 65. Springer, 2014.
- E. D. Kolaczyk, L. Lin, S. Rosenberg, J. Walters, J. Xu, et al. Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *The Annals of Statistics*, 48(1):514–538, 2020.
- E. F. Krause. *Taxicab geometry: An adventure in non-Euclidean geometry*. Courier Corporation, 1986.
- H. Lee, Z. Ma, Y. Wang, and M. K. Chung. Topological distances between networks and its application to brain imaging. *arXiv preprint arXiv:1701.04171*, 2017.
- J. M. Lee. *Smooth manifolds*. Springer, 2013.
- J. Lei and L. Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):71–96, 2014.

Bibliography

- J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.
- H. Liu, X. Chen, L. Wasserman, and J. D. Lafferty. Graph-valued regression. In *Advances in Neural Information Processing Systems*, pages 1423–1431, 2010.
- S. Lord, C. Després, and T. Ramadier. When mobility makes sense: A qualitative and longitudinal study of the daily mobility of the elderly. *Journal of Environmental psychology*, 31(1):52–61, 2011.
- D. Lusher, J. Koskinen, and G. Robins. *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press, 2013.
- A. Mallasto and A. Feragen. Wrapped gaussian process regression on riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5580–5588. IEEE, 2018.
- S. Manson and D. O'Sullivan. Complexity theory in the study of space and place. *Environment and Planning A*, 38(4):677–692, 2006.
- H. Maron, H. Ben-Hamu, N. Shamir, and Y. Lipman. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.
- J. S. Marron and A. M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.
- E. Miller, M. Owen, and J. S. Provan. Polyhedral computational geometry for averaging metric phylogenetic trees. *Advances in Applied Mathematics*, 68:51 – 91, 2015.
- N. Miolane, S. Holmes, and X. Pennec. Template shape estimation: correcting an asymptotic bias. *SIAM Journal on Imaging Sciences*, 10(2):808–844, 2017.
- T. Mizuno, H. Takayasu, and M. Takayasu. Correlation networks among currencies. *Physica A: Statistical Mechanics and its Applications*, 364:336–342, 2006.
- V. Mosco. *Becoming digital: Toward a post-internet society*. Emerald Group Publishing, 2017.
- E. Nagel. *The structure of science: Problems in the logic of scientific explanation*. Harcourt, Brace & World, 1961.
- Y. Ni, F. C. Stingo, and V. Baladandayuthapani. Bayesian graphical regression. *Journal of the American Statistical Association*, 114(525):184–197, 2019.

- T. M. W. Nye. Principal components analysis in the space of phylogenetic trees. *The Annals of Statistics*, 39:2716–2739, 2011.
- T. M. W. Nye. An algorithm for constructing principal geodesics in phylogenetic treespace. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 11(2):304–315, 2014.
- T. M. W. Nye, X. Tang, G. Weyenberg, and R. Yoshida. Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika*, 104: 901–922, 2017.
- C. O’neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.
- H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman. Inductive Confidence Machines for Regression. In *Machine Learning: ECML 2002*, Lecture Notes in Computer Science, pages 345–356. Springer, 2002.
- L. Pappalardo, D. Pedreschi, Z. Smoreda, and F. Giannotti. Using big data to study the link between human mobility and socio-economic development. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 871–878. IEEE, 2015.
- V. Patrangenaru and L. Ellingson. *Nonparametric statistics on manifolds and their applications to object data analysis*. CRC Press, 2015.
- X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision (IJCV)*, 66(1):41–66, 2006.
- P. Pucci, F. Manfredini, and P. Tagliolato. *Mapping urban practices through mobile phone data*. Springer, 2015.
- J. O. Ramsay and B. W. Silverman. *Applied functional data analysis: methods and case studies*. Springer, 2007.
- C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 33(5):727–748, 2006.
- J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive computing*, 6(3):30–38, 2007.
- Regione Lombardia. Open Data Regione Lombardia, 2019. Accessed: 2019-09-02.
- K. Richardson, P. Cilliers, and M. Lissack. Complexity science: A “gray” science for the “stuff in between”. *Emergence: Complexity and Organization*, 3(2):6–18, 2001.

Bibliography

- K. Riesen and H. Bunke. Iam graph database repository for graph based pattern recognition and machine learning. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 287–297. Springer, 2008.
- P. Riva, P. Sturla, A. Calissano, and S. Vantini. Landscape perception through complex data: Exploring george hargreaves’s queen elizabeth olympic park in london. *Journal of Digital Landscape Architecture*, 4:42–56, 2019.
- G. Robins, T. Snijders, P. Wang, M. Handcock, and P. Pattison. Recent developments in exponential random graph (p^*) models for social networks. *Social networks*, 29(2): 192–215, 2007.
- P. G. Røe. Qualitative research on intra-urban travel: an alternative approach. *Journal of Transport Geography*, 8(2):99–106, 2000.
- H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda. gboost: a mathematical programming approach to graph classification and regression. *Machine Learning*, 75(1):69–89, 2009.
- F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.
- M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607. Springer, 2018.
- C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.
- P. Secchi, S. Vantini, and V. Vitelli. Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of milan. *Statistical Methods & Applications*, 24(2): 279–300, 2015.
- K. E. Severn, I. L. Dryden, and S. P. Preston. Non-parametric regression for networks. *arXiv preprint arXiv:2010.00050*, 2020.
- N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt. Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*, 12:2539–2561, 2011.
- S. Skwerer. *Tree Oriented Data Analysis*. PhD thesis, University of North Carolina at Chapel Hill Graduate School., 2014.

- T. A. Snijders. Statistical models for social networks. *Annual review of sociology*, 37: 131–153, 2011.
- T. A. Snijders. Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application*, 4:343–363, 2017.
- A. Srivastava, S. H. Joshi, W. Mio, and X. Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(4):590–602, 2005.
- K. T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, 338:357–391, 2003.
- P. Suppes. Ernest nagel. *Biographical memoirs of the National Academy of Sciences*, 1994.
- J. L. Toole, M. Ulm, M. C. González, and D. Bauer. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*, pages 1–8. ACM, 2012.
- K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Discrete and Computational Geometry*, 52(1):44–70, 2014.
- S. Vantini, V. Vitelli, É. de France, and P. Zanini. Treelet analysis and independent component analysis of milan mobile-network data: Investigating population mobility and behavior. *Analysis and Modeling of Complex Data in Behavioural and Social Sciences*, 87:1–4, 2012.
- S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt. Graph kernels. *The Journal of Machine Learning Research*, 11:1201–1242, 2010.
- C. von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov. Public transport networks: empirical analysis and modeling. *The European Physical Journal B*, 68(2):261–275, 2009.
- V. Vovk, A. Gammernan, and G. Shafer. *Algorithmic learning in a random world*. Springer, 2005.
- H. Wang and S. J. Marron. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.
- J. Wang, X. Zuo, and Y. He. Graph-based network analysis of resting-state functional MRI. *Frontiers in systems neuroscience*, 4:1–16, 2010.

Bibliography

- P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. González. Understanding road usage patterns in urban areas. *Scientific reports*, 2:1–6, 2012.
- S. Wasserman and K. Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.
- W. H. Whyte et al. *The social life of small urban spaces*. Conservation Foundation Washington, DC, 1980.
- A. Willis. Confidence sets for phylogenetic trees. *Journal of the American Statistical Association*, 114(525):235–244, 2019.
- E. Wyly. Automated (post) positivism. *Urban Geography*, 35(5):669–690, 2014.
- K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.
- G. Zeni, M. Fontana, and S. Vantini. Conformal Prediction: a Unified Review of Theory and New Challenges. *arXiv:2005.07972*, 2020.
- M. Zhang and P. T. Fletcher. Probabilistic principal geodesic analysis. In *Advances in Neural Information Processing Systems 26*, pages 1178–1186. Curran Associates, Inc., 2013.
- M. Zhang, Z. Cui, M. Neumann, and Y. Chen. An end-to-end deep learning architecture for graph classification. In *Thirty-Second AAAI Conference on Artificial Intelligence*, pages 4438–4445. The AAAI Press, 2018.
- C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11):2178–2199, 2014.
- S. Zuboff. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1):75–89, 2015.

