Arbenore Luzhnica, Bsc.

Competency-Based Network Diagram for Visualising Students' Competencies

Master Thesis

for the attainment of the degree of

Master of Telematics

submitted to

Graz University of Technology

Supervisor:
Ebner, Martin, Priv.-Doz. Dipl.-Ing. Dr.techn.

Advisor:
Dr. Michael D. Kickmeier-Rust

Institute of Interactive Systems and Data Science

Graz, April 2018

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit/Diplomarbeit/Dissertation identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis/diploma thesis/doctoral dissertation.

Datum / Date

Unterschrift / Signature

Abstract

This thesis aims at providing visual means of representation of learners gained competences with respect to the taught topics. Its main goal is to present an easy overview of the learning or gained competences, identify strengths and weaknesses of collective groups (e.g. classrooms) as well as individual learners (students) in order to allow the teachers to adjust teaching strategies. This work is mainly focused on the presentation of information as well as on the implementation of such visual presentation. It first presents a concept for visualisation of students' competencies and their inner relationships for which a concept of mathematical and probabilistic representation of students data is created. Finally, it implements the concept using modern web technologies. The end result is a competency-based hierarchical network diagram for presenting students competencies and their achievement during the teaching/process.

Zusammenfassung

Diese Arbeit zielt darauf ab, visuelle Repräsentationsmöglichkeiten der erworbenen Kompetenzen der Lernenden in Bezug auf die gelehrten Themen zu bieten. Das Hauptziel ist es, einen leichten Überblick über das Lernen oder die erworbenen Kompetenzen zu geben, Stärken und Schwächen kollektiver Gruppen (z. B. Klassenräume) sowie einzelner Lernender (Schülerinnen und Schüler) zu identifizieren, damit die Lehrkräfte ihre Lehrstrategien anpassen können. Diese Arbeit konzentriert sich hauptsächlich auf die Präsentation von Informationen sowie auf die Umsetzung einer solchen visuellen Präsentation. Es stellt zunächst ein Konzept zur Visualisierung der Kompetenzen und innerer Beziehungen von Studierenden dar, mit dem ein Konzept der mathematischen und probabilistischen Repräsentation von Schülerdaten erstellt wird. Schließlich implementiert die Arbeit das Konzept mit modernen Web-Technologien. Das Endergebnis ist ein kompetenzbasiertes hierarchisches Netzwerkdiagramm zur Darstellung der Kompetenzen der Studierenden und deren Lernerfolge während des Unterrichts.

Acknowledgements

I would first like to express my gratitude to my supervisor Martin Ebner. I am grateful for his support, guidance, patience, positive appreciation and counsel during this process. My sincere thanks to my advisor Michael Kickemeier-Rust, for the continuous support, patience and the helpful critiques not only during the progression of this work but also for my almost four-year employment at the Institute of Interactive Systems and Data Science, respectively at the Cognitive Science Section (CSS) department. I would also like to acknowledge my colleagues in this group for their assistance during the time I was part of the group.

Nobody has been more important to me in the pursuit of this project than the members of my family. I would like to thank my beloved parents Suzana and Shani, my siblings Mimoza, Arber and Kaltrina, whose love and support sustained me throughout this journey. A special thank goes to Sahadete and other family members Nora, Dafina and Enxhell. I thank Donika for always being there and listening to me!

Finally, I like to express my very profound gratitude to my lovely and supportive husband Granit, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. Most importantly, to my little Jon. Thank you for coming in our lives and providing me with unending love and inspiration!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Latest technological advancements have touched and enhanced almost every aspect of our life including learning. The online accessibility of information, learning materials and especially the proliferation of the video based massive online open courses (MOOCs) [Khalil and Ebner, 2013b, Khalil and Ebner, 2013a] have drastically expanded learning opportunities to enhance educational performance for learners. When talking about education, one carefully consider also other aspects of it, such as teachers and environment. While numerous tools and online resources are made freely or commercially for learners, tools that target teaching especially in classroom settings are rather rare. Unfortunately, in most of the mainstream schools, technological support and enhancement for teaching in classrooms have not evolved along with other technological advancements.

Following such a technological gap in teaching, researchers around the world have been consistently investigating and proposing technological means to better support teachers. The use of digital traces of students in combination with data analytics and visualisation has emerged in a new research field commonly referred as **learning analytics**. It aims at providing an overall overview as well identifying specific aspects of the learning process in order to make it more effective. For instance, teachers could use utilise learning analytics in order to evaluate the strengths and weaknesses of each classroom with respect to a given topic in order to adjust teaching strategy, revisit or re-explain particular parts of a material in order to maximise the knowledge gained by learners.

This thesis aims at providing visual means of representation of learners gained competences with respect to the taught topics. Its main goal is to present an easy overview of the learning or gained competences, identify strengths and weaknesses of collective groups (e.g. classrooms) as well as individual learners (students) in order to allow the teachers to adjust teaching strategies. This work is mainly focused on the presentation of information as well as on the implementation of such visual presentation. It first presents a concept for visualisation of students' competencies and their inner relationships for which a concept of mathematical and probabilistic representation of students data is created. Finally, it implements the concept using modern web technologies. The end result is a competency-based hierarchical network diagram for presenting students competencies and their achievement during the teaching/process.

1.1 LEA's BOX

This work was done within the frame of the LEA's Box (Learning Analytics Box) project where I had a chance to be part of it, by contributing to developing the visualisations aspects of it. It is an already completed multi-partner research and development project funded by the European Commission.

LEA's BOX stands for LE-Learning Analytics Box, which offers educators a box containing several learning analytics tools for [Kickmeier-Rust and Albert, 2016a]:

- performing competence-centred and multi-source reasoning algorithms based on psycho-pedagogical models such as Competence-based Knowledge Space Theory (CbKST) and Formal Concept Analysis (FCA)
- intelligent model-based argumentation services
- innovative visualisation techniques

In daily school practice there is only scattered, heterogeneous, and incomplete data, which makes Learning Analytics on the class level hard. Lea's Box proposes to use a central competence (or curriculum) model as an anchor. The origin of CbKST lies in the field of the intelligent tutorial system where found application also in the contexts of educational games and the formative assessment and feedback. FCA try to formalise concept and concept hierarchies where its origin comes from applied mathematics. The application of FCA methodology takes place in different fields such as knowledge representation, visualisation, analysing of data, and data mining

[Kickmeier-Rust, 2016]. With the CbKST logic of separating latent competencies from observable performance, it aims at linking entire collections of evidences that can be obtained from real classrooms to this central anchor model. For this, the project developed and provided technical solutions which enable such a functionality [Kickmeier-Rust et al., 2016a].

Moreover, typical analytics dashboards are not always suitable and have clear weaknesses in reflecting learning processes. Hasse diagrams [Birkhoff, 1948] hold various potentially useful information for teachers. Specifically, they allow making clear recommendations (the so-called outer fringes [Kickmeier-Rust et al., 2015]), and they allow predictions. A prominent aspect of it is the Open learner Modelling [Ginon et al., 2016, Dimitrova et al., 2007] (OLM). The purpose of the OLM in Lea's Box is:

- making the results of analytics open to learners in a comprehensible way (e.g., in the form of Smileys [Johnson et al., 2016]) and most importantly

- constructing the processes including underlying hidden analytics algorithms in a way that would result in more transparent and understandable information. With the persuadable (or negotiated) learner models, developed in the project, learners are allowed to also disagree with the results and to negotiate them, for example by adding further evidence.

Learning Analytics tools in LEA's Box are designed to use data collected from multiple sources in order to monitor competency development. Through such analytics, the competencies students have, or the competencies they lack can be easily identified. Data from different sources are collected, analysed and then presented using comprehensible visualisation, including the one I will explain in this thesis.

1.2 The Structure of this Thesis

The Related Work follows the remainder of the thesis. Given that this thesis cross paths with different disciplines and research areas, the related work is divided in Information Visualisation (Chapter 2), Learning Analytics (Chapter 3) and Competence Oriented Learning Analytics (Chapter 4). The Related Work is then followed by the Concept (Chapter 5) which describes in great details the conceptual aspects of the competency based hierarchical network diagram, the mathematical modelling of how the information is encoded. The next chapter (Chapter 6) describes the implementation of the concept and the user interface and interaction design.

Last but not least, a summary and discussion (Chapter 7) are provided where I discuss the overall thesis, its challenges, limitations as well as recommendations for future work.

Chapter 2

Information Visualisation

2.1 Overview

According to Frank [Niles, 2011]:

"*Visualization works because neurons in our brains – those electrically excitable cells that transmit information – interpret imagery as equivalent to real-life action. When we visualize an act, the brain generates an impulse that tells our neurons to 'perform' the movement. This creates a new neural pathway – clusters of cells in our brain that work together to create memories or learned behaviours – that primes our body to act in a way consistent with what we imagined.*"[Niles, 2011].

One concern in the field the information society is the generation of the massive amount of data (big data), where at the end of the day, this data has to be comprehensible somehow. A good solution for gaining a quick insight from that massive amount of data is an appropriate visualisation. Visualisations have been used for centuries, where people have used graphs trying to transmit in this way different kind of information. Nowadays, with evolving technology, the process of visualisation became more prevalent thus finding universal application in almost every section of our lives. Unlike in the past, where visualisations are made mostly by hand (manually), recently this process holds associated to each other computer and data, which means that first, the data needs to be computed before being visualised. Today almost every one of us who are impacted by the modern technology, use data visualisations in a daily routine, without even realising it. Let's take an example and show that how a simple visual can make our life easier. Let's say you want to decide how to get dressed on a particular day, but you don't know how first you need to check a weather forecast. If you are using a smartphone, which have a weather app and if you see an icon of the sun on it, you will predict that is going to be a good and sunny day, without having to read any further about it. In this case, with the help of this simple visualisation, you could be able to have a quick overview of the weather forecast! [Yuk and Diamond, 2014]. So it is technology however that made it possible to rapidly evolve the process of graphical representation (visualisation) of data at lightning-fast speed. Another visual example provides an overview that how a good visualisation can provide a quick insight of data about the excel salaries around the world.

1	Unique I	clean Salary (in local currency)	Currency	Salary in USD	Job Type	clean Country	How many hours of a day you work on Excel	Years of Experience
2	ID0001	5846	USD	5846	Analyst	India	4 to 6 hours a day	n/a
3	ID0002	15000	USD	15000	Controller	Croatia	All the 8 hours baby, all the 8!	n/a
4	ID0003	58000	USD	58000	Analyst	USA	All the 8 hours baby, all the 8!	n/a
5	ID0004	48000	USD	48000	Controller	Pakistan	2 to 3 hours per day	n/a
6	ID0005	54000	USD	54000	Engineer	USA	All the 8 hours baby, all the 8!	n/a
7	ID0006	41731	USD	41731	Analyst	Iceland	All the 8 hours baby, all the 8!	n/a
8	ID0007	145000	EUR	184207,9187	Manager	Germany	1 or 2 hours a day	n/a
9	ID0008	12000	USD	12000	Analyst	Ukraine	All the 8 hours baby, all the 8!	n/a
10	ID0009	44000	USD	44000	CXO or Top Mgmt.	Portugal	1 or 2 hours a day	n/a
11	ID0010	1152000	PKR	12227,4302	Accountant	Pakistan	All the 8 hours baby, all the 8!	n/a
12	ID0011	51650	EUR	65616,13102	Specialist	Ireland	2 to 3 hours per day	n/a
13	ID0012	14000	USD	14000	Engineer	Hungary	4 to 6 hours a day	n/a
14	ID0013	749000	INR	13338,1296	Analyst	India	All the 8 hours baby, all the 8!	n/a
15	ID0014	49000	USD	49000	Analyst	USA	All the 8 hours baby, all the 8!	n/a
16	ID0015	85000	USD	85000	Engineer	USA	1 or 2 hours a day	n/a
17	ID0016	75000	USD	75000	Engineer	USA	All the 8 hours baby, all the 8!	n/a
18	ID0017	107000	USD	107000	Manager	Switzerland	4 to 6 hours a day	n/a
19	ID0018	45000	USD	45000	Reporting	South Africa	All the 8 hours baby, all the 8!	n/a
20	ID0019	550000	INR	9794,354178	Manager	India	2 to 3 hours per day	n/a
21	ID0020	50000	USD	50000	Manager	India	1 or 2 hours a day	n/a
22	ID0021	13500	USD	13500	Manager	India	4 to 6 hours a day	n/a
23	ID0022	96000	USD	96000	Analyst	USA	2 to 3 hours per day	n/a
24	ID0023	1000000	INR	17807,91669	Manager	India	4 to 6 hours a day	n/a
25	ID0024	75000	USD	75000	CXO or Top Mgmt.	USA	4 to 6 hours a day	n/a
26	ID0025	40000	USD	40000	Manager	USA	2 to 3 hours per day	n/a
27	ID0026	60000	USD	60000	Analyst	USA	All the 8 hours baby, all the 8!	n/a
28	ID0028	32400	EUR	41160,94182	Manager	Belgium	4 to 6 hours a day	n/a
29	ID0029	900000	INR	16027,12502	Engineer	India	1 or 2 hours a day	n/a
30	ID0030	600000	INR	10684,75001	Manager	India	4 to 6 hours a day	n/a
31	ID0031	41000	USD	41000	Manager	Russia	All the 8 hours baby, all the 8!	n/a
32	ID0032	360000	INR	6410,850007	Specialist	India	4 to 6 hours a day	n/a
33	ID0033	35000	GBP	55166,23952	Analyst	UK	All the 8 hours baby, all the 8!	n/a
34	ID0035	19200	USD	19200	Analyst	Poland	2 to 3 hours per day	n/a
35	ID0036	500000	INR	8903,958344	Consultant	India	All the 8 hours baby, all the 8!	n/a
36	ID0037	150000	USD	150000	Manager	USA	2 to 3 hours per day	n/a

Figure 2.1: Salaries in USD represented in excel sheet

Figure 2.1 indicates the listing of a big excel sheet file with approximately 1880 lines, where each line represent a specific information (data) about the salaries, such as dean salary, dean country, currency, job type, working hours and so on. If you get the information only in this form, it would be harder to get an overview at first sight of how these salaries have been used or what exactly the table contains. In this case, it is almost unachievable to make sense of all those rows by only taking a look at them. But if you get the same information represented in a visual form, including graphs, diagrams and so on, as shown in the Figure 2.2, it is much easier, to understand what precisely these data represent. In this way, this simple example makes us know that a good visualisation makes massive datasets more coherent, it makes information to appear more compact.

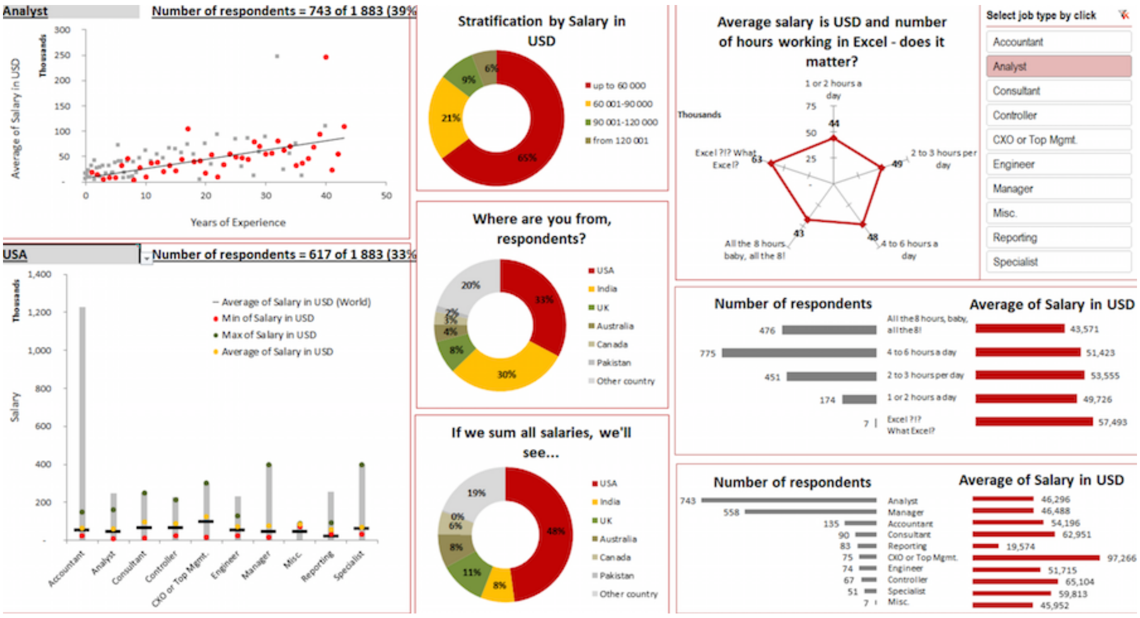

Figure 2.2: Salaries in USD represented with different visualisation diagrams

Nowadays visualisation techniques are bringing art and science together. A good visualisation is good to be aesthetic, but first of all, it must be informative and efficient [Steele and Iliinsky, 2010]:

- **Aesthetic**: According to [Salem and Rauterberg, 2005], aesthetics is named as "the measurement of beauty". Despite that fact, aesthetic is also something that entails pleasure in the human eye and not only about the vision and the beauty [Lau and Moere, 2007]. To create an aesthetic visualisation, first of all, the designer has to take in consideration that the user has to be able to read the data from that visualisation and not to create just a beautiful picture [Lau and Moere, 2007].

- **Informative**: Regardless of how beautiful the visualisation might be, the first thing that should take in consideration for a successful visualisation is informativeness. If you don't achieve to provide information to the user to gain knowledge, then that visualisation has failed [Steele and Iliinsky, 2010].

- **Efficient**: One of the most critical factors for a visualisation to be useful is simplicity. The simpler it is, the more efficient it will be. The straightforward graphs are always the smoothest one for the user to understand.

Edward Swan [Swan II et al., 1999] has also the same opinion, where he sad that: "*Effectively designed visual representations facilitate the understanding of complex phenomena by selectively emphasising the most important features and relationships while minimising the distracting effects of extraneous details.*" [Swan II et al., 1999]

2.2 Types of Information Visualisation

Data visualisation is about telling a story. It is about helping users to understand data by placing it in a graphical context. It is about gaining valuable insight data, by making complex information in this way more accessible, usable and of course making it much easier to understand.

Ben Shneiderman [Shneiderman, 1996] expressed a visual information seeking mantra, of how to deal with the representation of data:

"*Overview first, zoom and filter, details on demand.*"

Based on this approach, the interactions that a user might perform in information visualisation systems are divided into seven necessary tasks that are at a high level of abstraction [Shneiderman, 1996]:

- **Overview**: Gain an overview of the entire information.
- **Zoom**: Zoom in on items of interest.
- **Filter**: Filter out uninteresting information.
- **Details-on-demand**: Select an item or group and get details when needed.
- **Relate**: View relationships among items.
- **History**: Keep a history of actions to support undo, replay, and progressive refinement.
- **Extract**: Allow extraction of sub-collections and query parameters.
- **Organise**: Organise items manually for finding them later on easily.

Andrews [Andrews, 2002] states that information visualisation is based on abstract data that don't necessarily have a spatial dimension, unlike scientific visualisation which deals with physical data. The collected data at the information visualisation are not mapped and represented direct in 2D or 3D physical space like at the scientific visualisation, but they are visualised in any of a multitude way. This process is displayed in Figure 2.3.

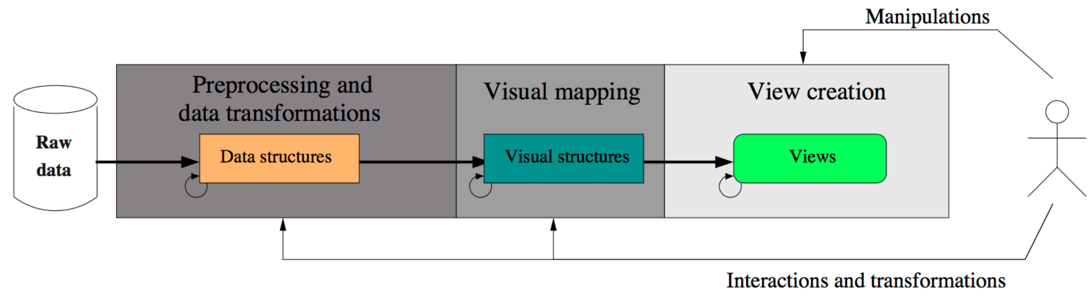

Figure 2.3: Diagram depicting the information visualisation reference model [Mazza, 2009]

Raw data is drafted into data structures that support a visualisation. These data structures are used to construct a visual data structures by forming in this way optical properties such as colour, geometry, and position. For creating interactive views of the data with user interaction, are used visual structures given in Figure 2.3. No matter how abstract the nature of data is, or what is the purpose of the information, there are several types of information visualisation. According to Shneiderman [Shneiderman, 1996] and Andrew [Andrews, 2002], information can be classified into the following categories:

- *Linear information*
- *Hierarchical information*
- *Networks information*
- *Multidimensional information*
- *Vector spaces information*
- *Spatial information*

Before deciding which type of visualisations to use, first we have to know the nature of the data, to understand the kind of the data and also the relationship between the data that will be visualised. For every type of information mentioned above, a significant number of visualisation exist. In the following, we will describe some of the more commonly used visualisations.

2.2.1 Line Chart

A line chart or also known as a line graph for the first time was used by William Playfair [Playfair, 1801]. It is a type of visualisation where the information is represented as a line, formed as a connection of series data points called ‘markers’. It is usually used for time series data such as financial data in a stock market, data for facilitating trend analysis. The first line chart is presented by Playfair [Playfair, 1801] in Figure 2.4 where the modern line chart is shown in Figure 2.5.

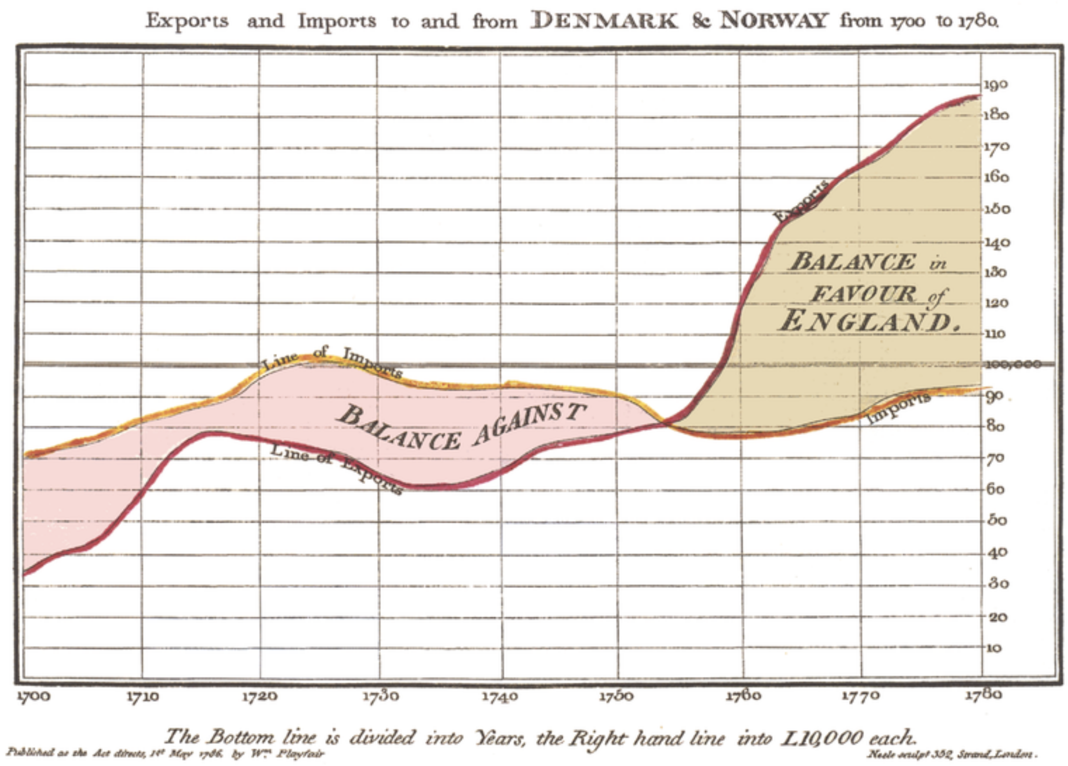

Figure 2.4: The exports/imports between Norway and Denmark from 1700 to 1780 Wikipedia [Wikipedia, 2018b]

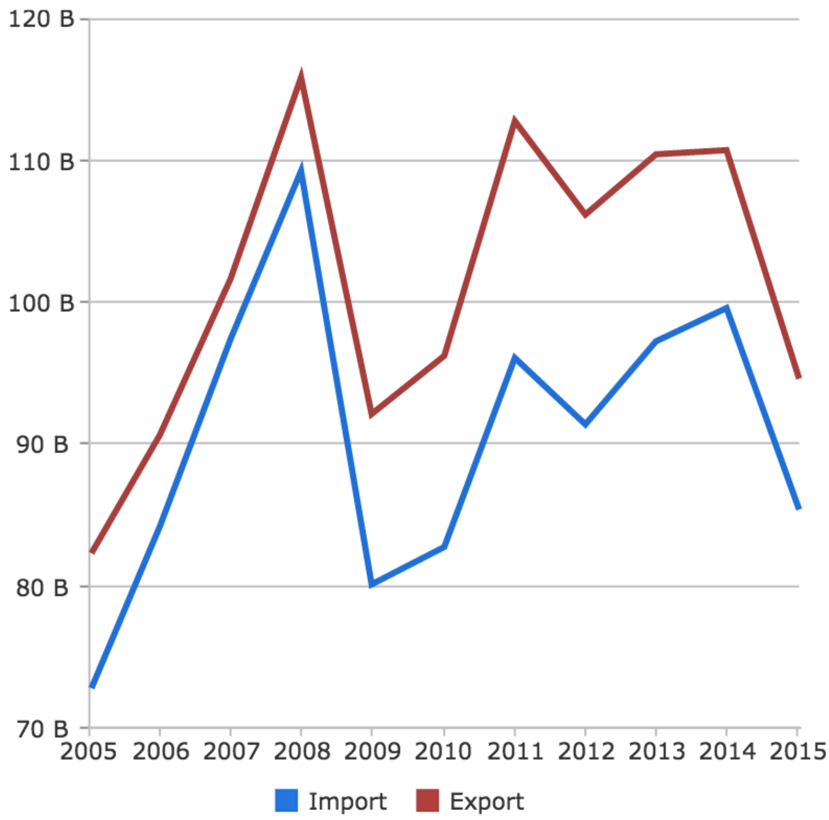

Figure 2.5: The exports/imports of Denmark from 2005 to 2015 [Solution, 2015]

2.2.2 Pie Chart

A pie chart or a circle chart is a type of visualisation graph in the form of a circle, divided into slices or sectors interpreting numerical proportion (relative size of the data). The information from this chart is straightforward to understand (if the number of slices is not significant) as every slice size of the circle shows the weight of each category. The sum of all slices in the pie charts must add up to a full circle (360 degrees), so each slice represents a percentage of 100 percent. The main disadvantage of pie charts is that by presenting a substantial number of slices where each slice will become tiny, it will be hard to identify the weight of each slice. In that case for making the information more significant, pie charts have to use other indicators such as colour, textures or arrows, which leads in this way to a creation of a new legend inside the pie chart, which makes them useless in the case of big

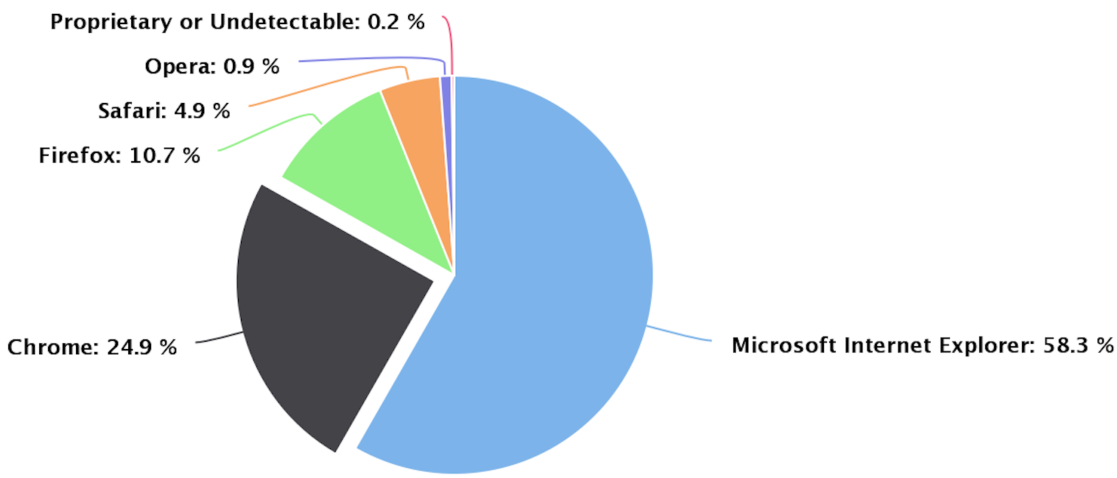

Figure 2.6: Pie chart representing a browser market shares from January 2015 to May 2015 [Highcharts, 2017b]

data. There are several alternatives to the pie charts, such as Donough chart, three-dimensional pie charts and, multi-level pie charts [Wikipedia, 2018a]. An example of a classic pie chart is shown in Figure 2.6.

2.2.3 Bar Chart

Bar charts deals with either horizontal or vertical rectangular bars that present quantitative data. It usually has two axis, where one of them displays the category, and the other one shows the discrete value. Most commonly they are used to compare the height and the width of different category measures which answers the question of "how many", and also for displaying the change of data over time. Sometimes when there is a grouping of categories colour could be used to indicate that group. Figure 2.7 demonstrates the world population by region represented by horizontal bars, and Figure 2.8 shows monthly average rainfall by region represented by vertical bars.

2.2.4 Area Chart

Similar to line charts, area charts also serve for displaying and comparing the quantitative progression of data over time. Here the area beneath a line, a line which

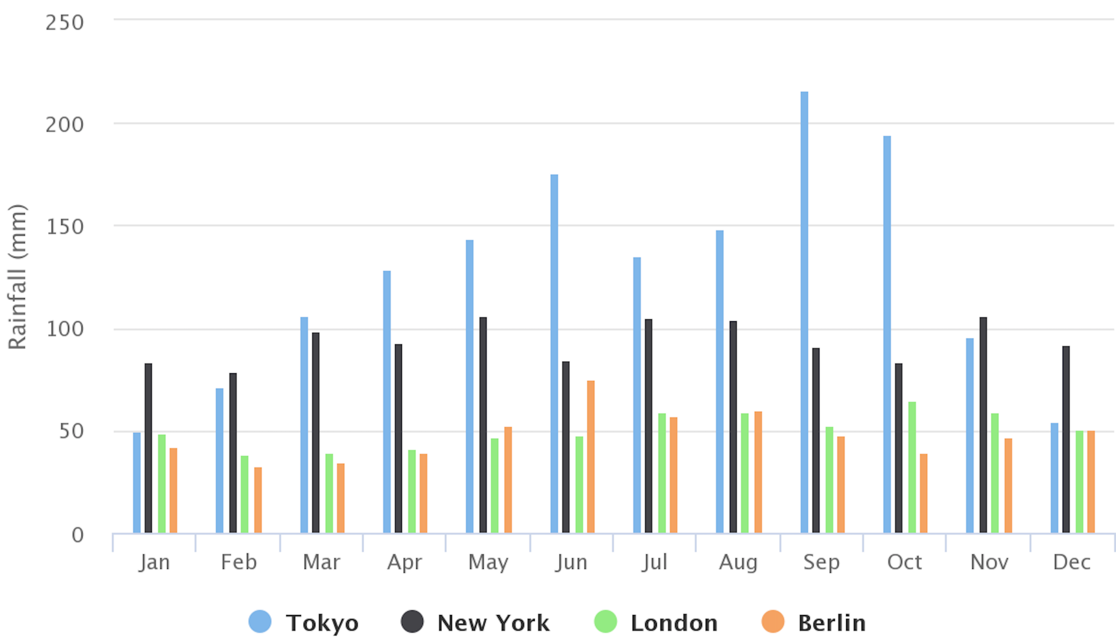

Figure 2.7: World population by region illustrated in horizontal bars [Highcharts, 2017a]

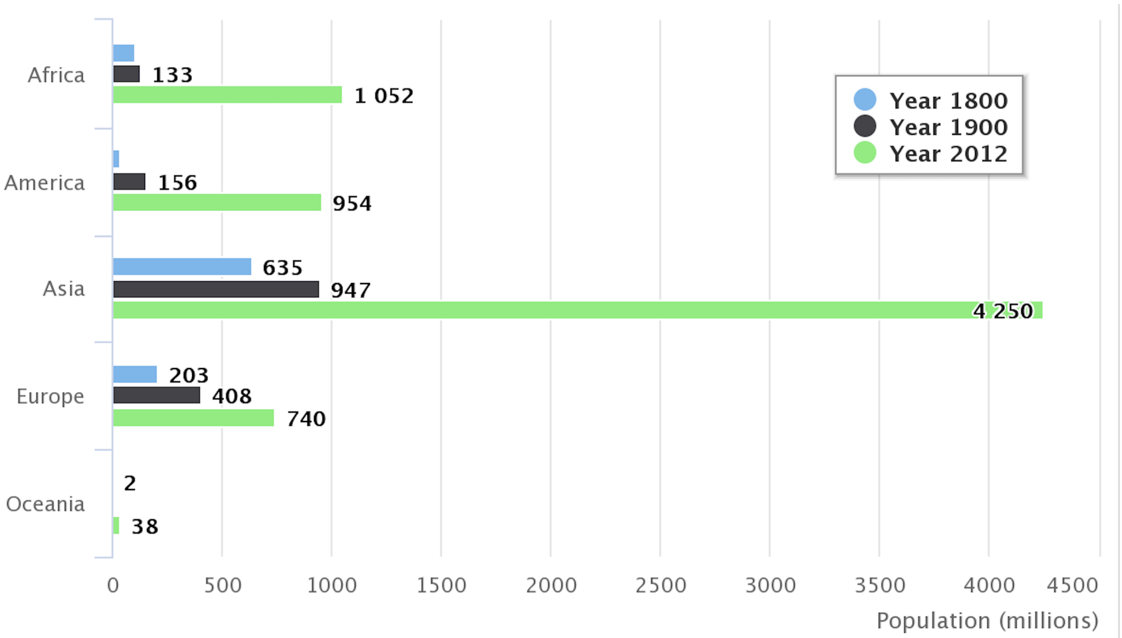

Figure 2.8: Monthly average rainfall by region illustrated in vertical bars [Highcharts, 2017e]

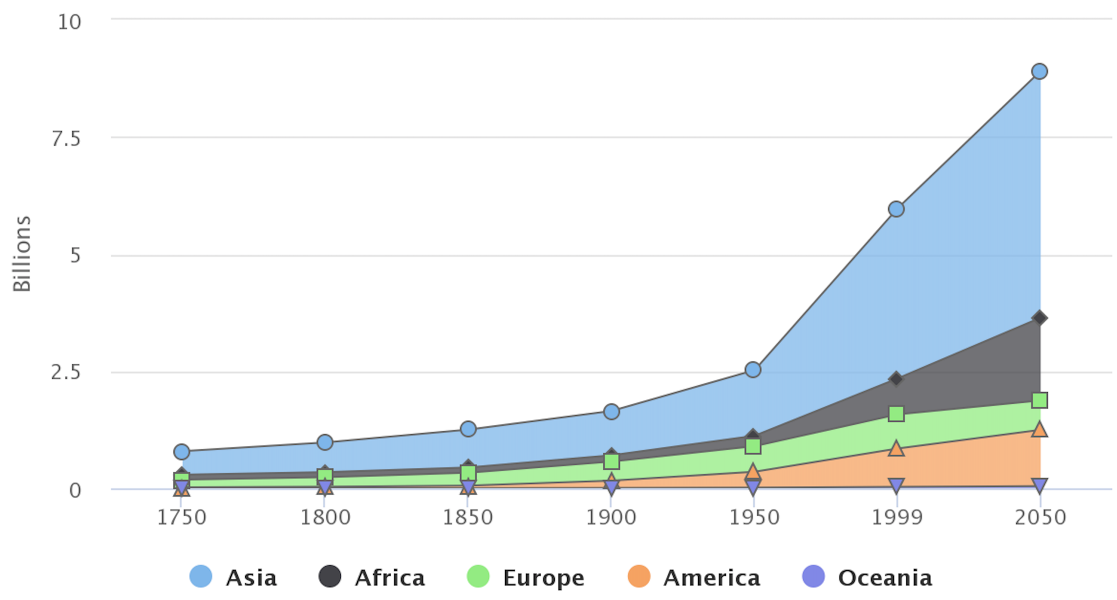

Figure 2.9: Stacked area chart of historic and estimated worldwide population growth [Highcharts, 2017d]

is created as a connection of data point of each entity, is filled in with colour. A very often used type of area charts is stacked area charts, in which each data entity starts from the point left by the previous data entity. An example of stacked area chart is displayed in Figure 2.9 which represent historical and estimated worldwide population growth by region.

2.2.5 Heat Map

Heat maps also are used to perform categorical data such as geographic data or data tables through variations in colouring or through changing the colour intensity of each data entity depending on their values. Figure 2.10 shows an example of heat map matrix used in the educational field for representing students competencies. The intensity of red colour expresses the level of competency achievement. The higher the intensity (the darker the colour), the stronger the competency [Johnson et al., 2016].

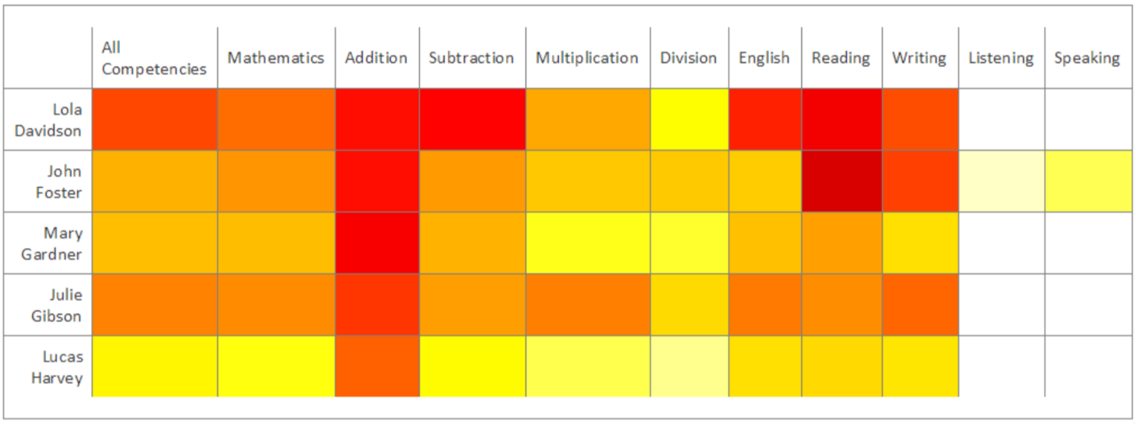

Figure 2.10: Heat map displaying multi-dimensional data

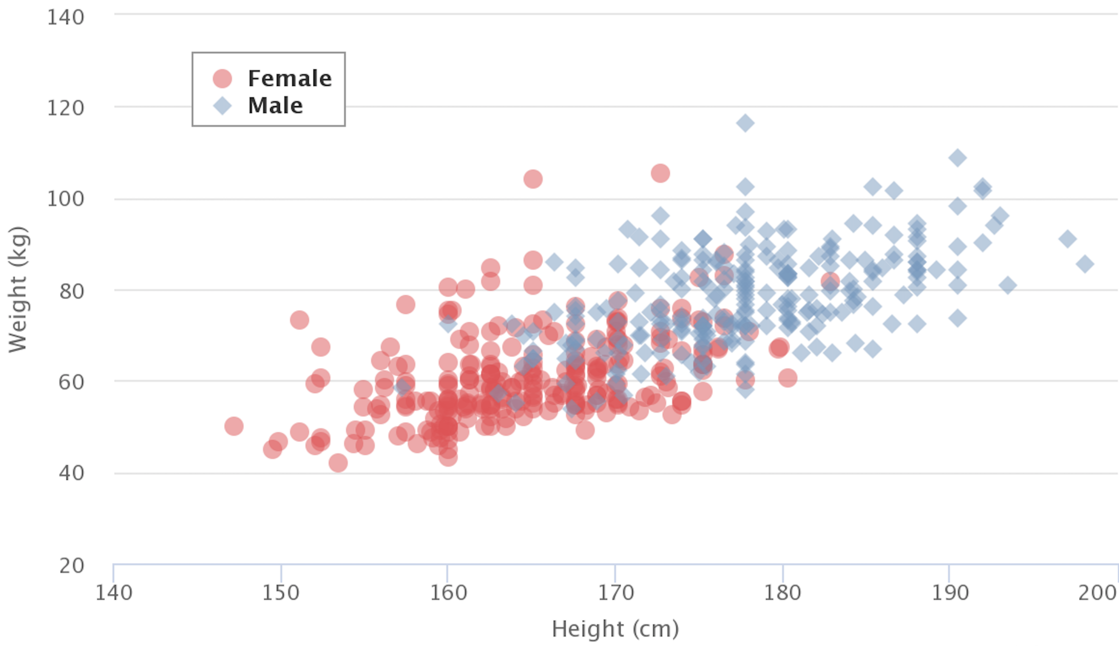

Figure 2.11: Height versus weight of 507 individual by gender represented by scatter plot [Highcharts, 2017c]

2.2.6 Scatterplot

It visualises the relationship between two variable in a set of data points. Using this type of chart one can observe how the other one observes one variable. Additionally, a single variable can be observed over time. An example of scatterplot is visualised in Figure 2.11, that shows the differences height versus weight of 507 individuals by gender.

2.2.7 Tree Diagram

A tree diagram is a visualisation method for mapping out hierarchical data in the form of a tree structure. The structure of the tree diagram consists of elements called nodes, where one node (the very first one) is considered as a root (source) node, which means it doesn't have any parent node. It continues with other elements which are connected to each other will lines called branches. Every node that is one step higher in the hierarchy is referred to as a parent node, and the others that are one step lower are named children nodes. Finally, nodes that have no children nodes are known as leaf nodes or as end-nodes. Characteristic for this diagram is the existence of a unique path from root element to any vertex (entity) of the graph. They find applications in various fields such as: in computer science and mathematics, in taxonomy, in family relations, in businesses and organisations. For instance, in a team, they're helpful for expressing a new idea about any subject, where the connections are directly visible. Figure 2.12 shows a case of the tree diagram, which is used in the decision making of the cost-effectiveness of computer-assisted surgery-CAS in Monte Carlo.

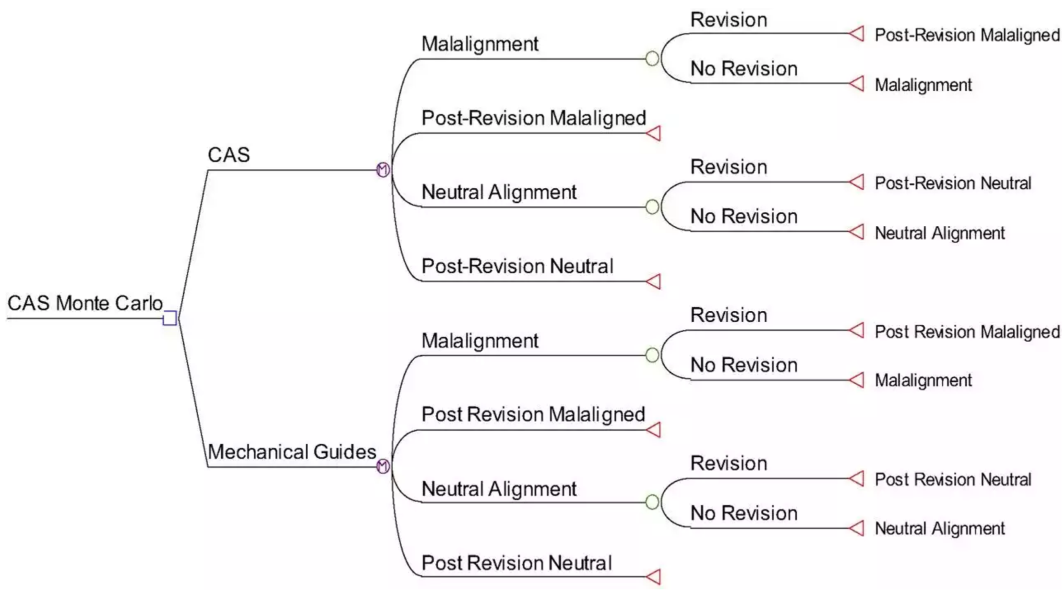

Figure 2.12: Tree diagram of the cost-effectiveness of computer-assisted surgery-CAS in Monte Carlo [Novak et al., 2007]

2.2.8 Network Graph

The network graph is an extended method of hierarchical visualisations which enables the drawing of the data sets and their relationships inside a network. It comprises elements such as nodes that present entities connected by edges. A circle illustrates a node, and edges are shown as lines or arrows. Network graphs may be either static or dynamic, representing in this way information continuously. They are used to show machine- and human-based systems. An example of a network graph is depicted in Figure 2.13 that represents the EU crisis from March 2011 till March 2013. The data visualised in the network chart are taken from European news media websites such as Spiegel Online, Guardian, and Ekathimerini. The size of the node means the appearance frequency, in this case, the more often the name of the country was referred, the bigger the node is displayed. Also, the edges and their thickness between nodes have special meaning such as the common naming of the countries in the analysis and how often that naming happened, which means the often the naming occurs, the thicker that line appears. For instance, from the graph, we can see that Germany and UK have a substantial relationship to each other which means that they were mainly linked together in the EU crisis.

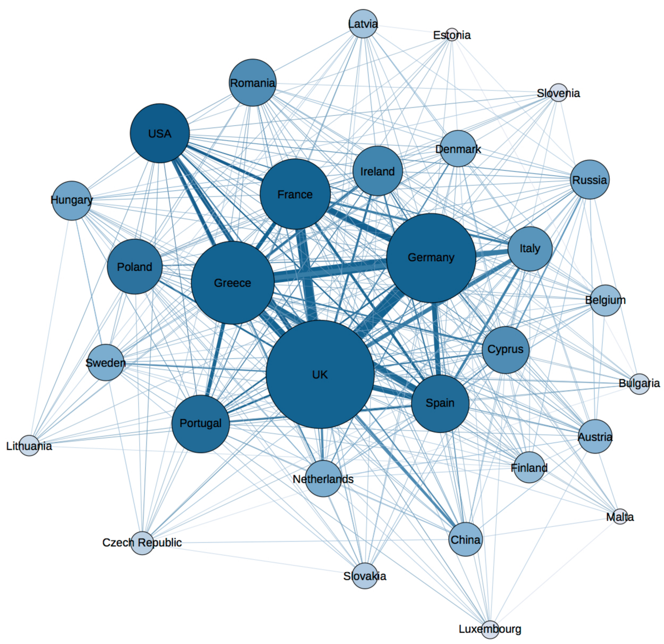

Figure 2.13: Network Graph: Convergence in the news media discourse on the EU crisis [Nguyen, 2015]

2.2.9 Force-Directed Graph

The force-directed graph is based on network graphs and node-link graphs. The aim of the graph is placing the nodes (entities) in 2 dimensional or 3-dimensional space. They use drawing algorithms for calculating the position of the nodes as well as the edges. The edges represent relationships between nodes. The most common usage of this graph is in social media such as Facebook and Google+, where social entities (friends or corporations) are represented by vertices and relationships or friendships between pairs of social entities are represented by edges or lines. A sample of the force directed graph is displayed in Figure 2.14, expressing the mobile patent suits between different provider. Dashed edges represent resolved suit where green edges represent licensing.

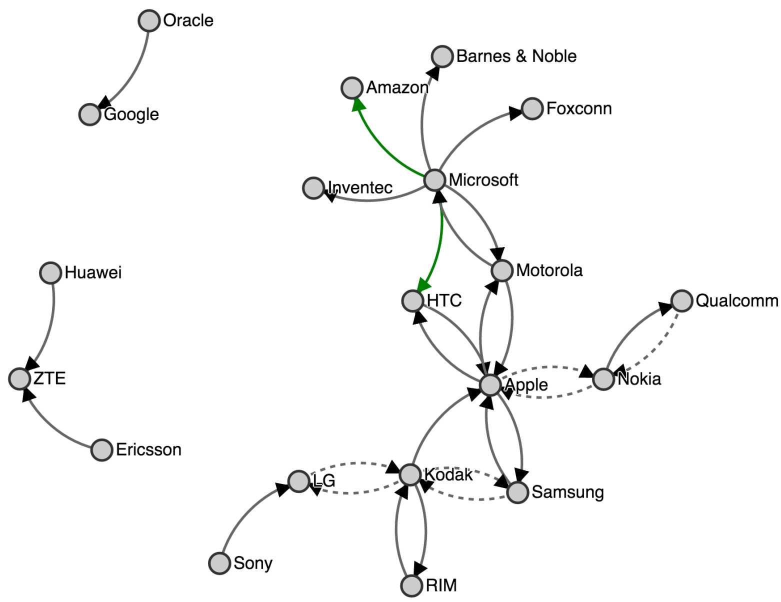

Figure 2.14: Network Graph showing mobile patent suits [Bostock, 2017]

2.2.10 Smiley Faces Visualisations

In the context of educational data, except for visualisation graphs and charts as mentioned above, the use of traffic light metaphor or the smileys, as well as text-based reports about achievements, or rubrics is becoming quite popular.

Such kind of visualisations is also used in Lea's Box Project, more concrete at the OLM (Open Learner Model) part. Those visualisations aim to help students to identify their strengths and their weaknesses, to identify an area of improvements, to help them plan what to do next, to think about the way how they are learning, and to allow them to work independently to make decisions about the learning. They show different visualisations of students competencies. Some of the visualisations display structure, some of them can show time, and the others can give an overview. One particular example of this is provided in Figure 2.15, which displays a smiley face visualisation. The smiley faces are used to indicate competencies of the student. The happier the face, the stronger the competencies. The confused face indicates little or no competence, and no face means there is no information.

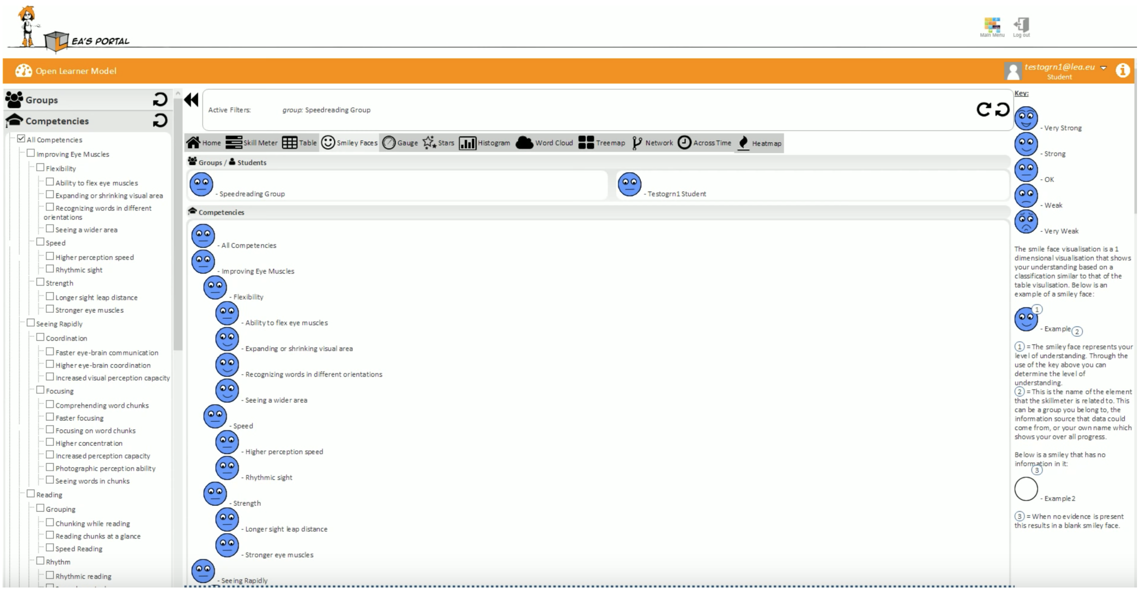

Figure 2.15: Smiley face visualisation from LEas'Box OLM [LEA'sBox, 2016]

2.3 Strengths and Weaknesses

Since the applications of the visualisations discussed earlier vary, it is clear that they do have advantages and disadvantages, depending on fact for what purpose they are being used. The tables 2.1, 2.2 and 2.3 will represent some of the strength and the weakness of these visualisations as well as provide simple examples of how they could be used for educational data.

Chart	Advantages	Disadvantages	Examples
Line Chart	(i) clear data trends, (ii) can compare multiple continuous data sets easily exact values retained, (iii) compare two data set, (iv) interim data can be inferred from graph line	(i) hard to read for a wide range of data, (ii) use only continuous data, (ii) not easy to compare different categories of data, (iii) hard to quantify partial icons, (iv) icons must be of consistent size	average grades of students over a month, semester or year
Pie Chart	(i) easy to understand and to set up and require minimal extra explanation, (ii) show data comparison at a glance, (iii) shows the percent of the total for each category, (iv) display relative proportions of multiple classes of data, (v) represents data visually as a fractional part of a whole	(i) represent one data sets, (ii) angles are difficult to estimate, (iii) difficult to compare data, (iv) less useful for too many pieces of data, (v) no exact numerical data	the portion of students that got different grades for a subject
Area Chart	(i) shows quantitative data in a relatively simple format, (ii) easy to interpret relative values	(i) hard to read the more segments each bar has, (ii) hard to compare each segment to other because of baselining (y-axis intersects the x-axis)	the portion of students that got different grades for a subject over a period of time

Table 2.1: A summary of different types of visualisations, their strengths, weaknesses and some example usages in the context of educational data (part 1)

Chart	Advantages	Disadvantages	Examples
Bar Chart	(i) summarise a large data set in a visual form, (ii) easier to highlight trends, (iii) each data category is represented in a frequency distribution, (iv) estimate key values at a glance	(i) fail to expose essential assumptions, causes, effects and patterns, (ii) require additional explanation, (iii) all the bars start at zero, at log scale isn't allowed	the number of students registered in each course
Heat Map	(i) good to visualise 4D data, (ii) ideally used with discrete variables, (iii) track user activities (e.g. user clicks on a website)	hard to map colour onto a continuous scale	compare activities of a student with his or her competencies, by visualising them in a heat map
Scatterplot	(i) straightforward and non-mathematical method is used, (ii) help to identify trends in the data by a positive and a negative correlation as well as no correlation, (iii) shows minimum and maximum of the data set, (iv) shows a trend in the data relationship	(i) use only continuously data, (ii) hard to map colour onto a continuous scale, (iii) hard to see the data because no graph lines exist to look at where exactly the point is, (iv) data on both axes should be continuous, (v) hard to visualise event in large data sets	student grades progress over time

Table 2.2: A summary of different types of visualisations, their strengths, weaknesses and some example usages in the context of educational data (part 2)

Chart	**Advantages**	**Disadvantages**	**Examples**
Tree Diagram	(i) simple to understand and interpret after a brief explanation, (ii) effective when used as a general guideline	has no flow and doesn't show processes and determinations made among the structure	displaying the relationships between different courses
Network Graph	(i) show precedence well, (ii) highlights a sequence of the task in the project, (iii) can calculate critical activities and path, (iv) shows possible existing floats	(i) hard to follow on a massive project with many activities, (ii) time-consuming to create, (iii) expensive to create	can represent how all competencies are linked to each other
Force-Directed Graph	(i) easily adapted and extended, (ii) easy to implement and to understand, (iii) interactive	(i) slow running time for the more significant graph, (ii) poor local minimum	relationships between different courses

Table 2.3: A summary of different types of visualisations, their strengths, weaknesses and some example usages in the context of educational data (part 3)

Chapter 3

Learning Analytics

Have you ever wondered how your brain is capable of absorbing new things, or how you approach problems, or what would you do next time different to achieve greater success by learning new things easier and quickly, or say it better, how to get an overview of what you did well and what went wrong during the learning process or even how to improve it? These questions are become the latest buzz in the educational field, within those frames for years are going on different research studies, which are trying to find answers in this fields. The teaching process and the skills that have been taught to the students from 20th century till now have been changed, most of all due to the rapid technological progress and its widespread in the education.

Nowadays, as a result of this rapid change teachers are being pushed by the researchers to teach students new skills such as problem-solving, critical thinking, collaboration, innovation, communication, creativity, self- discipline and so on, which are also called or known as the "21st-century skills" [Khalil and Ebner, 2015].

In education, students leave traces behind their learning process. Sometimes is too hard or too complicated to be able to capture or even to measure those skills of the students learning activities during their classes, by a human eye, in this case by only one teacher for a considerable number of students participating in one class. Using the current assessment techniques such as teacher-constructed test, multiple choice tests, open items and portfolios or quizzes, is hard to say that these techniques are the right ones since they are based on product and not on processes [Blikstein, 2011].

However, educational institutions emphasised the importance of new disciplines like educational data mining and learning analytics which are enabling detecting, collecting and analysing the massive amount of data generated from students to enhance learning and to improve student's performance.

Even though the data mining and learning analytics deal with the same problem by having similar goals and interests for providing higher insight into the educational process, deep inside them, based on their orientations, like technological, methodological and ideological they differ from each other [Siemens and d Baker, 2012].

Learning analytics is more about improving and shaping the learning process by collecting data from students' footprints left behind while studying, whereas the educational data mining provide different algorithm to treat those students' footprints for identifying patterns or estimating indicators [Duval, 2012, Greller et al., 2014].

Educational data mining EDM is defined from the International Educational Data Mining as:
"*Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. Whether educational data is taken from students' use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities, it often has multiple levels of meaningful hierarchy, which often need to be determined by properties and the data itself, rather than in advance. Issues of time, sequence, and context also play important roles in the study of educational data* "[Society, 2018].

On the other hand, the term of "Learning Analytics" for the first time was used by Bienkowski Mingyu and Means in 2009 [Bienkowski et al., 2012]. In 2011 the first international conference on learning analytics was held in Banff, Alberta. Since then, is being the recent buzzword in education systems. Learning analytics enables the collecting, tracking, aggregating, analysing and visualising of those data that take place in different environments, in a clear and useful way with the help of various smart tools [Duval, 2011], So, it is the way how we can use technology to better impact the teaching process. Bader-Natal and Lotze [Bader-Natal and Lotze, 2011] stated that learning analytics aim is to understand learning in a better and in a more profound way, as well as to foster it more efficiently.

A good definition of learning analytics that is worth quoting is from Learning

Analytics Research society as follows:

"*Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*"[Siemens, 2001]

George Siemens announced as well that:

"*Learning analytics is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.*" [Siemens, 2010]

George Siemens has described the process of the learning analytics as a flow of the learning data, starting from the learner's off-put data and intelligent data, thus continuing with proliferation, analysation, prediction, ending up with personalisation and adaptation of the data [Siemens and Long, 2011]. The whole process is displayed in Figure 3.1 [Siemens, 2010].

The figure gives us one potential approach to learning analytics. According to this approach, learners generate constantly data during their routine of the learning process. The origin of data could be various, such as data from different social media (Tweeter, Facebook, LinkedIn, blogs etc.), and log data from personal learning environments, learning management systems and so on. This data is collected into the database, and later on, this data could be analysed and transformed into intelligent ones using semantic processes, which could be used for adaptation, personalisation and building recommender systems.

3.1 Learning Analytics Cycle

According to Clow [Clow, 2012], the analytic learning cycle is expressed as a closed and iterative process consisting four components: learners, data, metric/analytics and intervention see Figure 3.2 [Clow, 2012].

The first component is learners. The learners could be students, pupils or students from online courses that generate data.

The generation and capturing of the data is represented as the second component of this cycle. Data could be various, such as log data, questionnaires, assessment data, activity data from online portals and so on.

The next component is considered the main element of the learning analytics cycle. It is the processing of the collected data from the previous step into metrics

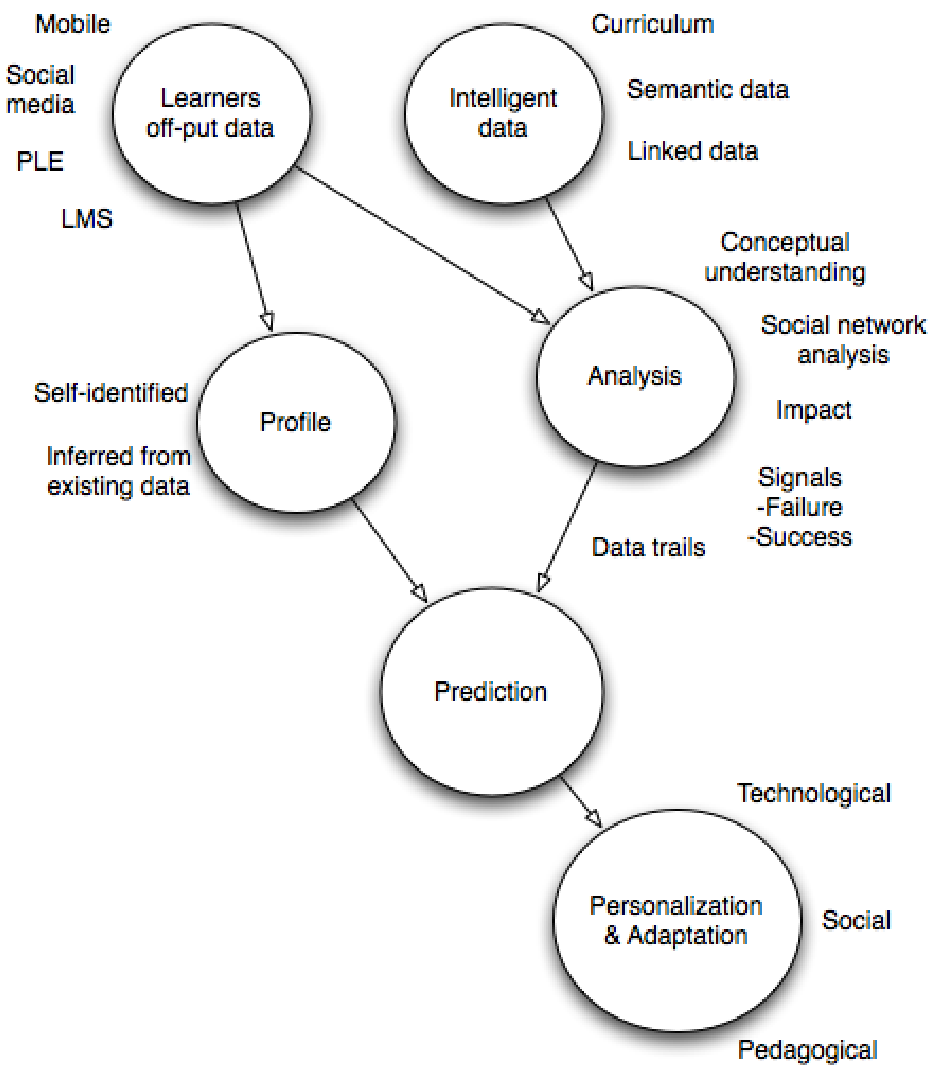

Figure 3.1: The process of learning analytics[Siemens, 2010]

or analytics. Using which are created different statistical analysis and visualisations, such as dashboards, recommenders, network analysis etc., which help teachers to better understand the data.

This is the cycle where the metrics result in concrete interventions. Here can be effected whether the learners or the teachers. When the learners are being affected, using self-learning dashboards and statistics, they can reflect and compare themselves to the others. When the teachers are affected, they use the dashboards to inform the students and to intervene in their learning process by helping them improve it.

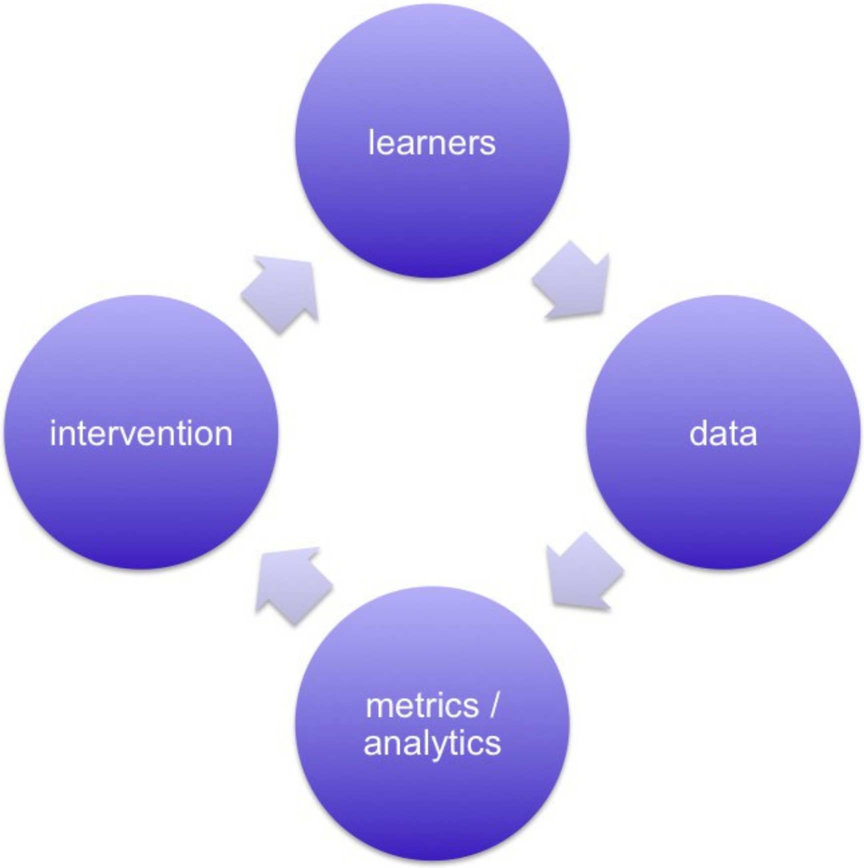

Figure 3.2: The learning analytics cycle [Clow, 2012]

3.2 Some Applications of Learning Analytics

Even though learning analytics is a relatively new study field, several types of research indicate that the use of learning analytics has prompted a significant transformation and has found application not only in education but also in various sectors such as sport, health, business intelligent or advertising. In the following, there will be mention some applications demonstrated with examples.

An example of learning analytics application in sport and health direction may be presented through applications like runkeeper (http://runkeeper.com/) or nikeplus (http://nikerunning.nike.com/) which is illustrated in Figure 3.3 [Duval, 2011]. The relation of this example in the learning context is that these kinds of apps help users (runners) to see the progress they have made, to set and achieve goals, to motivate them to continue further by looking at their progress, to follow a plan they have made by themselves, to track activities such as run, walk, climb stairs so in this way they could get a real view of their workout, etc.

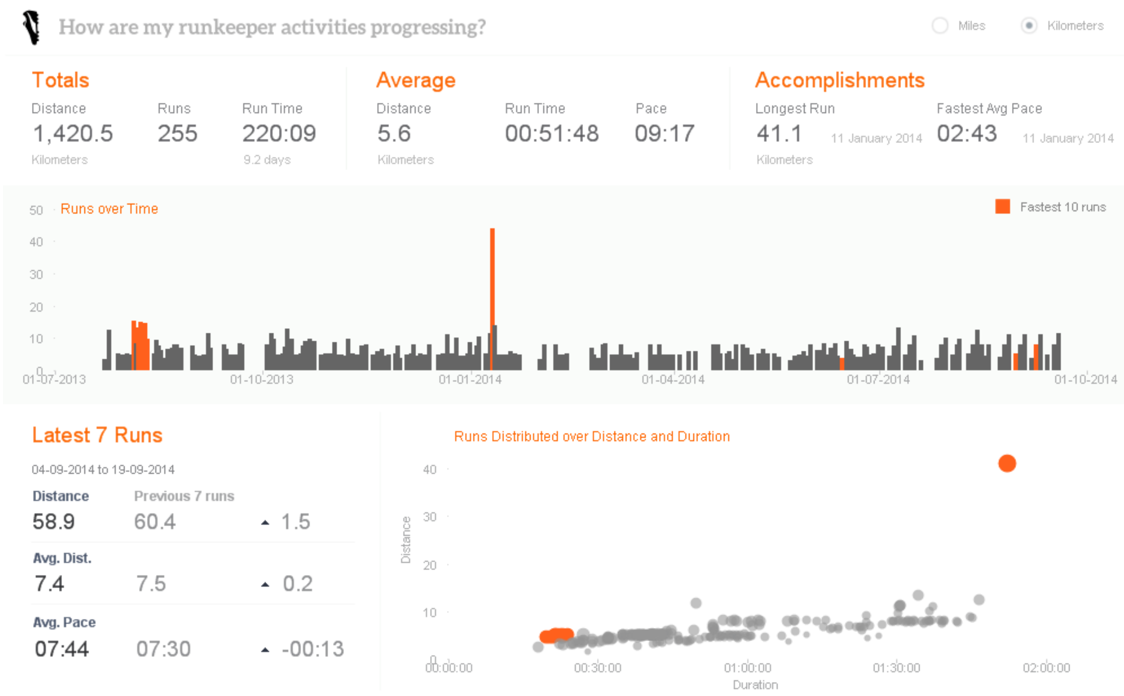

Figure 3.3: A runkeeper dashboard

Another positive aspect of this apps is on the online social networks, where they enable sharing the data with other runners who are using the same apps, challenge them, compare your data with others, so you can see how your runs stack up to each other [Duval, 2011].

Other applications of learning analytics in education which I would like to mention are some apps developed at the Graz University of Technology, for testing students competencies in different mathematical problems such as adding and multiplication, as well as multiplication and division. These apps emphasise the importance of the learning analytics and how interesting and useful could be its usage for education in primary schools [Ebner and Schön, 2013].

One of these applications is named "ein mal eins" [1]. Here the students are able to learn and to practice multiplication table [Ebner and Schön, 2013](see Figure 3.4). The main screen of the multiplication trainer is displayed in the Figure 3.4. It shows a problem that is given to the user as well as correctly answered problems listed at the left side of the screen. The correct answers distinguish between colours, where yellow means the well-known problems, and orange means known problems. The given problem has to be answered within a given time range.

The proposed apps collected data from users for a long period of time, where they were being tested with a significant amount of problems, e.g. on learning multiplication table. In the end, after entering a considerable number of repetitions, the app gave the users feedback. Due to this feedback, the users were able to improve their self by realising where did they perform good or bad, which affects the learning process in a positive manner. The end results of the analysed students data from the application show that these results could not be achieved by human teachers so easily which highlight once again the importance of learning analytics for teachers and learners [Schón et al., 2012, Ebner and Schön, 2013].

[1] `https://schule.learninglab.tugraz.at/math`

Figure 3.4: Screenshot of the multiplication trainer
`http://mathe.tugraz.at`

Chapter 4

Competence Oriented Learning Analytics

In this Section, is described Knowledge Space Theory KST and Competency-based Knowledge Space Theory and their approaches in Learning Analytics.

4.1 Competence based Knowledge Space Theory

Gaining knowledge and reaching goals (learning trajectories) in many education scenarios is like a mountain climbing. Some can climb very easy; some find it very hard, and some do not even try or fail.

Techniques like learning analytics and educational data mining allow identifying the individual strength and weaknesses, individual learning path and goals, competencies and competency gaps (that need to be filled) by analysing the students learning styles as well as their learning trajectories [Kickmeier-Rust and Albert, 2016b]. The growing use of these techniques is becoming an essential issue of nowadays educational life which aims to improve teaching and learning by providing different ongoing developments.

One of those development methodologies is competence-oriented learning analytics, which is also the fundamental methodology on which this project (thesis) is based. Thus, the competence oriented approach is significant for learning analytics. This approach brings together two psychological theories, basically from the intelligent tutorial system field [Kickmeier-Rust and Albert, 2016b]. The one is

Competency-based Knowledge Space Theory (CbKST), and the other one is Formal Concept Analysis (FCA).

CbKST is an extension of the behaviourist and descriptive knowledge space theory KST based on competencies [Doignon and Falmagne, 1985, Korossy, 1997]. KST and CbKST offer a theoretical underpinning for learning analytics that pushes analytics beyond being mere "learning stats". CbKST serves as the anchor to link multiple sources and evidence and allows building, slowly but steadily, a probability-based, believe model about students' abilities. While KST focuses exclusively on the observable solution performance, CbKST deals with learning objects and their associated skills and competencies.

Jean-Paul Doignon and Jean-Claude Falmagne introduced the term of KST in the year 1985 [Falmagne and Doignon, 2010]. Since then have been so many studies and research based on this theory. In mathematical psychology theory, KST is the way of representing the learner knowledge in a certain domain. A knowledge domain is represented as a set of problems denoted by Q. The knowledge state is defined as a subset of problems or items that a learner is capable of solving [Doignon and Falmagne, 2012]. So it describes the current knowledge of a person. The problems (items) of a knowledge domain can be associated with a value (0 1) which would indicate whether they are correctly solved or not. Those items can be dependent on each other such that from the correct solution of a specific problem the power of some other problems can be surmised. The so-named prerequisite relation captures these dependencies or surmises relation, which is a transitive and reflexive relation on problem types [Doignon and Falmagne, 1985]. So, the items are described as a domain of the knowledge body. A knowledge structure K is defined as a collection of knowledge states (items). Learners get these items as they are increasing knowledge, corresponding to prerequisite relation, including the empty set ϕ and the whole set Q = {a, b, c, d, e} [Reimann et al., 2013, Reimann et al., 2015, Nasar, 2016].

Formally, an example of a knowledge structure is represented as follows:

$$K = \{\phi, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, \{a, b, d\}, \{a, b, c, d\}, \{a, b, d, e\}, Q\} \quad (4.1)$$

A visualisation of this knowledge structure is represented in the figure Figure 4.1 a)[Kickmeier-Rust et al., 2016b, Heller et al., 2006].

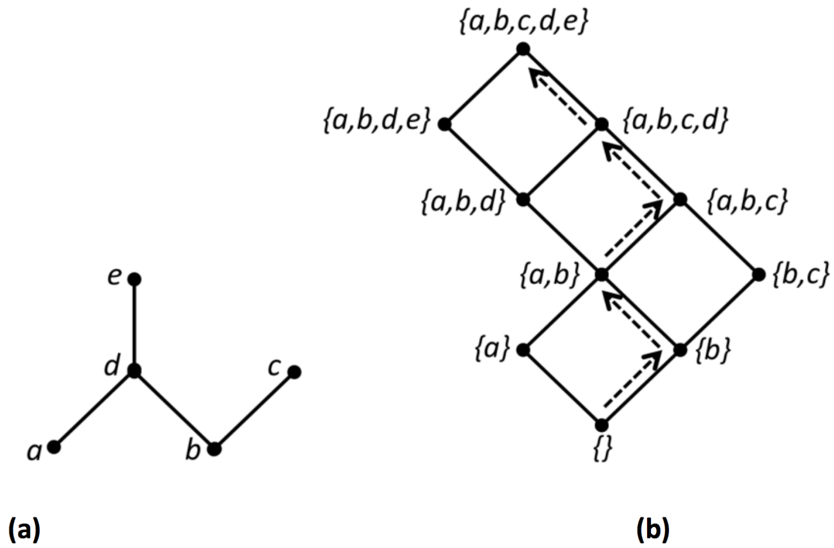

Figure 4.1: a) Prerequisite relations b) Knowledge structure [Kickmeier-Rust et al., 2016b]

A knowledge structure described in the Figure 4.1 shows the collection of knowledge states, where the links between each other can provide several possible learning paths, starting from the very naive knowledge state ϕ (no knowledge) navigating to a full mastery of knowledge [Doignon and Falmagne, 1985, Falmagne et al., 2013, Heller et al., 2006]. The surmise (prerequisite) relation can be visually represented by so-called Hasse diagram. From that prerequisite relation, the knowledge structure can be derived, which is illustrated in Figure 4.1 (b) where the dashed line through the knowledge states indicate one possible learning path. However, not all knowledge states will occur here due to the common dependencies that appear between knowledge domain, where the potential number of the knowledge states will be restricted because of these dependencies also expressed by so defined prerequisite relations [Reimann et al., 2013, Reimann et al., 2015, Nasar, 2016].

An example of making the prerequisite relations clear is described in Figure 4.2 [Kickmeier-Rust et al., 2006]. The author characterised a sample of five competencies about mathematical problems, such as addition subtraction, multiplication, division and competence of solving an equation. The set of all possible knowledge states is 2 to the power of 5, where the power 5 indicates the total number of problems. Assuming the fact that the competencies to add, subtract, multiply and divide are the prerequisite competencies of solving an equation, it is doubtful that a person can solve an equation without knowing how to add numbers. In this case, instead of 32 knowledge states, only 25 knowledge states will happen[Kickmeier-Rust et al., 2006].

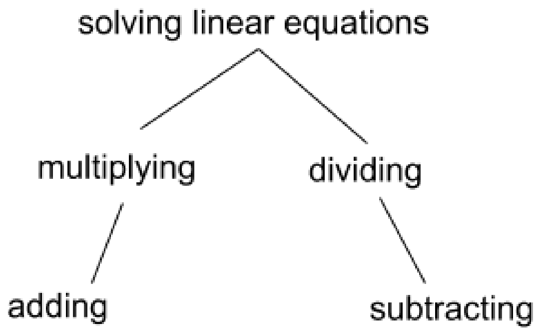

Figure 4.2: Prerequisite relations among competencies of an algebra problem [Kickmeier-Rust et al., 2016b]

CbKST is an extension of the behaviourist and descriptive KST based on competencies [Doignon and Falmagne, 1985, Korossy, 1997]. It provides a theoretical framework for knowledge and competency modelling. CbKST describes a knowledge state of a learner expressed as a set of competencies, that the learner has gained at a given time. So, its approach is based on structuring and representing domain and learner knowledge, by collecting and analysing the student's response. It expresses the learning domain in the form of competences and also presents the relationships between those competencies [Heller et al., 2006]. CbKST is essentially decomposing a domain knowledge into a catalogue of some aptitude (knowledge, skills, ability competencies). In a simple word, the prerequisite relations (the relationship between competencies) can be established by following this model, (as known as a latent model) saying that one competency is learned before the other, the one is necessary to understand the other, and so on and so forth. These relationships enable obtaining a global competency model which is essentially a competency structure.

CbKST deals with three distinct sorts of entities [Heller et al., 2006], such as:

- **knowledge structure**
- **learning structure** - the set L of learning objects
- **competency structure** - the set C of skills relevant for solving the problems, and taught by the learning objects

The knowledge structure establishes the set Q of assessment of a learners competency. Learning structure and competence structure are described equivalently to the knowledge structure as we explained earlier [Heller et al., 2006].

The competence state is a set of all possible competencies (skills) of a learner, which can be inferred from learner's performance on the set of problems that represent the domain. Similarly to KST, prerequisite relations represents the set of competencies constituting the competence structure C composed of alls competency states [Kickmeier-Rust et al., 2006].

A competence structure illustrated in Figure 4.3 describes the prerequisite relations among competencies in the form of a Hasse Diagram. Nodes signify the competencies where links express relationships between competencies. From the Figure 4.3 in the panel a), we can see that the prerequisite competencies for the competence X are competence Y and W., In this case, it can be assumed that if a learner holds a competence X, she or he holds either competence V or W, or both of them. The panel b) indicates the competence structure generated from the prerequisite relation from panel a) where the number of displayed likely competencies states is 13 from a total of 25 states

In order theory, a Hasse Diagram is a type of mathematical representation of a so-called semi-order, used to represent the structure of learning domain in the form of a directed graph that reads from bottom to top, as well as visualising the progress of a learner through this domain. The properties of a semi-order are: (i) reflexivity, (ii) antisymmetry, and (iii) transitivity [Kickmeier-Rust et al., 2016b]. Moreover, the Hasse diagram is frequently used to visualise knowledge or competence structures. It holds some properties that might be very useful for teachers, by offering more than only stats or chart-type visualisations.

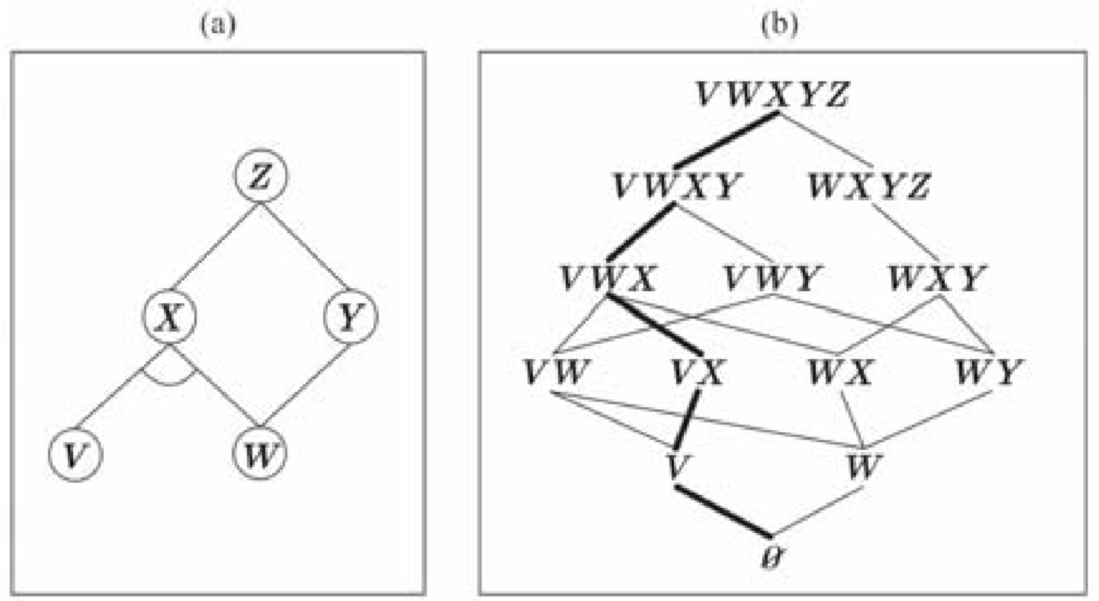

Figure 4.3: a)Prerequisite relations among competencies b) Competence structure set from prerequisite relation from the left. The bold line shows a possible learning path. The graph is taken from [Kickmeier-Rust et al., 2006]

Chapter 5

Visualisation Concept: Competency Based Network Diagram

The main goal of this thesis is to visualise students' competencies, the relation of those competencies in term of dependencies, based on psycho-pedagogical models such as competency-based knowledge space theory. The visualisation should be suited for a group of students (e.g. as an entire classroom) as well as for individual students. The so-called Hasse Diagram inspires the end result visualisation of this thesis invented in the 1960s by Helmut Hasse [Birkhoff, 1948]. It is a direct graph that contains nodes and edges which represent the competencies and their relationships. We will refer to it as competency-based network diagram (CBNK).

This chapter will describe only the visualisation concept. It will explain the meaning and the logic of how the information is encoded and the mathematical and probabilistic representation behind it. Additionally, it will present some low fidelity mockups which were developed during the conceptual phase. This concept and the mockups then are implemented, and the details of the user interface and implementation are described in the next chapter (see Chapter 6).

5.1 Definitions

First, let us define some important concepts. A competency could represent a single problem of a domain as explained in the Section 4. Competencies can depend on each other as one could be a prerequisite for the others. For instance, in order for students to be able to read words, first, they need to learn letters of the Alphabet. In this case, the words competency is depended on the letters competency. Competencies can depend (have prerequisites) on many other competencies. Let us denote the set of n competencies for a particular topic by $C = \{c_i\}, i \in [1, n]$. Additionally, let us denote the relationships between competencies as $\Theta(C) = \{\theta(c_i, c_j)\}, i \in [1, n], j \in [1, n]$ which is defined to be:

$$\theta(c_i, c_j) = \begin{cases} 1 & \text{if competency } c_j \text{ depends on } c_i \\ 0 & \text{Otherwise} \end{cases} \tag{5.1}$$

Note that $\{\theta(c_i, c_j)\}$ is not the same as $\{\theta(c_j, c_i)\}$ as the reflexivity does not apply.

Additionally, each competency can be acquired with a certain proficiency by a student. To model that, we use probabilistic modelling where probabilities are used to express the certainty that a particular student has acquired a given competency. In practice, one could think of such a probability as the percentage of acquired competency. So, let us define the set of m students belonging to a group as $S = \{s_i\}, i \in [1, m]$. For a set of competencies C and students S, we define the set of probabilities as:

$$P(C, S) = \{p(c_i, s_j)\}, i \in [1, n], j \in [1, m] \tag{5.2}$$

where $p(c_i, s_j)$ represents the probability that student j acquired the competency i.

Using the students' probabilities matrix, P one can derive a similar metric for the entire group which would express the probabilities that the entire group has acquired a certain competency c_i. This can be defined as:

$$p(c_i, S) = \sum_{j=1}^{m} p(c_i, s_j). \tag{5.3}$$

Furthermore, let us extend to the notation of C to simplify the modelling of dependencies by splitting competencies into hierarchical subsets called levels. We split them in such a way that all competencies of the level l can depend only on the competencies of the levels that are less or equal than $l-1$. This way a hierarchy of levels is created where the competencies can only depend on the competencies of lower levels. Such a structural organisation is mainly created for practical purposes, and it will be used throughout this thesis. Thus, I define the organised competencies to be the set $\hat{C} = \{c_{il}\}$ where l simply indicates that c_i belongs to the level l.

5.2 Visualisation Concept

The visualisation is a direct graph that contains nodes and edges. Both elements represent particular information which will be described in the subsections below.

5.2.1 Nodes

Each competency is represented by a node in the visualisation. The nodes are colour coded in order to represent the probabilities P as described by equations 5.6 and 5.7) If the node is coloured, it indicates that the student holds the given competency (see Figure 5.1). Otherwise, if the node appears empty (non-coloured) and in dashed lines as illustrated in Figure 5.1, it indicates that the student or the group (class) does not possess the given competency.

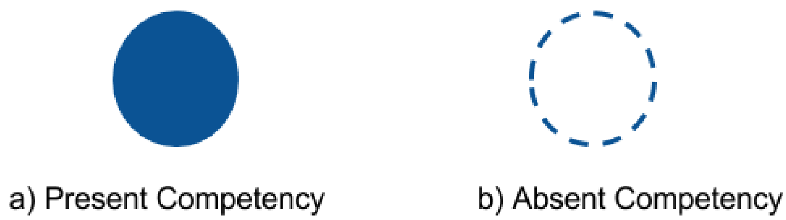

Figure 5.1: Nodes presenting the the presence or absence of a competency

Additionally, the intensity of the colour represents information as well. The higher the probability that the competency is possessed, the darker the associated node is coloured. Vice versa, the lower the probability it is, the brighter the colour appears as Figure 5.2 illustrates.

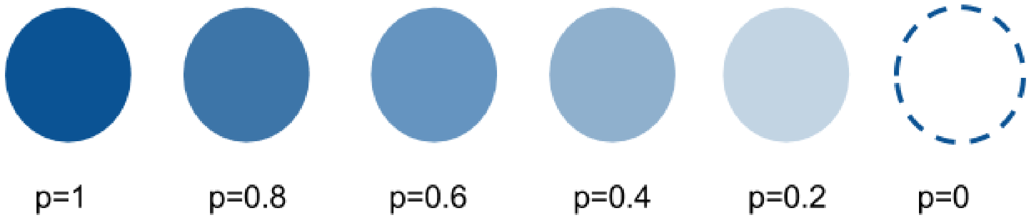

Figure 5.2: Colour coding of nodes based on probability values represented competencies

5.2.2 Edges (Node Connections) and Hierarchical Positioning

For reason of simplicity, because the visualisations of the competence structures are often too big by containing numerous competence states and thus are not scalable, I illustrate the characteristics and features of the visualisation method on the basis of the prerequisite relations. The details of the competence structures and the prerequisite relations are described in the section 4.

The edges or connections represent the inner dependencies or the prerequisites of the competencies. In the visualisation, the nodes (competencies) that are connected with lines that are dependent on each other. The lines do not indicate any direction as the direction is determined by the position of the nodes. The nodes are presented in a hierarchically structured way, with parents and children. The very beginning node, which is located at the bottom of the diagram, represents an empty set {} or a naive knowledge state. This could be the beginning of the semester or something that doesn't hold any information. We can say that this node is the parent of the all upcoming nodes.

The parent nodes, are visualised in the lower part where the children nodes are visualised one layer above. This means that the direction of the diagram reads from bottom to top. If a node one layer above is connected to a node one layer below, it means that the node in the lower layer is the prerequisite and the parent for the linked child node above. An abstract example of such a visualisation is provided in the Figure 5.3. The highest node in the hierarchy or the end node represents the end goal of the competencies which for instance could be the required knowledge at the end of a semester.

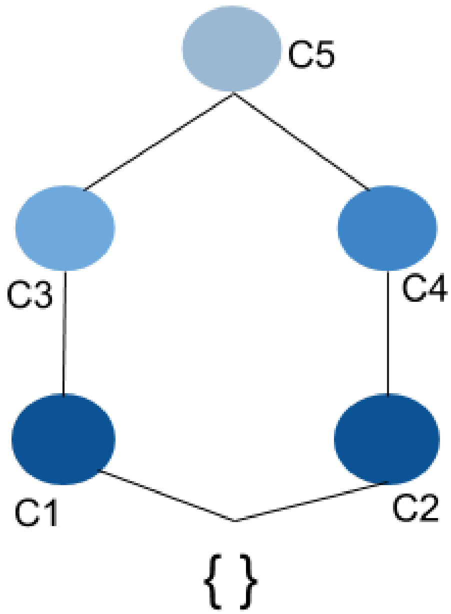

Figure 5.3: The prerequisite relations between competencies

5.2.3 Properties

Figure 5.3 elaborates how the prerequisite relations of competencies work. The competency c_1 is a prerequisite (dependency) for c_3 . Since c_1 is dependency for c_3 and c_3 is prerequisite for c_5, given the **transitivity** property, c_1 is also a prerequisite for c_5. Besides transitivity property, the digram has to other properties: reflexivity and anti-symmetry. **Reflexivity** means that the competency refers itself, so it is implicit, but it is not shown in the diagram [Kickmeier-Rust et al., 2015, Kickmeier-Rust et al., 2006]. The property of the **antisymmetry** says that the relationship between competencies is not invertible (e.g. if c_1 is prerequisite for c_5, it does not make c_5 prerequisite for c_1) [Kickmeier-Rust et al., 2015].

5.2.4 An Illustrative Example

The following example will provide an overview of how the diagram works. The idea is that students start from the essential competencies and proceed in acquiring the advanced one. For instance, let us consider the example of how the reading can be taught. It is unlikely to be that someone can read a sentence without knowing to read a word first. Similarly, users are less likely to be able to recognise words

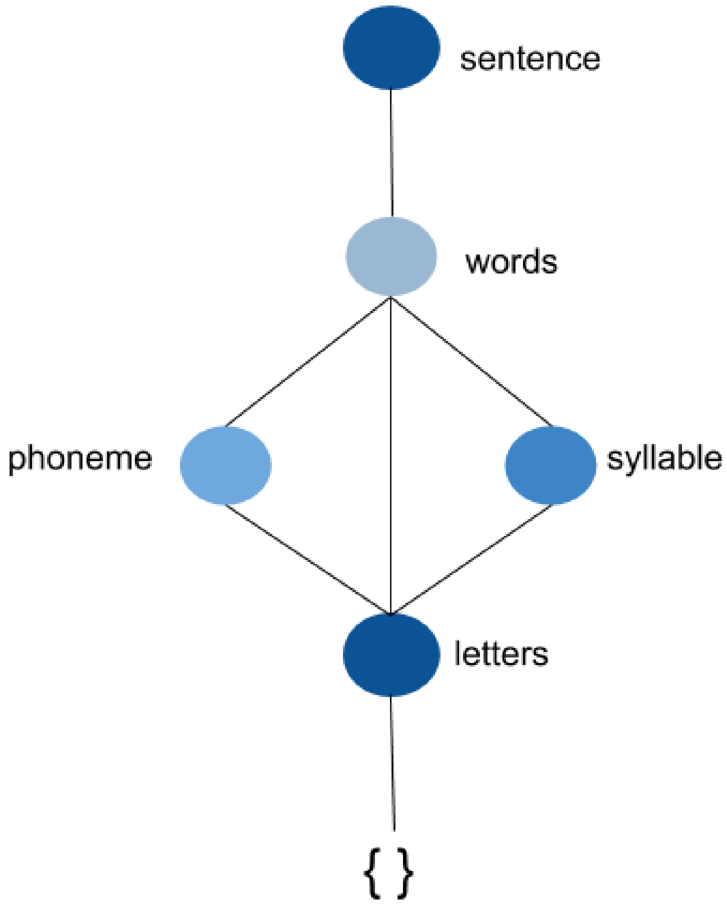

Figure 5.4: The prerequisite relations for reading

without being able to recognise either letters, phonemes or syllables.

Figure 5.4 visualises the competencies of reading. We can see that knowing the letters of the alphabet is the very first competency someone needs in order to be able to read. From letters, students could proceed either with phonemes, syllables or directly with words. It is important to note here that not all prerequisites should be fulfilled in order to move to the higher competency. For instance, for recognising words, depending on the student, one of its prerequisites would be sufficient to proceed with words.

5.2.5 Learning Path

Learning paths and learning path predictions are a property of the competence structure. However, because of the scalability purposes, we will integrate those features in the same competence based network diagram.

The competency-based network diagram presented in the above sections is a good way of grasping an overall overview of all competencies belonging to a topic. Teachers could explore in which competencies students lack the knowledge or are very competent. Nevertheless, there are many use cases, where students do not need to possess all competencies but only a subset of them. Let us reiterate the example of teaching to read presented in Figure 5.4 (see Section 5.2.4). There, the end goal is reading sentences and thus reading the text which requires recognising words. Recognising words requires either recognising syllables, phonemes or letters, meaning that not all of those three prerequisites are necessary to master the reading. In such situations, teachers might want to know how students master the word competency, which path they do follow. Such information might help teachers to reaffirm that the used teaching strategy is having success. Alternatively, teachers could use such information to adjust their strategy. Moreover, not all students learn using the very same approach. Different students might take different paths to mastering top layer competencies.

Thus, the learning path represents the history of learning, and it should display the 'road' of how the learners achieved the knowledge over a certain time. In term of competencies, in our case, the learning path at a certain competency shows simply the path through all competencies on lower levels that are achieved to achieve the current competency. The advantage of visualising such a path, is beneficial for an educator for documenting the student's accomplishments, let's say over a semester and also for planning the forthcoming activities. It helps the educators and the students to have a quick summary (overview) of how they progressed during a semester or a course seeing this way which competencies they handle at the end.

Formal Definition of the Learning Path

Let us formally define the learning path. First, let us consider a particular competency c_{il} (where l represents the hierarchical level) using notation of hierarchical competencies $\hat{C}$ presented in Section 5.1 Second, let us define a competencies propagation path related to a particular competency c_{il} which represents any path connecting the node of c_{il} with the zero node (c_{00}) containing a single node for each level below l. So a competency propagation of c_{il} is a subset of competencies which includes c_{il} and c_{00} as given in Equation 5.5:

$$CPP(c_{il}) = \{c_{00}, c_{x1}...c_{y(l-1)}, c_{il}\} = \{cpp_0, cpp_1...cpp_{l-1}, cpp_l\} \tag{5.4}$$

where l is the level and $c_{x1}...c_{y(l-1)}$ (which is the same as $cpp_1...cpp_{l-1}$) represent one competency for each level $1...l$. The notation cpp_k simply represents competency of the CPP which is located in the level k ($cpp_0 = c_{00}$ and $cpp_l = c_{il}$).

Note that every combination of sub level competencies can form a competency propagation path according to Equation 5.5. But not all of them are valid. Such a propagation path would be valid only if every subsequent competency are connected with each other through the dependency/prerequisite relation. So the given condition must hold:

$$\theta(cpp_k, cpp_{k+1}) = 1,\ \forall\ k \in \{0, l-1\} \tag{5.5}$$

Given that each competency is associated by a probability, either for a single student or a group, we can calculate the entire probability of competencies propagation path as a joint probability of each competency within the path. So, for a group of students S, such the probability of a CPP for a given competency c_{il} is defined as:

$$p(CPP(c_{il}), S) = \prod_{k=1}^{l} p(cpp_k, S)\ \theta(cpp_{k-1}, cpp_k) \tag{5.6}$$

where $p(cpp_k, S)$ represents the probability of the competency cpp_k for the set of students S as defined in Equation 5.3. Note that $\theta(cpp_{k-1}, cpp_k)$ represents the dependency of two successive competencies as defined in Equation 5.1 and it is used to filter out the CPPs that are invalid as in such cases $\theta(cpp_{k-1}, cpp_k) = 0$ and thus the entire probability would result in 0. Analogously, using Equation 5.2, the probability of a CPP can be defined for a single student s_j as:

$$p(CPP(c_{il}), s_j) = \prod_{k=1}^{l} p(cpp_k, s_j)\ \theta(cpp_{k-1}, cpp_k) \tag{5.7}$$

For a given diagram D and with a set of competencies C, for most of the competencies there exist many competencies propagation path each resulting in different probabilities. However, most of them are less probable or not valid at all. So we define the **learning path** for a competency c_{il} to be the CPP(c_{il}) with the highest probability. So for a set of all possible CPPs denoted as $CPPS = \{CPP_j(c_{il})\}$ belonging to the competency c_{il}, the learning path $LP(c_{il}, S)$ (for a set of students S) is defined to be:

$$LP(c_{il}, S) = \underset{CPP(c_{il})}{\arg\max}\ p(CPP(c_{il}), S) \tag{5.8}$$

Analogously, learning path can be defined for a single student s_j as:

$$LP(c_{il}, s_j) = \underset{CPP(c_{il})}{\arg\max}\ p(CPP(c_{il}), s_j) \tag{5.9}$$

Solving the equations 5.8 and 5.9 will result in choosing a competencies propagation ($CPP(c_{il})$ with the highest probability for a given set of students S or a single student s_j which is defined to be the learning path with respect to the given competency.

Visualisation

In the visualisation, the learning path is highlighted by reducing the visibility of all other nodes that do not belong to the learning path. Similarly, all edges of the diagram that do not belong to the learning path are made less visible. An example of a learning path visualisation is provided in Figure 5.5. We can see that the learning path (for the competency c_6) drawn here indicates the way of reaching competency c_6 through the competencies c_4 and c_1. The other competencies and edges are still visible enough to indicate that they exist, but with reduced visibility to indicate that they are not a part of the learning path.

Obviously, the user needs to indicate for which competency the learning path needs to be shown. Additionally, the user should be able to ask for the learning path of competency only if such competency has a probability greater than 0, which means that student(s) already posses it up to a degree. However, taking those considerations is a task of user interface implementation and interaction design which is out of the scope of this chapter but described in Chapter 6.

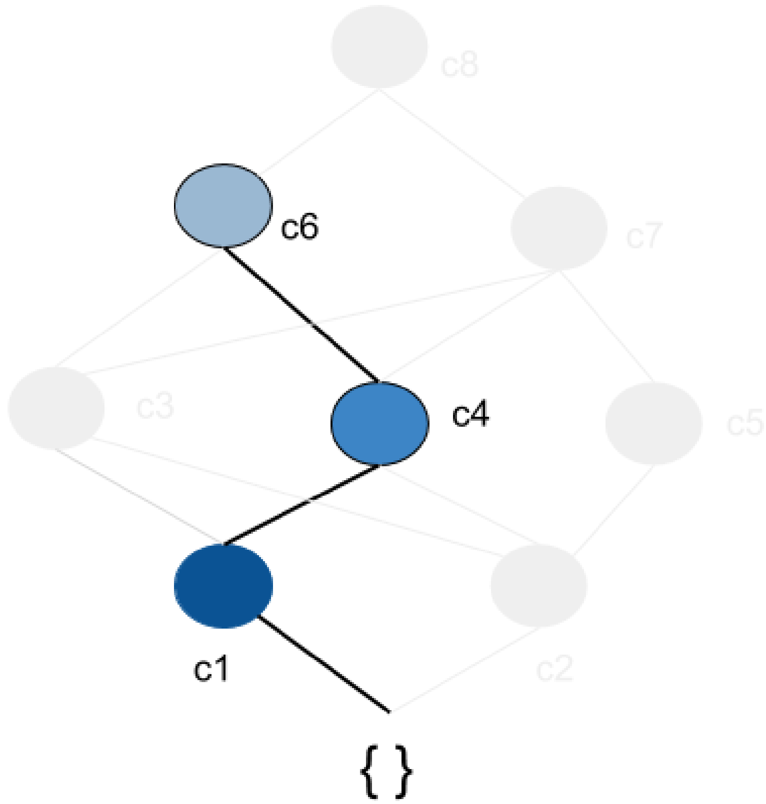

Figure 5.5: Learning path for the competency c_6

5.2.6 Forward Learning Path Prediction

Learning path is a good way to observe how users gain competencies through time. However, for teachers, it would be useful not only to show the learning path up to a particular competency c_{il} but also to show a prediction on how students would most likely continue from the c_{il} to the top competency c_{nm}, where m indicates the complete depth of the diagram or the most upper level. Such a prediction is a sort of forward learning path (from the given node to the top), but it could also be provided in the absence of the data. In case there exist data (assigned probabilities for some students) for the competencies in the levels $l+1$ to m, we can use the probabilities of such nodes and then utilise the same formula defined in equations 5.8 and 5.9 with some small adaptations. The only adaptation we need to do is that when building the set of all competencies propagation paths, instead of using the paths starting at node c_{00} and ending at node c_{il}, we use the paths that start at c_{il} and end at

c_{nm}. This small adaptation would be sufficient to calculate the path from c_{il} to c_{nm}, which is rather a forward calculation instead of a prediction of the learning path.

However, during the teaching process, in most of the time-points, there will be no data for all competencies as the teachers will probably not have covered the planed material. For instance, in the middle of the semester, the competency-based network diagram will likely be coloured only in the half bottom part. This means that there will be data (assigned probabilities for competencies) only for the nodes in the bottom half of the diagram. As the time passes and teachers cover more materials, the students will acquire more competencies. Thus resulting in the diagram being more coloured and thus as more nodes will have assigned probabilities. So, in such cases, where for some upper part of the diagram, no data is assigned yet, the forward path cannot be calculated as all competencies propagation paths will have 0 probability. Instead, a prediction is required.

Thus, we assign each to competency a default probability which is set by experts (teachers), and it reflects their opinion of how probably learners will achieve this in future. It is expected, that experts take into consideration how hard, time consuming and complex is the task of acquiring a particular competency when providing such a default probability. Such a default probability can be used to predict the forward path, in case the that the real probabilities (which is based on real evidence on data) are missing. So for the forward path, if a probability is missing for competency, the default probably will be used instead. Due to triviality and simple adjustment for the forward learning path prediction in comparison with the learning path, the formal definitions will be skipped.

The relevance of learning path prediction has been in depth discussed by [Kickmeier-Rust and Albert, 2016a]. The authors define weights for competencies which define the difficulty of achieving a competency. Such weights are used then to provide a learning path prediction. The concept of foreward learning path prediction used in this thesis is based on the aforementioned work [Kickmeier-Rust and Albert, 2016a]. However, it makes some small practical adjustments. First instead of using the pre-defined weights, in case there is data available for given nodes, it uses such data. Second, instead of defining weights it defines default probabilities for the nodes for implementation practical. This way the same implementation of learning path can be used to visualise path prediction by simply replacing probability with default probability in case the students' data is missing. However, one could think of default probabilities as a parameter which could be derived from the weights as described by [Kickmeier-Rust and Albert, 2016a], with an inverse relationship between them.

The forward learning path prediction will be highlighted in a similar fashion as the learning path, in the sense that the nodes are part of it are more visible the ones that are not part of it or part of the learning path. However, there will be a distinction in terms of edges. The edges of a foreword learning path prediction will be visualised using dashed lines as opposed to continuous lines used in learning path. An abstract example of the foreword learning path prediction is illustrated in Figure 5.6. It demonstrates the learning path prediction of how competency c_8 can be achieved by a student starting from the competency c_4, whereas down below c_4 is represented the learning path.

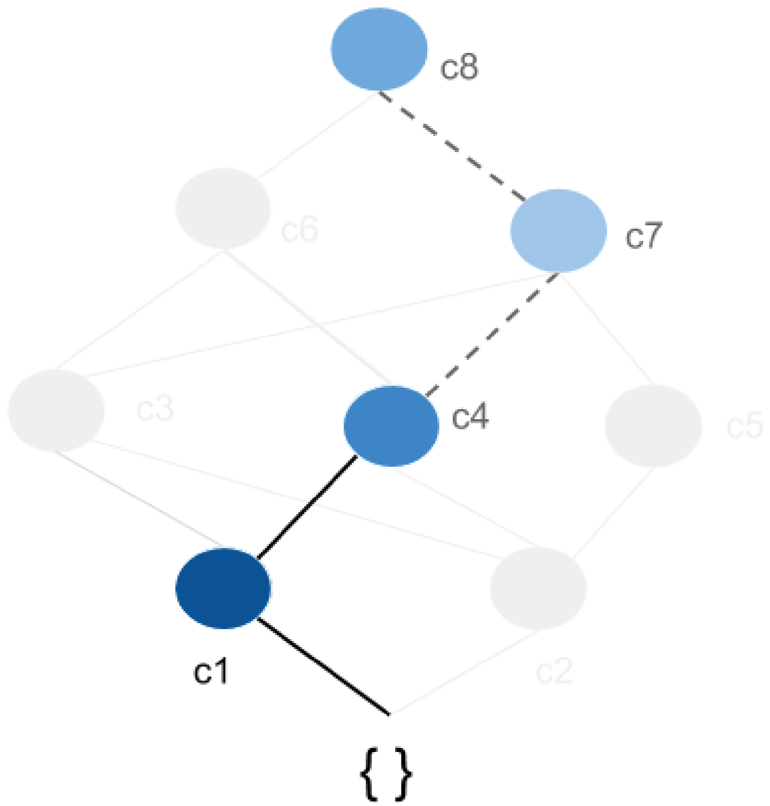

Figure 5.6: Learning path and learning path prediction for competency c_4

5.2.7 An Overview of Visualisation Concept

This chapter described a concept of competency bases network diagram. The diagram is designed to give overview your all competencies acquired by the learners as well as prerequisites and dependencies between competencies. This diagram is designed to enable teachers to inspect the achievement of competencies for a single student or a group of students by a single look at it. Using the colour coding, they can immediately spot which competencies are achieved and where the students are lacking. Such information can be used to adopt, continue with used teaching strategies, reiterate or skip some parts of the curriculum as teachers can react based on the information provided by the visualisation.

Besides the visualisation of competencies and their dependencies, the diagram is designed to provide the teachers with the learning path, which visualises the history of learning and it shows the path of how the learners propagated throughout the competencies over time. Additionally, it is designed to visualise the prediction of the foreword learning path which shows the prediction of how learners will continue to acquire competencies.

Besides developing the visualisation and appearance in this chapter, the encoding of information and necessary formal mathematical and probabilistic representation are created. Probability values for students are assigned to nodes or competencies, which represent the probability value that a student has acquired the competency. In practice, this can represent the percentage of the knowledge of that particular competency compared to the entire taught knowledge. Throughout this thesis, it will be assumed such probabilistic representation. The way to create such probabilistic values (how to know how much percent a student acquired a competency, e.g. using the score of a test) will not be discussed as it is out of the scope of this thesis. Using the probabilistic representation of the values of the nodes, one can use the visualisation for a single user or simply adopt it for a group of users by using the Equation 5.6. Additionally, the probabilistic representation allows the derivation of probabilistic calculations for a competencies propagation path as a joint probability of the single competencies. The equations for calculating probabilities are defined in equations 5.6 and 5.7. Such calculations are furthered utilised to define an optimisation problem (see equations 5.8 and 5.9) which enables finding the competencies propagation path with the highest probability which is used to find the learning path and foreword learning path prediction.

The visualisation diagrams presented in this chapter are only mockups, conceptual and thus no user interaction is provided in this phase. However, they together with mathematical and probabilistic formulation defined here, lay the foundation for implementation and user interaction development which will be described in the next chapter.

Chapter 6

User Interface, Interaction Design and Implemenation

6.1 User Interface and Interaction Design

This section describes the implemented user interface provided, implemented visualisation as well as the interactions that are provided with the user interface. The diagram visualisation is an implementation of the conceptual design. For details on the visualisation concept, the mathematical definitions and background regarding the visualisation, please refer to the Chapter 5.

6.1.1 Authentication

Before using the application, the user needs to register at the central user management system. After creating the credentials, the user can go to the homepage and authenticate using the username and the password as in shown the Figure 6.1 in order to proceed and use the application. The prototype is designed and developed to be used by teachers in a very easy and intuitive way. Thus we will use the following sections user and teacher interchangeably.

6.1.2 Visualisation and Interaction

After the authentication process, the teacher will reach the main page of the application from where she/he can start using the application as shown in the Figure 6.2.

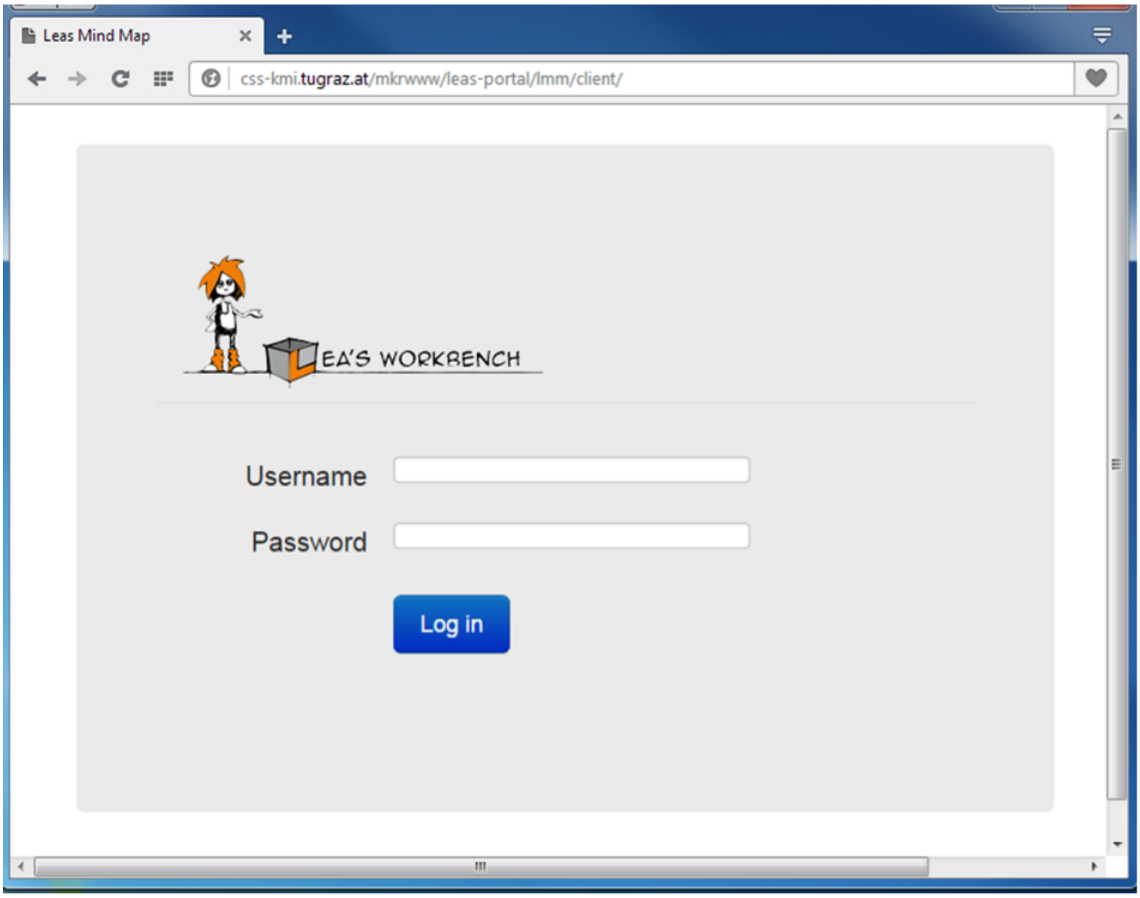

Figure 6.1: Authentication page

There are two dropdown menus shown on the screen for the user to select:

- Group - which represents the group of students. Typically a group represents all students within a classroom. Although, it does not have to, as groups can be defined according to the needs of the teachers.

- Subject - which represents a common topic for which a set of competencies are defined. Typically this represents course, but it could represent a single topic as well. Teachers are free to be more creative and play with the granularity of the subject according to their needs.

Each teacher is assigned to have access (in the database) only to certain groups, e.g. only classes they teach. This way teachers can access only to those groups and subjects that they are responsible for. The selection of the group is quite simple, by

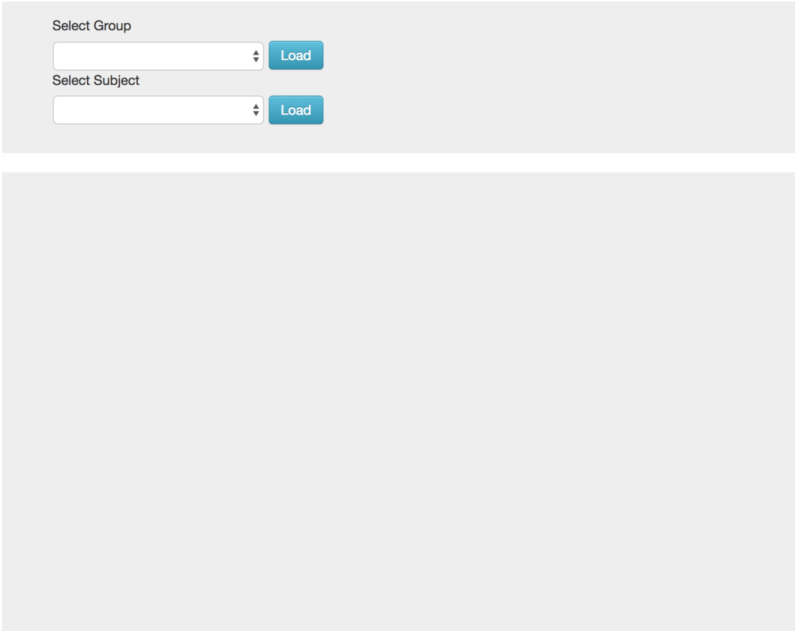

Figure 6.2: Main page

just clicking on the boxes where the user can choose between the groups, as shown in the Figure 6.3.

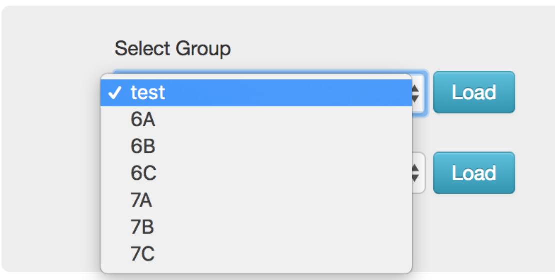

Figure 6.3: Selecting a group

After selecting a group, all students that are already assigned to the selected class will be listed on the right part of the window, each of them visualised by an avatar. This is illustrated in the Figure 6.4. The teacher can also choose one of the subjects that are related to the selected group. The process of the selection of the subject is same as for the group (see Figure 6.5).

By selecting the subject, the competency-based network diagram representing the competencies of that subject will be presented in the lower part of the page.

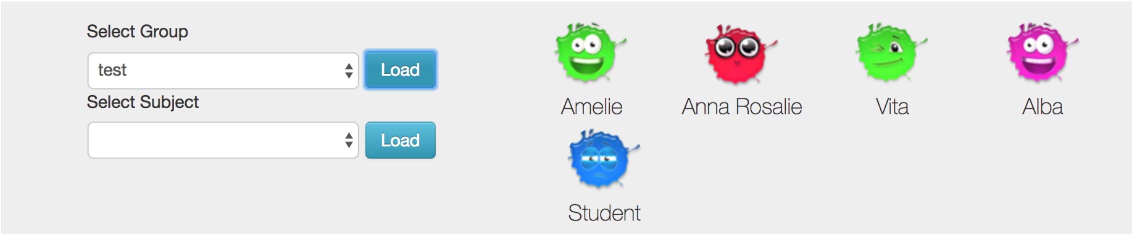

Figure 6.4: List of students that belong to the selected subject

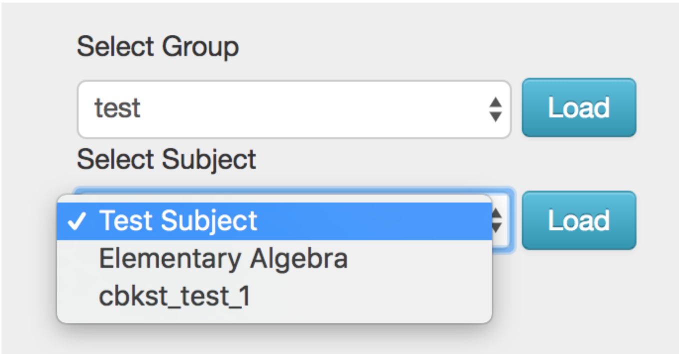

Figure 6.5: Selecting a subject

This is the most important part of this user interface. An example of such a diagram is provided Figure 6.6. As it is already described in the concept (see Chapter 5), the diagram is visualised as a set of hierarchical nodes connected to each other, where each node represent a competency. Note that the nodes have different colours as the colour encodes the achievement of that competency by a single student or a group of students. The intensity of the blue colour represents the probability of the competence achievement. It means, that the higher the intensity, the higher the probability of competency achievement and vice versa. When the node is not coloured (empty), then the student has no awareness or say it better does not own that competency at all.

The diagram can be used to indicate the competences of the whole group, or it can be used to visualise the competences of a particular student. The user chooses the data underlying the diagram through the interactions. When no student is selected, then the data of the entire group is used whereas when a student is selected then the data of only that particular student is used. To select a student, the user needs to simply click on the avatar. After selecting a student, the user can deselect him/her by clicking again on the selected student, in which case the data of all

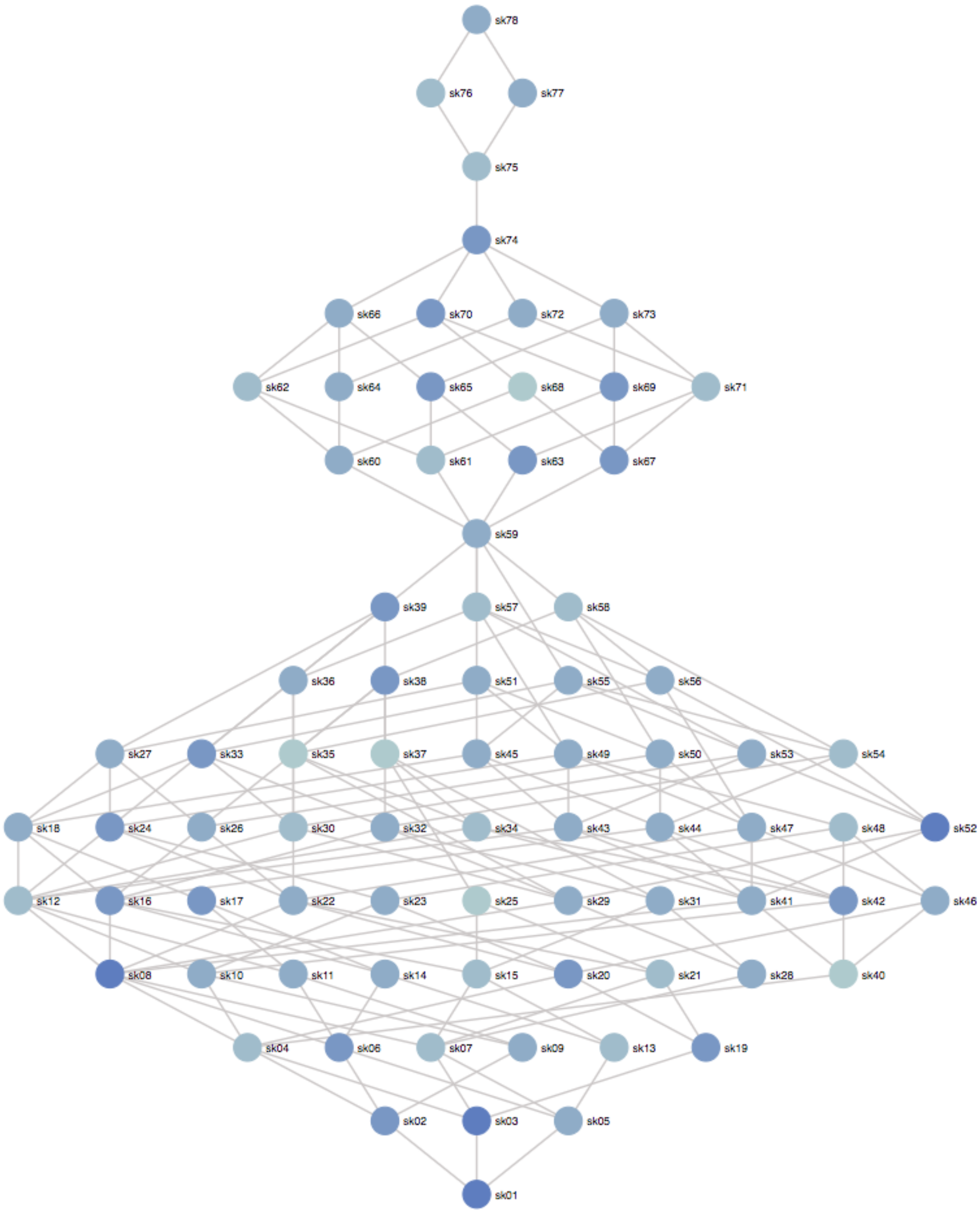

Figure 6.6: The competency based network diagram for all students of a class

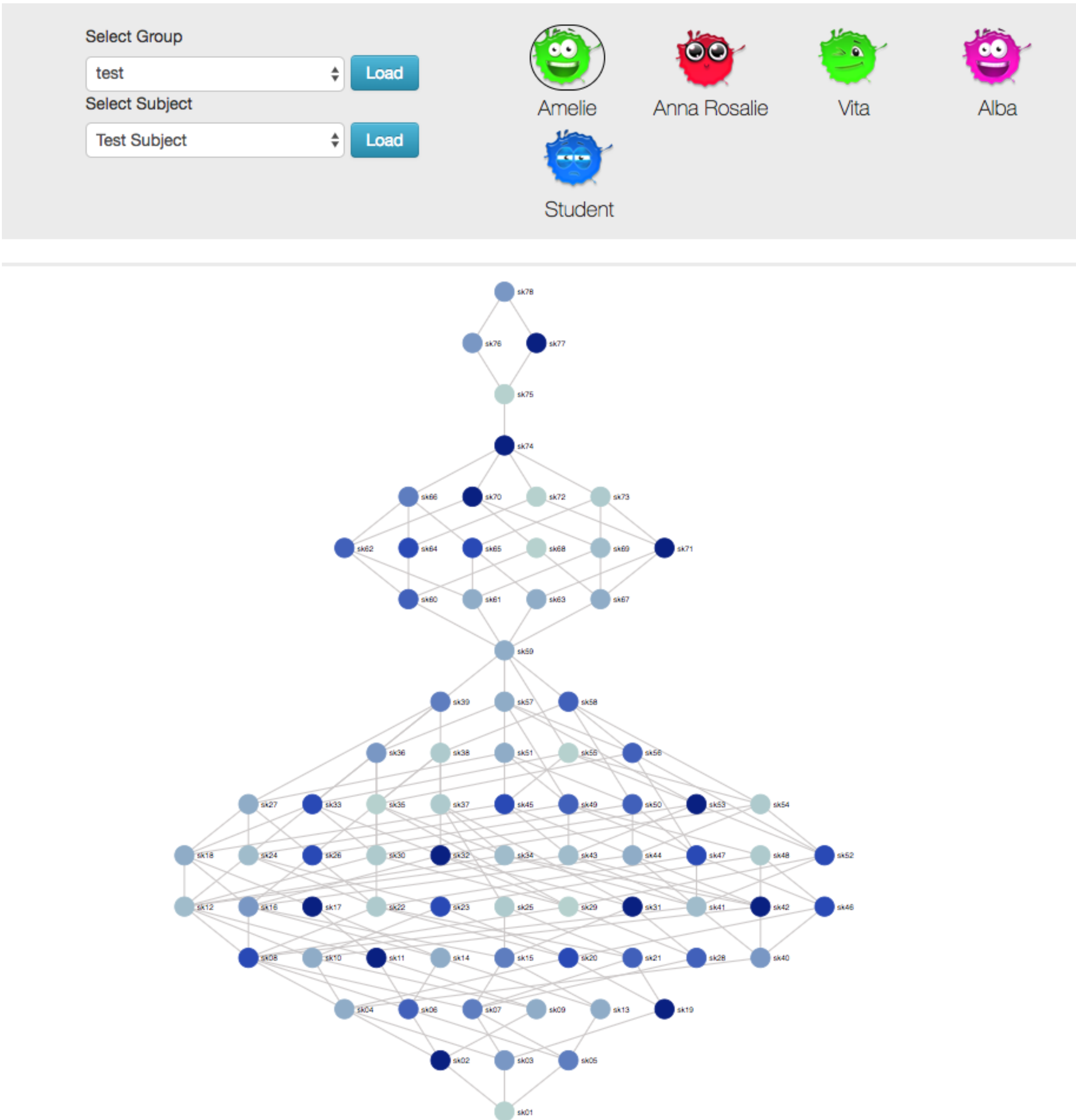

Figure 6.7: The competency based network diagram for a single student

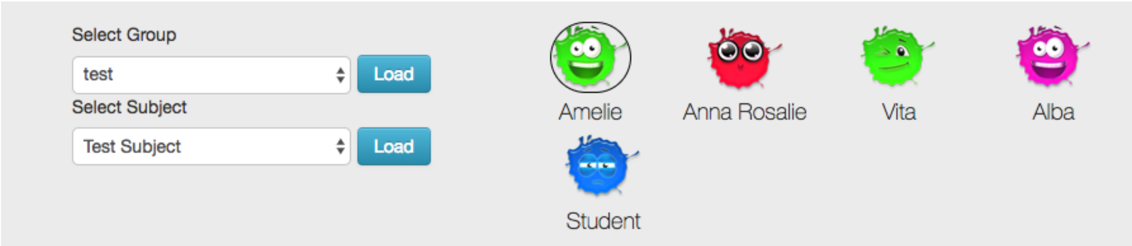

Figure 6.8: A student (Amelie) is selected which is visualised by the circle around the avatar

students is used, or select another student. When the student is selected, the avatar indicates the selection by a visualised circle around the avatar (see Figure 6.8). Additionally, the diagram will update to represent the data of the selected student only. The colours of the nodes change from student to student or to a group of them as the underlying data change. For instance, the diagram of the entire class, shown in Figure 6.6, looks different from the one of a single student which in shown in Figure 6.7.

6.1.3 Learning Path and Foreword Learning Path Prediction

Besides visualising competencies and their interconnections in terms of prerequisites, the competency-based network diagram is capable of visualising the leaning path and forward learning path prediction. Such paths are bound to a particular node/competency. In order to highlight the path, the user needs to select a node by clicking on it. Upon clicking on the node, both the learning path and the forward learning path prediction for that node and underlying data are highlighted as depicted in Figure 6.9. The learning path from the bottom node to the selected node is visualised using a line. Similarly, the forward learning path prediction is shown from the selected node to the top node where the edges are presented in dashed lines. All other nodes and edges that do not belong to both paths are displayed with reduced visibility, which is still visible but not highlighted. The user can select anywhere in the canvas to deselect the node, which would remove the highlight if the learning path and would return the view of the entire diagram as shown in figures 6.9 and 6.6. It is evident that their paths to the same competency are very different from each other.

The underlying data for the learning path, similar to the underlying data for the entire diagram, depends on whether a particular student is selected or not. If yes, then the data of that student is used. Otherwise, the data of the entire group is used. Note that, for the same diagram and same competency, different students will probably have different learning paths. For instance a learning path for two students for the same competency (sk_{65}) is shown in figures 6.9 and 6.10.

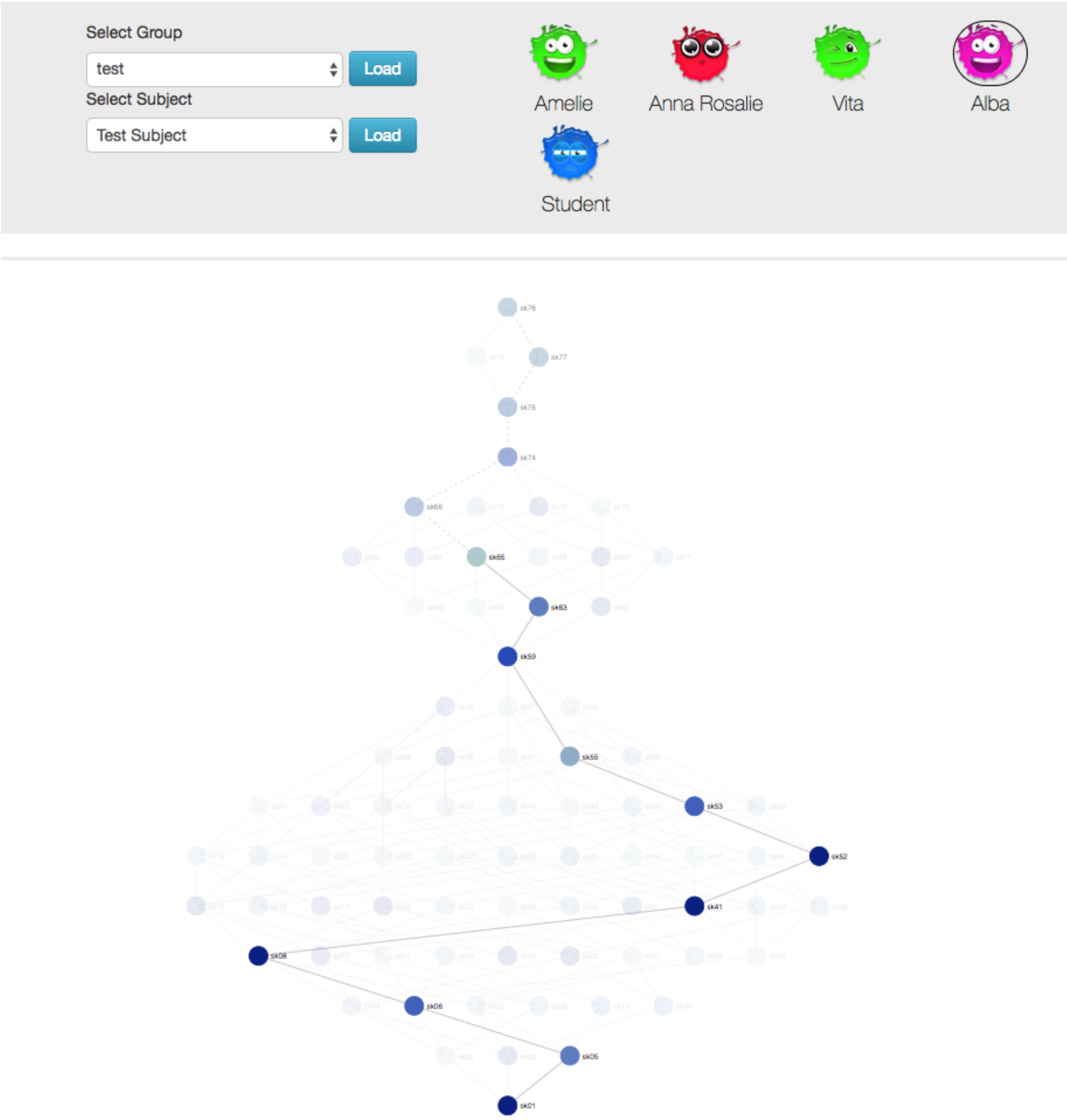

Figure 6.9: Learning path of student Alba

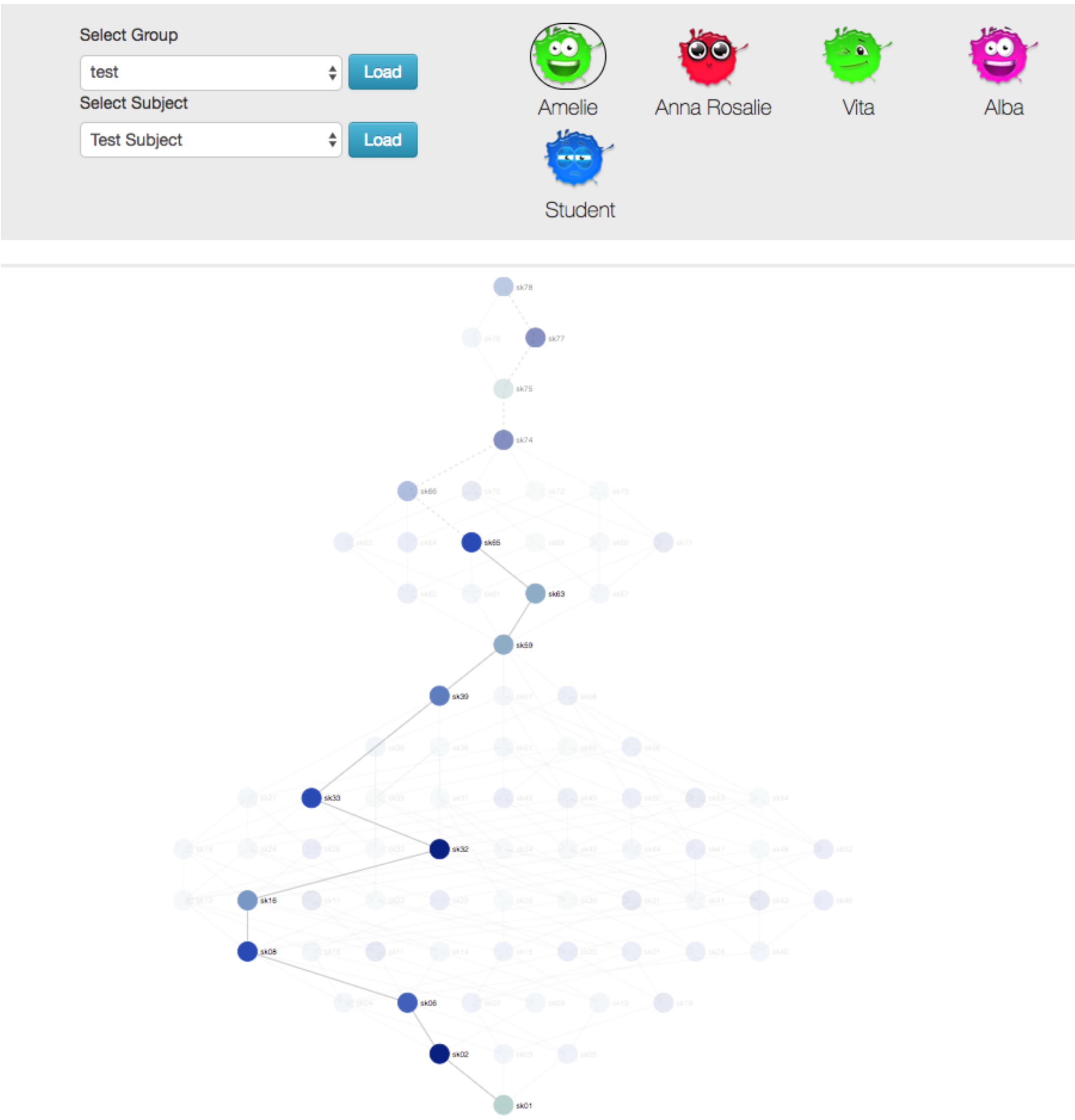

Figure 6.10: Learning path of student Amelie

6.2 Technical Implementation

The entire project is implemented using a client-server architecture as visualised in Figure 6.11. The server is provided as a RESTful service. Its job is to accept the GET request, fetch the data from database and then provide the response in the JSON format. Given that the manipulation(insert, delete, update) of the data is out of the scope of this thesis, only the GET request is used. The server is implemented in PHP, and it uses Slim Framework [1] for providing the RESTful service. The database uses MySQL [2] as its backend.

On the other hand, the role of the client is to provide a user interface to the user, detect the user interaction, send the proper request to the server, parse the response and finally present and visualise the result to the user. The user interface is implemented in HTML5/CSS3, and it uses the Bootstrap Framework [3] for providing theme, layout and user controls. Additionally, the competence-based network diagram is developed using D3.js Framework [4].

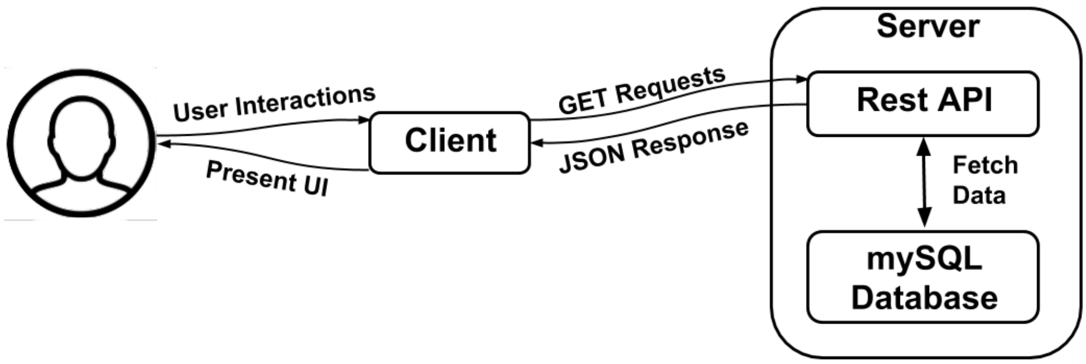

Figure 6.11: The client server architecture used in the developed prototype

In the following subsection, brief descriptions of some of the essential technical aspects of the prototype will be outlined.

[1] https://www.slimframework.com/
[2] https://www.mysql.com/
[3] https://getbootstrap.com
[4] https://d3js.org/

6.2.1 Database

In this section, it will be discussed the database storage used for storing the visualisation and student data. Although the database of the project contained dozens of tables (53), we will only focus on the tables that are used for the visualisation and ignore the rest. Other tables which store information regarding users, schools, subjects, etc. will not be discussed.

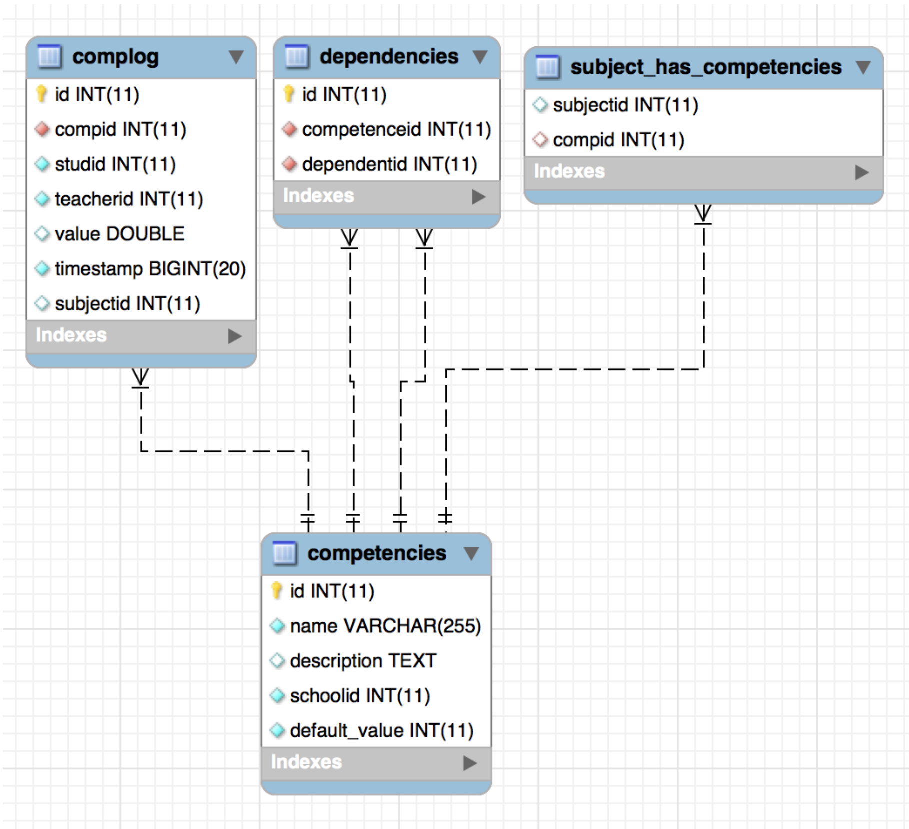

Figure 6.12: Database tables that are relevant for the visualisation.

Figure 6.12 shows the four relevant tables used for visualisation which include:

- competencies - which saves all competencies
- dependencies - which is used to define prerequisites meaning which competency depends on which competency
- subject_has_competencies - which is used to model the relation between subject and competencies. This way competencies can be assigned to subjects.

- complog - which is used to log students performance during the semester. Note that, as the teacher provides many entries during the semester, one should need to pay attention only to the latest entries in order to get the most accurate picture.

Please note that some of the relevant table columns are not displayed for readability purposes. To visualise the competency-based network diagram for a given subject, one needs information for competencies that belong to the subject, their dependencies and student achievements for the given competencies (for every user and aggregated). To collect such data, the three queries displayed in listenings 6.1, 6.2 and 6.3 are used.

```
SELECT *
FROM dependencies
WHERE competenceid IN (
          SELECT C.id
          FROM competencies C
          INNER JOIN subject_has_competencies SHC
          ON SHC.compid = C.id
          WHERE SHC.subjectid = $subjectid
          )
AND dependentid IN (
          SELECT C.id
          FROM competencies C
          INNER JOIN subject_has_competencies SHC
                    ON SHC.compid = C.id
          WHERE SHC.subjectid = $subjectid
)
```

Listing 6.1: SQL query to obtain all dependencies for the given subject identified by $subjectid.

```
SELECT
          C.id,
          C.name,
          C.description,
```

```sql
        C.default_value,
        AVG(IFNULL(CL.value,0)) AS value
FROM competencies C
INNER JOIN subject_has_competencies SHC
        ON SHC.compid = C.id AND SHC.subjectid = $subjectid
LEFT JOIN complog CL
        ON CL.compid = C.id AND CL.id IN (
                SELECT MAX(id)
                FROM complog
                GROUP BY compid, studid
        )
GROUP BY C.id
```

Listing 6.2: SQL query to obtain the competencies (values averaged per user) for the given subject identified by $subjectid.

```sql
SELECT
        C.id,
        C.name,
        C.description,
        C.default_value,
        IFNULL(CL.value,0) AS value,
        CL.studid,
        CL.timestamp
FROM competencies C
INNER JOIN subject_has_competencies SHC
        ON SHC.compid = C.id AND SHC.subjectid = $subjectid
LEFT JOIN complog CL
        ON CL.compid = C.id AND CL.id IN (
                SELECT MAX(id)
                FROM complog
                GROUP BY compid, studid
        )
```

Listing 6.3: SQL query to obtain the competencies of all users for the given subject

Url	**Prams**	**Method**	**Description**
/login	userid, password	POST	authenticates user
/logout		GET	logs out
/authenticated		GET	returns authentication status
/dependencies/:subjectid	subjectid	GET	returns all dependencies for a subject
/compdep/:subjectid	subjectid	GET	returns all competencies, dependencies and the students' probability values for a given a subject
/groups		GET	returns all groups that the teacher is allowed to access
/students/:groupid	groupid	GET	returns all students that belong to a group
/subjects		GET	returns all subject that the teacher is allowed to access

Table 6.1: Restful API for the prototype.

identified by $subjectid.

6.2.2 Server API

The server API offer the entire functionality needed for the prototype. It provides functionality related to the visualisation of the diagram as well as functionalities related to the website such as login and log out. The API functionality is given in the Table 6.1. The response is provided in JSON format.

6.2.3 Visualisation

The entire visualisation is built in HTML5/Javascript, and it uses D3.js as visualisation framework. In this section, we will shed light on two crucial algorithms used to construct the visualisation. The initial step on visualising is building the hierarchy as the data are delivered merely with anchors which describe the competency dependencies, but no hierarchy is stored in the database. Thus an algorithm for constructing the hierarchy and assigning each node to the appropriate level of hierarchy was developed.

The main intuition of the algorithm is to start with the head node and then proceed in with next levels wherein each level are placed all the competencies that have prerequisites only in lower levels of hierarchy. The algorithm is listed in Listening 6.4.

```
function generateLevels(competencies, dependencies){
  var competenciesDict = {};

  for (var i = 0; i < competencies.length; i++){
    var c = competencies[i];
    competenciesDict[c.id] = c.name;
  }

  var levels = {};
  var keys = Object.keys(competenciesDict);
  var j = 0;
  var previousKeys = [];
  while(keys.length > 0){
    levels[j] = keys.slice(0);

    for (var i = 0; i < dependencies.length; i++){
      var key = dependencies[i].dependentid;
      var cid = dependencies[i].competenceid;
      if(j == 0){
        var ix = levels[j].indexOf(key);
        if(ix >= 0){
          levels[j].splice(ix, 1);
        }
      }else{
        var ix = levels[j].indexOf(key);
```

```
        var cix = keys.indexOf(cid);
        var pix = previousKeys.indexOf(cid);

        if(ix >= 0 && pix < 0){
          levels[j].splice(ix, 1);
        }
      }
    }
    for(var k = 0; k < levels[j].length; k++){
      previousKeys.push(levels[j][k]);
      keys.splice(keys.indexOf(levels[j][k]), 1);
    }

    j++;
  }
  return levels;
}
```

Listing 6.4: Building the hierarchy for the competency-based network diagram for the given competencies and dependencies

Once the hierarchy is built, and nodes have been assigned, then the visualisation simply assigned the positions of each node according to the place in the hierarchy and some visualisation settings such as the horizontal distance between nodes, the vertical distance between levels of hierarchy, node size, etc... The algorithm for creating the visualisation is given in Listening 6.5.

```
function visualise(competencies, dependencies, width){
  var xStart = 20;
  var xDistance = 100;
  var yDistance = 80;
  var nodes = [];
  var edges = [];
  nodesDict = {};

  var levels = generateLevels(competencies, dependencies);
```

```
var levelsReversed = {};
var lKeys = Object.keys(levels);
var maxLevelSize = 0;

for (var j = 0; j < lKeys.length; j ++){
  var k = lKeys[j];
  for (var i = 0; i < levels[k].length; i ++){
    levelsReversed[levels[k][i]] = k;
  }

  if(levels[k].length > maxLevelSize){
    maxLevelSize = levels[k].length;
  }
}

var yStart = lKeys.length*yDistance-20;

for (var index = 0; index < competencies; index++) {
      var competency = competencies[index];
      var levelSize = levels[level].length;
  var levelIndex = levels[level].indexOf(competency.id);

  var cxStart = xStart + (width - xDistance*levelSize)/2

  var x = (cxStart + levelIndex*xDistance);
  var y = (yStart - level*yDistance);
  var probability = competency.value;
  var def_probability = competency.default_value;

  var node = createNode(x, y, probability, def_probability)
  nodesDict[node.id] = node;
  nodes.push(node);
}
```

```
  for (var index = 0; index < dependencies; index++) {
    var dependency = dependencies[index];
    var sourceid = dependency.competenceid;
    var targetid = dependency.dependentid;

    var source = nodesDict[sourceid];
    var target = nodesDict[targetid];
    var edge = createEdge(source, target);
    edges.push(edge);
  }

  draw(nodes, edges);
}
```

Listing 6.5: Algorithm for drawing nodes and edges between nodes.

An essential aspect of the visualisation and the diagram, in general, is the learning path. The algorithm for finding and drawing the learning path for a given node is provided in the Listening 6.6, which is an implementation of a solution the optimisation problem defined in Equation 5.8.

```
function visualiseLearningPath(currentNode, links){
  var lastLevelNodes = [currentNode.id];
  var pathNodes = [[currentNode.id]];

  while (true){
    var c = 0;
    var newLevelNodes = [];
    var pathNodesTemp = [];

    for (var i = 0; i < edges; i++) {
      var link = edges[i];
      if($.inArray(link.source.id, lastLevelNodes) > -1 &&
         $.inArray(link.target.id, lastLevelNodes) == -1){
        c++;
```

```
        newLevelNodes.push(link.target.id);

        for (var i = 0; i < pathNodes.length; i ++){
          arr = pathNodes[i];
          if(arr[arr.length-1] == link.source.id){
            narr =  arr.slice(0);
            narr.push(link.target.id);
            pathNodesTemp.push(narr);
          }
        }
      }
    }

    if (c == 0) break;
    pathNodes = pathNodesTemp;
    lastLevelNodes = newLevelNodes;
  }

  var proabbilities = pathNodes.map(function(pathNode){
    return pathNode.reduce(
      function(previousValue, currentValue, currentIndex) {
        var res = localVars.nodes.filter(function( n ) {
          return n.id == currentValue;
      });
      return previousValue* res[0].probability;
    }, 1);
  });

  var bestPathNodeIds = [];
  var maxProba = Math.max.apply(Math, proabbilities);
  var maxProbaIx = proabbilities.indexOf(maxProba);
  var bestPathNodeIds = pathNodes[maxProbaIx];

  var bestPathLinks = [];
```

```
  for (var i = 1; i < bestPathNodeIds.length; i++){
      var targetid = bestPathNodeIds[i-1];
      var sourceid = bestPathNodeIds[i];

    var res = localVars.links.filter(function( l ) {
        return l.source.id == sourceid &&
                l.target.id == targetid;
    });
    bestPathLinks.push(res[0]);
  }

  draw_highlight(bestPathNodeIds, bestPathLinks);
}
```

Listing 6.6: The algorithm for finding and drawing the learning path for a given node.

6.2.4 Technical Evaluation

This section will provide a technical evaluation of the prototype. First, it will evaluate the capabilities in terms of required resources for the visualisation of the diagram as well as for the learning path and learning path prediction. Second, it will provide some limitations of the prototype.

First, let us see some limitations of the prototype. The students are organised into rows and displayed in pairs of 4 for a row as shown in Figure 6.6. The space holding them will adjust depending on the number of students. However, in case the classrooms with many students, a better approach would be to reorganise the space of students by fitting them more within a row (e.g. smaller icons).

The visualisation itself adopts as well. At the initial zoom level (at which the diagram is initially visualised), ten nodes fit at one row of each hierarchical level. Whereas for the number of levels, the visualisation adjusts its height to fit them. Note that, users can still zoom in and out (by scrolling), to visualise the entire diagram in case levels contain more than ten nodes and thus in practice this is not a limitation of the visualisation.

Additionally, to provide an overview of the technical capabilities, we present measurements of the time it takes to visualise the entire diagram as well as the time it takes to find and visualise the learning path and forward learning path prediction. The visualisation of the diagram includes the organisation of nodes into levels of hierarchy which is given by Algorithm 6.4 and the presentation of nodes and connection in the hierarchy as given by Algorithm 6.5. On the other hand, the implementation of the learning path is given by Algorithm 6.6. The evaluation for discrete samples of the number of nodes, links and levels hierarchy is given in Table 6.2.

Levels	**Width**	**Nodes**	**Links**	**Visualisation**	**Learning Path**
2	2	6	8	4	1
3	3	11	24	4	2
5	5	27	70	6	3
5	10	52	140	9	4
10	10	102	290	19	134
15	10	152	440	37	540
20	10	202	590	65	10503
30	10	302	890	170	1504660
40	10	402	1190	370	
50	10	502	1490	668	
60	10	602	1790	1120	
70	10	702	2090	1740	
80	10	802	2390	2540	
90	10	902	2690	3590	
100	10	1002	2990	4950	

Table 6.2: Visualisation time, learning path finding and visualisation time depending on the size of diagram. Notation: levels - the number of levels in hierarchy, width - the max number of nodes per level, links - the total number of nodes, nodes - the total number of nodes, visualisation - the time in ms for building the hierarchy and visualising it (both algorithms 6.4 and 6.5) and visualisation - the time in ms for finding and visualising it the learning path and forward learning path prediction (Algorithm 6.6).

Note that, due to high memory and processing time requirements, not all configurations in Table 6.2 have values for learning path construction and visualisation. As shown in Table 6.2, up to 150 nodes and 440 links, both visualisation of diagram and learning path is quite fast. This structure contains 15 levels of hierarchy (17 if the first and last node are included) and 10 nodes per level. As the number of levels, nodes and links increases, the processing requirements increase, especially for the leaning path where after 15 levels (150 nodes, 440 links), the processing time increases to a state where the user experience could be considered to be suffering a lot. On the other hand, the visualisation of the diagram itself, even up 60 levels which contain around 600 nodes and 1800 links, can be still performed with a processing time of slightly above 1 seconds which could be still be considered acceptable. Nevertheless, in practice, it is not expected to have subjects that have more than 100 nodes as then teachers would have a hard time to grasp an overview. Thus, for such a tree the visualisation time would be around 20 ms and whereas the learning path including the forward learning path could be constructed and visualised in about 130 ms, which is entirely acceptable for the user experience.

Chapter 7

Summary and Discussion

The main goal of this thesis was to visualise the learners' gained competencies and their relationship. The visualisation mainly targets teachers with the intention of letting them explore students' gained knowledge, identify the strengths or weakness of a collective group (e.g. entire classroom) and adapt learning strategies accordingly, e.g. by reiterating some parts of curricula where students are having more difficulties.

Initially, this thesis identifies and thoroughly reviews the related work in order to grasp an overview of different related fields and thus be able to tackle the problem from different perspectives. As the topic of this thesis crosses paths with many disciplines, the related work first covers the basics of information visualisation in general (Chapter 2). Then it proceeds with the analytical aspect of data in the learning domain which is known as learning analytics (Chapter 3). Last but not least, it dives into the psychological theory of competency-based learning analytics (Chapter 4) for a better understanding of competence based learning theory.

Given the knowledge gained in related work and the better understanding of the problem, I created a visualisation concept along with the low fidelity mockups. The mockups define how visualisation will look like as well as how the information is presented in the visualisation. The chosen visualisation is a hierarchical network diagram, where the information is presented based on the colours of the nodes, connection of the nodes as well as the location of them in the hierarchy. Each node represents a competency which is associated one value for each learner. Such value represents the probability that the given students acquired the competency. Additionally, nodes are connected to each other by edges which define dependencies or prerequisites between them. For the connected nodes, the ones in the upper level in the hierarchy are depended on the lower ones meaning that the nodes or competencies on, the lower levels are prerequisites for the ones on the upper level. Additionally, the visualisation defines the learning path as a way to visualise the road of how leaner(s) acquire knowledge and propagate through competencies.

This thesis assumes that the students' probability values for competencies are evident and it does not take the task of defining or suggesting how such probabilities are assigned. However, common testing methods could be used as the probability values could be represented by a percentage of the knowledge for the given competency. Thus test scores expressed in percentage could be used for that purpose.

The probabilistic representation of knowledge has many advantages. Thus in this thesis, I defined formal mathematical and probabilistic representations (see Chapter 5) of competency values which allows the switching of underlying data of the visualisation from one learner to a group of students and vice versa. Additionally, the probabilistic representation of node values (probabilities) allows the definitions of probabilities for an entire competencies propagation path containing connected nodes (one per level) in different levels in the hierarchy. Such path probability can be calculated as a joint probability of the nodes contained within it. Finally, this allows formulating the finding of learning path as an optimisation problem which searches through all competencies propagation paths in order to find the one with the highest joint probability which is defined to be the learning path (which then later is implemented in the software prototype).

Given the mockups and the mathematical conceptual definitions for the visualisation, I implemented the visualisation as well a user interface which would allow users (teachers) to use it. The user interface allows the users to select groups of students or individual ones in order to visualise their data.

Overall, the hierarchical network competency-based diagram developed throughout this thesis presents a good opportunity for teachers to review students progress and adapt teaching strategy based on evidence (real data). However, the visualisation has not been evaluated with target users, which is also the main limitation of this thesis. While initially, an evaluation was planned, due to the lack access to real students' data and access to teachers, the evaluation was not feasible. Given that the tool targets teacher, I did not see if meaningful to evaluate with other users groups (outside of teaching profession).

Thus, an evaluation with professional end-users is something that should be addressed in future work. Additionally, in future, having access to real data might make it worth to explore changes in parts of the concept of the diagram as well. For instance, instead of showing one learning path, one could visualise more of them (e.g. top 2-3) which in turn would cluster the propagation of competencies in different subgroups. Perhaps, users progress patterns tend to cluster in some limited path. However, as mentioned, such exploration would be possible only when equipped with real data and then it should be evaluated whether such information would be useful for teachers. Nevertheless, the thesis provides a thorough technical evaluation of the visualisation, which identifies the limitations of the developed prototype.

Additionally, in this thesis, I added the concept of forward learning path which predicts users learning path, even when there is no data for some upper parts in the hierarchy of the visualisation as students are still in the process of graining competencies (for instance in the middle of the semester). The prediction is made by using default values assigned by experts instead of probability values obtain from students' data. One could use more sophisticated approaches of recommending the forward path, especially for individual students. For instance, one could suggest forward paths based on students who already have similar learning path up to the given competency and also have data for the upper part of the graph/visualisation as such students might be more advanced. For instance, for students who attended the course last year, probably there is evidence data for all nodes. Such individual (students in this case) similarly approaches are quite common in recommender system [Ricci et al., 2011, Linden et al., 2003] even in learning environments [Kopeinik et al., 2017a, Kopeinik et al., 2017b]. For instance, collaborative filtering [Ricci et al., 2011, Linden et al., 2003, Sarwar et al., 2001, Kopeinik et al., 2017a] or cosine similarly [Singhal et al., 2001, Tan et al., 2005] are very popular algorithmic choices for recommending based on individual similarities. However, to apply such algorithms and evaluate whether they provide meaningful recommendations, real data and user studies are necessary. But such an approach could be a good opportunity for exploration in the future work.

Bibliography

[Andrews, 2002] Andrews, K. (2002). *Visualising information structures: aspects of information visualisation*. TU Graz.

[Bader-Natal and Lotze, 2011] Bader-Natal, A. and Lotze, T. (2011). Evolving a learning analytics platform. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pages 180–185. ACM.

[Bienkowski et al., 2012] Bienkowski, M., Feng, M., Means, B., et al. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1:1–57.

[Birkhoff, 1948] Birkhoff, G. (1948). Lattice theory. new york. *American Mathematical Society*.

[Blikstein, 2011] Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. In *Proceedings of the 1st international conference on learning analytics and knowledge*, pages 110–116. ACM.

[Bostock, 2017] Bostock, M. (2017). Mobile patent suits. `http://bl.ocks.org/mbostock/1153292`.

[Clow, 2012] Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 134–138. ACM.

[Dimitrova et al., 2007] Dimitrova, V., McCalla, G., and Bull, S. (2007). Open learner models: Future research directions. *International Journal of Artificial Intelligence in Education*.

[Doignon and Falmagne, 1985] Doignon, J.-P. and Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. 23:175–196.

[Doignon and Falmagne, 2012] Doignon, J.-P. and Falmagne, J.-C. (2012). *Knowledge spaces*. Springer Science & Business Media.

[Duval, 2011] Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In *Proceedings of the 1st international conference on learning analytics and knowledge*, pages 9–17. ACM.

[Duval, 2012] Duval, E. (2012). Learning analytics and educational data mining.

[Ebner and Schön, 2013] Ebner, M. and Schön, M. (2013). Why learning analytics in primary education matters. *Bulletin of the Technical Committee on Learning Technology*, 15(2):14–17.

[Falmagne et al., 2013] Falmagne, J.-C., Albert, D., Doble, C., Eppstein, D., and Hu, X. (2013). *Knowledge spaces: Applications in education*. Springer Science & Business Media.

[Falmagne and Doignon, 2010] Falmagne, J.-C. and Doignon, J.-P. (2010). *Learning spaces: Interdisciplinary applied mathematics*. Springer Science & Business Media.

[Ginon et al., 2016] Ginon, B., Johnson, M. D., T"́rker, A., and Kickmeier-Rust, M. D. (2016). An open learner model used by teachers to monitor speed reading learners. In *IWTA@EC-TEL*.

[Greller et al., 2014] Greller, W., Ebner, M., and Schön, M. (2014). Learning analytics: From theory to practice–data support for learning and teaching. In *International Computer Assisted Assessment Conference*, pages 79–87. Springer.

[Heller et al., 2006] Heller, J., Steiner, C., Hockemeyer, C., and Dietrich, A. (2006). Competence-based knowledge structures for personalised learning. *International Journal on ELearning*, 5(1):75.

[Highcharts, 2017a] Highcharts (2017a). Horizontal bar charts. `https://www.highcharts.com/demo/bar-basic`.

[Highcharts, 2017b] Highcharts (2017b). Pie charts. `https://www.highcharts.com/demo/pie-basic`.

[Highcharts, 2017c] Highcharts (2017c). Scatter plot. `https://www.highcharts.com/demo/scatter`.

[Highcharts, 2017d] Highcharts (2017d). Stacked area charts. `https://www.highcharts.com/demo/area-stacked`.

[Highcharts, 2017e] Highcharts (2017e). Vertical bar charts. `https://www.highcharts.com/demo/column-basic`.

[Johnson et al., 2016] Johnson, M., Ginon, B., Masci, D., Baber, C., and Bull, S. (2016). Deliverable d4.3 second release of visualisation and olm services and tools.

[Khalil and Ebner, 2013a] Khalil, H. and Ebner, M. (2013a). *Interaction Possibilities in MOOCs – How Do They Actually Happen?*, pages 1–24. .

[Khalil and Ebner, 2013b] Khalil, H. and Ebner, M. (2013b). "how satisfied are you with your mooc?"-a research study on interaction in huge online courses. In *EdMedia: World Conference on Educational Media and Technology*, pages 830–839. Association for the Advancement of Computing in Education (AACE).

[Khalil and Ebner, 2015] Khalil, M. and Ebner, M. (2015). Learning analytics: principles and constraints. In *EdMedia: World Conference on Educational Media and Technology*, pages 1789–1799. Association for the Advancement of Computing in Education (AACE).

[Kickmeier-Rust, 2016] Kickmeier-Rust, M. (2016). Final report: Tools for competence-centred, multi-source learning analytics.

[Kickmeier-Rust et al., 2016a] Kickmeier-Rust, M., Khurshid, A., Debus, K., and Fessl, A. (2016a). Deliverable d2.6 final system release.

[Kickmeier-Rust et al., 2016b] Kickmeier-Rust, M., Steiner, C., Luzhnica, A., Drnek, M., and Matthew, J. (2016b). Deliverable d3.2 first release of la/edm services and algorithms.

[Kickmeier-Rust and Albert, 2016a] Kickmeier-Rust, M. D. and Albert, D. (2016a). Support teachers' predictions of learning success by structural competence modelling. In *IWTA@ EC-TEL*, pages 35–39.

[Kickmeier-Rust and Albert, 2016b] Kickmeier-Rust, M. D. and Albert, D. (2016b). Theory-driven learning analytics and open learner modelling: The teacher's toolbox of tomorrow? In *UMAP (Extended Proceedings)*.

[Kickmeier-Rust et al., 2006] Kickmeier-Rust, M. D., Albert, D., and Steiner, C. (2006). Lifelong competence development: on the advantages of formal competence-performance modeling.

[Kickmeier-Rust et al., 2015] Kickmeier-Rust, M. D., Steiner, C. M., and Albert, D. (2015). Uncovering learning processes using competence-based knowledge structuring and hasse diagrams. In *VISLA@ LAK*, pages 36–40.

[Kopeinik et al., 2017a] Kopeinik, S., Kowald, D., Hasani-Mavriqi, I., Lex, E., et al. (2017a). Improving collaborative filtering using a cognitive model of human category learning. *The Journal of Web Science*, 2(4):45–61.

[Kopeinik et al., 2017b] Kopeinik, S., Lex, E., Seitlinger, P., Albert, D., and Ley, T. (2017b). Supporting collaborative learning with tag recommendations: A real-world study in an inquiry-based classroom project. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, pages 409–418, New York, NY, USA. ACM.

[Korossy, 1997] Korossy, K. (1997). Extending the theory of knowledge spaces: A competence-performance approach. *Zeitschrift fur Psychologie*, 205(1):53–82.

[Lau and Moere, 2007] Lau, A. and Moere, A. V. (2007). Towards a model of information aesthetics in information visualization. In *Information Visualization, 2007. IV '07. 11th International Conference*, pages 87–92.

[LEA'sBox, 2016] LEA'sBox (2016). Lea's box. `http://css-kmi.tugraz.at/mkrwww/leas-box/project.html`.

[Linden et al., 2003] Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80.

[Mazza, 2009] Mazza, R. (2009). *Introduction to information visualization*. Springer Science & Business Media.

[Nasar, 2016] Nasar, S. (2016). Knowledge space framework: An api for representation, persistence and visualization of knowledge spaces. *arXiv preprint arXiv:1612.02111*.

[Nguyen, 2015] Nguyen, D. (2015). Convergence in the news media discourse (online) on the eu crisis shown in network graphs. [**?**].

[Niles, 2011] Niles, P. F. (2011). Why goal visualization works. `https://www.huffingtonpost.com/frank-niles-phd/visualization-goals_b_878424.html`.

[Novak et al., 2007] Novak, E. J., Silverstein, M. D., and Bozic, K. J. (2007). The cost-effectiveness of computer-assisted navigation in total knee arthroplasty. *JBJS*, 89(11):2389–2397.

[Playfair, 1801] Playfair, W. (1801). *The commercial and political atlas: representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of england during the whole of the eighteenth century.* T. Burton.

[Reimann et al., 2015] Reimann, P., Bull, S., Kickmeier-Rust, M., Vatrapu, R., and Wasson, B. (2015). *Measuring and visualizing learning in the information-rich classroom*. Routledge.

[Reimann et al., 2013] Reimann, P., Kickmeier-Rust, M., and Albert, D. (2013). Problem solving learning environments and assessment: A knowledge space theory approach. *Computers & Education*, 64:183–193.

[Ricci et al., 2011] Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.

[Salem and Rauterberg, 2005] Salem, B. and Rauterberg, M. (2005). Aesthetics as a key dimension for designing ubiquitous entertainment systems. In *The 2nd International Workshop on Ubiquitous Home-ubiquitous society and entertainment*, pages 85–94. NICT Keihanna and Kyoto.

[Sarwar et al., 2001] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA. ACM.

[Schön et al., 2012] Schön, M., Ebner, M., and Kothmeier, G. (2012). It's just about learning the multiplication table. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 73–81. ACM.

[Shneiderman, 1996] Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, pages 336–, Washington, DC, USA. IEEE Computer Society.

[Siemens, 2001] Siemens, G. (2001). Learning and academic analytics. `http://www.learninganalytics.net/uncategorized/learning-and-academic-analytics/`.

[Siemens, 2010] Siemens, G. (2010). What are learning analytics? `http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/`.

[Siemens and d Baker, 2012] Siemens, G. and d Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 252–254. ACM.

[Siemens and Long, 2011] Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics in learning and education. 5:30–32.

[Singhal et al., 2001] Singhal, A. et al. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 24(4):35–43.

[Society, 2018] Society, I. E. D. M. (2018). educationaldatamining.org. `http://educationaldatamining.org/`.

[Solution, 2015] Solution, W. W. I. T. (2015). Denmark trade summary. `https://wits.worldbank.org/CountryProfile/en/Country/DNK/Year/2015/SummaryText`.

[Steele and Iliinsky, 2010] Steele, J. and Iliinsky, N. (2010). *Beautiful visualization: looking at data through the eyes of experts.* " O'Reilly Media, Inc.".

[Swan II et al., 1999] Swan II, J. E., Rhyne, T.-M., Laidlaw, D. H., Munzner, T., and Interrante, V. (1999). Visualization needs more visual design! In *IEEE Visualization*, pages 485–490.

[Tan et al., 2005] Tan, P.-N., Steinbach, M., and Kumar, V. (2005). Introduction to data mining. 1st.

[Wikipedia, 2018a] Wikipedia (2018a). Pie chart. `https://en.wikipedia.org/wiki/Pie_chart`.

[Wikipedia, 2018b] Wikipedia (2018b). Statistical graphics. `https://en.wikipedia.org/wiki/Statistical_graphics`.

[Yuk and Diamond, 2014] Yuk, M. and Diamond, S. (2014). *Data visualization for dummies.* John Wiley & Sons.