UCL
Université catholique de Louvain

École polytechnique de Louvain (EPL)

Analysis of the Financial Times ranking *master in management* with machine learning

Mémoire présenté par
Arthur JANSEN

en vue de l'obtention du grade de master en
Ingénieur civil informatique

Promoteurs
John LEE , Frédéric VRINS

Lecteurs
Cyril DE BODT, Marco SAERENS

Année académique 2016-2017

With these few lines, I would like to thank anyone who helped me through the writing of my master thesis. Especially, I would thank my thesis advisors Mr John Lee and Mr Frédéric Vrins who constantly allowed this thesis to be my own work, but steered me in the right direction whenever they thought I needed it. They were always open to answer my questions when I had some trouble or provide me with technical resources to help me in my work.
I would also like to thank Mr Cyril De Bodt and Mr Marco Saerens for the time they will spend reading this thesis.

Abstract University rankings play nowadays a major role in the decision of many students with regards to their future schools. Nonetheless, these rankings often remain quite opaque: not all data are made available, the methodology behind the rankings is not well defined, etc. One of the main ranking centred on business schools is the *Master in Management* from the Financial Times. This work aims to study the relevance of this ranking and its possible flaws. Several techniques are conducted, as a robustness analysis to assess the sturdiness of the ranking when facing uncertainties or dimensionality reductions to visualize the data set in two or three dimensions and thus envision local neighbourhoods of schools. Finally, from these analyses, potential improvements to the ranking from the Financial Times are discussed, from the design of a new ranking from scratch to the addition of visualization tools to enhance the informative character of the ranking.

Contents

List of acronyms

BAP	Budget Allocation Process
BLVA	Bayesian Latent Variable Analysis
DR	Dimensionality Reduction
FT	Financial Times
FT MiM	Financial Times *Masters in Management*
GTM	Generative Topographic Mapping
JSE	Jensen-Shannon Embedding
LLE	Locally Linear Embedding
LVA	Latent Variable Analysis
Ms. JSE	Multi-scale Jensen-Shannon Embedding
NLDR	Non-linear dimensionality reduction
PCA	Principal Component Analysis
SBDR	Similarity-Based Dimensionality Reduction
SNE	Stochastic Neighbor Embedding
t-SNE	t-Distributed Stochastic Neighbor Embedding
VIF	Variance-Inflation Factor

Notations

The FT MiM ranking attempts to order 70 business schools based on 16 criteria. For conciseness, the schools and criteria are often named according to their FT rank in plots. The complete lists can be found in Tables A.1 and A.2. For clarity, the full name of schools and criteria is used in the text. To be able to match the words with the numbers, the following notation is adopted for schools: the University of St Gallen (school 1), and for criteria: the criterion *Weighted criteria* (1). Both examples represent the first school and the first criterion.

In dimensionality reduction, high- and low-dimensional spaces are handled. Following the idea proposed in [30], Greek letters are used for variables or operations in the high-dimensional space, while Latin letters are used in the low-dimensional space. A non-exhaustive list of examples of this rule can be:

- ξ_{ij} represents the value of the criterion j of the school i in high dimension. The equivalent is x_{ij} in low dimension.
- $\boldsymbol{\xi}_i$ represents the vector of the sixteen values of the school i in high dimension. The equivalent is $\boldsymbol{x}_i$ in low dimension. More generally, a vector will always be denoted with bold letters in this report.
- $\delta(i, j)$ or δ_{ij} represents the distance between two schools i and j in high dimension. The equivalent is $d(i, j)$ or d_{ij} in low dimension.
- σ_{ij} is the similarity (or probability) that the point i chooses the point j as its neighbour in the high-dimensional space. The equivalent is s_{ij} is in the low-dimensional space.

List of Figures

List of Tables

1 Introduction

1.1 Motivation

For some years now, different organizations have presented university rankings. They are either general or centred on a specific field of study. Concerning general rankings, one can cite the *QS World Universities Rankings* (originally in collaboration with the Times Higher Education magazine) [37], the *Academic Ranking of World Universities* (compiled by the Shanghai Jiao Tong University) [1]. Concerning more specific rankings, one can note the *U.S. News & World Report* (centred on U.S. universities) [10], the *Times Good University Guide* (centred on universities in the U.K.) [11] or the multiple *Financial Times rankings* [13] (focused on business and management schools). These rankings are eagerly awaited both by universities and by prospective students because it plays nowadays a major part in the decision of many students with regards to their future schools. They play a key role in defining the identities of schools and universities [45]. Consequently, there have been great concerns raised by universities that these rankings do a poor job of defining what a good school should be [46]. From this an effort has been made by official institutions to characterize the quality of a university and thus to enhance the transparency behind these rankings. For instance, the European Union developed a tool, called *U-Multirank*, to compare universities according to students' own view of a good education [12].

The Financial Times (FT) publishes several rankings of pre- and post-experience programme in the areas of business and management. Among them is the *Masters in management* ranking (FT MiM) which ranks schools delivering a master degree in management [14]. This ranking is one of the most prominent in the field of pre-experience programme in management. Its methodology is to apply a set of weights on sixteen criteria to all schools and derive from it a unique score that is used to order the schools. These criteria and related weights[1] are decided by the Financial Times and are thus subject to controversy [7]. Nevertheless, as the FT MiM ranking is perceived to have a legitimacy by a large part of the higher-education community, it could be interesting for schools and for students to understand how the rankings systems work. The former could then learn the potential effects of their decisions on the ranking while the latter could interpret results of the ranking more knowingly. The Financial Times publishes most of the data it used to construct its ranking [14], as well as some information on its methodology [34], which could ease the analyzing of the ranking.

The FT MiM ranking is obtained by sorting business schools according to aggregated scores, this aggregation being based on weights. Therefore, the FT MiM ranking can be seen as a projection of the 16-dimensional data to a one-dimensional set of natural numbers. It could then be tempting to compare the FT MiM ranking with more developed data projection methods, unfolding the same 16-dimensional data set to a one- or two-dimensional map that is easier to analyze. That way, it is possible to deliver a more in depth analysis of the ranking to universities and prospective students.

[1] These criteria and weights can be found in Table A.2.

The aim of this work is then to demonstrate clearly how the FT MiM ranking works. Different techniques are used, including a robustness analysis to assess the stability of the FT MiM ranking in the presence of uncertainties, or basic as well as more evolved data projections to discover if similarities between business schools in criteria are well translated into the FT MiM ranking. Conclusions of theses analyses are used to present alternatives to the FT MiM ranking that aim to solve some of the above-mentioned problems.

1.2 State of the art

As public interest in university rankings increases, the number of studies analyzing these rankings has been growing constantly. In this section, a digest of what currently exists on the subject of analyzing university rankings is exposed. On the one hand, there have been research papers focusing on discussing in a general way about the teaching quality defined behind these rankings [46], with [7] that centred its talks on the FT rankings. Other works focused on differences in the measurement of this quality between countries [8] or on how the schools have to respond to these rankings [45, 21]. These articles are of less importance in this work as a stress is put on analyzing a specific ranking, the FT MiM one.

On the other hand, some research papers concentrated on analyzing one or two particular rankings. A robustness analysis of the rankings of the Shanghai Jiao Tong University and the Times Higher Education magazine, based on a multi-modelling approach, has shown that statistical inferences on the rankings are unsound [40]. It also has been said that the FT rankings are somewhat vulnerable to volatility [7] even though a proper robustness analysis has not been conducted in this study.
Only basic data projection techniques have been used to assess the importance of the different criteria in university rankings. Almost all the research papers found on the subject ran a linear dimensionality reduction method called Principal Component Analysis (PCA). In short, a PCA aims to find new uncorrelated criteria based on linear combinations of the FT criteria. Then, only the new criteria with a high variance are kept. It allows to assess the proportion of each original criterion, giving some insight into which FT criterion plays a key role in discriminating the schools. For instance, a PCA has been conducted on the rankings from the U.S. News & World Report and this showed that the actual contributions of the criteria differ substantially from the official weighting scheme [44].
An extensive comparison between several rankings, including the ones from the Times Higher Education magazine and the Shanghai Jiao Tong University has been performed using similarity methods to show that there are reasonable similarities between these rankings even though they use different methodology and criteria [2].

From this, some research papers decided to define modified methodologies for these rankings that take into account the weaknesses noticed in their analyses. The article that conducted a robustness analysis of the rankings from the Shanghai Jiao Tong University and the Times Higher Education magazine also proposed an alternative ranking which is more dependant on the set of criteria than on the methodological choices, like the weights [40]. Concerning the FT rankings, an experiment at 'fairer' ranking has been done by allowing the official weights to be varied somewhat in each school's favour to construct what is called "divisions" rather than ranks [24].
Also, attempts at redefining the teaching quality described behind rankings have been done. For instance, a ranking more based on research, educational and environmental indicators has been designed, using the data set of the Times Higher Education magazine [31].
Finally, some works undertook the construction of new rankings by using statistical procedures

instead of arbitrary weights on criteria, to be able to assess the uncertainty related to each rank in the ranking. A study did a comprehensive analysis of this kind of methods to rank institutions in the areas of health and education [16]. Another research analyses statistical procedures using the data sets from the U.S. News & World Report and the America's Best Colleges [17].

Besides that, it is worth noting that official institutions began delivering guidelines to construct an efficient and robust ranking. For instance, the OECD[2] developed an extended handbook to build rankings (called composite indicators in there) [33]. This book is more focused on ranking countries than schools but can be quite easily translated to university rankings. As several articles mentioned it, uncertainty in rankings can sometimes be large and the European Union conducted some works to develop tools to assess it [41].

From all this, it can be seen that little research has been conducted on using evolved data projection methods to analyze rankings. In particular, it should be noted that no article has analyzed rankings using non-linear dimensionality reduction (NLDR) techniques. This is quite unfortunate as NLDR methods aim at offering a much faithful view of the data set in low dimensions if this data set can not be reduced to a linear model.
Nonetheless, literature about evolved data projection techniques actually exist. Some books describe the state of the art of dimensionality reduction methods [43, 27] while other ones focus more on non-linear methods [30].
In particular, concerning linear techniques (explained in Chapter 3), the main ones are the Principal Components Analysis (in Section 3.2), already used to describe rankings as seen above, and the Feature Selection. The former has already been explained before. The latter is based on selecting a subset of existing criteria without creating new ones. These two methods are also respectively explained in [19, 9] and in [18].
Concerning non-linear procedures (explained in Chapter 4), different kinds of methods can be defined depending on the type of preservation. Some techniques are based on preserving distances between points. An example of it is the Curvilinear Component Analysis [6]. Another class of methods focus on preserving the topology of the data sets, like the Self-Organizing Map (in Section 3.3) [23, 19, 38]. The third group is based on preserving the similarity between points. The Stochastic Neighbour Embedding (SNE) [20] or the t-Student Stochastic Neighbour Embedding (t-SNE) [32] are examples of it (in Section 3.4).

1.3 Presentation of the FT MiM ranking

	Criterion	Weights (%)	Units
1	Weighted salary	20	US$
2	Value for money	5	rank
3	Careers	10	rank
4	Aims achieved	5	rank
5	Placement success	5	rank
6	Employed at three months	5	%
7	Women faculty	5	%
8	Women students	5	%
9	Women board	1	%
10	International faculty	5	%
11	International students	5	%
12	International board	2	%
13	International mobility	10	rank

[2] The Organization for Economic Co-operation and Development (OECD).

	Criterion	Weights (%)	Units
14	International course experience	10	rank
15	Languages	1	
16	Faculty with doctorates	6	%

Table 1.1: List of criteria of the FT MiM ranking.

Since 2005, the Financial Times has published a yearly ranking, ordering universities and business schools delivering a master degree in management. The last edition of this ranking dates from 2016 [15]. In this work, the data coming from the 2014 FT MiM ranking is used [14]. These two rankings use the same methodology even though the 2016 ranking orders 93 schools while 70 were officially ranked in 2014.

The FT MiM ranking is officially international but a large majority of the business schools inside it are located in Europe. Only five universities are not European, thus representing only a little more than 7% of all the schools. Two of them are settled in India, two others in China and the last one in Canada. Figure 1.1 depicts the location of the European schools[3]. Each of them is coloured by their FT MiM rank. It can be seen that a lot of schools are established in the west part of Europe, especially in France, Germany, Belgium and the UK. Figure 1.2 presents a zoom on this region to be able to more easily distinguish all the schools. The main remark that could be made on this plots is that the geographic location does not seem to be a major criterion to well perform in the FT MiM ranking, even though there is a slight concentration of well-performing schools in cities like Paris and London.

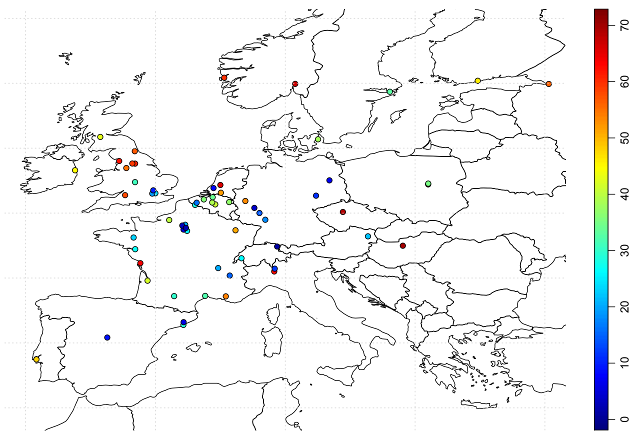


Figure 1.1: Map of the European schools present in the ranking, coloured by their FT MiM ranks.

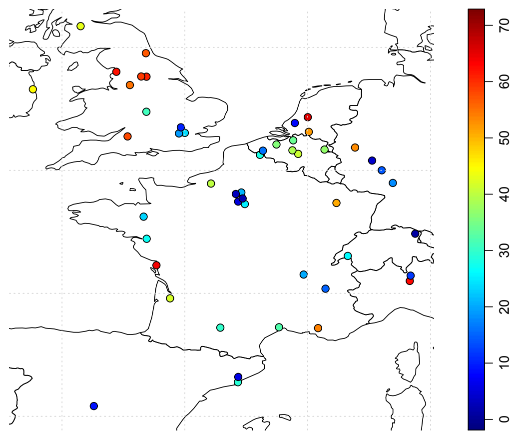


Figure 1.2: Zoom of the map of the European schools, coloured by their FT MiM ranks.

More particularly, the FT MiM ranking is based on sixteen criteria. There are listed in Table 1.1 and are explained below. Weights for ranking criteria are shown in parentheses as percentage. This information was found in [34].

*1. *Weighted salary* (20) is the average salary (in US dollars) three years after graduation, adjusted according to variations between industry sectors.

2. *Value for money* (5) is a ranking calculated according to alumni salaries today and other unidentified costs.

[3]Some schools have several campuses, located in different countries (and event continents). For instance, the Skema Business School (28) has campuses in France, US and China. When this was the case, the location of their main buildings was used to decide their position.

*3. *Careers* (10) is a ranking estimated according to alumni seniority and the size of their company in terms of the number of employees worldwide.

*4. *Aims achieved* (5) is a ranking measuring the extent to which alumni fulfilled their goals for doing a masters.

*5. *Placement success* (5) is a ranking describing the effectiveness of the careers service in supporting student recruitement, as rated by alumni.

6. *Employed at three months* (5) is a percentage of the most recent classes that found employment within three months of completing their course and for which the school was able to provide data.

7. *Female faculty* (5) is a percentage of the faculty that is female.

8. *Female students* (5) is a percentage of female students on the masters programme.

9. *Women board* (1) is a percentage of women on the school advisory board.

10. *International faculty* (5) is a percentage based on the mix of nationalities and the percentage of faculty members whose citizenship differs from their country of employment.

11. *International students* (5) is a percentage based on the mix of nationalities and the percentage of masters students whose citizenship differs from their country of study.

12. *International board* (2) is a percentage of the board whose citizenship differs from the school's home country.

*13. *International mobility* (10) is a ranking estimated to changes in the country of employment of alumni from graduation to today.

14. *International course experience* (10) is a ranking calculated according to whether the most recent graduating class undertook exchanges, company internships or study trips in countries other than where school is based.

15. *Languages* (1) is the number of extra languages required on graduation.

16. *Faculty with doctorates* (6) is a percentage of faculty with doctoral degrees.

Several remarks on this description of FT criteria could be made. From [34], it can be noticed that some criteria are based on data from the current year and the one or two preceding years where available. These criteria are marked with an asterik in the above list. The proportion of more than one-year-old data used for each concerned criterion remains unclear. It also can seem quite strange that these criteria span a different time periods than other ones, like the *Employed at three monts* (6) and the *International course experience* (14) which are only based on the most recent graduating class. Also, criteria based on more than one-year-old data could slow down the repercussions of new measures taken by deans and university administrators.

Furthermore, from the above explanation of FT criteria, it can be seen that some of them are based on alumni surveys while others are established on the basis of university responses. Actually, in terms of FT weights, 55% of the weights are based on alumni-related criteria[4]. The fact that this represents more than the majority of weights tells us that alumni surveys play a key role in the performance of universities in the FT MiM ranking. Some universities engaged in aggressive marketing to their alumni members to make sure they fulfill a good survey [7].

[4]There are criteria 1, 2, 3, 4, 5 and 13.

1.3.1 Quality defined by the FT MiM ranking

The criterion *weighted salary* (20) has the highest weight. If the criterion "value for money" is added to it, a quarter of the weights goes to how much money a former student will earn. The possibility of having a good salary afterwards is thus a primordial component of teaching quality according to the Financial Times.
A little less than a quarter of the weights (23%) is assigned to criteria[5] related to the composition of the faculty, the student pool and the board. Hence, the Financial Times seems to consider very important the diversity inside the school and a good guarantee of teaching quality. It should be noticed however that there is no attempt to measure the quality of the faculty members and the board, as well as there is no attempt to measure whether the board has any meaningful governance role [7].
A third category could gather career-related criteria, namely the *careers* (10), the *aims achieved* (5), the *placement success* (5) and the *employed at three months* (5). This category is worth a quarter of the FT weights and shows that one of the top priority of teaching quality according to the Financial Times is how a master degree will help former students to progress in their career.
Another group of criteria that could be defined contains the criteria associated to globalization. It is composed of the criteria *International mobility* (10), *International course experience* (10) and *Languages* (1). Hence, 21% of the FT weights are linked to the idea that having the possibility of travelling and working abroad is part of teaching excellence.
The only criterion left that did not fall within a previous category is the *Faculty with doctorates* (6). It has an average weight and tells us that the Financial Times find important that faculty members come from the research community.
Table 1.2 summarizes the weights assigned to each group of criteria.

	Category	Weight (%)
1	Future salary	25
2	School diversity	23
3	Career opportunity	25
4	Globalization	21
5	Faculty with doctorates	6

Table 1.2: Summary of the quality defined by the FT MiM ranking.

Now that a definition of teaching quality viewed by the Financial Times has been sketched, it should be noted that possible criticisms to this definition of teaching quality fall outside the scope of this work and the data set provided by the FT is used as is. The only remark one could do is to note that among the numerous alumni-related criteria, no one measures directly the alumni opinions about the teaching quality of their own school.

1.4 Goals and contribution

Now that the current situation and common criticisms of university rankings analyzes have been presented, it is time to formally present what is the goal of this work. This study aims to clearly demonstrate the relevance of the FT MiM ranking, its possible flaws and to finally discuss potential improvements.
To do so, the first part is devoted to the analysis of the data set and the reconstruction of the FT MiM ranking to show that the provided data is sufficient for further analyses. A robustness assessment is also conducted to reveal possible weaknesses in the methodology used.

[5] There are gender- and foreigner-related criteria.

The second part of this work is dedicated to data projections. Basic ones, like PCA, are tried so that results are compared with results of other research papers on university rankings. However, more evolved data projections, like NLDR methods, are also conducted so that relations between schools are more easily assessed.
Then, the last part of this study aims to gather all the flaws discovered and propose possible improvements to the FT MiM rankings. The improvements could be to modify somewhat the FT methodology, work on better visualizations of results.

1.5 Organization of this work

This work is divided into distinct chapters. In the present introduction, the motivation of this study is presented as well as the state of the art of the analysis of university rankings and a basic presentation of the FT MiM ranking.
In Chapter 2, the data set provided by the Financial Times is studied. First, a more statistical analysis of the data is conducted to gauge the differences between the criteria. Then, a reconstruction of the FT MiM ranking based on the provided FT data is performed and then compared to the ranking given by the Financial Times. The final section of this chapter examines the robustness of the ranking, when exposed to different types of uncertainties.
In Chapter 3, the theoretical frameworks of dimensionality reduction (DR) methods are presented. The DR methods studied in this work are the Principal Component Analysis (PCA), the Self-Organizing Map (SOM) and the Multi-scale Jensen-Shannon Embedding (Ms. JSE).
In Chapter 4, the projections and the results of DR methods presented in the last chapter are analyzed and compared.
In Chapter 5, possible improvements to the FT MiM ranking are displayed according to the results of the previous chapters.
In Chapter 6, a summary and a conclusion of this work are conducted.

2 Analysis of the data

Abstract This chapter is devoted to analyze the data set of the Financial Times, as it is. The ranking was presented in the last chapter. Here, a general presentation of the data will be provided by way of introduction, followed by some statistical characteristics of it. Then, a reconstruction of the 2014 FT MiM ranking based on the data set will be performed. The goal is to show that the provided data set is relevant enough to conduct further analyses on it. The last part of this chapter is centered on realizing a robustness analysis. The aim is to test the variability of the ranking when some perturbations are introduced, so that one can be sure that the ranking is not a "stroke of luck" but is rather based on solid grounds.

2.1 Presentation of data

The 2014 FT MiM ranking has been presented in the previous chapter. As a reminder, it consists of 70 business schools delivering a master degree in management ordered according to 16 criteria. The list of schools can be found in annex (Table A.1). The list of criteria was already shown in the previous chapter and it can also be found in annex (Table A.2).

The Financial Times assures these sixteen criteria are enough and self-sufficient to provide an accurate picture of what is a good business school, but do they ? Before analyzing the ranking in itself by attempting to perform a reconstruction of it, basic statistical analyses of the data are done.
Table A.6 displays some characteristics of the criteria, like the maximum and minimum values, the mean or the standard deviation.

2.1.1 Outliers

Figure 2.1 displays boxplots of z-scores for each criterion. Z-scores, or standard scores, are normalized values with zero mean and a standard deviation of one. Z-scores are explained in more details in the next section when the reconstruction of the ranking is tackled. The central box of a boxplot represents the interquatile range (IQR), meaning that this box spans from the first to the third quartile. The upper and lower ends of the queues respectively show the maximum and the minimum of the data, without taking into account the outliers. A point is defined as an outlier in this work if it falls outside 1.5 times the IQR above the upper quartile (called an upper outlier) or below the lower quartile (named a lower outlier).
As can be seen in Figure 2.1, four criteria have outliers. The criteria *Women board* (9) and *International faculty* (10) have each one one upper outlier and the criterion *Weighted salary* (1) has two upper outliers. The last criterion to have outliers is *Faculty with doctorates* (16) which has three lower ones.

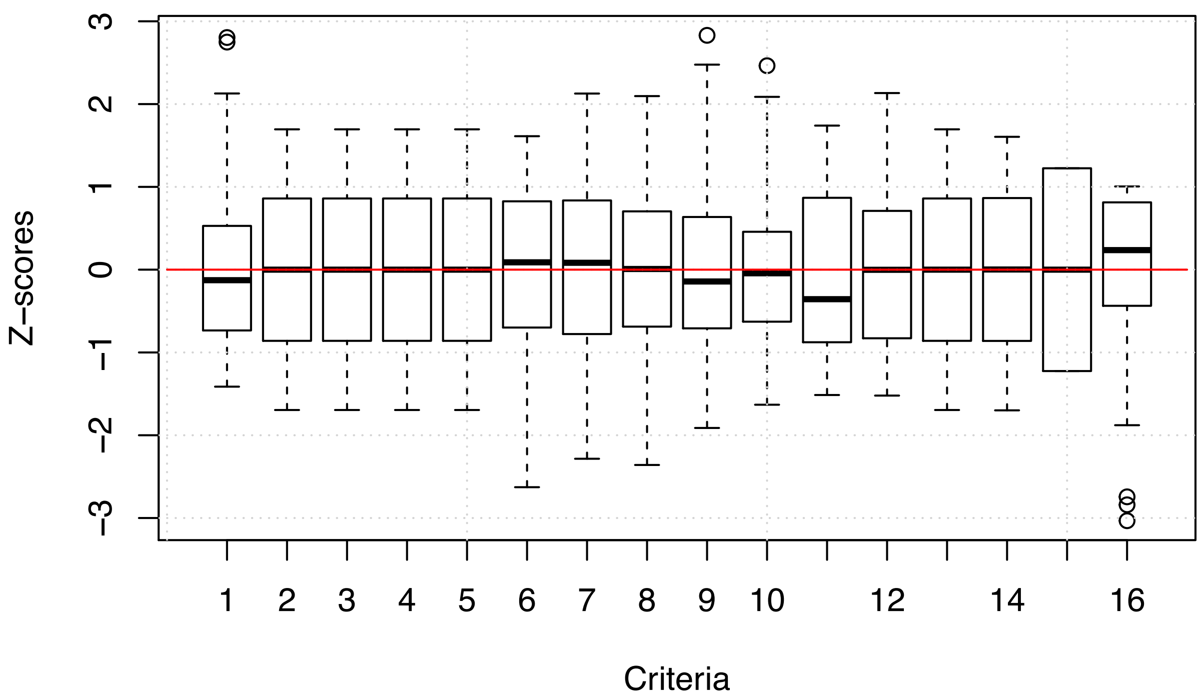


Figure 2.1: Boxplots for the sixteen criteria.

The two upper outliers for the criterion *Weighted salary* (1) are particularly interesting because of the high weight of this criterion. Indeed, the two schools that display these two outliers could benefit a lot of having a quite high value for this criterion, even though their scores for other criteria are average. The two schools are the Indian Institute of Management, Ahmedabad (17) and the WHU Beisheim (4). The second one belongs to the top schools of the ranking, while the first one stands a little lower but is still far above average. In any case, these two schools perform quite well in the ranking and a reason of this is probably their high score for *Weighted salary* (1).

The school that has the highest score for the criterion *International faculty* (10) is the Cems (5). This school is in the top-performing ones as well and this criterion most likely helps it. However, it should be noticed that the *International faculty* (10) presents a FT weight of 5%, four times less than *Weighted salary* (1).

Three schools present a lower outlier for the criterion *Faculty with doctorates* (16): the Nyenrode Business Universiteit (66), the Politecnico di Milano School of Management (63) and the Eada Business School Barcelona (29) (the latter displaying the lowest score). Apart from the Eada Business School Barcelona (29) which is above average in the ranking, it can be seen that the two other schools are quite low in the ranking. This criterion has a FT weight of 6% and it could be the reason of the low performance of these schools.

For the sake of completeness, the school that presents an upper outlier for the criterion *Women board* (9) is the Warsaw School of Economics (68). However, this measure is not relevant as the Financial Times reveals that schools with a 50:50 (male/female) composition receive the highest score, as it will be explained in the next section when the reconstruction of the ranking is tackled.

Overall, the FT data set does not present strange behaviours regarding the outliers. More importantly, it can be seen that only the WHU Beisheim (school 4) belongs to the top schools and displays an upper outlier for the criterion *Weighted salary* (1). The situation where the top schools are so well performing because of an upper outlier for the highest-weighted criterion does not appear in this case.

2.1.2 Correlation

An important aspect when analyzing data of a ranking is the correlation of it. Two kinds of correlation are distinguished here. The first one is between each criterion and the FT MiM ranking while the second one is between every pair of criteria. The correlation between criteria and the FT MiM ranking is interesting to analyze because it demonstrates the importance of each criterion in the ranking in accordance with their respective FT weights. Likewise, it is worth assessing the correlation between every pair of criteria. Indeed, if two criteria are highly correlated, one could think of the necessity to keep the sixteen criteria and why not throwing out one of them.

	Criteria	Correlation
1	Weighted salary	0.77
2	Value for money	(*) 0.18
3	Careers	0.52
4	Aims achieved	0.61
5	Placement success	0.59
6	Employed at three months	0.35
7	Women faculty	-0.30
8	Women students	(*) 0.10
9	Women board	(*) -0.12
10	International faculty	0.41
11	International students	(*) 0.16
12	International board	0.37
13	International mobility	0.59
14	International course experience	0.53
15	Languages	0.26
16	Faculty with doctorates	0.43

Table 2.1: Pearson correlations of variables compared to the FT ranking. Correlations with insignificant p-value (> 0.05) are marked with an asterisk symbol (*).

Table 2.1 shows correlations of each criterion with the FT ranking. The criterion *Weighted salary* (1) is the one with the highest correlation (0.77). This is partly due to the fact that this criterion has the maximal FT weight (20%). Other criteria that have a high correlation are *Aims achieved* (4) (0.61), *Placement success* (5) (0.59) and *International mobility* (13) (0.59). The first two criteria have a FT weight of 5% while the last one has a weight of 10%.
It should also be noticed that four criteria (*Value for money* (2), *Women students* (8), *Women board* (9) and *International students* (11)) have a too high p-value to have a significant correlation.
From criteria that have a significant correlation, only *Women faculty* (7) has a negative correlation (-0.30). It means that schools performing well in the FT MiM ranking tend to have a low gender diversity inside their faculty.

Table 2.2 displays Pearson correlations between each pair of criteria. It can be seen that in general, the pair of criteria are not highly correlated. The mean correlation is 0.09[1], which is close to zero.
The highest pairwise correlation is composed of the criteria *International course experience* (14) and *Languages* (15) with a Pearson correlation of 0.72. These two criteria have a FT weight of respectively 10% and 1%. If one thinks about the meaning behind these criteria, it can be

[1]Only taking into account pairs of **different** criteria.

realized that the high correlation between these two is not all that surprising, considering that students with a lot of international experiences tend to be able to communicate in multiple languages.
It could also be interesting to concentrate on the criterion *Weighted salary* (1) as it has the maximal FT weight. It can be noted that it is more correlated than average with three other criteria: *Aims achieved* (4), *Placement success* (5) and *Women faculty* (7). The first two ones have a positive correlation with the *Weighted salary* (1) of respectively 0.69 and 0.68. This can be explained by the fact a high salary is often an achievement and a good placement success for numerous of students. The third criterion has a negative correlation of -0.61. It means that a student can expect a higher salary in a business school with a low gender diversity among the teachers. This is quite surprising as the relationship between *Weighted salary* (1) and *Women faculty* (7) is not natural at all.

In general, correlations between each pair of criteria are looking good. Works conducted on other rankings display correlations between criteria that are relatively higher. This is the case for the rankings from the Times Higher Education magazine [40] or the U.S. News & World Report [44]. However, it will have to be kept in mind in further analyzes, however, that the criterion *Weigthed salary* (1) has correlations with *Aims achieved* (4) and *Placement success* (5) that are more than average because this criterion has the highest FT weight. Also, the negative correlation between *Weighted salary* (1) and *Women faculty* (7) has to be reminded as it could play a role in the constructing of a ranking.

Nonetheless, correlations only analyze relationships between a criterion and the ranking or between two criteria. It could also be interesting to look for connections between several variables at the same time. A possibility is to assess the multi-collinearity of criteria as proposed in [33].

2.1.3 Multicollinearity

A possible solution to overcome the shortcomings of correlation is to assess the multi-collinearity of each criterion. The multi-collinearity is a phenomenon between several criteria where one of them could be predicted with a high degree of accuracy using a linear combination of the other criteria [3].
The measurement used here to estimate the multi-collinearity is the Variance-Inflation Factor (VIF), as proposed in [33]. The VIF is the reciprocal of tolerance. The VIF for the criterion i can be calculated as follows:

$$VIF_i = \frac{1}{1 - R_i^2} \ , \tag{2.1}$$

where R_i^2 is the coefficient of determination for the criterion i [35]. A high value of VIF for a particular criterion means that this one presents a serious multi-collinearity with other criteria.

The general rule of thumb is that VIFs exceeding four warrant further investigation, while VIFs exceeding ten are signs of serious multi-collinearity requiring immediate correction [25]. Having a VIF greater than four can also be seen as a cut-off criterion for suggesting that there is a multi-collinearity problem [33]. Results of VIF are presented in Table 2.3. It can be noticed that two criteria have a VIF higher than four: *Weighted salary* (1) and *International course experience* (14), with a VIF of respectively 4.52 and 4.14. The two criteria that have the lowest VIF are *Employed at three months* (6) and *Women board* (9) (respectively 1.35 and 1.31). The mean VIF is 2.41.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00															
2	0.37	1.00														
3	0.40	-0.18	1.00													
4	0.69	0.33	0.19	1.00												
5	0.68	0.08	0.33	0.54	1.00											
6	0.30	0.18	0.16	0.14	0.32	1.00										
7	-0.61	-0.36	-0.02	-0.54	-0.41	-0.24	1.00									
8	-0.23	-0.14	-0.14	-0.09	-0.27	-0.17	0.20	1.00								
9	-0.26	-0.11	-0.10	-0.14	-0.28	-0.16	0.13	0.22	1.00							
10	0.05	-0.17	-0.03	0.23	0.05	-0.01	-0.06	0.40	0.13	1.00						
11	-0.11	-0.18	-0.10	0.22	-0.09	-0.15	-0.02	0.27	0.26	0.60	1.00					
12	0.03	0.12	0.18	0.14	0.09	0.21	0.08	0.13	0.03	0.29	0.25	1.00				
13	0.19	-0.08	0.20	0.30	0.15	-0.09	-0.05	0.33	0.04	0.61	0.59	0.39	1.00			
14	0.20	0.08	0.22	0.10	0.14	0.17	0.05	0.28	-0.06	0.17	-0.20	0.37	0.37	1.00		
15	-0.01	0.06	0.06	-0.10	0.10	0.11	0.20	0.15	-0.08	0.06	-0.30	0.22	0.16	0.72	1.00	
16	0.39	0.18	-0.04	0.28	0.32	0.26	-0.34	-0.03	-0.14	0.27	0.04	0.01	0.17	0.06	0.11	1.00

Table 2.2: Pearson correlations for every pair of criteria.

	Criteria	VIF
1	Weighted salary	4.52
2	Value for money rank	1.86
3	Careers rank	1.79
4	Aims achieved rank	2.57
5	Placement success rank	2.47
6	Employed at three months	1.27
7	Women faculty	2.04
8	Women students	1.62
9	Women board	1.24
10	International faculty	2.32
11	International students	3.38
12	International board	1.65
13	International mobility rank	3.01
14	International course experience rank	4.14
15	Languages	3.23
16	Faculty with doctorates	1.52

Table 2.3: Variance Inflation Factors (VIF) for the 16 criteria of the FT MiM ranking.

Take a closer look to the criteria *Weighted salary* (1) and *International course experience* (14). From Table 2.2, it can be seen that the first one is highly correlated to *Aims achieved* (4), *Placement success* (5) and *Women faculty* (7) while the second one is highly correlated to *Languages* (15). Several combinations of these criteria were tested to be cut off but the best one revealed to be to remove both criteria with the highest VIF, namely *Weighted salary* (1) and *International course experience* (14).
Figure 2.4 displays the new VIFs for the remaining criteria. It can be noted that the VIFs are generally smaller now. The maximum value is 2.97 for the criterion *Internation students* (11) and the mean VIF decreased from 2.41 to 1.84. No criterion exceeds a VIF of 4 anymore.

	Criteria	VIF
2	Value for money rank	1.70
3	Careers rank	1.49
4	Aims achieved rank	2.16
5	Placement success rank	2.09
6	Employed at three months	1.24
7	Women faculty	1.83
8	Women students	1.48
9	Women board	1.23
10	International faculty	2.31
11	International students	2.97
12	International board	1.51
13	International mobility rank	2.69
15	Languages	1.62
16	Faculty with doctorates	1.41

Table 2.4: Variance Inflation Factors (VIF) for the 14 criteria of the FT MiM ranking after removing the criteria *Weighted salary* (1) and *International course experience* (14).

In conclusion, this multi-collinearity analysis demonstrated that the criteria *Weighted salary* (1) and *International course experience* (14) were the ones with the highest VIF. It can be problematic that the criterion with the highest FT weight is also the one with the highest VIF. However, the deletion of these two criteria to obtain acceptable VIFs seems not to be reasonable as they represent together 30% of the FT weights. The definition of teaching quality as seen by the Financial Times would thus be greatly modified. Instead, in the following sections, the sixteen criteria will be kept while keeping in mind that correlations and multi-collinearity could effect some of the results.

2.2 Reconstruction

The Financial Times did not only publish the FT MiM ranking, it also made available some data related to the sixteen criteria [14]. Before beginning to analyze the ranking in more details, it could be a good idea to attempt to reconstruct the ranking based on published data. That way, the disparity between each school in the ranking could be observed. Indeed, the present FT MiM ranking gives no information on the distance that separates adjacent schools in the ranking. But more importantly, this reconstruction could legitimate the use of the FT data set for further analyses if it is possible to reconstruct the FT MiM ranking to a high degree of accuracy from this data set.

In this section, the methodology used by the Financial Times to create the FT MiM ranking is first introduced. Then, the best choice of correlation to compare the FT MiM ranking and the reconstructed one is discussed. The third part of this section is devoted to reconstruct the ranking, followed by a discussion of the results.

2.2.1 FT Methodology

The Financial Times does not perform a weighted sum on the criteria *per se*. First, it computes z-scores for each criterion. Z-scores are standardized scores with a mean of 0 and a standard deviation of 1. It can be calculated like this:

$$Z = \frac{X - \mu}{\sigma} \quad ,$$

where μ and σ are respectively the mean and the standard deviation of the criterion. Note that some criteria as *Value for money* (2) or *Aims achieved* (4) are ranked values. For these criteria, z-scores are estimated differently as schools with a low value perform better than the other ones:

$$Z = \frac{\mu - X}{\sigma} \quad .$$

Secondly, the weighted sum is performed on z-scores, giving a unique number ζ for each school, which is used to create the ranking:

$$\zeta = \sum_{i=1}^{16} w_i Z_i \quad .$$

It can be seen that the FT MiM ranking is a relative ranking. The z-scores show how well a school performs compared to the mean for a specific criterion [34]. If a school A had an excellent employment rate three months after graduation but the other schools did too, the school A would only get an average score for the criterion *Employed at three months* (6).

In the next section, an adequate measure of quality for the reconstructed ranking is defined, so that the new ranking can be compared with the FT MiM one.

2.2.2 Choice of correlation

Before trying to reconstruct the FT MiM ranking, one should agree on the best kind of correlation to use to compare a reconstructed ranking and the FT MiM one. Being able to compare them is essential so that it can be assessed which ranking is the best reconstruction. One possible choice is to use Spearman and Kendall correlations. They are the most commonly considered correlations [33] and are used in various studies on university rankings [40, 44], even though no work appears to use both in parallel.

For the correlation definitions below, it will be assumed that X and Y are two rankings and that x_i and y_i are the ranks of the university i.
The Spearman correlation ρ between two rankings can be expressed as:

$$\rho = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)} \ , \tag{2.2}$$

where n is the number of schools and d_i represents the difference between each rank of corresponding values x_i and y_i.

The Kendall correlation τ between two rankings can be written as:

$$\tau = \frac{4P}{n(n-1)} - 1 \ , \tag{2.3}$$

where n is the number of universities and P is the sum, over all the schools, of schools ranked after the given school by both rankings.

These two rankings are used in parallel in the next section where the reconstruction of the FT MiM ranking is tackled.

2.2.3 Reconstruction

Now that the Kendall and Spearman correlations have been defined as a measure of quality for the reconstructed ranking, the latter can actually be computed. The methodology followed by the Financial Times has already been explained in Section 2.2.1. The schools are ranked according to the weighted sum of their z-scores for each criteria. This new ranking will be called the `BA` ranking, as the basic one. The results are shown in Figure 2.2(a). A comparison between the newly reconstructed ranking and the FT MiM ranking is depicted in Figure 2.2(c). Some differences can definitely be found between these two rankings. The Kendall and Spearman correlations are respectively 0.9412 and 0.9915. Additionally, the mean absolute error is 1.71 in rank places and the maximum error in ranking is 12 places.

Some remarks can be done about the `BA` ranking. First of all, there is a large gap between the ten best placed schools and the other ones. This is represented well in Figure 2.2(a). There is a decrease in nearly 30 points between schools IE Business School (school 9) and HHL Leipzig Graduate School of Management (school 11). Secondly, the area composed of the schools from rank 10 to 20 present scores that are very diffuse. As a result, these schools are quite misplaced in the `BA` ranking.
Overall, the relation between the scores and the FT MiM ranking is not continuous. Some schools have a drop in scores in comparison to their direct neighbours. For instance, the school Indian Institute of Management, Calcutta (school 13) obtained a score lower than 20 units compared to its two neighbours.
The difference between the two rankings can not be explained by an incorrect methodology as the one defined by the Financial Times and explained in Section 2.2.1 was followed. It could

be related to the data however. So far, the data provided by the Financial Times has been used without modifying it but there is no guarantee that the Financial Times used it *per se* to construct their ranking.

L. Ortmans[2] (personal communication, February 14, 2017) assured that data given by the Financial Times does not display accurate scores for some criteria. Among other things, several criteria are expressed as a ranking but not as plain values. Besides, criteria for international diversity (*International faculty* (10), *International students* (11) and *International board* (12)) also take into account the diversity by country of origin and not only the percentage of foreigners, which is displayed online. It could be a good solution to attempt to estimate values hidden behind these criteria and create a new ranking based on that.

2.2.3.1 Ranked criteria

There are six ranked criteria: *Values for money* (2), *Careers* (3), *Aims achieved* (4), *Placement success* (5), *International mobility* (13) and *International course experience* (14). The fact that the Financial Times used these ranked criteria or the values behind them are unclear. In this subsection, an investigation is initiated to find these underlying values and reconstruct a new ranking based on them.

The values behind the ranked criteria are not available so they will have to be estimated. A basic assumption would be that they follow a normal distribution and that all criteria are independent from each other. With this in mind, for each ranked criterion, ten thousand sets of 70 points were generated, one point for each school. Each set is ordered and after that, the mean of the best values in each set is computed, then the mean of the second best values in each set, etc. At the end, a set of 70 ordered mean values that follow a normal distribution is obtained and each value is associated with a school according to its rank for that criterion. This is done for all the criteria.

Once all the ranked criteria have been adapted, the same methodology than with the BA ranking can be used. It gives us a new ranking that will be named the RA ranking, as ranked. The results are shown in Figure 2.2(b). A comparison between the RA ranking and the FT MiM ranking is depicted in Figure 2.2(d). The Kendall and Spearman correlations between the two rankings are respectively 0.9528 and 0.9951. Besides, the mean absolute error is 1.29 in rank places and the maximum error in ranking is 7 places. This is quite an improvement compared with the BA ranking. This shows that the Financial Times did not use the data they provide as it is. They rather most likely decided to use some criteria for the FT MiM ranking and to only publish the position of each school for these criteria. The choice of displaying some criteria as a ranking and others as plain values remains unclear, however.

The RA ranking better approximates the FT MiM ranking. Some discrepancies remain though. Again, a large gap between the ten best placed schools and the rest is shown in Figure 2.2(b). Also, it can be seen in Figure 2.2(d) that the schools ranked between 10 and 20 as well as the ones forming the tail of the ranking are generally misplaced.

A possible improvement of the RA ranking could be not to only generate points that follow a normal distribution but to consider a broader range of distributions by optimizing the skewness and the kurtosis while keeping the mean and standard deviation to 0 and 1 respectively. I optimized the skewness and the kurtosis of the distribution in order to maximize the correlation between the reconstructed ranking and the FT MiM ranking. The results were not encouraging.

[2]Laurent Ortmans is a business education statistician at the Financial Times and is the author of the article [34].

Kendall and Spearman correlations were nearly the same than for the RA ranking. Furthermore, the approach to optimize some criteria to maximize the correlation while there are other criteria that are not fully displayed online (like international diversity criteria) could seem a little odd. Indeed, the optimization of ranked criteria could overlap the effects of hidden information of international diversity criteria.

2.2.3.2 International diversity criteria

There are three international diversity criteria: *International faculty* (10), *International students* (11) and *International board* (12). The actual scores of these criteria are much more difficult to estimate than for the ranked criteria. Indeed, the Financial Times only provides in its data set the percentage of foreigners for each school. However, as it was already explained earlier, the Financial Times also takes into account the diversity of foreign countries represented among the students. For instance, a school located in a country A and having 20% of its members coming from a country B will receive a lower score than another school located in the same country and that has 20% of its members who come from several different countries. Hence, it does not seem like a good idea to try to estimate the actual scores of these criteria when there is no clue about the hidden information.

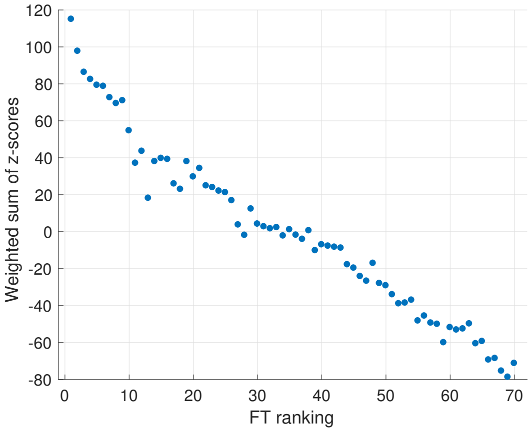
2.2.4 Conclusion

In this section, an attempt to reconstruct the FT MiM ranking from the data given by the Financial Times was conducted. It has been seen that this task was quite difficult as the Financial Times does not provide the data it used but rather publishes summarized data for some criteria. Some assumptions on the values hidden behind some of these criteria have then been made to generate points that are closer to the data used by the Financial Times. A perfect correspondence between the reconstructed ranking and the FT MiM one has not been reached though. The best one that was designed is the RA ranking. It has Kendall and Spearman correlations equal to 0.9528 and 0.9951 respectively.

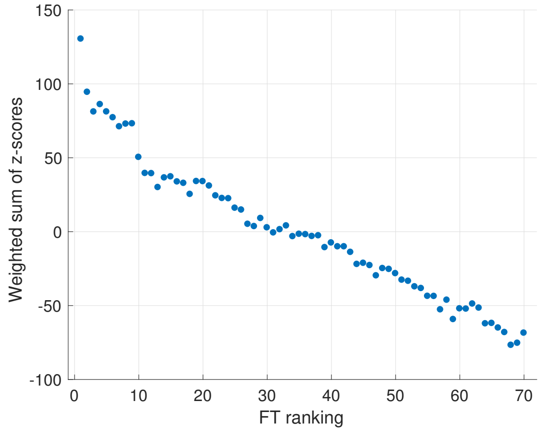
However, not being able to reproduce the exact same ranking than the Financial Times is not a major handicap. An imperfect (but good enough) ranking will still make possible to analyze the FT MiM ranking in more details. Among other things, it has been discovered while doing the reconstruction that there exists a clear gap in scores between the top ten schools and the other ones. In particular, the University of St Gallen (school 1) is well ahead of all the others. This will already allow us to have a better visualization of the position of each school in comparison to others.

Furthermore, related research papers faced the same problem. Another work [36] attempted to reconstruct the FT MiM ranking and reached a Spearman correlation equal to 0.990, which is lower than the correlation of the RA ranking. Besides, it is not specific to the Financial Times rankings. The problem that not all the data are made available is inherent to the majority of the rankings. An article that focused on the US News & World Report tier rankings [44] encountered the same situation.

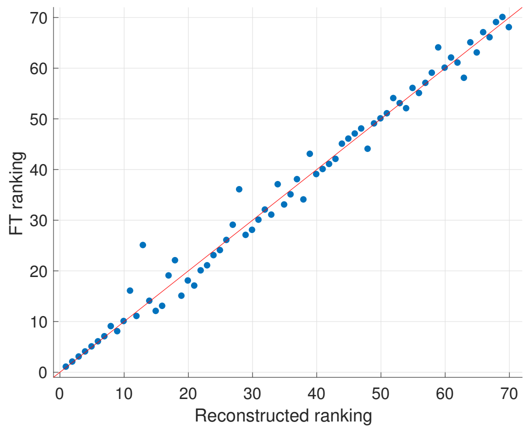
The goal of this section is then reached. It was proved that the FT data set was relevant enough to conduct further analyses as the correlation obtained with the new rankings are quite high. The RA ranking presenting the best correlation compared to the FT MiM ranking, values used for the ranked criteria will be calculated in the same way than in the RA ranking from now on.



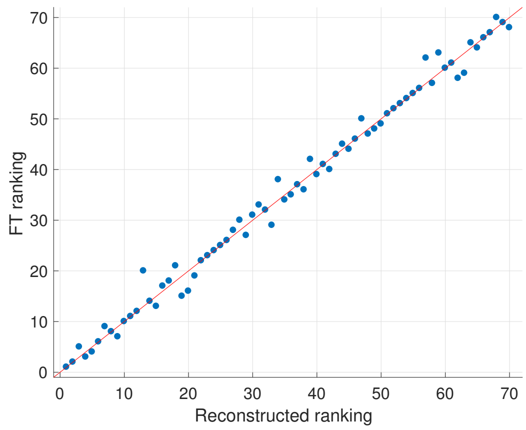
(a) Weighted sum of z-scores of schools in BA reconstruction ordered according to their FT MiM ranking.



(b) Weighted sum of z-scores of schools in RA reconstruction ordered according to their FT MiM ranking.



(c) BA ranking compared to the FT MiM ranking.



(d) RA ranking compared to the FT MiM ranking.

Figure 2.2: (a) and (c) depict the results of the BA reconstruction. (b) and (d) show the results of the RA reconstruction.

2.3 Robustness analysis

Robustness analysis is an indispensable element of a ranking analysis. There is always a part of subjectivity in designing a ranking. For instance, in the FT MiM ranking, the Financial Times selected sixteen predefined criteria to create its ranking even though it provides data for others (like the maximum course fee, the course length, the number of students enrolled, etc.) [14]. It also decided to normalize the data and use a weight-and-sum as aggregation method with arbitrary weights. All these choices could affect the robustness of the ranking. Besides, these methodological choices are often criticized by many [7]. Therefore, in this section, a focus will be put on analyzing the robustness of the ranking. The robustness analysis introduced in [33] will be used for this purpose. It is a combination of uncertainty and sensitivity analyses. Uncertainty analysis concentrates on how uncertainty in the input propagates throughout the ranking. Sensitivity analysis however focuses on the contribution of the individual source of uncertainty to the output. So, the main difference between these two analyses is that the former only looks at the changes in output, without taking into account the proportion of input uncertainties in the output variability. On the other hand, the latter concentrates more on which input uncertainty made the output vary the most.

Uncertainty analysis is often used alone in other research papers that study university rankings. It is the case in [40] for instance. However, [33] shows that a combination of both uncertainty and sensitivity analyses leads to better robustness assessment and gives an increased transparency in the design of a ranking. A detailed example of a robustness analysis of the United Nations' technology achievement index, also called TAI[3], is presented in [41]. Ideally, a robustness analysis should assess all possible sources of uncertainties: choice of criteria, values of criteria, weights, normalization, aggregation rules, etc. In this work, I focus on the choice of criteria, values of criteria, weights and normalization. The aggregation rule is fixed to a linear aggregation.

A Monte Carlo methodology has been used to gauge the uncertainty of the FT MiM ranking. A Monte Carlo algorithm relies on repeated random samplings to simulate deterministic processes. Indeed, a high number of repetitions of random samplings over a probability distribution tends to reproduce quite well deterministic processes when the results are aggregated. Several random variables are defined to imitate the uncertainties in the input factors, one for each source of uncertainties. Four random variables are thus defined, namely X_1, X_2, X_3 and X_4. They respectively model uncertainties in values of criteria, weights, the choice of criteria and the normalization method.
It is also important to define the output to be analyzed beforehand, for we look at changes in the output due to uncertain inputs. Here, the resulting rankings will be designated as output, even though the difference in ranks between two schools or the difference in values of criteria could also be defined as outputs, as seen in [41]. The resulting rankings will be named Y. So, the model that links inputs and outputs can be defined as:

$$Y = f(X_1, X_2, X_3, X_4) \ . \tag{2.4}$$

In what follows, the methodologies of uncertainty and sensitivity analyses are explained. The results of both analyses are presented after that. The end of this section is devoted to a comparison between only two schools and how the robustness analysis can enhance this comparison.

2.3.1 Methodology

2.3.1.1 Uncertainty analysis

As a reminder, the uncertainty analysis aims to make the inputs vary and then to study the output variability, without taking into account which input uncertainties were responsible for the output changeability.
As explained before, a Monte Carlo methodology is used to simulate the uncertainties in the inputs. Recall that four random variables X_1, X_2, X_3 and X_4 were defined for this purpose. Each of these random variables has its own probability distribution and at each iteration of the Monte Carlo algorithm, a number is generated from each of the probability distributions and is linked to the associated random variable. The different probability distributions are explained below:

- X_1: there is an equal probability of adding a noise to the data of one of the sixteen criteria. This noise has a mean of zero and a standard deviation equal to the standard deviation of this criterion.
- X_2: concerning the weight, there is a 20% chance of equalize each weight to one. Otherwise, noise with zero mean and standard deviation equal to the standard deviation of the weights is added to each weight.

[3]http://www.insme.org/glossary/technology-achievement-index-tai

- X_3: there is a 50% chance of removing a criterion. All criteria have an equal probability to be removed.
- X_4: for the data normalization, there is an equal probability to choose between a ranking only, a minmax or a z-scores.

$$I_{i,j} = Rank_j(x_{i,j}) \ , \tag{2.5}$$

$$I_{i,j} = \frac{x_{i,j} - min(x_j)}{max(x_j) - min(x_j)} \ , \tag{2.6}$$

$$I_{i,j} = \frac{x_{i,j} - mean(x_j)}{sd(x_j)} \ , \tag{2.7}$$

where $I_{i,j}$ is the normalized data of the school i for the criterion j. $Rank_j()$ means a ranking over the criterion j and $sd()$ is the standard deviation.

2.3.1.2 Sensitivity analysis

It can be recalled that a sensitivity analysis focuses on the contribution of the individual source of uncertainties X_1, X_2, X_3 and X_4 to the output, meaning that it concentrates more on which input uncertainty makes the output vary the most. A model for the analysis has been introduced earlier: $Y = f(X_1, X_2, X_3, X_4)$, where Y is the output of interest (the resulting rankings in this work). X_1, X_2, X_3 and X_4 respectively simulate uncertainties in values of criteria, weights, the choice of criteria and the normalization methods. The probability distribution of these random variables are the same than in the uncertainty analysis.
Different methods exist to analyze the sensitivity of a ranking model. In this work, variance-based techniques are adopted. They were used in [33] and [41]. These techniques focus on the contribution of sources of uncertainty (X_1, X_2, X_3 and X_4) to the output variance, written $V(Y)$.

To compute the variance-based sensitivity measure for a given input X_i, the fractional contribution of X_i to the model output variance $V(Y)$ is assessed, meaning that the variance of Y due to the uncertainty in X_i is calculated. To do this, the value of X_i is fixed and the mean output Y is computed while varying the other random variables. This is done an extensive number of times, performing thus a Monte Carlo algorithm on the other three sources of uncertainty. This procedure is then repeated for all possible values of X_i. There are thus one mean output Y for all possible values of X_i. The variance of these means, noted V_i, represents the contribution of the uncertainty in X_i to the output variance $V(Y)$. Mathematically speaking, V_i can be described as:

$$V_i = V_{X_i}(E(Y|X_i)) \ . \tag{2.8}$$

The range of V_i comes from 0 (if X_i does not make any contribution to Y) to $V(Y)$ (if all other random variables are non-influential to the considered output).

V_i represents a variance contribution of the first-order term, meaning that the variance in the output for only one random variable is taken into account at a time. However, higher-order variances can also be defined. They represent interactions among several inputs. For instance, a variance contribution of the second-order term is:

$$V_{ij} = V_{X_iX_j}(E(Y|X_i, X_j)) - V_{X_i}(E(Y|X_i)) - V_{X_j}(E(Y|X_j)) \ . \tag{2.9}$$

V_{ij} describes the variance due to interactions among uncertainties in inputs X_i and X_j. This is illustrated by the above formula. V_{ij} is the contribution of the uncertainties in X_i and X_j to the output variance $V(Y)$, minus V_i and V_j, so that only variance due to X_i and X_j varying at the same time is taken into account.

When all random variables X_i are independent, as it is the case in this sensitivity analysis, the output variance $V(Y)$ can be decomposed into first- and higher-order terms as follows:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \sum_i \sum_{j>i} \sum_{l>j,j>i} V_{ijl} + V_{1234} \ . \tag{2.10}$$

The meaning of third- and fourth-order terms can easily be guessed from the definition of V_{ij}. There is no fifth-order term in this case as there are only four random variables X_i.

Now that the variance contributions have been described, indices that will be used to assess the sensitivity of the ranking can be defined. Two types of indices are looked at in this analysis. First of all, a sensitivity index could be defined. A first-order sensitivity index is obtained through a normalization of a first-order term V_i by the total variance $V(Y)$:

$$S_i = \frac{V_i}{V_Y} \ . \tag{2.11}$$

The meaning of S_i is relatively similar of the first-order term V_i. It is simply the variance V_i when its range spans from 0 to 1.
The main advantage of this index is that it will be much more easier to compare the S_i from two different schools than their respective V_i. Indeed, if one of these two schools has a much higher output variance than the other one, its V_i will be likely higher too.

Likewise, a higher-order sensitivity index is obtained through normalization of a higher-order term $V_{i...l}$ by the total variance $V(Y)$:

$$S_{i..l} = \frac{V_{i..l}}{V_Y} \ . \tag{2.12}$$

The second type of index that could be interesting to look at is the total effect sensitivity index. It sums all the indices $S_{i..l}$ related to a given input factor. For instance, with four different random variables, the total effect sensitivity index for the random variable X_1 is calculated as:

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{14} + S_{123} + S_{124} + S_{134} + S_{1234} \ . \tag{2.13}$$

Analyzing S_i and S_{Ti} could be useful to understand the differences between the variance contribution that corresponds to one input and the one that is due to interactions among several inputs. Indeed, S_i represents the contribution of uncertainties in the input X_i to the output variance while $(S_{Ti} - S_i)$ describes the contribution of interactions among uncertainties in X_i and uncertainties in every other X_j to the output variance. Comparing these two values could help us understand if uncertainties in X_i have more effect on the output variance when they are alone or when they are coupled with uncertainties from other inputs.

2.3.2 Results

2.3.2.1 Uncertainty Analysis

Concerning the uncertainty analysis, the Monte Carlo algorithm presented earlier consisted in 2500 iterations, meaning that the random variables X_1, X_2, X_3 and X_4 were assigned each one a value 2500 times and 2500 rankings were created based on that. Figure 2.3 depicts the results as boxplots. Upper and lower parts of the boxes represent respectively the third and first quartiles, while upper and lower parts of the dashed lines illustrate respectively the 95^{th} and 5^{th} percentiles. Outliers are not displayed here. Schools are ordered on the x-axis according to their initial FT MiM ranking. Small red crosses show the FT MiM ranking for each school. The red boxplot represents the Louvain School of Management (school 41) while the green one depicts

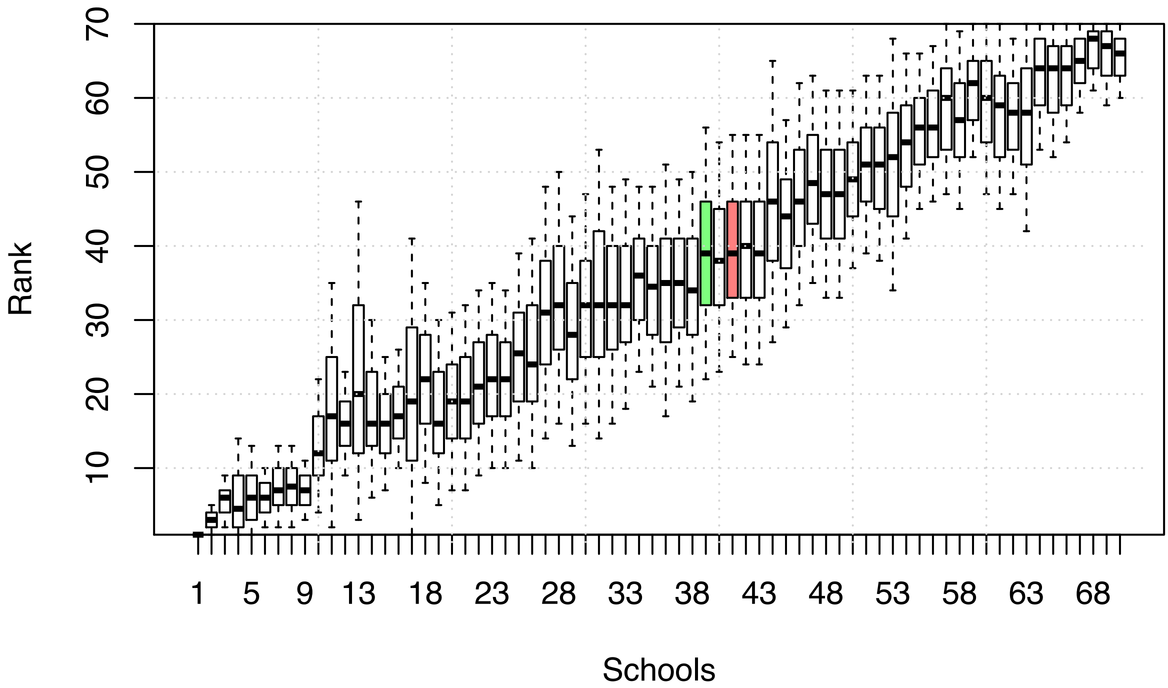


Figure 2.3: Rankings of schools during the uncertainty analysis.

the Solvay Brussels School of Economics and Management (school 39)[4].

Several comments of this plot can be made. First of all, it can be seen that some schools present lower variance. It is the case for the three worst-performing and the nine best-performing schools (even though the school WHU Beisheim (school 4) presents a variance a little higher than its neighbours). However, other schools span a much broader range. For instance, schools Indian Institute of Management, Calcutta (school 13) and Indian Institute of Management, Ahmedabad (school 17) were given a rank ranging over more than forty places. This is quite a big difference. Overall, the 5^{th} and 95^{th} percentiles of schools have a difference of twenty places in rankings.

Another aspect that stands out is the boxplot of the University of St Gallen (school 1). It is the only one to have a boxplot reduced to one rank (without the outliers). Figure 2.4 shows boxplots of z-scores that each school obtained during the ranking procedure[5]. It displays well the fact that the University of St Gallen (school 1) outperformed the other schools in the majority of rankings. This graph also explains why the nine best-performing schools had a lower variance in Figure 2.3. This is because these schools overall received a distinctively better score compared to other schools in the rankings.

2.3.2.2 Sensitivity Analysis

As a reminder, two indices were introduced to assess results of a sensitivity analysis: the sensitivity and total effect sensitivity indices S_i and S_{Ti}.

Figure 2.5 depicts the variance contributions for the first-order terms V_i. The high variances

[4] They are highlighted as they will be used later as a comparison example.

[5] Values resulting of rankings which normalized criteria differently (ranking only, min-max, z-score), they needed to be transformed into z-scores in order to be compared.

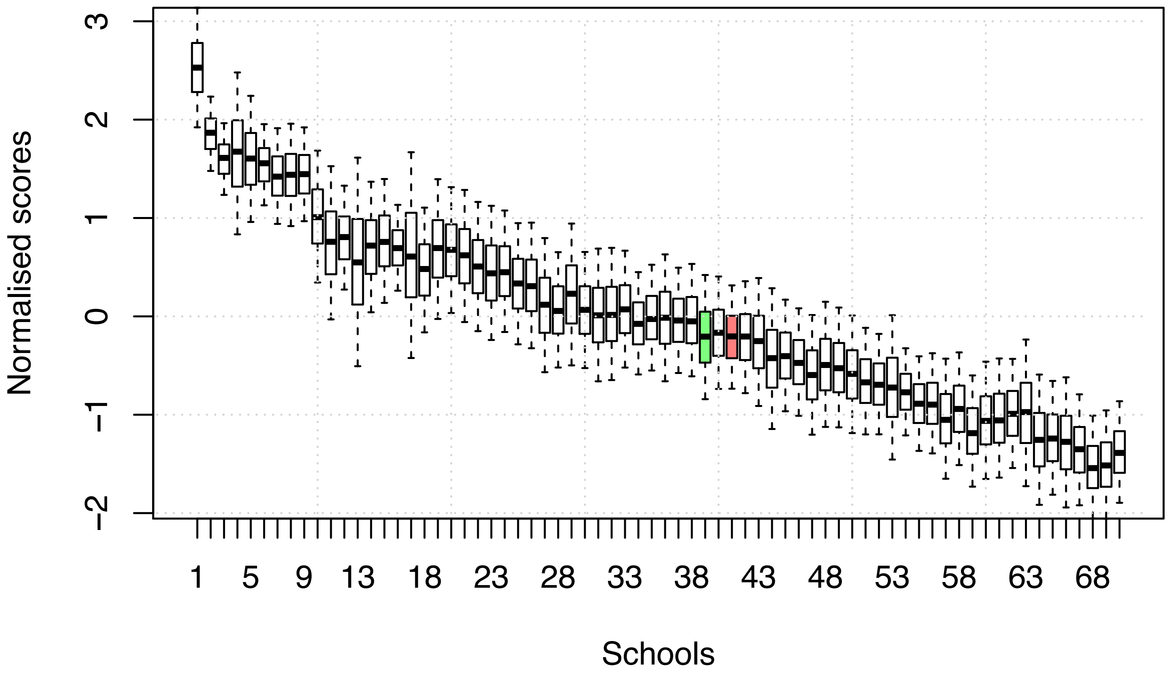


Figure 2.4: Z-scores of schools during the uncertainty analysis.

of schools Indian Institute of Management, Calcutta (school 13) and especially Indian Institute of Management, Ahmedabad (school 17) really stand out. They were the two schools with the widest boxplot in Figure 2.3. It is interesting to note that these two schools do not present the same variance of first-order terms even though the wideness of their respective boxplot was similar in Figure 2.3. It allows us to demonstrate that the variance observed in the uncertainty analysis does not necessarily have the same source. While the Indian Institute of Management, Ahmedabad (school 17) seems to have a lot of variance coming from first-order terms, Figure 2.5 indicates that the Indian Institute of Management, Calcutta (school 13) has more variance coming from higher-order terms, as both schools have the same overall amount of variance because of their same wideness boxplot.

Besides that, it can be noticed that the best- and worst-performing schools have a low variance for the first-order terms. Again, it really is in accordance with boxplots in Figure 2.3. Furthermore, it can be observed that the first-order terms for weights and exclusions have generally a higher contribution to the overall variance. This is backed up by the information in Table 2.5. This table displays the first-order and total effect sensitivity indices for this sensitivity analysis. It can be seen that X_2 has the highest first-order index and X_3 the second one. X_1, representing uncertainties in data, has the lowest first-order index.
Moreover, from the difference $(S_{Ti} - S_i)$, it can be seen that X_2 has the highest difference while X_3 has the lowest one. It means that a bigger proportion of variance relative to X_2 comes from higher-order terms than variance relative to other random variables.

	Random Variables	First-order index (S_i)	Total effect index (S_{Ti})	Difference ($S_{Ti} - S_i$)
Data	X_1	0.115	0.325	0.210
Weights	X_2	0.201	0.431	0.230
Exclusion	X_3	0.157	0.338	0.181
Normalization	X_4	0.150	0.375	0.225

Table 2.5: First-order and total effect sensitivity indices in the sensitivity analysis.

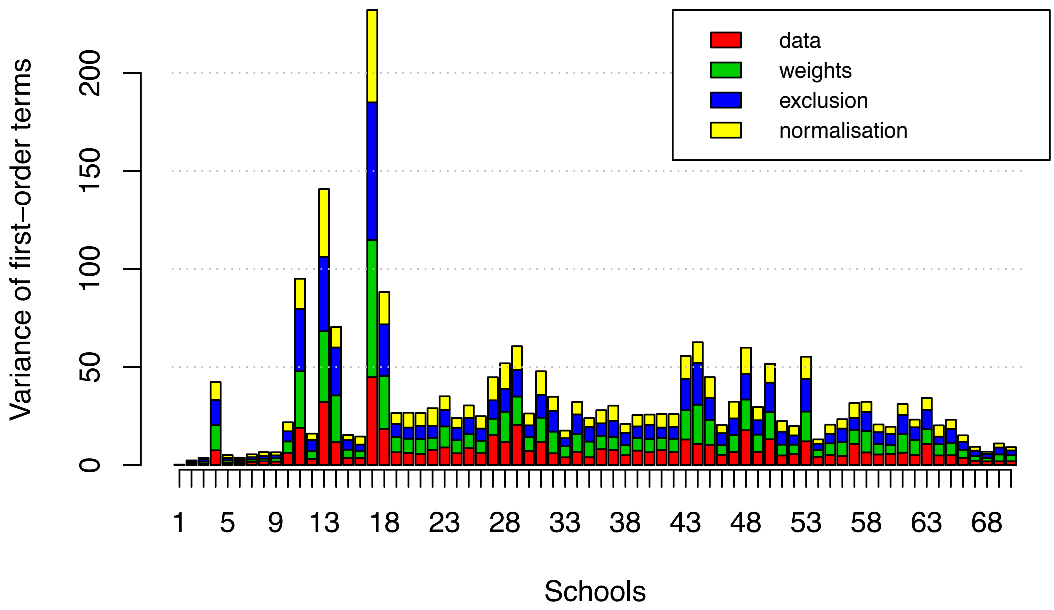


Figure 2.5: Barplots of variance contributions for the first-order terms V_i.

Another important aspect of Table 2.5 is the sum of S_i. It represents 62.3% of the total variance $V(Y)$. This means that 37.7% of the overall variance comes from higher-order terms which represent interactions between several random variables X_i. This is a non-negligible part and explains why the total effect indices S_{Ti} are of great importance too. The latter are also displayed in Table 2.5. It can be observed that uncertainties in weights have still the highest index. However, the second highest index is now uncertainties in normalization methods. Figure 2.6 shows the total effect sensitivity indices for each school. The first thing that can be noticed is the regularity of proportions of each total effect sensitivity index among schools. Besides, it can be seen that the sum of indices is greater than one. This is explained by the fact that interactions among several X_i are counted in each S_{Ti}.

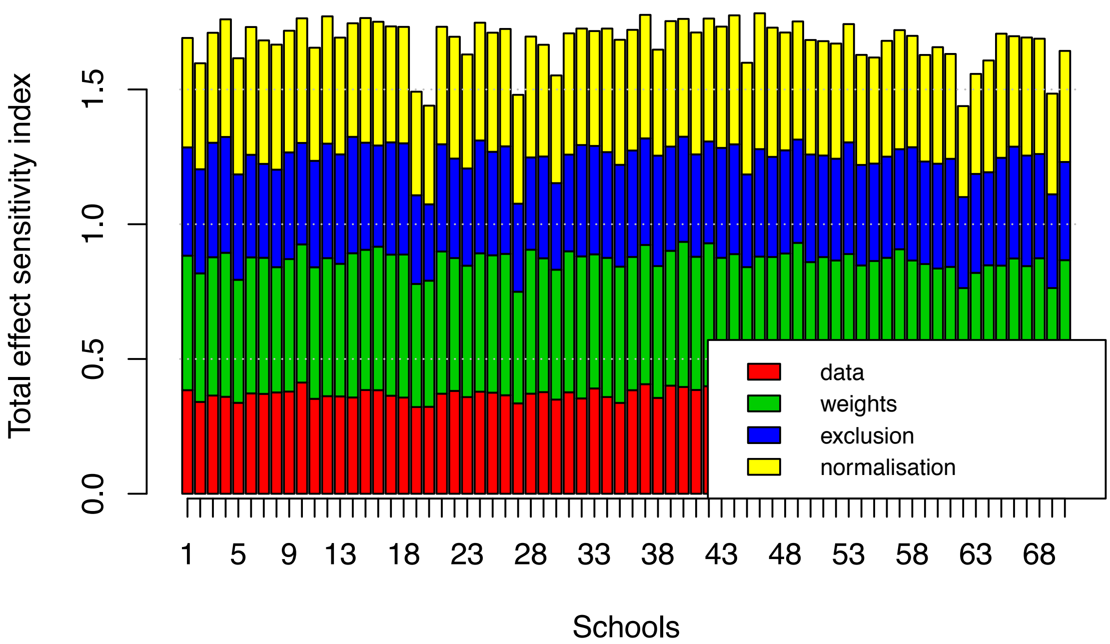


Figure 2.6: Barplots of the total effect sensitivity indices S_{Ti}.

2.3.2.3 Comparison between two schools

Before finishing the sensitivity analysis, it could be interesting to compare the sensitivity in ranking between two schools. Indeed, a global sensitivity analysis among all the schools was conducted so far and the focus was not really put on the order of schools in the ranking. A comparison will be carried out between the Louvain School of Management (school 41) and the Solvay Brussels School of Economics and Management (school 39). I arbitrarily chose the Louvain School of Management. The Solvay Brussels School of Economics and Management was picked up because it is the only other Belgian French-speaking school in the FT MiM ranking. These schools will be respectively called LSM and Solvay from now on and in this comparison only.
The LSM and Solvay are respectively ranked 41 and 39 in the FT MiM ranking. It could be interesting to know if this is still the case in the different rankings developed during the robustness analysis. Both schools are highlighted in Figure 2.3. The LSM is in green while Solvay is in red. In the general sensitivity analysis, the observed output was the ranking of schools. Here, a additional output will be considered that better suits our needs: the difference in ranks between the two schools:

$$D_{ij} = Rank_i - Rank_j = g(X_1, X_2, X_3, X_4) \ , \tag{2.14}$$

where i and j are the indices of two schools.

Figure 2.7 illustrates a histogram of D_{ij} for the LSM and Solvay. A negative D_{ij} means that Solvay is better ranked while a positive difference means that LSM performs better. This histogram is based on 2500 rankings generated with the method used in the uncertainty analysis. 1254 rankings ranked Solvay before LSM while the other 1246 ones ranked LSM first, meaning that 50.16% of the rankings decided that Solvay was better-performing. This is nearly an equality. However, it can be seen that there are more rankings where Solvay outperforms dramatically LSM. For instance, the former is ranked at least ten places before the latter in 16.36% of the rankings, while 14.36% of them present the opposite. Solvay beats LSM by twenty places or more in 3.28% of the rankings, while LSM outperforms Solvay by at least twenty places in 2.04% of the rankings.

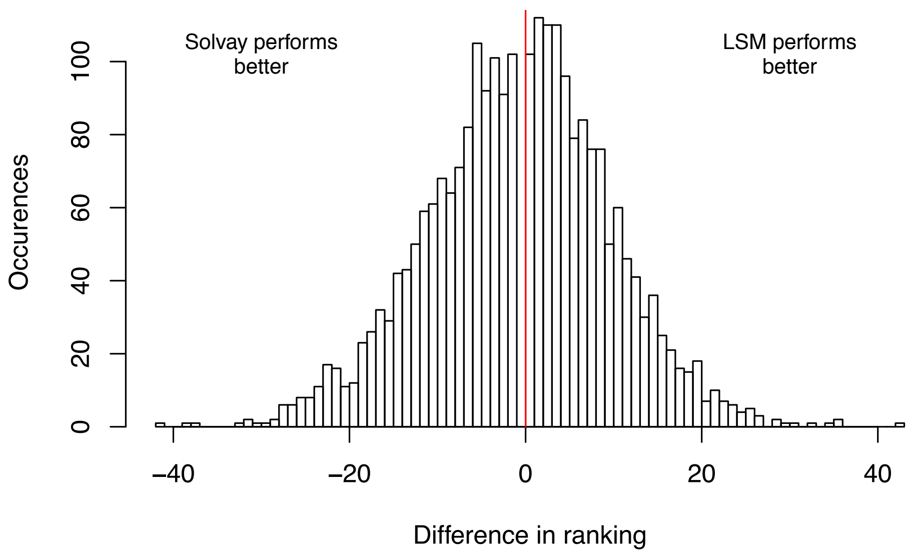


Figure 2.7: Rankings of schools during the uncertainty analysis.

Table 2.6 displays the first-order indices S_i and total effect sensitivity indices S_{Ti} of the output D_{ij} between the two schools. The input factors X_2 and X_4 (representing respectively uncertainties in weights and normalization rules) have the highest first-order indices while it was the input factors X_2 and X_3 that had the highest first-order indices for the output Y during the sensitivity analysis. Here, 39.5% of the total variance comes from interactions between several factors. This is more than for the output Y. For the total effect sensitivity indices S_{Ti}, the highest values are the ones from the input factors X_2 and X_4. The input factor X_3 (representing the uncertainties in data) has a high index too.

	Random Variables	First-order index (S_i)	Total effect index (S_{Ti})	Difference ($S_{Ti} - S_i$)
Data	X_1	0.111	0.357	0.246
Weights	X_2	0.215	0.408	0.193
Exclusion	X_3	0.123	0.312	0.189
Normalization	X_4	0.156	0.389	0.233

Table 2.6: First-order and total effect sensitivity indices of D_{ij} in the comparison between two schools.

Before finishing this comparison between the two schools, it could be interesting to give a little more attention to the uncertainties in weights and normalization rules. Indeed, there are the two factors that have both the highest first-order index S_i and the highest total effect sensitivity index S_{Ti}. The uncertainties in weights and normalization rules will be analyzed by attempting to find the combinations of weights and normalization rules that allow LSM or Solvay to beat the other in the ranking. Figures 2.8, 2.9 and 2.10 show the mean of the sixteen criteria for each normalization rules: ranking only, minmax and z-scores. The "similar" weights are the weights when LSM and Solvay are ranked one or two places from each other. The "better" weights are the weights when LSM is ranked three or more places before Solvay and the "worse" weights are the opposite. These data were computed based on 500 rankings with no uncertainty in data, no exclusion and with a linear aggregation rule.

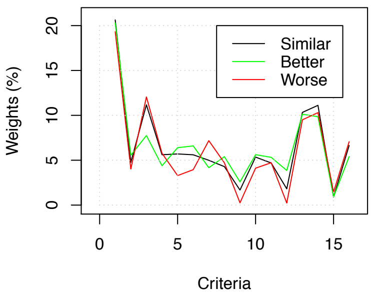


Figure 2.8: Weights of criteria with a ranking normalization.

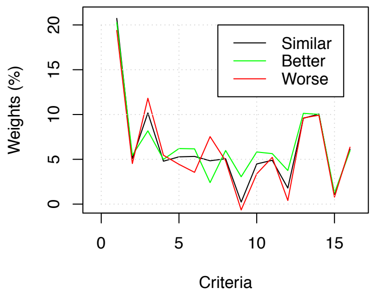


Figure 2.9: Weights of criteria with a minmax normalization.

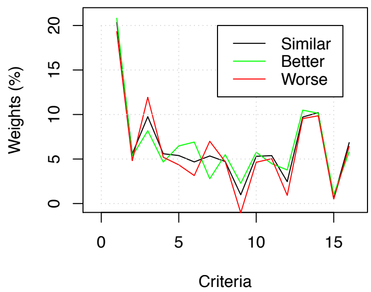


Figure 2.10: Weights of criteria with a z-score normalization.

It can be seen from these line graphs that the mean for some weights are not modified when the normalization method changes. For example, it is the case for the criteria *Value for money* (2) and *Careers* (3) where the weights for the "similar", "better" and "worse" cases remain steady. However, the difference between these two criteria is that the weights for the three cases are nearly equal for the criterion *Value for money* (2) while it can be noticed that the criterion *Careers* (3) has a clearly lower weight when LSM performs better. Apart from that, there are other criteria which are much more interesting in our study of the uncertainties in weights and normalization rules as their weights for each case vary a lot depending on the normalization rule. It is the case for the criterion *Women board* (9). The weights for this criterion can be quite different among the three cases depending on the normalization method. For instance, it can be observed that the weight in the "better" case with a minmax normalization is much more discriminated than with other types of normalization rules.

Figure 2.11 depicts boxplots for the weights of the criterion *Women board* (9), depending on the normalization method. Outliers are not displayed here. The FT weight for this criterion is 1% (the blue dashed line on the graph). The first three boxplots represent weights when a ranking-only normalization is applied, the following three ones when a minmax normalization is adopted and the last three ones when z-scores are used. It can be seen that there is a clear gap between the "better" case and the two other cases with a minmax normalization. This could mean that this criterion plays a role in the fact that LSM is better ranked when this normalization rule is applied. Overall, the weight for this criterion when Solvay performs better is lower. This also could be induced by the fact that the criterion *Women board* (9) is a weakness for Solvay.

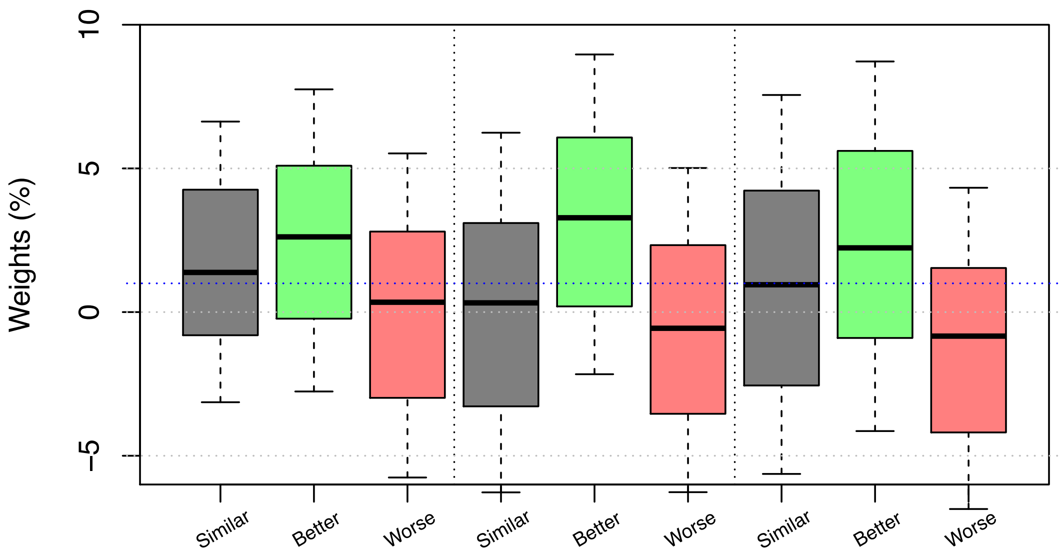


Figure 2.11: Boxplots of weights for the criterion 9 depending on the normalization rule: only ranking (left), minmax (center) and z-score (right).

In conclusion, in this section, a more thorough analysis was conducted between two schools. It allowed us to understand how a robustness analysis can enhance the comparison between two schools in the ranking. I had to choose only two of them for the sake of conciseness: the Solvay Brussels School of Economics and Management (school 39) and the Louvain School of Management (school 41). They are both very similar according to the FT MiM ranking and the uncertainty analysis draws the same conclusion (as could be seen in Figure 2.7). A more precise study of the variability in weights and normalization methods demonstrated that which arrangement of weights was more advantageous for one school or the other.

2.3.3 Conclusion

The robustness analysis was decomposed into two parts: uncertainty and sensitivity analyses. The first one aimed to assess the variability of the ranking when some inputs changed: values of criteria, weights, the choice of criteria and the normalization method. It showed which business school tended to vary the most but more importantly, which part of the ranking managed to stay the same even in presence of uncertainties. It was the case for the top ten schools and the five worst-performing ones. This illustrated the fact that there exist two groups that are respectively far ahead and far below the main group representing schools that are average. This was already partly discovered during the reconstruction of the ranking (in Section 2.2).

While other research papers on university rankings generally stop their robustness analysis here, a sensitivity analysis was also conducted in this work. Whereas the uncertainty analysis only focused on the output when facing uncertainties in the inputs, the sensitivity analysis allowed us to understand which uncertainty tended to make the output vary the most. It showed that it was the weights and the normalization method.

The last part of this section was devoted to a thorough comparison between two schools, to explain how the results of a robustness analysis could be used to enhance a pairwise comparison. In particular, it showed which set of weights was more advantageous for one school of the other. A future work could be to analyze more deeply the comparison between each pair of schools. Indeed, the only comparison made here was between Solvay Brussels School of Economics and Management (39) and Louvain School of Management (41), as this work aims to be general. It could be interesting to conduct this study for two schools among the best-performing schools in the FT MiM ranking and two schools among the worst-performing ones. Also, it could be useful to analyze other pairs of input factors than the uncertainties in weights and normalization methods.

3 Dimensionality reduction methods

Abstract This chapter is devoted to the introduction of dimensionality reduction methods and the definition of their theoretical frameworks. They are of particular interest as they allow to visualize the schools in 2D or 3D, so that the relationships between them will be more easily described. The first introduced method is the Principal Component Analysis (PCA). It is based on a linear model and is the method that is traditionally used to analyze university rankings in other research papers. Then, two non-linear dimensionality reduction (NLDR) methods are broached: the Self-Organizing Map (SOM) and the Multi-Scale Jenhsen-Shannon Embedding (Ms. JSE). They are more complex but could grasp non-linear manifolds in the data.

3.1 Introduction

In the previous chapter, it has been seen that pairwise correlations between criteria and multicollinearity are present in the FT data set. Besides, it has been discovered throughout the robustness analysis that the choice of weights was the methodological design that makes the ranking vary the most. However, criticisms on the fact that university rankings do not represent truthfully the distance separating each school or that they do not allow to define neighbourhoods of schools, have not been tackled yet. Therefore, this chapter aims to present possible methods to visualize the data in low dimensions (for instance, 2D or 3D), so that it will be easier to understand the relationships between schools and if the FT MiM ranking truthfully translate them.
Furthermore, knowing universities that present the same characteristics could be very useful to deans and university administrators. Indeed, knowing from which schools a university has to distinguish itself could help it to climb up the ranking faster.

The first introduced method is the Principal Component Analysis (PCA). It is a linear method that attempts to create from N variables of a data set, N linearly uncorrelated variables, via an orthogonal transformation while maximizing the variance of each axe after projection. These new variables are called principal components. From this, a dimensionality reduction can be implemented by only taking into account a couple of principal components. The chosen ones are preferably the ones that present the highest variance after projection. A dimensionality reduction using principal components is also called *reversed PCA*.
The PCA is introduced in this work as it is the consensual method used to analyze university rankings in other research papers, like in [44]. Its popularity could be explained by the fact that it is rather easily implemented and the signification of the principal components are quite straightforward as they are formed via a linear combination of old variables. This last characteristic is actually appealing as the meaning of axes created with non-linear dimensionality reduction methods will be much harder to fathom.

The two other introduced methods are non-linear dimensionality reduction (NLDR) ones.

They are more complex than PCA but could grasp non-linear models in data. Besides, they have never been used to analyze university rankings in other research papers, to my knowledge. The first NLDR method presented in this work is the Self-Organizing Map (SOM). It focuses on the topology preservation. In short, it implies that it attempts to preserve the topology of the data set instead of more usual measures like the distances between points of the data set, meaning that it tries to preserve neighbourhood relationships between subregions of the manifold.
The interest of SOM is that it creates a discrete mapping model in low dimensions *a priori* and then tries to map this model in high dimensions while linking schools to nodes of the mapping model. Thus, by creating a model with fewer nodes than the number of schools, several universities are linked to the same node so that it is possible to compare the resemblance between schools.

The last presented method is the Multi-Scale Jensen-Shannon Embedding (Ms. JSE). It is a more state-of-the-art NLDR method based on similarities. It centers a Gaussian on each point of the data set in high dimensions and uses the densities under the Gaussian to define a probabilistic distribution over a neighbourhood of fixed size. Then, it attempts to approximate the probabilities (called similarities) between points in low dimensions. The method contains the word *multi-scale* in its designation as several neighbourhoods of different sizes are designed during the execution.
The interest of Ms. JSE is that it is a state-of-the-art NLDR method, based on similarity preservation. The expectation of sterling results is thus high. At least, it is supposed that Ms. JSE will outperform PCA, even though the latter presented practical results due to its linear model.

Before beginning to explain these methods, it is important to get on with notations. Usually, the high-dimensional space is considered to have N dimensions while the low-dimensional space has M dimensions. Also, the data set is supposed to be composed of K points. Furthermore, a particular vocabulary is adopted when considering non-linear dimensionality reduction methods. For instance, the term *manifold* is sometimes used to express the high-dimensional space in which the points of the data set lie. This has to be differentiated from the data set itself, which is a set of points belonging to a manifold. Also, the term *embedding* in NLDR methods can be seen as a synonym of *dimensionality reduction*.

3.2 Principal Component Analysis (PCA)

3.2.1 Basic idea

The Principal Component Analysis (PCA) is a linear method that attempts to create from N variables of a data set, N linearly uncorrelated new variables, via an orthogonal transformation while maximizing the variance of each axe after projection. These new variables are called principal components. A dimensionality reduction, called *reverse PCA*, can be performed by only taking into account the M first principal components, with $M \leq N$. The chosen principal components are preferably the ones that present the highest variance after projection.

3.2.2 Theory

In this description of a PCA, a data set, named Ξ, in N dimensions (so, characterized by N variables) is assumed. The goal of a Principal Component Analysis is to infer a new set of N variables from the N basic dimensions. These new variables, called principal components, are such that they present some useful statistical properties. In this section, the N variables of the data set Ξ are written x_i while the N new variables are written z_i.

The covariance matrix of the variables x_i can be defined as:

$$\boldsymbol{C_{XX}} = \begin{bmatrix} C_{x_1x_1} & \cdots & C_{x_1x_N} \\ \vdots & \ddots & \vdots \\ C_{x_Nx_1} & \cdots & C_{x_Nx_N} \end{bmatrix} \quad . \tag{3.1}$$

Remember that covariance is based on centred values. For instance, $\boldsymbol{C_{X_iX_j}}$ is obtained via the expression $E[(\boldsymbol{X_i} - \bar{X}_i)(\boldsymbol{X_j} - \bar{X}_j)]$.

One of the statistical properties that principal components z_i should have is that the data expressed following the z_i should be white. It means that the covariance matrix should be equal to the identity matrix and that the expectation for each z_i should be zero. This implies that the z_i form an orthogonal coordinate system. The covariance matrix of z_i can be written as:

$$\boldsymbol{C_{ZZ}} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & 1 & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \quad . \tag{3.2}$$

The relation between the x_i and z_i can then be expressed as a whitening process. In matrix notation, it gives:

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{X} \quad , \tag{3.3}$$

where $\boldsymbol{V}$ can be decomposed as:

$$\boldsymbol{V} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Theta}^T \quad . \tag{3.4}$$

$\boldsymbol{\Lambda}$ contains the eigenvalues of the covariance matrix $\boldsymbol{C_{XX}}$ in its diagonal while $\boldsymbol{\Theta}$ holds the eigenvectors of $\boldsymbol{C_{XX}}$ in columns. They can be made explicit:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \lambda_i & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \tag{3.5}$$

and

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_1 & \cdots & \boldsymbol{\Theta}_N \end{bmatrix} \quad . \tag{3.6}$$

Besides attempting to design the z_i to form an orthogonal coordinate system, the axes that maximize variance after projection are looked for. For instance, considering this linear function connecting the x_i and z_i:

$$z_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{iN}x_N \quad , \tag{3.7}$$

or in matrix notation: $z_1 = \boldsymbol{a}_1\boldsymbol{X}$, the maximization of the variance after projection could be written as follows:

$$Var(z_1) = E[z_1z_1] = \max_{\boldsymbol{a}}\ \boldsymbol{a}_1^T\ E[\boldsymbol{X}\boldsymbol{X}^T]\ \boldsymbol{a}_1 = \max_{\boldsymbol{a}}\ \boldsymbol{a}_1^T\ \boldsymbol{\Theta}^T\boldsymbol{\Lambda}\boldsymbol{\Theta}^T\ \boldsymbol{a}_1 \quad . \tag{3.8}$$

It can be proved that the best choice for $\boldsymbol{a_1}$ is the eigenvector $\boldsymbol{\Theta}_1$ associated with the largest eigenvalue λ_1 of the covariance matrix $\boldsymbol{C_{XX}}$. By generalizing it, it can be demonstrated that the best choice for $\boldsymbol{a_i}$ is the eigenvector $\boldsymbol{\Theta}_i$ associated with the eigenvalue λ_i of the covariance matrix $\boldsymbol{C_{XX}}$, when the eigenvalues are ordered in decreasing order.

Equation 3.7 well shows that principal components z_i are based on a linear model. Once the vectors $\boldsymbol{a_i}$ have been defined, the PCA can be used to assess real distributions of variables x_i by controlling the loadings of each variable in the first principal components. This will be study in the next chapter devoted to the results.

Quite often, a PCA is coupled with a regression analysis, meaning that a dimension reduction is performed. Obviously, when a reduction to M dimensions is conducted, only the M principal components presenting the highest variance are kept. One way of assessing the loss of information during a reduction to M dimensions is to sum the M remaining eigenvalues λ_i and to compare them to the sum of all the eigenvalues:

$$\%_{variance\ kept} = \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \ . \tag{3.9}$$

However, it is important to notice that this implies a strong assumption of the equivalence between variance and information kept. Even though the dimension reduction keeps 95% of variance, the useful information could be hidden in the discarded 5%.

3.3 Self-Organizing Map (SOM)

This section is dedicated to introduce an NLDR method that attempts to preserve the topology of the high-dimensional manifold. Early NLDR methods focused more on distance preservation. However, this last kind of preservation has some drawbacks, including the fact that describing a manifold with distances turn out to bolt it with rigid steel beams [30]. However, quite often, the embedding of a manifold requires some adaptability to perform a dimensionality reduction. Some subregions should be stretched while other ones should be shrunk. The NLDR methods that are studied in this section allows more flexibility, for it attempts to preserve the topology instead of the distances, meaning that it tries to preserve neighbourhood relationships between subregions of the manifold, thus allowing different subregions to be stretched or shrunk depending on their underlying data.

One of the major difficulties of topology-preserving NLDR methods is the characterization of the topology of the manifold in high dimensions. Indeed, only few points belonging to this manifold are available in the data set, making its characterization troublesome. Yet, the precise description of a manifold is crucial to preserve the topology during the dimensionality reduction. Most of the topology-preserving NLDR methods define a discrete mapping model (or *lattice*) to characterize the topology. Early methods construct the lattice *a priori*. It is the case for methods like Self-Organizing Maps (SOM) or Generative Topographic Mapping (GTM). Other methods work with a data-driven lattice, meaning that the structure of the lattice evolves while the method is running. An example of this kind of methods is the Locally Linear Embedding (LLE). In this work, only the SOM is broached. More information can be found in [4] for GTM and in [39] for LLE. They are also both deeply described in [30].

3.3.1 Basic idea

A discrete mapping model, called *lattice*, is defined in the low-dimensional space. Its dimension, size and topology are arbitrarily defined by the designer *a priori*. The links of the schools to the nodes of the lattice, as well as the positions of the nodes in the high-dimensional space are determined iteratively until reaching convergence.
Starting from the low-dimensional space is somewhat surprising as the majority of DR methods begin from the high-dimensional space.
If the lattice presents less nodes than the number of schools, several universities are linked to a same node, thus facilitating the determination of close neighbourhoods among the schools.

3.3.2 Theory

Self-organizing maps, also called SOM, were first introduced by von der Malsburg in 1973 [42] and Kohonen in 1982 [23]. The former proposed a biologically motivated model to explain the problem of mapping from the retina to the visual cortex while the latter provided a more general model that captured the essential features of computational maps in the brain [19]. Because of this source of inspiration, self-organizing maps have neurobiologically-based concepts. For instance, the components of a self-organizing map are sometimes called neurons, in reference of the components of the brain.

A SOM is an artificial neural network. Its algorithm is designed in several steps. First of all, a grid (or lattice) of size S is defined in the low-dimensional space: $\mathbf{G} = [\mathbf{g}_i]_{1 \leq i \leq S}$. The dimension, the topology and the number of nodes of the lattice are decided arbitrarily. Usually, the dimension is one or two and the topology of the lattice is a rectangle or an hexagon. The number of nodes (or neurons) of the lattice, S, is generally lower than the number of high-dimensional data. That way, a vector quantization can be performed. The grid nodes also have high-dimensional coordinates: $\mathbf{\Gamma} = [\boldsymbol{\gamma}_i]_{1 \leq i \leq S}$. The grid can be initialized randomly.
Figure 3.1 represents a basic two-dimensional lattice. Input data are three-dimensional and the lattice is structured as a four-by-four grid. It can be seen that each input data influence all the sixteen nodes of the lattice.

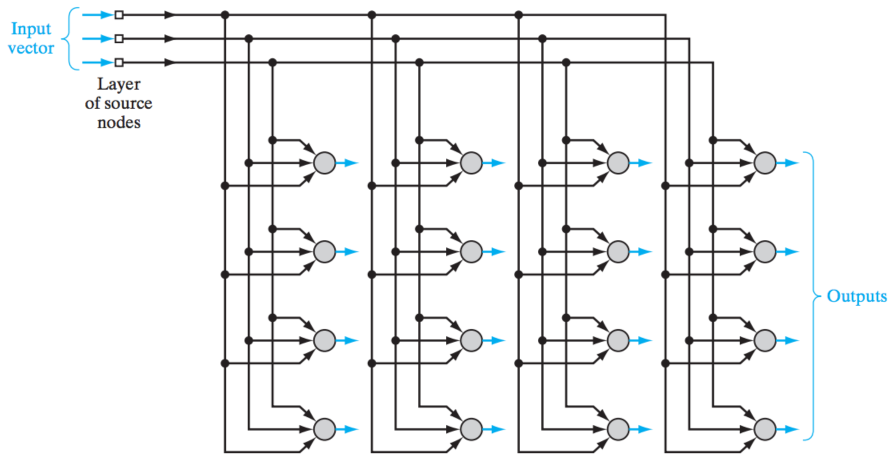


Figure 3.1: Two-dimensional lattice with a three-dimensional input and four-by-four dimensional output [19].

The second step is an adaptive procedure that updates iteratively the high-dimensional coordinates $\boldsymbol{\gamma}_i$. It includes three processes involved in the development of the self-organizing map: a competitive process, a cooperative process and an adaptive process. At each iteration, all high-dimensional data vectors ξ_k are selected one by one in random order.

During the **competitive** process, the closest SOM node i is calculated for each data vector $\boldsymbol{\xi}_k$:

$$i = \arg\min_j ||\boldsymbol{\xi}_k - \boldsymbol{\gamma}_j||_2 \quad .$$

It can be noticed that the closest node is computed according to the high-dimensional coordinates. This process is said to be competitive as each SOM node calculates the distance separating it from the data vectors and tries to be closer than the other SOM nodes.

During the **cooperative** process, a topological neighbourhood of excited neurons centred on the winning neuron i is computed. Indeed, in a human brain, it has been found that an excited neuron tends to excite the nearby neurons. In the rest of this section, the topological neighbourhood centred on winning neuron i for a data vector $\boldsymbol{\xi}_k$ will be denoted by $h_{j,i(\boldsymbol{\xi}_k)}$, where j represents a possible neighbour of neuron i. To ensure that nearby neurons are more excited than distant ones, it seems logical to define $h_{j,i(\boldsymbol{\xi_k})}$ to decrease with distance between i and j. Early works (as [42] or [23]) constructed the topological neighbourhood function as:

$$h_{j,i(\boldsymbol{\xi_k})} = \begin{cases} 0 & \text{if } \delta(i,j) > \lambda \\ 1 & \text{if } \delta(i,j) \leq \lambda \end{cases} , \tag{3.10}$$

where λ is the neighbourhood width. Usually, λ is reduced after each iteration. However, this function decreases with $\delta(i,j)$ quite abruptly: it equals to zero when this distance exceeds a given threshold. This last bizarre effect pressed more recent works (as [19]) to define the topological neighbourhood function as:

$$h_{j,i(\boldsymbol{\xi_k})} = \exp(-\frac{\delta^2(i,j)}{2\lambda^2}) , \tag{3.11}$$

which looks more like a Gaussian function with the neighbourhood width λ instead of the variance. It can be seen that the maximum value of $h_{j,i(\boldsymbol{\xi_k})}$ is attained when $\delta(i,j) = 0$, i.e. when j is the winning neuron i. Furthermore, it can be noticed that this function decreases more smoothly with $\delta(i,j)$ than the previous presented topological function, as $h_{j,i(\boldsymbol{\xi_k})}$ tends to zero when $\delta(i,j)$ tends to the infinity. The neighbourhood width λ is also constructed so as to decrease over time. A possible definition for λ has been proposed in [38]:

$$\lambda(n) = \lambda_0 \exp(-\frac{n}{\rho_1}) , \tag{3.12}$$

where n represents the n^{th} iteration. λ_0 is the initial neighbourhood width and should be assigned to a rather large value. ρ_1 is a time constant to be chosen by the designer. This definition of $\lambda(n)$ can be seen as gradually increasing the selectivity of the individual neurons to be excited in the course of the learning process. The topological neighbourhood function becomes:

$$h_{j,i(\boldsymbol{\xi_k})}(n) = \exp(-\frac{\delta^2(i,j)}{2\lambda^2(n)}) . \tag{3.13}$$

As a comparison, Figure 3.2 depicts the rectangular and the Gaussian neighbourdhood functions. It can be seen that the latter is much smoother than the former.

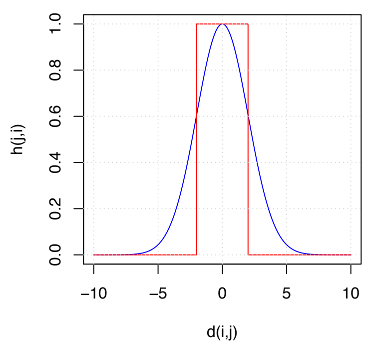


Figure 3.2: Neighbourhood function in two dimensions. The red line represents the rectangular function used before while the blue line displays the Gaussian neighbourhood function.

During the **adaptative** and final process, the low-dimensional coordinates of neurons $\boldsymbol{\gamma}$ have to be updated. A possible updating algorithm, defined in [30] and [19], could be:

$$\boldsymbol{\gamma_j} \leftarrow \boldsymbol{\gamma_j} + \alpha h_{j,i(\boldsymbol{\xi_k})}(\boldsymbol{\xi_k} - \boldsymbol{\gamma_j}) , \tag{3.14}$$

where α represents the learning rate. This updating step has to be done for each data vector $\boldsymbol{\xi_k}$. The learning rate α can be defined in a similar way to the neighbourhood width λ:

$$\alpha(n) = \alpha_0 \exp(-\frac{n}{\rho_2}) \ , \tag{3.15}$$

where n represents the n^{th} iteration. The updating algorithm thus becomes:

$$\boldsymbol{\gamma_j}(n+1) = \boldsymbol{\gamma_j}(n) + \alpha(n) h_{j,i(\boldsymbol{\xi_k})}(n)(\boldsymbol{\xi_k} - \boldsymbol{\gamma}_j(n)) \ . \tag{3.16}$$

These three processes are repeated iteratively until the convergence is achieved for all neurons. In brief, the SOM algorithm can be seen as simultaneously performing the combination of two concurrent subtasks: vector quantization and topographic representation [30].

3.4 Multi-scale Jensen-Shannon Embedding (Ms. JSE)

The Self-Organzing Map was a rather consensual NLDR method. In this last section, however, a state-of-the-art NLDR method is presented, suggested in [29] and further explained in [28]. This method is based on similarities. In the followings, a basic similarity-based NLDR method is first introduced, called Stochastic Neighbour Embedding (SNE). It is then shown how this method has been modified and enhanced to obtain the state-of-the-art NLDR method.

3.4.1 Basic idea

Ms. JSE is an NLDR method based on similarities. It attempts to project the data into lower dimensions while conserving neighbourhoods between each point. It is called *multi-scale* as it considers neighbourhoods of different sizes. It also performs a shift-invariant to fight the phenomenon of norm concentration.

3.4.2 Theory

An early method using similarities for non-linear dimensionality reduction is the Stochastic Neighbour Embedding (SNE) method and has been introduced in [20]. This method preserves similarities by centring Gaussian on each item of the data set in the high-dimensional space and by using the densities under this Gaussian to define a probabilistic distribution over all the potential neighbours of the item. Then, a dimensionality reduction is performed in order to approximate this distribution as correctly as possible when the same procedure is carried out on items in the low-dimensional space. The similarities are approximated over a neighbourhood of fixed size. The size of this neighbourhood is K from now on in this section.

The similarity (or probability) σ_{ij} that the point i chooses j as its neighbour in the high-dimensional space is:

$$\sigma_{ij} = \frac{\exp(-\delta_{ij}^2/(2\lambda_i^2))}{\sum_{k,k\neq i} -\delta_{ik}^2/(2\lambda_i^2)} \ , \tag{3.17}$$

where λ_i is either set by hand or found by a binary search for the value λ_i that makes the entropy of the distribution over K neighbours equal to $log(K) = -\sum_{j=1} N\sigma_{ij} log(\sigma_{ij})$ [20].
Likewise, the similarity (or probability) s_{ij} that the point i chooses j as its neighbour in the low-dimensional space is:

$$s_{ij} = \frac{\exp(-d_{ij}^2/2)}{\sum_{k,k\neq i} -d_{ik}^2/2} \ , \tag{3.18}$$

where the variance has been fixed to 1 without loss of generality [27]. These two similarities σ_{ij} and s_{ij} are also called softmax similarities.
The aim of the dimensionality reduction is to match σ_{ij} and s_{ij} as well as possible. SNE method enforces this by minimizing a cost function which is a sum of Kullback-Leibler divergences between σ_{ij} and s_{ij} over K neighbours for each object:

$$D_{KL}(\sigma_i||s_i) = \sum_{j=1}^{N} \sigma_{ij} \ log(\frac{\sigma_{ij}}{s_{ij}}) \ . \tag{3.19}$$

Improvement 1. An enhancement of SNE was introduced by [32] and was called t-SNE. The change was motivated because of the weakness of SNE to sometimes capture the local structure of the high-dimensional space. The main difference is the computation of similarity in low-dimensional space. t-SNE uses a Student-t distribution rather than a Gaussian to calculate the similarity s_{ij}. The formula becomes:

$$s_{ij} = \frac{(1+d_{ij}^2)^{-1}}{\sum_{k,l,k\neq l}(1+d_{kl}^2)^{-1}} \ . \tag{3.20}$$

One drawback of t-SNE, however, is the use of different formulas for similarities in high and low dimensions.

Improvement 2. Another enhancement used in [28] compared to SNE is the modification of the computation of the cost function. While the latter was a sum of Kullback-Leibler divergences in SNE, [28] uses a type 2 mixture of Kullback-Leibler divergences, introduced for NLDR methods in [26]:

$$D_{KLs2}^{\beta}(\sigma_i||s_i) = (1-\beta)D_{KL}(\sigma_i||z_i) + \beta D_{KL}(s_i||z_i) \ , \tag{3.21}$$

where $z_i = (1-\beta)\sigma_i + \beta s_i$. The parameter β balances both terms $D_{KL}(\sigma_i||z_i)$ and $D_{KL}(s_i||z_i)$. The use of this mixture of Kullback-Leibler divergences is also called Jensen-Shannon embedding (JSE). Results described in [26] between SNE and JSE show that a change in the formula of the cost function can noticeably improve the preservation of similarity during a dimensionality reduction.

Improvement 3. Until now, the first step of a similarity-based NLDR method was to fix K, the size of the neighbourhood on which the similarity calculation was based on. To capture structures and schemas on different levels and with different sizes of neighbourhoods, it could be interesting to make K vary. This is particularly important for high-dimensional data that lie on several different, but related, low-dimensional manifolds. For instance, a procedure introduced in [27] to make K vary using JSE is the following. It is called Multi-scale JSE or Ms. JSE. The method iterates over several values of K: $K_l = 2, 4, ..., 2^{L_{max}-l+1}$, with $1 \leq l \leq L \leq L_{max} \leq log(N/2)$. Multi-scale similarities become non-weighted averages of single-scale similarities. They can be respectively expressed in high and low dimensions as:

$$\sigma_{ij} = \frac{1}{L}\sum_{l=1}^{L} \sigma_{ijl} \ , \ \sigma_{ijl} = \frac{\exp(-\pi_{il}\delta_{ij}^2/2)}{\sum_{k,k\neq i}\exp(-\pi_{il}\delta_{ik}^2/2)} \tag{3.22}$$

and

$$s_{ij} = \frac{1}{L}\sum_{l=1}^{L} s_{ijl} \ , \ s_{ijl} = \frac{\exp(-p_{il}d_{ij}^2/2)}{\sum_{k,k\neq i}\exp(-p_{il}d_{ik}^2/2)} \ , \tag{3.23}$$

where π_{il} and p_{il} are precision terms to ensure entropy equalisation. As for the variance of s_{ij} in SNE, p_{il} can be fixed *a priori*. Then, a multi-scale minimization of divergences is performed, from $L = 1$ to $L = L_{max}$.

Improvement 4. An additional enhancement that was already used in some SNE, but that was absent from distance-based DR methods, is shift-invariant similarities. Indeed, a handicap of distance-based NLDR methods is their weak ability to fight the phenomenon of norm concentration. This phenomenon is illustrated in 3.3. The two left plots represent a uniform distribution in 2D and 20D. The two right plots display a gaussian distribution in 2D and 20D. It can be seen that different distributions tend to have the same norm frequency in high dimensions. Besides, norms in higher dimensions present a shift to higher values. For instance, distances for a uniform distribution in 2D are located between 0 and 4, while in 20D, the scale has shifted to be between 4 and 9. Therefore, comparing δ_{ij} and d_{ij} seems like comparing apples and pears. A possible solution could be to add a negative shift to high-dimensional distances. The shift should be at the same time high enough to thwart the norm concentration and low enough to ensure the shifted distances remain positive. A possible value for the shift could then be:

$$S_i = \min_{k, k \neq i} \delta_{ik} \quad . \tag{3.24}$$

The formula of similarity in the high-dimensional space becomes:

$$\sigma_{ij} = \frac{\exp((S_i^2 - \delta_{ij}^2)/(2\lambda_i^2)}{\sum_{k,k \neq i}(S_i^2 - \delta_{ik}^2)/(2\lambda_i^2)} \quad . \tag{3.25}$$

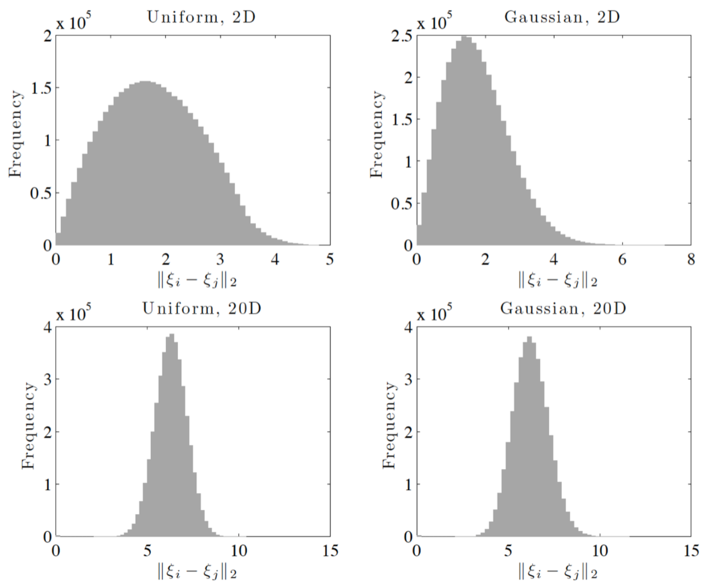


Figure 3.3: Phenomenon of norm concentration. The two left plots represent a uniform distribution in 2D and 20D. The two right plots display a gaussian distribution in 2D and 20D [27].

The basic theoretical framework of SNE and these four improvements characterize the NLDR method described in [28]. The full name of this method could be a multi-scale similarity-based dimensionality reduction, with shift-invariant softmax-normalized similarities and sums of alpha-beta-divergences as a cost function. In this work, it will be called Ms. JSE.

3.5 Conclusion

In this chapter, the theoretical backgrounds of three DR methods have been introduced: the Principal Component Analysis (PCA), the Self-Organizing Map (SOM) and the Multi-scale Jensen-Shannon Embedding (Ms. JSE). While the first one is based on a linear model, the last two ones are more complex non-linear techniques. The three methods presented distinct interesting characteristics.

PCA is one of the most used DR methods in the field of university rankings and is therefore appealing to carry out to compare its results with conclusions of other works in this domain. Furthermore, its linearity makes easier to discuss the new axes.

On the other hand, the particularity of SOM is that it focuses more on topology preservation, so that it tries to preserve neighbourhood relationships between subregions of the manifold. Besides, defining *a priori* a lattice with fewer points than the number of schools could allow to conduct an analysis on vector quantization, which could be a good starting point to find clusters among schools.

Finally, Ms. JSE is a NLDR method based on similarity preservation. The interest of this technique is that it is an advanced method and it is expected to show a quite good level of performance when projecting FT data into lower dimensions.

4 Results

Abstract This chapter is dedicated to the application of DR methods seen in Chapter 3 on FT data. First, a Principal Component Analysis (PCA) is considered. An analysis of the linear combinations forming the new axes as well as a dimensionality reduction as such, called a reversed PCA, are performed. Then, the Self-Organizing Map (SOM) is studied. A two-dimensional mapping and a clustering on the new map are conducted. Finally, the Multi-scale Jensen-Shannon Embedding (Ms. JSE) is examined. A two- and three-dimensional embeddings are carried out. For each method, a quality assessment is performed to gauge the efficiency of the DR method on FT data. After that, a comparison between the three DR methods is realized and a projection into one dimension closes this chapter.

4.1 Introduction

This chapter is devoted to the application of DR methods on FT data. The three methods explained in the previous chapter are considered. It has to be kept in mind that the main goal of concentrating on DR methods was to be able to visualize the FT data in two or three dimensions, so that it would be easier to grasp the connections between schools and if the FT MiM ranking truthfully translate them.
Before analyzing results of the projections, it could be interesting before conducting the dimensionality reductions to consider how the projections are evaluated. Different techniques exist and two of them are used in this work: the Shepard diagram and the NX-scores.
After these two introductory parts, results using the DR methods are displayed. It has to be remembered that these techniques presented different aspects even though they shared the same goal: projecting data from a high-dimensional space to a lower-dimensional embedding. Because of their different characteristics, results of each method are explained with a particular structure.

The first considered DR method is the Principal Component Analysis (PCA). It has been seen that it is a method based on a linear model, that attempts to create from N variables of a data set, N linearly uncorrelated variables, via linear combinations of the old variables while maximizing the variance of each axe after projection. These new variables are called principal components. From this, a dimensionality embedding can be implemented by only taking into account a couple of principal components. A dimensionality reduction using principal components is also called *reversed PCA*.
Two PCA are conducted in this chapter, one considering the FT data as it is and another one also taking into account the FT weights. The two PCA are respectively called *non-weighted PCA* and *weighted PCA* in this chapter. First, the analysis of linear combinations defining the principal components of both PCA are presented and compared. Then, projections into lower dimensions, taking only into account principal components from the weighted PCA, are performed. Results of projections into two and three dimensions are introduced and their quality assessed.

The second illustrated DR method is the Self-Organizing Map (SOM). It focuses on topology preservation. The interest of SOM is that it creates a discrete mapping model in low dimensions and attempts to map this model in high dimensions while linking schools to nodes of the mapping model. By creating a map with fewer nodes than the number of schools, several universities will be linked to the same node so that it will be possible to compare the resemblance between schools.
The first part of the analysis of the results using SOM on the FT data set is thus devoted to inspect the schools grouped in a same node and examine their correspondence. Then, a quality assessment of the mapping model is conducted. After that, the self-organizing map is clustered adopting a k-means clustering technique and the results are discussed. The last part of this section is dedicated to the analysis of the scores for each criteria of neurons constituting the map in high dimensions. This could explain why some schools that do not look similar at first sight are gathered into one node.

The last presented DR method is the Multi-scale Jensen-Shannon Embedding (Ms. JSE). It is based on similarities and attempts to preserve local neighbourhoods of schools when projecting them to lower dimensions. A two- and three-dimensional embeddings are conducted and their quality assessed.

After having presented the three DR methods and their result on the FT data set, they are examined together. A first comparison is done, only focusing on the results of projections. Then, a second comparison is performed, interested more in the quality assessments of the different methods. From this, a DR technique is found as preserving the best the information embedded in the FT data set while projecting it into lower dimensions.

The last part of this chapter is devoted to a more unusual analysis. DR methods were used in this chapter to visualize the FT data in two or three dimensions, so that it would be easier to enhance the connections that exist between schools and to highlight possible local neighbourhoods of universities. The logic could be pushed further and one could try to conduct a one-dimensional embedding of the FT data set.
For that, the DR method that performed the best in the previous quality comparison is used to create a one-dimensional projection of the data. Obviously, a comparison between this embedding and the FT MiM ranking is conducted.

4.2 Evaluation of dimensionality reduction

Before reducing dimensionality, it could be interesting to design tools to assess the quality of the reduction, to be able to compare the methods that are studied in this work. These methods are of a great variety: from linear ones like PCA to non-linear ones like SOM or Ms. JSE, they will all focus on different characteristics of the high-dimensional space to perform a dimensionality embedding. The difficulty of a quality assessment is thus to construct fair assessment tools. The solution that is adopted in this report is the use of two quality indicators: the Shepard diagram and the NX-scores. The former is based on a scatter plot comparing the high- and low-dimensional distances while the latter focuses on quantifying the topology preservation. The Shepard diagram was traditionally used to assess the quality of a dimensionality reduction. The NX-scores is more recent and is more and more adopted as a quality indicator for NLDR methods. In particular, it is used in [28] where the NLDR method Ms. JSE is introduced. It also comes as a good alternative in [5].

4.2.1 Shepard diagram

The Shepard diagram is a simple quality indicator. An example of it can be found in Figure 4.1. It compares the distances in high- and low-dimensional spaces by constructing a scatter plot where each point is associated to a pairwise distance and has as coordinates its value in high and low dimensions. Clearly, a good distance-preserving method will have a Shepard diagram that forms a straight line, meaning that high-dimensional distances are all reduced in low dimension in the same way.

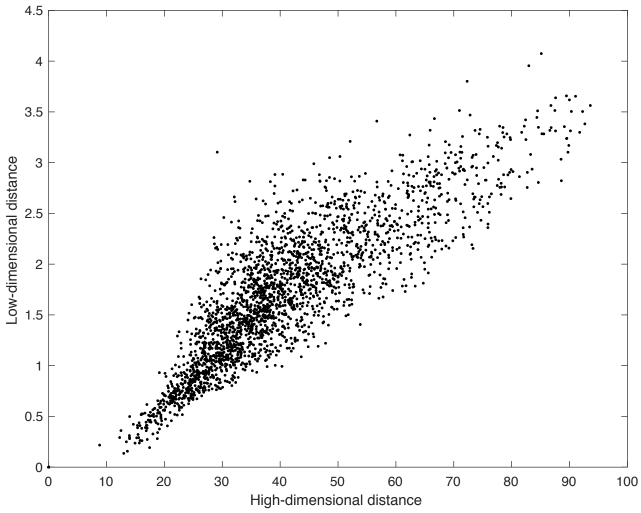


Figure 4.1: Example of a Shepard diagram.

4.2.2 NX-scores

Instead of manipulating distances, the NX-scores[1] use the notions of ranks and neighbourhoods. A rank between two points i and j is the number of points that are closer to i than j. In the high-dimensional space, it can be written as follows:

$$\rho_{ij} = |\{k : \delta_{ik} < \delta_{ij} \text{ or } (\delta_{ik} < \delta_{ij} \text{ and } k < l)\}| \ , \tag{4.1}$$

and in the low-dimensional space:

$$r_{ij} = |\{k : d_{ik} < d_{ij} \text{ or } (d_{ik} < d_{ij} \text{ and } k < l)\}| \ . \tag{4.2}$$

On the other hand, a neighbourhood of size K centered on point i is defined as the set of indices of points whose rank with i is lower or equal to K. Mathematically speaking, it can be seen in high dimensions as:

$$\nu_i^K = \{j : 1 \leq \rho_{ij} \leq K\} \ , \tag{4.3}$$

and in low dimensions as:

$$n_i^K = \{j : 1 \leq r_{ij} \leq K\} \ . \tag{4.4}$$

From these definitions, a multi-scale quality assessment could be defined as:

$$Q_{NK}(K) = \sum_{i=1}^{N} \frac{|\nu_i^K \cap n_i^K|}{KN} \ , \tag{4.5}$$

where Q stands for quality, K is the maximal size of the neighbourhood and N the number of data points. This quality assessment can be seen as an average agreement of the K-ary

[1]NX standing for *intrusion and extrusion.*

neighbourhoods. An example of Q^K_{NX} can be found in Figure 4.2. The x-axis represents K, the maximal size of the neighbourhood while the y-axis represents the percentage of overlapping between the neighbourhoods in high- and low-dimensions. It can be seen in this figure that the red line displays a random embedding. As a good DR method should never go down a random guessing, one could seem weird that half of the plot is pointless. A possible solution is to use relative quality assessment instead, so that the random embedding is shifted to the horizontal axis. Figure 4.3 depicts the transformation. The new quality score becomes:

$$R^K_{NK} = \frac{(N-1)Q^K_{NK} - K}{N-1-K} \quad , \tag{4.6}$$

where R stands for relative. R^K_{NK} can thus be seen as a relative quality between a perfect embedding and a random one. Figure 4.4 displays possible examples of R^K_{NK}. It can be noticed that the random embedding has now merged with the x-axis.
In this chapter, the term *NX-scores* is used to designate R^K_{NK}.

As a last remark on NX-scores, it should be noted that a logarithmic scale is adopted for the x-axis in the future quality assessments. Indeed, the similarity when a small neighbourhood is considered seems more important in this work, as the local neighbourhoods among schools are studied.

4.3 Principal Component Analysis (PCA)

Before presenting results of the dimensionality reduction, the linear combinations of new variables created by the PCA are first introduced. Indeed, as it was indicated in the last chapter, one of the main advantages of PCA is that its design of new variables, called principal components, via linear combinations of old ones allows to understand the meaning of these principal components.

Two different PCAs are conducted in this section. One takes into account the FT weights while the other one does not consider them. Surely, ranging from 1% to 20%, these weights could strongly affect the results of PCA and it could be interesting to compare them. In what follows, the FT data set without taking into account the weights is first used, then a second analysis considering the FT weights is conducted. A comparison of the two studies is done before moving on the reversed PCA.

An important remark that has to be noted before describing the results is that while the FT criteria are constructed so that best-performing schools receive high scores for these criteria, it is not the same for the principal components. Having a low score along a principal component does not mean that the school is mediocre. So, in order to analyze the principal components in the following work, the loading of criterion for a specific principal component will often be refered to. For instance, for a criterion i and a principal component j, the value of interest is:

$$l_{ij} = \frac{|a_{ij}|}{\sum_{k=1}^{n} |a_{kj}|} \quad , \tag{4.7}$$

where n is the number of criteria. The percentage of this value will also often be used. This value is also called the absolute coefficient of the criterion i on the principal component j in the following PCAs.

4.3.1 Non-weighted PCA

First of all, a principal component analysis was conducted on the FT data set without taking into account the FT weights. However, the data set were normalized and the values of ranked

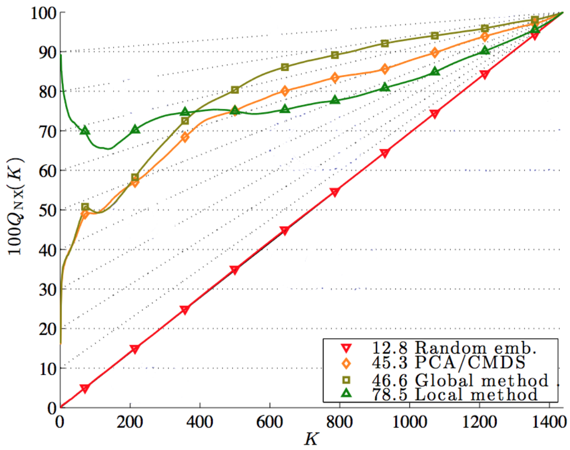


Figure 4.2: Examples of Q_{NX}^K [27].

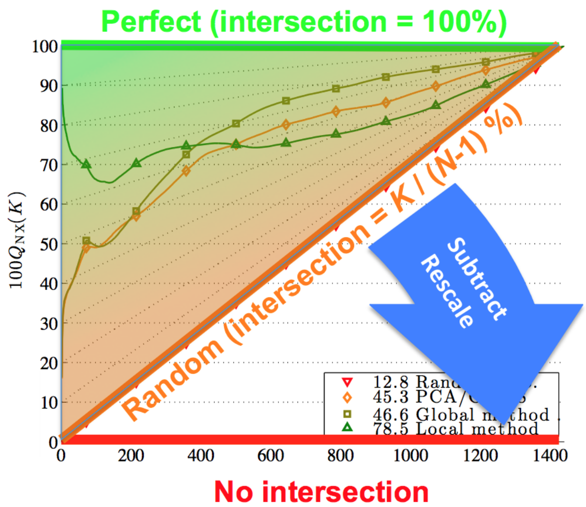


Figure 4.3: Transformation from Q_{NX}^K to R_{NX}^K [27].

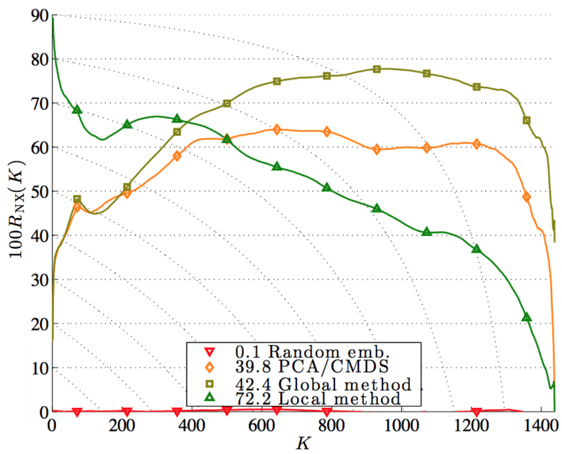


Figure 4.4: Examples of R_{NX}^K [27].

criteria were modified as explained in Section 2.2. Table 4.1 summarizes the sixteen principal components and their proportional explanatory contributions applied to the FT MiM ranking. Figure 4.5 depicts the variances of the principal components.
It can be seen that the first four principal components present a variance rather higher and more distant than from the others. They explain 62.6% of the variance in the FT data set. Nevertheless, the proportion of variance explained by the first principal components remains quite low compared to other analyses conducted in other research papers. For instance, the first principal component gathered more than half of variance in [44], where a PCA has been performed on the rankings from the U.S. News & World Report.

Table 4.2 depicts the coefficients for the first four principal components. The interest here is to evaluate if the criteria that form these first principal components are the ones that present the higher FT weights or not. The table reveals that the loading of the criteria *Weighted salary* (1), *Aims achieved* (4), *Placement success* (5) and *Women faculty* (7) have the highest absolute coefficient values: respectively 13.3%, 11.6%, 11.5% and 10.0%. The first criterion has the highest FT weight (20%) while the others have an average one (5%).
The second principal component present high absolute coefficients for the criteria *Women students* (8), *International faculty* (10), *International students* (11) and *International mobility* (13) of respectively 11.9%, 14.3%, 12.6% and 14.5%.
The third and fourth principal components show high absolute coefficients respectively for the criteria *International students* (11), *International course experience* (14) and *Languages* (15), and for the criteria *Value for money* (2), *Careers* (3) and *Faculty with doctorates* (16).

It can be seen from this description of the coefficients for the first four principal components that almost all the criteria have at least one high absolute coefficient for one principal component. Only three criteria were not listed here: *Employed at three months* (6), *Women board* (9) and *International board* (12). It is interesting to note that these two last criteria have the lowest FT weight (1%).
Besides, the fact that the criteria *Weighted salary* (1), *Aims achieved* (4), *Placement success* (5) and *Women faculty* (7) present the highest absolute coefficients for the first principal component can be partly explained their high pairwise correlations, as seen in Table 2.2.

4.3.2 Weighted PCA

A second principal component analysis was conducted on the FT data set, taking into account the FT weights this time. However, the data set were normalized and the values of ranked criteria were modified as explained in Section 2.2. Table 4.3 summarizes the sixteen principal components and their proportional explanatory contributions applied to the FT MiM ranking. Figure 4.6 depicts the variances of the principal components.
The first element that really stands out is the variance of the first principal component, compared to the others. It represents more than half of the explained proportion of variance. Overall, the following three principal components seem interesting to look at because they describe respectively 15.8%, 10.2% and 8.3%, while the fifth principal component explains less than 5% of the variance of the FT data set.

The first four principal components, which represent 84.8% of the variance in the FT MiM ranking, are summarized in Table 4.4. The coefficients of criteria for the first principal component is of particular importance as this component surpasses all the others. The table reveals that the loading of the criterion *Weighted salary* (1) has the highest absolute coefficient value. It represents 41.0% of the sum of the absolute coefficients. The second criterion with the highest absolute coefficient is *Careers* (3) and is worth only 9.8% of the total absolute coefficients. The criteria *Aims achieved* (4), *Placement success* (5) and *Women board* (7) are respectively ranked

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	1.9170	1.6957	1.4557	1.1621	1.0138	0.9835	0.9132	0.7961
Proportion of Variance	0.2297	0.1797	0.1324	0.0844	0.0642	0.0605	0.0521	0.0396
Cumulative Proportion	0.2297	0.4094	0.5418	0.6263	0.6905	0.7510	0.8031	0.8427
	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	0.7364	0.6723	0.6182	0.5788	0.5348	0.4833	0.3917	0.3643
Proportion of Variance	0.0339	0.0283	0.0239	0.0209	0.0179	0.0146	0.0096	0.0083
Cumulative Proportion	0.8766	0.9048	0.9287	0.9496	0.9675	0.9821	0.9917	1.0000

Table 4.1: Standard deviation and proportion of variance in the data explained by each principal component, when the FT weights are not taken into account. The principal components are ordered according to their respective proportion of variance.

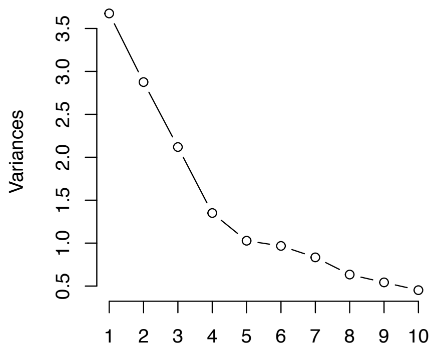


Figure 4.5: Variances of the ten first principal components, when the FT weights are not taken into account.

	Criteria	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
1	Weighted salary	-0.46	0.10	-0.05	0.07	13.3	3.3	1.6	2.5
2	Value for money	-0.21	0.15	0.01	-0.54	6.0	4.8	0.3	18.9
3	Careers	-0.19	-0.01	0.17	0.64	5.6	0.3	5.6	22.7
4	Aims achieved	-0.40	-0.05	-0.21	0.00	11.6	1.4	6.9	0.0
5	Placement success	-0.40	0.08	0.01	0.23	11.5	2.6	0.4	8.0
6	Employed at three months	-0.24	0.10	0.15	-0.06	6.8	3.2	5.1	2.2
7	Women faculty	0.34	-0.13	0.26	0.19	10.0	4.1	8.7	6.6
8	Women students	0.10	-0.38	0.06	-0.21	3.0	11.9	2.1	7.6
9	Women board	0.17	-0.18	-0.13	-0.07	5.0	5.7	4.4	2.6
10	International faculty	-0.11	-0.46	-0.16	-0.06	3.3	14.3	5.3	2.1
11	International students	-0.00	-0.40	-0.41	0.06	0.1	12.6	13.5	2.1
12	International board	-0.13	-0.29	0.18	0.03	3.8	9.2	6.2	1.1
13	International mobility	-0.18	-0.47	-0.04	0.11	5.2	14.5	1.3	3.7
14	International course experience	-0.17	-0.23	0.52	-0.09	4.8	7.0	17.3	3.1
15	Languages	-0.07	-0.13	0.56	-0.18	2.0	4.1	18.8	6.3
16	Faculty with doctorates	-0.28	-0.03	-0.07	-0.30	8.0	1.0	2.4	10.6

Table 4.2: The first four columns present coefficients of each criterion for the first four principal components, when the FT weights have not been taken into account to conduct the PCA. The last four columns display the percentage of the contribution of each coefficient to the sum of the absolute coefficients, defined as l_{ij} in Equation 4.7.

third, fourth and fifth regarding their absolute coefficient value for this principal component. This is not totally surprising as these three criteria were the ones that were the most correlated with the criterion *Weighted salary* (1), as seen in Table 2.2.

The second principal component presents a high percentage of the absolute coefficients for the criteria *International mobility* (13) and *International course experience* (14): respectively 26.4% and 20.4%. It is not unexpected that these two criteria display a high loading on one of the first principal components as they have the second highest FT weight (10%), with the criterion *Careers* (3).
The third and fourth principal components show a non-negligible percentage of the absolute coefficients for the criteria *Careers* (3), *International mobility* (13) and *International course experience* (14). The only surprise is that the third principal component presents a high loading of 10.4% on the criterion *International students* (11), even though this criterion has a lower FT weight than the others. This can be partly explained by the fact that it is seriously correlated with the criterion *International mobility* (13), as can be noticed in Table 2.2.

4.3.3 Comparison

The results of weighted and non-weighted PCA are quite distinctive. While thirteen out of the sixteen criteria presented a high absolute coefficient for one of the first four principal components for the non-weighted PCA, only eight of them did for the weighted PCA. Out of these eight criteria, four have a high FT weight (*Weighted salary* (1), *Careers* (3), *International mobility* (13) and *International course experience* (14)) while the four other criteria presented a high pairwise correlation with at least one of the criteria with a high FT weight.
The criteria that were underlined in the non-weighted PCA but were missing in the weighted PCA are the ones that are not heavily correlated with high-weighted criteria and that have average to low FT weights. Of course, it was obvious that the two analyses would present some discrepancies. However, the fact that no criterion that did well in the non-weighted PCA but that has average FT weight, was present in the weighted PCA really demonstrates that the arbitrary weights of the Financial Times shape the ranking.

A last detail that could be highlighted is the fact that the first four principal components in the non-weighted PCA represented 62.6% of the variance in the FT data set, while they described in the weighted PCA 84.8% of the variance in the weighted data set. This again shows that the application of weights can help few criteria to discriminate the FT data more efficiently.

4.3.4 Reversed PCA

A possible application to use results of PCA is to perform a principal component regression analysis, also called reversed PCA. The objective is to only keep a subset of the principal components, thus conducting a dimensionality reduction based on a linear model. The number of principal components to conserve depends on the proportion of variance one wants to preserve. Maintaining a lot of principal components could allow us to keep a high percentage of variance present in the FT data set, but would also reduce effects of the dimensionality reduction.

In the previous section, a detailed analysis was performed on the first four principal components. The criteria that drive these components were presented but it could be interesting to determine if these components manage to well approximate the FT MiM ranking. Using the data set redefined by the non-weighted and weighted reversed PCAs, the reconstruction of the ranking gives a Kendall correlation of respectively 0.8137 and 0.8443 with the FT MiM ranking. This is significantly lower than the correlation obtained when using the entire FT data

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	21.8184	12.2030	9.7896	8.8503	5.6174	4.6218	4.3633	4.2906
Proportion of Variance	0.5053	0.1581	0.1017	0.0832	0.0335	0.0227	0.0202	0.0195
Cumulative Proportion	0.5053	0.6634	0.7652	0.8483	0.8818	0.9045	0.9247	0.9443
	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	3.7280	3.3854	3.2681	2.7459	2.3460	1.5006	0.9042	0.6040
Proportion of Variance	0.0147	0.0122	0.0113	0.0080	0.0058	0.0024	0.0009	0.0004
Cumulative Proportion	0.9590	0.9712	0.9825	0.9905	0.9963	0.9987	0.9996	1.0000

Table 4.3: Standard deviation and proportion of variance in the data explained by each principal component, when the FT weights are taken into account. The principal components are ordered according to their respective proportion of variance.

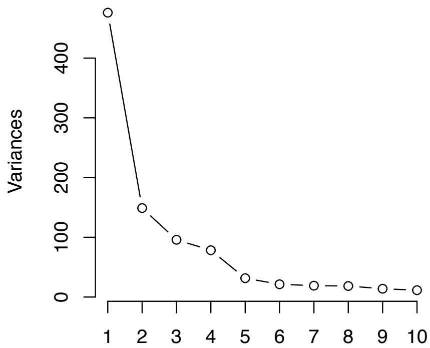


Figure 4.6: Variances of the ten first principal components, when the FT weights are taken into account.

	Criteria	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
1	Weighted salary	-0.91	-0.18	-0.09	0.04	41.0	6.8	3.3	1.6
2	Value for money	-0.08	-0.10	-0.11	0.23	3.6	3.7	3.8	9.0
3	Careers	-0.22	0.18	0.69	-0.55	9.8	7.0	24.2	21.5
4	Aims achieved	-0.16	0.02	-0.15	-0.04	7.3	0.7	5.1	1.5
5	Placement success	-0.16	-0.02	0.01	-0.03	7.3	0.9	0.2	1.2
6	Employed at three months	-0.08	-0.04	0.07	0.12	3.4	1.6	2.5	4.5
7	Women faculty	0.14	0.10	0.17	-0.03	6.2	3.7	5.8	1.1
8	Women students	0.05	0.22	-0.09	0.09	2.1	8.2	3.1	3.3
9	Women board	0.01	0.01	-0.01	-0.01	0.5	0.4	0.3	0.3
10	International faculty	-0.02	0.24	-0.23	-0.08	1.0	9.2	8.0	3.3
11	International students	0.02	0.18	-0.30	-0.26	0.9	6.7	10.4	10.3
12	International board	-0.01	0.08	0.01	0.01	0.4	2.9	0.3	0.3
13	International mobility	-0.12	0.69	-0.34	-0.26	5.6	26.4	11.7	10.0
14	International course experience	-0.13	0.54	0.36	0.68	5.7	20.4	12.5	26.3
15	Languages	-0.00	0.03	0.03	0.06	0.1	1.3	1.1	2.4
16	Faculty with doctorates	-0.11	0.00	-0.22	0.09	5.1	0.1	7.7	3.4

Table 4.4: The first four columns present coefficients of each criterion for the first four principal components, when the FT weights have been taken into account to conduct the PCA. The last four columns display the percentage of the contribution of each coefficient to the sum of the absolute coefficients, defined as l_{ij} in Equation 4.7.

set (0.9528) but this remains quite good when considering that only four principal components were considered.

More generally, Figure 4.7 depicts the Kendall correlations of the non-weighted and weighted reversed PCAs when the number of principal components taken into account varies. It can be seen that the latter is invariably well ahead the former, except when all the principal components are considered because the two lines merge. As already said before, the fact that fewer principal components in the weighted reversed PCA can simulate quite well the entire FT data set could be explained by the application of weights that help criteria to discriminate the FT data more efficiently.
As a last remark on this plot, it could be noticed that the weighted reversed PCA presents a Kendall correlation approaching 0.85 when only two principal components are taken into account. This is quite impressive.

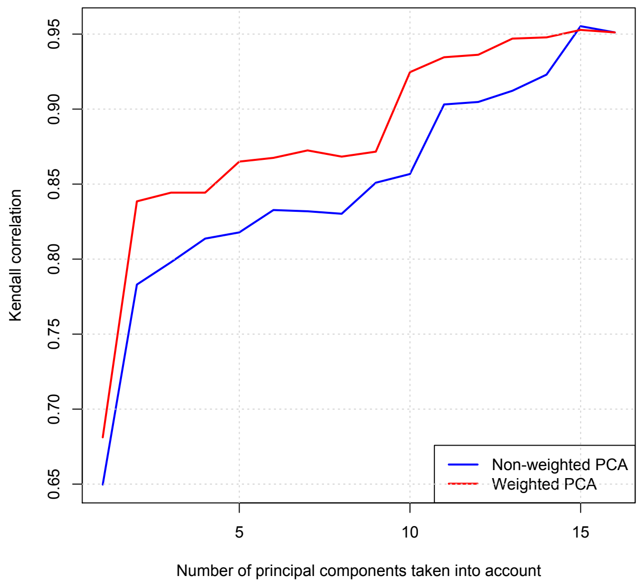


Figure 4.7: Kendall correlations of the non-weighted and weighted reversed PCAs when the number of principal components considered varies.

Before terminating the presentation of results of reversed PCA, it could be interesting to present the projection in 2D or 3D. Indeed, a two-dimensional embedding has the advantage to be able to be represented on a plot. A three-dimensional embedding could be interesting to look at to assess the discrepancies in quality coming from adding one dimension to the projection.
In the followings, only the weighted reversed PCA will be considered as it has been shown that it obtained the highest correlation with the FT MiM ranking.

4.3.4.1 Two-dimensional embedding

Results. A projection into two dimensions using the weighted reversed PCA can be found in Figure 4.8 and 4.9. In the first plot, the schools are coloured according to their FT rank while in the second figure, the schools are coloured according to the criterion *Weighted salary* (1)[2]. Each school is numbered according to its FT rank. Each circle encompasses ten schools and are centered on the University of St Gallen (school 1). It can be seen that the innermost circle is the widest, suggesting that the best-performing schools are also the most dissimilar among each other. Conversely, it can be noticed that slices containing schools that are in the middle of the

[2]This criterion is chosen as it represents the highest FT weight.

ranking are the thinnest, meaning that this region includes a lot of schools that are located at a similar distance from the University of St Gallen (school 1).
Besides, despite the fact that the reconstruction of the FT MiM ranking in Section 2.2 revealed that the top ten schools were well ahead from the other ones, it can be observed that the WHU Beisheim (school 4) is situated quite far from the University of St Gallen (school 1). It is interesting to note that the WHU Beisheim (school 4) was the one among the top ten schools presenting the highest variance during the robustness analysis, in Section 2.3.
The last remark that could be done is that three schools are very distant from the other ones: the Indian Institute of Management, Calcutta (school 13), the Indian Institute of Management, Ahmedabad (school 17) and the Shanghai Jiao Tong University: Antai (school 44). The first two schools are close to each other. It is worth noting that they are the only two Indian universities in this ranking, suggesting that the localization of schools could play a role in the ranking, even though Figure 1.1 implied the contrary for European schools. Furthermore, the Shanghai Jiao Tong University: Antai (school 44) is also not European: it is located in China. However, it has to be noted that the other Chinese school in this ranking, the Tongji University School of Economics and Management (school 65), is projected among the main group of schools.

Figure 4.9 depicts the schools coloured according to the criterion *Weighted salary* (1). It can be seen that the schools are tremendously discriminated according to this criterion. It has been explained when presenting the linear combinations of principal components that the first principal component displayed a high absolute coefficient for the criterion *Weighted salary* (1). However, it can be noticed that the second principal component also plays a role in the localization of schools with high score for this criterion. Indeed, the schools Rotterdam School of Management, Erasmus University (school 8) and Mannheim Business School (school 18) have more or less the same value for the first principal component. Nonetheless, they present very different score for the criterion *Weighted salary* (1).
As a final commentary on this plot, this figure is a good example to demonstrate the fact that it does not need to do a high score for each principal component to be well ranked, as opposed to the FT criteria. Indeed, it can be seen that the schools with the highest score for the *Weighted salary* (1) tend to have a very negative score for the first principal component.

Quality Assessment. Now that the results of the projection in two dimensions have been analyzed, an evaluation of the quality assessment should be conducted. Figure 4.10 shows the comparison of the distances between points in high and low dimensions while Figure 4.11 depicts the NX-scores. It can be seen that a structure of a straight line can be guessed from the cloud of points, even though the width of this line is large. However, it can also be observed in Figure 4.11 that the similarity between the projection and the FT data set is quite low, especially when the size of the considered neighbourhood is low. When its size is equal to one, the similarity is as small as 25%. Nonetheless, the similarity increases when the size of the neighbourhood grows, to reach 90% when the neighbourhood attains 55 schools.
One could explain the mediocre similarity of weighted reversed PCA by the fact that the first two principal components only account for 66.3% of the total variance.

4.3.4.2 Three-dimensional embedding

The second projection performed using the weighted reversed PCA is a three-dimensional embedding. Indeed, it could be worth taking a look at a projection into three dimensions to assess the differences in quality coming from adding one dimension to the projection. In this part, only a plot presenting the comparison of the distances in the high- and low-dimensional spaces (Figure 4.12) and a plot depicting the NX-scores (Figure 4.13) are shown, for a three-dimensional embedding could hardly be understandable when presented on a report. Overall, Figure 4.12

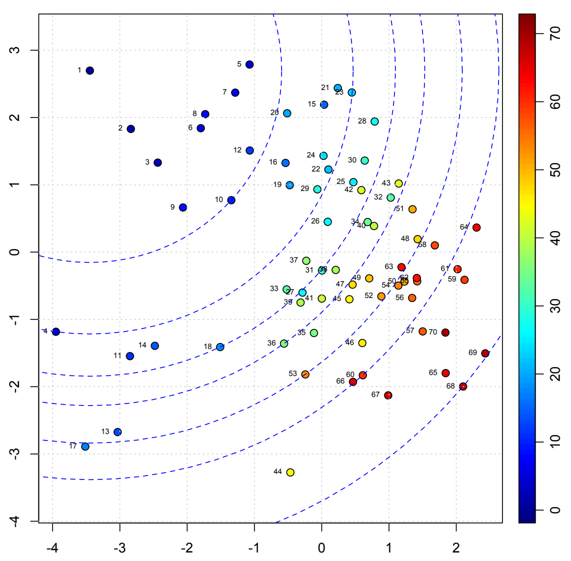


Figure 4.8: Projection in 2D of the schools following the weighted reversed PCA. The x- and y-axes represent respectively the first and second principal components. The schools are coloured according to their FT MiM ranking.

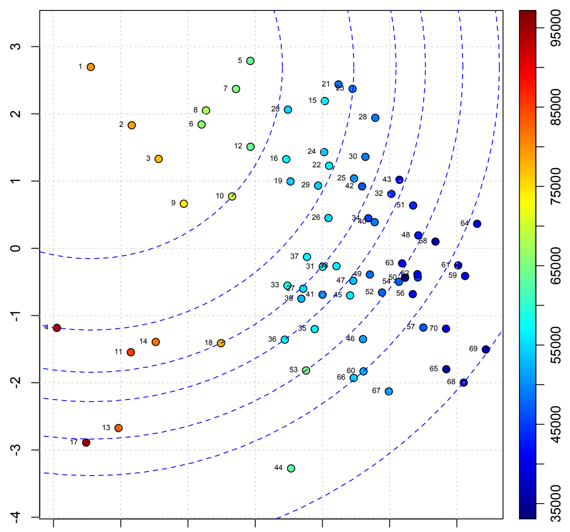


Figure 4.9: Projection in 2D of the schools following the weighted reversed PCA. The x- and y-axes represent respectively the first and second principal components. The schools are coloured according to their score for the criterion *Weighted salary* (1).

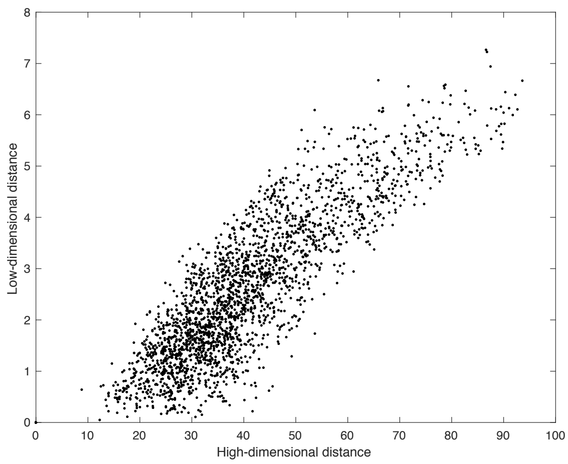


Figure 4.10: Shepard diagram of the projection in 2D using weighted reversed PCA.

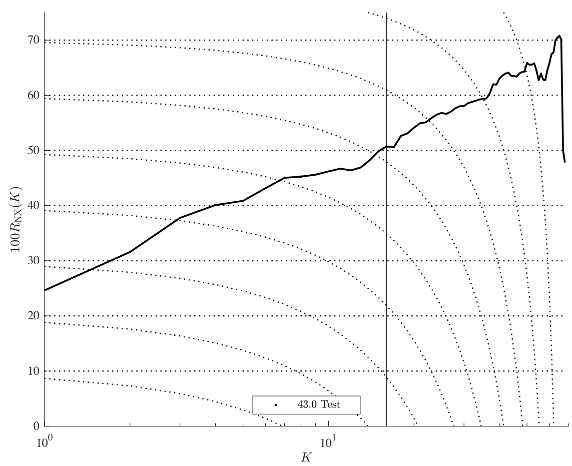


Figure 4.11: NX-Scores of projection of schools in 2D using weighted reversed PCA.

presents a less blurred cloud of points when the distances in high dimensions are low. For larger distances, the difference with the projection into two dimensions is less clear. However, the contrast between two- and three-dimensional embeddings in Figure 4.13 is much sharper. Even though the similarity is still far from good, beginning at around 35% when the size of neighbourhood is equal to one, it is much better than for the projection into two dimensions. This can be explained by the fact that the first three principal components represent 76.5% of the total variance, compared to 66.3% for the first two ones.

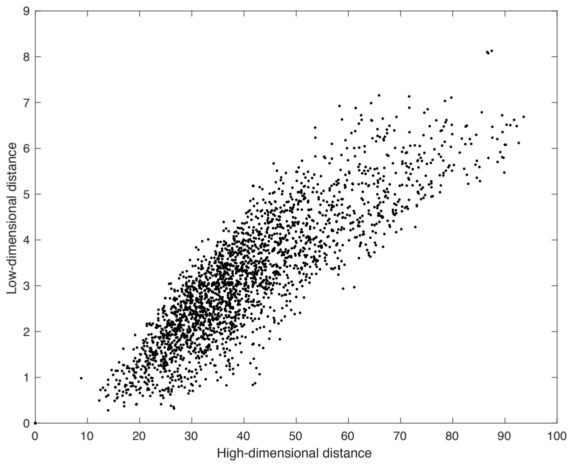


Figure 4.12: Shepard diagram of the projection in 3D using weighted reversed PCA.

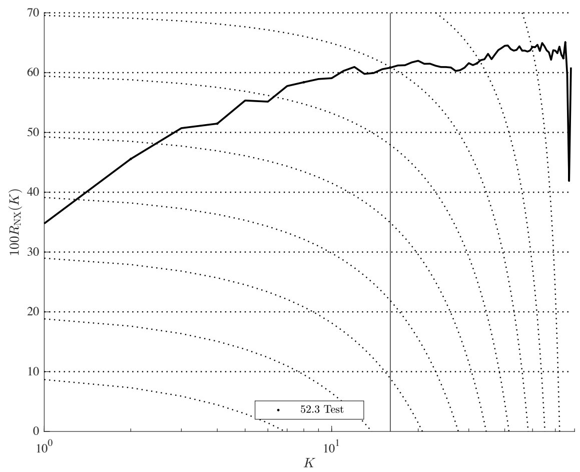


Figure 4.13: NX-Scores of projection of schools in 3D using weighted reversed PCA.

4.4 Self-Organizing Map (SOM)

4.4.1 Results

A self-organizing map has been trained with 25 nodes using the FT data set and taking into account the weights for each criterion. The nodes are distributed into a 5×5 hexagonal grid. The distribution of the schools among the SOM nodes are shown in figure 4.14. Being located in the North-West corner or the South-East one does not make any differences. It is the neighbouring SOM nodes that are important to look at. It can be seen that the best schools are positioned in the West-Northwest part of the grid. However, the least rated schools are more scattered throughout the grid. Some are located in the East-Southeast part of the grid while schools NHH (school 60), Nyenrode Business Universiteit (school 66) and BI Norwegian Business School (school 67) are more located in the Northeast part of the grid.
Furthermore, some nodes present curious association of schools. It is the case for a node in the North part of the grid that contains both the Mannheim Business School (school 18) and University of Cologne, Faculty of Management (school 53), which are quite distant from each other in the FT MiM ranking. Also, it should be noted that the top three schools in the FT MiM ranking are all located in the same node.

Another aspect of a self-organizing map that is interesting to look at is the distance between the nodes. It is shown in Figure 4.15, which displays the sum of the distances with all immediate neighbouring SOM nodes. It can be seen that the SOM nodes in the North-Northwest part of the grid are more far apart from each other than the other nodes. Especially, the node including the schools Indian Institute of Management, Calcutta (school 13) and Indian Institute of Management, Ahmedabad (school 17) is the most distant from its neighbours.

4.4.2 Quality assessment

For the self-organizing map, the quality assessment is a little different than for reversed PCA and Ms. JSE. Instead of analyzing the projection of the seventy schools from high to low dimensions, it is the projection of the twenty-five neurons forming the map that are considered. Indeed, only focusing on the seventy schools would not have been suitable as several schools are mapped to the same node in low dimensions.
Furthermore, the fact that nodes of the self-organizing map in low dimensions have a discrete distribution made the results of NX-scores irrelevant for this projection. Therefore, only the

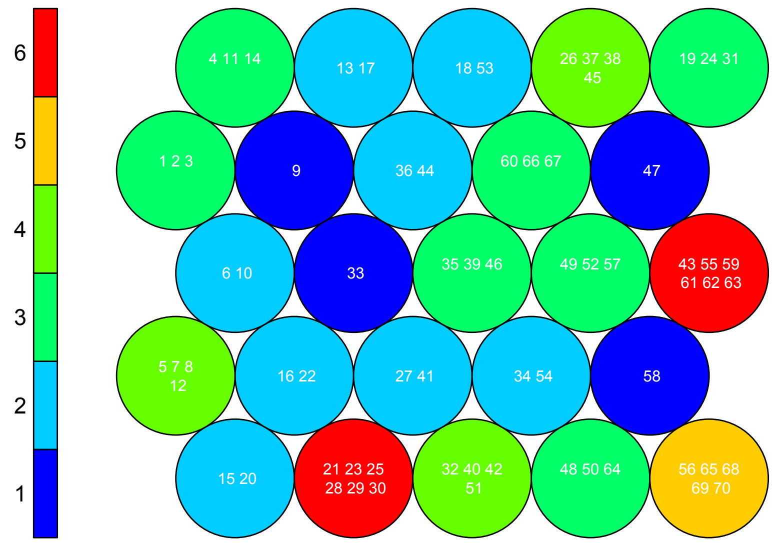


Figure 4.14: Self-organizing map based on the FT data set. The ranks of schools in the FT MiM ranking are written in their respective node. The nodes are coloured according to the number of schools it contains.

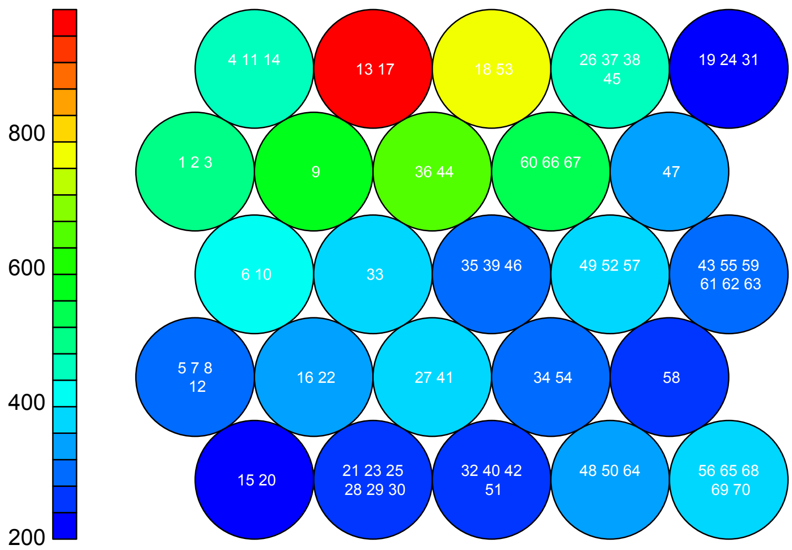


Figure 4.15: Self-organizing map based on the FT data set. The rank of schools in the FT MiM ranking are written in their respective node. The nodes are coloured according to the sum of the distances to all the immediate neighbouring nodes.

Shepard diagram is displayed here. It can be found in Figure 4.16. The different plateaus due to the discrete distribution of SOM nodes can clearly be noticed. Overall, it can be seen that a plateau representing large distances in low dimensions covers a larger range of distances in high dimensions than a plateau describing lower distances.

4.4.3 Clustering

It could be appealing to try to divide SOM nodes into clusters. That way, large groups of schools that share similar properties could be created. It could help us understand more further the connections between the schools and the resulting FT ranking. A k-means clustering has been used for this purpose. Figure 4.17 shows the sum of squares in each group relative to the

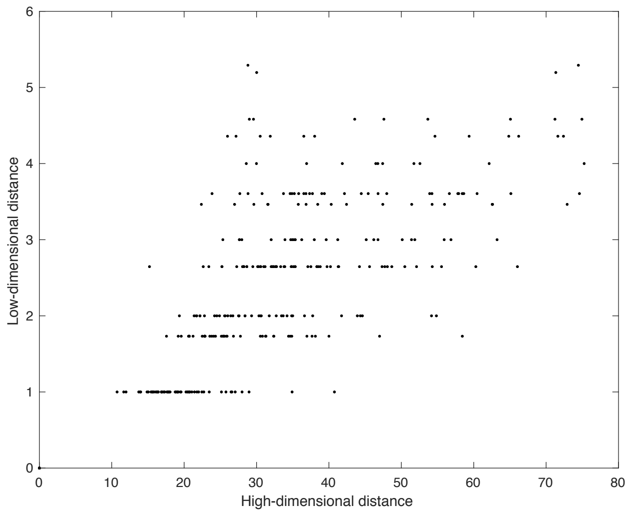


Figure 4.16: Shepard diagram of the projection in 2D using Self-Organizing Map.

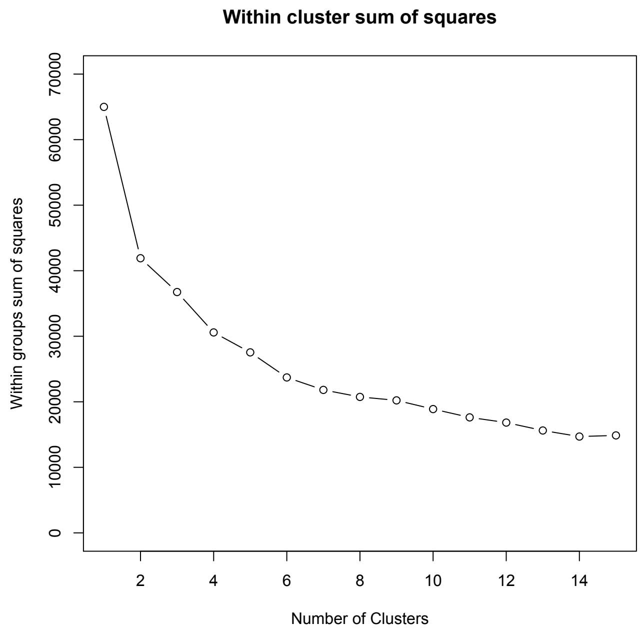


Figure 4.17: Within groups sum of squares relative to the number of clusters dividing the SOM nodes.

number of clusters. It can be seen that the sum of squares decreases steadily with the number of clusters. However, there is a sharper decline when the number of clusters is lower than six. After that, the decrease is slighter.

Figures 4.18 to 4.23 depict the self-organizing map where nodes are divided into two to six clusters. The first interesting thing to notice is the separation of schools belonging to the top 20 and the other ones when there are only two clusters (Figure 4.18). This really demonstrates the fact that the top schools are quite different from the other ranked schools. What is even more surprising is that it is the cluster containing the top schools that is branched into two to give the third cluster (Figure 4.19). It means that there is still a difference inside the top schools group, even though there are themselves far apart from the other schools. These findings are consistent with the fact that the tails of a distribution (represented here by the best- and worst-performing schools) are sparse by definition and will fragment first. For four clusters, it is finally the group of less-performing schools that are divided in two, compared to the map with three clusters (Figure 4.20). The new cluster contains the worst-performing schools, like the University of Economics, Prague (school 69) or the Corvinus University of Budapest (school 70), but also schools that are located in the middle of the FT MiM ranking, like the Montpellier Business School (school 32) or the Antwerp Management School (school 34). This is thus quite a disparate group. Figures 4.21 to 4.23 show the SOM where nodes are divided into five to seven clusters. It can be seen that the top schools are not separated any further.

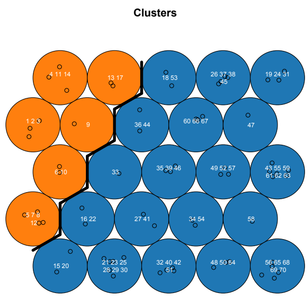


Figure 4.18: SOM nodes divided into two clusters.

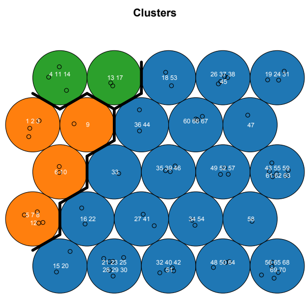


Figure 4.19: SOM nodes divided into three clusters.

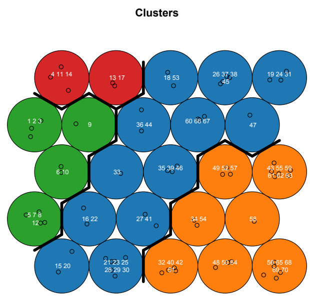


Figure 4.20: SOM nodes divided into four clusters.

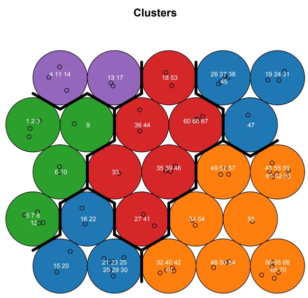


Figure 4.21: SOM nodes divided into five clusters.

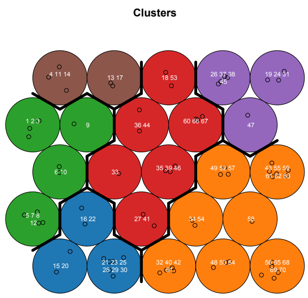


Figure 4.22: SOM nodes divided into six clusters.

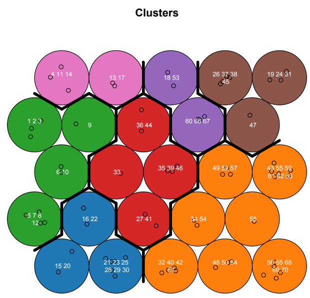


Figure 4.23: SOM nodes divided into seven clusters.

4.4.4 Characteristics of SOM nodes

Until now, it has been seen how the schools organize themselves into SOM nodes and how to divide these nodes into clusters to minimize the sum of squares. It could be interesting to look at the values of the SOM nodes for each criterion. Figures 4.24 to 4.39 depict the values for each of them. For all these figures, a red colouring means that the node presents a high score for this criterion while a blue one means the opposite. It can be seen that the criterion that presents the best continuity of colour between a node and its neighbours is the *Weighted salary* (1) (Figure 4.24). This is not totally surprising as this criterion has the highest FT weight (20%). Besides this criterion, it can be noted that the criteria *Women faculty* (7) (Figure 4.30) and *International course experience* (14) (Figure 4.37) also discriminate the nodes quite well. The former reveals that the top schools tend to have a lower value for this criterion (in particular the first four schools in the FT MiM ranking) while the less-performing schools, like the University of Economics, Prague (school 69) or the Corvinus University of Budapest (school 70) which represent the bottom of the ranking, display a much higher score for this criterion. The latter reveals that nodes in the South-West part of the grid present a higher value for the criterion *International course experience* (14), while nodes in the north-east part of the grid display lower score for this criterion. This is interesting as this distinction does not necessarily follow the FT MiM ranking. Overall, the criteria *Women faculty* (7) and *International course experience* (14) have a correlation with the FT MiM ranking of respectively -0.30 and 0.53, as can be seen from Figure 2.1. Remember that the criterion *Women faculty* (7) was the only criterion with a negative (and significant) correlation.

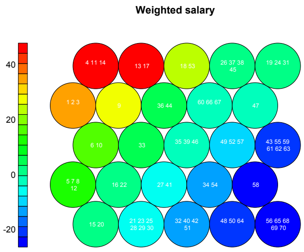


Figure 4.24: SOM coloured with *Weighted salary* (1).

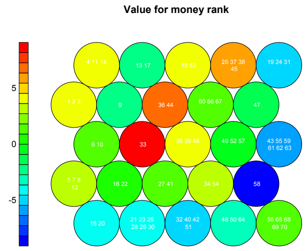


Figure 4.25: SOM coloured with *Values for money rank* (2).

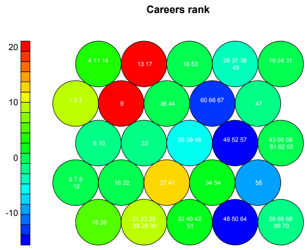


Figure 4.26: SOM coloured with *Careers* (3).

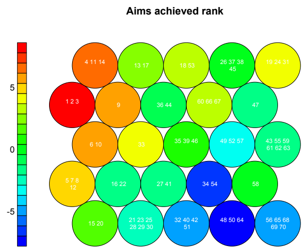


Figure 4.27: SOM coloured with *Aims achieved* (4).

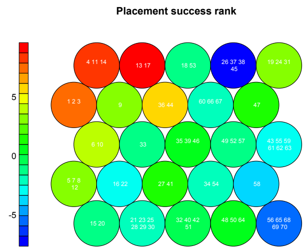


Figure 4.28: SOM coloured with *Placement success* (5).

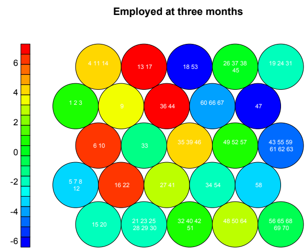


Figure 4.29: SOM coloured with *Employed at three months* (6).

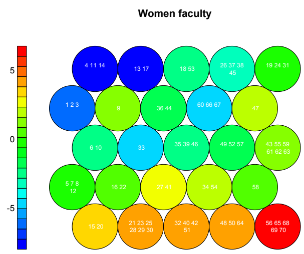


Figure 4.30: SOM coloured with *Women faculty* (7).

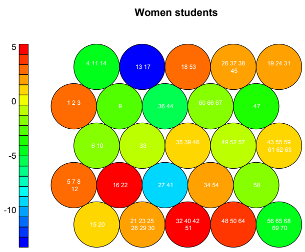


Figure 4.31: SOM coloured with *Women students* (8).

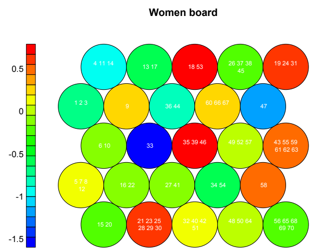


Figure 4.32: SOM coloured with *Women board* (9).

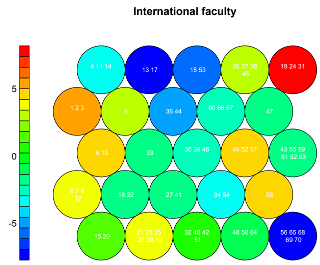


Figure 4.33: SOM coloured with *International faculty* (10).

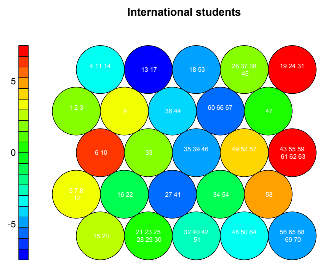


Figure 4.34: SOM coloured with *International students* (11).

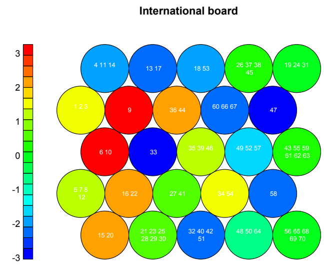


Figure 4.35: SOM coloured with *International board* (12).

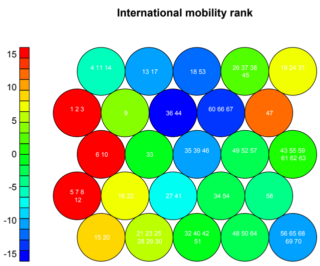


Figure 4.36: SOM coloured with *International mobility* (13).

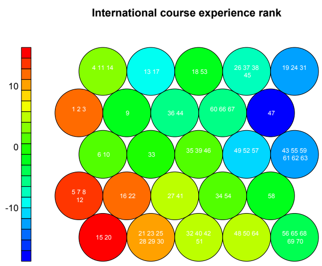


Figure 4.37: SOM coloured with *International course experience* (14).

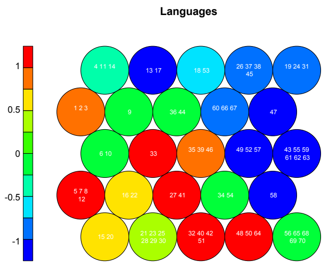


Figure 4.38: SOM coloured with *Languages* (15).

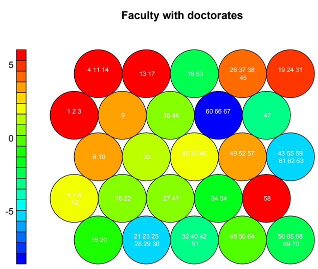


Figure 4.39: SOM coloured with *Faculty with doctorates* (16).

4.5 Multi-scale Jensen-Shannon Embedding (Ms. JSE)

In this section, two Ms. JSE projections have been conducted using the weighted data from the Financial Times. Like for reversed PCA, the first projection is a two-dimensional embedding while the second one is a three-dimensional embedding. The motivations for these projections are the same than for reversed PCA: a projection into two dimensions has the advantage to be able to be represented on a plot whereas a projection into three dimensions is valuable as the discrepancies of adding one dimension could be studied.

4.5.1 Two-dimensional embedding

4.5.1.1 Results

Results of a two-dimensional embedding are first presented presented. Figures 4.40 and 4.41 show the projection of the data. In the first figure, the schools are coloured according to their FT rank while in the second one, the schools are coloured according to the criterion *Weighted salary* (1). Each school is numbered according to its FT rank. Each circle encompasses ten schools and are centered on the University of St Gallen (school 1). As in reversed PCA, it can be seen that the innermost circle is the widest, meaning that the best-performing schools are also the most dissimilar among them. On the contrary, the narrowest slice is the sixth (beginning from the center). Also, some schools that are fairly well ranked in the FT MiM ranking belong to the outermost area in the Ms. JSE projection. This is the case for the Imperial College Business School (school 19) and the City University: Cass (school 24).
Besides, it can be noticed that nearly every school in the top ten of the FT MiM ranking belongs to the innermost circle. Only ESCP Europe (school 7) lies in the second innermost slice, but it remains quite close to the other top schools. A final remark that can be done regarding Figure 4.40 is the high equivalence between two German schools: HHL Leipzig Graduate School of Management (school 11) and EBS Business School (school 14). These two schools were already close to each other in the projection using reversed PCA, but there are almost overlapping in the Ms. JSE projection.

Figure 4.41 depicts the projected schools coloured according to the criterion *Weighted salary* (1). It can be seen that even though the schools are less discriminated according to this criterion than they were with reversed PCA, a general scheme can be defined. Schools with a high value for this criterion tend to lie on the South-East part of the projection while the opposite schools tend to be located on the North-West part. Particularly, it can be noticed that the two schools presenting the highest score for this criterion (the WHU Beisheim (school 4) and Indian Institute of Management, Ahmedabad (school 17)) are situated in the South-East part of the projection.

4.5.1.2 Quality assessment

Now that the results of the projection in two dimensions have been presented, an evaluation of the quality assessment should be conducted. Figure 4.42 shows the Shepard diagram while Figure 4.43 depicts the NX-scores. It can be seen that the cloud of points constituted by the comparison of the distances in the high- and low-dimensional spaces is quite fuzzy, especially when the distances in high dimension grow. A structure of a straight line can not easily be guessed. However, as MS. JSE attempts to preserve the similarity, Figure 4.43 could seem more suitable for assessing the quality of the dimensionality reduction. It can be observed that the overall similarity is quite high. It begins at around 85% when there is no neighbourhood, goes up and down between 70% and 80% until the size of the neighbourhood reaches 45 and stands above 80% after that.

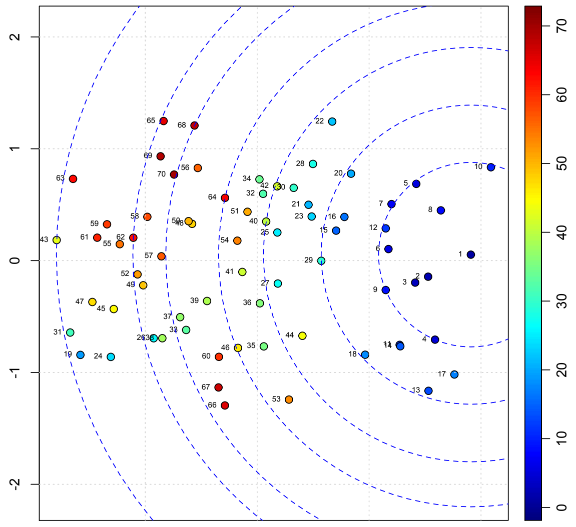


Figure 4.40: Projection of schools in 2D using Ms. JSE coloured with their rank.

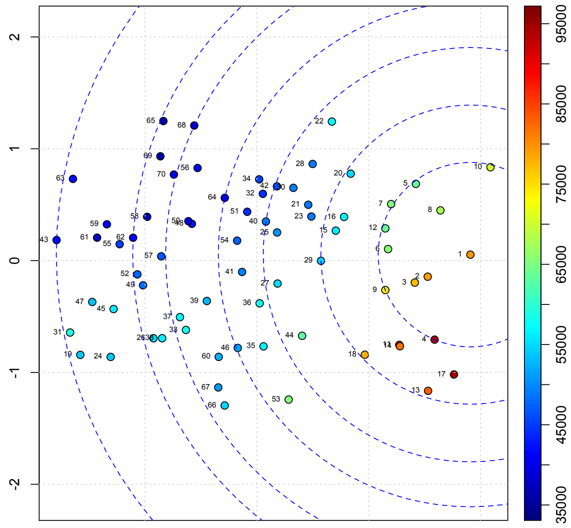


Figure 4.41: Projection of schools in 2D using Ms. JSE coloured with their salary.

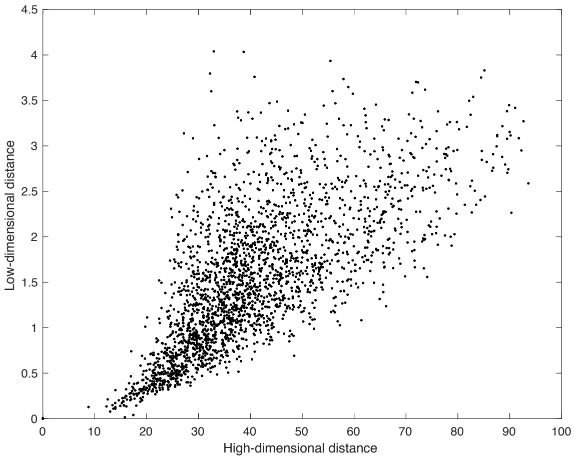


Figure 4.42: Shepard diagram of the projection in 2D using Ms. JSE.

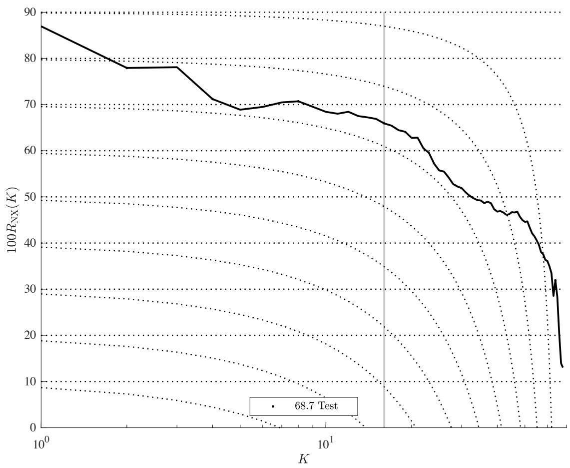


Figure 4.43: NX-Scores of projection of schools in 2D using Ms. JSE.

4.5.2 Three-dimensional embedding

It could be interesting to focus on a projection into three dimensions to assess the differences in quality coming from adding one dimension to the projection. As in reversed PCA, only a plot presenting the comparison of the distances in the high- and low-dimensional spaces (Figure 4.44) and a plot depicting the NX-scores (Figure 4.45) are shown here, as a three-dimensional projection could hardly be understandable when presented on a report. Overall, Figure 4.44 presents a much less blurred cloud of points. The structure of a straight line can be guessed from this scatter plot. Again, small distances in the high dimensional space presents a lower variance when translated in three dimensions.

As for the two-dimensional embedding however, Figure 4.45 is much more valuable to assess the quality of the reduction, for MS. JSE is a similarity-based NLDR method. It can be seen from NX-scores that this projection preserves the most the similarity of the data set among embeddings that were presented so far. The similarity is as high as 95% when there is a neighbourhood of size 1. It decreases at around 80% until the size of the neighbourhood reaches 25. It keeps increasing for bigger neighbourhoods.

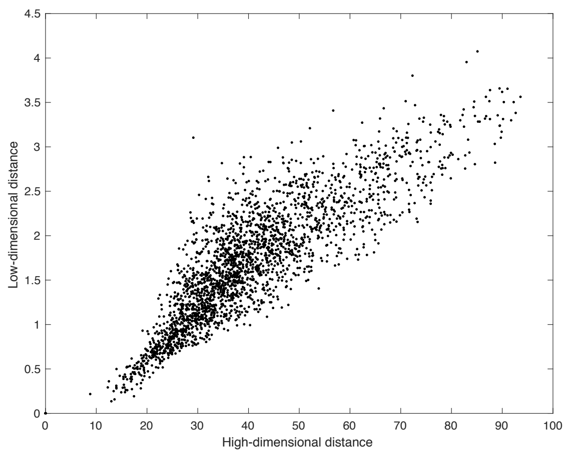


Figure 4.44: Shepard diagram of projection in 3D using Ms. JSE.

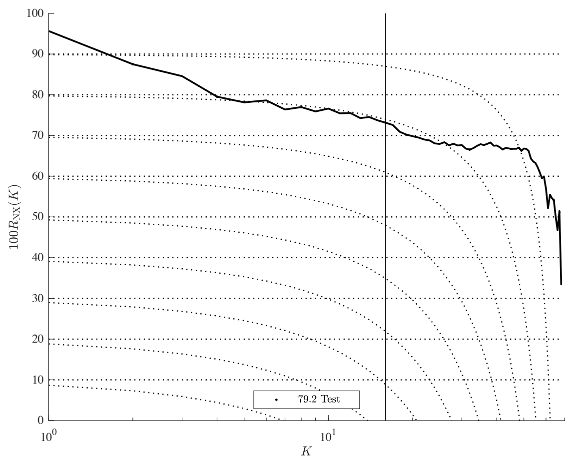


Figure 4.45: NX-Scores of projection of schools in 3D using Ms. JSE.

4.6 Comparison between the three DR methods

Now that results for the three DR methods have been presented, a comparison between them is conducted. First, only the results of projections are compared. Then, the quality assessments of the DR methods are considered.

4.6.1 Comparison of projections

These DR methods have been defined in Chapter 3 and their results presented at the beginning of this chapter. While the respective goals of these methods were different, it could still be interesting to compare these techniques and examine if they allowed to extract common features from the FT data set. The comparison concentrates on Figures 4.8, 4.14 and 4.40. For convenience, they are also displayed on a single page in Figures 4.46, 4.47 and 4.48.

One of the main observation when discussing results of these DR methods was to determine if the top ten schools were still packed together after projection, to represent the fact that they were well ahead of other schools in the reconstruction of the FT MiM ranking, in Section 2.2. The clustering of the self-organizing map clearly demonstrated the fact that the top schools were quite distant from the other ones. Indeed, when using two clusters to divide the SOM nodes, the top schools were grouped into one cluster while the rest of the schools shared the other cluster. For the weighted reversed PCA, it has been seen that the top ten schools were gathered in 2D around the University of St Gallen (school 1), except for the WHU Beisheim (school 4). Also, for Ms. JSE, this latter school seems a little distant from the other best-performing universities even though this is less distinct than for the reversed PCA.
A last remark that could be done on the top schools is the proximity of the Universite Bocconi (school 12) in reversed PCA and Ms. JSE. This can also be noticed in SOM where this school shares its node with schools Cems (school 5), ESCP Europe (school 7) and Rotterdam School of Management, Erasmus University (school 8).

At the opposite, the five worst-performing schools were shown to present low variance when facing uncertainties in the robustness analysis, in Section 2.3. It could then be appealing to

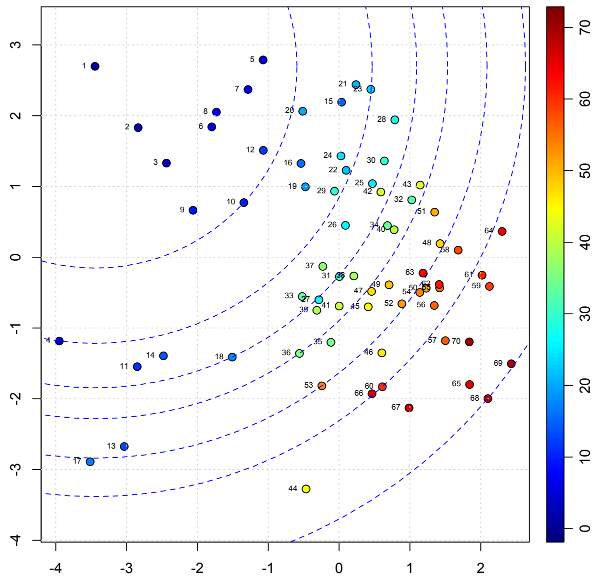


Figure 4.46: Projection in 2D of the schools following the weighted reversed PCA coloured with their FT rank.

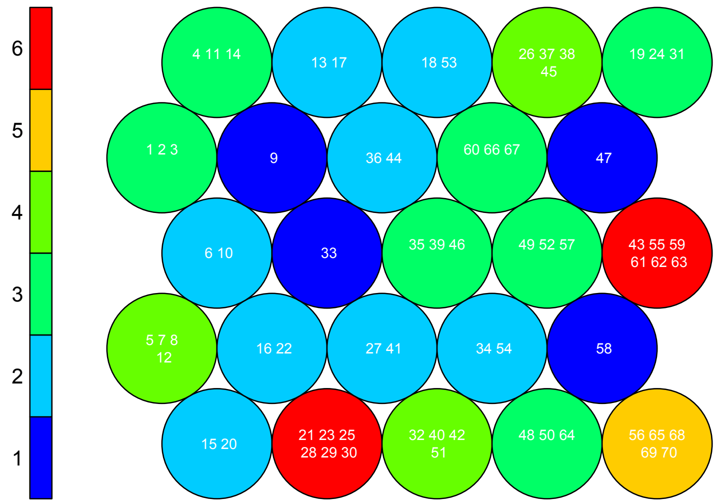


Figure 4.47: Self-organizing map based on the FT data set.

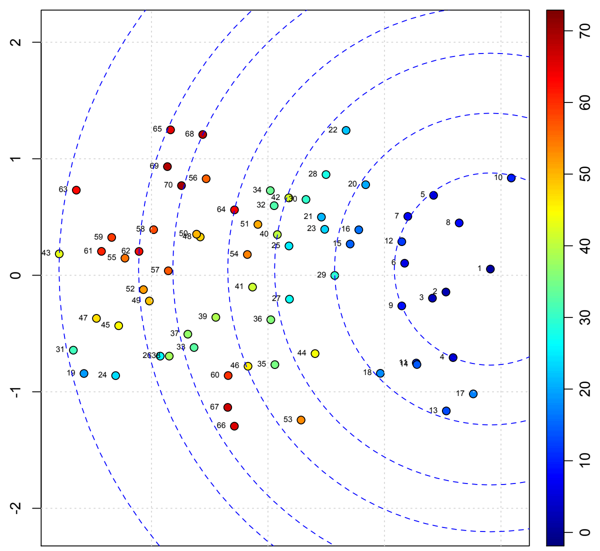


Figure 4.48: Projection of schools in 2D using Ms. JSE coloured with their FT rank.

compare this behaviour in the different projections.
In SOM, these universities are packed in two different nodes that are somewhat quite distant from each other. Indeed, these two nodes do no even belong to the same cluster when the self-organizing map is divided into four parts. One of the node contains schools Nyenrode Business Universiteit (school 66) and BI Norwegian Business School (school 67), while the other one includes schools Tongji University School of Economics and Management (school 65), Warsaw School of Economics (school 68), University of Economics, Prague (school 69) and Corvinus University of Budapest (school 70).
In the projection obtained with reversed PCA, it can be seen that these two groups of less-performing schools can be identified, even though the Corvinus University of Budapest (school 70) is a little apart from the other schools of its SOM node. Furthermore, it can be noticed that the two groups of schools are quite close to each other in this projection.
Finally, in the Ms. JSE, the situation is noticeably different from reversed PCA. Indeed, even though the two groups can definitely be noticed, they are also very far apart from each other, with the main group of schools separating them.
It is quite interesting that the worst-performing schools are discriminated into two groups. It could be explained by the fact that there are two possible reasons that a school belongs to the bottom of the ranking. From the colouring of SOM nodes according to FT criteria (Figures 4.24 to 4.39), it could be suggested that the group composed of schools ranked 65, 68, 69 and 70 present a poor rank because of their low score to the criterion *Weighted salary* (1) (which represents 20% of FT weights), while the second group constituted of schools ranked 66 and 67 could accuse their mediocre score in the criterion *Careers* (3) (which accounts for 10% of weights) for their poor rank. These two groups could obviously also be discriminated according to other criteria, but *Weighted salary* (1) and *Careers* (3) were cited here as they represent ones of the highest FT weights and could therefore have a great effect on the ranking.

Another aspect that is worth taking a look at is the SOM node that gathered the quite disparate schools Mannheim Business School (school 18) and University of Cologne, Faculty of Management (school 53). From the other projections, it could be interesting to find out if these two schools aggregated into one node because there were no more available SOM nodes in the map or because they really share common characteristics. In both projections of reversed PCA and Ms. JSE, it can be seen that, even though these two schools are near each other, they are not quite close. For instance, the Mannheim Business School (school 18) seems more distant to the University of Cologne, Faculty of Management (school 53) than to schools HHL Leipzig Graduate School of Management (school 11) and EBS Business School (school 14) in both projections. This demonstrates that this SOM node constituted of these two schools was not created because of the extreme proximity between them, even though they remain unusually near to each other in the light of their FT rank.

A final discussion that could be considered is the position of schools Solvay Brussels School of Economics and Management (school 39) and Louvain School of Management (school 41) in these projections. As a reminder, they were examined during the robustness analysis. In reversed PCA, they are close to each other. They remain quite neighbouring in Ms. JSE, even though they are a little more distant. In the self-organizing map, they belong to different SOM nodes but these nodes are adjacent and they are associated with the same cluster even when the map is divided into seven clusters (Figure 4.23).
All this confirm the real proximity of these two schools that was suggested in the robustness analysis.

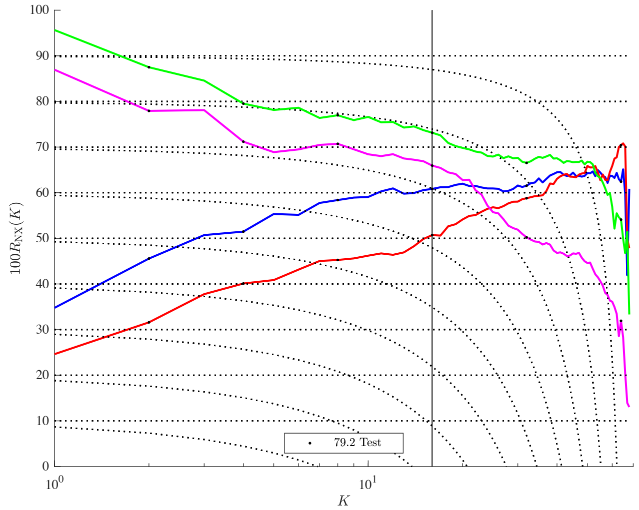


Figure 4.49: Comparison of NX-scores for different projections. Blue and red lines match respectively projections into 2D and 3D using reversed PCA, while magenta and green lines correspond to respectively projections into 2D and 3D using Ms. JSE.

4.6.2 Comparison of the quality assessment

Now that an extensive comparison of the results of the projections has been conducted, a focus is put on quality assessments of the different projections. It should be noted that the comparison is done using the NX-scores. Indeed, one of the advantages of these scores is that NX-scores from multiples projections can be plotted on the same figure. Therefore, this comparison concentrates on quality assessments of reversed PCA and Ms. JSE as it has been seen in Section 4.4 that NX-scores for SOM are irrelevant because of the discrete distribution of SOM nodes in low dimensions.
Figures 4.11, 4.13, 4.43 and 4.45 depict the NX-scores for respectively projections into 2D and 3D using reversed PCA and projections into 2D and 3D using Ms. JSE. To ease the comparison, they are all gathered into one plot displayed in Figure 4.49. Blue and red lines match respectively projections into 2D and 3D using reversed PCA, while magenta and green lines correspond to respectively projections into 2D and 3D using Ms. JSE.

To assess the quality of each projection, a scalar score, called AUC, is assigned to each of them, obtained by computing the area under the curve of the NX-scores. To reflect the fact NX-scores are displayed in log plots, a mathematical definition of AUC used in this comparison could be:

$$AUC(R_{NX}(K)) = \frac{\sum_{K=1}^{N-2} \frac{R_{NX}(K)}{K}}{\sum_{K=1}^{N-2} \frac{1}{K}} \tag{4.8}$$

The AUC thus assesses the average quality of projection on all scales, with a focus on small neighbourhoods.

It can be seen from Figure 4.49 that there is a marked difference between the two DR methods. While the similarity when the neighbourhood is small is quite high for Ms. JSE - it nearly reaches 90% of similarity for the two-dimensional embedding - , reversed PCA presents a poor similarity when the neighbourhood is narrow.

Table 4.5 depicts the AUC scores for each projection. It can be observed that projections using Ms. JSE received significantly better AUC score. More importantly, it can even be noticed that the projection into 2D using Ms. JSE obtains a better score than the projection into 3D using reversed PCA. This really shows the advantage of using non-linear models to project data into a lower-dimensional space, in this case.

Projection	$AUC(R_{NX}^{K})$
Reversed PCA (2D)	0.4299
Reversed PCA (3D)	0.5232
Ms. JSE (2D)	0.6870
Ms. JSE (3D)	0.7918

Table 4.5: $AUC(R_{NX}^{K})$ scores for each projection.

4.6.3 Conclusion

In this section, the projections of the three DR methods were first compared. A focus was put on both the best- and worst-performing schools in the FT MiM ranking, a SOM node that aggregated two schools far apart in the ranking and the two schools that were extensively compared in the robustness analysis. What really stands out of this comparison is that these three projections never contradict each other but rather offer a vision a little different than the two others.
Furthermore, it has been shown that these projections could be used together to extract more information. For instance, starting from the assessment that the worst-performing schools presented a low variance in the robustness analysis (which could be explained by the fact that they were far below the other universities), it has been shown that they were discriminated into two well-distinct groups in the Ms. JSE projection. Then, using the self-organizing map, it has been demonstrated that one group was weak because of their low score in *Weighted salary* (1) while the second one did not perform well because of their low score in *Careers* (3).

After that, a comparison has been conducted on the NX-scores, to assess the best DR method for the FT data set. An *area under curve* (AUC) score has been used to compare projections. Reversed PCA clearly loses against Ms. JSE. The latter presented a better AUC score even for a projection into 2D, compared to projection into 3D using reversed PCA.

In the next section, a one-dimensional embedding will be carried out. The reduction to one dimension could be interesting to look at as the FT MiM ranking can also be viewed as a one-dimensional embedding: the information about schools according to sixteen criteria are aggregated into one score to form the ranking. The DR method Ms. JSE has been chosen to perform this dimensionality reduction as it displayed the best preservation during the comparison, in Subsestion 4.6.2.

4.7 One-dimensional embedding

Only two- and three-dimensional embeddings were considered so far. As designing a ranking can be seen as a projection into one dimension - summarizing information embedded in the data as a unique number - , the last dimensionality reduction that is conducted in this chapter is a one-dimensional embedding. The DR method Ms. JSE has been chosen to perform this dimensionality reduction as it displayed the best preservation during the comparison, in Subsestion 4.6.2. Figure 4.50 shows the result of this projection. The University of St Gallen (school 1) is identified by a red marker on the plot. It can be seen that this school does not obtain the highest score in the only dimension left. However, as already said, the meaning of dimensions in a non-linear dimensionality embedding is much more difficult to explain. Therefore, this projection does not necessarily mean that the Indian Institute of Management, Calcutta (school 13) is the best one according to this reduction. It can be noticed from this plot that the top schools according to the FT MiM ranking are still packed around the University of St Gallen (school 1), while the least-performing schools according to the FT MiM ranking are all quite distant from it. Furthermore, as it was observed in two-dimensional projections using reversed PCA and Ms. JSE, the WHU Beisheim (school 4) is the most isolated among the top ten schools. The latter is even nearer in Figure 4.50 to schools HHL Leipzig Graduate School of Management (school 11) and EBS Business School (school 14) with whom it shared a node in the self-organizing map. Besides, the Universite Bocconi (school 12) is projected quite close to the top ten schools, as it was already remarked in other projections using Ms. JSE.

Figure 4.51 displays the NX-scores for this one-dimensional embedding, as well as the NX-scores for the two previous projections using Ms. JSE and the two other projections using reversed PCA. On this plot, light blue line corresponds to the projection into 1D using Ms. JSE. Deep blue and red lines match respectively projections into 2D and 3D using reversed PCA, while magenta and green lines correspond to respectively projections into 2D and 3D using Ms. JSE.
Naturally, the projection into 1D achieves a poorer similarity with respect to the two- and three-dimensional embeddings using the same method. However, its results are quite surprising compared to projections using reversed PCA. This is supported by the fact that this one-dimensional embedding present a $AUC(R_{NX}(K))$ score equal to 0.6154. This is even better than the AUC for the projection into 3D using reversed PCA.

A last analysis that could be conducted on this one-dimensional embedding is to attempt to extract a ranking from this projection. As it was mentioned before, the meaning of new axes defined by a NLDR method are not easily guessed. Therefore, in order to create this ranking, the University of St Gallen (school 1) is assumed to be the best-performing school and the other universities are ranked according to their distance from it. A comparison between this new ranking and the one from the Financial Times can be found in Figure 4.52. The first thing that can be noticed is the relative dissimilarity with the FT MiM ranking. The Spearman correlation between these two is 0.865. The highest difference in place between the two rankings is 30.
The difference could be partly explained by the fact that concerning the top twenty schools, they are projected on both sides of the University of St Gallen (school 1) along the unique axis.

4.8 Conclusion

This chapter was devoted to analyze the application of three DR methods on the FT data set. The main goal to use DR methods was to visualize the FT data in two or three dimensions, so that it would be easier to grasp the connections between schools and if the FT MiM ranking truthfully translate them.
It has been shown that the three DR methods generally agree on the global structure of the projection, even though they present distinct differences. However, the comparison between the quality assessments clearly demonstrated the dominance of Ms. JSE compared to reversed PCA. This was even more obvious when the $AUC(R_{NX}(K))$ were measured: the one-dimensinal embedding using Ms. JSE received an even better score than the projection into three dimensions using reversed PCA.

The DR methods, independently from each other, taught us many things. For instance, the difference between linear combinations in non-weighted and weighted PCA really showed the heavy impact that arbitrary weights could have on the data set.
More importantly however, what truly stands out from this chapter is the advantage of using together the different characteristics of the DR methods to extract better information. For example, it has been seen in the projection into 2D using Ms. JSE that the worst-performing

schools were packed into two very distant groups. This phenomenon was confirmed by the self-organizing map and the scores of each criterion associated to SOM nodes allowed us to figure out that one group of schools was not well performing in the ranking because of their low scores for the criterion *Weighted salary* (1) while the second group did not achieved to get a good score for the criterion *Careers* (3). This information would have been more difficult to extract if only a projection using Ms. JSE was conducted.
This can most likely be explained by the fact that each DR method presented here focus on a different aspect of the dimensionality reduction and they each have their advantages.

At the end, it has been learned that Ms. JSE could be a tremendous projection tool but it has its limits. Indeed, when performing a one-dimensional embedding, the Ms. JSE could not tell us the best-performing school so that it has been assumed that the University of St Gallen (school 1) was the best one and a ranking could thus be created from it.

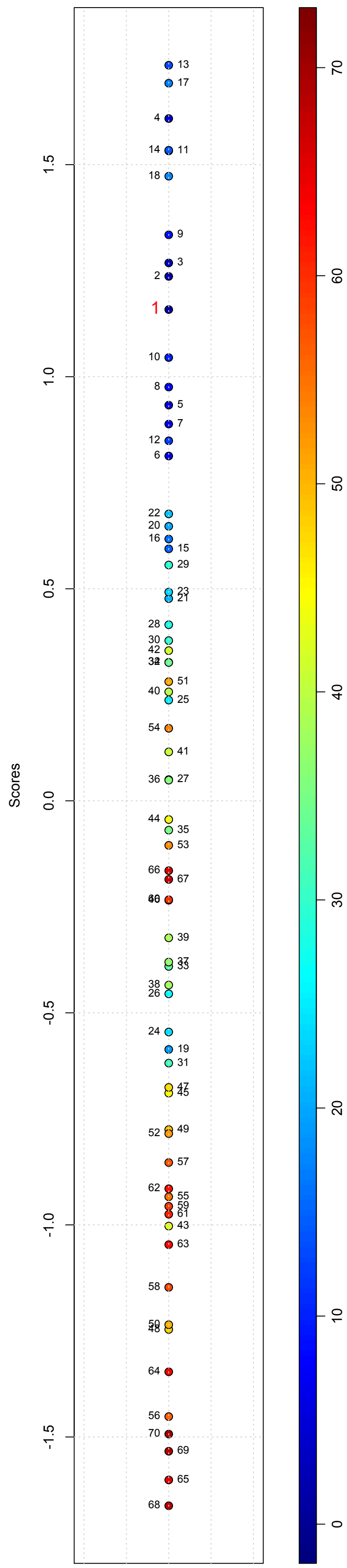


Figure 4.50: Projection of schools in 1D using Ms. JSE coloured with their rank.

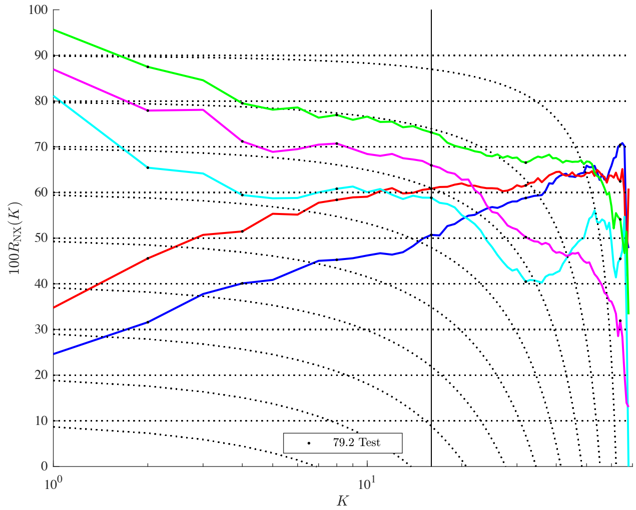


Figure 4.51: Comparison of NX-scores for different projections. Light blue line corresponds to the projection into 1D using Ms. JSE. Deep blue and red lines match respectively projections into 2D and 3D using reversed PCA, while magenta and green lines correspond to respectively projections into 2D and 3D using Ms. JSE.

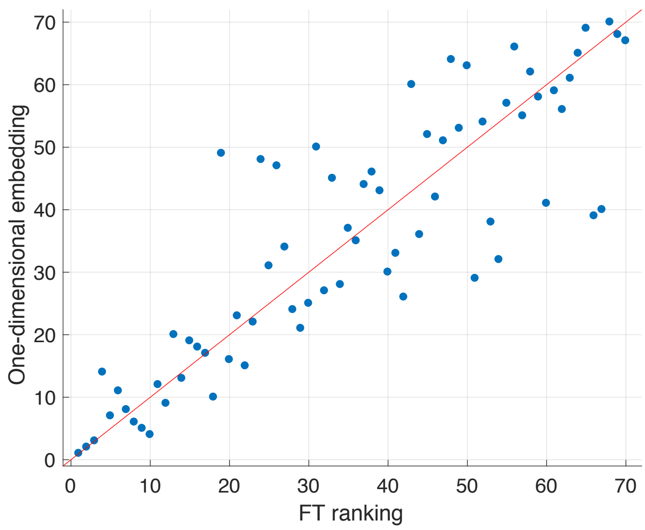


Figure 4.52: Comparison between the FT MiM ranking and the ranking obtained via a one-dimensional projection using Ms. JSE. The straight line represents the identity function.

5 Design of a new ranking

Abstract Throughout these first chapters, the FT MiM ranking as well as data used to construct it have been presented and deeply analyzed. In this chapter, conclusions made from these analyses are gathered and become a starting point to design new rankings. The first one, called Bayesian Latent Variable Analysis (BLVA), aims to re-create the weights based on information extracted from data instead of arbitrary decisions. The second ranking, called Budget Allocation Process (BAP) shares the same philosophy: redesigning the weights. However, BAP focuses on teaching quality defined by students and universities. Finally, the third ranking attempts to enhance the existing FT MiM ranking, without modifying it, by adding two-dimensional projections of the FT data.

5.1 Introduction

This chapter is devoted to proposals of new rankings based on conclusions of what have been done in this work. Indeed, throughout the first chapters, the FT MiM ranking as well as data used to construct it have been presented and deeply analyzed. It could be worth focusing on gathering information from these analyses to form a starting point to design new rankings.
Among problems that were met in this work, some of them can be named as they will be used to explain the choice of the following new rankings. First, in the reconstruction phase in Section 2.2, it has been seen that the computed z-scores revealed much more information than just ranked numbers, as in the FT MiM ranking, which has already been the source of some criticisms [7]. Secondly, in the robustness analysis, it has been discovered that the FT weights were a great source of uncertainty. Finally, it could also be mentioned that the analyses of local neighbourhoods of schools by projections into two or three dimensions with various methods, brought a lot of information on universities that were quite different from what could have been understood when only looking at the FT MiM ranking.

Three rankings are presented in this chapter and result from conclusions made in the previous chapters. The first one, called Bayesian Latent Variable Analysis (BLVA), is based on latent variables and Bayesian inference. It attempts to re-create the weights of the ranking from information extracted from data rather than arbitrary decisions. Due to the use of Bayesian inference, it also provides some information on the uncertainty in the new ranking. It is motivated by the fact that the weights were a source of great uncertainty in the robustness analysis and that the teaching quality defined by the weights has already been a cause of criticisms. However, as it will be seen, results of BLVA provide a teaching quality that is not totally acceptable. This can partly be due to high pairwise correlations in data.

The second ranking presented in this work, called Budget Allocation Process (BAP), attempts to resolve the problem seen in BLVA (that the teaching quality was not acceptable) by allowing students and universities to define themselves what is a good teaching. Two case studies are conceived to represent two possible students that look for different aspects in a business school. The two resulting rankings are then analyzed and compared.

The last ranking focuses on enhancing the FT MiM ranking by adding some information to it, rather than re-creating a new ranking from scratch. In this work, the z-scores computed in Section 2.2 and the two-dimensional projection using Ms. JSE seen in Section 4.5 are used. The choice of these two enhancements are driven by the fact that some criticisms have been made saying that the FT MiM ranking does not represent well the differences between schools (hence the z-scores) and that the FT MiM ranking fails to describe faithfully local neighbourhoods around schools (hence the Ms. JSE projection). An analysis of the information added by the z-scores and the Ms. JSE projection is then conducted. The last part of the presentation of this ranking is dedicated to a discussion of why adding only one projection and not more.

5.2 Bayesian Latent Variable Analysis (BLVA)

In the robustness analysis, it has been demonstrated that the uncertainties in weights could heavily affect the ranking. Considering that the FT weights were fixed arbitrarily by the Financial Times, it could be interesting to look at new rankings that attempt to overcome these shortcomings.

The ranking explained in this section is based on a latent variable analysis (LVA). It considers that two kinds of variables underlying data can be distinguished. First, the *manifest* variables. They can be directly observed. The second type of variables is the latent variables. They are normally unobserved, implicit. Generally, there are fewer latent variables than manifest ones. The goal of a LVA is to find the latent variables from the manifest ones. That way, the phenomenons described by the data set could be explained with fewer variables than before [22]. For instance, in the FT data set, the sixteen criteria are *manifest* variables and the aim of a LVA could be to find a single (latent) variable, so that a school could be described by a unique score. The value of this latent variable could describe as accurately as possible the quality of a school. A parallel can be done with the score of the weight-and-sum in the FT MiM ranking. Indeed, the latter attempts to find a unique criterion describing schools that could be compared to the single latent variable.

A Bayesian Latent Variable Analysis (BLVA) was performed on the *U.S. News & World Report* and the *America's Best Colleges* rankings in [17]. The BLVA that is conducted here on the FT MiM ranking follows the same methodology. The usual approach to design university rankings is to gather data on criteria found relevant to assess the quality of a university, to fix weights for each of these criteria (often based on subjective opinions) and then to perform a weight-and-sum algorithm, which gives an ordinal ranking of universities. The approach backed by the BLVA is to conduct a statistical procedure based only on data formed by criteria relevant to assess the quality of an institution (and thus not on subjective weights on these criteria). The output of this procedure is a sequential ordering of schools that also contains some information on the uncertainty in the ranking for each school. This new approach allows to solve some problems that were previously raised, as mentioned in [17]:

- How to take into account incertitude in the rankings when there is some uncertainty in the input or in the methodology ?
- How to remove from the ranking subjective opinions like the weights applied to criteria ?
- When to announce that a given school is statistically better than another one ?

Solving these problems could have very practical consequences. Answering to the first and third questions could help people understand the true difference between two universities and to know if these schools are really statistically different. In a regular ranking, two schools ranked adjacently could be nearly identical but are ranked differently anyway because of the intrinsic characteristics of a ranking. Answering to the second question could help people designing rankings to avoid some criticism by decreasing the part of subjectivity. However, it has to be noted that BLVA does not eliminate subjectivity either. BLVA is based on data formed by criteria relevant to assess the quality of an institution. Choosing a given criterion rather than another is still the decision of people designing the ranking.

Apart from [17], other studies analyzed the advantages and disadvantages of using a BLVA to rank universities or other educational institutions. A good example is [16] that conducted an extensive study on the use of a BLVA to rank health and educational institutions.

5.2.1 Methodology

As already said, the ranking presented here is based on the one developed in [17] which relies on a Bayesian Latent Variable Analysis (BLVA). The latent variable procedure differs from other weight-and-sum approaches because it determines the weights for each criterion based on information embedded in data rather than subjectively fix them *a priori*. Another difference of the latent variable approach is that it also computes the uncertainty for each school in the rankings and not only an ordinal ranking.
In the BLVA developed in this work, it is supposed that each school is characterized by an unobserved, latent quality-related measure, noted z_i. The goal of the BLVA is to find out these z_i from the sixteen observable criteria given in the FT data set. A linear model is used to associate the observed criteria with the latent quality measure z_i, which gives a system of sixteen regression models that can be expressed as:

$$\begin{aligned} feature_i^{(1)} &= \beta_0^{(1)} + \beta_1^{(1)} z_i + \epsilon_i^{(1)},\ \epsilon_i^{(1)} \sim N(0, \sigma_1^2)\ , \\ feature_i^{(2)} &= \beta_0^{(2)} + \beta_1^{(2)} z_i + \epsilon_i^{(2)},\ \epsilon_i^{(2)} \sim N(0, \sigma_2^2)\ , \\ &\vdots \\ feature_i^{(16)} &= \beta_0^{(16)} + \beta_1^{(16)} z_i + \epsilon_i^{(16)},\ \epsilon_i^{(16)} \sim N(0, \sigma_{16}^2)\ , \end{aligned}$$

where $feature_i^{(j)}$ represents the value of the school i for the criterion j. The sixteen $\beta_0^{(j)}$ can be seen as intercept terms and the sixteen $\beta_1^{(j)}$ as coefficients on z_i. $\epsilon_i^{(j)}$ represent error terms. These error terms follow a normal distribution with a mean of 0 and a standard deviation of σ_i where σ_i is the standard deviation of values for the criterion j. The goal of the Bayesian approach is to find the posterior distribution of z_i and calculate a ranking from that. As the joint posterior distribution of z_i, β_0, β_1 and σ is quite complex, the posterior mean of z_i, given all the other parameters and observed data, will rather be computed. Indeed, it is reduced to a linear combination of the observed data:

$$E(z_i|.) = \sum_{j=1}^{J} \frac{\beta_1^{(j)}}{\sigma_j} \frac{x_j - \mu_j}{\sigma_j}\ . \tag{5.1}$$

In practice, to calculate $E(z_i|.)$, all the parameters β_0 were initialized to 0 and β_1 to 1. Then, the posterior mean $E(z_i|.)$ is assessed from these parameters. Parameters β are then estimated by resolving the linear regression models and finally, the posterior mean $E(z_i|.)$ are computed again.
The seventy $E(z_i|.)$ are then normalized (with mean of 0 and standard deviation of 1) and a ranking is constructed based on this. It can be seen that a weight-and-sum approach with weights $w_j = \frac{\beta_1^{(j)}}{\sigma_j}$ is used in the formula 5.1. This method thus determines a ranking of schools

as well as the weights this ranking is based on. This allows this new ranking to be less dependent to subjective opinions.

There is a probabilistic aspect in this ranking as an error term was added in the sixteen linear regression models. A Markov Chain Monte Carlo integration was then used to assess the uncertainty in the ranks, as proposed in [17] and [16]. In practice, the Monte Carlo algorithm consisted in launching 20.000 times the calculation of the posterior mean $E(z_i|.)$ with different error terms. The results of this procedure are presented in the following section.

5.2.2 Results

Figure 5.1 depicts boxplots of estimated ranking for each school, ordered according to the FT MiM ranking. Boxplots display the 95% confidence interval and outliers are not shown in this plot for clarity. Red crosses represent the placement of each school in the FT MiM ranking. It can be seen that the predicted ranking seems quite different to the FT MiM ranking. If only the mean of each estimated ranking are taken into account, the Kendall and Spearman correlations between this ranking and the FT MiM ranking are respectively 0.763 and 0.914. It can be noticed that the schools that presents the highest difference between the two rankings is the Indian Institute of Management, Calcutta (school 13), which is ranked 32.5 places behind its position in the FT MiM ranking. The school that shows the biggest increase in the new ranking is the Solvay Brussels School of Economics and Management (school 39). It can also be observed that only two schools present a 95% confidence interval contained in one rank place: the University of St Gallen (school 1) and the IE Business School (school 9).

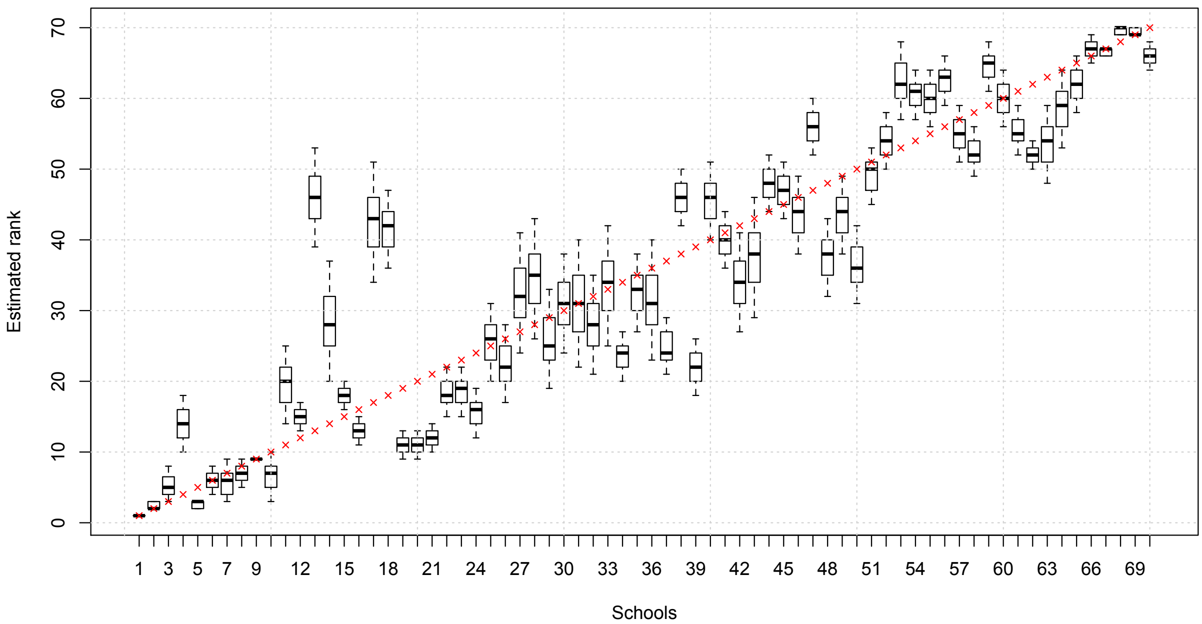


Figure 5.1: Estimated ranking for schools ordered according to the FT MiM ranking in BLVA.

Figure 5.2 depicts boxplots of normalized estimated value of the latent variable rather than ranking. Again, schools are ordered according to the FT MiM ranking, boxplots display the 95% confidence interval and outliers are not shown in this plot for clarity. It could be interesting to look at this figure to determine if there exists a significant gap between two schools adjacent in the newly created ranking. It can be seen that the gap between the WHU Beisheim (school 4) and schools adjacent to it is quite important, which results in a lower ranking as well. The same is true for the Indian Institute of Management, Calcutta (school 13).

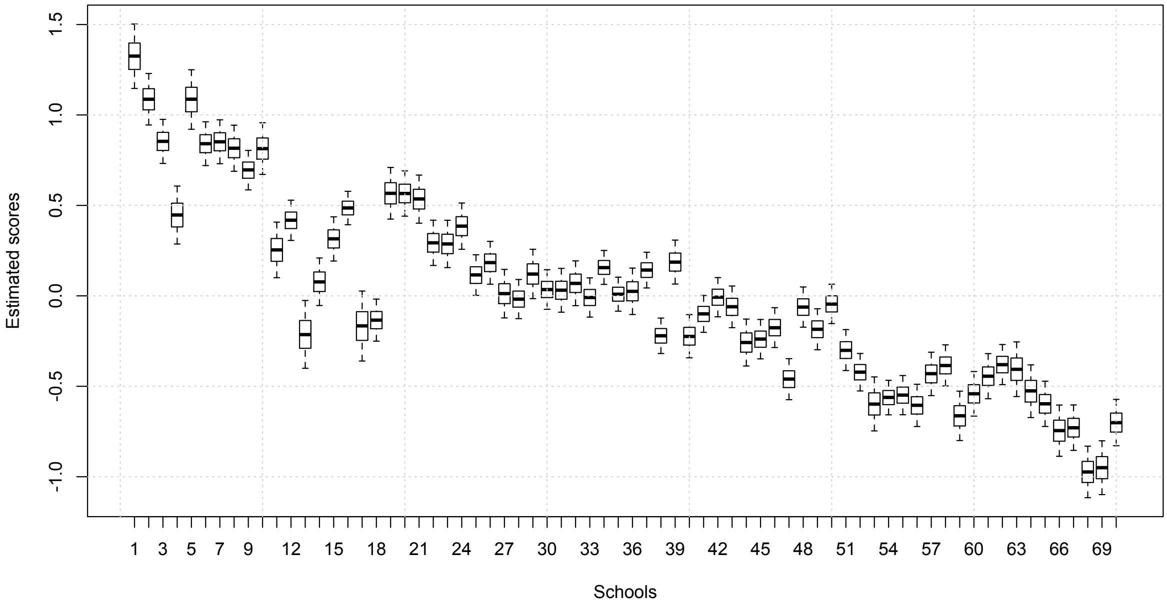


Figure 5.2: Estimated scores for schools in BLVA, ordered according to the FT MiM ranking.

Until now, the ranking and the uncertainty behind BLVA have been described but there is still an important aspect of this ranking that needs to be analyzed, namely the estimated weights. As a reminder, these weights were estimated as $w_j = \frac{\beta_1^{(j)}}{\sigma_j}$, where w_j stands for the weight of the criterion j, $\beta_1^{(j)}$ for the coefficient in front of z_i in the linear regression model and σ_j the standard deviation for values of the criterion j. Figure 5.3 depicts boxplots of the estimated weights. As for the previous figures, boxplots display the 95% confidence interval and outliers are not shown in this plot for clarity. Red crosses represent the FT weights. Several remarks can be done. First of all, the criterion *Weighted salary* (1), which has the highest FT weight (20%) only presents a weight of 7.5% here. It has not even the highest weight any more. The first place is held by the criterion *International mobility* (13) at around 12% (it has a FT weight of 10%). Secondly, the estimated weight of the criterion *Women faculty* (7) is quite surprising, with a negative mean, suggesting that a high value for this criterion could harm the ranking of a school. Two other criteria have negative values in their 95% confidence interval: *Value for money* (2) and *Women board* (9).
Overall, a majority of weights ranges from 5% to 10% and there is no more extremely highly-weighted criterion, like the *Weighted salary* (1).

5.2.2.1 Use of these results

Results being given in the form of boxplots, one could think about how the schools can be compared with each other. Indeed, when looking at Figure 5.1, one could consider using the means to compare schools, or the minimum rank contained in the boxplot, etc. A solution proposed by [17] is to compare schools by using their boxplots. It suggests that a school is statistically superior to another one if their respective boxplots do not overlap and the one of the first school is better ranked. For instance, if the two schools Solvay Brussels School of Economics and Management (school 39) and Louvain School of Management (school 41) compared in the robustness analysis are considered, the first one is statistically superior to the second one because their boxplots do not overlap. However, the schools Vlerick Business School (school 36) and Louvain School of Management (school 41) can not be statistically differentiated as their boxplot do overlap, even though the first one presents a mean rank better than the second one.

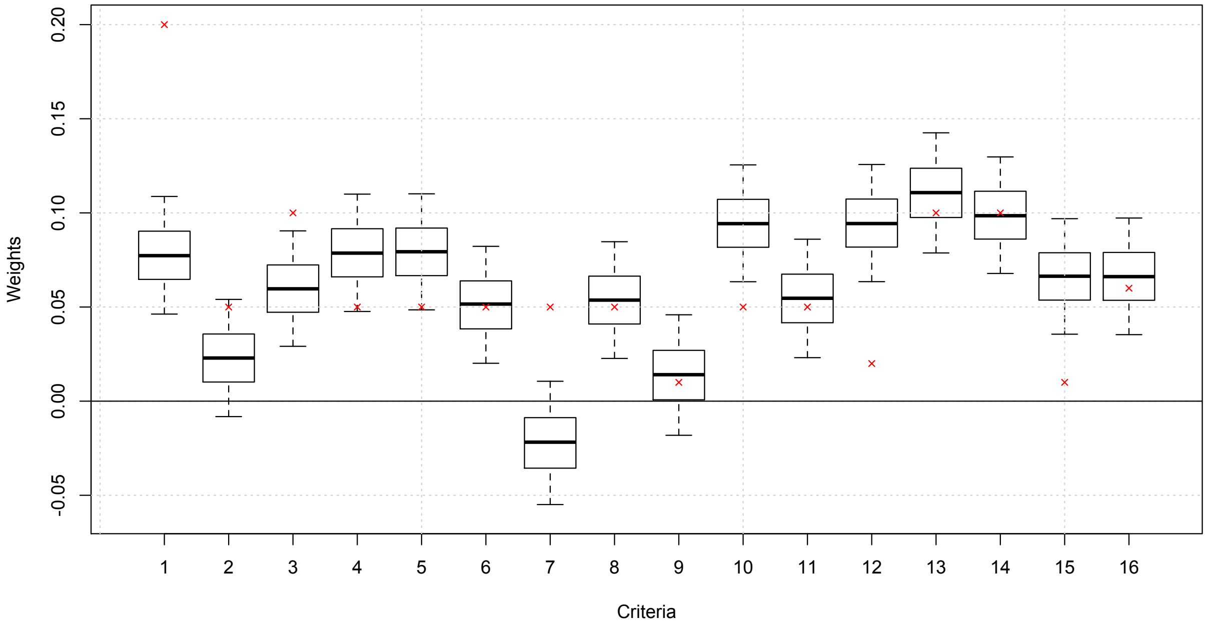


Figure 5.3: Estimated weights for the sixteen criteria in BLVA.

5.2.2.2 Teaching quality defined by these new weights

It could be interesting to consider the teaching quality definition according to the new weights. To do so, the mean weights are used, after being normalized[1]. Table 5.1 depicts these means.

	Criterion	FT Weights (%)	BLVA weights (%)
1	Weighted salary	20	7.4
2	Value for money	5	2.2
3	Careers	10	5.7
4	Aims achieved	5	7.4
5	Placement success	5	7.6
6	Employed at three months	5	4.9
7	Women faculty	5	-2.1
8	Women students	5	5.1
9	Women board	1	1.3
10	International faculty	5	9.1
11	International students	5	5.3
12	International board	2	9.1
13	International mobility	10	10.6
14	International course experience	10	9.5
15	Languages	1	6.3
16	Faculty with doctorates	6	6.3

Table 5.1: List of criteria of the FT MiM ranking, with their FT and mean BLVA weights.

In the introduction, five categories of criteria had been described and were summarized in Table 1.2: future salary, school diversity, career opportunity, globalization and faculty with doctorates. They are re-used in this section. The *future salary* group, composed of the criteria *Weighted salary* (1) and *Value for money* (2), only accounts for 9.6% of the total weights. This

[1]The mean weights are divided by the sum of the absolute value of the sixteen mean weights.

is much less than with the FT weights. This is due to a reduction of weights for both criteria. The future salary thus seems to be less important in terms of teaching quality.
The second category is *school diversity* and consists in the criteria *Women faculty* (7), *Women students* (8), *Women board* (9), *International faculty* (10), *International students* (11), *International board* (12). It receives a global BLVA weight of 27.8% This is higher than with FT weights (23%). It can be seen that despite the fact that women-related criteria obtained much lower weights than before (especially the criterion *Women faculty* (7) that receives a negative weight), the global BLVA weight of this category is valuable as international-related criteria got a much better weight.
Career opportunity is the third group and represents the criteria *Careers* (3), *Aims achieved* (4), *Placement success* (5) and *Employed at three months* (6). It receives a global BLVA weight of 25.6% compared to a global FT weight of 25%. This is nearly the same.
The fourth category is *globalization* and is composed of the criteria *International mobility* (13), *International course experience* (14) and *Languages* (15). From 21% for FT weights, it accounts now for 26.4% of the total BLVA weights. This increase is mostly due to the criterion *Languages* (15) as the first two criteria kept a similar weight.
The last group, *faculty with doctorates*, consists in only one criterion: *Faculty with doctorates* (16). Its weight remained nearly unchanged.
Table 5.2 summarizes what has just been said.

	Category	FT Weight (%)	BLVA Weight (%)
1	Future salary	25	9.6
2	School diversity	23	27.8
3	Career opportunity	25	25.6
4	Globalization	21	26.4
5	Faculty with doctorates	6	6.3

Table 5.2: Summary of the quality defined by BLVA, compared to the FT MiM ranking.

If only criteria that presented the highest BLVA weights are considered, namely *International faculty* (10), *International board* (12), *International mobility* (13) and *International course experience* (14), it can be seen that BLVA believes it is of great importance to have an international school that pushes its students to have international experience.

5.2.3 Discussion

This section was devoted to the development of a new ranking. This was based on a Bayesian Latent Variable Analysis (BLVA). The need for a new ranking methodology was due to the conclusion that FT weights played a key role in the FT MiM ranking. It has been seen in the robustness analysis that uncertainties in weights could greatly affect the ranking, or that principal components of non-weighted and weighted PCAs were very different depending on the use of weights or not. Considering that FT weights were arbitrarily fixed by the Financial Times and that some criticisms of the teaching quality defined by these weights were made [7], BLVA presented the valuable advantage of designing the weights by extracting information embedded in the FT data set.

This attempt to design a new ranking exhibited the fact that some schools presented a very different rank than in the FT MiM ranking. It is the case for WHU Beisheim (school 4) or Indian Institute of Management, Calcutta (school 13) for instance. This new ranking also demonstrated that international mobility-related criteria seemed to be much more important when information embedded in the data set was taken into account. Conversely, the criterion *Weighted salary* (1) presented a much lower score.

Nonetheless, however good the methodology behind BLVA seems to be, this new ranking is based on a data set that consists in an arbitrary choice of criteria. The fact that the criterion *Women faculty* (7) received such a negative BLVA weight is problematic and seems to suggest that there are some issues with the FT data set. Indeed, when the correlations between *Women faculty* (7) and other criteria were analyzed in Section 2.1, it appeared that the pairwise correlation with *Weighted salary* (1) is highly negative (-0.60). Furthermore, *Weighted salary* (1) is quite positively correlated with two other criteria: *Aims achieved* (4) (0.69) and *Placement success* (5) (0.68). This could possibly explain the fact that *Women faculty* (7) does not manage well: in front of three other criteria that seem to be negatively correlated with it, the criterion receives a negative score in BLVA.
If one thinks about a possible relationship between criteria like *Weighted salary* (1) and *Women faculty* (7), it seems logical not to understand why they are related. That a former student would earn less because he/she had more female teachers does not appear very reasonable indeed. However, whether the relationship between the two criteria is coherent or not, the very fact that it does exist most likely affects results of BLVA.

A solution to this problem could be to change the data set, by removing or adding criteria. Certainly, it can be seen that the main problem here is the choice of criteria and their pairwise correlations. Designing a data set with less correlation could thus seem beneficial. Still, this could be hard to push universities to give more information about their alumni or their students. Another possible alternative to keep the same data set could be to prevent some criteria from having negative weights. This would obviously affect the Bayesian analysis and its extraction of information embedded in the data set, but at least, it could somewhat show the characteristics of a teaching quality considered as acceptable to BLVA.

5.3 Budget Allocation Process (BAP)

It has been seen that the weights can have a great impact on the ranking. For instance, the robustness analysis displayed great uncertainties when the weights varied. Beside that, some criticisms of the teaching quality defined by the FT weights were put forward [7]. In the last section, an attempt to resolve this last problem was made by conducting a Bayesian Latent Variable Analysis (BLVA). It sought to create new weights that are designed from information embedded in the data rather than arbitrary choices. Its methodology seemed promising but a shadow was cast over BLVA because the teaching quality defined by its weights were not acceptable.

Given all this, one could be tempted to directly work on weights, by defining them itself. Indeed, the prime focus of a ranking is to help institutions and prospective students to decide the position of a particular school compared to other ones. Different institutions or students could have a distinct definition of a quality and it could seem quite strange to present an unique ranking. They could therefore be able to build a particular ranking based on their expectation.

The Budget Allocation Process (BAP) is an aggregation method where a person is asked to allocate a budget of one hundred points among the sixteen criteria. This technique is one of the recommended aggregation method by the OECD [33]. Besides, this aggregation method could help institutions to better concentrate on what they want to improve in their way of teaching without having to improve specific criteria as some institutions complained about [7]. For instance, the European Union already developed a tool, called *U-Multiank*, where prospective students could choose their future university by using a similar method than BAP.

In the followings of this section, two case studies representing possible students (or university institutions) are defined and the weights are decided using BAP accordingly. The aim of this is to assess the practicality of BAP: do the resulting rankings present distinctive differences or do they look quite similar ?

5.3.1 Case study 1

The first prospective student is a young man[2] that wants to meet very different people during his cursus. The mixity is thus quite important to him. Beside that, he wants to have the ability to travel and work abroad. Naturally, he is eager to step up the ladder in its future company. To a lesser extent, he wants to live well and earn enough money. He does not care about having teachers with PhD degrees.

5.3.2 Case study 2

The second prospective student is a young woman that would pursue a brilliant career and lead a quite comfortable life. She wants to have numerous career opportunities, including abroad. The mixity in her school is not her top priority, as long as teachers with PhD degrees populate her faculty.

5.3.3 Results

Table 5.3 shows a possible list of weights according to the two case studies. Figures 5.4 and 5.5 depict the two resulting rankings. It can be directly noticed that they appear quite dissimilar. While the second ranking seems to be comparable to the FT MiM ranking, the first one presents some tremendous differences with the FT MiM ranking. For instance, the Indian Institute of Management, Ahmedabad (school 17), ranked 17 in the FT MiM ranking, is ranked 58 in the ranking representing the first case study. This could be partly explained by the fact that this school received the highest score for the criterion *Weighted salary* (1), which is low in the first case study.
Another aspect that is worth being noted is the resilience of the University of St Gallen (school 1). Indeed, this school is ranked first in both rankings, besides being the best-performing school in the FT MiM ranking.
On the other hand, it could also be interesting to look at universities belonging to the middle of the ranking. Indeed, it has been shown that it was this part of the ranking that was more subject to variations in the robustness analysis. Consider the schools Solvay Brussels School of Economics and Management (school 39) and Louvain School of Management (school 41) that were investigated in the uncertainty analysis. Their respective ranks in the resulting ranking due to the first case study are 36 and 37. However, when the second case study is studied, their respective ranks become 34 and 29. Besides the fact that they did better in these two rankings than in the FT MiM ranking, it can be observed that in one ranking, one school is better while in the other one, the other school performs better.

5.3.4 Discussion

This section was dedicated to the development of a new ranking based on a Budget Allocation Process (BAP) to assign the weights to criteria. The need for a new methodology to design weights was due to the conclusion that a lot of criticisms were made on the teaching quality defined by the FT weights. Furthermore, even though the BLVA methodology seemed quite

[2] A parallel can easily be done for a university institution.

	Criterion	Weight 1 (%)	Weight 2 (%)
1	Weighted salary	5	16
2	Value for money	4	6
3	Careers	8	16
4	Aims achieved	8	10
5	Placement success	4	6
6	Employed at three months	5	5
7	Women faculty	6	3
8	Women students	6	3
9	Women board	2	1
10	International faculty	6	3
11	International students	6	3
12	International board	2	1
13	International mobility	15	9
14	International course experience	15	9
15	Languages	6	3
16	Faculty with doctorates	2	7

Table 5.3: List of weights of the two case studies. The column *Weight 1* represents the first case study while the column *Weight 2* represents the second one.

promising, its procedure failed to produce a set of weights that was acceptable, due to the design of the FT data set. Thus, BAP focuses on letting the choice to the designer to fully determine the new weights.
Two case studies were created, each one representing an imaginary student that were both quite different. The analysis of the resulting rankings really demonstrated that the weights can heavily modify the overall structure of a ranking.

The main advantage of BAP is that the student or the university institution eager to rank the business schools can do so, establishing his own criteria. This should silence criticisms proclaiming that the teaching quality is not representative of expectations of the population.
Besides, if BAP were used instead of classical school rankings, deans and university administrators should not have to worry anymore about which university sector to fund in order to boost their ranking, because there would not be a single ranking anymore, but rather a list of criteria that a student could weight himself.

Nevertheless, some flaws in BAP exist. The fact that there is no more an unique ranking makes it more difficult for the designer (the Financial Times for the FT MiM ranking) to publicize it. Thus, this kind of ranking, being less lucrative, could be less used, making it a bit utopian.
Furthermore, to be able to define the teaching quality as precisely as possible, more than sixteen criteria would be needed, making the work to create the data set heavier. Besides, the description of the different criteria should be very specific, so that a person wanting to create a ranking know the implications of his choice, but it has been seen in the Introduction that the Financial Times had already some difficulties in explaining correctly its sixteen criteria.

5.4 A better visualization

So far, in this chapter, the problematic behind the definition of the teaching quality as well as issues in the high variation due to uncertainties in weights were considered. From these, only solutions involving new rankings were suggested. It could be worth focusing on an existing ranking, for instance the FT MiM one, and attempt to enhance it in the light of what were

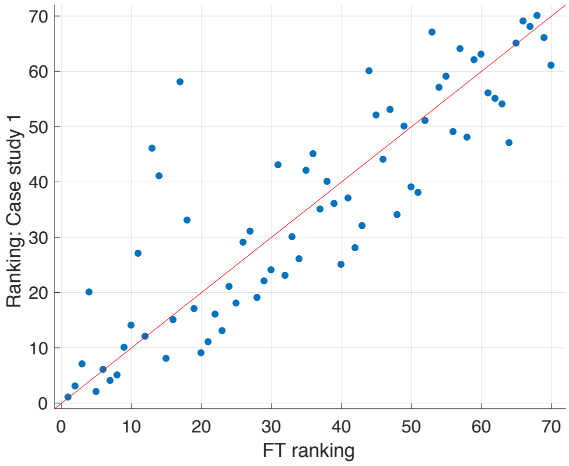


Figure 5.4: Comparison between the FT MiM ranking and the resulting ranking from case study 1.

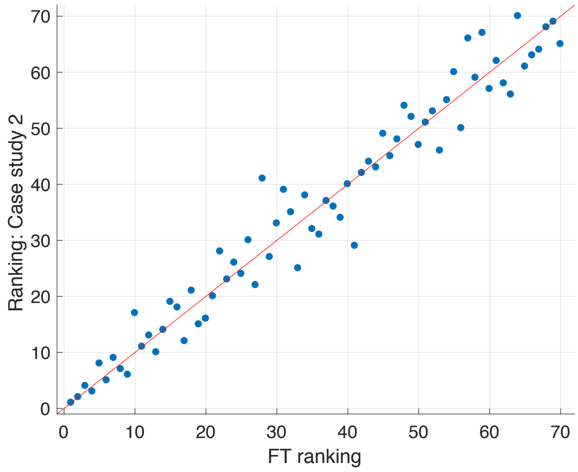


Figure 5.5: Comparison between the FT MiM ranking and the resulting ranking from case study 2.

discussed in this work.

Furthermore, one of the main critique directed to university rankings in general is that they only present a ranked list of schools but they say nothing about the way in which two schools differ or about the local neighbourhoods involving similar schools. It could have been thought from the FT MiM ranking that universities are equally distant from each other when that is clearly not the case, as it was depicted in Figure 2.2 during the reconstruction phase, in Section 2.2. The top ten schools appeared to far outperform the rest of the ranking while the worst-ranked schools came out to be distinctively below the other universities. Meanwhile, the middle of the ranking consisted in more or less similar schools, which lead this part of the ranking to heavily vary when facing uncertainties during the robustness analysis, 2.3.

In this section, the problem of accurately depicting differences between schools is thus tackled. A solution to keep the FT MiM ranking could be to add some information to this ranking. It has been seen in Chapter 4 that the two-dimensional embedding using Ms. JSE preserved quite faithfully local neighbourhoods. An attempt to extend the FT MiM ranking with a two-dimensional projection is then made. Figures 5.4 and 5.6 depict a possible presentation of results of the FT MiM ranking. The two figures are placed in front of each other as it could be in a booklet advertising the ranking. Figure 5.4 presents the FT MiM ranking, together with scores for each school computed in Section 2.2[3], while Figure 5.6 displays the two-dimensional embedding using Ms. JSE that was built in Section 4.5.

5.4.1 Results

The rest of this section is devoted to a commentary of the added value brought by the use of such projection as well as computed z-scores. Most of remarks done below have already been made in previous chapters. First of all, z-scores can be very helpful to understand differences between adjacent universities in the ranking. For instance, the University of St Gallen (school 1) stands far ahead of all the other schools - an information that was not available in the FT MiM ranking. More generally, the top ten schools perform significantly better than the rest of universities. More than thirty points separate the schools IE Business School (school 9) and

[3] As a reminder, a perfect concordance between the FT MiM ranking and the reconstructed one was not met, so the scores are not in strict descending order.

FT rank	Schools	Z-scores
1	University of St Gallen	115.0
2	HEC Paris	97.7
3	Essec Business School	86.3
4	WHU Beisheim	82.4
5	Cems	79.3
6	Esade Business School	78.7
7	ESCP Europe	72.5
8	Rotterdam School of Management	69.4
9	IE Business School	70.9
10	London Business School	54.6
11	HHL Leipzig Graduate School of Management	37.1
12	Universite Bocconi	43.5
13	Indian Institute of Management, Calcutta	18.1
14	EBS Business School	38.0
15	Grenoble Graduate School of Business	40.0
16	Edhec Business School	39.1
17	Indian Institute of Management, Ahmedabad	25.9
18	Mannheim Business School	22.9
19	Imperial College Business School	37.9
20	EMLyon Business School	29.6
21	Ieseg School of Management	34.2
22	WU (Vienna University of Economics and Business)	24.8
23	ESC Rennes	23.9
24	City University: Cass	22.0
25	Telecom Business School	21.1
26	HEC Lausanne	16.8
27	Audencia Nantes	3.7
28	Skema Business School	-1.9
29	Eada Business School Barcelona	12.3
30	Toulouse Business School	4.1
31	Warwick Business School	2.7
32	Montpellier Business School	1.6
33	Stockholm School of Economics	2.2
34	Antwerp Management School	-2.3
35	Kozminski University	1.1
36	Vlerick Business School	-1.9
37	Maastricht University School of Business and Economics	-4.1
38	Copenhagen Business School	0.5
39	Solvay Brussels School of Economics and Management	-10.3
40	Neoma Business School	-7.1
41	Louvain School of Management	-7.8
42	Kedge Business School	-8.4
43	University of Strathclyde Business School	-8.9
44	Shanghai Jiao Tong University	-17.9
45	University College Dublin	-19.7
46	Aalto University	-24.2
47	University of Sydney Business School	-26.8
48	Nova School of Business and Economics	-17.1
49	University of British Columbia	-28.0
50	Catolica Lisbon School of Business and Economics	-29.2
51	ICN Business School	-34.1
52	Tias Business School	-39.0
53	University of Cologne	-38.6
54	IAE Aix-en-Provence	-37.1
55	Alliance Manchester Business School	-48.3
56	St Petersburg State University Graduate School of Management	-45.7
57	Durham University Business School	-49.4
58	University of Bath School of Management	-50.1
59	Bradford University School of Management	-60.1
60	NHH	-51.9
61	Leeds University Business School	-53.2
62	Lancaster University Management School	-52.6
63	Politecnico di Milano School of Management	-49.9
64	La Rochelle Business School	-60.6
65	Tongji University School of Economics and Management	-59.4
66	Nyenrode Business Universiteit	-69.4
67	BI Norwegian Business School	-68.6
68	Warsaw School of Economics	-75.5
69	University of Economics, Prague	-78.8
70	Corvinus University of Budapest	-71.3

Table 5.4: List of schools in the FT MiM ranking, with their calculated z-scores.

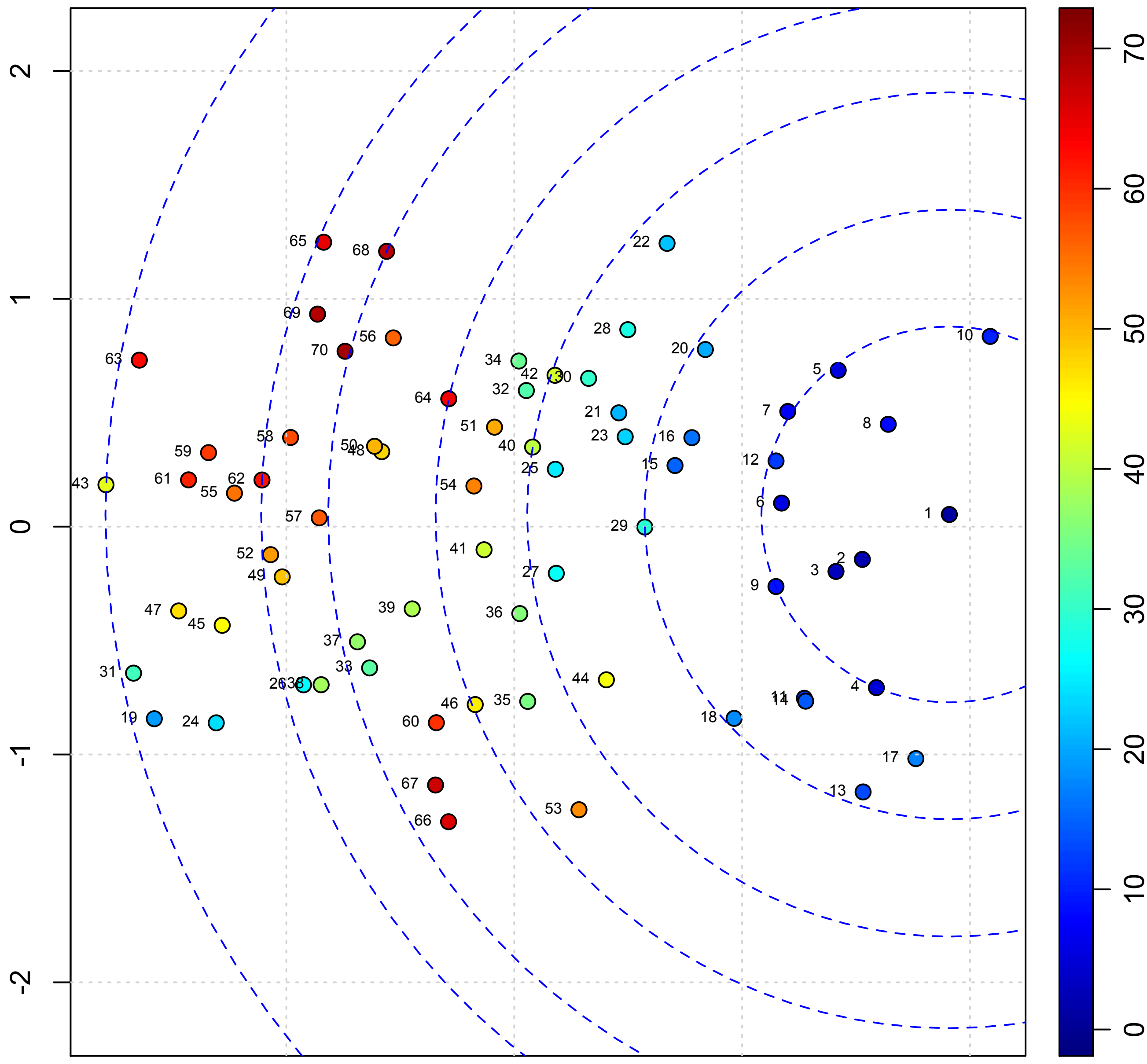


Figure 5.6: Projection of schools in 2D using Ms. JSE coloured with their FT rank.

HHL Leipzig Graduate School of Management (school 11), while these two universities seemed quite close in the FT MiM ranking due to their proximity in the ranking.

On the other hand, a two-dimensional projection can give more information on similarities and differences between schools. Figure 5.6 depicts a two-dimensional embedding using Ms. JSE. For instance, it can be observed that the WHU Beisheim (school 4) seems quite far from the other schools from the top ten. The particularity of this university could not have been seen in z-scores as it got a similar score compared to its neighbours.
In addition, it can be noticed that some well-performing universities (like the schools Imperial College Business School (school 19) and City University: Cass (school 24)) are quite distant from the University of St Gallen (school 1) in the projection while it could not be observed in z-scores.
A last information that could be noted in the projection but not in the FT MiM ranking is the division into two groups of the worst-performing universities. It can be seen that the schools Warsaw School of Economics (school 68), University of Economics, Prague (school 69) and Corvinus University of Budapest (school 70) lie on one side of the projection while the schools Nyenrode Business Universiteit (school 66) and BI Norwegian Business School (school 67) are located on the other side. This phenomenon was deeply explained in the previous chapter, and a parallel with the self-organizing map has even been done.

5.4.2 Discussion

The previous presentation of the FT MiM ranking using z-scores and Ms. JSE projections demonstrated the fact that a ranking could gain a lot of information by depicting data into higher dimensions. Even though the two-dimensional embedding using Ms. JSE provided more knowledge of similarities between schools, z-scores were also important as they only focus on the FT MiM ranking.

One could also think to add more projections to the ranking, like the self-organizing map that proved to be quite helpful in defining clusters of universities. It must be remembered though that the first goal of a university ranking is to summarize the available information on schools by providing a unique number characterizing these universities. The use of the Ms. JSE projection was very valuable, but more projections could make the ranking somewhat confusing, in my opinion. However, other projections like the self-organizing map could still be of great help for university administrators to analyze more thoroughly the position of their school compared to the others.

Future works could thus focus on how to extend a ranking with more than one projection, without overflooding it with too much information.

5.5 Conclusion

In this chapter, three rankings were created to address some problems that were met in this work. These problems were rather diverse and so are the created rankings. The first one adopted the Bayesian Latent Variable Analysis (BLVA). The main goal was to extract information from FT data to create the weights, instead of designing them arbitrarily. Using a Bayesian inference, this ranking also gave us some indications on uncertainty on the resulting ranking. This is actually a tremendous advantage compared to the FT MiM ranking, for it allows the reader to understand if a school is statistically superior to another one or not.
However, the computed weights being based on provided data, the method is heavily dependent on it. It has been seen that a highly-negative pairwise correlation between criteria *Women*

faculty (7) and *Weighted salary* (1) forced the first criterion to have a negative weight. If one thinks in terms of teaching quality defined behind these weights, being less ranked for having a lot of female teachers is somewhat unacceptable.
The methodology of this ranking being quite promising, two solutions were proposed to address this problem of negative weights for sensible criteria. The first one was the more radical choice: modify the data provided so that pairwise correlations are minimized. The second solution was to fix a minimal value for sensible criteria, like gender-related ones.

The second ranking was proposed to address one problem already tackled by BLVA: the teaching quality defined behind the FT weights. If results of BLVA did not give an acceptable ranking, it could be interesting to directly describe the teaching quality so that one can be sure it matches his expectations. The uses of this kind of rankings could be numerous, from the student that wants to know the best university according to his own criteria, to a university that wants to assess itself compared to other schools according to criteria it finds important.
This ranking is based on a Budget Allocation Process (BAP). It consists in distributing a hundred points (that represent the weights) among the different criteria and construct the ranking from it. Two fictional students were created to assess the differences between the rankings created by distinctive people. The main result was that universities' ranks changed dramatically between the two rankings. This showed the importance of the definition of teaching quality lying behind each ranking.
Nonetheless, the main disadvantage of this kind of rankings is that there is not a unique ranking, but rather a multitude. Designers of rankings (like the Financial Times) could not be very fond of BAP because they could not advertise one ranking anymore.

As each of the two previous solutions presented a ranking designed from scratch, the last proposal explains how to enhance the FT MiM ranking. In this work, it was suggested to add z-scores (computed in Section 2.2) and a two-dimensional projection using Ms. JSE (displayed in Section 4.5). It has been analyzed how these two improvements can provide significantly more information than the FT MiM ranking alone. This new ranking tackled the problem that the FT MiM ranking did not represented local neighbourhoods of schools well in an one-dimensional ranking.
The last part of the analysis of this ranking was spent to justify why these enhancements were enough and why not adding a self-organizing map or other projections: the main goal of a university ranking is to summarize information on schools and arrange it in a readable way. Adding more projections would surely increase the quality of the ranking, but it would also blur the principal messages of the ranking.

These three rankings proved that it is certainly feasible to enhance the FT MiM ranking, even though it could be done in many different ways. A combination of these rankings could even be thought of. For instance, a BLVA ranking adjoined to a two-dimensional projection of data could give a lot of information on the incertitude behind the ranking (thanks to BLVA) as well as local neighbourhoods between schools (thanks to Ms. JSE projections).
However, the BLVA ranking, which represented the most promising one, faces a major drawback: the pairwise correlations between non-linked criteria (like *Weighted salary* (1) and *Women faculty* (7)) severely reduced the quality of results. In this work, it was decided to take the FT data as it is, but a future work could clearly be to attempt to enhance the quality of the provided data.

6 Conclusion

6.1 Summary

This work aims to study the relevance of the FT MiM ranking, its possible flaws and to discuss potential improvements. It has been seen that the FT MiM ranking presents some high pairwise correlations between criteria. Even though this phenomenon is more limited than in other rankings, it could still play a role. The fact that the criterion *Weighted salary* (1) (which has the highest FT weight) has multiple high pairwise correlations is particularly significant.
Before undertaking analyses on the ranking, a reconstruction of the FT MiM ranking from the data set provided by the Financial Times has been done, so that it was clear that there was enough information embedded in this data set to pursue further analyses, even though the data was not entirely transparent. Results showed that the reconstructed ranking presented a higher correlation with the FT MiM ranking than in related works [36].
Furthermore, a robustness analysis has been conducted and proved that the FT MiM ranking could vary significantly in the presence of uncertainties, especially when the weights and the normalization methods were modified. This demonstrated that the teaching quality defined by the FT weights was of primordial importance in constructing the ranking. The conception of the teaching quality should thus be taken very seriously as it could heavily affect the ranking.

The second part of this work was devoted to introduce different dimensionality reduction methods. Indeed, in order to study the relevance of the FT MiM ranking, it could be useful to assess local neighbourhoods of each school and projecting data into two or three dimensions could help visualize the connections between schools and how they differ from each other. Actually, one of the principal criticisms done to the FT MiM ranking (and university rankings in general) is that they do not reflect the fact that schools could excel in the ranking in various ways. For instance, they could perform well because they allow their former students to earn a lot of money (via the criterion *Weighted salary (1)*) or because they allow their former students to travel around the world (via the criterion *International mobility* (13)).
Three DR methods were introduced in Chapter 3 and their results presented in Chapter 4 showed that each method brought some distinct information. The interest of Principal Components Analysis (PCA) was that it is a LDR method, thus the linear combinations linking the FT criteria and the principal components could be deeply studied to understand which criteria played a major role in the ranking. The Self-Organizing Map (SOM) seemed appealing as it attempts to preserve neighbourhood relationships between subregions of the manifold. A vector quantization could also easily be undertook on the SOM if the number of SOM nodes is less than the number of schools. The last method was a NLDR state-of-the-art one, called Ms. JSE. It was based on similarity preservation and used multi-scale and shift-invariant techniques to respectively address the problems of neighbourhoods of multiple sizes and norm concentration. As these three DR methods presented somewhat different characteristics, it has been concluded that the use of different visualizations in parallel could then be necessary to fully understand the neighbourhoods underpinning the relationships between schools.

Chapter 5 was devoted to potential improvements to the FT MiM ranking. It served as a pre-conclusion by gathering results found in this work and by attempting to enhance the FT MiM ranking. Following the conclusions that the weights were a source of great uncertainty in the ranking (in Section 2.2), a first solution was proposed via a Bayesian Latent Variable Analysis (BLVA). Instead of arbitrarily determining the weights, it computed them based on information extracted from the data. Furthermore, by using a Bayesian inference, it was able to suggest uncertainty on the resulted ranking. However, it was shown that the teaching quality defined by BLVA was not acceptable and was due to weird pairwise correlations in the data provided by the Financial Times.
BLVA could have silenced some criticisms but the fact that a curious teaching quality is still defined and that it could not be the same than our own view of what should a school be, a second solution is proposed via the Budget Allocation process (BAP). It allows anyone to design its own weights (and thus its own teaching quality) and the ranking would be created from these weights. An advantage of this technique is that universities should not be worried anymore to follow a teaching quality defined by ranking designers. Nonetheless, the main disadvantage of this kind of rankings is that there is not an unique ranking, but rather a multitude. Designers of rankings (like the Financial Times) could not be very fond of BAP because they could not advertise one ranking anymore.
This previous solution being a bit utopian, the last improvement is more down-to-earth. Instead of designing the ranking from scratch, one could add visualization tools to it so that the relationships between schools or local neighbourhoods formed among them could be highlighted. The z-scores (computed in Section 2.2) and a two-dimensional projection using Ms. JSE was added to the FT MiM ranking and it has been demonstrated how only these two enhancements could greatly increase the amount of information given to the public.

In general, in this work, the relevance of the FT MiM ranking was challenged. It has been seen that while the FT data presented some high pairwise correlations, these were less pervasive than in other rankings [44]. Nevertheless, this study allowed us to understand that a lot of somewhat crucial information was hidden by the FT MiM ranking and that the use of two-dimensional projections could be of great help to grasp the pertinence of the FT MiM ranking. However, the design of a new ranking is something very complex and the balance between available information and complexity of the ranking should not be underestimated. There is no use to add multiple projections to the ranking if the latter becomes too complicated to be understable.

6.2 Future works

For the sake of conciseness, this work limited its search by only focusing on certain aspects. Firstly, the aim of this work was to be general. Possible future developments could be to concentrate more on certain schools to compare them more precisely. University administrators could find very helpful to conduct an extensive comparison on their university and other similar schools. It could describe them how to spend a budget in order to efficiently go up in the ranking.

Secondly, the data provided by the Financial Times was used as it is, whereas it could be worth taking a look at this data set to attempt to enhance it. Indeed, regarding results of BLVA in Chapter 5, it could be questioned if the FT data should not be modified to represent a more "uncorrelated" situation.

Furthermore, only the FT data from 2014 was considered in this work. It could also be interesting to analyze the ranking over time. Indeed, it has been seen that the FT MiM ranking was particularly prone to variability when the weights and normalization methods fluctuated.

The same phenomenon could occur when concentrating on different years, which could decrease the impact of the FT MiM ranking if it can not be stable from year to year. Especially, editions from 2013 to 2016 could be interesting to look at as they share the same FT methodology.

Finally, this work focused on a ranking constructed by the Financial Times, hence its title. It could be very appealing to perform the same methods conducted in this work on other rankings, like for example the *U.S. News & World Report*, the *America's Best Colleges* or the *QS World Universities Rankings.* That way, for instance, an extensive comparison of benefits of using NLDR projections could be completed.

A Data sets

A.1 List of schools

FT rank	Schools	Country
1	University of St Gallen	Switzerland
2	HEC Paris	France
3	Essec Business School	France
4	WHU Beisheim	Germany
5	Cems	France
6	Esade Business School	Spain
7	ESCP Europe	France
8	Rotterdam School of Management	Netherlands
9	IE Business School	Spain
10	London Business School	UK
11	HHL Leipzig Graduate School of Management	Germany
12	Universite Bocconi	Italy
13	Indian Institute of Management, Calcutta	India
14	EBS Business School	Germany
15	Grenoble Graduate School of Business	France
16	Edhec Business School	France
17	Indian Institute of Management, Ahmedabad	India
18	Mannheim Business School	Germany
19	Imperial College Business School	UK
20	EMLyon Business School	France
21	Ieseg School of Management	France
22	WU (Vienna University of Economics and Business)	Austria
23	ESC Rennes	France
24	City University: Cass	UK
25	Telecom Business School	France
26	HEC Lausanne	Switzerland
27	Audencia Nantes	France
28	Skema Business School	France
29	Eada Business School Barcelona	Spain
30	Toulouse Business School	France
31	Warwick Business School	UK
32	Montpellier Business School	France
33	Stockholm School of Economics	Sweden
34	Antwerp Management School	Belgium
35	Kozminski University	Poland
36	Vlerick Business School	Belgium
37	Maastricht University School of Business and Economics	Netherlands

FT rank	Schools	Country
38	Copenhagen Business School	Denmark
39	Solvay Brussels School of Economics and Management	Belgium
40	Neoma Business School	France
41	Louvain School of Management	Belgium
42	Kedge Business School	France
43	University of Strathclyde Business School	UK
44	Shanghai Jiao Tong University	China
45	University College Dublin	Ireland
46	Aalto University	Finland
47	University of Sydney Business School	Australia
48	Nova School of Business and Economics	Portugal
49	University of British Columbia	Canada
50	Catolica Lisbon School of Business and Economics	Portugal
51	ICN Business School	France
52	Tias Business School	Netherlands
53	University of Cologne	Germany
54	IAE Aix-en-Provence	France
55	Alliance Manchester Business School	UK
56	St Petersburg State University Graduate School of Management	Russia
57	Durham University Business School	UK
58	University of Bath School of Management	UK
59	Bradford University School of Management	UK
60	NHH	Norway
61	Leeds University Business School	UK
62	Lancaster University Management School	UK
63	Politecnico di Milano School of Management	Italy
64	La Rochelle Business School	France
65	Tongji University School of Economics and Management	China
66	Nyenrode Business Universiteit	Netherlands
67	BI Norwegian Business School	Norway
68	Warsaw School of Economics	Poland
69	University of Economics, Prague	Czech Republic
70	Corvinus University of Budapest	Hungary

Table A.1: List of schools in the FT MiM ranking.

A.2 List of criteria

	Criterion	Weights (%)	Units
1	Weighted salary	20	US$
2	Value for money	5	rank
3	Careers	10	rank
4	Aims achieved	5	rank
5	Placement success	5	rank
6	Employed at three months	5	%
7	Women faculty	5	%
8	Women students	5	%
9	Women board	1	%

	Criterion	Weights (%)	Units
10	International faculty	5	%
11	International students	5	%
12	International board	2	%
13	International mobility	10	rank
14	International course experience	10	rank
15	Languages	1	
16	Faculty with doctorates	6	%

Table A.2: List of criteria of the FT MiM ranking.

A.3 Raw data

Raw data taking up a lot of space, they will be split into three tables. The first one displays the first six criteria, the second one the following six criteria and the last table the rest of the criteria.

FT rank	Weighted salary	Value for money rank	Careers rank	Aims achieved rank	Placement success rank	Employed at three months
1	79572	1	30	1	4	88.00
2	78825	28	11	5	5	63.05
3	77451	40	8	8	9	67.34
4	93948	5	9	3	1	90.00
5	63468	2	44	12	35	57.66
6	65647	29	6	7	16	93.12
7	65404	48	25	11	11	52.29
8	67696	11	52	9	23	84.48
9	74263	46	2	10	18	83.60
10	70414	27	68	15	13	94.08
11	85238	26	45	2	7	79.20
12	63986	38	17	19	15	39.48
13	83085	41	1	18	2	99.00
14	81734	25	24	27	6	86.00
15	56048	49	13	37	59	66.00
16	56651	54	18	32	19	91.18
17	94721	55	3	30	3	95.00
18	78088	7	50	6	27	74.26
19	54031	63	16	4	10	71.40
20	54771	61	29	20	24	57.95
21	48639	67	10	57	43	70.52
22	56839	17	53	56	68	94.09
23	49162	60	5	58	66	73.87
24	53734	45	46	13	31	46.00
25	50633	23	14	29	25	48.60
26	54718	10	31	41	69	85.36
27	55174	33	4	53	20	93.06
28	48971	57	43	60	45	63.00
29	54290	44	7	24	40	77.44
30	49381	59	21	62	46	48.60

FT rank	Weighted salary	Value for money rank	Careers rank	Aims achieved rank	Placement success rank	Employed at three months
31	58963	66	40	39	22	67.90
32	44295	64	32	59	34	92.15
33	58410	3	41	14	42	64.00
34	45076	19	34	51	30	77.40
35	56621	12	33	26	28	87.22
36	57768	9	19	52	14	93.06
37	56871	14	54	31	62	76.80
38	56470	6	37	23	70	72.80
39	52766	15	56	38	17	88.27
40	49162	52	20	54	29	73.92
41	49329	21	15	33	37	67.89
42	46708	58	27	63	44	73.87
43	41790	70	12	40	41	71.89
44	62797	4	55	28	8	100.00
45	56042	13	63	46	49	47.20
46	51529	16	61	25	47	81.00
47	53452	42	39	45	26	46.17
48	42562	39	69	68	21	80.75
49	48212	30	59	65	36	82.32
50	39062	43	66	47	12	95.06
51	44041	62	42	55	39	47.60
52	47918	51	60	34	55	83.52
53	65463	32	23	43	53	21.08
54	47562	22	28	69	57	47.20
55	45067	53	49	49	60	56.98
56	40025	35	35	48	48	57.62
57	46817	36	70	50	38	51.24
58	36900	68	57	35	54	58.50
59	40785	50	47	64	58	32.68
60	52392	8	67	21	56	40.67
61	37185	69	38	36	50	56.70
62	40507	56	58	16	52	52.48
63	41946	37	22	44	32	46.02
64	39770	65	62	70	64	68.89
65	35547	34	26	61	61	99.00
66	54861	47	51	22	33	60.68
67	50998	31	65	17	51	62.56
68	38260	20	36	67	63	37.84
69	36177	24	48	66	67	83.30
70	39640	18	64	42	65	61.60

Table A.3: List of raw data for criteria 1 to 6.

FT rank	Women faculty	Women students	Women board	International faculty	International students	International board
1	11	46	25	77	92	67
2	22	45	13	65	42	65
3	30	45	14	51	31	52

FT rank	Women faculty	Women students	Women board	International faculty	International students	International board
4	19	39	14	21	21	26
5	33	48	31	98	94	95
6	31	39	18	35	72	82
7	36	46	35	67	71	52
8	20	48	30	42	67	30
9	35	38	28	55	65	82
10	24	40	30	86	97	75
11	20	27	10	25	25	15
12	36	41	14	27	31	45
13	18	24	19	1	1	0
14	16	38	20	26	28	0
15	43	44	53	44	92	53
16	32	52	17	39	35	67
17	17	22	20	3	0	20
18	36	45	20	19	20	20
19	31	55	38	89	84	50
20	33	58	0	50	32	83
21	40	47	10	81	34	70
22	35	51	33	21	44	72
23	36	54	60	84	52	50
24	30	49	47	70	93	53
25	50	50	36	50	31	45
26	27	50	27	80	48	55
27	41	27	21	40	14	64
28	44	55	32	38	33	55
29	32	33	43	48	88	43
30	40	53	40	41	33	30
31	36	61	12	76	92	12
32	46	51	53	42	35	33
33	23	40	7	31	58	0
34	31	49	20	28	55	90
35	32	49	18	22	18	64
36	29	30	17	24	13	92
37	17	37	21	51	65	64
38	32	46	27	38	54	9
39	17	32	41	38	9	55
40	48	54	12	43	28	0
41	33	30	27	23	8	27
42	23	50	33	40	37	0
43	37	49	35	32	84	47
44	29	38	14	3	25	41
45	31	42	20	47	56	41
46	35	46	43	18	15	43
47	36	63	12	30	47	0
48	44	58	33	29	33	33
49	22	59	22	79	63	19
50	33	46	29	40	32	24
51	41	51	15	44	17	15
52	28	46	17	42	58	0

FT rank	Women faculty	Women students	Women board	International faculty	International students	International board
53	19	46	50	6	9	10
54	40	61	18	19	23	27
55	34	56	18	33	94	18
56	52	70	11	2	18	28
57	36	65	38	64	89	38
58	33	61	31	63	74	12
59	41	45	36	29	92	36
60	24	38	36	26	11	9
61	40	56	40	43	95	47
62	30	63	25	46	91	31
63	25	36	33	0	82	73
64	41	50	17	31	12	33
65	38	60	17	4	19	58
66	20	38	11	25	16	11
67	25	42	62	30	11	12
68	44	61	67	1	11	11
69	50	67	44	9	27	28
70	43	68	11	9	13	61

Table A.4: List of raw data for criteria 7 to 12.

FT rank	International mobility rank	International course experience rank	Languages	Faculty with doctorates
1	1	4	1	100
2	6	10	2	100
3	12	14	2	98
4	40	33	0	100
5	5	7	2	90
6	3	18	1	92
7	13	3	2	95
8	2	6	2	100
9	24	41	1	96
10	8	47	1	100
11	52	22	2	100
12	4	17	2	89
13	60	56	0	99
14	53	30	0	100
15	7	15	1	80
16	18	13	1	87
17	61	55	0	99
18	57	35	0	85
19	11	66	0	100
20	22	1	2	95
21	15	8	2	98
22	20	5	2	95
23	14	11	1	81
24	10	46	0	96
25	26	32	1	76
26	31	48	0	100

FT rank	International mobility rank	International course experience rank	Languages	Faculty with doctorates
27	58	23	2	81
28	17	2	1	76
29	23	24	1	58
30	29	9	2	92
31	30	68	1	100
32	42	19	2	94
33	39	36	2	92
34	45	20	1	85
35	66	34	1	88
36	56	40	1	90
37	27	42	1	97
38	32	50	0	92
39	37	29	2	98
40	47	21	2	75
41	49	25	2	100
42	28	16	2	91
43	16	59	0	86
44	70	57	1	90
45	21	67	0	100
46	64	37	2	94
47	9	68	0	84
48	25	26	2	100
49	33	60	0	98
50	55	28	2	98
51	34	27	2	78
52	41	58	0	95
53	65	49	1	85
54	43	39	1	88
55	38	64	0	88
56	44	31	1	92
57	54	63	0	93
58	46	43	0	99
59	35	68	0	82
60	63	38	1	93
61	51	65	0	82
62	36	62	0	88
63	19	61	0	60
64	50	12	2	71
65	69	45	1	87
66	62	53	0	61
67	67	51	0	70
68	68	52	1	95
69	59	54	1	71
70	48	44	1	80

Table A.5: List of raw data for criteria 13 to 16.

A.4 Basic characteristics of data

It should be noted that in Table A.6, ranked criteria were replaced by normalized values as explained in Section 2.2. Also, there are two entries for each gender-related criteria. The first one is the one found in the FT data set while the second one (marked with an asterisk symbol) represents statistical characteristics of values after ranging from 0 (minimum) to 50 (maximum) as the Financial Times gives a higher score for schools having a 50:50 (male/female) composition [?].

	Criterion	Max value	Min value	Mean	Standard deviation
1	Weighted salary	94’721	35’547	55’376	14’032
2	Value for money	2.355	-2.367	0.004	0.982
3	Careers	2.390	-2.379	0.007	0.992
4	Aims achieved	2.364	-2.390	-0.002	0.987
5	Placement success	2.371	-2.385	0.001	0.989
6	Employed at three months	100	21	69.992	18.614
7	Women faculty	52	11	32.229	9.294
7 (*)	Women faculty	50	11	32.171	9.182
8	Women students	70	22	47.414	10.772
8 (*)	Women students	50	22	41.329	6.822
9	Women board	67	0	27.014	14.127
9 (*)	Women board	50	0	25.729	11.745
10	International faculty	98	0	39.043	23.933
11	International students	97	0	45.129	29.806
12	International board	95	0	39.543	25.996
13	International mobility	2.374	-2.397	-0.009	0.990
14	International course experience	2.378	-1.785	0.019	0.967
15	Languages	2	0	1	0.816
16	Faculty with doctorates	100	58	89.543	10.402

Table A.6: Basic characteristics of data.

Bibliography

[1] Academic Ranking of World Universities. http://www.shanghairanking.com/. Accessed: 8 June 2017.

[2] Aguillo, I. F., Bar-Ilan, J., Levene, M., and Ortega, J. L. Comparing university rankings. *Scientometrics 85*, 1 (2010), 243–256.

[3] Belsley, D. A., Kuh, E., and Welsch, R. E. *Regression diagnostics: Identifying influential data and sources of collinearity*, vol. 571. John Wiley & Sons, 2005.

[4] Bishop, C. M., Svensén, M., and Williams, C. K. Gtm: The generative topographic mapping. *Neural computation 10*, 1 (1998), 215–234.

[5] Chen, L., and Buja, A. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association 104*, 485 (2009), 209–219.

[6] Demartines, P., and Hérault, J. Cca:" curvilinear component analysis". In *15° Colloque sur le traitement du signal et des images, FRA, 1995* (1995), GRETSI, Groupe d'Etudes du Traitement du Signal et des Images.

[7] Devinney, T., Dowling, G. R., and Perm-Ajchariyawong, N. The financial times business schools ranking: What quality is this signal of quality? *European Management Review 5*, 4 (2008), 195–208.

[8] Dill, D. D., and Soo, M. Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher education 49*, 4 (2005), 495–533.

[9] Dreyfus, G. *Neural networks: methodology and applications.* Springer Science & Business Media, 2005.

[10] European Union. https://www.usnews.com/rankings. Accessed: 8 June 2017.

[11] European Union. http://www.milkround.com/staticpages/12780/good-university-guide/. Accessed: 8 June 2017.

[12] European Union. http://www.umultirank.org/. Accessed: 8 June 2017.

[13] Financial Times. http://rankings.ft.com/businessschoolrankings/rankings. Accessed: 8 June 2017.

[14] Financial Times. http://rankings.ft.com/businessschoolrankings/masters-in-management-2014, 2014. Accessed: 8 June 2017.

[15] Financial Times. http://rankings.ft.com/businessschoolrankings/masters-in-management-2016, 2016. Accessed: 8 June 2017.

[16] Goldstein, H., and Spiegelhalter, D. J. League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* (1996), 385–443.

[17] Guarino, C., Ridgeway, G., Chun, M., and Buddin, R. Latent variable analysis: A new approach to university ranking. *Higher Education in Europe 30*, 2 (2005), 147–165.

[18] Guyon, I., and Elisseeff, A. An introduction to variable and feature selection. *Journal of machine learning research 3*, Mar (2003), 1157–1182.

[19] Haykin, S. S. *Neural networks and learning machines*, vol. 3. Pearson Upper Saddle River, NJ, USA:, 2009.

[20] Hinton, G., and Roweis, S. Stochastic neighbor embedding. In *NIPS* (2002), vol. 15, pp. 833–840.

[21] Hopwood, A. G. Comment: The rankings game: Reflections on devinney, dowling and perm-ajchariyawong. *European Management Review 5*, 4 (2008), 209–214.

[22] Knott, M., and Bartholomew, D. J. *Latent variable models and factor analysis.* No. 7. Edward Arnold, 1999.

[23] Kohonen, T. Self-organized formation of topologically correct feature maps. *Biological cybernetics 43*, 1 (1982), 59–69.

[24] Köksalan, M., Büyükbaşaran, T., Özpeynirci, Ö., and Wallenius, J. A flexible approach to ranking with an application to mba programs. *European Journal of Operational Research 201*, 2 (2010), 470–476.

[25] Kutner, M. H., Nachtsheim, C., and Neter, J. *Applied linear regression models.* McGraw-Hill/Irwin, 2004.

[26] Lee, J. A. Type 1 and 2 mixtures of divergences for stochastic neighbor embedding. In *Proceedings of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (2012), pp. 525–530.

[27] Lee, J. A. Unsupervised dimensionality reduction: State of the art, quality assessment, and scalability. Molecular imaging, radiotherapy and oncology, 2016.

[28] Lee, J. A., Peluffo-Ordóñez, D. H., and Verleysen, M. Multi-scale similarities in stochastic neighbour embedding: Reducing dimensionality while preserving both local and global structure. *Neurocomputing 169* (2015), 246–261.

[29] Lee, J. A., Renard, E., Bernard, G., Dupont, P., and Verleysen, M. Type 1 and 2 mixtures of kullback–leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing 112* (2013), 92–108.

[30] Lee, J. A., and Verleysen, M. *Nonlinear dimensionality reduction.* Springer Science & Business Media, 2007.

[31] Lukman, R., Krajnc, D., and Glavič, P. University ranking using research, educational and environmental indicators. *Journal of Cleaner Production 18*, 7 (2010), 619–628.

[32] Maaten, L. v. d., and Hinton, G. Visualizing data using t-sne. *Journal of Machine Learning Research 9*, Nov (2008), 2579–2605.

[33] Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E. Handbook on constructing composite indicators.

[34] Ortmans, L. Key and methodology for masters in management 2016. `https://www.ft.com/content/aa2b8c86-6f1a-11e6-9ac1-1055824ca907`, 2016. Accessed: 4 March 2017.

[35] Pardoe, I. Lecture notes in "stat 501: Regression methods", 2017.

[36] Proudlove, N. Cracking the rankings part (i): Understanding the financial times mba rankings. *OR insight 25*, 4 (2012), 221–240.

[37] QS World Universities Rankings. `https://www.topuniversities.com/university-rankings/world-university-rankings/2016`. Accessed: 8 June 2017.

[38] Ritter, H., Martinetz, T., Schulten, K., Barsky, D., Tesch, M., and Kates, R. *Neural computation and self-organizing maps: an introduction*. Addison-Wesley Reading, MA, 1992.

[39] Roweis, S. T., and Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *science 290*, 5500 (2000), 2323–2326.

[40] Saisana, M., d'Hombres, B., and Saltelli, A. Rickety numbers: Volatility of university rankings and policy implications. *Research policy 40*, 1 (2011), 165–177.

[41] Saisana, M., Saltelli, A., and Tarantola, S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 168*, 2 (2005), 307–323.

[42] von der Malsburg, C. Self-organization of orientation sensitive cells in the striate cortex. *Biological Cybernetics 14*, 2 (1973), 85–100.

[43] Wang, J. *Geometric structure of high-dimensional data and dimensionality reduction*. Springer, 2011.

[44] Webster, T. J. A principal component analysis of the us news & world report tier rankings of colleges and universities. *Economics of Education Review 20*, 3 (2001), 235–244.

[45] Wedlin, L. The role of rankings in codifying a business school template: Classifications, diffusion and mediated isomorphism in organizational fields. *European Management Review 4*, 1 (2007), 24–39.

[46] Zimmerman, J. L. Can american business schools survive?

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve **www.uclouvain.be/epl**