GALATASARAY UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

HIERARCHICAL STRUCTURES IN DATA SCIENCE

Halime Beyza Küçükdağ

May 2019

HIERARCHICAL STRUCTURES
IN DATA SCIENCE

(VERİ BİLİMİNDE
HİYERARŞİK YAPILAR)

by

Halime Beyza Küçükdağ, B.S.

Thesis

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

in

MATHEMATICS

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

May 2019

This is to certify that the thesis entitled

HIERARCHICAL STRUCTURES
IN DATA SCIENCE

prepared by **Halime Beyza Küçükdağ** in partial fulfillment of the requirements for the degree of **Master of Science in Mathematics** at the **Galatasaray University** is approved by the

Examining Committee :

Doc. Dr. Ayşegül Ulus Yıldız (Supervisor)
Department of Mathematics
Galatasaray University ____________________

Prof. Dr. Meral Tosun
Department of Mathematics
Galatasaray University ____________________

Prof. Dr. Ayşe Hümeyra Bilge
Department of Engineering and Natural Sciences
Kadir Has University ____________________

Date: ____________________

Acknowledgements

First of all, I would like to say a heartfelt thanks to my advisor Dr. Ayşegül Ulus who supported me to prepare my thesis in an intensive understanding. She worked meticulously with me on every subject. With her correct orientation, we have achieved useful resources and results. I also thank her for her patience and for all efforts.

I would also like to thank Professor Meral Tosun for her questions that pushed me to think and to advance, for her advices and for her precious time.

Also thanks to Yağmur Yavuz for supporting us in many articles and especially for her psychological support.

Finally, I express my gratitude to everyone who contributed to this thesis.

15 May 2019
Halime Beyza Küçükdağ

Table of Contents

List of Symbols

$E[X]$	: Expectation of X
μ	: Mean Value
$var(X)$	: Variance of X
σ_X	: Standard deviation of X
$cov(X,Y)$	: Covariance
$\rho_{X,Y}$	: Pearson Correlation Coefficient
d	: Distance
$\|.\|$	: Norm
MST	: Minimum Spanning Tree
$corr$	: Correlation

List of Figures

Abstract

In recent years, the increase of studies analyzing data as complex systems lead clustering to play key role. Hierarchical clustering is one of the most popular clustering method in data science. It is a useful method with its comprehensible application, graphical analysis and with its resulting hierarchical tree.

This thesis aims to study the mathematical background of the hierarchical clustering structures of a particular data by using metric and ultrametric spaces' features as well as graph theoretical tools.

First of all, we study metric spaces, normed spaces and ultrametric spaces. Besides some examples, including the remarkable p-adic spaces, the topological properties of these spaces are studied.

Then, we study how to interpret a particular data by means of a metric and ultrametric space. Ultrametric tree models of similarity and association are used to produce the representation of the data. We gave the equivalence of agglomerative hierarchical clustering model using single linkage and the graph theoretical model using minimal spanning tree. We tackled here some notions of Graph Theory which helps us to visualize the data and mainly the question how to obtain a Minimum Spanning Tree (MST) from a graph which represents the optimization process.

Finally, we analyze the data obtained from PISA-mathematical and PISA-reading performance evolution over 4 years for 10 OECD countries. We analyze these particular data by using minimum spanning tree model which are obtained by using certain algorithms (Prim& Kruskal) and programs (Python& Sage). The results of our data analysis allow us to make a meaningful conclusion about the evolution of mathematics and reading performance in the considered 10 OECD countries.

Keywords: Metric Space, Normed Space, Ultrametric Spaces, p-adic numbers, Hierarchical Clustering, Minimum Spanning Trees.

Özet

Son yıllarda, veri analizinin karmaşık sistemler olarak ele alınması, kümelenme yönteminin bu konularda kilit rol oynamasına neden olmaktadır. Hiyerarşik kümelenme, veri biliminde en önemli veri analizi yöntemlerinden biri haline gelmiştir. Anlaşılır uygulaması, grafik analizi ve sonuçta ortaya çıkan hiyerarşik ağacı ile yararlı bir yöntemdir.

Bu tezin amacı verilerin hiyerarşik kümelenme yapılarını metrik ve ultrametrik uzay özellikleri ve çizge kuramı yöntemlerini kullanarak ele almaktır. Bu sebeple, bu tezde, veri üzerinde yapılacak uygulamalara matematiksel bir iskelet oluşturabilmek için öncelikle metric uzaylar, normlu uzaylar ve ultrametrik uzaylar çalışılmıştır. P-sel uzaylar gibi özel öneme sahip örneklerin yanı sıra başka örnekler de çalışılmış ve adı geçen uzayların topolojik karakterizasyonları da verilmiştir.

Daha sonra hiyeraşik yapılar ele alınarak kümelenme kavramının matematiksel içeriği üzerine çalışılmıştır. Veri üzerinde benzerlik ve bağlam ifadelerden nasıl bir metric ve ultrametrik uzay elde edildiği anlatılmıştır. Çizge Teorisi'nin bazı tanımlamaları sayesinde veriyi görselleştirebilme yöntemleri ifade edilmiştir ve çizge üzerinde bir optimizasyon yapılarak bize anlamlı bir çizge çıkaran Minimal Geren Ağaç yöntemi ele alınmıştır.

En son olarak yapılan çalışmalar bir veri üzerine uygulanır. Üzerinde çalışılan veri 10 tane OECD ülkesinin PISA-matematik ve PISA-okuma performanslarının 4 yıllık bir zaman serisidir. Bu tez, tek bağla toplanabilir hiyerarşik kümelenme yöntemi ve çizge kuramsal bir yöntem olan Minimal Geren Ağaç yönteminin eşitliğinden yola çıkarak veriyi analiz etmemizi sağlamıştır. Tezden elde ettiğimiz sonuçlar sözkonusu veriden ilgili 10 ülkedeki matematik performansları hakkında anlamlı bir sonuç çıkarmamızı sağlamaktadır.

Anahtar Sözcükler : Metrik Uzay, Normlu Uzay, Ultrametrik Uzaylar, p-sel sayılar, Hiyerarşik Kümelenme, Minimal Geren Ağaçlar.

1 Introduction

In mathematics, a space is a set which consists of selected mathematical objects like points and selected relationships between these points which formulate the structure of the space, for instance: Euclidean spaces, linear spaces, probability spaces, etc. In ancient Greek mathematics, space was a geometric abstraction of the three dimensional reality observed in everyday life. *Euclid* built all of mathematics on these geometric foundations, going so far as to define numbers by comparing the lengths of line segments to the length of a chosen reference segment and he gave axioms for the properties of spaces. The method of coordinates which is used for knowledge of similarity of objects in natural science, was adopted then by *René Descaretes* in 1637. The period between the late 17th century to the beginning of 20th century was the golden age of geometry where the axiomatization of Euclidean space started. The mathematicians followed the idea of *Bernhard Riemann* given by the inaguaral lecture in 1854 which stated that every mathematical object parametrized by n real numbers may be treated as a space of n $-$dimensional space so that the terminology of Euclid can be used for different mathematical concepts, for example for function spaces that will be elaborated in 20th century by functional analysis.

In 1906, the French mathematician *Maurice Fréchet* initiated the study of *abstract spaces* in his work "Sur Quelques Points du Calcul Fonctionnel" (Fréchet, 1906). An abstract space is a set of objects which satisfy certain axioms. The nature of objects are unspecified but the axioms define the space and the general theorems can then be given for various sets.

Metric spaces are fundamentals in abstract spaces whose idea was given by *Fréchet* even though the name is due to *Felix Hausdorff*. Metric spaces can be seen as the basis of abstract spaces such as normed spaces, Banach spaces, inner product spaces or Hilbert spaces. A metric space is a space for which a distance function is defined for any pair of elements belonging to the space. The distance function takes non-negative values and satisfies the axioms which are mainly symmetry and the triangular inequality.

In this thesis, we shall consider metric spaces and normed spaces in a great deal as they provide a basis for our study. However, our special interest will be on an other abstract space which is called *ultrametric space*. Ultrametric spaces were first topologically characterized by *J. de Groot* (Chapter 10, Theorem 2.1 in Nagata, 2004), that theorem was extended to higher dimension by *J.Nagata and P.Ostrand* (Nagata, 2004). Before seen as a special abstract space in the mathematics literature, in 1897, a famous example of an ultrametric space, the p-adic space (or the notion of p-adic numbers) introduced in algebraic number theory by *Kurt Hensel*(Hensel, 1897). Hensel's idea gave birth to topological and geometrical notions in the field of p-adic numbers. The p-adic absolute value was first defined on the field of rational numbers: Given a fixed prime number p, for any $x \neq 0$, let $|x|_p = r$ where r is the highest power of p dividing x. This definition of p-adic absolute value satisfies a stronger version of triangular inequality axiom:

$$|x+y|_p \leq \max\{|x|_p, |y|_p\}$$

However, it unobeys the rule of Archimedes[1], that is beacuse it is sometimes called as *non-Archimedean absolute value*.
Through the p-adic absolute value, the p-adic distance was defined as $d(x,y) = |x-y|_p$. Two rational numbers are p-adically close if their difference is divisible by a high power of p.

The most important fact behind the pioneering work of Hensel in the viewpoint of this thesis is two-folded: Firstly, from this particular example, we study the topological properties of ultrametric spaces in a general way. These are mainly, in an ultrametric space, every point inside a ball is itself at the center of the ball, and the diameter of a ball is equal to its radius. Two balls are either disjoint or contained one within the other. In different terms, the stronger triangular inequality implies that, in an ultrametric space, all triangles are either equilateral or isosceles with a small base. Secondly, the natural geometrical ordering of p-adic numbers is not along the real line but on a hierarchical generating tree. Actually, not only the particular example of p-adic spaces but notion of ultrametric spaces provide us to consider different distances than usual geometrical distance between the same set of objects such as genetical distance, or any distance which measures the similarity, relations or kindship.

After 1950's, ultrametricity appeared in scientific fields outside mathematics such as taxonomy, biological classification or linguistics. The idea was to classify a set of objects. The idea of this classification can be explained for example for a set of living

[1] Replacing y with x, one has $|x+x|_p \leq \max\{|x|_p, |x|_p\} = |x|$ contradicting Archimedean rule.

species given by a dendogram or an inverted tree where the objects, which are in fact the end lines of leaves, are listed at the bottom of the inverted tree (evolution tree). Starting from the bottom, several leaves which can be interpreted as species are merged into a branch to form genus, genus are merged into an higher branch to form a family, and etc. until the most common ancestor, taxa within taxa where higher taxa comprise a larger diversity of species than lower taxa. In other words, going from the bottom to the top, one obtains a clustering or an inclusion mechanism. Distances or metric spaces are used to measure the hierarchy by associating a real number to the branching points which can be interpreted as the time elapsed since they evolved and branced off in the evolution tree.

By the improvement of measurement of distances in the related fields such as molecular biology, lexicology and physics, the question of how to classify ideally for example a set of biomolecules strings of mammals or a set of words of a language raised. Several procedures for tree construction were proposed and started to be used in the literature related to ultrametrics. (Sørensen, 1948 and Florek et al., 1951)

In the last quarter of the 20th century, the development of the statistical methods and computer technology achieved a significant breakthrough which can be called ”data analysis” (or as its popular name says nowadays “data science”) by the works of Benzecri (1965) and Benzecri (1984) , Shephard (1980) , Gordon (1981) , Murtagh (1983). Since 2000, many other techniques are now commonly used in data analysis, in order to extract information from all sorts of data.

2 Literature Review

In this thesis we will focus on one of the aggregation procedure which is called single-linkage clustering (dated back to Florek et al., 1951), Sokal and Sneath (1963) and as given in our main references Rammal et al. (1983), (1985). Given a collection of objects with a distance matrix, the idea of the procedure can be explained as follows: begin by aggregating the two closest objects into one cluster, define the (renormalized) distances between two clusters which is the minimum distance between any member of one cluster and any member of the other. This sytematical procedure leads to a unique tree and a unique renormalized distance matrix. Renormalized distances which are smaller than or equal to the original distances are the subdominant ultrametric. This procedure provides the hierarchical arrangement of particular data studied recently by econophysicists to detect for example the hierarchical structure in a portfolio of n stocks traded in a financial market. There was a need to quantify a distance between different stocks which was satisfied by synchronous correlation coefficient of the daily difference of logarithm of closure price of stocks. The correlation coefficient is computed between all the possible pairs of stocks present in the portfolio in a given time period (Mantegna, 1999).

In this thesis, we analyze the data obtained from PISA-mathematical performance evolution over 4 years for 10 OECD countries. Following Mantegna (1999), we quantify a distance by correlation coefficient of the yearly difference of logarithm of closure mathematical performance points of countries. We study how to interpret a particular data by means of a metric and ultrametric space. Ultrametric tree models of similarity or association are given producing the representation of the data which are obtained from the data as correlation coefficients. We give the equivalence of agglomerative hierarchical clustering model using single linkage and the graph theoretical model using minimal spanning tree. We tackle here some notions of Graph Theory which helps us to visualize the data and mainly the question how to obtain a Minimum Spanning Tree (MST) from a graph which represents the optimization process given by certain algorithms (Prim& Kruskal) and programs (Python& Sage). The results of our data analysis allow us to make a meaningful conclusion about the evolution of mathematics performance in the considered 10 OECD countries.

This thesis is composed of three main parts. Part 1 consists of two chapters, namely, Chapter 2 and Chapter 3 which give the mathematical preliminaries of the hierarchical structures where we studied metric spaces, normed spaces and ultrametric spaces.

In the second part, in Chapter 4, we give the equivalence of agglomerative hierarchical clustering using single linkage and the graph theoretical models using minimal spanning tree model. We tackle here some notions of Graph Theory and mainly the question how to obtain a Minimum Spanning Tree (MST) from a graph. Examining the mathematical background lying behind the hierarchical clustering is the objective of this chapter.

Finally, in Chapter 5, we analyze the data obtained from PISA-mathematical and PISA-reading performance evolution over 4 years for 10 OECD countries. Ultrametric tree models of similarity and association are used to produce the representation of this data. Together with the statistical analysis given in the Appendix, we produce the representation of our data as an hierarchical organization by using minimal spanning tree model given by certain algorithms (Prim& Kruskal) and programs (Python, R). We also obtained a hierarchical clustering of the same data by using a program with ready-made algorithms in Orange which is a data visualization toolkit. We conclude with the comparative results.

3 Metric Spaces and Normed Spaces

In this chapter, we will give the definitions and the examples of a metric space and normed space which will be the mathematical preliminaries for the terms of closeness, distance and similarity used in the following chapters of this thesis.

3.1 Metric Space

Definition 3.1.1. *Let X be a non-empty set. A metric on X is a function $d : X \times X \to \mathbb{R}_+$ satisfying the following axioms for all $x, y, z \in X$:*

i. $d(x,y) = 0$ *if and only if* $x = y$,

ii. ***(Symmetry)*** $d(x,y) = d(y,x)$,

iii. ***(Triangle Inequality)*** $d(x,z) \leq d(x,y) + d(y,z)$.

The pair (X,d) is then called a metric space.

Remark. Let X be an Euclidean space, consider the map $d : X \times X \to \mathbb{R}_+$
defined by $d(x,y) = |x-y|$ for $x, y \in X$.
Therefore, d defines a metric on an Euclidean Space which satisfies the minimal properties we might expect of a distance, illustrated in the following examples:

Example 3.1.1. *I. Let $X = \mathbb{R}$. Consider the map $d_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ defined by $d_1(x,y) = |x-y|$ for all $x, y \in \mathbb{R}$.*
$(\mathbb{R}, d_1)$ is a metric space, since the map d_1 satisfies the axioms for all $x, y, z \in \mathbb{R}$:

i. $d_1(x,y) = |x-y| = 0 \Leftrightarrow x = y$.

ii. $d_1(x,y) = |x-y| = |y-x| = d_1(y,x)$. *Symmetry is satisfied.*

iii. *The triangle inequality holds for all $x, y, z \in \mathbb{R}$:*

$$\begin{aligned} d_1(x,z) &= |x-z| = |x-z+y-y| \\ &\leq |x-y| + |y-z| \\ &= d_1(x,y) + d_1(y,z) \end{aligned}$$

Therefore, d_1 is a metric which is called ***usual metric*** *on $\mathbb{R}$.*

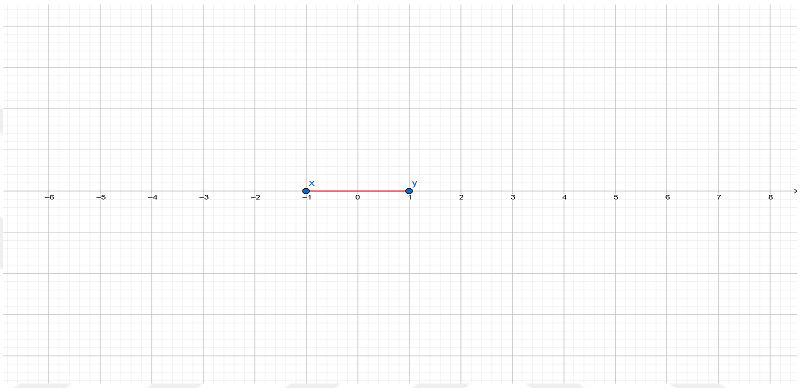


Figure 3.1: $d_1(x,y) = d_1(-1,1) = 2$

II. *Let $X = \mathbb{R}^2$. Consider the map d_1 on $\mathbb{R}^2$ for $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R}^2$ such that:*

$$\begin{aligned} d_1 : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R}_+ \\ (x,y) &\mapsto |(x_1,x_2) - (y_1,y_2)| \\ &\mapsto |(x_1 - y_1, x_2 - y_2)| = |x_1 - y_1| + |x_2 - y_2| \end{aligned}$$

where $x = (x_1, x_2), y = (y_1, y_2)$.

i. *If we have $d_1(x,y) = |x_1 - y_1| + |x_2 - y_2| = 0$,*
we have $|x_1 - y_1| = 0$ and $|x_2 - y_2| = 0$ which implies that $x_1 = y_1$ and $x_2 = y_2$.
Thus, $x = y$ if $d_1(x,y) = 0$.
If we have $x = (x_1, x_2),\ y = (y_1, y_2)$ and $x = y$, then $d_1(x,y) = 0$

ii. *For all $x, y \in \mathbb{R}^2$, we have:*

$$d_1(x,y) = |x_1 - y_1| + |x_2 - y_2| = |y_1 - x_1| + |y_2 - x_2| = d_1(y,x).$$

Thus, symmetry is satisfied.

iii. Let $x = (x_1, x_2), y = (y_1, y_2), z = (z_1, z_2) \in \mathbb{R}^2$*, we have*

$$\begin{aligned} d_1(x,z) &= |x_1 - z_1| + |x_2 - z_2| \\ &= |x_1 - z_1 + y_1 - y_1| + |x_2 - z_2 + y_2 - y_2| \\ &\leq |x_1 - y_1| + |-z_1 + y_1| + |x_2 - y_2| + |-z_2 + y_2| \\ &= d_1(x,y) + d_1(y,z). \end{aligned}$$

Hence, the triangle inequality is also satisfied.

$(\mathbb{R}^2, d_1)$ *is thus a metric space and* d_1 *is also called* ***taxicab metric*** *or* ℓ^1.

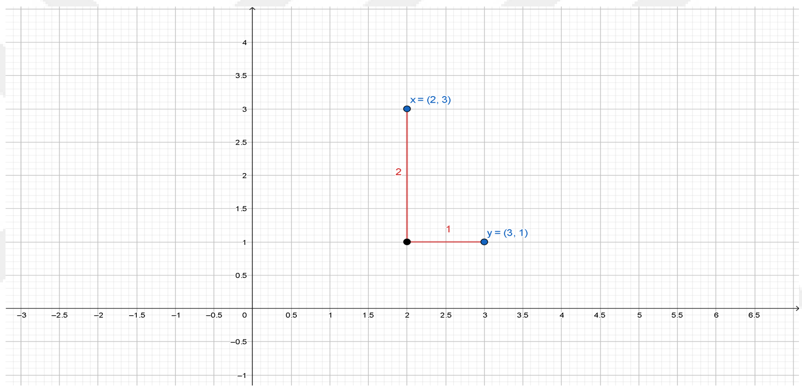


Figure 3.2: The Taxicab distance of $(2,1)$ is $d_1(x,y) = 2 + 1 = 3$

III. The metric d_1 *can be extended to the space* $X = \mathbb{R}^n$ *as follows:*
Consider the map $d_1 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ *defined by*
$d_1(x,y) = |x - y|$ *for all* $x, y \in \mathbb{R}^n$ *where* $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$*. Then:*

$$\begin{aligned} d_1 : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R}_+ \\ (x,y) &\mapsto |(x_1, \ldots, x_n) - (y_1, \ldots, y_n)| \\ &\mapsto |(x_1 - y_1, \ldots, x_n - y_n)| = |x_1 - y_1| + \ldots + |x_n - y_n| \\ &\mapsto = \sum_{i=1}^{n} |x_i - y_i| \end{aligned}$$

$(\mathbb{R}^n, d_1)$ *is a metric space.*

Example 3.1.2. *Consider the map* $d_2 : X \times X \to \mathbb{R}_+$ *defined by*

$$d_2(x,y) = |x - y| = \sqrt{(x-y)^2}$$

for all $x, y \in X$.

I. $(\mathbb{R}, d_2)$ *is a metric space.*

II. Let $X = \mathbb{R}^2$. *Consider the map* d_2 *on* $\mathbb{R}^2$ *such that:*

$$\begin{array}{rcl} d_2 : \mathbb{R}^2 \times \mathbb{R}^2 & \to & \mathbb{R}_+ \\ (x,y) & \mapsto & |(x_1,x_2)-(y_1,y_2)| \\ & \mapsto & |(x_1-y_1,x_2-y_2)| = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} \end{array}$$

where $x = (x_1,x_2), y = (y_1,y_2)$ *for all* $x,y,z \in \mathbb{R}^2$, *we have:*

i. If we have $d_2(x,y) = 0$, *then;*

$$d_2(x,y) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} = 0$$

Since $|x_1-y_1| = 0$ *and* $|x_2-y_2| = 0$ *implies that*
$x_1 = y_1$ *and* $x_2 = y_2$. *Thus,* $x = y$.
Corollary, if we have $x = y$ *and* $x_1 = y_1$, $x_2 = y_2$, *then* $d_2(x,y) = 0$.

ii. We have,

$$\begin{array}{rcl} d_2(x,y) & = & \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} \\ & = & \sqrt{(y_1-x_1)^2+(y_2-x_2)^2} = d_2(y,x) \end{array}$$

iii. Let $x = (x_1,x_2), y = (y_1,y_2), z = (z_1,z_2)$, *then*

$$\begin{array}{rcl} d_2(x,z) & = & \sqrt{(x_1-z_1)^2+(x_2-z_2)^2} \\ & = & \sqrt{(x_1-z_1+y_1-y_1)^2+(x_2-z_2+y_2-y_2)^2} \\ & = & \sqrt{((x_1-y_1)+(y_1-z_1))^2+((x_2-y_2)+(y_2-z_2))^2} \\ & \leq & \sqrt{(x_1-y_1)^2+(y_1-z_1)^2+(x_2-y_2)^2+(y_2-z_2)^2} \\ & \leq & \sqrt{(x_1-y_1)^2+(x_2-y_2)^2}+\sqrt{(y_1-z_1)^2+(y_2-z_2)^2} \\ & = & d_2(x,y)+d_2(y,z). \end{array}$$

Hence, $(\mathbb{R}^2, d_2)$ *is a metric space called Euclidean or* ℓ^2 *metric.*

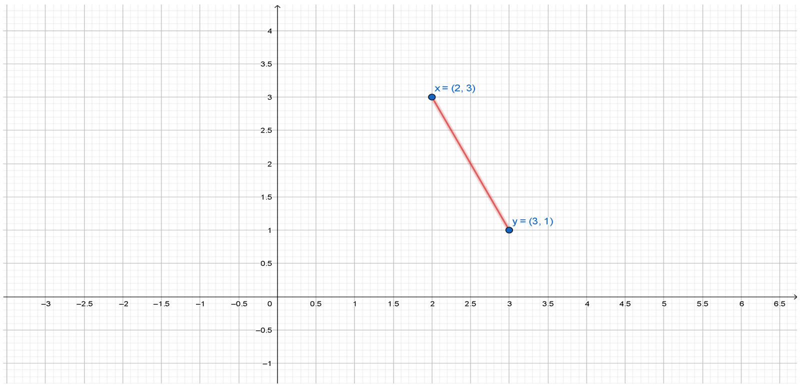


Figure 3.3: $d_2(x,y) = |(2,3)-(3,1)| = \sqrt{(2-3)^2+(3-1)^2} = \sqrt{5}$

III. *The metric d_2 can be extended to $\mathbb{R}^n$ as follows:*
Consider the map $d_2 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ defined by $d_2(x,y) = |x-y|$ for all $x,y \in \mathbb{R}^n$.

$$\begin{aligned} d_2 : (x_1,\ldots,x_n) \times (y_1,\ldots,y_n) &\mapsto |(x_1,\ldots,x_n)-(y_1,\ldots,y_n)| \\ &= \sqrt{(x_1-y_1)^2+\ldots+(x_n-y_n)^2} \end{aligned}$$

$(\mathbb{R}^n, d_2)$ is a metric space.

Example 3.1.3. I. *Let $X = \mathbb{R}$. Consider the map $d_\infty : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ defined by $d_\infty(x,y) = max(|x-y|)$ for all $x,y \in \mathbb{R}$.*
$(\mathbb{R}, d_\infty)$ is a metric space, since the map d_∞ satisfies the axioms for all $x,y,z \in \mathbb{R}$:

i. $d_\infty(x,y) = max(|x-y|) = 0 \Leftrightarrow x = y$.

ii. $d_\infty(x,y) = max(|x-y|) = max(|y-x|) = d_\infty(y,x)$.

iii. *The triangle inequality holds for all $x,y,z \in \mathbb{R}$:*

$$\begin{aligned} d_\infty(x,z) &= max(|x-z|) = max(|x-z+y-y|) \\ &\leq max(|x-y|) + max(|y-z|) \\ &= d_\infty(x,y) + d_\infty(y,z) \end{aligned}$$

Therefore, d_∞ is a metric on $\mathbb{R}$.

II. Let $X = \mathbb{R}^2$. Consider the map $d_\infty : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$ defined by

$$
\begin{array}{rcl}
d_\infty : \mathbb{R}^2 \times \mathbb{R}^2 & \to & \mathbb{R}_+ \\
(x,y) & \mapsto & |(x_1,x_2) - (y_1,y_2)| \\
& \mapsto & max(|x_1 - y_1|, |x_2 - y_2|)
\end{array}
$$

where $x = (x_1,x_2)$ and $y = (y_1,y_2)$.

$(\mathbb{R}^2, d_\infty)$ is a metric space if d_∞ satisfies all the axioms of a metric for all $x,y,z \in \mathbb{R}^2$.

i. If $d_\infty(x,y) = max(|x_1 - y_1|, |x_2 - y_2|) = 0$, then we have $|x_1 - y_1| = 0, |x_2 - y_2| = 0$ implying that $(x_1 = y_1),\ (x_2 = y_2)$. Thus, $x = y$. If we have $x = y$ and $(x_1 = y_1),\ (x_2 = y_2)$, then $d_\infty(x,y) = 0$.

ii. The symmetry axiom is satisfied as follows:

$$
\begin{array}{rcl}
d_\infty(x,y) & = & max(|x_1 - y_1|, |x_2 - y_2|) \\
& = & max(|y_1 - x_1|, |y_2 - x_2|) = d_\infty(y,x)
\end{array}
$$

iii. Let $z = (z_1,z_2) \in \mathbb{R}^2$, then:

$$
\begin{array}{rcl}
d_\infty(x,z) & = & max(|x_1 - z_1|, |x_2 - z_2|) \\
& = & max(|x_1 - z_1 + y_1 - y_1|, |x_2 - z_2 + y_2 - y_2|) \\
& \leq & max((|x_1 - y_1|, |y_1 - z_1|), (|x_2 - y_2|, |y_2 - z_2|)) \\
& \leq & max(|x_1 - y_1|, |x_2 - y_2|) + max(|y_1 - z_1|, |y_2 - z_2|) \\
& = & d_\infty(x,y) + d_\infty(y,z)
\end{array}
$$

Finally, (iii) is also satisfied.

Thus, $(\mathbb{R}^2, d_\infty)$ is a metric space called ℓ^∞ or maximum metric.

III. Let $X = \mathbb{R}^n$. Consider the map $d_\infty : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ defined by

$$\begin{array}{rcl} d_\infty : \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R}_+ \\ (x,y) & \mapsto & |(x_1,\ldots,x_n)-(y_1,\ldots,y_n)| \\ & \mapsto & max(|x_1-y_1|,\ldots,|x_n-y_n|) \end{array}$$

where $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$.

We will show that d_∞ satisfies all the axioms of a metric for all $x,y,z \in \mathbb{R}^n$.

i. If $d_\infty(x,y) = max(|x_1-y_1|,\ldots,|x_n-y_n|) = 0$, then we have $|x_1-y_1| = 0,\ldots,|x_n-y_n| = 0$ implying that $(x_1 = y_1),\ldots,(x_n = y_n)$. Thus, $x = y$. If we have $x = y$ and $(x_1 = y_1),\ldots,(x_n = y_n)$, then $d_\infty(x,y) = 0$.

ii. We will have the symmetry as follows:

$$\begin{array}{rcl} d_\infty(x,y) & = & max(|x_1-y_1|,\ldots,|x_n-y_n|) \\ & = & max(|y_1-x_1|,\ldots,|y_n-x_n|) = d_\infty(y,x) \end{array}$$

iii. Let $z = (z_1,\ldots,z_n) \in \mathbb{R}^n$, then:

$$\begin{array}{rcl} d_\infty(x,z) & = & max(|x_1-z_1|,\ldots,|x_n-z_n|) \\ & = & max(|x_1-z_1+y_1-y_1|,\ldots,|x_n-z_n+y_n-y_n|) \\ & \leq & max((|x_1-y_1|,|y_1-z_1|),\ldots,(|x_n-y_n|,|y_n-z_n|)) \\ & \leq & max(|x_1-y_1|,\ldots,|x_n-y_n|)+max(|y_1-z_1|,\ldots,|y_n-z_n|) \\ & = & d_\infty(x,y)+d_\infty(y,z) \end{array}$$

Finally, (iii) is also satisfied.

Hence, $(\mathbb{R}^n, d_\infty)$ is a metric space.

Example 3.1.4. *Let X be a set of all bounded sequences. Each element in the space ℓ^1 is a sequence $x = (x_i) = (x_1,x_2,\ldots)$ of numbers such that $|x_1|+|x_2|+\ldots$ converges; thus*

$$\sum_{i=1}^{\infty} |x_i| < \infty$$

and the metric is defined by

$$d(x,y) = (\sum_{i=1}^{\infty} |x_i - y_i|)$$

where $y = (y_i)$ *and* $\sum |y_i| < \infty$.

i. *If* $d(x,y) = (\sum_{i=1}^{\infty} |x_i - y_i|) = 0$, *we have* $|x_i - y_i| = 0$ *and* $x_i = y_i$. *Thus,* $x = y$. *Inversely, if* $x = y \ \to x_i = y_i$, *then* $d(x,y) = (\sum_{i=1}^{\infty} |x_i - y_i|) = 0$.

ii. $d(x,y) = (\sum_{i=1}^{\infty} |x_i - y_i|) = (\sum_{i=1}^{\infty} |y_i - x_i|) = d(y,x)$.

iii. *Let* x,y,z *in* ℓ^1 *the series in converges. Note that* $z = (z_i)$ *and* $\sum |z_i| < \infty$.

$$\begin{aligned} d(x,y) &= (\sum_{i=1}^{\infty} |x_i - y_i|) \\ &\leq (\sum_{i=1}^{\infty} [|x_i - z_i| + |z_i - y_i|]) \\ &\leq (\sum_{i=1}^{\infty} |x_i - z_i|) + (\sum_{i=1}^{\infty} |z_i - y_i|]) \\ &= d(x,z) + d(z,y) \end{aligned}$$

This completes that l^1 *is a metric space.*

Example 3.1.5. *Let X be a set of all bounded sequences. Each element in the space* ℓ^2 *is a sequence* $x = (x_i) = (x_1, x_2, \dots)$ *of numbers such that* $|x_1| + |x_2| + \dots$ *converges; thus*

$$\sum_{i=1}^{\infty} |x_i| < \infty$$

and the metric is defined by

$$d(x,y) = \sqrt{\sum_{i=1}^{\infty} |x_i - y_i|^2}$$

where $y = (y_i)$ *and* $\sum |y_i| < \infty$.

i. *If* $d(x,y) = \sqrt{\sum_{i=1}^{\infty} |x_i - y_i|^2} = 0$, *we have* $|x_i - y_i| = 0$ *and* $x_i = y_i$. *Thus,* $x = y$. *Inversely, if* $x = y \ \to x_i = y_i$, *then* $d(x,y) = \sqrt{\sum_{i=1}^{\infty} |x_i - y_i|^2} = 0$.

ii. $d(x,y) = \sqrt{\sum_{i=1}^{\infty} |x_i - y_i|^2} = \sqrt{\sum_{i=1}^{\infty} |y_i - x_i|^2} = d(y,x)$.

iii. *Let x, y, z in ℓ^2 the series in converges. Note that $z = (z_i)$ and $\sum |z_i| < \infty$.*

$$\begin{aligned} d(x,y) &= \sqrt{\sum_{i=1}^{\infty} |x_i - y_i|^2} \\ &= \sqrt{\sum_{i=1}^{\infty} (|x_i - z_i + z_i - y_i|^2)} \quad \textit{by using Minkowski Inequality} \\ &\leq \sqrt{\sum_{i=1}^{\infty} |x_i - z_i|^2} + \sqrt{\sum_{i=1}^{\infty} |z_i - y_i|^2} \\ &= d(x,z) + d(z,y) \end{aligned}$$

This completes that l^2 is a metric space. [1]

Example 3.1.6. *Let X be a set of all bounded sequences. Every element of X is a sequence $x = (x_i) = (x_1, x_2, \ldots)$ such that for all $i = 1, 2, \ldots$ and we have $|x_i| \leq c_x$ where c_x is a real number which may depend on x, but does not depend on i. We choose the metric defined by*

$$d(x,y) = sup|x_i - y_i| \ (i \in \mathbb{N})$$

where $y = (y_i) \in X$. The metric space thus obtained is generally denoted by ℓ^∞.

i. $d(x,y) = sup|x_i - y_i| = 0 \Leftrightarrow x = y.$

ii. $d(x,y) = sup|x_i - y_i| = sup|y_i - x_i| = d(y,x).$

iii. *The triangle inequality holds for all $x, y, z \in X$ with $z = (z_i) \in X$:*

$$\begin{aligned} d(x,z) &= sup|x_i - y_i| = sup|x_i - z_i + z_i - y_i| \\ &\leq sup|x_i - z_i| + sup|z_i - y_i| \\ &= d(x,z) + d(z,y) \end{aligned}$$

Hence, ℓ^∞ is a metric.

Remark. Let X be a non-empty set. Two distances d and d' with $d, d' < \infty$ defined on X are equivalents if there exists $(\alpha, \beta) \in \mathbb{R}^2,\ 0 < \alpha < \beta$ such as for all $(x,y) \in X^2$;

$$\alpha d(x,y) \leq d'(x.y) \leq \beta d(x,y)$$

[1] Minkowski Inequality:

$$(\sum_{i=1}^{\infty}(|x_i + y_i|^p)^{\frac{1}{p}} \leq (\sum_{i=1}^{\infty}(|x_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^{\infty}(|y_i|^p)^{\frac{1}{p}}$$

Proposition 3.1.1. *Let X be a non-empty set and $d_1 : \mathbb{R}^n \to \mathbb{R}$. If $d_\infty(x,y) \le d_1(x,y) \le n.d_\infty(x,y)$ i.e. $y = (y_1,...,y_n), x = (x_1,...,x_n) \in \mathbb{R}^n$, $d_\infty(x,y)$ and $d_1(x,y)$ are equivalents.*

Proof. Let $y = (y_1,...,y_n)$, $x = (x_1,...,x_n) \in \mathbb{R}^n$. For each x_i and y_i $(1 \le i \le n)$, we have

$$|x_i - y_i| \le (|x_1 - y_1| + \cdots + |x_n - y_n|) = d_1(x,y)$$

.

So, $d_\infty(x,y) = max|x_i - y_i| \le d_1(x,y)$ and $d_1(x,y) \le n.d_\infty(x,y)$.

Proposition 3.1.2. *Let X be a non-empty set and $d_2 : \mathbb{R}^n \to \mathbb{R}$. If $d_\infty(x,y) \le d_2(x,y) \le n^{\frac{1}{2}} d_\infty(x,y)$ i.e. $y = (y_1,...,y_n), x = (x_1,...,x_n) \in \mathbb{R}^n$, $d_\infty(x,y)$ and $d_2(x,y)$ are equivalents.*

Proof. Let $y = (y_1,...,y_n)$, $x = (x_1,...,x_n) \in \mathbb{R}^n$. For each x_i and y_i $(1 \le i \le n)$, we have

$$|x_i, y_i| \le (|x_1 - y_1|^2 + \cdots + |x_n - y_n|^2)^{\frac{1}{2}} = d_2(x,y)$$

.

So, $d_\infty(x,y) = max|x_i - y_i| \le d_2(x,y)$ and $d_2(x,y) \le n^{\frac{1}{2}} d_\infty(x,y)$.

Example 3.1.7. *Consider the map $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ defined by*

$$d(x,y) = \begin{cases} 1 & if \;\; x \neq y \\ 0 & if \;\; x = y \end{cases}$$

where $x, y \in \mathbb{R}$. Verify that d satisfies all the axioms of a metric for all $x, y, z \in \mathbb{R}$:

i. *If $d(x,y) = 0$, then we have $x = y$. If $x = y$, it become $d(x,y) = 0$.*

ii. *The symmetry is clearly satisfied.*

iii. *For $x \neq y$; we have 3 cases:*
Case 1: If $z = x$, we have $d(x,z) = 0$ and $d(x,y) = d(y,z) = 1$ i.e. $d(x,z) \le d(x,y) + d(y,z)$.
Case 2: If $z = y$, we have $d(y,z) = 0$ and $d(x,z) = d(x,y) = 1$ i.e. $d(x,z) \le d(x,y) + d(y,z)$.
Case 3: If $z \neq x$ and $z \neq y$, we have $d(x,z) = d(x,y) = d(y,z) = 1$ i.e. $d(x,z) \le d(x,y) + d(y,z)$.
For $x = y$; we have 2 cases:
Case 1: If $z = x$, in this situation $d(x,z) = d(x,y) = d(y,z) = 0$, then $d(x,z) \le d(x,y) + d(y,z)$.

Case 2: If $z \neq x$: in this situation $d(x,y) = 0$ and $d(x,z) = d(y,z) = 1$, so $d(x,z) \leq d(x,y) + d(y,z)$.
*Finally, $(\mathbb{R}, d)$ is a metric space and d is called **discrete metric**.*

3.1.1 Balls in Metric Spaces

Definition 3.1.2. *Let (X,d) be a metric space. The closed ball of radius $r > 0$ centered at a point a in X, denoted by $B(a,r)$, is defined as follows:*

$$B(a,r) = \{x \in X |\ d(x,a) \leq r\}.$$

Example 3.1.8. *Consider the map in $\mathbb{R}^2$ with $x = (x_1, x_2),\ y = (y_1, y_2)$. We will construct the unit closed balls with respect to d_1, d_2, d_∞:*

i. *The unit ball $B(a,1)$ around $a = (a_1, a_2) \in \mathbb{R}^2$ with respect to d_1:*

$$\begin{aligned} B(a,1) &= \{x \in \mathbb{R}^2 |\ d_1(x,a) \leq 1\} \\ &= \{y \in \mathbb{R}^2 :\ |x - a| \leq 1\} \\ &= \{(x_1,x_2) \in \mathbb{R}^2 :\ |(x_1,x_2) - (a_1,a_2)| \leq 1\} \\ &= \{(x_1,x_2) \in \mathbb{R}^2 :\ |(x_1 - a_1, x_2 - a_2)| \leq 1\} \\ &= \{(x_1,x_2) \in \mathbb{R}^2 :\ |x_1 - a_1| + |x_2 - a_2| \leq 1\} \end{aligned}$$

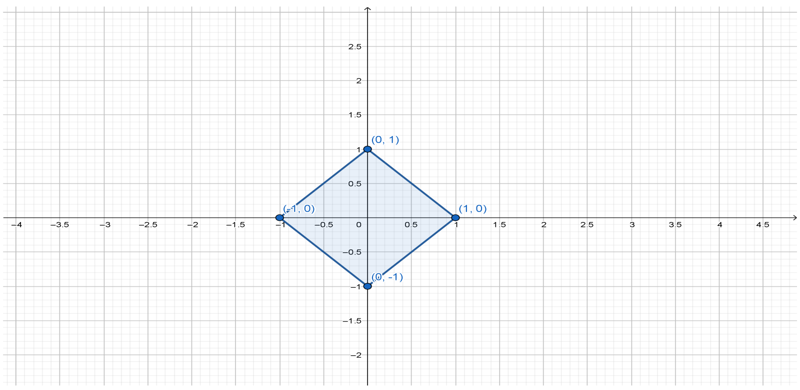


Figure 3.4: Unit closed ball in $(\mathbb{R}_2, d_1)$ is $B(0,1) = \{(x_1,x_2) \in \mathbb{R}^2 :\ |x_1| + |x_2| \leq 1\}$

ii. *The unit closed ball $B(a,1)$ centered $a = (a_1, a_2) \in \mathbb{R}^2$ in metric with respect to d_2*

is:

$$\begin{aligned} B(a,1) &= \{x \in \mathbb{R}^2 | \ d_2(x,a) \le 1\} \\ &= \{x \in \mathbb{R}^2 : \ |x-a| \le 1\} \\ &= \{(x_1,x_2) \in \mathbb{R}^2 : \ \sqrt{(x_1-a_1)^2+(x_2-a_2)^2} \le 1\} \\ &= \{(x_1,x_2) \in \mathbb{R}^2 : \ ((x_1-a_1)^2+(x_2-a_2)^2) \le 1\} \end{aligned}$$

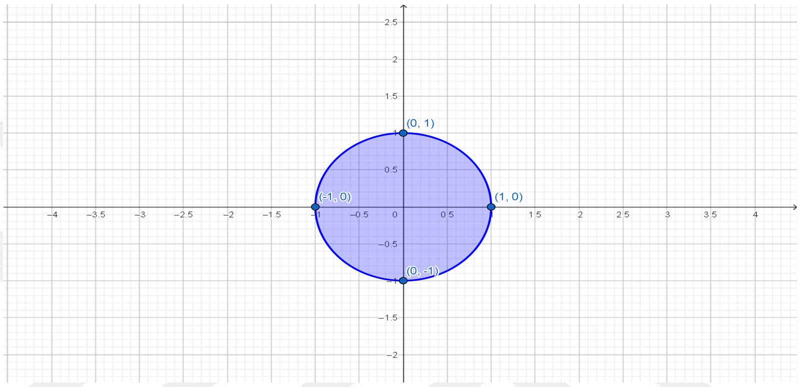


Figure 3.5: Unit ball in $(\mathbb{R}_2, d_2)$, $B(0,1) = \{(x_1,x_2) \in \mathbb{R}^2 : \ x_1^2 + x_2^2 \le 1\}$

iii. *The unit closed ball $B(a,1)$ around $a = (a_1,a_2) \in \mathbb{R}^2$ with respect to d_∞:*

$$\begin{aligned} B(a,1) &= \{x \in \mathbb{R}^2 | \ d_\infty(x,a) \le 1\} \\ &= \{y \in \mathbb{R}^2 : \ |x-a| \le 1\} \\ &= \{(x_1,x_2) \in \mathbb{R}^2 : \ max(|x_1-a_1|,|x_2-a_2|) \le 1\} \end{aligned}$$

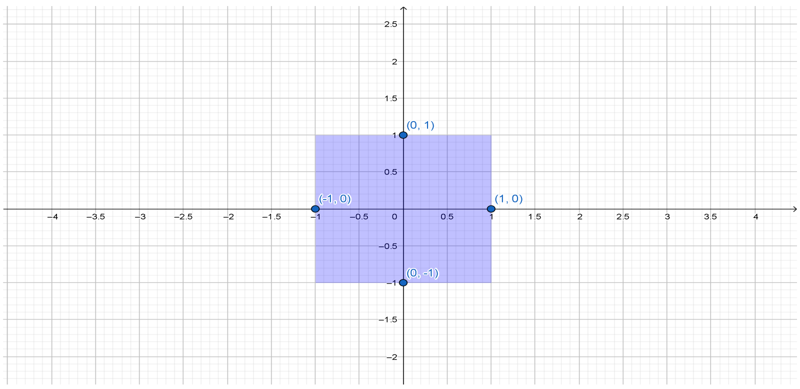


Figure 3.6: Unit closed ball in $(\mathbb{R}_2, d_\infty)$, $B(0,1) = \{(x_1,x_2) \in \mathbb{R}^2 : \ max(|x_1|,|x_2|) \le 1\}$

3.2 Normed Space

Definition 3.2.1. *Let V be an $\mathbb{R}$-vector space*[2] *of finite dimension. A norm on V is a map $\mathcal{N} : V \to \mathbb{R}_+$ which satisfies the following axioms for all $x, y \in V$:*

(N1) $\mathcal{N}(x) = 0$ *if and only if* $x = 0$,

(N2) $\mathcal{N}(kx) = |k|.\mathcal{N}(x),\ k \in \mathbb{R}$,

(N3) $\mathcal{N}(x+y) \leq \mathcal{N}(x) + \mathcal{N}(y)$

The pair $(V, \mathcal{N})$ is then called a normed space.

Example 3.2.1. *I. Let $V = \mathbb{R}$. Consider the map $\mathcal{N} : \mathbb{R} \to \mathbb{R}_+$ defined by*

$$\mathcal{N}(x) = |x| \ \text{for all} \ x, y \in \mathbb{R}.$$

We will show that $\mathcal{N}$ is a norm in $\mathbb{R}$, i.e. $\mathcal{N}$ satisfies the axioms $N1, N2, N3$;

(N1) *If we have* $\mathcal{N}(x) = |x| = 0$, $x = 0$.
$x = 0$, *we have* $\mathcal{N}(x) = |0| = 0$.

(N2) $\mathcal{N}(kx) = |kx| = |k|.|x| = |k|.\mathcal{N}(x)$ *for all* $k \in \mathbb{R}$.
Thus, the condition $(N2)$ is satisfied.

(N3) *Let* $x, y \in \mathbb{R}$.

$$\mathcal{N}(x+y) = |x+y| \leq |x| + |y| = \mathcal{N}(x) + \mathcal{N}(y)$$

Hence, $(\mathbb{R}, \mathcal{N})$ is a normed space.

II. Let $V = \mathbb{R}^2$ and $\mathcal{N} : \mathbb{R}^2 \to \mathbb{R}_+$ be defined as

$$\mathcal{N}((x_1, x_2)) = |x_1| + |x_2| \ \text{where} \ x = (x_1, x_2) \in V.$$

---

[2] A vector space is a collection V, of vectors and two operators, "$+$" and ".", such that the following hold for $u, v, w \in V$ and $c, d \in \mathbb{R}$:

1. $u + v \in V$
2. $u + v = v + u$
3. $(u+v) + w = u + (v+w)$
4. $0 \in V$ where $0 + u = u$
5. For all $u \in V, u + (-u) = 0$
6. $c.u \in V$
7. $c(u+v) = cu + cv$
8. $(c+d).u = cu + du$
9. $c(du) = (cd)u$
10. $1.u = u$

We will show that $\mathcal{N}$ is a norm in $\mathbb{R}^2$, satisfying the axioms $N1, N2, N3$ for all $x, y \in \mathbb{R}^2$;

(N1) *If we have $\mathcal{N}((x_1,x_2)) = |x_1| + |x_2| = 0$, we get $|x_1| = 0$ and $|x_2| = 0$; $x_1 = 0$ and $x_2 = 0$. Thus, $x = (x_1,x_2) = 0$*
If we have $x = (x_1,x_2) = 0$, then
$\mathcal{N}((x_1,x_2)) = |0| + |0| = 0.$

(N2)

$$\begin{aligned}
\mathcal{N}(kx) &= \mathcal{N}(k.(x_1,x_2)) = \mathcal{N}((k.x_1,k.x_2)) \\
&= |k.x_1| + |k.x_2| = |k|.|x_1| + |k|.|x_2| = |k|.(|x_1| + |x_2|) \\
&= |k|.\mathcal{N}((x_1,x_2)) = |k|\mathcal{N}(x)
\end{aligned}$$

where $k \in \mathbb{R}$ and $x = (x_1,x_2) \in \mathbb{R}^2$.

(N3) *Let $x, y \in \mathbb{R}^2$.*

$$\begin{aligned}
\mathcal{N}(x+y) &= |x+y| = |(x_1,x_2) + (y_1,y_2)| \\
&= |(x_1+y_1),(x_2+y_2)| \\
&\leq |x_1+y_1| + |x_2+y_2| \leq |x_1| + |x_2| + |y_1| + |y_2| \\
&\leq \mathcal{N}(x) + \mathcal{N}(y)
\end{aligned}$$

where $y = (y_1,y_2) \in \mathbb{R}^2$ and $x = (x_1,x_2) \in \mathbb{R}^2$.

Hence, $(\mathbb{R}^2, \mathcal{N})$ is a normed space.

III. *Let $V = \mathbb{R}^n$. $\mathcal{N}$ can be extended to the space $\mathbb{R}^n$ as follows:*
Consider the map $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}_+$ defined by

$$\mathcal{N}(x) = \|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$

for any $x \in \mathbb{R}^n$ and $x = (x_1, \ldots, x_n)$.

Thus, the pair $(\mathbb{R}^n, \|.\|_1)$ is a normed space.

Example 3.2.2. *(Kreyszig) Let* $V = \mathbb{R}^n$*. Defined norms by*

$$\|x\|_2 = (\sum_{i=1}^{n} |x_i|^2)^{\frac{1}{2}} = \sqrt{|x_1|^2 + \ldots + |x_n|^2}.$$

$$\|x\|_\infty = sup_i |x_i|$$

where $x \in \mathbb{R}^n$*.*

Proposition 3.2.1. *If the pair* $(V, \mathcal{N})$ *is a normed space, then the following map defines a metric* $d : V \times V \to \mathbb{R}_+$ *where*

$$d(x,y) = \mathcal{N}(x-y).$$

Proof. Let $(V, \mathcal{N})$ is a normed space. We will show that d satisfies $(i)-(iii)$ for all $x,y,z \in V$:

i. We have $d(x,y) = \mathcal{N}(x-y) = 0$ if $x = y$ and also if we have $x = y$, we conclude that $d(x,y) = \mathcal{N}(x-y) = \mathcal{N}(0) = 0$

ii. $d(x,y) = \mathcal{N}(x-y) = \mathcal{N}(-(y-x)) = \mathcal{N}(-1)\mathcal{N}(y-x) = \mathcal{N}(y-x) = d(y,x)$

iii. For all $x,y,z \in V$:

$$\begin{aligned} d(x,y) = \mathcal{N}(x-y) &= \mathcal{N}(x-y+z-z) \\ &\leq \mathcal{N}(x-z) + \mathcal{N}(z-y) = d(x,z) + d(z,y) \end{aligned}$$

Hence, the triangle inequality holds.
Thus, the metric d is induced by the norm $\mathcal{N}$ on V as above and is called **the induced metric**.

Example 3.2.3. *(Kreyszig)* $(\mathbb{R}^n, d_1)$, $(\mathbb{R}^n, d_2)$, $(\mathbb{R}^n, d_\infty)$ *are metric spaces whose metrics are all induced by* $\|x\|_1$, $\|x\|_2$, $\|x\|_\infty$ *respectively.*

$$d(x,y) = \|x-y\|_1 = |x_1 - y_1| + \ldots + |x_n - y_n|$$

$$d(x,y) = \|x-y\|_2 = \sqrt{|x_1 - y_1|^2 + \ldots + |x_n - y_n|^2}$$

$$d(x,y) = \|x-y\|_\infty = max(|x_1 - y_1| + \ldots + |x_n - y_n|)$$

Remark. We can induce a metric space from all norm space but there exits metric space which is not necessarily induced by a norm.

Example 3.2.4. *Let d be discrete metric with $x, y \in \mathbb{R}$.*
Let us check the condition (N2) for the discrete metric space:
Take $x \neq y$ and $|k| \neq 0, 1$ for all $k \in \mathbb{R}$.

$$d(kx, ky) = d(k(x, y)) = |k| d(x, y) = |k| \neq 1 = d(x, y)$$

The discrete metric is not induced from a norm $\mathcal{N}$ as the condition (N2) is not satisfied.

4 Ultrametric Spaces

In this chapter, we will define ultrametric spaces and we will study some examples including the remarkable p-adic spaces and the topological properties of these spaces. Ultrametric spaces provide us to consider different distances than usual geometrical distance between the same set of objects such as genetical distance, or any distance which measures the similarity, relations or kindship. They construct the mathematical bridge between the notions of distance and similarity and the data visualization that we will analyze in the next chapters.

Definition 4.0.1. *Let X be a non-empty set and u be the map $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ satisfying the following conditions for all $x, y, z \in X$:*

i. $u(x,y) = 0$ *if and only if* $x = y$,

ii. ***(Symmetry)*** $u(x,y) = u(y,x)$,

iii. ***(Strong Triangle Inequality)*** $u(x,z) \leq max\{u(x,y), u(y,z)\}$.

Such a pair (X,u) is called an ultrametric space.

Remark. A metric space (X,d) is an ultrametric space if d satisfies for each $x, y, z \in X$:

$$d(x,z) \leq max\{d(x,y), d(y,z)\}.$$

Remark. Every ultrametric space is a metric space.

Example 4.0.1. *The discrete metric space (X,d) is ultrametric space.*
It's sufficient to show that d is an ultrametric if the third axiom "Strong Triangle Inequality" is satisfied:
As we have $d(x,y) = 1$, if and only if $x \neq y$ and d(x,y)=0 if and only if $x = y$.
Let $x, y, z \in X$, We have 2 cases:
Case 1: Assume $x = z$, we have also 2 cases $y = x$ or $y \neq x$:

i. *If $y = x$; $d(x,z) = d(x,y) = d(y,z) = 0$ and $d(x,z) = max\{d(x,y), d(y,z)\}$.*

ii. If $y \neq x$; we have $d(x,y) = 1$, then

$$0 = d(x,z) \leq max\{d(x,y), d(y,z)\} = 1$$

Case 2: Assume $x \neq z$, we have two cases:

i. If $y \neq x$ and $y \neq z$; $d(x,z) = 1 = max\{d(x,y), d(y,z)\} = 1$.

ii. If $y = x$ or $y = z$, $d(x,y) = 1$ or $d(y,z) = 1$ and $d(x,z) = max\{d(x,y), d(y,z)\}$.

Hence, d is an ultrametric.

Example 4.0.2. *Let X be a set of all the words in a dictionary of an arbitrary language. Let us define the distance function d such that*

$$\begin{array}{rcl} d : X \times X & \to & \mathbb{R}_+ \\ (x,y) & \mapsto & d(x,y) = \begin{cases} \frac{1}{2^{n-1}} & if\, x \neq y \\ 0 & if\ x = y \end{cases} \end{array}$$

where n is the order of the letter where the two words of any length x and y differ[1]*.*
Let us show that (X,d) is an ultrametric space. The proprieties $(i)-(ii)$ are satisfied clearly. We will provide the axiom (iii):
If $x = z$, $d(x,z) = 0 \leq max\{d(x,y), d(y,z)\}$.
If $y = z$, $d(y,z) = 0 \leq max\{d(x,y), d(x,z)\}$.
If $x = y$, $d(x,y) = 0 \leq max\{d(x,z), d(z,y)\}$.
When $x \neq y$, $y \neq z$ and $x \neq z$; we will have $d(x,y) = \frac{1}{2^{n-1}}$,
$d(y,z) = \frac{1}{2^{m-1}}$, $d(x,z) = \frac{1}{2^{k-1}}$ respectively.
For all $j < n$, $x_j = y_j$;
For all $j < m$, $y_j = z_j$;
For all $j < k$, $x_j = z_j$;
And for all $j < min\{n,m\}$, $x_j = z_j$ and $k \geq min\{n,m\}$, then:

$$\begin{array}{rcl} \frac{1}{2^{k-1}} & \leq & max\{\frac{1}{2^{n-1}}, \frac{1}{2^{m-1}}\} \\ \\ d(x,z) & = & max\{d(x,y), d(y,z)\}. \end{array}$$

Therefore, (X,d) is an ultrametric space, called often as the alphabetical metric.

[1] We accepted each word as a sequence. For example: Let us note $(a = 1, b = 2, \ldots, z = 29)$, then $x = (c,a,l,l) = (3,1,11,11,0,0,...)$.

To illustrate this metric for English alphabet
, let us consider three words x, y, z as follows:

$$x = (call),\ y = (calling),\ z = (car)$$

Then; $d(x,y) = \frac{1}{2^{5-1}} = 2^{-4}$ *(The* 5*. letter is different),*
$d(x,z) = \frac{1}{2^{3-1}} = 2^{-2}$, $d(x,x) = 0$.
For satisfying the axiom (iii)*:*
$x \neq y$ *and* $d(x,y) = \frac{1}{16} \leq max\{d(x,z) = \frac{1}{4}, d(z,y) = \frac{1}{4}\}$.

4.1 p-Adic Norm and p-Adic Metric

In this section, we will study an important examples of an ultrametric: p-adic metric.

Definition 4.1.1. *Let p a prime and $x \in \mathbb{Q}$ noted as $x = rp^n/s$ for $r,s,n \in \mathbb{Z}$ i.e. p divides neither r nor s. Then $ord_p(x) := n$. The p-adic norm on $\mathbb{Q}$ is defined by*

$$|x|_p = \begin{cases} \frac{1}{p^{ord_p x}} & if\, x \neq 0 \\ \\ 0 & if\ x = 0 \end{cases}$$

Example 4.1.1. *Let $x = \frac{28}{16}$. We will calculate the* 2*-adic order of x:*

$$x = \frac{28}{16} = \frac{2^2.7}{2^4} = 2^{-2}.7$$

Then $ord(x)_2 = (-2)$ *and* $|x|_2 = 2^{-(-2)} = 4$.

Proposition 4.1.1. *Let p a prime. The p-adic norm $|.|_p : \mathbb{Q} \to \mathbb{R}_+$ satisfies the following properties for all $x, y \in \mathbb{Q}$:*

(**N1**) $|x|_p = 0$ *if and only if* $x = 0$

(**N2**) $|xy|_p = |x|_p|y|_p$

(**N3**) $|x+y|_p \leq max\{|x|_p, |y|_p\}$ *with equality if* $|x|_p \neq |y|_p$

Remark. The p-adic norm $|.|_p$ which satisfies (**N1**), (**N2**), (**N3**) of proposition is often called ultranorm or non-Archimedean[2]norm on $\mathbb{Q}$.

[2] A totally ordered group $(G, +, \leq)$ is said to be **archimedean** if for all elements $x, y \in G$ satisfying $0 < x < y$, there exits a $n \in \mathbb{N}$ such that $nx > y$.
For example:
$|625|_5 = \frac{1}{5^4} < |1|_5 = 1$ but there is not any $n \in \mathbb{N}$ such that $n.|625|_5 > |1|_5$. Hence, p-adic is non-Archimedean.

Proof. For all $x,y \in \mathbb{Q}$, we will show that $(\mathbb{Q}, |.|_p)$ is a (ultra)norm space:
Note that $x = rp^n/s$ and $y = kp^m/l$ for $r,s,n,m,k,l \in \mathbb{Z}$.
For (**N1**), $|x|_p = |rp^n/s|_p = p^{-n} = 0$ if and only if $x = 0$.
For (**N2**), $|xy|_p = |rk.p^{n+m}/s.l|_p = p^{-(n+m)} = p^{-n}.p^{-m} = |x|_p|y|_p$.
For (**N3**), if $n < m$, $\ x+y = p^n(r/s + p^{m-n}.k/l)$.
Then $|x+y|_p \leq max\{|x|_p, |y|_p\}$.
Hence, $|.|_p$ is ultranorm (non-Archimedean norm) on $\mathbb{Q}$.

Example 4.1.2. *Calculate $|7|_5$ and $|1|_5$:*
$x = 7 = \frac{7}{1}.5^0$*, then* $ord_5(7) = 0$ *and* $|7|_5 = \frac{1}{5^0} = 1$
$x = 1 = \frac{1}{1}5^0$*, then* $ord_5(1) = 0$ *and* $|1|_5 = \frac{1}{5^0} = 1$

Definition 4.1.2. *The metric on $\mathbb{Q}$ which is induced by the $p-$adic norm, d is called $p-$adic metric with*

$$d_p(x,y) = |x-y|_p.$$

Example 4.1.3. *Let $x,y \in \mathbb{Q}$ and $p = 5$, we will give the $5-$adic distance for (x,y):*

$$\begin{aligned}
d_5(x,y) &= |x-y|_5 \\
d_5(7,6) &= |7-6|_5 = |1|_5 = \frac{1}{5^0} = 1 \\
d_5(7,1) &= |7-1|_5 = |6|_5 = \frac{1}{5^0} = 1 \\
d_5(626,1) &= |626-1|_5 = |625|_5 = \frac{1}{5^4}
\end{aligned}$$

Proposition 4.1.2. *The $p-$adic metric on $\mathbb{Q}$ is an ultrametric.*

Proof. Consider the map $u : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}_+$, we will satisfy the following conditions for all $x,y,z \in \mathbb{Q}$:

i. If $u(x,y) = |x-y|_p = 0$, $\ x-y = 0$ therefore $x = y$. Conversly, if $x = y$, then $x-y = 0$ due to the definition of p-adic norm; $u(x,y) = |x-y|_p = 0$.

ii. **(Symmetry)** As $|x-y|_p = \frac{1}{p^{ord_p(x-y)}}$ and noted $ord_p(x-y) = n$ which is obtained from $(x-y) = rp^n/s$ for $r,s,n \in \mathbb{Z}$ i.e. p divides neither r nor s. Then also $(y-x) = -rp^n/s$ and $ord_p(y-x) = n$. Consequently;
$|x-y|_p = \frac{1}{p^{ord_p(x-y)}} = \frac{1}{p^{ord_p(y-x)}} = |y-x|_p$.

iii. **(Strong Triangle Inequality)**
As $x-z=(x-y)+(y-z)$, we have:

$$ord_p(x-z) \geq min\{ord_p(x-y), ord_p(y-z)\}$$

which implies
$|x-z|_p = \frac{1}{p^{ord_p(x-z)}} \leq max\{\frac{1}{p^{ord_p(x-y)}}, \frac{1}{p^{ord_p(y-z)}}\} = max\{|x-y|_p, |y-z|_p\}$.
Thus, $u(x,z) \leq max\{u(x,y), u(y,z)\}$.

Corollary 1. $(\mathbb{Z}, |.|_p)$ is an ultrametric space.

Example 4.1.4. *For $p=5$, we observe 626 and 1 are much closer to each other than 7 and 1:*

$$|626-1|_5 = |625|_5 = \frac{1}{5^4}$$

$$|7-1|_5 = |6|_5 = \frac{1}{5^0} = 1$$

$$1 > \frac{1}{5^4}$$

We will give the next expansion using the Holly(2001) in the references.

Definition 4.1.3. *(Holly, 2001) Let p a prime number. All rational number can be written as a series of the form*

$$\sum_{k=n}^{\infty} b^k p^k \texttt{ for some } n \in \mathbb{Z}, \texttt{ and } b^k \in \{0,1,...,p-1\} \texttt{ for each } k \geq n.$$

*This series is called the **p-adic expansion** of the number.*

Remark. Positive integers have finite expansions, even though negative integers and many non-integers are represented by infinite series.

Example 4.1.5. *i. The $3-$adic expansion of 26 is:*

$$26 = 2 + 2.3 + 2.3^2$$

ii. The $3-$adic expansion of $\frac{14}{3}$ is:

$$\frac{14}{3} = 2.3^{-1} + 1.3^0 + 1.3^1$$

iii. The $3-$*adic expansion of* -1 *is:*

$$-1 = 2 + 2.3 + 2.3^2 + 2.3^3 + \ldots$$

It suffices to show that $\lim_{n\to\infty} |(2 + 2.3 + 2.3^2 + \ldots + 2.3^n) - (-1)|_3 = 0$

$$\begin{aligned}
& \lim_{n\to\infty} |3 + 2.3 + 2.3^2 + \ldots + 2.3^n|_3 = \lim_{n\to\infty} |1 + 2.3^0 + 2.3^1 + 2.3^2 + \ldots + 2.3^n|_3 \\
= \ & \lim_{n\to\infty} |1 + 2.(3^0 + 3^1 + 3^2 + \ldots + 3^n)|_3 \\
= \ & \lim_{n\to\infty} |1 + 2.(\frac{1 - 3^{n+1}}{1 - 3})|_3 = \lim_{n\to\infty} |1 - 1 + 3^{n+1}|_3 \\
= \ & \lim_{n\to\infty} |3^{n+1}|_3 \\
= \ & 0
\end{aligned}$$

Theorem 4.1.1. *(Gouvêa, 1997, Koblitz, 1984) Each* $p-$*adic number has a unique* $p-$*adic expansion of the form*
$\sum_{k=n}^{\infty} b^k p^k$ *for some* $n \in \mathbb{Z}$*, and* $b^k \in \{0, 1, \ldots, p-1\}$ *for each k.*

Example 4.1.6. *(Holly,2001) (Picture of a 3-adic metric space). Consider* $(\mathbb{Z}, |.|_3)$*. The 3-adic metric between any two integers x and y can be at most* 1*, corresponding to* $ord_3(x-y) = 0$ *and all possible distances are* $1/3, 1/3^2, 1/3^3$*, etc. Thus, this ultrametric space can be viewed as a rooted tree, where the root represent the least possible order* $ord_3 = 0$ *and highest possible distance* $1/3^0 = 1$*. There will be infinitely many orders and distance levels.*
The 3-adic metric between any two integers x and y is the height, (as labeled in the distance column) which one must climb in traversing a path from x's leaf to y's leaf.

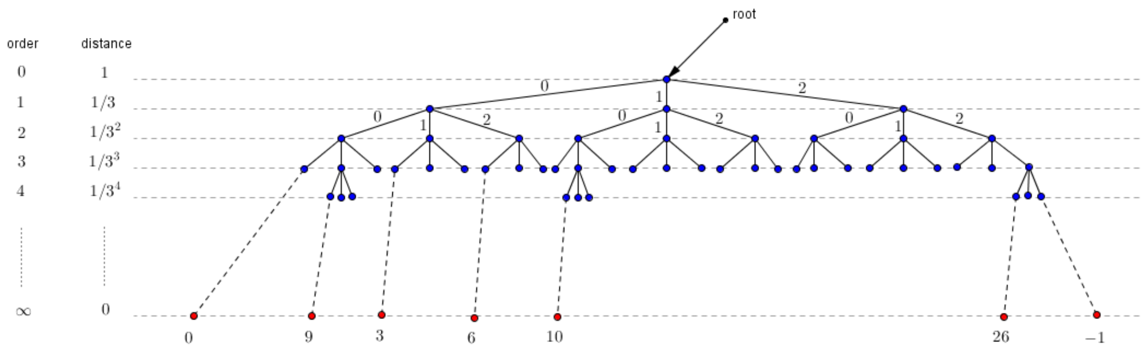


Figure 4.1: This figure is adapted from [2].

4.1.1 Balls in Ultrametric Spaces

In this section, we give some topological properties of an ultrametric space following (Holly,2001, Kirk, 2001).

Definition 4.1.4. *Let (X,u) be an ultrametric space. The closed ball of radius $r>0$ centered at a point a in X, denoted by $B(a,r)$, is defined as follows:*

$$B(a,r)=\{x\in X|\ u(x,a)\leq r\}.$$

Remark. Let $x,y,z\in X$. Each three points x,y,z of any ultrametric space represent either vertices of an equilateral triangle or vertices of an isoceles triangle with the unequal side being the shortest one. So each triangle in an ultrametric space is isoceles or equilateral. We will show that $u(x,z)=u(x,y)=u(y,z)$ or $u(x,z)=u(x,y)$ (or $u(x,z)=u(y,z)$). When $u(x,y)=u(y,z)$ by the condition **(iii)** of ultrametric, we have

(1) $u(x,z)\leq u(x,y)$ or $u(x,z)\leq u(y,z)$

In the case (1), we have two cases

(1a) $u(x,z)<u(x,y)$

(1b) $u(x,z)=u(x,y)$

If $u(x,y)\neq u(y,z)$ by the condition **(U3)** we have

(2) $u(x,z)\leq u(x,y)$ or $u(x,z)\leq u(y,z)$

In the case (2) we have two cases

(2a) $u(x,z)=u(x,y)$

(2b) $u(x,z)<u(x,y)$

But also $u(x,y)\leq max\{u(x,z),u(z,y)\}$. If $max\{u(x,z),u(z,y)\}=u(x,z)$ or $max\{u(x,z),u(z,y)\}=u(z,y)$, therefore $u(x,y)<u(x,z)$. This case is a contradiction with (2b). So, $u(x,z)=u(x,y)$.

Proposition 4.1.3. *Let (X,u) be an ultrametric space if $y\in B(x,r)$ then $B(x,r)=B(y,r)$. That is, every element of a closed ball is also center of the ball. In that sense, the ultrametric can be interpreted as a democratic metric.*

Proof: By definition, $y \in B(x,r)$ if and only if $u(x,y) < r$. Let $z \in X$ such that $u(x,z) < r$. We have $u(x,y) \leq max\{u(x,z),u(z,y)\} < r$. This implies that $z \in B(y,r)$ which shows that $B(x,r) \subset B(y,r)$. In a similar way, we have
$B(x,r) \supset B(y,r)$. Thus $B(x,r) = B(y,r)$.

Proposition 4.1.4. *Let (X,u) be an ultrametric space. Every closed balls of X are either distinct or nested.*

Proof. Let $a \neq b$. Let $B(a,r_1) = \{x \in |u(a,x) \leq r_1\}$ and $B(b,r_2) = \{x \in |u(b,x) \leq r_2\}$ two closed balls in X with the property that $r_1 \leq r_2$. Therefore, there are two possibilities either $B(a,r_1) \cap B(b,r_2) = \emptyset$ or $B(a,r_1) \subseteq B(b,r_2)$. In particular, if $r_1 < r_2$ and if
$a \in B(b,r_2)$ then $B(a,r_1) \subseteq B(b,r_2)$.
Two closed balls in (X,u) are contained in each other, i.e., $B(a,r_1) \cap B(b,r_2)$ is non-empty then either $B(a,r_1) \subseteq B(b,r_2)$ or $B(a,r_1) \supseteq B(b,r_2)$.

Proposition 4.1.5. *Let (X,u) be an ultrametric space. Let $B(a,r_1)$ and $B(b,r_2)$ two closed balls in X. If $B(a,r_1) \subseteq B(b,r_2)$ and if $b \notin B(a,r_1)$ then $u(b,a) = u(b,z)$ for each $z \in B(a,r_1)$.*

Proof. Since $u(b,a) > r_1$ and as $u(a,z) \leq r_1$ for every $z \in B(a,r_1)$, we have
$u(b,a) > u(a,z)$. However, $u(b,z) = max\{u(b,a),u(a,z)\} = u(a,z)$.

Example 4.1.7. *Remember the discrete metric is an ultrametric. Let us define the closed balls in discrete metric. In particular, if $r < 1$ then the closed ball $B(x,r) = \{x\}$ and if $r \geq 1$ then $B(x,r) = X$. Moreover, if $y \in B(x,r)$ then $B(x,r) = B(y,r)$.*

Definition 4.1.5. *Any collection of balls in B which is totally ordered by inclusion is called a nest of ball.*

The following figure gives the ball structure of 3-adic space whose picture is given by 5.1.

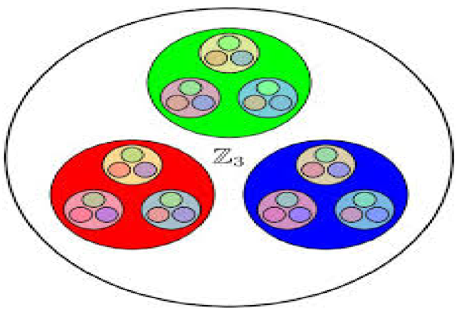


Figure 4.2: $\mathbb{Z}_3$

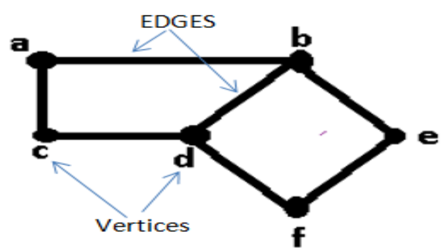
5 Hierarchies and Minimum Spanning Tree

In this chapter, we will first give some definitions about basic graph theory. Next, we will explain our method of data analysis. In order to give a meaningful representation of the data, we will use our knowledge of metric for illustrating of a graph and then we will obtain an ultrametric via a graphical theoretical tool, MST. Finally we make inferences of hierarchy and clustering of the data via ultrametric.

5.1 Basics of Graph Theory

In this section, we will give some basic definitions of graph theory following the reference (Ruohonen, 2013).

Definition 5.1.1. *A graph $G = (V,E)$ is a pair where V is a non-empty set of elements and E is a subset of $V \times V$. The elements of V are called* ***vertex*** *and the elements of E are called* ***edges***.



Definition 5.1.2. *Let $(u,v) \in V \times V$ be a pair of vertices.* ***A loop*** *from u to v, denoted by $P(u,v)$ which is a sequence of vertices $< u, v_0, ..., v_k, v >$ such that for all $i, 0 \leq i < k$ and there is an edge between the vertices v_i and v_{i+1} in G.*

Definition 5.1.3. *The graph $G' = (V',E')$ is a* ***subgraph*** *of G when $V' \subseteq V$ and $E' \subseteq E$.*

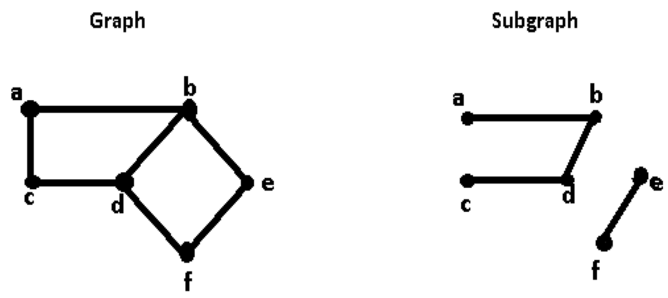


Figure 5.1: An example for one graph and its subgraph

Definition 5.1.4. *The graph $G = (V,E)$ is called connected if and only if there is at least one loop between all pairs of vertices $u,v \in V$.*

Definition 5.1.5. *Let $G = (V,E)$ be a graph and $u,v \in V$ two vertices of G. The distance $d(u,v)$ between the vertices u and v in G is called sometimes* ***weight*** *between u and v.*

Proposition 5.1.1. *Let T = (V, E) be a graph. T is called* ***a tree*** *if and only if there is a single loop between all pairs of vertices $u,v \in V$.*

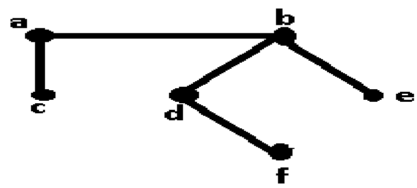


Figure 5.2: An example for a tree

5.2 Clustering and Hierarchies

A clustering is an ordering of inputs into groups on the basis of their relationships. We will use clustering algorithms for data analysis.
The measure of a similarity or dissimilarity index is a major decision. The clustering can be explained as a hierarchy, generally pictured as an inverted tree where the inputs(the leaves of the tree) are indicated horizontally on the bottom line. Going from the bottom up, several leaves merge into a branch; several such branches merge into a higher branch. Lower class comprise a closer distance of inputs than higher class.

Definition 5.2.1. *Let $\Omega = \{x_1, \ldots, x_n\}$ be a finite set with $n \in \mathbb{Z}_+$.* ***A Hierarchy*** *$\mathcal{H} = \{H \in P(\Omega)\}$ is a subset of $P(\Omega)$ such that*

i. $\Omega \in \mathcal{H}$,

ii. For every $x_i \in \Omega, \{x_i\}$ belongs to H for $1 \leq i \leq n$,

iii. For each pair $H, H' \in \mathcal{H}$, we have $H \cap H' = \emptyset$; or if $H \cap H' \neq \emptyset$, we have $H \subset H'$ or $H' \subset H$.

Example 5.2.1. *Let $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ be a finite set. Suppose $\mathcal{H} = \{H_1 = \{x_1\}, ..., H_5 = \{x_5\}, H_6 = \{x_1, x2\}, H_7 = \{x_1, x_2, x_3, x_4\}$, $H_8 = \Omega\}$. We will show that $\mathcal{H}$ is a hierarchy:*

i. $\Omega \in \mathcal{H}$,

ii. For every $x_i \in \Omega$, we have $\{x_i\} \in \mathcal{H}$ for $1 \leq i \leq 5$,

iii. For each pair $H_k, H_j \in \mathcal{H}$, we have $H_k \cap H_j = \emptyset$; or if $H_k \cap H_j \neq \emptyset$, we have $H_k \subset H_j$ or $H_k \subset H_j$ where $1 \leq k, j \leq 8$.

$$H_1 \cap H_6 = \{x_1\} \quad \Rightarrow \quad H_1 \subset H_6$$

$$H_6 \cap H_7 = H_6 \quad \Rightarrow \quad H_6 \subset H_7$$

$$H_7 \cap H_8 = H_7 \quad \Rightarrow \quad H_7 \subset H_8$$

.

Hierarchy can be quantified, that means, when a positive real number can be associated with each class, the hierarchy becomes **an indexed hierarchy**.

Definition 5.2.2. *Let $\Omega = \{x_1, \ldots, x_n\}$ be a finite set and $f : \mathcal{H} \to \mathbb{N}$ be a function satisfying the following conditions:*

i. $f(H) = 0$ if and only if H is reduced to a single element of Ω (singleton),

ii. If $H \subset H'$, then $f(H) \leq f(H')$.

The pair $(\mathcal{H}, f)$ is an indexed hierarchy on Ω where $\mathcal{H}$ denotes a given hierarchy on Ω. Note that for a given H, $f(H)$ corresponds to the level of aggregation.

Example 5.2.2. *Let $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ be a finite set.*
Suppose
$\mathcal{H} = \{H_1 = \{x_1\}, ..., H_5 = \{x_5\}, H_6 = \{x_1, x2\}, H_7 = \{x_3, x_4, x_5\}$,

$H_8 = \Omega\}$.
We will show that the pair $(\mathcal{H}, f)$ is an indexed hierarchy:
First of all, we will define f as:

$$f(H_i) = \begin{cases} 0 & if\ \#H_i = 1 \\ \#H_i & sinon \end{cases}$$

where $i = 1, ...5$.

i. $f(H_i) = 0$ *for all* $i = 1, ...5$, $f(H_6) = 2,\ f(H_7) = 3,\ f(H_8) = 5$

ii. As;

$$H_7 \subset H_8, \Rightarrow f(H_7) \leq f(H_8)$$

$$H_6 \subset H_8 \Rightarrow f(H_6) \leq f(H_8)$$

$$H_1 \subset H_6 \Rightarrow f(H_1) \leq f(H_6)$$

Hence, the pair $(\mathcal{H}, f)$ is an indexed hierarchy on Ω.

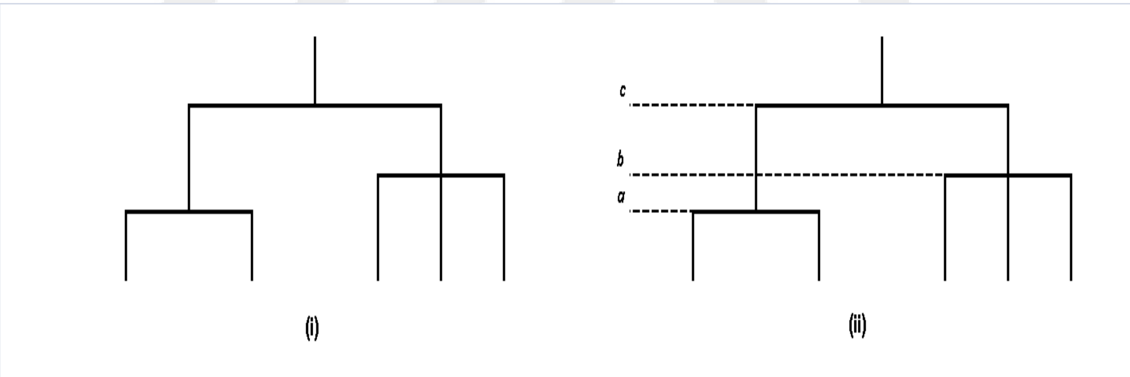


Figure 5.3: Hierarchical Trees: (i) hierarchy; (ii) indexed hierarchy, $a < b < c$

5.2.1 Metric Spaces versus Hierarchies

Definition 5.2.3. *Let (Ω, d) be a metric space.*
Let $\delta : \Omega \to \mathbb{N}$ be the aggregation or dissimilarity index such that
$\delta(H_i, H_k) \geq 0$ *and* $\delta(H_i, H_k) = \delta(H_k, H_i)$ *with* $i, k \in \mathbb{N}$.
Index δ is called Single-linkage which is defined as the distance between two points H_i and H_k in Ω:

$$\delta(H_i, H_k) = min\{d(x, y) |\ x \in H_i,\ y \in H_k\}$$

where $i, k \in \mathbb{N}$. [1]

Definition 5.2.4. *Let* Ω *be a set with partition* $\Omega = \cup_i \Omega_i$, d *be a metric on* Ω. *Then, the given partition induces a k of trivial ultrametric defined as follows:*

$$d_k(x,y) = \begin{cases} 0 & if \quad x = y \\ 1 & if\, x \in \Omega_i,\ y \in \Omega_j\ (i \neq j) \\ k & if\, i = j,\ 0 < k < 1 \end{cases}$$

This trivial ultrametric is used for the general connection between indexed hierarchies and ultrametrics which is also visible on the hierarchical trees.

Example 5.2.3. *Let* $\Omega = \{a,b,c\}$ *be a finite set.*
$\Omega_1 = \{a\},\ \Omega_2 = \{b\},\ \Omega_3 = \{c\},\ \Omega_4 = \{a,b\},$
$\Omega_5 = \{a,c\},\ \Omega_6 = \{b,c\},\ \Omega_7 = \{a,b,c\}$
As $a \in \Omega_1,\ b \in \Omega_2$ *we have* $d(a,b) = 1$.
As $a,b \in \Omega_4$; $d(a,b) = k$ *where* $0 < k < 1$.

Proposition 5.2.1. *Let* $(\mathcal{H}, f)$ *be an indexed hierarchy on* Ω.
$\varphi : \omega \times \omega \to \mathbb{R}$ *which is defined as:*

$$\varphi(x,y) = min_{H \in \mathcal{H}}\{f(H) \mid x \in H,\ y \in H\}$$

is an ultrametric. Conversely, each ultrametric φ *correspond to one and only one indexed hierarchy.*

Example 5.2.4. *Let* $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ *be a finite set.*
Suppose;
$\mathcal{H} = H_1 = \{x_1\}, ..., H_5 = \{x_5\}, H_6 = \{x_1, x_2\}, H_7 = \{x_3, x_4, x_5\},$
$H_8 = \{\Omega\}$.
We built an indexed hierarchy with f *such as*
$f(H_i) = 0$ *where* $i = 1, \ldots, 5$; $f(H_6) = 2$; $f(H_7) = 3$ *and* $f(H_8) = 5$.

[1] Index δ' is called Complete-linkage that is defined by

$$\delta'(H_i, H_k) = max\{d(x,y) | \ x \in H_i,\ y \in H_k\}$$

Let us examine the ultrametric distance $\varphi(x_i,x_j)$ *i.e.* $i,j=1,...5$:

$$\begin{aligned}\varphi(x_1,x_2) &= min_{H\in\mathcal{H}}\{f(H)\mid x_1\in H,\ x_2\in H\}\\ &= min_{H\in\mathcal{H}}\{f(H_6),f(H_8)\}=2\ ,\\ \varphi(x_1,x_3) &= min_{H\in\mathcal{H}}\{f(H_8)\}=5\\ \varphi(x_1,x_4) &= 5\ ,\ \varphi(x_2,x_3)=5\ ,\ \varphi(x_1,x_5)=5\ ,\\ \varphi(x_2,x_4) &= 5\ ,\ \varphi(x_3,x_4)=3\ ,\ \varphi(x_4,x_5)=3\ .\end{aligned}$$

5.2.2 Subdominant Ultrametric

Definition 5.2.5. *Let* $\mathcal{U}$ *be a set of ultrametrics on* Ω.
Let $\mathcal{U}^{<}=\{\delta\in\mathcal{U}\mid\delta\leq d\}$. ***The subdominant ultrametric*** $d^{<}$ *is defined as the upper limit of* $\mathcal{U}^{<}$. *This is the maximal element in* $\mathcal{U}^{<}$ *and by definition;*

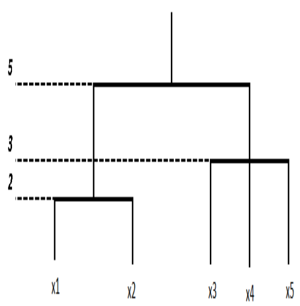
$$d^{<}(x,y)=max\{\delta(x,y)\mid\delta\in\mathcal{U},\delta\leq d\}.$$

The definition of $d^{<}$ *becomes clear via that above optimization method.*

Remark. We obtain the subdominant ultrametric $d^{<}$ like as:

1. The hierarchy $(\mathcal{H},f)$ is constructed.
2. Using by Proposition [4.2.1], $d^{<}$ is deduced from $\mathcal{H}$.

Example 5.2.5. *If we will examine the latest example: Firstly, we have established a hierarchical structure. After that, we deduced a subdominant ultrametric matrix from this hierarchical tree.*



	x1	x2	x3	x4	x5
x1	0	2	5	5	5
x2		0	5	5	5
x3			0	3	3
x4				0	3
x5					0

Figure 5.4: Hierarchical Tree with $d^{<}$ matrix

5.2.3 The minimum-spanning-tree (MST)

Definition 5.2.6. *It is called an undirected graphs which have edges that do not have a direction.*

Definition 5.2.7. *For a connected undirected graph $G = (V, E)$, a spanning tree is a tree $T = (V, E')$ with $E' \subseteq E$.*

Example 5.2.6. *A graph on the left and one possible spanning tree on the right in the figure* 4.3*:*

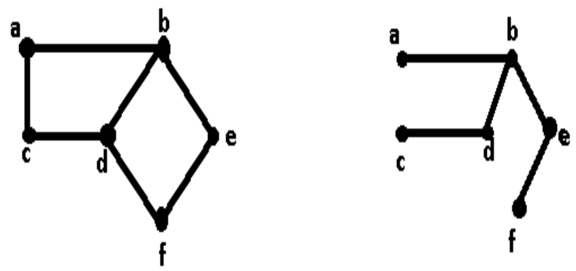


Figure 5.5: Graph and spanning tree

Definition 5.2.8. *The minimum spanning tree (MST) is given an connected undirected weight graph $G(V, E, w)$ with non-negative weights, find a spanning tree of minimum weight.*

5.2.4 MST Construction Method

Remark. MST which is a graphical tool for transition, allows us to do an optimization for obtaining $d^{<}$ from the weighted graph.

Remark. We can give the output of the subdominant ultrametric $d^{<}$ from the input of the metric d. This is called **the minimal-spanning-tree (MST)** construction method.
Let (Ω, d) be a metric space.

1. Construct the non-directed graph with the elements of Ω as vertices and the edge (x, y) has a length equal to $d(x, y)$ i.e. $x, y \in \Omega$.

2. Obtain $d^{<}$ via the construction of an MST on the connected graph so obtained. This MST will have the same vertices as Ω, but with minimal total length (Single-linkage).

3. When there is an MST on Ω, the distance $d^{<}(x,y)$ is given by:

$$d^{<}(x,y) = Max\{d(w_i, w_{i+1}),\ 1 \leq i \leq n-1\}$$

where $C_{xy} = \{(w_1,w_2),(w_2,w_3),\ldots,(w_{n-1},w_n)\}$ denotes the unique chain in MST, between x and y $(w_1 = x, w_n = y)$.

Example 5.2.7. *In this example, we will apply our construction method for obtaining hierarchical tree from the knowledge of metric with using MST. Let us explain the steps as follows:*

1. *We have a metric information and we are using matrix for showing the distance of each elements* (a,b,c,d,e)

2. *We use the some algorithms on "cocalc" for obtaining the non-directed graph from our d metric.*

```


B = [
[0,0.5,3,2,1],
[0.5,0,4,3,4],
[3,4,0,5,1],
[2,3,5,0,3],
[1,4,1,3,0]
]

G = nx.from_numpy_matrix(np.array(B))
m={0:'a',1:'b',2:'c',3:'d',4:'e'}
H=nx.relabel_nodes(G,m)
nx.draw(H, with_labels=True)
labels = nx.get_edge_attributes(H,'weight')
pos=nx.spring_layout(H)
nx.draw_networkx_edge_labels(H,pos,edge_labels=labels)
plt.show()
```

Figure 5.6: Graph Algorithms

3. *By using the one more algorithms (Kruskal), we can obtain MST from our undirected graph. MST will be a very useful transition tool for analyzing of subdominant ultrametric.*

```
B = [
[0,0.5,3,2,1],
[0.5,0,4,3,4],
[3,4,0,5,1],
[2,3,5,0,3],
[1,4,1,3,0]
]

G = nx.from_numpy_matrix(np.array(B))
m={0:'a',1:'b',2:'c',3:'d',4:'e'}
H=nx.relabel_nodes(G,m)
T = nx.minimum_spanning_tree(H)
nx.draw(T, with_labels=True)
plt.show()


labels = nx.get_edge_attributes(T,'weight')
pos=nx.spring_layout(T)
nx.draw_networkx_edge_labels(H,pos,edge_labels=labels)


plt.show()
```

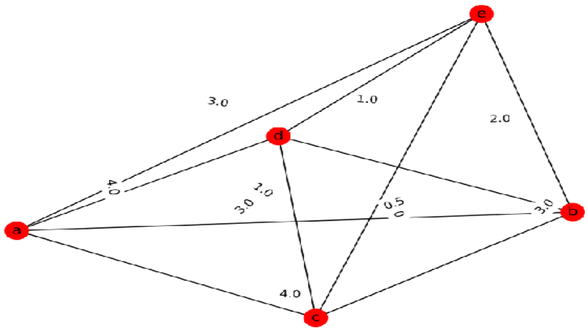
Figure 5.7: MST Algorithms

4. *We applied the Single-Linkage Method on MST and we obtained the $d^{<}$ matrix.*

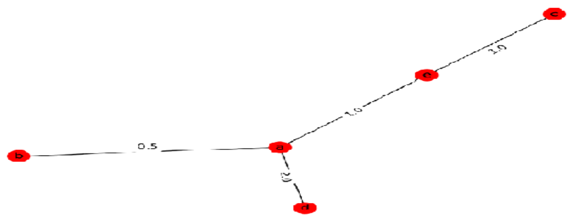
5. *From the $d^{<}$ matrix, we thus have the hierarchical tree.*

d	a	b	c	d	e
a	0	0,5	3	2	1
b		0	4	3	4
c			0	5	1
d				0	3
e					0

1.Matrix of metric.



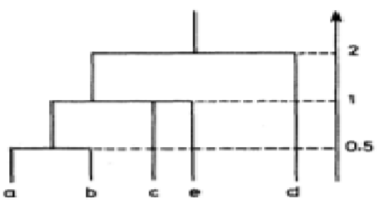
2. Non-Directed Graph



3. MST

$d^{<}$	a	b	c	d	e
a	0	0,5	1	2	1
b		0	1	2	1
c			0	2	1
d				0	2
e					0

4. Matrix of Subdominant Ultrametric



5. Hierarchical Tree

Figure 5.8: The subdominant ultrametric $d^{<}$ from (Ω, d)

6 An Application

In this chapter, we apply our results to a data of PISA[1]-Mathematical$([8]Holly, 2001)$ and Reading performances for 10 OECD[2] countries. Mathematical and reading performances, for PISA, measures the mathematical, reading literacy of a 15 year-olds to recognize, to use and to interpret mathematics, reading in various contexts.It gives an overall assessment in order to recognize the roles that mathematics, reading skills play in the world. In this context, we will try to observe and interpret the relationship between PISA mathematics and reading results of countries by revealing the hierarchical structure of this data. We deduce that which country has a closed or far relation with which country by using the topology of graph in these results. Finally, we will observe similarities and differences between mathematics and reading outcomes.

6.0.1 Data

The data used in this thesis consist of yearly data(3-year intervals) from the period 2003 to 2015. We selected 10 countries are: France, Netherlands, Brazil, Canada, Turkey, Ireland, Czech Republic, Russia, Norway, Mexico. This data contain the measures of 15-years-old school pupils' scholastic performance on mathematics and reading in these countries.

6.0.2 Methodology

First of all, we will do time evaluation table with the logarithmic return $(ln(t+1) - ln(t))$ where t represent the time (Figure 5.1). We adapt the logarithmic return process which is often used in financial data analysis for easy calculation in small data ranges.

[1]The Programme for International Student Assessment is a worldwide study by the OECD in member and non-member nations intended to evaluate educational systems by measuring 15-years-old school pupils' scholastic performance on mathematics, science and reading.

[2]The Organisation for Economic Cooperation and Development is a group of 34 member countries that discuss and develop economic and social policy.

Secondly, we will find the distance matrix with the using Pearson correlation (for the details, check the appendix) [3] of the countries with the entries of the table of time evaluation. Then to determine edges, we introduce a distance function respect to correlation coefficients as $d = (1 - corr^2)$ (Figure 5.2) i.e. the Pearson Correlation:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{E[X^2] - E[X]^2}\sqrt{E[Y^2] - E[Y]^2}}$$

After getting the distance matrix, we use the algorithm (via "cocalc") for illustration of graph ((Figure 5.3) and MST (review the appendix) via the weights (we use the single-linkage method) (Figure 5.4) for catching the optimized relation of countries. From MST, we calculate the matrix of subdominant ultrametric $d^<$ (Figure 5.5) and we able to observe the hierarchical structure (Figure 4.6).

	MATH				
	2003	2006	2009	2012	2015
France	511	496	497	495	493
Netherlands	538	531	526	523	512
Brazil	356	370	386	389	377
Canada	532	527	527	518	516
Turkey	423	424	445	448	420
Ireland	503	501	487	501	504
Czech Republic	516	510	493	499	492
Russia	468	476	468	482	494
Norway	495	490	498	489	502
Mexico	385	406	419	413	408

Time Evoluation	2006-2003	2009-2006	2012-2009	2015-1012
ln				
France	-0,02979	0,00201	-0,00403	-0,00405
Netherlands	-0,01310	-0,00946	-0,00572	-0,02126
Brazil	0,03857	0,04233	0,00774	-0,03133
Canada	-0,00944	0,00000	-0,01723	-0,00387
Turkey	0,00236	0,04834	0,00672	-0,06454
Ireland	-0,00398	-0,02834	0,02834	0,00597
Czech Republic	-0,01170	-0,03390	0,01210	-0,01413
Russia	0,01695	-0,01695	0,02948	0,02459
Norway	-0,01015	0,01619	-0,01824	0,02624
Mexico	0,05311	0,03152	-0,01442	-0,01218

Figure 6.1: Data and Time evaluation

DISTANCE1				ln						
	FR	NE	BR	CA	TR	IRE	CZE	RUS	NOR	MEX
France	0	0,983	0,90213	0,91503	0,99012	0,99651	0,97017	0,90628	0,77243	0,60423
Netherlands		0	0,608	0,75485	0,33709	0,98574	0,90569	0,96004	0,49157	0,99004
Brazil			0	0,993	0,187165	0,685987	0,880451	0,586525	0,838083	0,316828
Canada				0	0,997	0,24357	0,073131	0,445067	0,229354	0,899378
Turkey					0	0,772	0,91747	0,501602	0,501602	0,726202
Ireland						0	0,057	0,142397	0,721331	0,527038
Czech Republic							0	0,244	0,487352	0,71723
Russia								0	0,841	0,702067
Norway									0	0,981
Mexico										0

Figure 6.2: Distance Matrix ($d = (1 - corr^2)$)

[3]Correlation is the statistical method for giving the data entries relations and similarities.

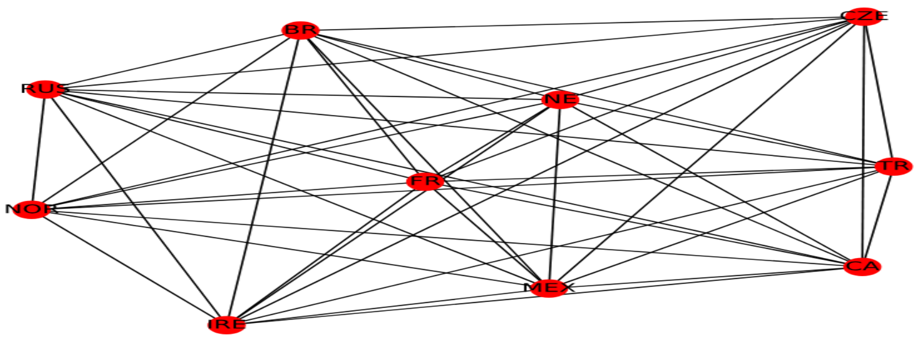


Figure 6.3: Complete Graph

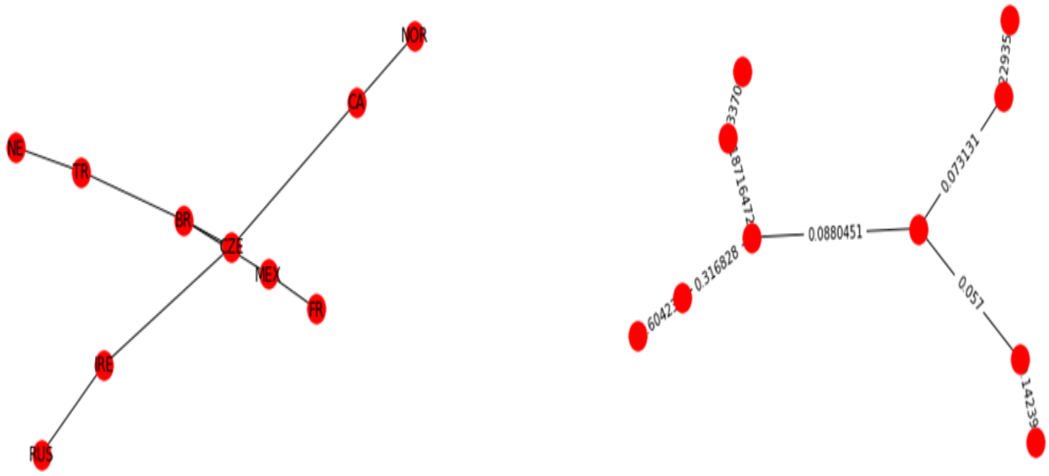


Figure 6.4: MST with the weights

				In						
	FR	NE	BR	CA	TR	IRE	CZE	RUS	NOR	MEX
France	0	0,229	0,14200	0,60400	0,14200	0,18700	0,14200	0,33700	0,60400	0,14200
Netherlands		0	0,229	0,60400	0,22900	0,33700	0,22900	0,33700	0,60400	0,22900
Brazil			0	0,604	0,073	0,187	0,088	0,337	0,604	0,057
Canada				0	0,316	0,316	0,316	0,337	0,604	0,073
Turkey					0	0,187	0,088	0,337	0,604	0,073
Ireland						0	0,187	0,337	0,604	0,187
Czech Republic							0	0,337	0,604	0,088
Russia								0	0,604	0,337
Norway									0	0,604
Mexico										0

Figure 6.5: Matrix of Subdominant ultrametric $d^{<}$

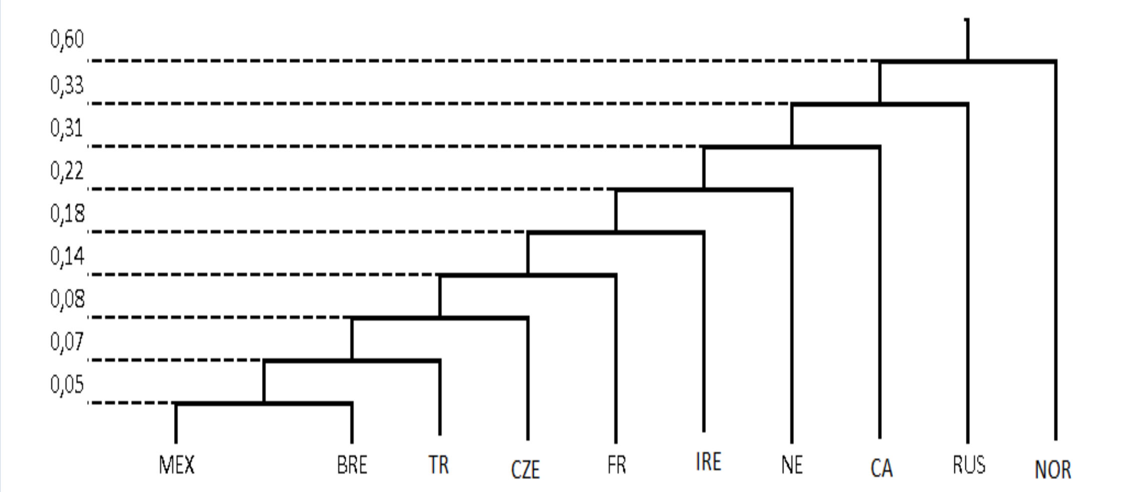


Figure 6.6: Hierarchical Tree

In the second part, we apply our all procedures on data of PISA Reading Performance. We use same methodologies with Mathematical Performance of PISA.

	Reading				
	2003	2006	2009	2012	2015
Brazil	507	501	506	509	499
Canada	528	527	524	523	527
Czech Republic	489	483	478	493	487
France	496	488	496	505	499
Ireland	515	517	496	523	521
Mexico	400	410	425	424	423
Netherlands	513	507	508	511	503
Norway	500	484	503	504	513
Russia	442	440	459	475	495
Turkey	441	447	464	475	428

Time Evoluation	2006-2003	2009-2006	2012-2009	2015-1012
ln				
Brazil	-0,01190	0,00993	0,00591	-0,01984
Canada	-0,00190	-0,00571	-0,00191	0,00762
Czech Republic	-0,01235	-0,01041	0,03090	-0,01225
France	-0,01626	0,01626	0,01798	-0,01195
Ireland	0,00388	-0,04147	0,05301	-0,00383
Mexico	0,02469	0,03593	-0,00236	-0,00236
Netherlands	-0,01176	0,00197	0,00589	-0,01578
Norway	-0,03252	0,03851	0,00199	0,01770
Russia	-0,00454	0,04228	0,03426	0,04124
Turkey	0,01351	0,03733	0,02343	-0,10419

Figure 6.7: Data and Time evaluation

DISTANCE					In					
	FR	NE	BR	CA	TR	IRE	CZE	RUS	NOR	MEX
France	0	0,07609	0,12598	0,64378	0,64972	0,98663	0,59228	0,97098	0,98991	0,98330
Netherlands	0,076	0	0,07564	0,47253	0,42818	0,93865	0,48621	0,87650	0,88644	0,98330
Brazil	0,125	0,075	0	0,29072	0,31740	0,99981	0,75116	0,89372	0,82638	0,85225
Canada	0,643	0,472	0,29	0	0,03709	0,97515	0,96133	0,96607	0,99915	0,47932
Turkey	0,649	0,428	0,317	0,037	0	0,99906	0,88487	0,93215	0,98924	0,63831
Ireland	0,986	0,938	0,999	0,975	0,999	0	0,057	0,97098	0,74592	0,45356
Czech Republic	0,592	0,486	0,751	0,961	0,884	0,057	0	0,244	0,99551	0,71233
Russia	0,97	0,876	0,893	0,966	0,932	0,97	0,244	0	0,841	0,92497
Norway	0,989	0,886	0,826	0,999	0,989	0,745	0,995	0,841	0	0,981
Mexico	0,983	0,983	0,852	0,479	0,638	0,453	0,712	0,924	0,981	0

Figure 6.8: Distance Matrix ($d = (1 - corr^2)$)

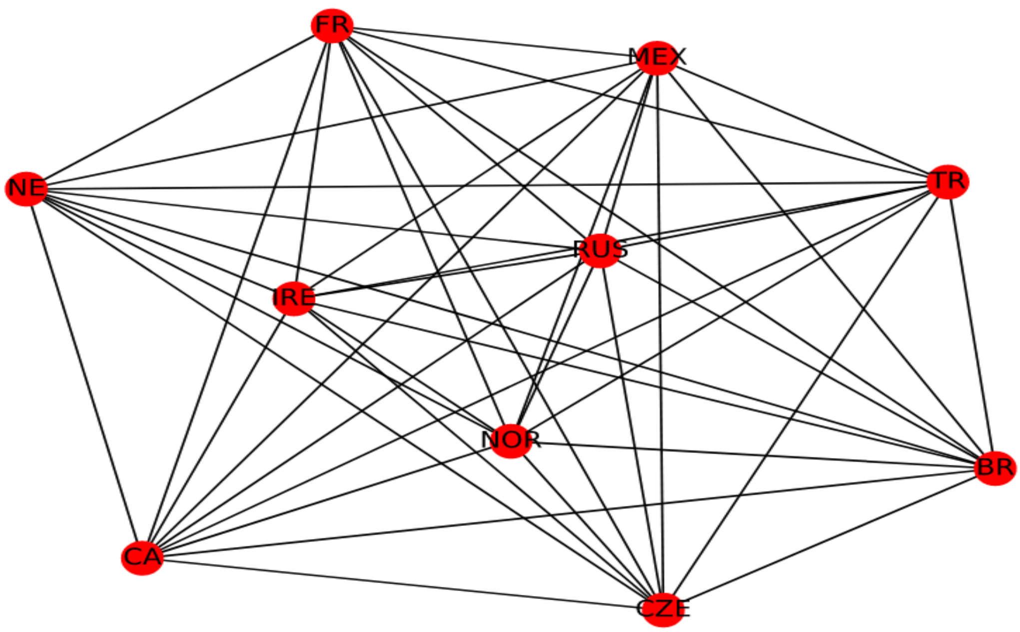


Figure 6.9: Complete Graph

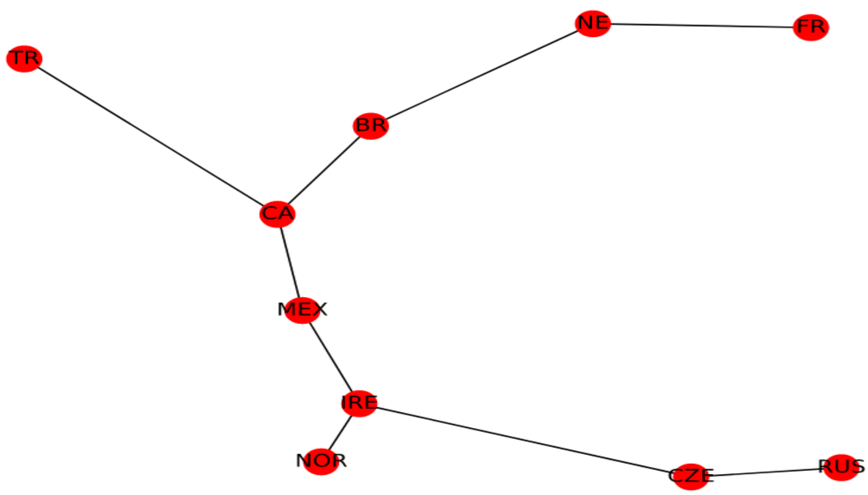


Figure 6.10: MST of PISA Reading Performance

Noticed that the weights of MST: $d('NE','BR') = 0.075$, $d('IRE','MEX') = 0.453$, $d('RUS','CZE') = 0.244$, $d('IRE','NOR') = 0.745$, $d('IRE','CZE') = 0.057$, $d('CA','TR') = 0.037$, $d('FR','NE') = 0.076$, $d('CA','MEX') = 0.479$, $d('CA','BR') = 0.29$

DISTANCE				In						
	FR	NE	BR	CA	TR	IRE	CZE	RUS	NOR	MEX
France	0	0,07600	0,07600	0,29000	0,29000	0,47900	0,47900	0,47900	0,74500	0,47900
Netherlands		0	0,07500	0,29000	0,29000	0,47900	0,47900	0,47900	0,74500	0,47900
Brazil			0	0,29000	0,29000	0,47900	0,47900	0,47900	0,74500	0,47900
Canada				0,00000	0,03700	0,47900	0,47900	0,47900	0,74500	0,47900
Turkey					0,00000	0,47900	0,47900	0,47900	0,74500	0,47900
Ireland						0,00000	0,05700	0,24400	0,74500	0,45300
Czech Republic							0,00000	0,24400	0,74500	0,45300
Russia								0,00000	0,74500	0,45300
Norway									0,00000	0,74500
Mexico										0,00000

Figure 6.11: Matrix of Subdominant ultrametric $d^{<}$

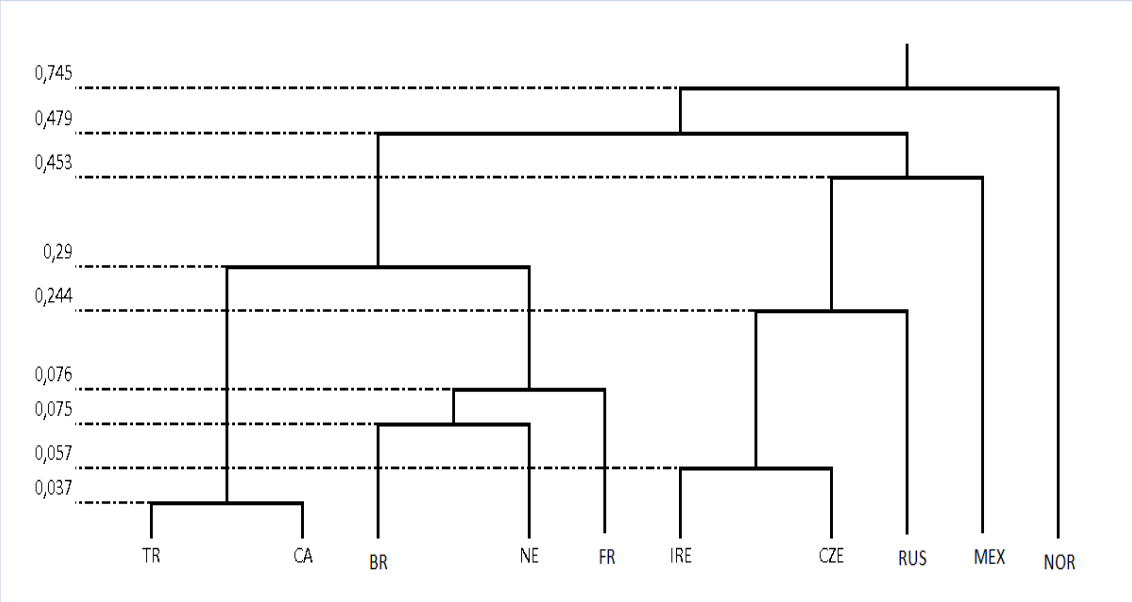


Figure 6.12: Hierarchical Tree

6.0.3 Results

In this section, following questions are answered in order to examine the hierarchical trees : How many clusters are obtained? Which countries' mathematical or reading performance evolution are associated more with each other? Which countries are in a separate cluster? What can the image of countries leaning against each other mean to us?

Results of Mathematical Performance:
Firstly, we can observe that Brazil and Mexico are in the same cluster and after then Turkey joins them. PISA data shows similarities in these countries of showing the changes in time intervals. Given the fact that the arguments that affect the exchange of mathematical knowledge in these countries and the common ones are omitted, it gives us a useful factor information.

Next, the information obtained from the hierarchical tree shows us the nested character of the similar countries as it seen by subclusters of Mexico, Brazil and Turkey as well as France, Ireland and Netherlands.

On the other hand, the evolution of mathematical performance of Russia sit between the ones of Canada and Norway that was not expected at first. This result shows us that similar characteristics can be deduced between unrelated countries. We also observe that the evolution of mathematical performance in Norway has a more different evaluation from the others.

Results of Reading Performance:

We obtain a tree structure in reading performance different from the one of mathematics. There are many different nested clusters. We can that countries' time evaluations in the different performance skills.

First of all, tree main clusters are scoped and under these clusters we have other nested sub-clusters:

Turkey and Canada which are in the same cluster, have a more closed relation than the other clusters. We can observe that Brazil, Netherlands and France in the same cluster also join them.

Ireland and Czech Republic are in same sub-cluster which reveals that the evolution of reading performance are highly correlated. In this sub-cluster, Russia enters and after joining Mexico, the cluster is completed.

As in the tree of mathematical performance, Norway again differs from all the countries in other clusters.

These results demonstrate us that the mathematical and reading knowledge in Norway may be key to us. These evaluation clusters can be a benchmark for the between of countries, for education and issues affecting of mathematics, reading knowledge in the considered countries.

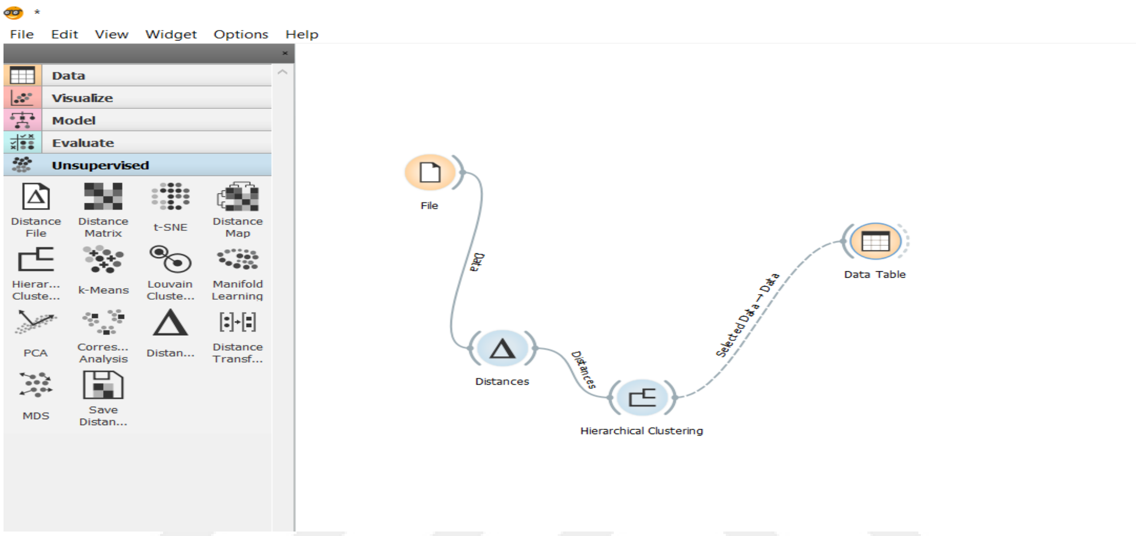
By comparing analogies between trees in different cases, common inferences or differing situations can be examined which may help to develop some education policies with respect to different performance skills.

The ultrametricity in locally minimal spanning tree that is constructed based on mathematical and reading performances can help us extract the information concealed in 10 countries by using the correlation coefficients of time evaluations of them. Geometrically, by visualizing the ultrametric space of data, we can see the clusters and sub-clusters' structures of countries in the different skills. On the other hand, from the hierarchical diagram of the ultrametric space we can visualize more clearly how a country specifically correlated to one another.

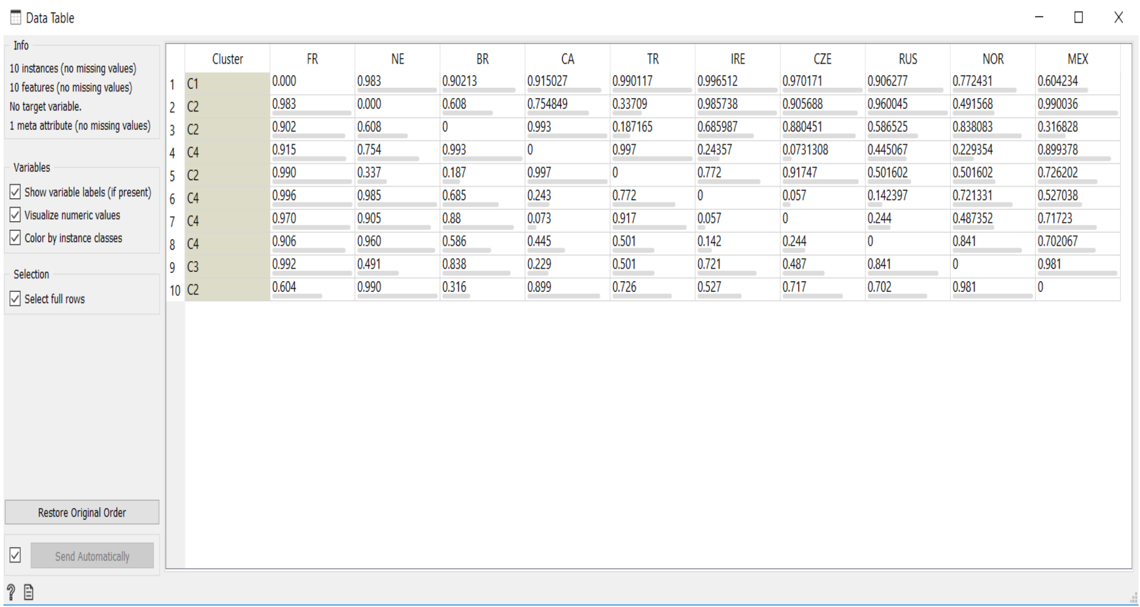
To summarize, the hierarchical clustering method using the procedures of our analysis performed on the correlation matrix helps us to detect statistically reliable aspects of the given issue which give us clues in comparative analysis.

6.0.4 Benchmarking

This section is devoted to the analysis of the same data by using "Orange"[4] which is a program with ready-made algorithms. We aimed to compare the results of our method presented in this thesis with a ready program.
The interface that is used in the program Orange is given as follows:



1. We used firstly our data of PISA-Mathematical Performance. We have arranged this data according to the structure of the program in software Excel:



	Cluster	FR	NE	BR	CA	TR	IRE	CZE	RUS	NOR	MEX
1	C1	0.000	0.983	0.90213	0.915027	0.990117	0.996512	0.970171	0.906277	0.772431	0.604234
2	C2	0.983	0.000	0.608	0.754849	0.33709	0.985738	0.905688	0.960045	0.491568	0.990036
3	C2	0.902	0.608	0	0.993	0.187165	0.685987	0.880451	0.586525	0.838083	0.316828
4	C4	0.915	0.754	0.993	0	0.997	0.24357	0.0731308	0.445067	0.229354	0.899378
5	C2	0.990	0.337	0.187	0.997	0	0.772	0.91747	0.501602	0.501602	0.726202
6	C4	0.996	0.985	0.685	0.243	0.772	0	0.057	0.142397	0.721331	0.527038
7	C4	0.970	0.905	0.88	0.073	0.917	0.057	0	0.244	0.487352	0.71723
8	C4	0.906	0.960	0.586	0.445	0.501	0.142	0.244	0	0.841	0.702067
9	C3	0.992	0.491	0.838	0.229	0.501	0.721	0.487	0.841	0	0.981
10	C2	0.604	0.990	0.316	0.899	0.726	0.527	0.717	0.702	0.981	0

[4]Orange is an open-source data visualization toolkit. It features a visual programming front-end for explorative data analysis and interactive data visualization, and can also be used as a Python library.

The hierarchical trees obtained by the ready tools are as follows:

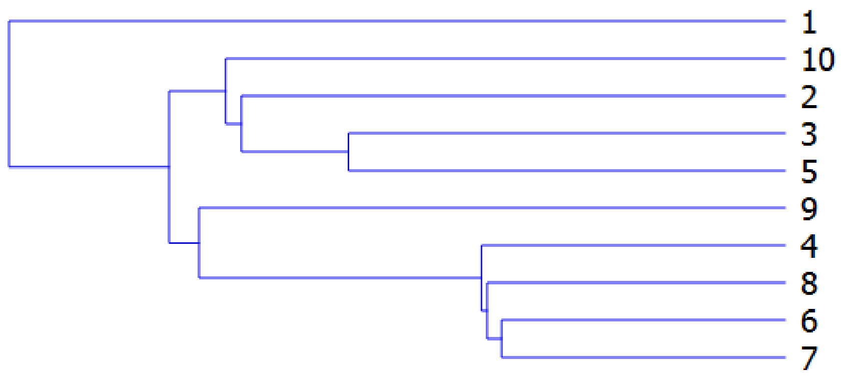


Figure 6.13: 1 :FR, 2 :NE, 3 :BR, 4 :CA, 5 :TR, 6 :IRE, 7 :CZE, 8 :RUS, 9 :NOR, 10 : MEX

We can observe that there are 4 separate clusters which are specified as follows:

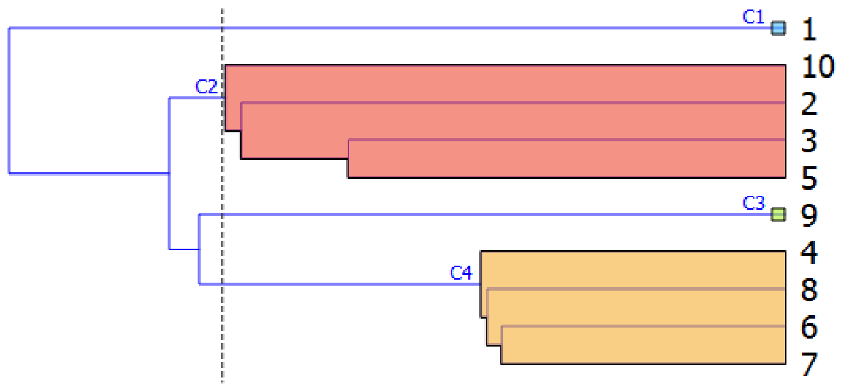
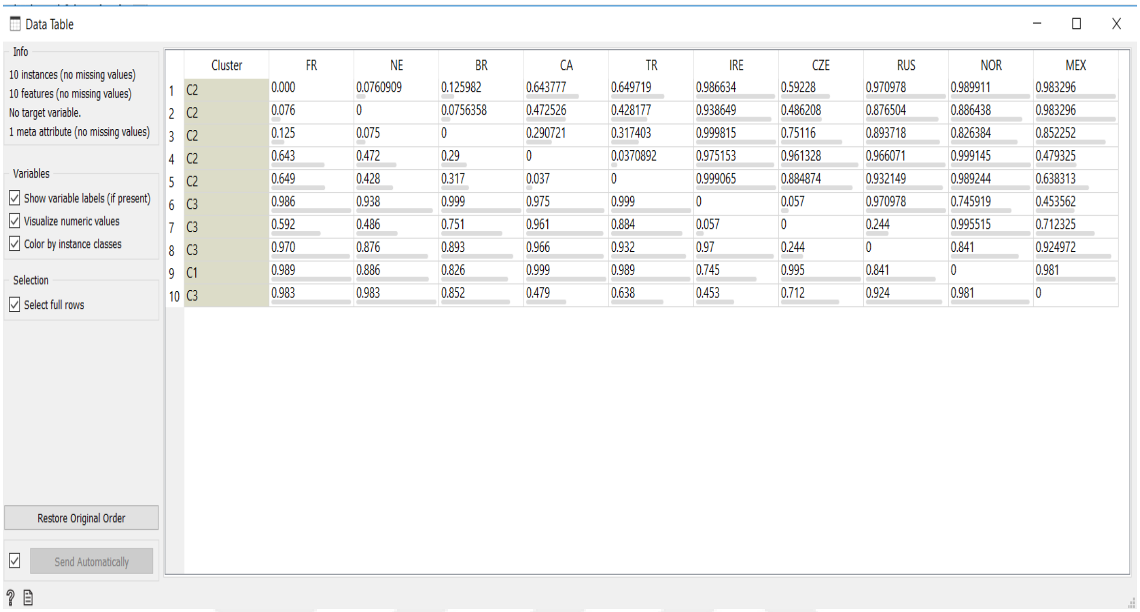


Figure 6.14: 1 :FR, 2 :NE, 3 :BR, 4 :CA, 5 :TR, 6 :IRE, 7 :CZE, 8 :RUS, 9 :NOR, 10 : MEX

We obtained a completely different hierarchical tree than the one that is obtained by our method.

2. Secondly, we edit and load the data of PISA-Reading Performance into the program Orange.



	Cluster	FR	NE	BR	CA	TR	IRE	CZE	RUS	NOR	MEX
1	C2	0.000	0.0760909	0.125982	0.643777	0.649719	0.986634	0.59228	0.970978	0.989911	0.983296
2	C2	0.076	0	0.0756358	0.472526	0.428177	0.938649	0.486208	0.876504	0.886438	0.983296
3	C2	0.125	0.075	0	0.290721	0.317403	0.999815	0.75116	0.893718	0.826384	0.852252
4	C2	0.643	0.472	0.29	0	0.0370892	0.975153	0.961328	0.966071	0.999145	0.479325
5	C2	0.649	0.428	0.317	0.037	0	0.999065	0.884874	0.932149	0.989244	0.638313
6	C3	0.986	0.938	0.999	0.975	0.999	0	0.057	0.970978	0.745919	0.453562
7	C3	0.592	0.486	0.751	0.961	0.884	0.057	0	0.244	0.995515	0.712325
8	C3	0.970	0.876	0.893	0.966	0.932	0.97	0.244	0	0.841	0.924972
9	C1	0.989	0.886	0.826	0.999	0.989	0.745	0.995	0.841	0	0.981
10	C3	0.983	0.983	0.852	0.479	0.638	0.453	0.712	0.924	0.981	0

The hierarchical trees with specifying clusters are revealed as follows:

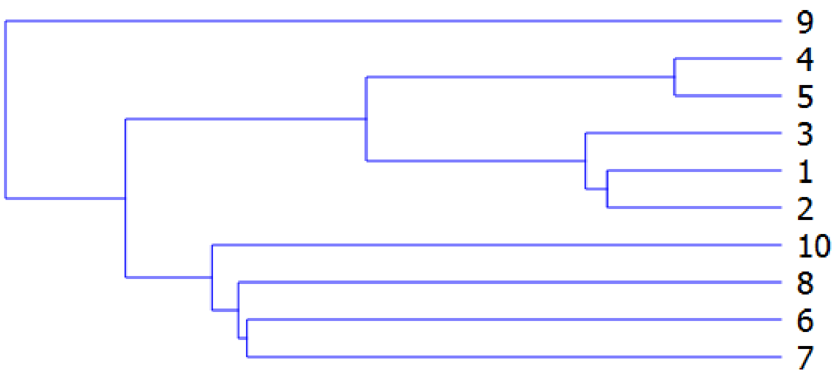


Figure 6.15: 1 :FR, 2 :NE, 3 :BR, 4 :CA, 5 :TR, 6 :IRE, 7 :CZE, 8 :RUS, 9 :NOR, 10 : MEX

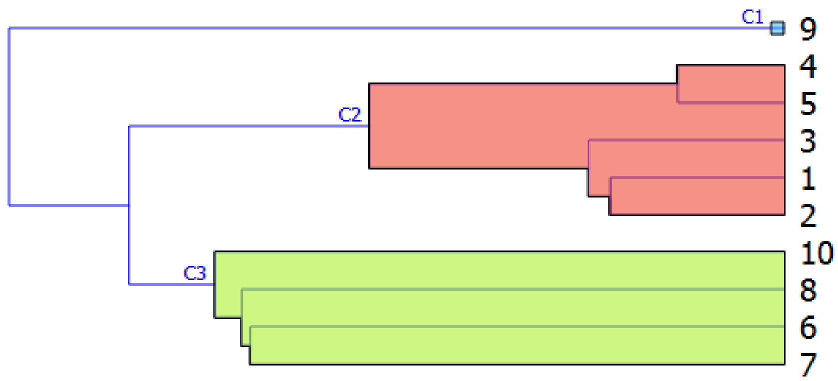


Figure 6.16: 1 :FR, 2 :NE, 3 :BR, 4 :CA, 5 :TR, 6 :IRE, 7 :CZE, 8 :RUS, 9 :NOR, 10 : MEX

This figure shows us that there are 3 main clusters. The result of hierarchical clustering by means of MST and Kruskal Algorithm and this clustering are almost identical. The only difference from the first tree structure is the order of the countries Brazil, Netherlands, France in the same cluster. Achieving a similar result might demonstrate the accuracy and reliability of our method.

7 Conclusion

This thesis aimed to obtain the hierarchical structures from big data by using metric and ultrametric spaces' features as well as graph theoretical tools.

This research consisted of three main parts. Part 1 consisted of two chapters, namely, Chapter 2 and Chapter 3 which give the mathematical preliminaries of the hierarchical structures where we studied metric spaces, normed spaces and ultrametric spaces.

In the second part, we gave the equivalence of agglomerative hierarchical clustering using single linkage and the graph theoretical models using minimal spanning tree model. We tackled here some notions of Graph Theory and mainly the question how to obtain a Minimum Spanning Tree (MST) from a graph.

Finally in Part 3, we analyzed the data obtained from PISA-Mathematical Performance and PISA-Reading Performance evolution over 4 years for 10 OECD countries. Ultrametric tree models of similarity and association are used to produce the representation of this data. We analyzed these particular data by using minimal spanning tree model which are obtained by using certain algorithms (Prim& Kruskal) and programs (Python, R).

In addition to statistical analysis given in the Appendix and the application given in Chapter 5, examining the mathematical background lying behind the hierarchical clustering (given in Chapter 2 – 5) was one of the important objective of this thesis.

References

[1] Benzécri, J. P. *Lecture notes, reproduced in Benzécri* 1965.

Benzécri, J. P. *L'analyse des données 1, La taxinomie* 1984, Dunod Paris.

https://cocalc.com

https://data.oecd.org/pisa/mathematics-performance-pisa.htm

Eric A.Hanushek & Marc Piopiunik & Simon Wiederhold *The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance* March 20, 2018. hanushek.stanford.edu

Erwin Kreyszig *Introductory Functional Analysis with Applications* University of Windsor.

Fernando Q.Gouvêa *P-adic numbers: An introduction, 2nd ed.* Berlin 1997 Springer-Verlag.

Florek, K., J. Lubaszewicz, J. Perkal, H. Stainhaus, S. Zubrzyeki 1951, Colloq. Math. 2, 282.

Gordon A. D. *Classification (Chapman and Hall, London)* 1981.

Hensel, Kurt *Über eine neue Begrund̈ung der Theorie der algebraischen Zahlen* 1897.

Hokky Situngkir & Yohanes Surya *On Stock Market Dynamics through Ultrametricity of Minimum Spanning Tree* May 10, 2005.

Jan E. Holly *Pictures of Ultrametric Spaces, the p-adic Numbers, and Valued Fields* October 2001.

Jan-Hendrik Evertse *P-ADIC NUMBERS* March 2011. http://www.math.leidenuniv.nl/ evertse/dio2011-padic.pdf

Keijo Ruohonen, Translation by Janne Tamminen, Kung-Chung Lee, Robert Piché *Graph Theory* 2013.

Kirk, W.A. And Khamsi, M.A. *An Introduction to Metric Spaces and Fixed Point Theory* 2001, John Willey and Sons, New York.

K.P. Hart, Jun-iti Nagata, J.E. Vaughan *Encyclopedia of General Topology* Elsevier 2004.

Maurice Fréchet *Sur Quelques Points du Calcul Fonctionnel* 1906.

Mehmet Ali Balci Hierarchies in Communities of Borsa Istanbul Stock Exchange

Murtagh, F. *Comput. J. 26, 354.* 1983.

Neal Koblitz *P-adic Numbers, P-adic Analysis, and Zeta Functions* New York 1984. Springer-Verlag.

Rammal, R. and G. Toulouse J. Phys. (Paris) Lett. 46, L945 1983.

Rammal, R., J. C. Angles d'Auriac, and B. Douqot J. Phys. (Paris) Lett. 44, L13. 1985.

R. Rammal & G. Toulouse & M. A. Virasoro *Ultrametricity for physicists* The American Physical Society Reviews of Modern Physics, Vol.58, No.3, July 1986.

R. N. Mantegna Hierarchical Structure in Financial Markets 1999.

Shepard, R. N. *Science 210, 390.* 1980.

Sokal, R. R, and P. H. A. Sneath Principles of Numerical Taxonomy (Freeman, San Francisco) 1963.

Sørensen Biol.Skr. 5, 1. 1948.

Appendix

Algorithms For illustrating our data outputs, we use some algorithms via program Sagemath (cocalc). The first algorithm is showing complete graph which is given by the distance matrix.

```
A = [ ] #matrix
G = nx.from_numpy_matrix(np.array(A))
m={ } #corners labeling
H=nx.relabel_nodes(G,m)
nx.draw(H, with_labels=True)
plt.show()
```

The second algorithm which is called Kruskal is for having Minimum Spanning Tree. This algorithm use also the distance matrix and typing labels for the corners and edges.

```
import matplotlib.pyplot as plt
import networkx as nx
import numpy as np
A = [ ] #matrix
G = nx.from_numpy_matrix(np.array(A))
m={ } #corners labeling
H=nx.relabel_nodes(G,m)
T = nx.minimum_spanning_tree(H)
nx.draw(T, with_labels=True)
plt.show()
labels = nx.get_edge_attributes(T,'weight')
pos=nx.spring_layout(T)
nx.draw(T,pos)
nx.draw_networkx_edge_labels(H,pos,edge_labels=labels)
plt.show()
```

The outcomes of these algorithms, we can easily transform data with large inputs into a meaningful graphs and trees.

Mean, Variance, Pearson Correlation

We will need to use statistical expressions in the following sections. Therefore, we have to start the subject with some terms.

Definition 7.0.1. *Let X be a random variable with a finite number of finite outcomes* $x_1,\dots,x_n$ *occurring with the probabilities* $p_1,\dots,p_n$*, respectively.* ***The expectation (mean)*** *of X is defined as:*

$$\mu = E[X] = \sum_{i=1}^{n} x_i p_i$$

Since $p_1+\dots+p_n = 1$.

Definition 7.0.2. *The variance of a random variable X is the expected value of the squared deviation from the mean of X,* $\mu = E[X]$*:*

$$Var(X) = E[(X-\mu)^2]$$

The variance is denoted by $\sigma_X^2, \sigma^2, Var(X)$*. Expressed by:*

$$Var(X) = E[(X-E[X])^2] = E[X^2 - 2XE[X] + E[X]^2]$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

Remark. In probability theory, the variance is the square of **the standard deviation**.

Example 7.0.1. *A fair six-sided can be modeled as a discrete random variable, X with outcomes* 1 *through* 6*, each with equal probability* $\frac{1}{6}$.
The expected value of X is

$$\mu = E[X] = \sum_{i=1}^{6} x_i p_i = \frac{(1+\dots+6)}{6} = \frac{7}{2}$$

Therefore, the variance of X is:

$$\sigma_X^2 = \sum_{i=1}^{6} \frac{1}{6}(i-\frac{7}{2})^2 = \frac{35}{12}\ ;$$

The standard deviation of X is: $\sigma_X = \sqrt{\frac{35}{12}}$.

In statistic, **covariance** is a measure of the joint variability of the two random variable.

Definition 7.0.3. *The covariance between two jointly distributed real valued random variable X and Y is defined as*

$$cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

noted that $E[X], E[Y])$ is the expected value of X, Y, respectively. Noted as $\sigma_{XY}, \sigma(X,Y)$. Expressed by:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] + E[X]Y + E[X]E[Y]]$$

$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]$$

The Cauchy-Schwarz Inequality in Probability

Proposition 7.0.1. *The Cauchy-Schwarz Inequality states that for all vectors u and v of a inner product space, it is true that:*

$$| < u, v > |^2 \leq < u, u > . < v, v >$$

where $< ., . >$ is the inner product: $< u, v > \leq \|u\|\|v\|$

Proof. As, $< u, v >= \|u\|\|v\|.cos(\delta)$
$| < u, v > | = |\|u\|\|v\|.cos(\delta)| = \|u\|\|v\|.|cos(\delta)| \leq \|u\|\|v\|.1$
Hence, $< u, v > \leq \| u\|\|v\|$.

In Probability: Let X, Y be random variables, then the covariance inequality is given by

$$var(Y) \geq \frac{cov(Y,X)cov(Y,X)}{var(X)}$$

After defining an inner product on the set of random variables using the expectation of their product, $< X, Y >:= E[XY]$, then the Cauchy-Schwarz Inequality becomes $|E[XY]|^2 \leq E[X^2]E[Y^2]$.
To prove the covariance inequality using the Cauchy-Schwarz Inequality;
Let $\mu = E[X]$ and $\nu = E[Y]$, then

$$\begin{aligned}
|cov(X,Y)|^2 &= |E[(X-\mu)(Y-\nu)]|^2 \\
&= | < X-\mu, Y-\nu > |^2 \leq < X-\mu, X-\mu >< Y-\nu, Y-\nu > \\
&= E[(X-\mu)]^2 E[(Y-\nu)]^2 \\
&= var(x)var(y)
\end{aligned}$$

Pearson Correlation Coefficient

In statistic, the Pearson Correlation Coefficient or bivariate correlation is a measure of the linear correlation between two variables X and Y. According to the Cauchy-Schwarz inequality it has a value between $+1$ and -1

$$+1: \texttt{positive linear corr.}$$

$$0: \qquad \texttt{no linear corr.}$$

$$-1: \texttt{negative linear corr.}$$

Definition 7.0.4. *The* ***Pearson Correlation Coefficient*** $(\rho_{X,Y})$ *between two random variables* X *and* Y *with the expected values* μ_X, μ_Y *and the standard deviations* σ_X, σ_Y *is defined as:*

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{E[X^2] - E[X]^2}\sqrt{E[Y^2] - E[Y]^2}}$$

Lemma 7.0.1. *The correlation cannot exceed* 1 *in absolute value.*

Proof. This proof is a result of Cauchy-Schwarz Inequality:

$$\begin{aligned} \rho_{X,Y} &= \frac{cov(X,Y)}{\sigma_X \sigma_Y} \\ &\leq \frac{\sqrt{var(x)var(y)}}{\sigma_X \sigma_Y} = \frac{\sqrt{\sigma_X^2 \sigma_Y^2}}{\sigma_X \sigma_Y} \\ &= \frac{|\sigma_X \sigma_Y|}{\sigma_X \sigma_Y} \leq |1| \end{aligned}$$

Hence, $-1 \leq corr(X,Y) \leq 1$.

BIOGRAPHICAL SKETCH

Halime Beyza KÜÇÜKDAĞ

Konya, Turkey, 1994

Education

2012-2017	Bachelor of Science in MATHEMATICS **Galatasaray University**, Istanbul, Turkey
AUTUMN 2015	Exchange Semester **Bourgogne University**, Dijon, France
2008-2012	High School Diploma in SCIENCE DEPARTMENT **Osman Nuri Hekimoğlu High School**, Konya, Turkey