

Master thesis in Design and Multimedia

Dissertation final report

# Visualizing Patterns of Flow and Usage in Wireless Networks

Ruslan Kamolov    ruslan@student.dei.uc.pt

Supervisors    Penousal Machado  
Pedro Miguel Cruz

Department of Informatics Engineering  
Faculty of Sciences and Technology of the University of Coimbra  
10 September 2013



FCTUC FACULDADE DE CIÊNCIAS  
E TECNOLOGIA  
UNIVERSIDADE DE COIMBRA



## Abstract

Flows are records of data related to movement of subjects or objects, which have a direction, an origin and a destination point. This dissertation focuses on the development of scientific visualization methods for representing flows and usage patterns on the wireless networks of four Portuguese universities.

The research has a theoretical and a practical part. The theoretical part analyses and synthesizes the theory of information visualization, referencing distinct types of graphic layouts known from the early days to the present. The graphic semiology and the interaction principles for visualization are explained and illustrated with examples based on different topologies and approaches to visualization of flows. The practical part presents the developed techniques for visualizing flows within wireless networks, based on examples studied during this research. Four interactive visualizations displaying the patterns of flow and usage in wireless networks are constructed. The available data is presented with four different approaches, each one showing particular features of networks, such as connectivity, transitionality and flow, and focusing on specific aspects of the data.

## Keywords

Information visualization, Scientific visualization, Visualization of flows, Visualization of wireless networks, User connectivity, User transitionality, Movement representation, Patterns of flow, Patterns of usage



*Special gratitude*  
*to Cláudia for the constant support,*  
*and to my supervisors for assistance and wise advices.*

*Dedicated to my family.*



# Contents

1. Introduction	11
1.1. Scope	11
1.2. Research Statement	12
1.3. Objectives	13
1.4. Context	14
1.5. Resources	14
1.6. Contributions	15
1.7. Summary	15
2. State of the Art	19
2.1. What Information Visualization is about	20
2.2. Visual Representations: Layouts and Applications	22
2.2.1. Cartography	22
2.2.2. Cosmography	28
2.2.3. Illustrations and Icons	29
2.2.4. Statistical Charts	31
2.2.5. Trees	33
2.2.6. Cartograms, Matrices and Small Multiples	34
2.3. Techniques for Visual Communication	38
2.3.1. Semiology of Visualizations	38
2.3.2. Principles of Graphic Excellence	41
2.4. Interactivity in Graphs	42
2.4.1. Levels of Interaction	43
2.4.2. Interactive Techniques	43
2.4.3. Data Exploration and Interactivity	44
2.5. Summary	45
3. Related Work	47
3.1. Scientific Visualization	47
3.2. Techniques for Representation of Flows	50
3.3. Practical Approaches to Flows	56
3.4. Summary	64

4. Research Goals and Methodology	67
4.1. Objectives	67
4.2. Modus Operandi	69
4.3. Summary	70
5. Preparatory Work	73
5.1. Interpretation of the Data	73
5.2. Experiments with the Graphic Layout	81
5.3. Summary	87
6. Approaches for Visualization of Flows and Usage in Wireless Networks	89
6.1. The Circuit of Transitions	90
6.2. The Matrix of Connections	96
6.3. The Panels of Flow Shapes	101
6.4. The Map of the Wireless Network	107
6.5. Summary	113
7. Experimental Results	115
7.1. The Transitionality in Networks	115
7.2. The Connectivity of Users	131
7.3. The Patterns of Network Usage	139
7.4. The Relative Geographic Flow of Users	149
7.5. Summary	156
8. Conclusions and Future Work	159
References	163

## Appendixes

- A. Visualizations with Gephi
- B. The Transitionality in Networks
- C. The Connectivity of Users
- D. The Patterns of Network Usage
- E. The Relative Geographic Flow of Users
- F. Siggraph abstract and presentation







## CHAPTER 1.

# Introduction

Information visualization is present in Science, in Art and in the social domain. The conceptions made in this field are concerned with the presentation of information. Treating information graphically can offer a better understanding of the data, generally more effective than table sheets. This advantage is achieved because the human's visual perception system senses and recognizes the hierarchically organized data and the related visual patterns more easily. The visualization of information coming from the scientific domain can provide an evident perception of the data. The crucial objective of scientific visualizations is to apply an appropriate graphic model to the required dataset.

The visualization of patterns of flow and usage in wireless networks is the object of this study. The main goals of the research are to make an overview of the state of the Art of the fields of information visualization and of visualization of flows, to create graphic models that show various aspects of the wireless network data, and to analyze the obtained results in order to understand the efficiency of the created visual techniques. The graphic expression in the construction of artifacts is also explored. A graphic language easy to read, interpret and comprehend, a reduced visual load and the clarity of communication of the complex information are some of the topics of this exploration. Another target of this work is to explore the representational potential of flows in wireless networks in a scientific way based on the principles of graphic design.

### 1.1. Scope

Scientific visualization is the study focused on displaying of scientific data. The main purpose of this field is to gain understanding and insight into the data and the information being studied. Its strategy involves ideas that can be usable as tools

for real applications. This goal is achieved by combining data parsing, system design, computer graphics and interactivity (BRODIE *ET AL*, 1992). This research uses the same techniques across many areas of study. The type of data associated to flows is typically collected by electronic devices, that inspire the ways of graphic representation. Such representation is similar to the existing scientific visualizations of graphs and networks, such as knowledge, social, transportation, semantic and business networks, and the *World Wide Web* [CHAPTER 3].

The visualization of flow in networks fits into a concept that depends on three aspects: data visualization, user interaction and visual aesthetics. Such framework, called *information aesthetics* (LAU AND MOERE, 2007), consists of the fusion of these techniques that form different branches of data visualization. Thus, it is essential to take into account these aspects for the conception of visualizations in this research: the first aspect requires the accuracy in providing quick and easy reading of information. The second aspect ensures that the user has an overview of the data and the control over the individual parts of the visualization, such as exploring and filtering the information for different detail levels. The third aspect involves the appellative and simplified appearance that, at the same time, should create instinctive associations with graphical elements that represent the data. In this research, combinations of these methods are used to get a variety of experimental results and to determine which of them work better, in what proportion and why.

In this specific case, the visualization of flows belongs to scientific visualization. This field deals with concrete data for specific purposes, implies lesser exploration of representational techniques and prefers the use of existing technologies. Graphical representation of scientific data may help to understand the complex nature of the phenomenon being studied. Combination of the concrete approach with the principles of information aesthetics can provide an easier understanding and interpretation of the data. There is an important difference between information visualization and scientific visualization. The first concerns with the simplicity of communication of information. The second term seeks to turn data understandable. Yet, both methods are often intertwined.

## 1.2. Research Statement

This dissertation presents new applications of techniques and models for visualization of patterns of flow and usage in wireless networks. This area of research is important because the analysis of the movement inside a system helps to understand and maximize the whole system's performance. Also, visualizations of large systems are impossible to be performed without proper tools. Visualizing the overall movement of an entire isolated system provides a global vision of its inner streams and routes of information. That vision can be used to understand the system's dynamics and the structural aspects of particular elements and general

aspects, to reveal movement patterns and to optimize the system. The nature of this research is experimental. It involves an analysis of the state of the Art of information visualization that tapers into the area of scientific visualization. The interaction principles applied to graphs are also studied. Afterwards, the techniques and respective examples related to representations of flows are analyzed to comprehend different processes and results. At the last phase, the graphic system for presentation of the provided technical data is developed. That system demonstrates the possibility of representation and understanding of flows within wireless networks. The design of visualizations is based on the simplification of information by the use of graphic and textual languages. Using the computational programmatic approach, the models and applications for presentation and communication of information are created. These deliverables are related to the movement of people and/or objects from various origins to multiple destinations. The goal of this process is to produce analytical tools for studying and visualizing movement patterns, along with static and dynamic artifacts, to facilitate the analysis of large volumes of data. The contribution of this research is the review of the state of the Art of the fields of information visualization and scientific visualization, and the developed visual dynamic techniques for visualization of flows in networks.

### 1.3. Objectives

The main goal of this research is to elaborate methods and techniques for visualization of flows in wireless networks. This objective implies the understanding of the data and the definition of aspects to be represented graphically. We provide ways of visualization of wireless network structures, of movement patterns and of directed streams of information traffic inside these networks. The visual techniques are inspired by the examples described in the state of the Art [CHAPTERS 2 AND 3]. The applications are developed from the ground up accordingly to the specific aspects of the data. The produced graphic results are analyzed to prove their efficiency and also to exemplify the ways of reading of these.

An analysis focused on graphic semiology and work related to scientific visualization distinguishes pertinent representation strategies. An efficient visual synthesis of large amounts of data is enabled to facilitate understanding of information. Creation of a scientific visualization implies data treatment and programmatic implementation, resulting from the extraction of relevant information, the accurate coding and the right graphic decisions.

This research is not about making a flow visualization in fluid dynamics. The objective of this dissertation is to create techniques for visualizing the flow of subjects and objects as they move in time and space, proving their legitimacy and effectiveness with functional applications.

## 1.4. Context

The dissertation *Visualizing Patterns of Flow and Usage in Wireless Networks* is developed in the context of the Master in Design and Multimedia, at Department of Informatics Engineering of the Faculty of Sciences and Technology of University of Coimbra (FCTUC).

The research also fits in the context of a scientific research grant under the project *SUM – Sensing and Understanding human Motion dynamics* of the Portuguese Foundation for Science and Technology. That project pursues two main goals: to explore the *Wi-Fi* technology to capture data about the human spatial behavior, and to use of the gathered data to detect motion patterns, finding its correlation with the spatial characteristics. The project is held at the Computational Design and Visualization lab of the Center of Informatics and Systems of the University of Coimbra (CISUC).

The supervision is performed by Penousal Machado, professor in the Department of Informatics Engineering of the FCTUC and researcher of Cognitive and Media Systems at CISUC, and Pedro Cruz, PhD student in Information Sciences and Technology of University of Coimbra, professor and a researcher at CISUC.

## 1.5. Resources

The following generic resources are necessary to proceed with the representation of flows in wireless networks.

*Existing bibliography* about interactive media, data visualization and scientific representations is the primal source of knowledge for this research. That literature aids to perceive the vastness of works and studies made in these areas and to be aware of what fits better in the scope of scientific visualization of flows.

*Datasets* available for analysis and graphic representation are the one of the key requirements for this research. The quality, content, reliability and relevance of data decide the progress flow, the artifacts to produce and the results to achieve. Data may accrue from hard and soft sciences or fields of social life, since these areas involve flows of subjects or objects. For this research, the data comes from the four universities of Portugal: the University of Coimbra, the University of Lisbon, the University of Minho and the University of Porto. The information describes the user connections between the access points of the universities' wireless network called education roaming, or *eduroam* [SECTION 5.1].

*Computer* is used as a medium of technological concretion. The capabilities of data and image processing of computational devices offer such opportunities as the performing of complex calculations, the rendering of composite high-resolution graphics and the creation of evolved systems in real time with great accuracy.

*Software tools* provide programmatic environments with methods and functions that make the development effective and efficient. In this case, the Processing programming language (FRY AND REAS, 2004) is utilized. It is a simple and powerful free-ware tool widely known among multimedia designers and programmers. Withal, other programming languages and auxiliary software may be required for obtaining specific results.

## 1.6. Contributions

This dissertation gives origin to a number of contributions as follows.

*The bibliographic review* of the general history and techniques of the field of information visualization is presented in CHAPTER 2. The analysis of the area of scientific visualization is explained in CHAPTER 3.

A system of *techniques and methods for visualization of flows* within wireless networks is produced. It allows the application of the defined graphic rules and algorithms to various cases of the similar data that represent movement streams. The graphic models are described in CHAPTER 6.

*Interactive software* for the visualization of flows and usage patterns is implemented to prove the practical application of proposed methods. As a result, the extraction of conclusions on the obtained visual output and its posterior analysis are possible. The implementation details are explained in CHAPTER 6.

*Produced artifacts and supporting materials*, both image and video based. The obtained results are described in CHAPTER 7 and can be seen in larger resolution in the APPENDIXES section [PP. II-51].

*A scientific abstract* reporting and depicting the work progress and its effects was submitted and accepted at *ACM SIGGRAPH*, an international conference on computer graphics and information visualization. The paper can be consulted in the APPENDIXES section [P. 54].

*The dissertation final report* that documents and describes the entire process of this research, as well as the analysis of the results and the conclusions drawn.

## 1.7. Summary

A review of information visualization accompanied with various historical pieces is presented in CHAPTER 2. The different types of existing visualizations and layouts are described to reveal their applicability in this research. The main concepts and methods of construction and analysis of visualizations are described, relying on the work of Jacques Bertin and Edward Tufte. The use of interactivity

in graphs is described to provide a series of graph exploring and data manipulating functionalities that became possible with the technological progress. Many of these techniques are used in the visualizations conceived in the context of this study.

CHAPTER 3 emphasizes the related work in the field of scientific visualization. The general conception of the area is explained, along with significant methods for representation of flows. The examples of complex data systems, such as networks, graphs and flows, are shown and described. The first two chapters illustrate and synthesize the theoretical research related to this dissertation.

CHAPTER 4 demonstrates in detail the objectives of the study. The problem solving approach, the research goals and the main reasons of this research are discussed. The applied methodology and the ways that objectives entwine with work methods are clarified, tracing their relationships to the theoretical approaches explained in the State of the Art and the Related Work chapters.

CHAPTER 5 describes the preparatory work phase, made before starting the construction of interactive applications. The data is described and the respective process of treatment is explained. Then, the experimental sketches of possible graphic layouts are implemented. Each model is analyzed in order to understand its use and value in the visual representation of networks and patterns of flow.

CHAPTER 6 describes the work produced on the visualization of patterns of usage and flow in wireless networks. The practical approaches are based on user connections and transitions within the networks of the biggest portuguese universities. The implementation of each prototype follows the methodology specified in CHAPTER 4.

CHAPTER 7 presents the experimental results obtained after the construction of visualizations, described in CHAPTER 6. All generated artifacts are analyzed and the conclusions about their efficiency are extracted.

CHAPTER 8 provides an overview of all work made during the research, resumes the objectives, process and contributions, and specifies the future work.







## CHAPTER 2.

# State of the Art

In this chapter, the meaning and the main purpose of information visualization are clarified – to make data clear and universally understandable, revealing submerged information to any kind of public in a self-explanatory way. Information visualization is a relatively recent field of study and research. Its existence has begun since humans felt the necessity to represent surrounding things. Pictures and images have a long and rich history that begins with primitive cave depictions and war paintings, and continues until the current era of multimedia and technology. The graphic nature of visual representations has been developing with the advent of new needs and trends. The actual emergence of the study of information visualization is justified by the accessibility and the attractiveness of images, being affected by the invention of novel approaches, as the variety of examples demonstrates below. These examples are grouped by the graphic layout similarity while describing the applicability of each category to the visualization of flows and networks.

The construction of a visualization usually relies upon the designer's decisions. Yet, there are basic notions of the theory of the graphic construction that help to make those decisions. Visual elements and layouts have different properties and meanings. Represented in a proper way, these convey the information of the graph's message for a clear perception and understanding. The theoretic principles for information design of two remarkable pioneers in data visualization, Jacques Bertin and Edward Tufte, provide the guidelines to the development and the analysis of graphics in the **TECHNIQUES FOR VISUAL COMMUNICATION** section.

Interactivity is a bidirectional flow of information between a software application and a user. The interactive principles and techniques evolved along with the advance of computer technology. Nowadays, user-directed applications provide an effective use of the graphic interface with intuitive functionalities to provide an easier understanding of the software's purpose and to offer possibilities of manipulation,

management and storage of information. These aspects are also applied to visualizations. The interactive graphs allow the exploration and the handling of the data, and the comparison of different samples and results through the interaction with the user. The most common techniques are categorized and reported in the INTERACTIVITY IN GRAPHS section.

## 2.1. What Information Visualization is about

Information visualization is a field that involves graphic representation of data in order to extract and amplify associated knowledge. Data can be abstract or concrete. Abstract data includes numerical and nonnumerical values, such as text, financial data, business information, geographic data – all these types of information do not manifest an evident spatial character. Concrete, or scientific, data deals with the values containing physical spatial information (CARD *ET AL*, 1999, PP. 6-7). These two concepts trace the difference between information graphics and scientific visualizations.

One of the first publications focusing on the representation of information is “The Commercial and Political Atlas,” written by William Playfair. The main objective of the book is to organize and display statistical data in graphs and charts to facilitate the people’s reading of the abundant data and make decisions based on the achieved cognition. Playfair understood the power of visualization of the data, as well as the rules of its proper presentation:

*“Conveying information is accurate in principle, though in execution it may be liable sometimes to error.”* (PLAYFAIR, 1801)

However, Playfair did not provide a theory about how and why his graphic symbols and forms work. Jacques Bertin in his “Sémiologie Graphique” made the monumental study of the use of graphics for analytical and presentational purposes [SUBSECTION 2.3.1]. The author described how quantities can be represented on paper for statistical purposes. In his theory, any graphic representation can be polysemic, i.e., the meaning of individual signs is deduced with consideration to a collection of signs (BERTIN, 2011, P. 2), and the meaning becomes subjective. This concept explains the polychotomy of reading and perceiving graphical information. The importance reveals in the process of conception of the graphic artifact when analyzing the impact it should provoke on the audience. Also, Bertin accentuates that spatial systems (including graphics) communicate various relationships among the visual variables. This statement stands for the main purpose of graphics – the maximum utilization of spatial perception.

Card, Mackinlay and Shneiderman define the term information visualization in the book “Readings in Information Visualization – Using Vision to Think.” The authors investigated the two-dimensional presentation forms of relational information, such as bar charts, node-link diagrams, scatter plots, and an evaluation system of these

graphic artifacts. Later, they advanced to three-dimensional computer graphics and animation. In this scope, comes the definition:

*“Information visualization: the use of computer-supported, interactive, visual representations of abstract data to amplify cognition.”* (CARD ET AL, 1999, P. 7)

With this expression, the authors accentuate the fact that visualization is not about the images, but gaining knowledge about what these images show. Displayed information is an object of study, and a meticulous study provides insights of the information. The current age of digital technology allows a faster handling and showing of the data (specially the complex data collected by electronic devices), so the authors give importance to the interactive computer-supported representation. Computer graphics stimulated the modern study of visualization. Computational approaches have been always used to resolve scientific problems. The machines that firstly allowed graphic output, initiated the era of digital representation of information. Nevertheless, the lack of graphic capabilities impaired the usefulness of the visual artifacts. With the technological advance in computer hardware and software, information visualization also took shape and impact. Its importance became recognizable by worldwide scientific society. The simplicity of introducing and processing large and intricate amounts of data with the computer substituted mountains of archives and paper. Since then, the media and specialized conferences started to spread the word about the new tendency of digital information visualization.

*“Information visualization uses graphical primitives ... to reveal patterns and structures in the data objects that these primitives represent.”*

(MANOVICH, 2011)

The field of information visualization embraces a great range of applications nowadays. The most popular themes are scientific study, data mining, information graphics, financial analysis, market studies, manufacturing production control and crime mapping. Normally, the progress dictates rules and trends. From the time when information visualization became a fashion, a new language of modern world, its expansion was incredible. Today, it is everywhere. From television and movies to music clips and presentations, from newspapers and magazines to advertising and product packages, from social networks to games – all kinds of visual media have adopted a fancy stylish look and techniques of the new way to show things.

*“The instrumental function of visualization has generated much attention; many disciplines have started to recognize the efficacy of visualization as an aid for research and these have begun to require the development of visual tools tailored for their disciplinary fields.”* (SCAGNETTI, 2012)

With this recent manifestation of interest in visualization as a tool to explore and comprehend complex phenomena, urges the necessity to define the taxonomy, the methodology and the limits of this developing area. On the other hand, information visualization had to adapt to target audiences. Its main consumers are common

people. With that enlargement, the graphic style became perceptible due to the novel figurative and abstract visual languages, while communicating information to the ordinary public. The public learn to interpret these languages and the evolution is obvious. If scientific visualizations are made by experts for understanding by experts, the mainstream visualizations seek to be friendly to any individual, regardless of age group, occupation or interests.

Information visualization has been always evolving in graphic form, thematic aspect and design approaches. The recent graphics present the simplified design and new communication languages. Abstract and geometric visual elements, vivid yet sober color themes, carefully organized layouts and good typographic decisions are conquering the media world. The presentation of information inherited techniques present in the modern graphic design. This visual evolution is partly due to the popularization of design courses and a growing amount of graduated employed designers. The audience becomes more aware of popular design trends and products. These tendencies supply and feed the demand for well made graphics.

## 2.2. Visual Representations: Layouts and Applications

The history of graphic representations had its roots from the time when the human being could firstly hold a painting material in hand to depict something on a drawable surface. Since then, the graphics evolved with the human's natural, cultural and technological progress. There is a great variety of graphical forms displaying information. Obviously, it is impossible to include all kinds of representations in this document. The examples shown in this section are chosen with the categorical criterion to sample different graphic approaches to information visualization. Maps, technical illustrations, iconic figures, statistical charts and tree diagrams are well-known visual techniques. Some appeared thousands of years ago, another originated more recently. Yet, the majority of these representational methods are used to exhibit information of all aspects of the present world.

This section presents multiple examples in the field of information visualization. The objective is to demonstrate the diversity of works made in this field and explain which visualizations are suitable for the representation of *movement in space and time*. This visual contextualization helps to determine which types of layout can be possibly used for the representation of flows in wireless networks.

### 2.2.1. Cartography

The evolution of information visualization begins from the prehistoric age. Early humans conceived the first representations of information: cave paintings and maps. Later, the use of icons was applied to the registry of cattle and stock. American Indians resorted to imagery to depict the journeys of ancestors. Illegible while alone, icons served as a support to memory and storytelling. Ancient Egyptians used hieroglyphic symbols to describe stories of life, work and religion.



Figure 2.1. Wall painting from Çatalhöyük, probably representing the Anatolian settlement and a landscape with a volcano, Turkey, 7500–6200 B.C.



Figure 2.2. A Stone Age map showing a line of dwellings beside a river, engraved on a mammoth tusk, Mezhirich, Ukraine, 12000–11000 B.C.

*Cartography* is the practice of map-making, depicting the real world on different supports: paper, display, wood, iron and even stone. The cartographic artifacts are ones of the eldest creations that can be called information visualizations. The traits represented on maps may be physical, such as terrain and roads, or abstract, such as words and conceptions. The paintings displayed in FIGURE 2.1 (NORMAN, 2012A) and FIGURE 2.2 (NORMAN, 2012B) are probably the oldest maps yet known. There are visible attempts to create an order and a positioning of graphic elements. Mainly, cave paintings depict the everyday life's aspects: animals, habitations and communal actions. Historicists do not know for what these drawings exactly served. Some theories hold that ancient people could use them to communicate with each others, another assign them ceremonial or religious meanings. The main purpose of maps remains unchanged – to depict spatial relationships between objects and areas.

The art of cartography had an important role in the development of the geography. Maps representing geographic locations were helpful in navigation and exploration. That evolution has allowed the representation of places on the two-dimensional plane in accordance with their real localizations. In the High Middle Ages, many cartographers, such as Fra Mauro, Alberto Cantino, Jerônimo Marini and Gerardus Mercator, were working on the maps' improvement and perfecting



Figure 2.3. Steel engraving "Map of Western Hemisphere," by unknown artist, 1626.

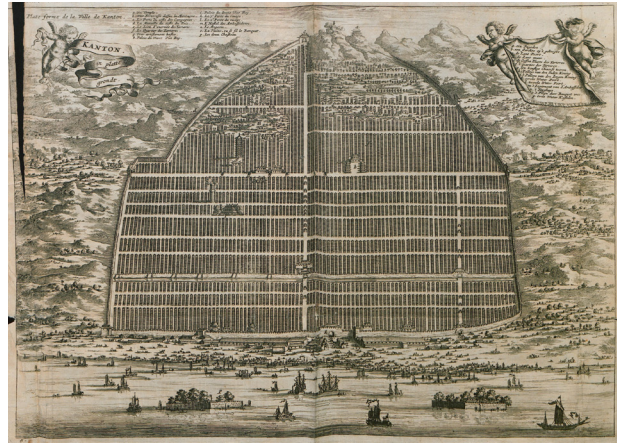


Figure 2.4. "Kanton in Platte Grondt," by Historic Urban Plans, 1665.

techniques of the world's representation. Most of them were from the countries with a strong navy, what permitted marine discovery expeditions. The map of the New World, engraved in 1626 [FIGURE 2.3], demonstrates the discovered parts of North and South Americas. The poorly explored part of the north is evident, as well as the state of California is shown as an island (OAKLAND MUSEUM OF CALIFORNIA).

Canton, one of the first Chinese cities visited by European travelers, is represented in decorative view on FIGURE 2.4. The illustration comes from "Nieuhoff's L'Ambassade de la Compagnie Orientale des Provinces Unies vers l'Empereur de la Chine" (HISTORIC URBAN PLANS, 1965). The house units are distributed on the geometrical grid well-marked by streets shown as horizontal and vertical lines.

The "Map of a Square and Stationary Earth" [FIGURE 2.5] tries to disprove the spherical Earth theory with the biblical map of the world. This version is a four-cornered three-dimensional concept. The continents and oceans are disposed in a circle within a square framework. The northern hemisphere is a heap in the middle of the southern hemisphere's bowl. Presumably, this structure resolves the problem of the oceans' falling off the edges of the flat earth (THE HISTORY BLOG, 2011).

FIGURE 2.6 presents a different kind of the cartographic application – the map of the London's subway system *Underground*. This version is of the authorship of Harry Beck, an engineering draftsman from the same company. His design proposed a more readable and understandable map of subway with horizontal, vertical and 45° diagonal lines (BRITTON). The geographic arrangement is very close to the real routes. The colorful system was applied to aid the passengers orient themselves in lines and stations. The geographical accuracy was not important, so the stations were more or less evenly spaced and identified with text. The Beck's radical proposal was firstly introduced to the public in a small flyer. It gained immediate popularity, and the *Underground* started using topological maps to illustrate the subway network.

These two examples use the principle of mapping the concrete physical information. Yet, each map had it applied in a different way. The map of the square Earth uses the 3D projection to depict the perspective and space, proving the legitimacy of the proposed model. The subway map uses the same principle with a considerable distortion





Figure 2.5. “Map of the Square and Stationary Earth,” by Orlando Ferguson, 1893.



Figure 2.6. The London Underground subway map, by Harry Beck, 1933.

of the reality. It was made wittingly to make the perceptible spatial arrangement of elements – a great solution, because the subway passengers do not have a clear idea of the exact geographic locations between the stations, as well as of the length of the lines that link them together.

FIGURE 2.7 illustrates a cartographic visualization drawn by John Snow. He made it in the context of problem analysis, a simple representation that came up from a complex dataset (McLEOD, 2000). While investigating the cholera outbreak in London in 1853-1854, he plotted the quantities of casualties (black bars) died in first 10 days. With his graph, Snow discovered the connection between the disease and the city water supply. Some days later, the London government confirmed that the indicated water pumps (black dots) were infected.

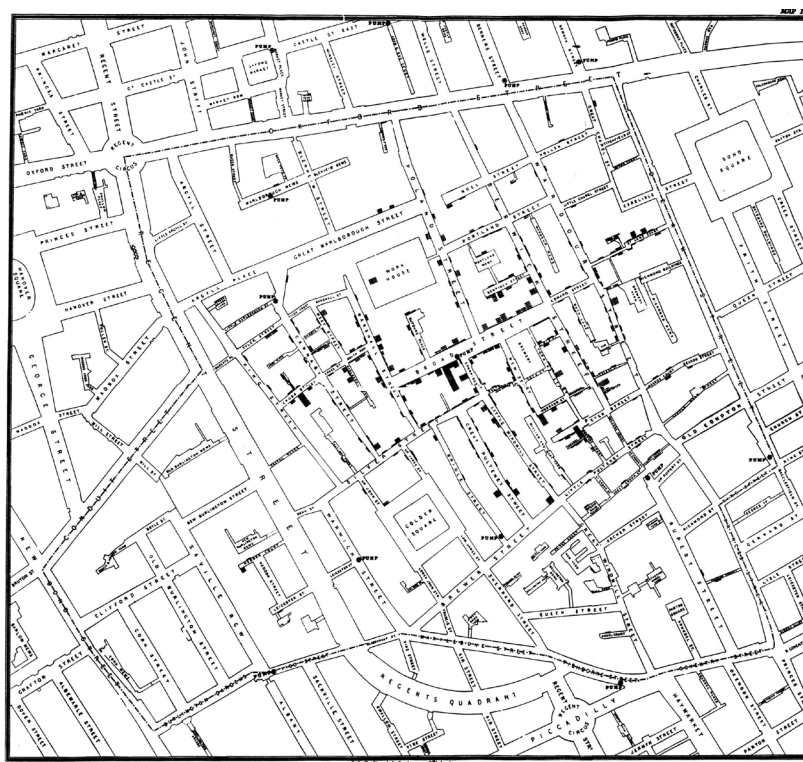


Figure 2.7. The map representing the clusters of cholera cases in the London epidemic, by John Snow, 1854.

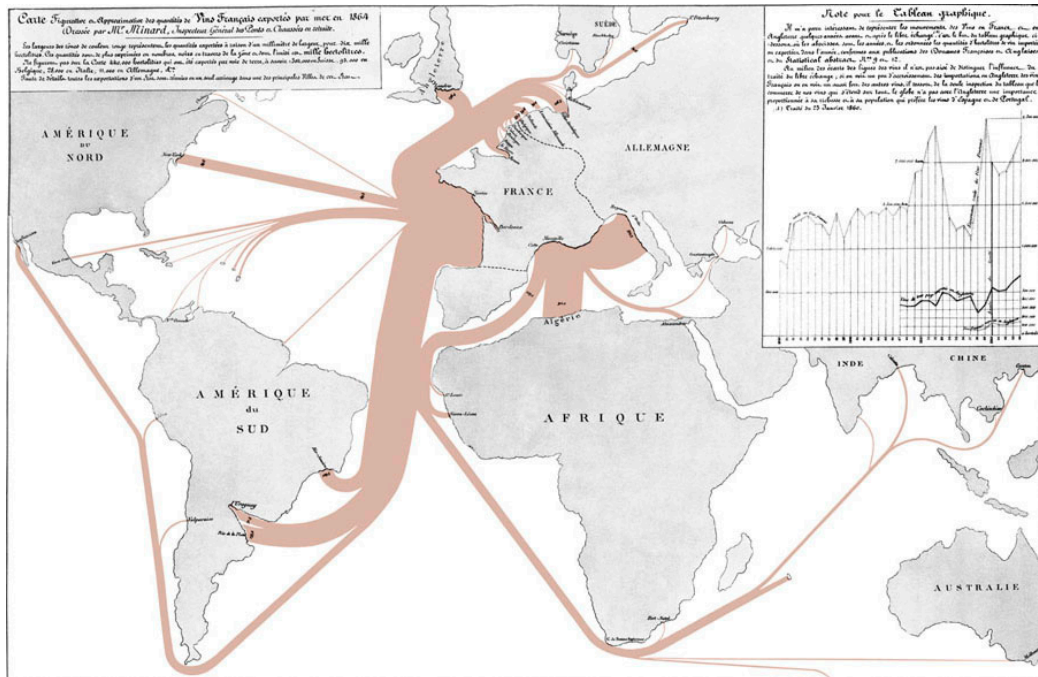


Figure 2.8. “Carte Figurative et Approximative des Quantités de Vin Français exportés par Mer en 1864,” by Charles Joseph Minard, 1864-1869.

The maritime exports of the French wine in 1864 are represented on FIGURE 2.8. The graph, crafted by Charles Minard, shows the world map with minimalistic grey-painted continents and the brown traces indicating the export routes. Each line starts from France and ends in the country of destination. Minard assigned the amount of the sold wine to the thickness of the respective route. There can be seen four major streams that lead to North Europe, South American countries, Algeria and Italy. A line depict both direction of the flow and the quantity of exported goods (TUFTTE, 2001, P. 25).

These two last maps adopt the cartographic approach. The multivariate data is shown along with a real world representation. The cholera map precisely marks the specific case data on the layout and shows the solution for the problem, eliminating the deadly threat. The map of French wine exports uses the world scale locations and lines to trace origins and destinations. The flows that point to the nearby destinations join together to demonstrate the bigger amount of the traffic of goods and highlight the main routes.

Technological advance has continually changed the way of making maps. Yet, the representation preferences were always up to their creators. Visual metaphors associated to geography became a standard for many graphs, defining displaying methods for general types of information beyond cartography. The map metaphor became popular and was used in many other contexts and representations, such as comic maps of political relations between countries, mind maps of ideas’ brainstorming, heat maps in scientific visualization and concept maps of project sketches.

The tradition of making maps with anthropomorphic and zoomorphic figures began in the medieval times. The creators of *satirical maps* presented their view of the



Figure 2.9. “John Bull and his Friends – a Serio-Comic Map of Europe,” by Fred W. Rose, 1900.

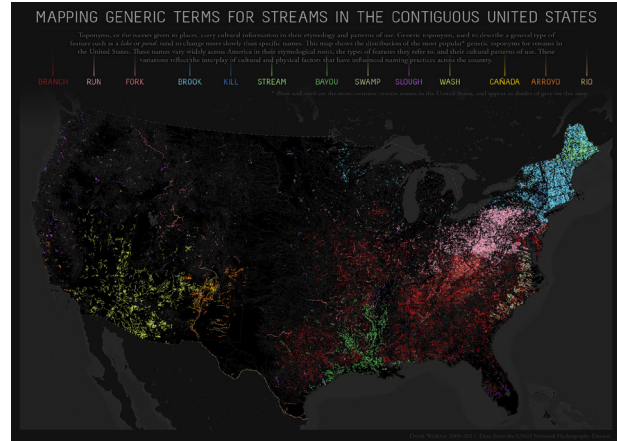


Figure 2.10. “Mapping Generic Terms for Streams in the Contiguous United States,” by Derek Watkins, 2011.

contemporary world in a metaphoric, allegorical and symbolic forms. Meanings and interpretations can be reached using imagination and anthropomorphism. Political cartoon maps appeared in the 19th century. It reflected the significant political and cultural changes. These maps were cleverly done and had brilliant commentaries on events and facts as an act of propaganda. The European nations were represented in a novel serio-comic way (BARRON, 2007). The political caricature map made by Frederick Rose is displayed on FIGURE 2.9. It is also known as the *Octopus Map* because of the Russian Empire depicted as a big octopus extending its tentacles towards Europe. Poland, China and Persia are under the Russian domain; France and Spain are attractive women; Italy, Germany and England are shown as military commanders. Hostile and allied nations are portrayed with strong stereotypes and preconceptions (ROSE, 1900).

The *toponyms map* [FIGURE 2.10] displays the variation of generic names for *streams* across the United States. Toponyms are names given to geographic places, reflecting cultural, physical and linguistic information that influences the naming practice and cartography. Generic toponyms tend to change slower than specific names (WATKINS, 2011). In this case, *creek* and *river* designations are gray colored due to their ubiquity, bright colors symbolize another toponyms. For example, *bayous* came from the old French settlements, *kill* derived from Dutch, *rios*, *arroyos* and *cañadas* are Mexican names. Watkins uses the USGS National Hydrography Dataset to map streams' terms to the layout with 48 states.

The previous visualizations evidence a small part of the great variety of the existing maps and their applications. Maps are a standard layout common to any spectator. Even when represented with some kind of distortion or stylistic modification, the maps remain recognizable and universal. This great advantage results from a simple fact: maps represent the natural habitat of the human beings, the places we all live in, the environment that we all belong to. With this, the cartographic layouts do not require an in-depth study to understand how the data is mapped, because the answer is simple and obvious: geographic locations. This type of layouts is very useful in the representation of movement from one point to another, the viewers will

easily associate the paths and routes to their real locations. From all maps shown above, the Minard's map of wine exports is the most appropriate to display the movement and the flows. He shows the multivariate data in a simple and accurate way. The visualization of flows in wireless networks can adopt the same paradigm, since the exact locations of the origins and destinations are available.

### 2.2.2. Cosmography

*Cosmography* is another scientific field that uses visualizations to represent the collected data. Like geography, its roots come from the ancient history. Human beings were always fascinated with the dark space surrounding them, but only with the invention of the telescope it become possible to make profound studies of the skies. Before that, the astronomers could examine and describe the celestial bodies and their phenomenons visible to naked eye. Two archaic cosmographics, shown on FIGURES 2.11 AND 2.12, exemplify different visual approaches from two distant epochs. The first is a geometric diagram of lunar phases. It displays the Moon's visible sides in different angles of view from the Earth that is represented by an unfilled circle on the right. The Sun is the red-filled circle on the left side (SALIM-ATCHEKZAI, 1974). The second illustration is a scheme of the Ptolemaic geocentric model of the Universe. Each body comes with the designation, the sign, the distance to the Earth and the period of temporal rotation around our planet (VELHO, 1568).

The graph on FIGURE 2.13, created by Sean McNaughton and the *National Geographic* staff, shows the space exploration in the last 50 years (NATIONAL GEOGRAPHIC, 2008). The central part presents the celestial bodies of the Solar System, positioned relatively by neighborhood to our planet. The curves indicate the performed various space missions, stacking them together around the mission's destination. The line colors belong to the past missions of different countries (yellow is for the United States, red is for the Soviet Union) and to the various currently existing spacecrafts (dark green – *Voyager*, lime – *Pioneer*, light green – *New Horizons*, gray – *Cassini* and blue – *Galileo*). The successful missions have the brighter color, compared to the failed ones. At the bottom, the scale demonstrates the distances of planets and four farther located spacecrafts from Earth.

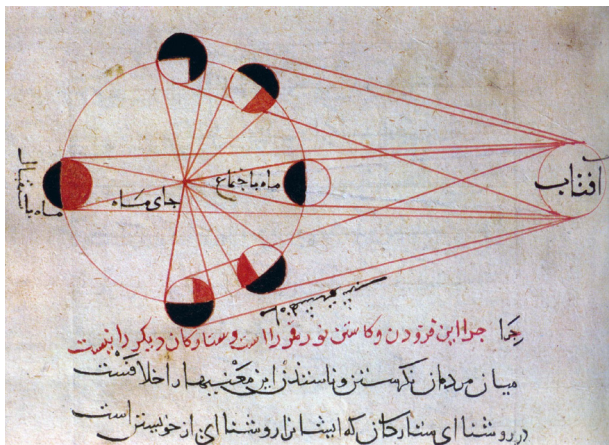


Figure 2.11. Diagram showing the phases of the Moon, by Abu Rayhan Al-Biruni, 1000.

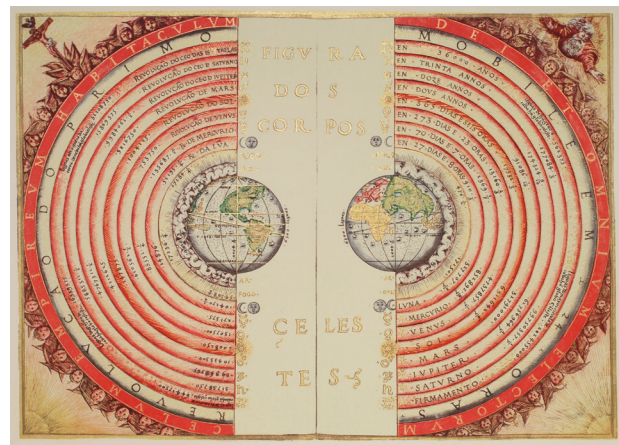


Figure 2.12. Illustration "Figure of the Heavenly Bodies," by Bartolomeu Velho, 1568.

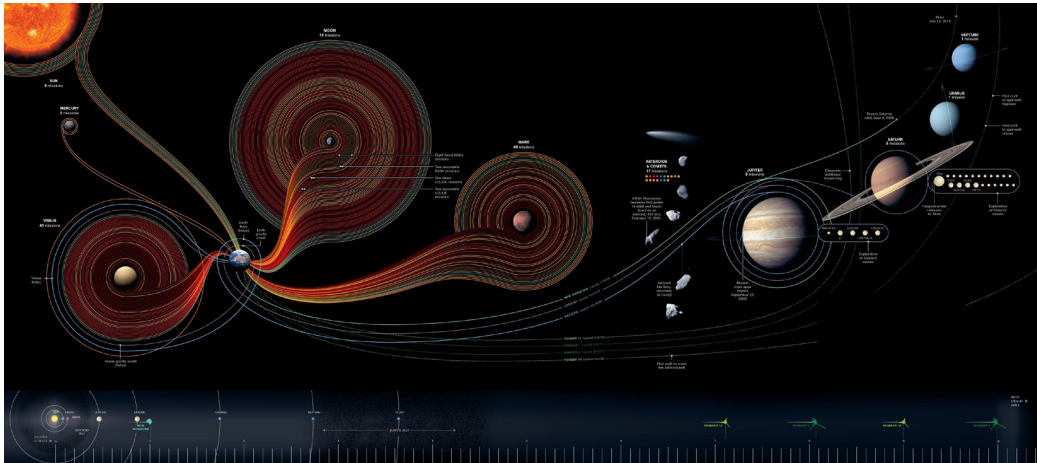


Figure 2.13. "Fifty Years of Exploration," by Sean McNaughton and National Geographic, 2008.

The cosmographic visualizations are similar to maps in cartography. The difference is in the represented area: the cartography shows the Earth's terrain, the cosmography displays the outer space. Moreover, both types of maps depict the exact positions of objects that belong to the respective field of study. The cosmographic paradigm can also be used in the representation of flows in wireless networks.

### 2.2.3. Illustrations and Icons

In the course of time, the graphic languages were enriching the forms of communication. Technical illustrations, simplified icons, geometric grids and brief descriptive captions became common graphic elements. The engraving was a widely used technique of representation on wood and metal from 15th to 18th centuries. The invention of printing press in the middle of 15th century has actuated the transition of graphics to the paper and its popularization.

*"The aim of illustration is to generate expressive images that effectively convey certain information via the visual channel to the observer."*

(VIOLA AND GRÖLLER, 2005)

*Technical illustrations* visually communicate technical information to a nontechnical observers. These can include drawings and diagrams, precise in terms of proportion and dimension of represented objects. This type of drawings can be regarded a more exact representations of things than, for example, photographs, because tend to clearly show unique traces of given figures. The detailed illustration, taken from the book "Fasiculus Rariorum" written by the German botanist Basilius Besler [FIGURE 2.14], represents five marine creatures in different views to highlight the distinctive parts of each animal (ISELBURG, 1616). The use of illustration is not convenient for

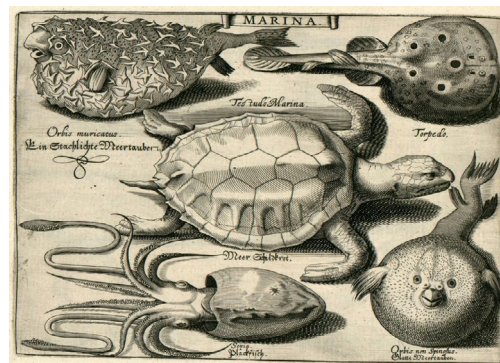


Figure 2.14. Illustration, "Fasiculus Rariorum," by Petrus Iselburg, 1616.

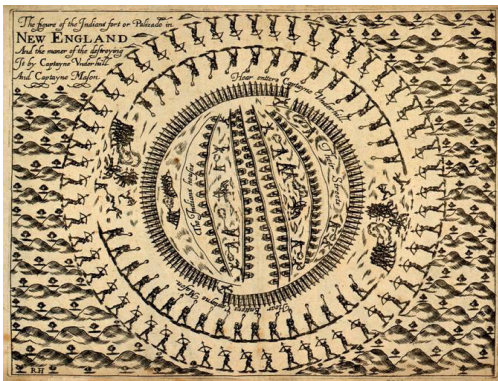


Figure 2.15. A woodcut print illustrating the Mystic Massacre, from Captain John Underhill, 1638.

an historical military event that took place during the Pequot War. Captain John Mason with the English settlers and indigene allies set fire to a fortified Pequot tribe village near the Mystic River. They killed those who tried to escape and then exterminated the entire village, consisted mostly of women and children. That massacre was a consequence of previous Pequot attacks. The xylograph shows a bird's-eye view of the settlement in the center, the surrounding attackers and the landscape (MASHANTUCKET PEQUOT MUSEUM, 2009). All graphic elements are exhibited as icons.

Propagandistic posters also inherited this persuasive graphic language to affect the society. The one on FIGURE 2.16 represents the worldwide known goods manufactured in and exported from Portugal. The information on the poster attempts to motivate the Portuguese people that the country is small but renowned (DIAS QUE VOAM, 2012). In many cases, those messages augmented the industrial and rural productivity and employment rates.

*International System of Typographic Picture Education*, or *Isotype*, is a method of showing social, historical, biological and technological contexts in a pictorial way. It was developed between 1925 and 1934 by Otto Neurath. The term Isotype was appropriated around 1935. The principle of Isotype pictograms is to represent things preserving the clarity with instructions for graphic configuration and color use. It always was the graphic language that helped and informed common people about their place in the world. The maximum exposition

the visualization of complex data, since illustration is usually handmade. It can be applied to the stylistic elements of a graph, depending on the concept of the visual theme.

*The iconographic language* is used in many types of supports and contexts. The imagistic synthesizing capacity and the power of visual communication are widely recognized. Icons are a simple way to pass a message of many words.

The *Mystic Massacre* [FIGURE 2.15] is

The *Mystic Massacre* [FIGURE 2.15] is an historical military event that took place during the Pequot War. Captain John Mason with the English settlers and indigene allies set fire to a fortified Pequot tribe village near the Mystic River. They killed those who tried to escape and then exterminated the entire village, consisted mostly of women and children. That massacre was a consequence of previous Pequot attacks. The xylograph shows a bird's-eye view of the settlement in the center, the surrounding attackers and the landscape (MASHANTUCKET PEQUOT MUSEUM, 2009). All graphic elements are exhibited as icons.



Figure 2.16. Poster "Portugal Manufactures and Exports World Renown Quality Products," by unknown author, 1959.

of these techniques can be found in Neurath's book "International Picture Language." The statistical diagram on FIGURE 2.17 represents relative unemployment rates in the United States, Great Britain, France and Germany in five years (BURKE, 2009). One man figure represents 500,000 lost work places and each of four included diagrams is a horizontal bar chart. The author, Gerd Arntz, depicts the differences between amounts of unemployed people.

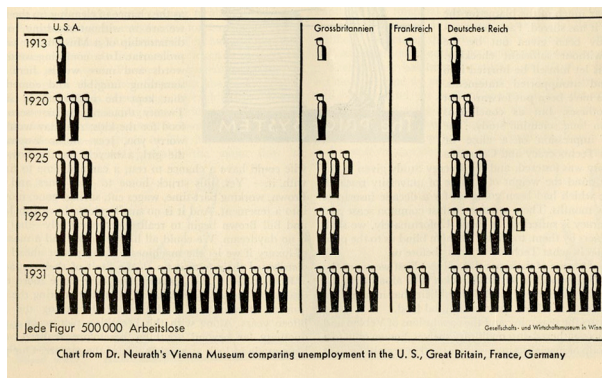


Figure 2.17. Chart comparing relative unemployment in various countries, by Gerd Arntz, Isotype, 1931.

In the era of Futurism and Constructivism, there can be found a countless diversity of charts and diagrams representing aspects of the working class. FIGURE 2.18 illustrates a series of business questions, designed by Irving Geis, known by its scientific paintings and drawings (GEIS, 1939). The survey includes such fields as legislation, labour units, foreign trade and industrial progress. The options and responses are hierarchically organized in each section. Both of them are accompanied by icons and charts to communicate the results visually. This kind of graphics became popular and is used in mass media to call attention and to highlight the public opinion.

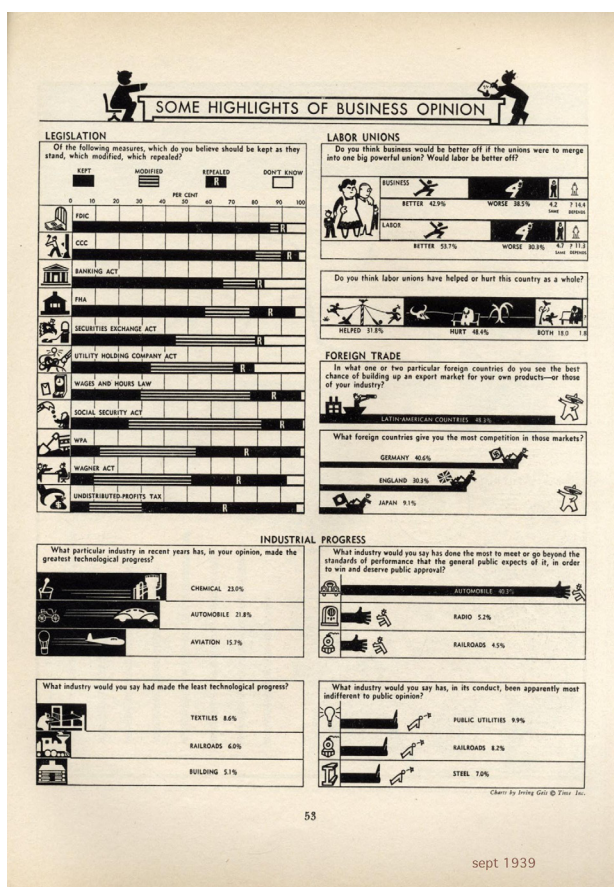


Figure 2.18. "Some Highlights of the Business Opinion," by Irving Geis, Survey Graphic magazine, 1939.

Iconic language is a universal form of communication. Simple style and common figures make the icons a strong tool in all design areas. In information visualization, the icons can be applied to the graph's variables, to descriptions and labels, and to elements of the interface.

### 2.2.4. Statistical Charts

The statistics had its origin in 1749. The German scholar Gottfried Achenwall proposed that the study of society by counting should be designated *statistik*, because it deals with the natural *states* of society. A Scottish Presbyterian minister introduced the word into the English language in his first volume of "Statistical Account of Scotland," in 1791. The harbingers of this new discipline, neither mathematicians nor scientists, began to call themselves statisticians – collectors, organizers, analysts, interpreters and presenters of numbers (MANOVICH, 2011).

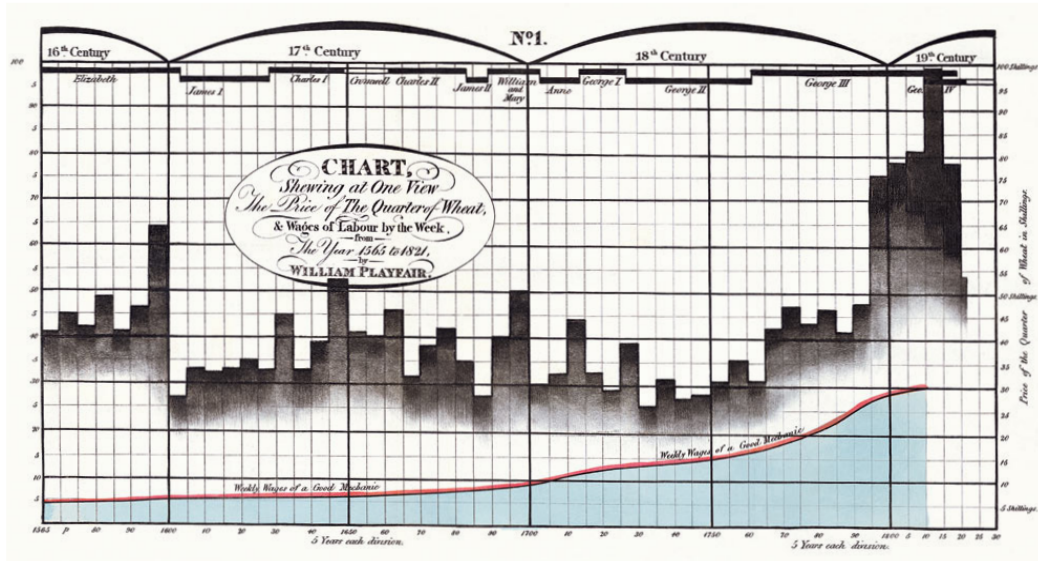


Figure 2.19. “Chart Showing at One View the Price of the Quarter of Wheat, & Wages of Labour by the Week from the Year 1565 to 1821,” by William Playfair, 1821.

The statistical chart on FIGURE 2.19 is the William Playfair’s most famous work. It shows values of weekly wages of labor and prices of the quarter of wheat along a timeline. The combination of the horizontal and vertical axes to map time and money was such a novelty that he had to add a textual explanation:

*“This method has struck several persons as being fallacious because geometrical measurement has not any relation to money or to time; yet here it is made to represent both.”* (PLAYFAIR, 1801)

In this diagram, two main variables are mapped on the two-dimensional graph. The wheat prices and the labor wages are in shillings. The timeline’s base unit is year, grouped by periods of reigns of various monarchs. This chart adds some design drama. The accentuated climb of the wheat price obfuscates the almost plain slope of weekly wages. Yet, this is an illusion. The cost of wheat relative to wages fell during two centuries until 1800. Playfair obtains this illusion by plotting the values in two different ways. He takes a five-year average for the wheat data and runs a horizontal line between the two vertical lines which mark the beginning and end of a five-year period. The position of the ladder plot over the wages’ curve and its dark fill contribute to the illusion (COSTIGAN-EAVES AND MACDONALD-ROSS, 1990).

Florence Nightingale was the British nurse. Attending wounded and sick soldiers, she made several observations of the terrible conditions of the British Army hospitals. In 1858, she wrote a report to the government depicting conditions and possible improvements. These conditions were shown in a rose-shaped diagrams (so called *Nightingale’s roses*). The graphs [FIGURE 2.20] display monthly deaths in the hospitals in the Crimea and provide a comparison with the Manchester hospitals in the same period of time. The diagrams are divided in segments that represent temporal line. The area of each of them corresponds to the number of deaths. The sanitary improvements applied in March 1855 are clearly visible (GILL AND GILL, 2005).



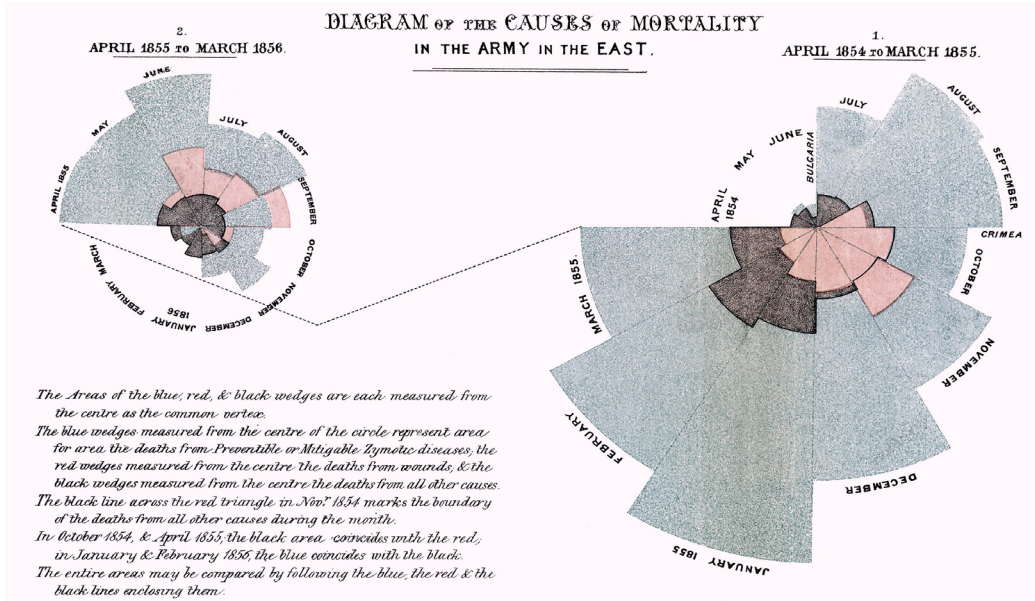


Figure 2.20. "Diagram of the Causes of Mortality in the Army in the East," by Florence Nightingale, 1855.

Statistical charts are the classical way to represent complex and concrete data. Generally, these graphs present information in conventional information graphics' types, such as bar charts, pie charts, scatter plots, box plots and time-series diagrams. All these layouts are useful for display, comparison and study of multivariate data, but usually present the data in a static manner. These can be applied to the punctual parts of visualization of movement, where a simple form of visualization is required.

### 2.2.5. Trees

Tree structures encode hierarchical data using connection or containment (CARD ET AL, 1999, P. 149). Their name and visual appearance are inspired by real trees. The origin is compared to the root, ramifications to branches and terminations to leaves. The connections are utilized to create node-link diagrams to encode relationships between elements, or nodes. Usually, the trees present data organized by levels of filiation, i.e., the child-nodes are connected to the father-node with an unidirectional link, and a child-node can affiliate its own child-nodes. The spatial arrangement in tree structures is merely organizational.

Ernst Haeckel's "Stammbaum des Menschen," shown on FIGURE 2.21, represents the evolution of Man – the unique tree with the thick trunk drawn by the author. He used a self-invented reconstruction method to illustrate the phylogeny of humans instead of the detailed series of human ancestors. Each stage corresponds to a particular period of time in the descent of humans. During these stages, Haeckel assumes

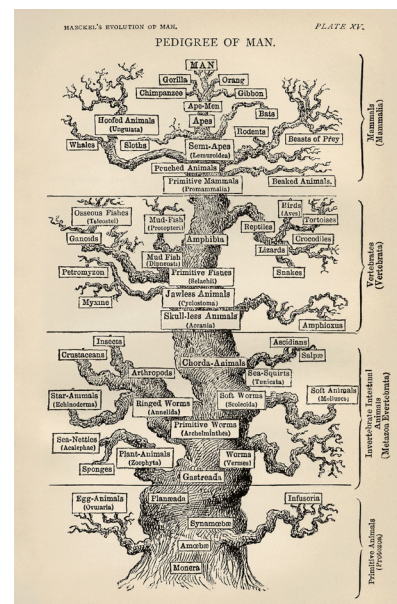


Figure 2.21. Lithography "Pedigree of Man," by Ernst Haeckel, 1874.

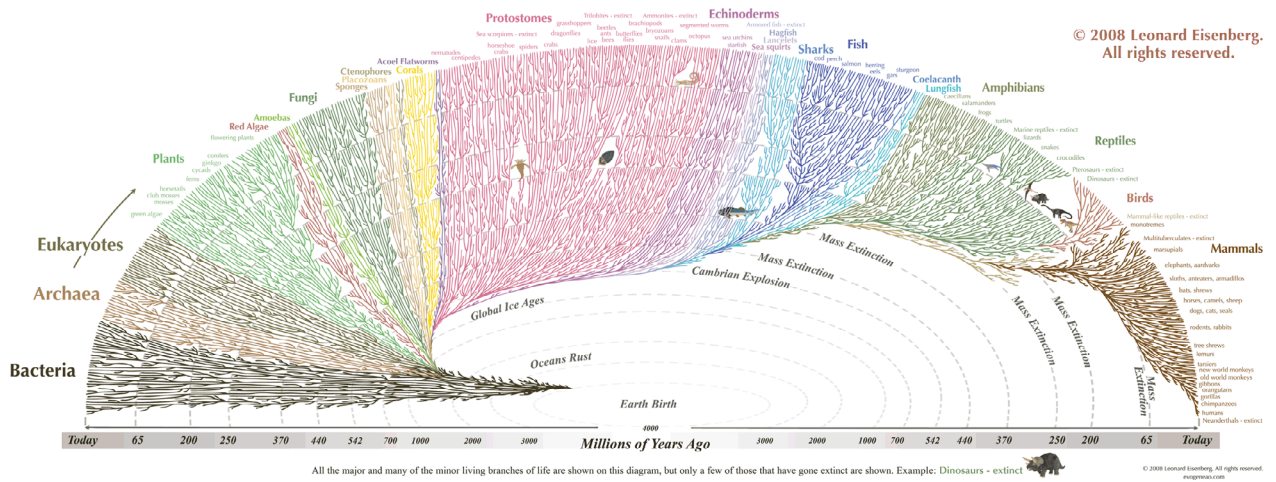


Figure 2.22. "The Great Tree of Life," by Leonard Eisenberg, 2008.

that the ancestors of Man had a similar morphology (DAYRAT, 2003). The trunk indicates the main evolution line with the Man at the top. The branches show lineages non-related to humans.

"The Great Tree of Life" diagram [FIGURE 2.22] is a ramification system based on the evolutionary relationships. The smaller branches illustrate the effect of mass extinctions and changes in diversity through time (EISENBERG, 2008). The graph tries to teach a lesson of evolution – all creatures are related to every other living and ever lived creatures. The vertebrates occupy a bigger part of the tree. The human-kind is located at the right extremity, what falsely involves that humans are the last goal of evolution. The bacteria holds a vast genetic diversity, yet occupies a relatively small section of the tree.

The tree layout is a visual structure suitable to display multilayered data. The quantity of roots can be infinite, as well as the amount of the branches linked to a root. With this, one can clearly demonstrate a hierarchical *father–children* distribution and place the similar elements on the same level. Since the complexity of the data in the context of the flows in wireless networks can rise greatly, the ramifications can be useful to represent partial elements or the entire graph.

### 2.2.6. Cartograms, Matrices and Small Multiples

The bunch of grapes [FIGURE 2.23], made by Alexandre Suannes, represents the levels of red wine consumption in various countries with a circular cartogram (DORLING, 1996). That type of cartograms replace the map objects with the same shape type (in this case, a circle) with sizes proportional to the data values. In this graphic, the layout is also *context-based*, where the visual appearance refers to the direct context of the data (the wine consumption). Each country is shown as a cutie and its size corresponds to the consumption in liters per capita in 2006. The graph shows the relative wine consumption. By the visual estimation, there can be seen that Luxembourg holds the first place, then appears France, Italy, Portugal, Swiss, Croatia, Spain and so on (BIG THINK, 2009). Also, some relative conclusions can

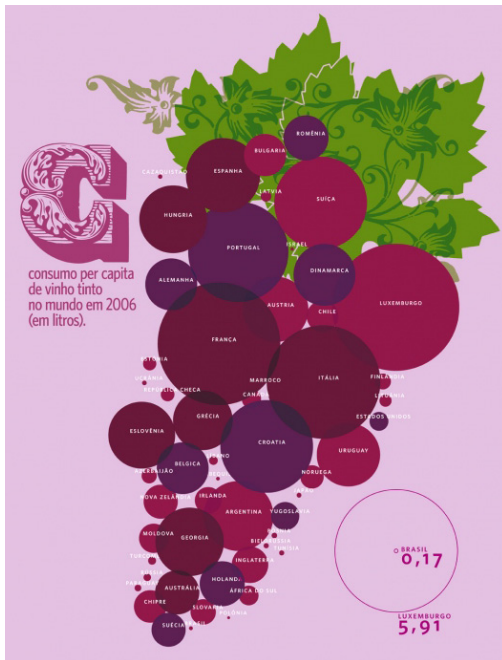


Figure 2.23. "World Wine Consumption," by Alexandre Suannes, 2006.

be derived. Firstly, red wine consumption can vary hugely between countries—neighbors (comparing Paraguay to Uruguay, for example). Secondly, some countries that are big viticultural producers are tiny consumers (Chile and Moldova). Lastly, red wine consumption should not be equaled to overall alcohol consumption. Another alcoholic drinks might have been ruling the market.

*Matrix plot* is a layout that represents the direct correspondences between variables, depending on the dimensions of the data. In a two-dimensional matrix, the components are mapped on two axes and the relationships between them are mapped on the intersection points of the corresponding entries.

The reorderable matrix is the invention of Jacques Bertin. The *adjacency matrices* displaying multidimensional data must be reordered to reveal interesting patterns. Bertin built several mechanical devices, called *Domino*, to conduct experiments with various reordering techniques.

The matrix on FIGURE 2.24 shows the laureates of Nobel Prizes since 1901 until 2011. *The Nobel Prize* is a set of annual international awards in different categories of cultural and scientific advances. The awards are represented in dots, divided in six categories and grouped by country of nomination. Germany dominated the field until the Second World War, then the United States have taken leadership. The graphic was made using data from the website *nobelprize.org*. The winners are assigned to the current country of affiliation. In cases where affiliations were not listed, the birth country is utilized. The rows from the United States to Russia display the seven countries ordered by the most Nobel laureates. Japan and China show acceleration trends in the winning process (BRUNER, 2011).

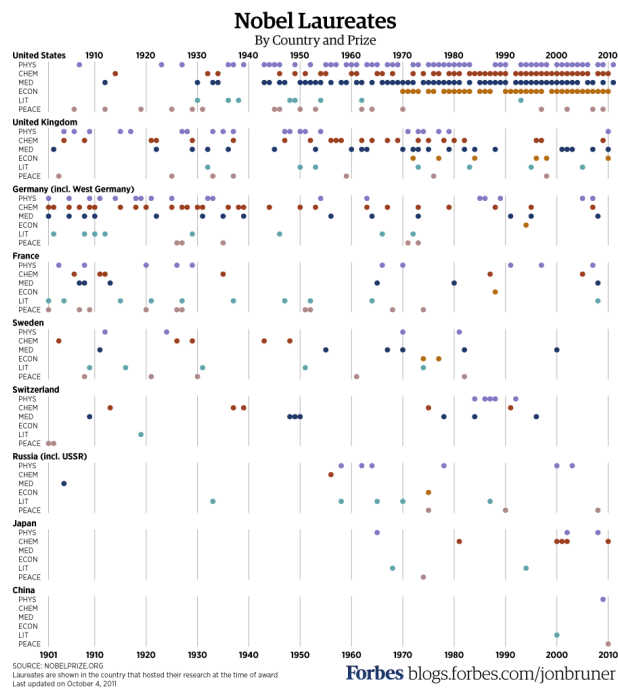


Figure 2.24. "Nobel Laureates by Country and Prize," by Jon Bruner, 2011.

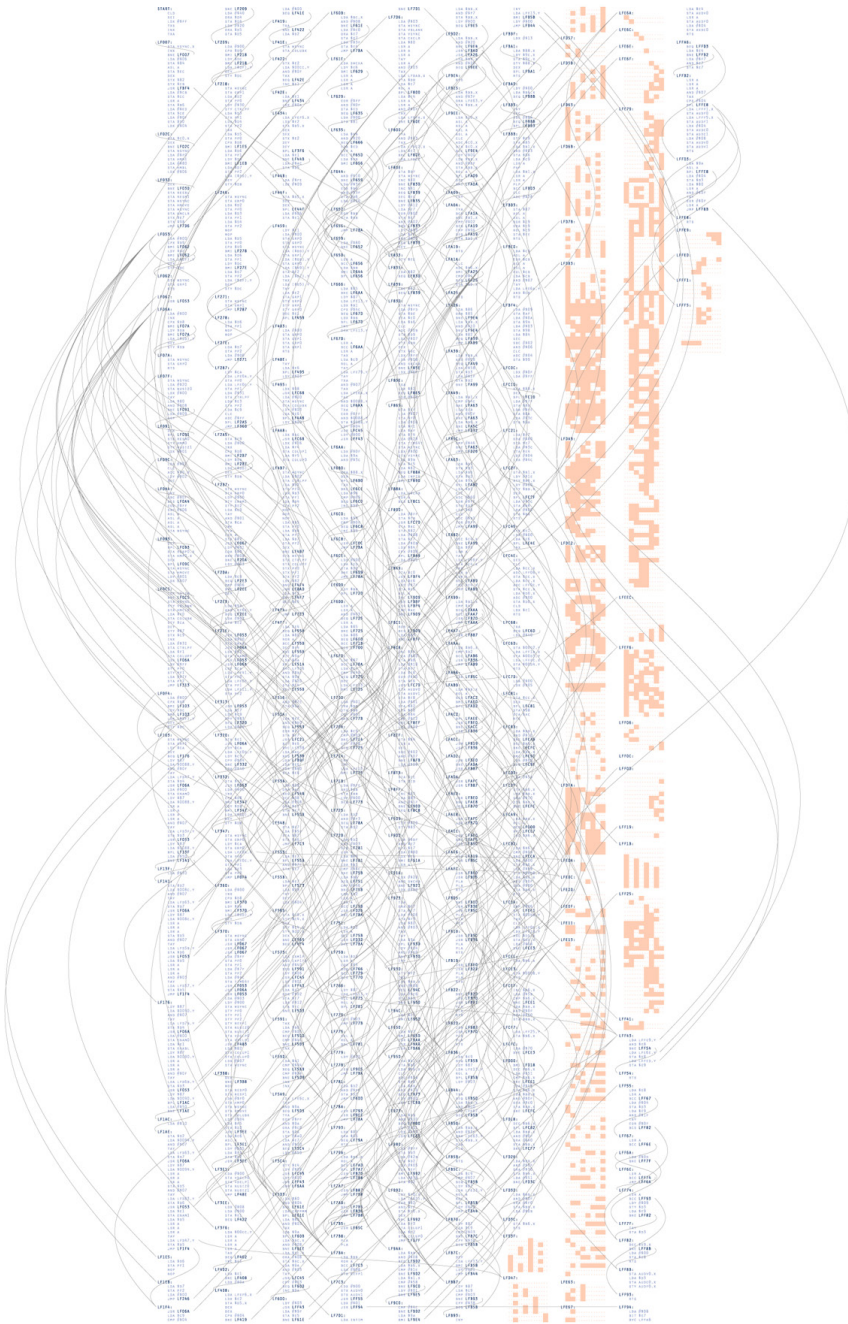


Figure 2.25. "Distellamap of Pac-Man," by Ben Fry, 2005.

The "Distellamap" [FIGURE 2.25], developed by Ben Fry, is an approach to visualize the operational code in *Atari 2600* games. These games contained executable code commingled with data. The code statements in *Assembly* language are listed in the graph's columns. The curves represent the *go-to* commands, linking the conditional statement and its destination. Bytes of data found in the cartridge are shown on the right with an orange row with eight binary elements: a block for a '1' or a dot for a '0'. The graph shows the complexity of the *Pac-Man* game with a huge count of *go-to* loops, due to the implementation of AI of ghosts (FRY, 2005).

*Small multiples* layout refers to the sequential lists of different states of the same data. Multiples give depth to the vision and understanding by stacking information

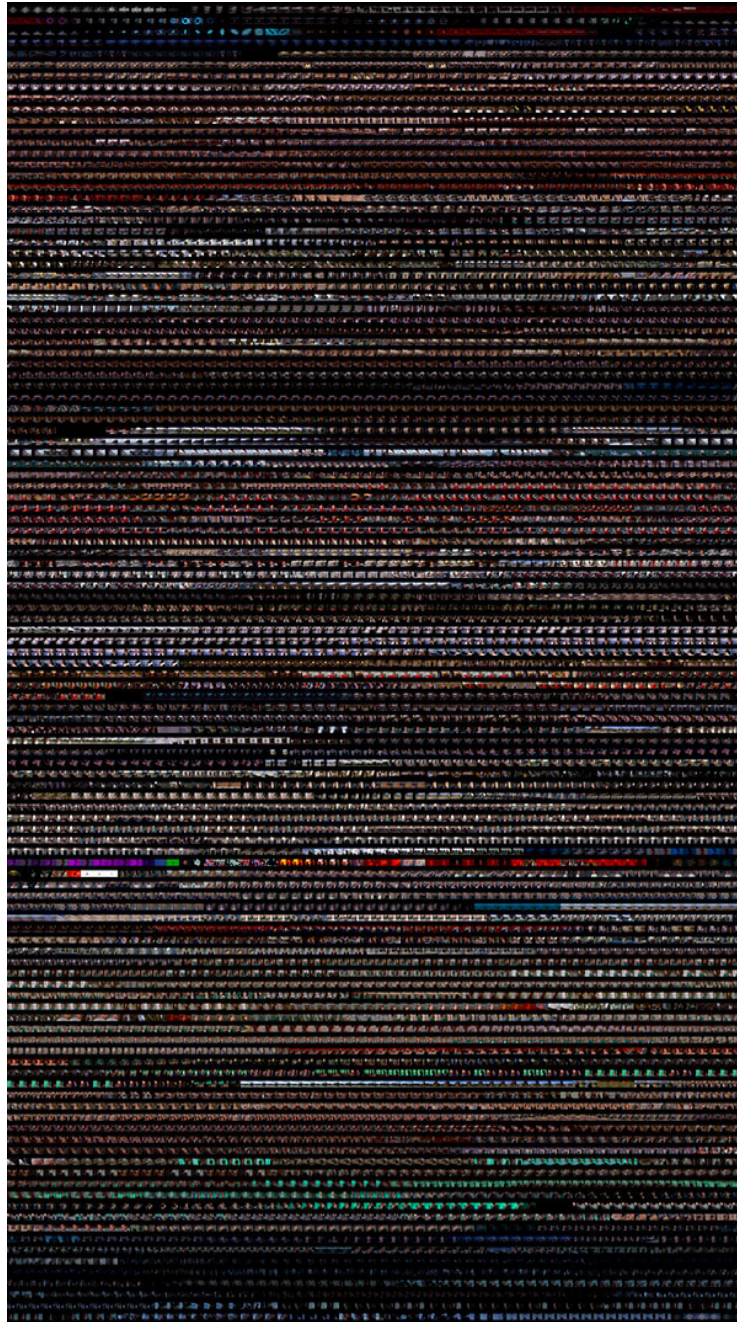


Figure 2.26. "Cinema Redux – Vertigo," by Brendan Dawes, 2004.

in visual arrays, revealing repetitions, changes and patterns. The layout based on small multiples helps to analyze, differentiate and compare the meaning of the continuous data (TUFTE, 2005, PP. 105-119).

The "Cinema Redux" project, created by Brendan Dawes, explores the idea of reducing a whole cinematographic piece to a single image. The approach consists of sampling a movie once per second and getting the miniature of the current frame. The process analyzes the entire piece, stacking miniature frames in rows. A row represents one minute of film time. As the result, the fingerprint of the film is achieved, a movie DNA displaying the colors and the rhythm of the editing process. The variations of the brightness and the hue of colors indicate scene changes (DAWES, 2004).

## 2.3. Techniques for Visual Communication

As any technical field, information visualization has patterns that lead the development to successful results. These rules are not rigorous, normally the authors choose the techniques appropriated for their representations. Information visualization is not an exact science, but there are norms and standards for an efficient communication of information. Several authors were describing these representation techniques in their lifework to define the better ways of graphing different sorts of data. The study of these important methods and norms is called *graphic semiology*.

This section explains the theoretic principles of data visualization and good practices of the construction of graphs. Jacques Bertin (1918-2010) was a French cartographer and one of the first who joined and organized the subjective matter of the graphic representation of information. The first part of this section exemplifies the base notions of his study and the second depicts the work of Edward Tufte (born 1942), an American statistician.

### 2.3.1. Semiology of Visualizations

The graphic treatment and organization helps to perceive and understand the information more efficiently. Accordingly to Bertin, a viewer needs at least 20,000 instants of perception to compare two tables with 100 rows and 100 columns, while graphically displayed data can be understood almost instantaneously.

The graphic system of a visualization consists of the *content*, or the information to transmit, and the *container*, or the properties of the system. The content is usually not discussed, the data is mined and treated by the professionals in that area. The visual transcription of the information, its quality and efficacy are the aspects that should concern the data designers. The information is constituted by the components and the invariant. The *invariant* is the main idea, the concept or the notion to be transmitted using the variable *components*, or visual variables, in the planar dimensions of the graph. The components have a *finite quantity* and are divided in categories (for example, “car,” “ship,” “train” are categories of a variable “vehicle type”). The number, or *length*, of categories is also finite. Visual variables can be organized in different *level types*: the qualitative, or nominal, level is based on a perceptual distinction of different elements and a separation of these in distinct groups; the ordered level implies a ranking of the elements in a universal manner; the quantitative level is performed using a countable unit in the sorting of the elements. Each element also has a unique current *position*. The position determines the reading of the displayed information (BERTIN, 2011, PP. 3-13).

*“One cannot study a graphic intelligently without knowing the invariant and the components displayed in it.”* (BERTIN, 2011, P. 19)

Bertin defined six *visual variables* in the two-dimensional plane [FIGURE 2.27]. These are: size, value, color, texture, shape and orientation. *Size* is a variable which variation reflects on the dimensions of an element. *Value*, or variation of gray, is a visual

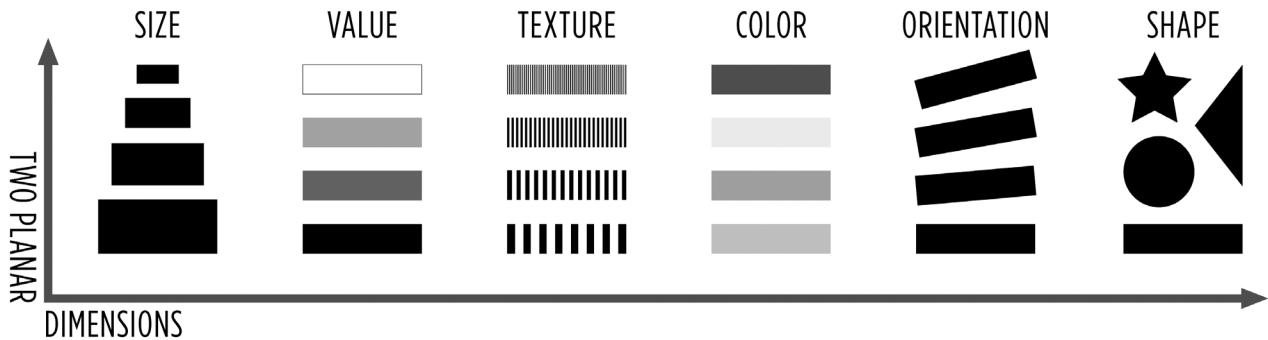


Figure 2.27. Six visual variables (size, value, texture, color, orientation and shape), represented in two planar dimensions.

ratio between the total amounts of black and white colors. *Color* can be defined as the perceptible difference between uniform areas with the same value. *Texture* is the number of separable marks in a given unitary area. *Orientation* is the angular difference between parallel marks. *Shape* is the element of similarity recognized in different variables (BERTIN, 2011, PP. 71-97).

Each variable has perceptual properties, which determine its level [FIGURE 2.28]. A variable is *selective* if the viewer can isolate all correspondences that belong to the category of this variable. These isolated elements form a family of the same selected variables. A variable is *associative* when it allows grouping of correspondences differentiated by this variable. Shapes are associative (polygons of the same size and color are similar), values are not associative (polygons of different color hues are not seen as similar). A variable is *ordered* if the classification of its categories is universal and immediate. A light-gray value is nearer to the lower limit (white) than a dark-gray value, that is nearer to the upper limit (black). This specification of different locations of the same scale produces an order – the same cannot be said about the blue, red or green hue at equal values. A variable is *quantitative* if the visual distance between two categories of an ordered component can be expressed by a numerical ratio (BERTIN, 2011, P. 48).

*Imposition*, or the utilization of the planar dimensions, depends on the nature of the displayed components. Bertin highlights four different impositions: diagrams, networks, maps and symbols. *Diagram* consists of two variables mapped on perpendicular axes, and all values of one variable can be established for all values of another variable. *Network* usually has one variable and the correspondences can be established between all elements of that variable, i.e., each value can correspond to all other existing values. *Map* is a network in which the elements can be arranged

PLAN.DIM.	SIZE	VALUE	TEXTURE	COLOR	ORIENTATION	SHAPE
A	<del>A</del>	<del>A</del>	A	A	A	A
S	S	S	S	S	S	
O	O	O	O			
Q	Q					

Figure 2.28. Perceptual properties and levels of organization of the visual variables: associative A, selective S, ordered O, and quantitative Q.


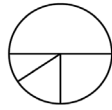
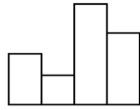
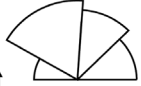

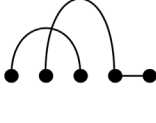
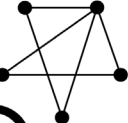
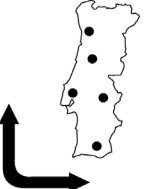

	ARRANGEMENT	RECTILINEAR	CIRCULAR	ORTHOGONAL	POLAR
DIAGRAMS					
NETWORKS					
MAPS					
SYMBOLS					

Figure 2.29. The impositions (rows) and the respective types of imposition (columns).

in a geographic order. *Symbol*, also known as icon, does not have an established correspondence on the plane, but outside the graphic, for example, with the viewer. SUBSECTION 2.2.3 demonstrates examples of visualizations with icons. Symbols are also based upon figurative analogies (BERTIN, 2011, PP. 50-51).

In the construction of graphs, the variables can be inscribed over the plane accordingly to a specific *arrangement* or to a *type of imposition*: linear – the reading of elements is made in a line, as in a 100% bar chart (BRINTON, 1939, PP. 92-97); circular – the reading is made in a radial way, as in a pie chart; orthogonal – the reading is performed combining two axes, horizontal and vertical; or polar – the information can be read both in orthogonal and circular ways [FIGURE 2.29]. The use of visual variables in the layout produces elevations, what combined with the types of imposition produces different *types of construction* (BERTIN, 2011, P. 52).

Jacques Bertin made a huge and scrupulous study of a graphical representation of statistical data. His detailed explanations and examples describe the construction of graphs and the ways of their reading and understanding. Despite being thorough, his study is, above all, technical. The study of Edward Tufte is also technical, but more focused on the functional aspects of the graphs. Principally, he described good and bad practices in information design, rich in classic and contemporary examples of different areas of information visualization. The main ideas and a series of principles are described in the next subsection.



### 2.3.1. Principles of Graphic Excellence

*Graphical excellence*, explained by Edward Tufte (2001, p. 51), is a well-designed presentation of the truth about data, where complex ideas are communicated with clarity, precision and efficiency. The concept consists of providing the greatest number of ideas to the viewer in the shortest time of communication with the graph, having the least quantity of elements distributed within the smallest space possible.

Graphics that show data need to attract attention to the sense and the substance of the data, representing clear, truthful and direct quantitative contents. *Showing the pure data* is the fundamental principle of well-made graphs. This principle relies on several aspects. Graphics should not contain non-related information to the subject data. Often, additional elements need to be included, such as grid lines, ticks and scale numbers, to create a *context* around the data. The objective is to minimize the quantity of these visual elements that distract users from the observing, studying and understanding the important information, presented in the graph. Erasing the non-information, minimizing the graphic redundancy and maximizing the use of crucial elements are ways of making visualizations cleaner (TUFTE, 2001, pp. 91-105). This will focus the attention on the graph's essence and its principal message.

The visual perceiving is the main way of communication between the viewer and the graph. The composition's properties of graphs reflect their integrity, trustfulness and clarity for the users. The *distortion* of data representations occurs when the information is not consistent with its exact numerical values. This phenomenon can be denominated as the *veracity ratio* of a graph. The greater the difference between the effect presented in the graphic and the effect provided in the dataset, the more the graph lies. Tufte provides two aspects of preserving graphical integrity. The first consists of the representation of data in the graph proportionally to numerical values. This involves a mapping of information to the layout in contextual and focused ways, assuring that perceivers understand the data correctly (TUFTE, 2006, p. 45). The second aspect stays for the use of detailed and complete labels. Explained and labeled information reduces graphical distortion and equivocation.

*Visual decoration*, used in editorial purposes, should not interfere with the data shown in a graph. Decorative elements can be applied around the graphic, without affecting the values. The suitable example for this situation are patterns with *moiré* effect. These graphic elements generate optical illusion that distracts the eye from the data. Another aspect that prejudices the graphical perception is the variation of design instead of the variation of data. This happens when the information is distorted in favor of the design decisions. For example, the use of perspective views (false 3D projections) in 2D bar charts creates additional visual cargo. The second or third visible sides lead to misconceptions, because the data variables are also being associated to these graphic elements, while should not be. Likewise, one-dimensional data should not be represented in two or three dimensions, unless there is a valid reason to do it. The 1D variable values represented by 2D shape areas are common examples of the violation of this principle. Firstly, each axis must have an information associated

to it, accordingly to the minimization of the graphic redundancy. Secondly, the value of the shape's area should not exceed the data value. The variation of shape areas that represent values with a small difference can be barely noticeable, and there is always a tendency to expand the shape to make it look bigger. Having this, the designer breaks the proportionality between the graphic element and the data it represents. The creation of informational context in the graph provokes a major level of understanding of its essence and importance. For graphics with small amounts of data, the context can be derived from the correct explanation and labeling. Once the quantity of pure data is very small, the additional data can be provided in textual form. Gathering of supplementary data enrich the values with comparability and loquacity.

The *use of color* in information visualization (TUFTE, 1990, pp. 81-95) is an important aspect of design. The color theme for a visualization should be created accordingly to the context and the purpose of the graph. While simple and direct data must be represented with a sober, or even monochromatic, color palette, complex information should be separated by hierarchical layers. These layers of information establish the order of importance of displayed elements. For example, bright and contrasted colors bring data to the foreground of reading, muted and gray colors can represent secondary information. Bright colors on a neutral and light backgrounds highlight the data. Colors can also be used in labeling and categorizing the data – the same types of elements painted with the same colors. The quantifying visual meanings can be applied to graphic elements in a given context – the darker color tones signify deepness of a sea or highness of a mountain in topographic maps. Yet, the color decisions must be taken carefully. User unfriendly color combinations, such as composed only with rich and bright colors or only with light non-contrasted colors, may prejudice the graphic appearance and convey the viewer to a false visual perception. The colors found in nature are user-friendly, since humans are accustomed to see them daily. Also, we should not forget that colors interact with each other, when represented in a layout. Certain combinations of colored elements produce optical distortions that produce undesirable reading of represented information.

## 2.4. Interactivity in Graphs

As was mentioned above, one of the most important advantages of the use of computational approaches to create visualizations of information is the capability of *handling and encoding of multivariate data* in classifiable variables, within the programmatic software framework. Another great advantage brought by the technological progress is the interactivity. *Interactivity* is the behavior of the application experienced by the user, referring to the functionalities offered by the artifact and to their control via the graphic user interface. Interactive software receives an *input* from the user and responds with *output* results. Intuitive and rapid interaction fundamentally changes the process of understanding the information displayed in a visualization. It stands for a tool of exploration of the data, depending on the provided interactive possibilities, as well as a control center of the graph's parameterization, allowing the user to examine what happens when variables are modified.

### 2.4.1. Levels of Interaction

There are three levels of interaction: an immediate response, one second response and 10 second response. The first level, also called *the psychological moment*, is at 0.1 second. Stimuli of perception within 0.1 second of each other merge into one percept: two similar pictures fuse into a motion perception, two sounds fuse into a single sound, an action and its perception occurred within that interval exhibit cause-effect relationship. The same happens in interactive graphs: if an interface element changes its property, such as color or position, within 0.1 second after being activated, the action will seem to have caused the change; if the data changes within 0.1 second of moving a slider, it will seem that the slider is directly changing the data. In these cases, the performance of the application is the crucial aspect. The second level is at 1 second, the time of an *unprepared response*. Events that happen in less than one second interval occur too quickly for the user to respond, unless he is prepared for it. This level can be applied for smooth interaction. For example, one second animations can convey information without slowing the user down. The third level is *the unit task*, with a common cyclic duration from 5 to 30 seconds. This is the typical pace of the routine cognitive skill, in which the user can perform a routine procedure. The unit task is the time to do the minimal unit of cognitive work (CARD ET AL, 1999, P. 231).

### 2.4.2. Interactive Techniques

There is a great variety of *interactive techniques* that can be implemented in an interactive visualization. Each technique has its mode of activation, presenting and exploring capabilities and benefits. These methods are incorporated accordingly to the purpose of the transformation of visual variables and of the data. While interacting with a graph, the user may have different interaction possibilities, such as data transformations, view transformations and visual mappings of elements (CARD ET AL, 1999, PP. 233-234).

The *data transformations'* group defines six types of interaction. *Dynamic queries* are used to select and filter the data to obtain more specific results. The possible choices can be presented with the user interface controllers, such as slider bars, simple buttons and radio buttons, to select the desired conjunctions of the existing information and display only the necessary data. *Direct walk* technique is based on a set of linkages, that vary from case to case. Generally, the mouse pointer is used to perform the navigation through links, searching and modifying the information. *Details-on-demand* is an exploration technique that the user can activate to expand an object and reveal its hidden variables. This allows the saving of the layout space and, consequently, mapping more elements. *Attribute walk* is a searching procedure, in which the user selects an intended case and searches for similar attributes in other cases. It is also used for comparison of the picked results. *Brushing* is a selection of the same objects or data from multiple visualizations to compare results. The user highlights a case from one visualization and the other views show the same case's information. This technique is accompanied by range selectors that brush the data. *Direct manipulation* is used to modify transformations, allowing the definition of various parameterizations for the displayed elements and variables.

The *view transformations*' group presents another six techniques. *Direct selection* is a set of schemes for selecting objects and groups of objects. These schemes can enhance the visual structure of a graph, highlighting the chosen objects that may be used for different posterior actions. *Camera movement* refers to the change of the observer's position. It is applied when there are benefits in exploring a visualization from different points and angles of view. This method reveals the details that are otherwise invisible. *Magic lens* is a combination of two procedures: the selection of objects by position on the layout and the application of other techniques, such as dynamic queries, direct manipulations or data/view transformations, on all selected items. *Overview+detail* technique consists of two or more layers of connected visualizations. One of them shows an overview layout with all objects or the objects of main importance. Another view is detailed, overlaying the general view and showing more information of the selected object. *Focus+context* is a variation of the previous technique, in which the highlighted information on the top layer may represent other types of information instead of the variables. A fish-eye lens is the common tool used in this technique. It visually distorts the layout in order to emphasize and focus the information presented in its center. *Zooming* technique involves the reduction of the amount of visible objects, transforming the view in some way. More variables and informations can be shown or a scale transformation can be applied to reveal smaller details.

The interaction with *visual mappings of elements* does not have a finite number of applications. There can be implemented an uncountable amount of types of transformation of all possible shapes/structures/layouts into another ones. Two of the earliest techniques are the *dataflow diagram*, that consists of the mapping of an existing layout to the explicit system representation, for example, a node-link diagram, and the *pivot table*, that lets the user manipulate presented data and easily map it to a standard table layout.

### 2.4.3. Data Exploration and Interactivity

Interaction is a real-time process. The visual elements of a visualization system must present different interactive states to in order to not cause the misinterpretation by the user. For example, if the system is running fast, the element that changes its state from the starting point to an ending point without an animated transition can simply lead the user to an error.

Interactive methods of the presentation of information are usually entwined with the interactive functions of the data exploration (SPENCE, 2001). For example, if the user zooms in the graph, he might be interested in moving across the graph in the scaled view to visualize and explore other parts of the graph. Zooming in and out to relocate to another points of view is a dull task. *Scrolling* of the data is an obvious solution. This is a navigational action that allows to move the data within its maximum boundaries, presenting the scrolled information in the application's window. It is usually performed with horizontal and vertical slider bars or with the mouse dragging action. The main problem of the scrolling is the difficulty in orientation in the data as a whole. The bigger the scrolling boundaries are, the harder is the user

orientation. This problem can be resolved by a *context map*. That map consists of an abstract miniature view of the entire visualization, with a lower amount of details, and a current view indicator, for example, a rectangle representing the application's window. As the name suggests, this technique is used to provide a contextual localization of the detailed view.

## 2.5. Summary

This chapter provided a definition of the field of information visualization and presented an overview of different types of visualizations. The examples are grouped by the layout type and each category was analyzed regarding its applicability to the visualization of movement and flow in networks. The purpose and the context of the displayed examples were explained.

The main concepts and principles of the construction and analysis of the visualizations of information were described in this section. The basis of this theoretic knowledge is the work and studies of Jacques Bertin and Edward Tufte. They are pioneers in the graphic analysis and representation of data, as well as in the information design. These notions are used in the analysis of the work related to the visualization of flows in wireless networks [CHAPTER 3] and in the development of visualizations within this dissertation [CHAPTER 6].

The interactive methods, described in this section, are base techniques of the *user-application* interaction. These techniques were developed since the beginning of the era of digital visualization. Some of them are implemented in the practical applications developed in the context of this research [CHAPTER 6]. The next chapter continues the theoretical part of this research describing the specific area of information visualization, in which the visualization of flows falls within.



## CHAPTER 3.

# Related Work

Complex graphics and visualizations emerged long before the Computer Age. The previously discussed works of Charles Minard, Florence Nightingale and John Snow [SECTION 2.2] are notable examples of two-dimensional visualizations of complex graphical and informational nature.

The main purpose of the visualization of flows is to make the flow patterns visible. The study firstly started from the visualization of fluid dynamics, in the field of Physics. These were animated computer simulations of the behavior of fluids in different ambients using analytical methods based on mathematical formulas and represented by vector fields. Later, the flows were being represented by the textures with the visual appearance similar to the real material's look. That modification made the directions and the streamlines noticeably clearer, principally because the fluids were mixing together.

This chapter focuses on the clarification of the object of study of scientific visualization and on the work related to the visualization of flows in networks. Different techniques and approaches to that area are presented and accompanied by minutely explained examples, both early and recent ones.

### 3.1. Scientific Visualization

With the technological evolution of computer graphics and data processing capabilities of digital devices (mainly computers), the graphical representation of computer-collected large datasets became possible. The visualization emerged as a new tool of discovery and understanding of scientific information, allowing further research and communication of formerly unknown fields and areas of Science.

Scientific visualization is a branch of the area of information visualization that deals with scientific and technical data, presenting it in conventional graphic ways applied to narrow and specific studies.

There is no specific definition of what is scientific visualization, since the taxonomy of visualization does not exist. Every author that works in that area defines the term in a different way, and all of these designations are correct. For example, one of them says that *scientific visualization* is an amalgam of tools and methods that seeks to promote new dimensions of insight into problem-solving using current technology. The purpose is the graphic exploration and visual analysis of data to gain understanding of information and processes, with the aid of hardware and software systems and techniques (EARNSHAW AND WISEMAN, 1992). This definition perfectly applies to this specific research: we have a problem of development of techniques for visualization of flows to solve; we explore the data graphically with the construction of graphs; we analyze the visual results produced by the implemented applications; and we use a combination of hardware (computer) and software (the *Processing* language) system to solve the problem and to gain the insight of the available network data.

Scientific visualization was originated in late 80s. The earliest report on the scientific visualization was made during a workshop on “Visualization in Scientific Computing,” in February 1987. The document summarized the results of a Panel Meeting on “Graphics, Image Processing and Workstations” sponsored by the National Science Foundation. The author of the report, Bruce McCormick, has drawn several conclusions about the importance and use of the visualization in Science (BROADIE *ET AL*, 1992, PP. 3-4). The contemporary data sources, such as supercomputers, satellites, medical equipment, space and radio stations, produce large volumes of data. The density of those sources increases as technology moves forward, and the volumes of gathered data increase exponentially. Scientific visualization was named an essential tool for handling large amounts of information. Also, the value of interdisciplinary teams was highlighted, once scientific visualization systems benefit from having collaborating disciplines from different areas. Engineers and computational scientists could combine in areas of molecular modeling and fluid dynamics, visualization scientists and engineers could collaborate in areas of visualization software and hardware, artists and cognitive scientists could ensure the best forms of use of the visual communication.

McCormick (1987) recognized a series of issues in visualization to be developed and supported in the future. Some of these are the use of supercomputers as workstations, graphic-oriented programmatic environments, visualization of scalar, vector and tensor fields of high dimensions, dynamic visualization of fields and flows, imaging networks with high bandwidth, signal and image processing, handling terabyte datasets, etc. The described advantages of the development of scientific visualization were also identified. The integrated set of portable tools, the increase of the scientific progress and productivity, the standardization of tools, the improvement of market competitiveness and of overall usefulness of computing facilities are the main benefits extracted by McCormick.



Construction of complex graphics representing large datasets is a hard and rigorous task. The challenge is to find the valid and precise compromise of the use of space and graphic elements. An accurate principle of creation of visualizations says:

*“... It is not how much empty space there is, but rather how it is used.  
It is not how much information there is, but rather how effectively  
it is arranged.”* (TUFTE, 1990, P. 50)

Withal, the large datasets imply high density of information in the graph's space and the quantity of displayed information can be immense. But the purpose of information visualization is not about simplicity of the graph or its reading, it is about the clarity of understanding of data.

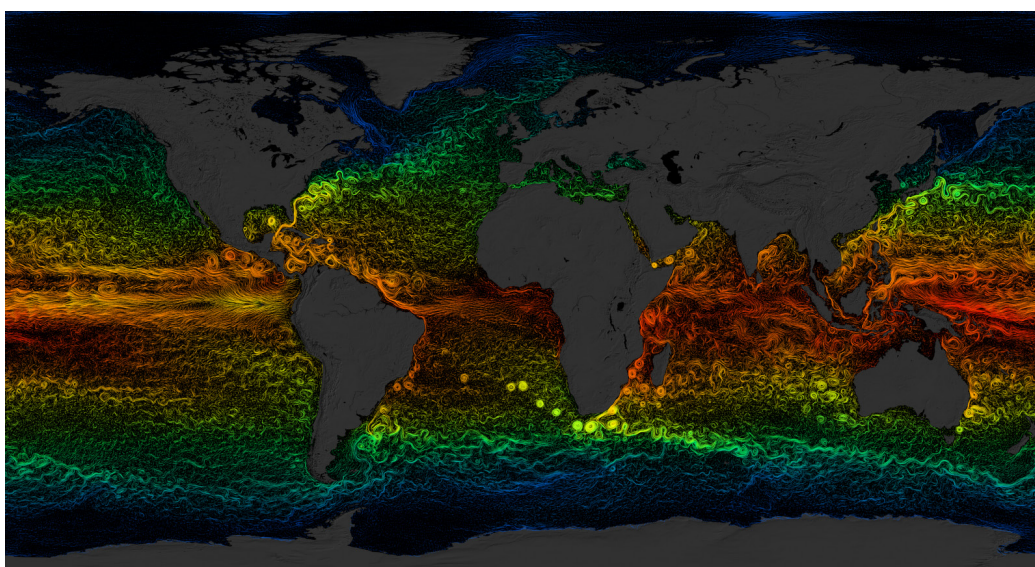


Figure 3.1. The streams on the water surface of the Earth, by NASA GSFC, 2012.

The visualization on FIGURE 3.1 shows the flows on the water surface of our planet (NASA GODDARD SPACE FLIGHT CENTER, 2012). The currents are colored by corresponding temperature (warm waters are red, cold waters are dark blue). This artifact was produced at the *NASA Scientific Visualization Studio*, it synthesizes satellite and local data of the ocean. This is a typical example of scientific visualization, combining a flow visualization and a cartographic layout as the data representation method, and a heat map as a coloring technique to display the information.

Primary visualizations using computer graphics were vector and scalar fields, applied to simulation and presentation of measured data. These used two-dimensional planes for drawing geometric primitives and color mapping. The advent of 3D graphics gave birth to the volume and flow visualization forms. Nowadays, scientific visualization is present in a wide range of areas of study, such as mathematics, medicine, physics, meteorology, astronomy, astrophysics. The main applications of the visualization in these fields are object modeling, data imaging, computation and simulation of system behaviors. The produced results are used for exploration, analysis, extraction of cognition and dissemination of the obtained aftermaths in media.

## 3.2. Techniques for Representation of Flows

The visual arrangement is an important aspect to consider in the creation of the graph. The spatial distribution of elements reflect the nature of the displayed data and the way of studying and understanding the insight of the information. A technique of visualization is typically based on the underlying physical phenomenon of the scientific data to be shown. The aim of the graph's designer is to visualize and make that phenomenon (and not the data itself) understandable for the users (BROADIE *ET AL*, 1992, P. 37).

The construction process of a visualization technique can be divided in three parts. The first part is the *creation of an empirical model*, that consists of an internal model of the physical entity of the provided data. For example, if the given data contains a set of points representing heights of several buildings, a continuous function that interpolates all points can be drawn. This reconstruction helps to distinguish the dependent and independent variables, often implying the use of mathematic formulas. The available variables are transformed in a set of separate samples. Numerical analysis provides different interpolation algorithms for continuous functions through sampled data. The second part involves the *representation of abstract objects*. The empirical model is analyzed and represented abstractly, accordingly to the variables and their properties. In this phase, the choice of abstraction relies on the physical phenomenon, defined above. The third part implies the *realization of the visualization as an object*. The visual primitives are generated and the variables are mapped to the layout in a graphic sequence. This phase can be defined as production of the data visualization artifact (BROADIE *ET AL*, 1992, PP. 38-40).

Flows typically occur in large systems, where exists exchange of data between network's constituents. This research is based on the analysis and representation of flow in networks, since the provided data portrays the user connectivity within the wireless campuses of the four portuguese universities. The further examples demonstrate and describe a selection of most common methods of visualization techniques and layouts for different types of flows. Typically, a graph combines a data arrangement technique and a topology of spatial layout, suitable to visualize the invariants and the components of the flow.

The term *spatial layout* stands for the base scheme of representation. The layout is a static container that holds the information and provides the context to the visualization as a whole. The classic examples of layouts are bar charts, pie charts, plots, trees and maps. All these layouts were discussed in SECTION 2.2. The specific layouts related to the visualization of flows are examined and explained below.

The visual structures called *networks* represent frameworks within various contexts, such as communication, politics and semantics. A network layout consists of nodes (subjects or objects) interconnected with edges (connections). The links can have associated information, such as weight, length or label. Also, networks have inner system complexities, such as cycles, paths and correlations. The links can



Figure 3.2. Force-directed layout (Bostock, 2010).

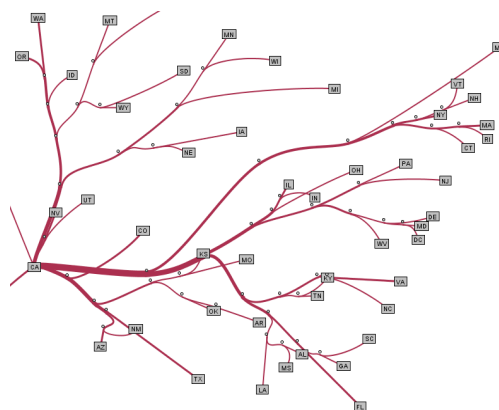


Figure 3.3. Flow-map layout (Phan et al., 2005).

be directed and undirected. Nodes and links can be unlabeled, labeled, ordered or quantitative (CARD *ET AL*, 1999, p. 187). Networks use space as a dimension of presentation, showing relationships between elements through their interconnectivity.

One of the most common methods of displaying networks is a *force-based layout* [FIGURE 3.2], also called *hairball*. The nodes, usually represented by simple geometric shapes, are interconnected with edges, typically displayed as lines or curves. The purpose of this layout is to distribute nodes in the space with minimal edge intersections, assigning forces among edges and nodes to simulate of a physical system. During several iterations, the whole system tends to ease and come to a mechanical equilibrium state. The most straightforward method is to assign the attractive force to edges (spring-like behavior) and the repulsive force to all nodes (electrically charged particles). Thus, the edges tend to have uniform length due to the attraction, the linked nodes tend to separate due to the repulsion between them, and single nodes are repelled.

Hairball is a problematic representational technique and its result depends on many factors, mainly the data and the proportion between values assigned to the edge lengths. There are various algorithms implemented to optimize the spatial distribution of force-directed layouts, such as circular-by-degree, edge-weighted, spring-embedded, inverted self-organizing map, orthogonal, random and manual node distribution. Some of the following types of spatial arrangement solve the problem in a more efficient manner.

*Flow-map layout* [FIGURE 3.3] is used to visualize the movement of objects between places. Sources are linked to targets by curves with variable thickness that corresponds to the amount of the flow. These maps reduce visual clutter by visually merging the curves and avoiding self-intersections (BUCHIN *ET AL*, 2011). Additionally, flow maps try to avoid covering important map features with curves to preserve recognizability of geographic locations. The wine exports graphic of Charles Minard described in SECTION 2.2 is the example of a flow map. It depicts the flow of components (army count) between various origins and destinations.



Figure 3.4. Radial convergence layout (Bostock, 2010).

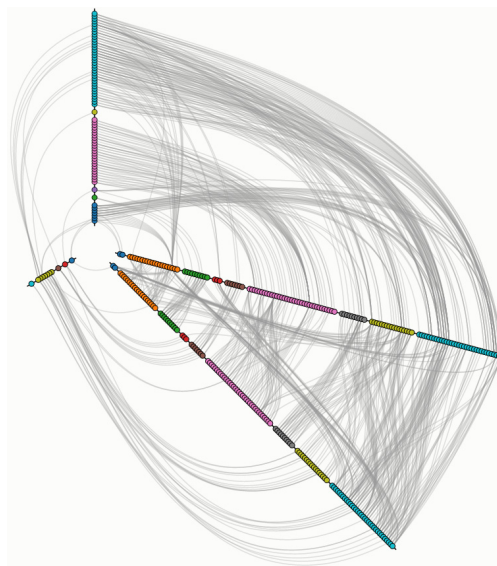


Figure 3.5. Hive plot layout (Bostock, 2010).

*Radial convergence layout* [FIGURE 3.4] shows relationships between elements of distinct typologies. These elements are distributed along a circumference and the relations are depicted by lines or curves, inside the formed circle (ARTMEDIA, 2012). The similar elements can be visually grouped to form families. The grouping allows a clearer identification of the elements of the same type and a faster distinction of relationships between different groups.

*Radial centralized layout* is similar to the radial convergence. It consists of one central point, called *host*, which attaches nodes in a circular manner. The center is linked to every surrounding node, or *satellite*, which is connected to the host only. This layout establishes relationships between the point of attention of the system and all another existing elements. On the other hand, a *decentralized layout* has various hosts, each having its own set of satellites. In this case, a satellite may be connected to one or more hosts, but not with other satellites. Generally, the communication is unidirectional within both these layouts: from the central points to the radial branches (GALLOWAY, 2004).

*Hive plot* [FIGURE 3.5] is a layout for visualization of networks. The nodes are constrained to radially distributed linear axes, accordingly to the structural properties of the network and its components. As in a radial convergence diagram, the nodes are connected by curved links. Hive plots allow a quantitative understanding of important aspects of the given network, managing the complex data and exposing network's trends and patterns. The advantage of this layout comparing to classic network representation is in the interpretation of the network. The direct visual comparison of hive plots is possible, while in the hairball layout it is a complicated task. Also, a hive plot can embed various states of a network in only one representation, regarding to the multiple axes representation. Each axis holds all nodes of the network and the multiple states are shown between these divisions (KRZYWINSKI, 2011).

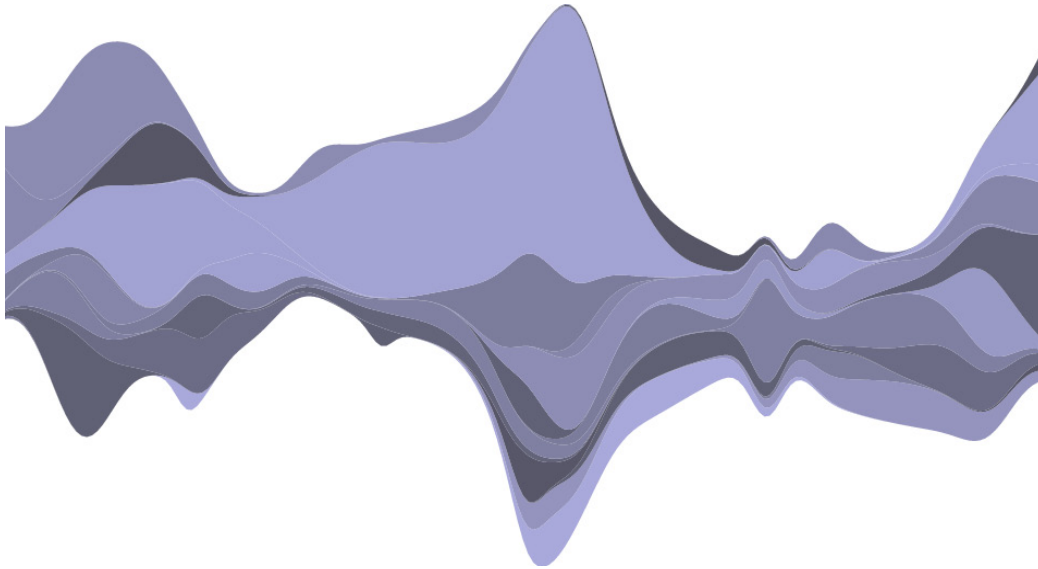


Figure 3.6. Streamgraph layout (Bostock, 2010).

*Arc diagram* is a one-dimensional planar layout using arcs to represent connections. The nodes are distributed linearly and the connections are drawn as semicircles in one or both half-planes. An arc diagram is based on the idea of visualizing all possible pairs of connections (WATTENBERG, 2002). This layout may not convey the overall structure as effectively as a two-dimensional layout, but it allows a strong node ordering, a minimization of intersections of curves, a clear identification of relationships and the displaying of multivariate data.

*Streamgraph* [FIGURE 3.6] is a layout created by Lee Byron and Martin Wattenberg (2008). It consists of a complex graph with stacked layers that result in an organic geometric shape. The horizontal axis represents a timeline. Each layer has an assigned color and represents a variable. The variation of the data related to that variable progresses from left to right through a finite time span. The height of the shape is proportional to the quantitative value of the variable, so the peaks are generated by big values. This layout tends to maximally reduce the amount of data distortion and is responsible for creating smooth asymmetrical curved forms.

*Data arrangement* refers to the principle of organization, categorization and representation of the variables in a visualization. It describes how the information is presented and can be based on the dependencies of variables, the mathematical formulas, the visual techniques and the programmatic algorithms. The data arrangement is the way the data is shown on the spatial layout. The common techniques of visual encoding and displaying the data related to flows are explained below.

*Coloring* of flows is a direct mapping of the flow attributes, such as velocity, pressure or temperature, to a color scale. The color plot techniques are widely known and this approach provides intuitive depictions and readings. The restriction is a careful choice of a color scale for mapping, respecting the perceptual differentiation. The color coding also results with temporal data, producing moving color plots accordingly to temporal changes of flow's properties (POST ET AL, 2002).

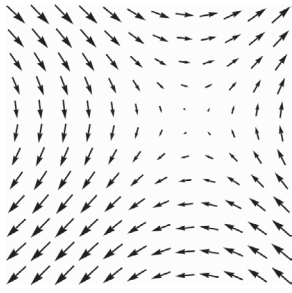


Figure 3.7. Arrow plot  
(WIKIPEDIA CONTRIBUTORS, 2004).



Figure 3.8. Streamlets  
(TURK AND BANKS, 1996).

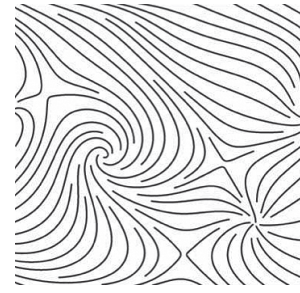


Figure 3.9. Streamlines  
(TURK AND BANKS, 1996).

*Glyph plots* is a technique based on vector visualization that represents the sample points of the flow field with markers, such as arrows [FIGURE 3.7], lines or another glyphs. Commonly, one marker is chosen to show all components or particles of the same type of flow. Two variants of glyph plots are often used: glyphs with normalized length to display the flow's direction and glyphs with varying length proportional to the velocity of the flow (POST ET AL, 2002).

*Texture-based techniques* are effective and versatile, being applied to a vast diversity of vector fields. This method uses the textures to encode vector flow fields with textures, resulting in images with a dense spatial resolution. There are various techniques of texturing of flows. *Spot noise* technique generates a texture by distributing a set of spots of intensity over the flow. *Line integral convolution* receives a vector field on a Cartesian grid and a white noise texture of the same size as an input, then the texture is locally smoothed along the streamlines to create a dense visualization of the field. *Texture advection* technique uses a motion map data structure to represent a dense flow and the information required to animate the flow (POST ET AL, 2002).

*Geometric flow technique* is based on the extraction of geometric shapes directly related to the underlying data. Various types of the geometric flow are available. *Contouring* is a natural extension of coloring technique, in which a contour is a boundary between two distinct regions that shows transitions with a color change and a drawing of an explicit transitional curve. *Isosurface* is a 3D analog of a contour that represents points within a volume of space. Integral objects are also used in geometric flow visualizations. *Streamlets* [FIGURE 3.8] are generated when flow vectors are integrated for a very short time, communicating temporal evolution. *Streamlines* [FIGURE 3.9] are utilized in longer integration, these are extension of glyph-based techniques and offer intuitive perception of the evolution of flows. A *pathline* represents the trajectory of a particle in a flow. A *streakline* is traced by a set of particles that have passed a given point in the domain, relating to continuous injection of foreign material. A *timeline* unites the positions of particles released at the same temporal instant from different points in the domain. Some of the extensions of these five techniques are *stream polygons*, *streamballs*, *streakballs*, *stream surfaces*, *stream ribbons*, *streamtubes*, *time surfaces* and *flow volumes*. Each type is used to show different combinations of particles (vectors, tensors, projections) and of their properties (acceleration, rotation, divergence, convergence) in a vector field (POST ET AL, 2002).

*Feature-based technique* is a flow visualization approach at a high level of abstraction. The data is described by features that represent the interesting information of the dataset. The huge data reduction is achieved because the percentage of extracted data amount is small and the original data is no longer needed. This enables the visualization of large sets interactively. The feature extraction procedure is the determination, quantification and description of the features in the dataset. The common examples of features are vortices (turbulence flows), shock waves (earthquakes, explosions, jet simulations) and separation and attachment lines (aerodynamic design). This technique can be divided in three approaches. The first, *image processing*, is the transformation of features in an analytic image. The second, *vector field topology*, implies a topological analysis of linear vector fields based on detection and classification of critical points. The third approach, *selective visualization*, is an analysis of physical characteristics of features (POST ET AL, 2002).

*Clustering technique* is utilized to discover natural groups in the data and to identify abstract structures that might be there, yet undefined (HAREL AND KOREN, 2001). This type of grouping can be applied to spatial data to find contiguously localized elements and group them in an abstract container (cluster). By other words, cluster is a number of similar objects collected together. A cluster can be defined by the following statements: a cluster is a set of entities which are alike, and entities from different clusters are not alike; a cluster is an aggregation of points in the space in the way that the distance between any two points in the cluster is lesser than the distance between any point in the cluster and any point out of it; clusters are connected regions of a multidimensional space with a relatively high density of points, separated from other such regions by a region with a relatively low density of points (JAIN AND DUBES, 1988).

*Edge bundling* [FIGURE 3.10] is an organizational technique based on the principle of visual merging of curves. This method can be compared to the bundling of electrical wires together along their joint paths in order to make the cable web more manageable.

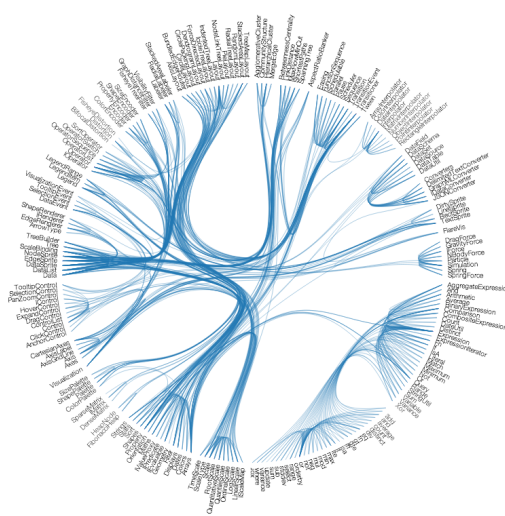


Figure 3.10. Hierarchical edge bundling (Bostock, 2010).

Edge bundling provides and intuitive and continuous form to control the strength of merging and reduces visual intersections when dealing with large amount of adjacent curves (HOLTEN, 2006).

*Swarm system* is a collection of autonomous agents that exhibit complex behavior as a group. Multiple examples of this paradigm can be found in Nature, such as ants, bees, fish and birds. The high density and chaotic motions of agents increase the complexity of swarm visualizations, but these can provide

the insight into the local (agent-level) and the global (group-level) behaviors (MINER AND KASCH, 2008). In the visualization of flows, the swarms can be used to represent streams, similarly to the geometric flow technique. Each particle follows the variation of a data variable, tracing a path or moving in space accordingly to that variation. The difference is that the swarm particles present a group behavior. Each agent in the swarm follows various principles, such as cohesion, alignment and separation (REYNOLDS, 1987). Having this, the inner-group interaction can slightly distort the data, but produces realistic movement and organic visual results.

This section summarized the layouts and techniques used in the visualization of flows. These principles can be combined in different ways, accordingly to the purpose of the graphic, the visualizing approach, the data to use and the results to achieve, to create an effective graph and reveal the insight of the flow. The next section presents a diversity of examples in the visualization of flows, referencing the layouts and techniques explained above and the semiological principles described in SUBSECTION 2.3.1.

### 3.3. Practical Approaches to Flows

The majority of visualizations of flows are created using computers, due to large amounts of data and the complex nature of the flow phenomenon. Each approach adopts a particular layout and one or more representation techniques to visualize the data in its original context. This section presents the work related to practical examples of visualization of flows, referencing several techniques and layouts from the previous section and showing some contemporary approaches to flows.

FIGURE 3.11 shows a snapshot of an animated *simulation* of a turbulent flow in the field of fluid dynamics. The application, implemented by Wim de Leeuw and Robert van Liere, maps numerical data directly to the screen to create the flow of particles. The flow goes from left to right and collides with a block placed in the field (the round mark on the left). Behind the block, the flow transits from laminar to turbulent, creating vortices and helicity. The animation uses *textures* and *spot noise* to represent the stream of *particles*, mapping the *pressure* variation accordingly to a *color scale*: blue shows low pressure, green shows zero pressure and red color shows high pressure in the flow (DE LEEUW AND VAN LIERE, 1997).

FIGURE 3.12 displays a *vector field* of wind over the *map* of Australia, calculated using a numerical weather model. The graph was crafted by Greg Turk and David Banks. Chains of *arrows* indicate wind's *streamlines*. The wind's *direction* is shown with the arrows' *orientation* and the *magnitude* is represented by their *size* and *opacity*. High wind's velocity results in larger and opaquer arrows, while weak wind streams are shown with smaller and more transparent streamlines. The *associative property* of arrows creates a coherent context of all represented data. The *glyphs* are aligned head-to-tail to allow a continuous visual following (TURK AND BANKS, 1996).



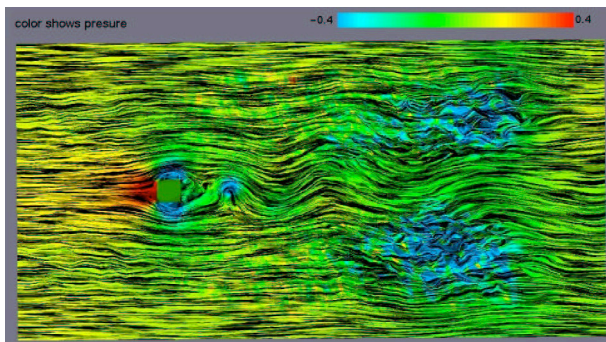


Figure 3.11. Direct simulation of a turbulent flow, by W. de Leeuw and R. van Liere, 1997.

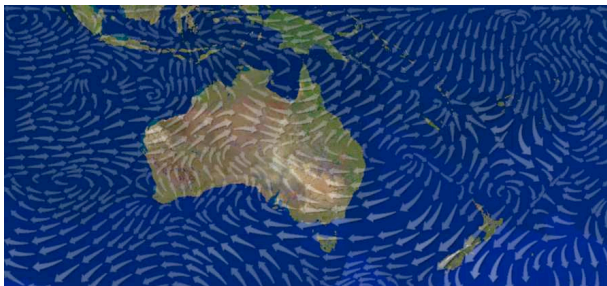


Figure 3.12. Wind direction and magnitude over Australia, by G. Turk and D. Banks, 1996.

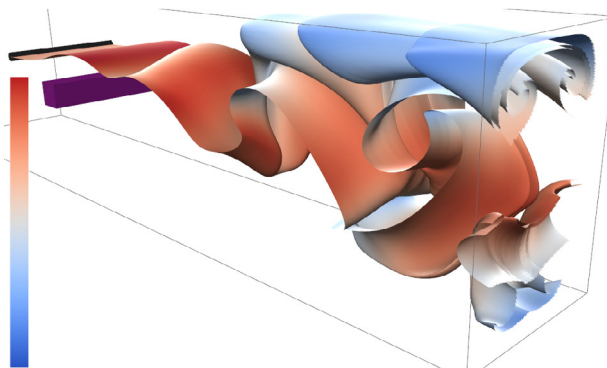


Figure 3.13. A streak surface flow past a cuboid, by T. McLoughlin et al, 2010.

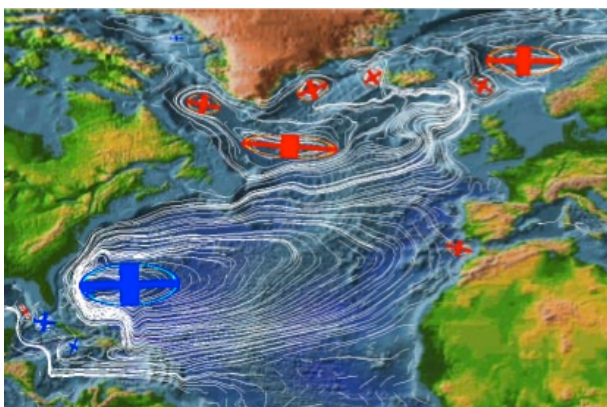


Figure 3.14. Currents in the North-Atlantic Ocean, by A. Sadarjoen and F. Post, 2000.

Tony McLoughlin, Robert Laramee and Eugene Zhang created a *streak surface* showing a *geometric flow* past a cuboid is depicted on FIGURE 3.13. Streak surfaces are mathematical models based on curves that demonstrate the evolution of the flow over time. This technique is misused due to computational complexity of generating dynamic meshes, despite the advantages. A snapshot from the animated visualization reveals complex regions and structures. The *velocity* of the surface is *colored* accordingly to its magnitude, with blue fragments representing low velocity and red parts showing high velocity flow (McLOUGHLIN ET AL, 2010).

The *feature-based* visualization [FIGURE 3.14] of the water currents in the North-Atlantic Ocean was created by Ari Sadarjoen and Frits Post. The graphic presents a technique of vortex detection, mapped on a *cartographic layout*. *Streamlines*, shown with white curves, are *clustered* accordingly to the common center of density. The *features* of the visualization are quantified ocean's *vortices*, represented with blue and red crosses. These artifacts were noticed after revealing the groups of currents with the same density centers. The vortices rotate in clockwise and counterclockwise directions, respectively to colors (SADARJOEN AND POST, 2000).

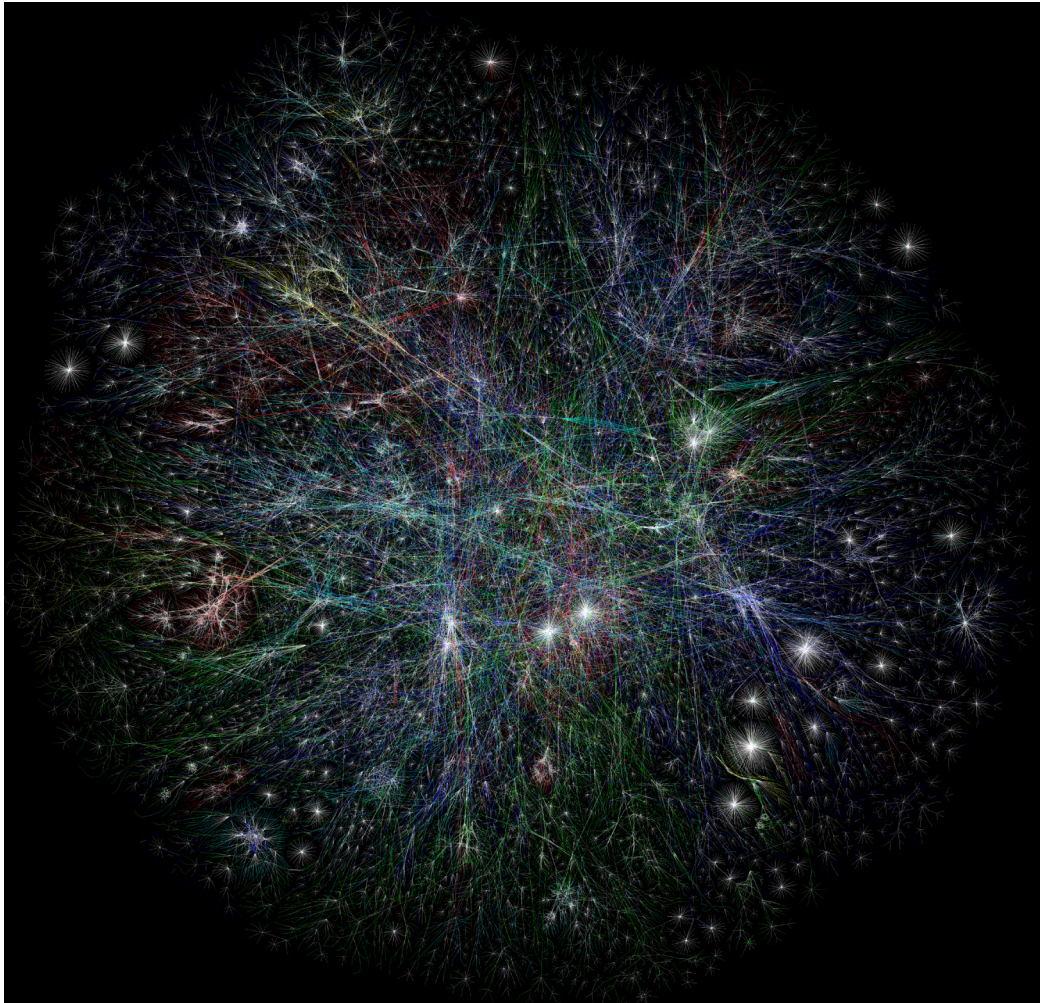


Figure 3.15. "The Opte Project," by Barrett Lyon, 2003.

The graph on FIGURE 3.15, created by Barret Lyon, represents a *map* model of the *Internet*. It is based on a *traceroute* technique that analyzes every sequence of the network and draws a *relative path*. Each path has an assigned *color* that represents the *domain* it belongs to: *Asia Pacific* is red, *Europe, Middle East, Central Asia and Africa* are green, *North America* is blue, *Latin America and Caribbean* are yellow, *RFC1918 IP addresses* are cyan, and unknown domains are white. Over 5,000,000 edges were used to create the whole visualization with an estimated 50,000,000 hop count (THE OPTE PROJECT, 2003). There can be identified major green, blue and red clusters of servers and many punctual white-colored stars of unknown public networks.

The "History Flow" visualization, developed by Fernanda Viégas, Martin Wattenberg and Kushal Dave in 2004, aims on understanding of the patterns of cooperation and conflict between authors of *Wikipedia*. Any page of the online encyclopedia can be introduced as input to demonstrate the flow of all contributions and editions of the article. The flow is compound by bending lines of different colors on the black background. The lines are mapped horizontally on a timeline, from the left to the right. Each author has an assigned color and the height of the lines of the same color represent his contribution to the article. Therefore, the total height of colored lines represents the length of the text in the article accordingly

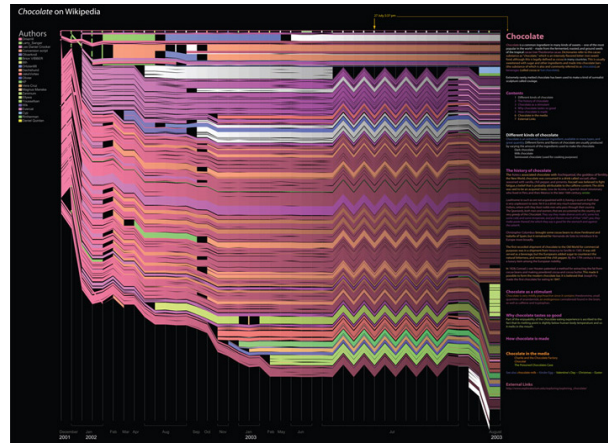


Figure 3.16. The history flow of “Chocolate,” by F. Viégas, M. Wattenberg and K. Dave, 2004.

to the timeline. The bend points of the lines mainly occur when the modifications in the article are made. The graph related to the “chocolate” page is shown on FIGURE 3.16. The pink author have written the article in the beginning, then the violet has extended the text. Other authors (green, white and blue) contributed with new information and modified the existing text. The zig-zag patten on the right represents an “edit war” when all authors tried to modify the information accordingly to their point of view (VIÉGAS *ET AL*, 2004).

“Flight Patterns” is an animated visualization of Aaron Koblin that displays more than 200,000 aircraft flying over the United States on August 12, 2008. Each *flight* is represented by a *curve* corresponding to the path of the aircraft, accordingly to the invisible *cartographic layout*. The *coloring* represents different *types of flight*. Starting in the morning, the country is illuminated with fly paths. By evening, the lines are fewer and the nocturnal flights from the East Coast to Europe become a major activity. In the static visualization [FIGURE 3.17], all flights are mapped at once. The whole image reveals the map of the United States and the overlapping semi-transparent paths trace the common routes. The author made several variations of the piece, highlighting airplane types, altitudes, regions and airline hubs. The viewer cannot interact with the application, what limits the study and the exploration of the artifact (FLOW, JUST FLOW, 2013).



Figure 3.17. “Flight Patterns,” by Aaron Koblin, 2008.

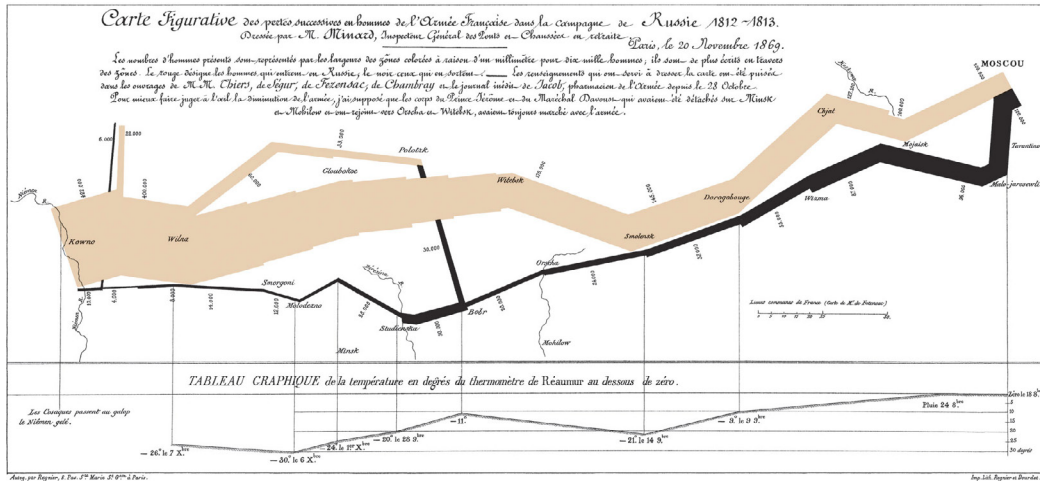


Figure 3.18. “Figurative Map of Successive Loss in Men of the French Army in the Russian Campaign in 1812-1813,” by Charles Joseph Minard, 1869.

The *flow map* on FIGURE 3.18, produced by Charles Minard, is not a recent creation, but it claims to be the best statistical graph ever made, accordingly to Edward Tufte (2001, p. 41). It tells the story about the French war march on Moscow in 1812-1813, depicting the loss of life in Napoleon’s Grand Army in the advance from Kowno to Moscow (yellow flow) and also in the retreat (black flow). The *army’s size*, initially of 422,000 men, is represented by the *line* with a decreasing *thickness*. The thickness’ variation clearly shows that the army was constantly losing forces. At a half-way to Moscow, Napoleon has lost more than a half of his troops. The whole campaign, overlaid on a real, yet not visible *map*, ends with 10,000 survivors that returned home. A *line chart* of the temperature is displayed below the graph, demonstrating the conditions of the Russian winter which accompanied the soldiers on their withdrawal. This graphic seems to defy the words of the historian by its eloquence (MAREY, 1885).

FIGURE 3.19 demonstrates a cyclic approach to flows. This visualization was developed by Fernanda Viégas and Martin Wattenberg in order to represent the world

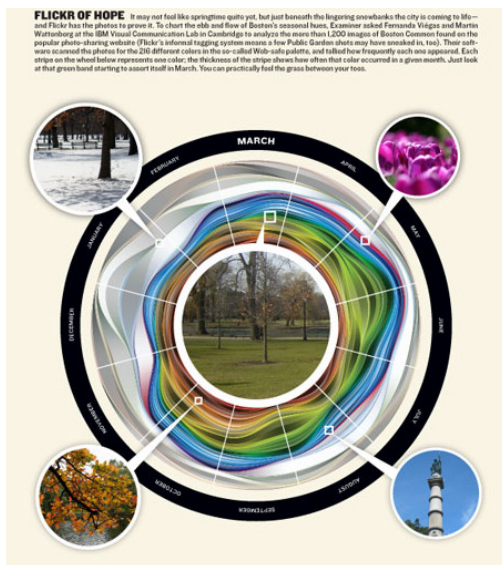


Figure 3.19. “Flicker Flow,” by F. Viégas and M. Wattenberg, 2009.

as a stream of color. The graph analyzes a collection of *photographs* taken from *Flickr*, sorted by months of the year. The relative *proportions of colors* existing in the photos are calculated and plotted on a *circular layout*. The result is a diagram of ebb and flow of common seasonal colors (HINT.FM, 2009). This work appeared in the *Boston Magazine*, with several *contextual* modifications. It includes month labels and grid ticks to provide a visual distinction of each month’s color palette. An original photograph of each season is also shown to compare the images with the corresponding color flow.

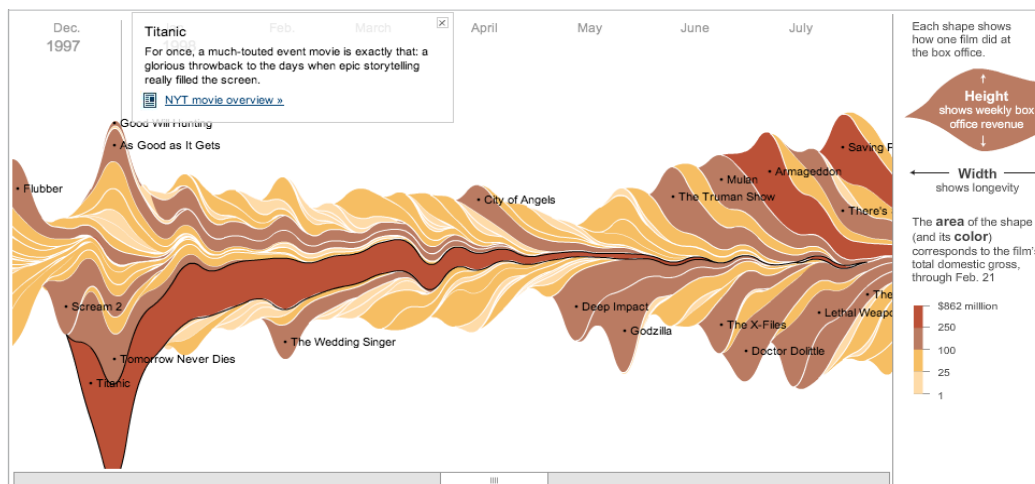


Figure 3.20. “The Ebb and Flow of Movies: Box Office Receipts 1986–2008,” by Lee Byron, 2008.

FIGURE 3.20 is a snapshot of an interactive *streamgraph* showing the receipts of movies. The horizontal axis is temporal and each *layer*, or stripe, represents a time series of a *movie*. The *height* of a stripe corresponds to the weekly box office *revenue* of the movie. The *area* and the *color* are determined by the total *domestic revenue* in USD. The shape’s *tail length* shows the movie’s temporal *longevity*. The color scale is situated at the right side of the application, along with the explanation of visual variables. The interactive graph uses the dataset which covers over 20 years of box office revenue. It reveals the repetitive pattern of the movie seasons and the change in blockbuster trends. Pointing with the mouse cursor at a stripe highlights it with a black stroke and shows its title (*direct selection*). Clicking on a layer shows an informational message about the chosen movie (*details-on-demand*), providing a link to the *NYTimes* online movie database. The user can also *navigate* through the timeline horizontally using the *scrolling bar* at the bottom of the application or the *mouse dragging* functionality (BYRON, 2008).

The *radial convergence diagram* shown on FIGURE 3.21 (the next page) is a work of Moritz Stefaner. This graph provides an overview of the whole citation network in Science. The represented dataset was extracted from *Thomson Reuters’ Journal Citation Reports 1997–2005*, with journals sorted by their *Article Influence* in 2005 and the journal count limited to 25 for each science field. The selected journals state in all seven years of reports and the top 10 journals with the highest score are included, that makes the subset coherent. The *colored* nodes, distributed along a circle, represent four main groups of journals. There are two layers: the individual journals (the inner ring) *clustered* by fields of Science (the outer ring). The nodes are scaled by the *Eigenfactor* knowledge ranking system. The general view presents the top 1,000 citation *links* shown with *curves*. The *thickness* and the *opacity* represent the *strength* of connections. The curves are grouped with the *edge bundling technique*. This graph also provides interactive possibilities. Pointing a cursor over a node shows its designation (science field or journal name) and the ranking score (*details-on-demand*). Selecting a single journal or a whole field highlights all citation flow coming to or from the selection (*direct selection*). The color of inner nodes is based on the color of outer clusters with the variation of *hue* accordingly to the *score level* (STEFANER, 2005).

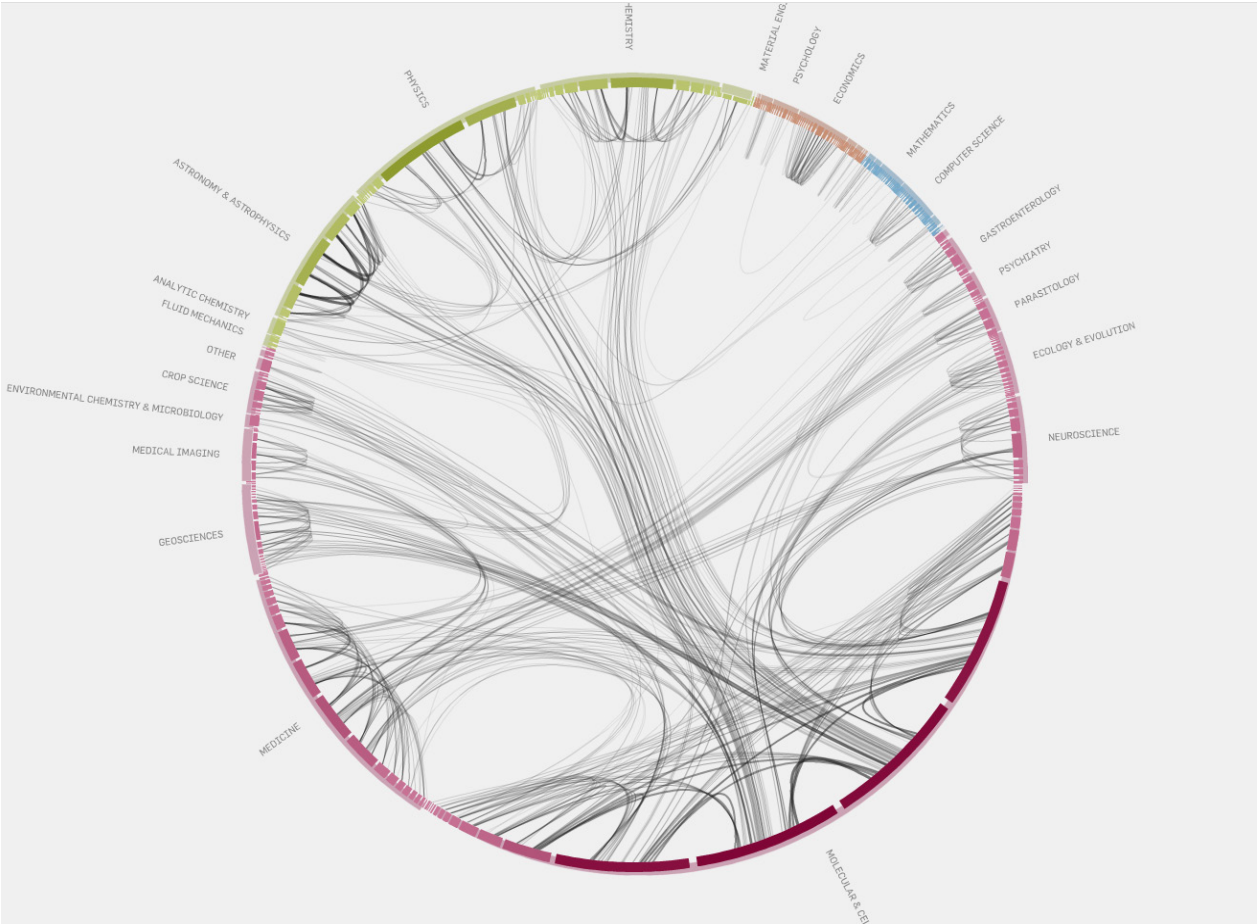


Figure 3.21. "Visualizing Information Flow in Science – Citation Patterns," by Moritz Stefaner, 2005.

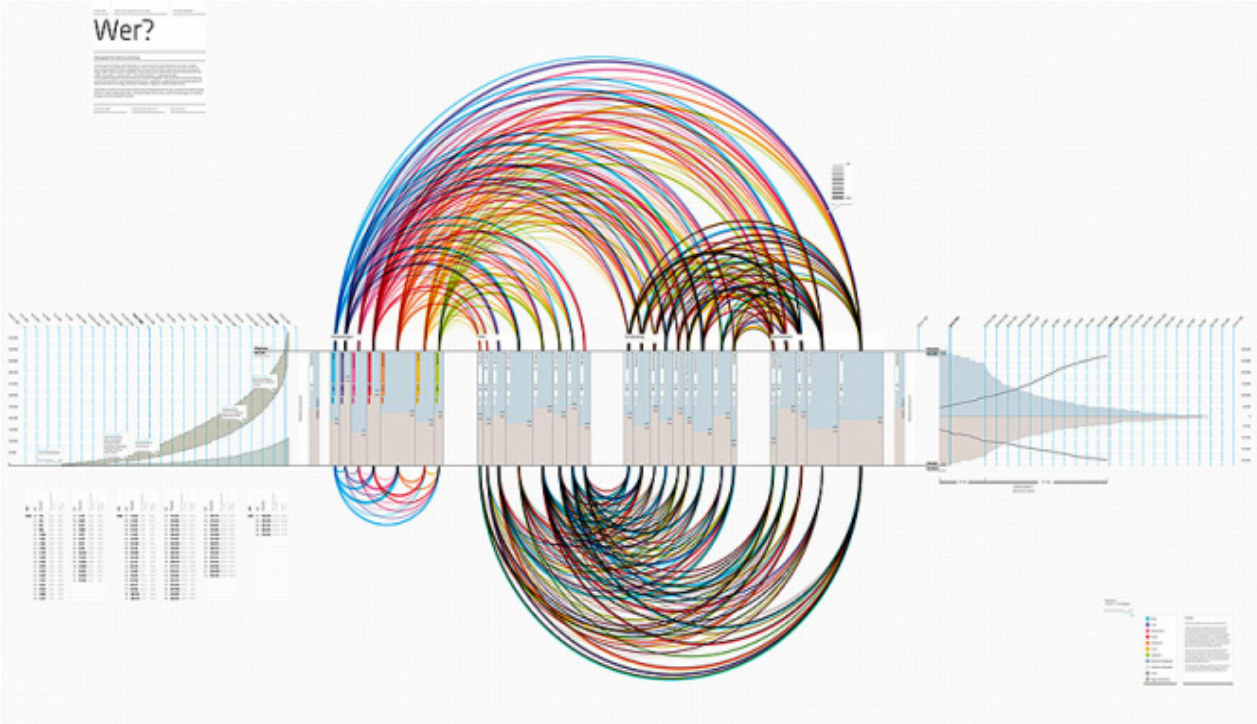


Figure 3.22. Visualization of a social network, by Felix Heinen, 2007.

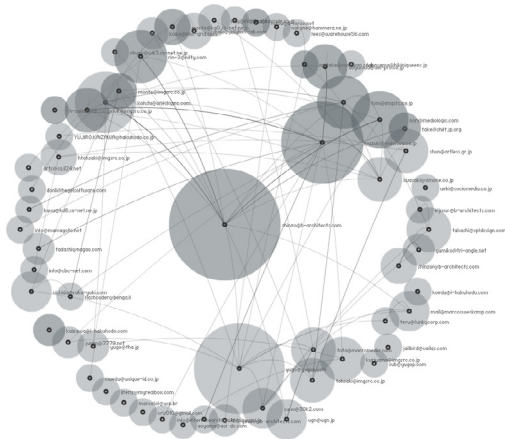


Figure 3.23. "Social Circles," by Marcos Weskamp, 2003.

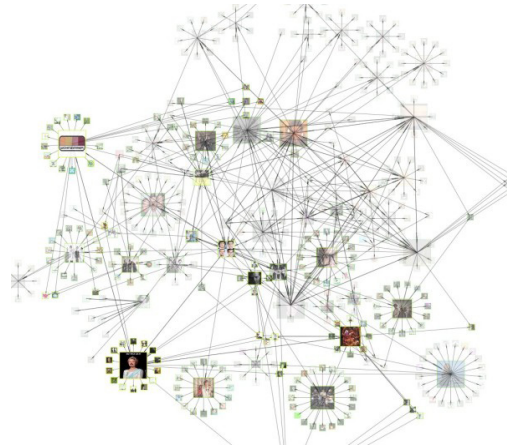


Figure 3.24. "Comment Flow," by D. Offenhuber and J. Donath, 2007.

FIGURE 3.22 is a complex graph, created by Felix Heinen, showing the variety and attitudes of *Facebook* members. It depicts the dependencies between demographic information, such as website functions used by people, age, gender, educational background and family status. The *numerical quantities* of that data is shown on the *normalized bar chart* that divides the arc diagram in two planes. Each *component* shows *percentage amounts* of the given demographic aspect for *male* and *female* users. On the right side of the diagram, a *scatterplot* shows the amount of registered German users. The numerical *statistics* are shown below the plot. On the left side, the log-in quote and frequency are shown with a *bar chart*. The *arc diagram* represents relationships between the components of the graph. The website's add-ons are *colored* and the arcs the link to other fields inherit the color of the respective function. The black curves refer to the fields not related to add-ons (HEINEN, 2007).

The purpose of the visualization "Social Circles" [FIGURE 3.23] is to reveal the social networks that emerge in mailing lists. The author Marcos Weskamp intended to display the hierarchies and the main subjects of conversation. Based on the *radial centralized layout*, the graph provides an overview of the whole network. The *node circles* represent the *center* of the group, the *participants* and the *topics* the people are talking about. Each component is accompanied by a textual description. The *size* of a circle depends on the relational *ratio* to the central subject. The *curved links* show the *relationships* between components (WESKAMP, 2003).

The *force-directed* graph called "Comment Flow" is the creation of Dietmar Offenhuber and Judith Donath [FIGURE 3.24]. It is a tool for a content-driven exploration and visualization of social networks. The layout consists of *nodes* that represent *user profiles*, and *edges* that represent *friendships*. The system displays the *activity* and the exchange of commentaries between people, depending on the *sequence* and *age* of messages. This approach emphasizes the visualization of individualized network, avoiding the representations of network-only structures. There are commercial, celebrity, spam and fake profiles that produce immense connections with insignificant amount of information about the actual social bonds and behaviors of the given person. The authors go beyond the connectivity analysis and look at the flow of information between the individual actors (SOMMERER ET AL, 2009).

The visualizations presented in this section make an overview of both classic and modern techniques of representation of networks and flows. The first of them started with the advent of computer graphics, the recent ones are contemporary approaches of effective presentation of scientific information. These methods were appearing accordingly to the demand and supply of the era of information and technology. It is impossible to include all existing approaches to visualization of flows and networks in this dissertation. The selected principles are remarkable and powerful foundations of graphic representation, keeping their meritorious place in the state of the Art of scientific visualization.

### 3.4. Summary

None of the existing techniques of visualization is flawless, so the visual forms and layouts will continue to emerge. This tendency is caused by specific requirements of information to be communicated, optimization of the classic methods and novel contribution to the Science and information design. Many of the techniques, layouts and graphs described in this chapter served as a source of inspiration for the visualizations developed during this research. As researchers, we intend to contribute with our approaches to visualization of the specific information, which display the patterns of flow, the patterns of usage, and the relative physical structures of wireless networks. The next chapter presents the objectives of this research and the methodology of achieving the goals.







## CHAPTER 4.

# Research Goals and Methodology

The visualizations of patterns of flow and usage, developed in the context of this dissertation, are representations of network data with four different approaches. The datasets contain information about user connections between the access points of the eduroam networks of four portuguese universities. The developed methodology is directed to accomplish the main goals of this research accordingly to the particularities of the available resources. The objectives and respective implementation techniques are explained in this chapter.

### 4.1. Objectives

The project that affiliates this research aims on two goals: the exploration of the *Wi-Fi* technology as a proxy to capture data about the human spatial behavior, and the use of the gathered data for detection of human motion patterns and its correlation with the characteristics of the space. The first objective concerns with data mining, being the University of Minho is responsible for that part. The second objective implies the construction of visualization models, and this is the main objective of this study.

In the initial phase of the work, the tasks and the deliverables to be produced are discussed. The milestones are established and the work methods are determined. To trace the boundaries of the area of the research, the objectives are formulated as explained below.

*Analyze and understand the available data* to define which aspects of the network can be represented. In this research, we work with the structure of networks and the movement inside these networks. The movement refers to the connectivity of users, which can be shown through general or particular connectivity and the patterns of movement. The data parsing phases are minutely described in SECTION 5.1.

*Visualize the physical structure of the wireless networks* to provide a full and clear vision of the scope. The wireless network of a university is a complex structure with a big quantity of devices involved. The complete system cannot be seen with the naked eye because the network's campus covers a considerable geographic area. This objective concerns with the creation of tools that represent the given network as a whole.

*Visualize the overall and particular connectivity of network's users* to understand the usage factor of the given network by real people. This objective concerns with the development of tools that show when and where the people connect to the given networks.

*Visualize the movement on the wireless networks* to analyze the usage of these networks. Flows are directed streams of data that run within a system. In the case of this study, the systems are wireless networks, the access points stand for static data, the eduroam users represent dynamic data, and the user connections between access points are streams of information. This objective concerns with the construction of tools that display the flow in the given networks.

*Analyze the obtained results* to reveal the patterns of flow and usage in wireless networks of four Portuguese universities and draw respective conclusions.

These four goals merge into one big objective: *the elaboration of proper visualization techniques for each case of study*. The specific visual methods are needed to make the flow perceptible, interpretable and understandable. As was demonstrated in CHAPTERS 2 AND 3, each approach is based on the context and on the data. We cannot simply use existing visualization tools to depict the wireless network data. These software applications can have limited data mapping and interactive functionalities, the visual appearance of the graphic elements is often fixed and non-customizable, and the results can be generalist, imprecise and fallacious. In some cases, the generic software can be utilized to preview the data and make visual sketches. Our experiments with the application *Gephi* (2008) can be found in the APPENDIXES section [PP. 5-9]. We prefer to create our own applications with the appearance, parameterization and interactive functions that we consider effective and relevant to use. In this research, we develop visual techniques inspired by relevant examples from the state of the Art of scientific visualization, implementing the graphs from the ground up.

Once the construction of a visualization is complete, the graphic results are analyzed to show their legitimacy, to indicate the effectiveness of the visual techniques and to make conclusions about the processed information. This is a usual task of data analysts. Considering that there is no people of such profession in our research team, we will accomplish the analysis of the experimental results by ourselves.

## 4.2. Modus Operandi

The scientific nature of the available datasets and of the visualization of flows itself determine certain work methods. Some of them are present in any visualization project development, yet there are proper stages to this specific case. The process of research and implementation of experimental visualizations is divided into four sequential phases. Our methodology is an adapted version of the system originally proposed by Ben Fry (2008, pp. 5-6).

*Data gathering, analysis and synthesis* is performed in order to begin the work on visualization of patterns of flow and usage in wireless networks. The analysis and treatment of the datasets is an important task. The information to be represented needs to be understood, restructured and filtered accordingly to the defined goals. The relationships between different types of information existing in a set are identified. Due to the imperfection of recording devices, the errors can be found and, if possible, corrected. After that, the data records are chosen to be visually treated.

*Visual representation of the information* consists of the selection of a visual layout that shall hold the components and the invariant of the graph. Firstly, the graphic language of a visualization is defined, depending on association of the selected data variables to visual components. This language needs to be simple (what is usually obtained with geometric abstraction) and perceptible (to represent data according to its common semantic meaning, e.g., places as circles, directions as arrows). All data needs to have a graphic element associated. This provides the readability of the obtained result and the ability to create visual connections and dependencies between different variables. The choice of language will also depend on the graphic artifact to produce. Secondly, a programmatic implementation is used to link data with objects in a graphic layout. The aftermath of this procedure is a visual environment with all elements, connections and dependencies included in a computer application. That result is called visualization or graph and permits to draw conclusions, based on its reading. These software models allow the analysis of a network with a large amount of circulating information.

*Refinement of parameters* is a further configuration of variables after the first implementation. It aims on the improvement of the graphic layout to make it more understandable, more readable and balanced, accentuating determined points and details. A correct parameterization, as well as different types of parameterizations, can reveal hidden meanings and various ways of reading.

*Interaction* is added to enable the control of the application. Interaction is an important feature in the scientific visualization, as well as in the general software development. The user interface provides the exploration of the data and the application itself, the customization of the layout and the data manipulation.

The order of these steps may change during the construction of a graph. The phases may entwine accordingly to the requirements of an experimental process, some tasks may be ignored if already performed and other work steps may be added if necessary. Usually, there are professionals that assure the quality of the information regarding the first phase. In this research, we will perform the necessary tasks of the data treatment by ourselves. Also, tasks related with information processing and analysis, such as pre-processing of data, clustering, estimation of distances and identification of places, are not our responsibility. Nevertheless, due to the unavailability of the results of such processes and of the necessary source code, we are forced to assume these tasks in order create the conditions for the development of our visualizations.

### 4.3. Summary

This chapter presented the research goals and the methodology used in this work. The ways that objectives relate to work methods are explained and the different phases of the creation of the visualizations is clarified. The next chapter describes the work performed within this dissertation, describing in detail the data treatment procedures and the implementation of the visualization prototypes.







## CHAPTER 5.

# Preparatory Work

This chapter describes the preliminary work phases of this research. The first section reports on the data we worked with and the pre-processing stages performed in order to create clean and consistent data files. The datasets contain records of user accesses to various routers of the eduroam network of the universities of Portugal. The information is analyzed and the aspects to be graphically displayed are extracted. That extraction implies a computational treatment of the raw records. It is explained step by step along with the purpose of the processed data in each phase. The second section consists of the development of the experimental graphic layouts to understand what types of spatial arrangement and impositions are suitable to the representation of the wireless networks and the respective inner flow.

### 5.1. Interpretation of the Data

This research fits into the context of the scientific research grant under the project *SUM – Sensing and Understanding human Motion dynamics* of the *Portuguese Foundation for Science and Technology*. The organization accountable for the data mining was the University of Minho (UM). We started testing visualization models with dummy data. Those dummy records included the information about user connections within the access points of the UM campus. The temporal period covered 17 hours of the network’s usage, so we could not make any reasonable conclusions about the flow. After several months of gathering the actual data from the UM, we started its treatment to continue the work on visualization models.

The data records include the information about user connections within access points (APs) of the wireless network campuses of four portuguese universities. That raw data was collected by the network servers of each university and aggregated in the data log format. *Computer data logging* is a registration of the activity of

the network as a system in order to understand its activity and to diagnose problems. The provided logs may contain errors of registry due to the technical problems of the network's equipment. These errors should be treated by the entities concerned with the quality of the data, but there was not any. We perform the correction by ourselves to keep the information consistent to produce valid results.

The *data treatment* is a slow procedure that must be done before the further construction of visualizations. This work phase is subdivided in 18 steps, applied to each university's dataset. Each phase uses one or more scripts that we implemented in the *Processing* programming language to parse and manage the given data. The list below describes the data treatment process step by step. All these phases are executed in the order of appearance.

### Step 1. Download files

First, we download all the provided datasets from the SQL server of the University of Minho and save them in our workstation in separate plain-text files. The record amount in seven sets is approximately 137,000,000 with a total size of 14 gigabytes.

### Step 2. Clean fields

A review of the datasets shows that three of them carry inconsistent information (mixed source networks, different entry formatting for the same fields, empty values). We discard these datasets and retain the remaining four, that include records from the University of Coimbra (UC), the University of Minho (UM), the University of Lisbon (UL) and the University of Porto (UP) networks, covering variable periods from 2010 to 2012. This data documents the connectivity of users within the wireless network campus of the respective university. The fields contain mainly technical information. After analyzing the content of the logs, we realize that the majority of the fields are empty or useless for the visualization of flows, such as the connecting port numbers, the internal network's equipment identifiers, the tunnel types and the operator names. We keep only the four fields regarding the user connections between the APs. These are:

Access point's IP address – the static IP address of the AP located within the network's range

User name/user IP address – the unique string identifier of the user that connected to the AP

Timestamp of the stop event – the temporal instant registered when the user's connection ended (the Unix timestamp format)

Session time – the temporal interval showing the amount of time that the user was connected to the AP (in seconds)

Each connection has an associated state of activity, with three possible types: start, alive and stop. The *start* event is assigned to the connection when the user connects to an AP, the *alive* event is registered periodically while the connection runs, and the *stop* event occurs when the connection ends. Each activity is registered together with the respective *timestamp*. The timestamp is a sequence of numbers that encodes date and time of an event. To make the datasets smaller and to save the computational time, we use the *stop* events only, being able to calculate the *start* using the formula:

$$\text{start\_timestamp} = \text{stop\_timestamp} - \text{session\_time}$$

In this phase of the data treatment, we face some particular situations to resolve. The UC dataset does not contain the session time field, so we are forced to extract it by ourselves by calculating a difference between the *start* and *stop* events. A series of these events were registered out of order (the habitual order is alternate, the *start* is first, then goes the *stop*), so an unexpected ordering treatment is required to prevent the creation of invalid connections. In the UL dataset, we replace the user IP address by the user login credentials, because the field of IP address has a lot of null values. Yet, each data type is valid to unequivocally identify a user.

### Step 3. Separate files by academic year

We resolve to subdivide the datasets in *academic years* instead of regular years due to the academic context of the data. In Portugal, an academic year begins in mid-September, depending on each university’s scholar calendar, and ends in August when the exams period is over. Our academic year separation starts from the very first connection registered in September and ends with the last connection of August, for every year. This procedure allows us to create a stronger context around the data and to reduce the amount of information for the next processing steps. Taking into account that some sets contain incomplete regular year periods due to the lack of data registry, we skip several “holes” in the datasets. Also, we discard the separate files with a period lesser than six months to make a broader temporal range analysis. The resultant amounts of connections are listed below.

	academic year	# of connections	covered period (DD.MM.YYYY)
UC dataset	2010-2011	5,857,776	12 months (01.09.2010-31.08.2011)
	2011-2012	6,335,611	9 months (01.09.2011-31.05.2012)
UL dataset	2011-2012	3,844,086	12 months (01.09.2011-31.08.2012)
UM dataset	2010-2011	9,730,241	12 months (01.09.2010-31.08.2011)
	2011-2012	11,947,523	10 months (01.09.2011-30.06.2012)
UP dataset	2010-2011	2,001,726	9 months (01.09.2010-01.06.2011)
Total		39,716,963	

### Step 4. Remove null values

We notice that the datasets contain a considerable amount of *null values* in all four fields. These values spoil the final results of the data treatment process: the null AP identifiers create connections with no origin, the connections with null user identifiers belong to no one and cannot be categorized, the null timestamps or session times cannot be localized on a timeline. We solve this issue with the filtering of the records based on the removal all existent null values. This procedure does not eliminate many records, so it preserves the data consistency and prevents the errors in the next processing steps and in the posterior data mapping.

### Step 5. Collect all users

This phase is dedicated to the extraction of the *users* existing in each university's network. The procedure is applied to all records in the given dataset, extracting the entries from the second field (user name/IP address). Then, the duplicated entries are removed. The amounts of active users in the networks are:

UC dataset	23,609	UM dataset	35,914
UL dataset	25,846	UP dataset	24,090

These quantities suggest the diversity of connectivity in each network. Three sets have similar numbers of users, except the UM set which contains a bigger number of users. Without a visualization, we cannot say anything about the connectivity, because it is a subjective matter. One network can have many different users with a little number of connections, and another network can reveal a small amount of users with a high connection periodicity.

### Step 6. Collect all APs

Similarly to the previous step, we extract the unique *APs* from each dataset to obtain lists of the active connection points. The lists have the following numbers of entries:

UC dataset	433	UM dataset	4,162
UL dataset	3,576	UP dataset	1,714

The numbers indicate two things: the physical scope of each network's campus and the concentration of the routers in each university. We do not possess this type of information (for example, the geographic location of each AP) to foresee any possible problem that can surge during the conception of the visualizations. The obtained user and AP lists will be used in visualizations as indexes.

### Step 7. Index users and APs

In this step, the *indexing* of the users and APs is performed to speed up the file reading and processing time. We create indexes using the lists obtained in previous two steps by adding a numeric identifier to each row. For example, the UC user index file has the user name field and a unique ordinal number field, from 1 to 23,609. Then, the user names/AP names were substituted by the respective identifier in each academic year file. The indexing also saves a considerable amount of disk space, since the index ID is shorter than the associated name or address.

### Step 8. Order connections by user ID and timestamp

We *sort* the connections in each academic year by timestamp, grouping the data by user index. This procedure creates the academic year files with a new structure: the connections of each user are consecutively ordered accordingly to the instant in which occurred, and all user IDs are in ascending order, starting with a smaller numbers. The structure allows the identification of the amount of connections of each user and the temporal density of his performed connections. These files will be used in the visualization of user connectivity.

### Step 9. Remove invalid connections

The file ordering completed in the previous step reveals the *temporal overlapping* of connections of a given user. There are two situations that invalidate connections: a connection was registered while the previous connection was still running and the same connection was registered several times within a small temporal interval. That could possibly happen due to an error of registry in the network's servers. We search for overlapping connections in the datasets and remove them. The information cleaning eliminates 53% of the data, what means that the servers do not work properly, registering wrong log files. From now on, we work only with the valid data, with lesser connections though. In the end of this procedure, the quantities of connections are reduced as numbers show below. The covered periods remain unchanged.

	academic year	# of connections	% of removed data
UC dataset	2010-2011	2,833,255	51.6
	2011-2012	3,130,781	50.6
UL dataset	2011-2012	2,055,214	46.5
UM dataset	2010-2011	4,227,654	56.6
	2011-2012	5,135,001	57
UP dataset	2010-2011	1,242,783	38
Total		18,624,658	53.1

### Steps 10-11. Recollect users and APs

Having smaller datasets, we also need to clean up the index files. The APs without connections are removed accordingly to the content of the new files, reducing the total quantity of active network routers. The user lists for four universities are maintained, because no user is deleted after the verification of files.

UC dataset	431 APs	UM dataset	3,885 APs
UL dataset	3,173 APs	UP dataset	1,624 APs

### Step 12. Create transitions

The connections' information can be used to represent the network's usage and the connectivity of users. Yet, it does not represent movement. In order to fulfill the main goal of the research – to visualize the flow on a network – we analyze the data and discover that we are able to extract the information about the transitions of users between the APs. We define a *transition* as movement of a user from one AP to another during a certain period of time. To obtain this information, we merge two successive connections X and Y of a given user into one transition belonging to the same user. That transition is not the real spatial movement of a user between two APs, but a relative movement. Usually, a transition of this kind is registered when the user shuts down or puts to sleep his mobile device (when a disconnection is registered) and connects to the network from another location. The other option of the

transition's registry is when the user relocates his active mobile device being connected to the network (in this case the farther he moves from the currently connected AP, the greater is the possibility of connection to another AP). This happens less frequently, because the university's users go online on class times or on intervals between classes. Each transition has the following fields: the AP from which the user started the transition (of the connection X), the AP to which the user transitioned (of the connection Y), the timestamp of the beginning of the transition (of the transition X, since it is the instant the user disconnected from the first AP), the transition time calculated with the formula

$$\text{transition\_time} = \text{stop\_timestamp\_connection\_Y} - \text{session\_time\_connection\_Y} - \text{stop\_timestamp\_connection\_X}$$

and the ID of the user that performed the transition. The amounts of extracted transitions are:

	academic year	# of transitions	covered period (DD.MM.YYYY)
UC dataset	2010-2011	2,819,264	12 months (01.09.2010-31.08.2011)
	2011-2012	3,114,470	9 months (01.09.2011-31.05.2012)
UL dataset	2011-2012	2,029,368	12 months (01.09.2011-31.08.2012)
UM dataset	2010-2011	4,207,973	12 months (01.09.2010-31.08.2011)
	2011-2012	5,108,867	10 months (01.09.2011-30.06.2012)
UP dataset	2010-2011	1,218,693	9 months (01.09.2010-01.06.2011)
Total		18,498,635	

It is normal that the number of transitions is smaller than the amount of connections, in the same academic year. As we merge two consecutive connections of a user to create a transition, some connections are not included for two reasons: the last connection of a user does not produce a transition and there are users that perform only one connection during the whole year. The data related to user transitions will be utilized to represent the movement flow of users within the networks and the usage of the given network as a system.

### Step 13. Remove invalid transitions

After creating the transitions, we realize that some of them contain *negative transition times*. Even having eliminated the overlapped connections, there is always a probability of existing another types of unknown registry errors. We discard the transitions with negative time values, because any spatial movement requires at least a moment to be performed. Even if the quantity of deleted records is low, we can observe the persistence of errors in the dataset. We prefer to eliminate these erroneous records to prevent the future inconsistency and errors in the visualizations. The resultant amounts of transitions are shown at the top of the next page.

	academic year	# of transitions	% of removed data
UC dataset	2010-2011	2,770,912	1.7
	2011-2012	3,035,056	2.5
UL dataset	2011-2012	1,931,710	4.8
UM dataset	2010-2011	4,147,777	1.5
	2011-2012	5,033,195	1.5
UP dataset	2010-2011	1,183,372	2.9
Total		18,102,022	2.2

The data cleaning process is over. Now we have the processed sets with connections and transitions, separated by academic year in each university's dataset.

#### Step 14. Order connections by timestamp

Having the connection files grouped by user and ordered by timestamp inside each group [STEP 9], we decide to create another version of the same file and order all connections by timestamp, independently of the user ID.

#### Step 15. Order transitions by timestamp

The transition files are also sorted by timestamp, like the files in the preceding phase. The ordered files from two last steps are used in animated visualizations based on a timeline, and all connections/transitions will appear in the temporal order.

#### Step 16. Aggregate data

The aggregated data describes combined information accordingly to a grouping criterion. We aggregate data accordingly to temporal measurements (months, weeks, days, hours and minutes). *Six types of groupings* are made: months/weeks/days, months/days/hours, months/hours/minutes, weeks/days/hours, weeks/hours/minutes and days/hours/minutes. Each aggregation type has *three levels* of depth, from maximum to minimum measurements. We only use time units with defined constant length (weeks, days, hours) for the second and the third aggregation levels (one week has seven days, one day has 24 hours, etc.) to create uniform data structures. The algorithm makes an in-depth analysis of connection and transition data, evaluating each timestamp and inserting it in the respective level. For example, the unix timestamp *1,283,295,761* can be translated to *Wednesday, 1 September 2010, 12:02:41 a.m. GMT+1*, if adjusted to the Portuguese timezone. In the aggregation by months/weeks/days, this value goes to the third day of the week (depth level 3) of the first week (depth level 2) of the ninth month (depth level 1). When all records of a dataset are distributed by the aggregation levels, we calculate the defined statistics, such as minimum, maximum, average and total connection or transition times, for each data group of the level 3.

### Step 17. Create AP clusters

In the absence of the spatial information about geographical locations of APs, the *clustering* algorithm is implemented. It groups nodes by transition proximity so their spatial *positioning* can be related to neighborhood. The term *clustering* refers to the process of grouping a set of objects in order to create groups, or clusters, with similar elements that differ from the elements of other clusters [SECTION 3.2]. This technique merges deeply connected APs, granting the visual clarity of the artifact and providing additional information to the viewer when applied to a graph. The grouping uses the average transition time associated to every transition. First, the most accessed point is chosen to become the central node of the new cluster. Then, all nodes reachable within a time interval of 60 seconds are recursively added to the cluster. This temporal interval indicates that users can transit from one AP to any another AP in the given cluster within one minute. The process is repeated until all attainable nodes, including nodes that were not added to any cluster (lonely nodes), are processed and added to clusters. When the clustering is complete, we have a new organization of nodes: the clusters (groups) with children (APs). The IDs are assigned to clusters automatically and incrementally. The IDs of children APs remain unchanged to preserve the source-target relations of transitions. The results of clustering depend on the read data: different datasets may produce distinct organizations of clusters.

We opt for a static clustering, i.e., the process is performed once, operating on all records of the given dataset to produce valid general results. This is useful when the data is dynamically modified (in an animated artifact or on data transformations), the organization of clusters is fixed in every state of the network, that means that the nodes do not jump from one cluster to another (it would happen if we were reorganizing clusters in every state of the graph) and the viewer can recognize and follow the differences in the flow.

UC dataset	431 APs grouped in 172 clusters relying on 5,805,968 transitions
UL dataset	3,173 APs grouped in 1,297 clusters relying on 1,931,710 transitions
UM dataset	3,885 APs grouped in 1,570 clusters relying on 9,180,972 transitions
UP dataset	1,624 APs grouped in 637 clusters relying on 1,183,372 transitions

### Step 18. Order APs and clusters

The *node ordering* algorithm allows the attribution of a meaning to the *spatial arrangement* of nodes. First of all, the most accessed cluster is found, becoming a central target node, and added to the *ordered list*. Then, all nodes reachable from the target node through existing transitions are detected and added to a list. In every iteration of the recursive comparison by *proximity*, the node with the smallest transition interval to the central cluster is chosen from the list of reachable nodes. Then, the chosen node is added to the left or to the right side depending on the total time route to the target node. We perform this type of insertion because the circular node formation



implies the addition of the elements to the both side of the list. This way, all reachable nodes are analyzed and recursively added to the ordered list. When the ordering is complete for the first most accessed cluster, the next cluster with the biggest number of accesses is chosen from the *unordered list* of the remaining nodes and the process repeats until that list is empty. After finishing with the clusters, the algorithm acts on the inner nodes of each cluster, sorting them contiguously inside the group. When all these procedures are finished, the clusters and their children are ordered by proximity. Similarly to the previous step, the sorting is also performed once, using all records of a dataset to produce valid general results.

Once we completed the data treatment, the work on the visualizations begins. Firstly, we start to develop small animated applications to discover the valid understandable models and graphic techniques to depict the flow within the given networks. The next section describes that experimental process.

## 5.2. Experiments with the Graphic Layout

This phase of the research focuses on an effective arrangement of graphic elements in the visualization of the given network data and flow. Inspired by examples examined in the State of the Art [SECTION 2.2] and in the Related Work [SECTION 3.3], several ideas are implemented as visualization prototypes to demonstrate their possible applicability in this research. The programmatic sketches are named accordingly to the inspiration for their conception. Every layout should fulfill the requirement of being both a static and a dynamic representation of a flow. Taking this into account, all approaches are animated visualizations and any given still frame has to represent a static visualization. The number in the upper left corner of each model indicates the current timestamp. The developed sketches<sup>1</sup> use the UC dataset regarding the academic year 2010-2011.

In the majority of these prototypes, the APs are blue circles or bars and the transitions are lines. Both of them have different levels of transparency, depending on the method of representation of a transition in the model. The background is white to contrast with the blue elements. Every visualization runs accordingly to a temporal scale, starting from the timestamp of the first transition in the dataset and incrementing one hour. The transitions are shown in the order of appearance in the sorted data file. The developed models are illustrated and explained on the following pages.

<sup>1</sup> The animated artifacts can be seen through the hyperlink: <https://vimeo.com/72588252>

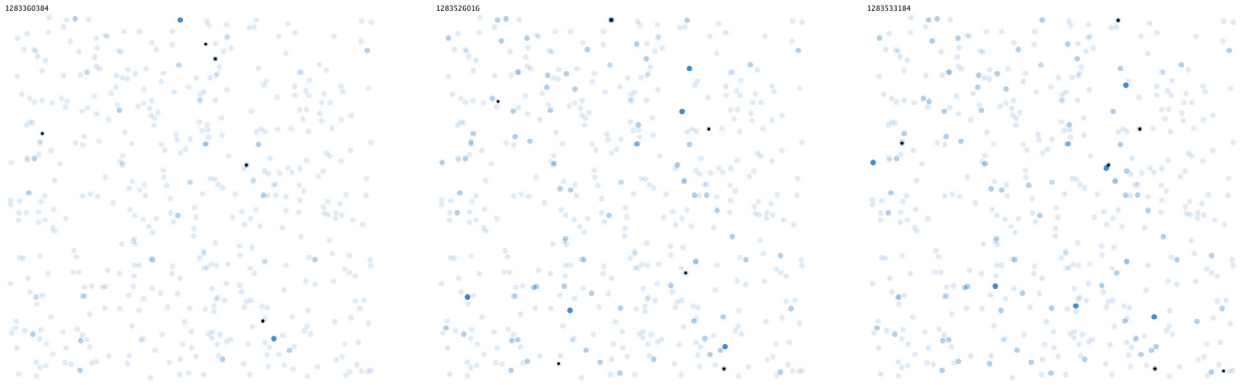


Figure 5.1. Three frames from the Starry Sky animation.

The *Starry Sky* layout [FIGURE 5.1] is inspired by the nocturnal sky full of glowing stars. The APs are randomly distributed in the window and their positions are fixed. When a transition occurs, the source and the target nodes become opaque. The target node gets the small black circle in its center to show the direction of the transition. When the next transition happens, the nodes of this transition gain opacity and the APs of the previous transition become transparent over time. This model does not work because only the current data is shown. Showing the direction with the point is not sufficient to demonstrate the flow. The random distribution of nodes in the space does not create any visible order or dependencies of the nodes.

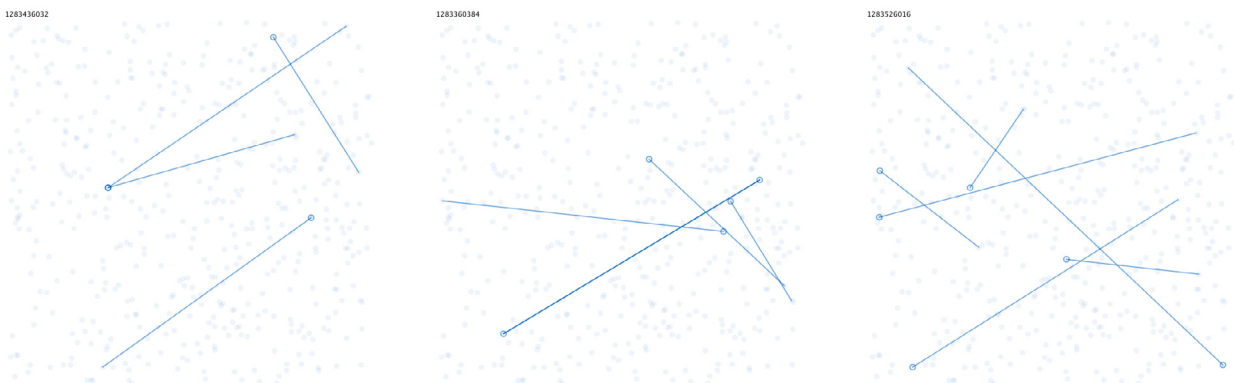


Figure 5.2. Three frames from the Starry Sky Paths animation.

The *Starry Sky Paths* [FIGURE 5.2] is an extension of the previous example. The difference is in animation during a transition. When a transition happens, an opaque line between the source and the target APs is drawn. The target node gains an opaque stroke showing the direction of the transition. The same animation is performed on the next transition, while the line and the node's stroke from the previous transition lose opacity. The line helps to understand the direction of the transition, but the remaining problems are unsolved, making the visualization limited in the communication of information.

In the *Pulsing Grid* layout on FIGURE 5.3, the APs are aligned to a modular quadratic grid and ordered by their ID number. This model is developed to examine the visual effects of the ordered *Starry Sky* and inspired by the *matrix plot* representation.

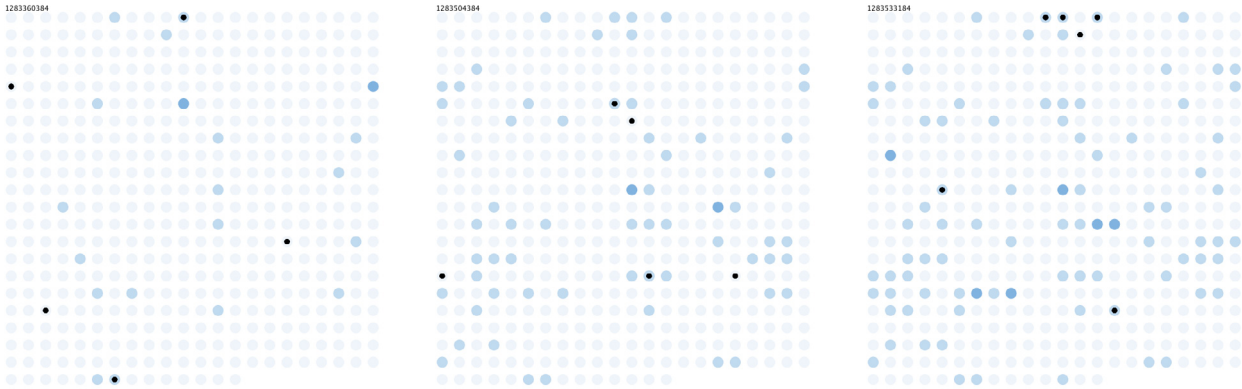


Figure 5.3. Three frames from the Pulsing Grid animation.

Thus, the transition animation is the same as in the *Starry Sky*: the source and the target nodes are opaque and the target is marked with the black point to show the direction of the transition. The grid helps to visualize the transition patterns, while the nodes are flickering with small temporal intervals. Yet, showing only the ordered nodes is not sufficient to understand the overall flow.

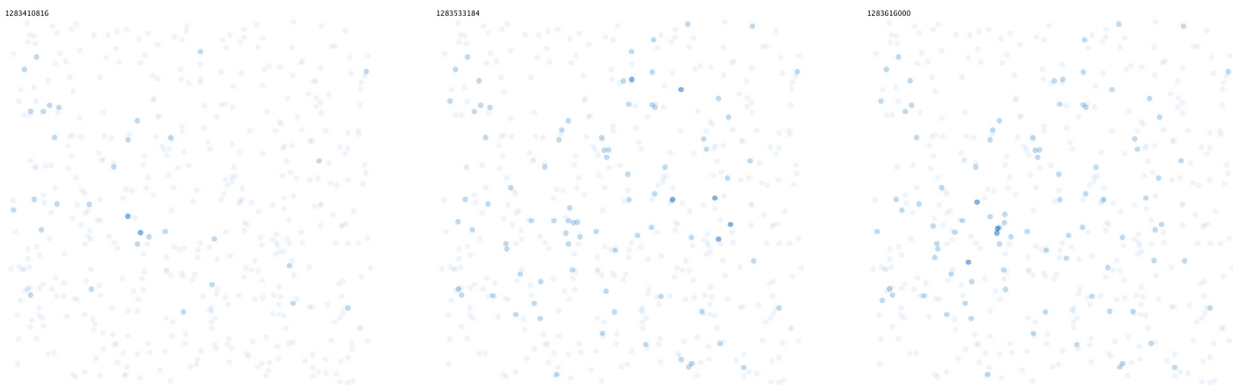


Figure 5.4. Three frames from the Target Magnets animation.

The *Target Magnets* layout [FIGURE 5.4] is inspired by the *force-directed layouts*, but with magnetic attraction instead of the mechanical spring force. The APs are randomly distributed with fixed positions. When a transition occurs, the source and the target nodes become opaque. The source node moves in the direction of the target node as it was attracted by a magnetic force. This movement indicates the direction of the transition. In this prototype, the movement time is not proportional to the transition time, because the temporal increment is big and the animation of particles would not be noticeable. The source node stops when it reaches the destination and remains there while not attracted by another node. When the next transition occurs, the nodes of this transition perform the animation and the APs of the previous transition lose opacity. The dislocation of nodes creates a visual grouping of most accessed points, since these remain close. The problem of this prototype is that at the end of the animation the majority of nodes gathers together. That means that almost all APs were accessed in the given academic year. This is a useful observation, but the nodes need to be spaced to show that even a relative structure of the network can be more or less stable.

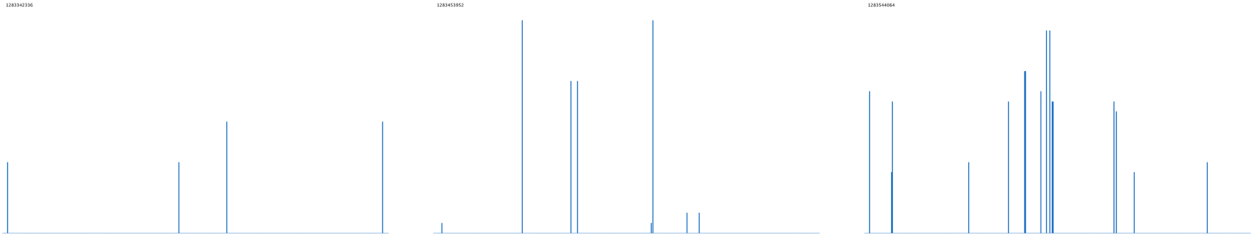


Figure 5.5. Three frames from the Equalizer x1 animation.

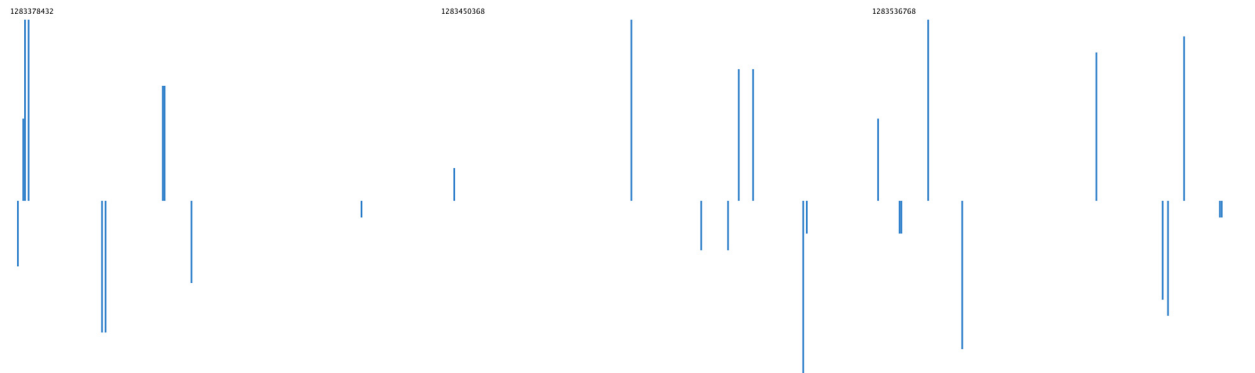


Figure 5.6. Three frames from the Equalizer x2 animation.

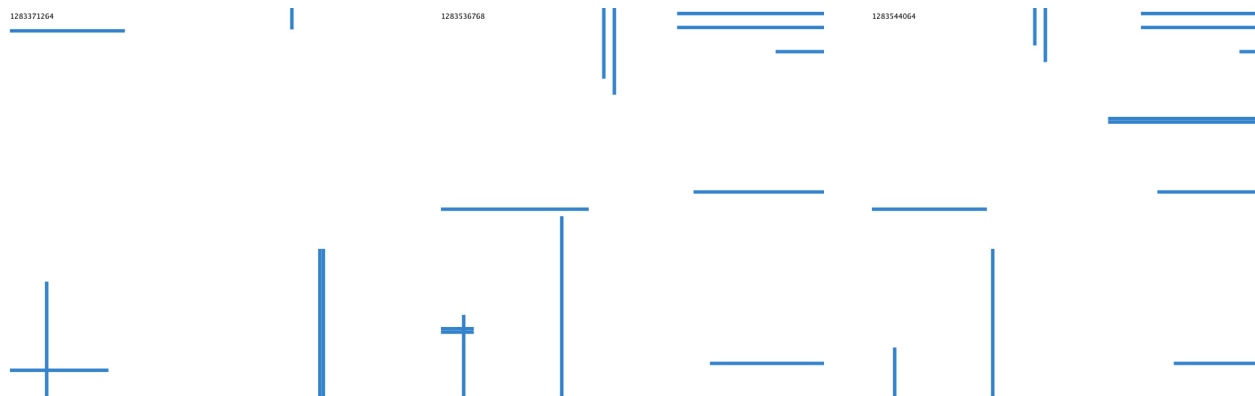


Figure 5.7. Three frames from the Equalizer x4 animation.

The *Equalizer* series is inspired by the *bar chart* representations. In the *Equalizer x1* version [FIGURE 5.5], the APs are aligned to the bottom of the graph and sorted by their IDs numerical order. When a transition occurs, two bars that represent the source and the target nodes, are equally scaled to the top of the window, mapping their height accordingly to the minimum value of the transition time. After reaching the limit, their heights reduce over time. The *Equalizer x2* and the *Equalizer x4* versions vary in the distribution of APs. The first [FIGURE 5.6] has two parallel and mirrored layers of nodes, aligned to the center of the window. A half of APs is represented on the upper layer and another half – on the lower layer. In the second version [FIGURE 5.7], the nodes are divided in four layers aligned to the four sides of the window. These layouts use more space to show transition lines and make the bars wider. The animation of transitions is the same that in the *Equalizer x1*. This model presents the same limitations as the *Starry Sky* and the *Pulsing Grid* layouts. On the other hand, the mapping of the transition time (in this case, to the

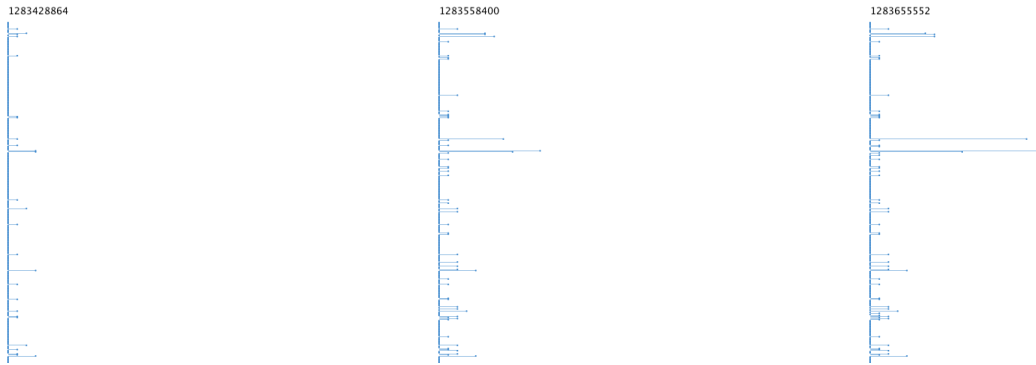


Figure 5.8. Three frames from the Horizontal Bars animation.

height of the bars) helps to visualize short and long transitions, what can be useful in calculation of the proximity of the APs regarding transition times between them.

The *Horizontal Bars* layout [FIGURE 5.8] is inspired by the classic *bar chart*. The APs are aligned to the right side of the window and ordered by ID number from the top to the bottom. When a transition occurs, the source and the target nodes advance to the right side a certain amount of distance. The bar lengths represent the number of transition occurrences of the corresponding nodes. This layout preserves the information previous to the current frame. The bigger bars belong to the most frequently accessed APs at the moment. The sum of all bar lengths represents the double amount of transitions, because both nodes advance when a transition occurs between them. Anyway, one cannot see the flow between APs in this prototype.

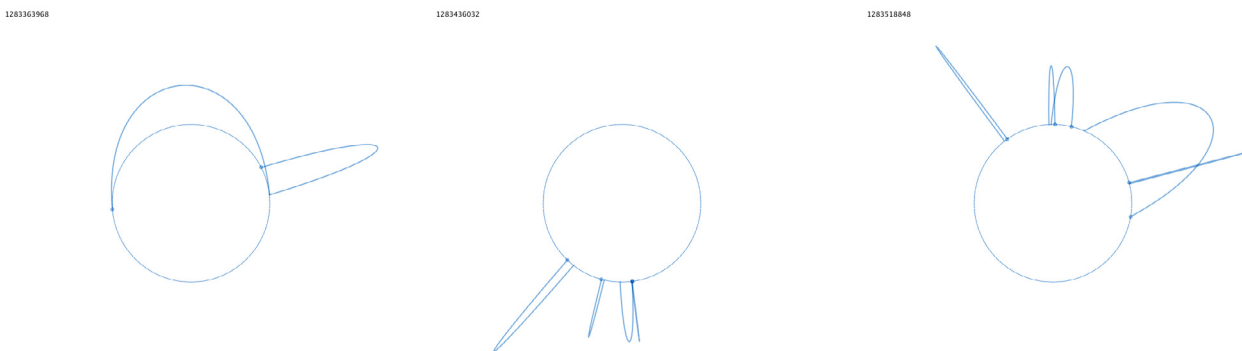


Figure 5.9. Three frames from the Flower animation.

The inspiration for the *Flower* [FIGURE 5.9] is the *radial convergence layout*. The APs are distributed along a circumference in the center of the window and sorted by ID number. When a transition occurs, an opaque curve is drawn outside the circle (in the radial convergence it is drawn inside the circle) between the source and the target APs. The curve's height is proportional. The target node is featured with a stroke line to show the direction of the transition. When the next transition happens, the same animation is performed, while the lines that belong to the previous transition lose opacity over time. The radial distribution of the nodes saves space in the application's window, so we can make the nodes larger and more visible. The curves indicate the flow between nodes at the current instant. The height of the bend

point of a curve helps to see the slow and fast transitions, and the space between two extremities depicts the relative distance between nodes in the circumference.

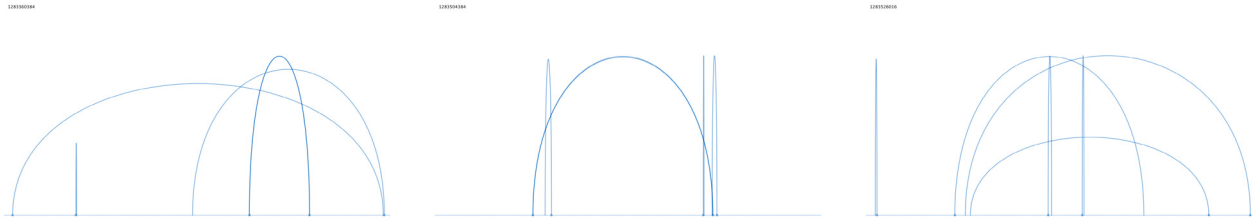


Figure 5.10. Three frames from the Humps animation.

The *Humps* layout [FIGURE 5.10] is a stretched version of the *Flower*. The APs are aligned to the bottom of the window and ordered by ID. When a transition happens, an opaque curve is drawn connecting the source-target node pair. Its height represents the minimum transition time value. The remaining graphic properties are the same as in the previous sketch. This prototype proves that the radial disposition is a better way to place the nodes in space than the linear one.

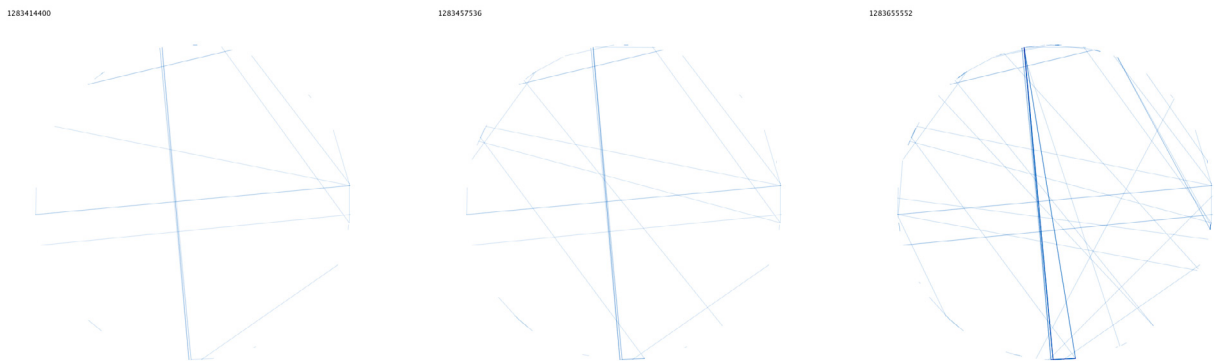


Figure 5.11. Three frames from the Round Web animation.

The *Round Web* layout [FIGURE 5.11] is inspired by the *circular layout*. The APs remain distributed on a circle and ordered by their IDs. When a transition occurs, a line with low opacity is drawn inside the circle between the source and the target. This model presents a cumulative effect of data, overlapping the transition lines. The common routes appear over time, because the respective lines overlap. Also, the circular layout can save the data-ink on the representation of transitions.

The *Soft Ball* layout [FIGURE 5.12] is inspired by the *force-directed layout*. The APs are distributed along the perimeter of a blue-filled circle and ordered by their IDs. Each node is connected to its neighbors and to the center of the circle with mechanical springs [FIGURE 5.13]. When a transition occurs, the springs that link the source and the target APs to the center become stiffer, what makes the two nodes move towards the center. These also pull neighboring nodes due to the spring connections. On the next transition, the nodes perform the same movement and the nodes of the previous transition tend to restore their initial positions, as well as their neighbors, because the springs lose the stiffness. This prototype shows the accessibility



Figure 5.12. Three frames from the Soft Ball animation.

of the APs. The closer is the node to the center of the ball, the more accesses it has. The effect is semi-cumulative, because the springs are constantly pull the node back to its original position and the successive accesses push the node to the center.

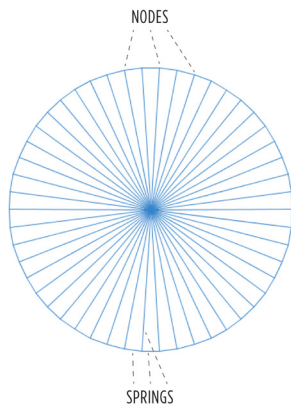


Figure 5.13. The structure of the Soft Ball.

The visualizations of transitions presented in this section vary in graphical layout and animation type. If we eliminate the temporal dimension and the traces of the performed transitions, all of these prototypes cease to work. The graphic marks of the mapped data shall exist to show the overall network flow. Even if these elements remain present on the layout, some models, such as *Starry Sky*, *Pulsing Grid* and *Target Magnets*, do not display any movement within the system. These do not represent the transition directly, but highlight the APs that belong to a transition. These models only work as animations.

Several prototypes work in both static and dynamic ways. The Horizontal Bars is a good example. The animation can be stopped in any moment and the current state shows that the bigger node bars represent most accessed APs. The same happens with the Round Web model. An important aspect to be considered is the spatial distribution of nodes. The effectiveness of the layout of nodes provides an intelligent and perceptible graphic representation of transition lines.

### 5.3. Summary

This chapter described the preparatory work of this research. The available datasets are analyzed, restructured and organized in processed files that are used in the creation of pragmatic visualizations, described in the next chapter. A variety of graphic layouts is created, being inspired by the existent models of visualization, explained in the previous chapters. The applicability of these experimental sketches to the visualization of network flows is discussed and the conclusions are traced. The next chapter explains the development of practical applications and graphs that represent the patterns of flow and usage in wireless networks.





## CHAPTER 6.

# Approaches for Visualization of Flows and Usage in Wireless Networks

This chapter reports the process of construction of graphic approaches to the visualizations of networks, related flows and usage patterns. This work phase consists of the construction of graphs accordingly to the objectives and the methodology described in CHAPTER 4. The developed interactive applications present experimental approaches inspired by the visualization models, explained in CHAPTERS 2 AND 3, accompanied by detailed descriptions of used techniques and principles of visualization and interaction. The utilized data files are described in the previous chapter.

The following visualizations depict the connections and transitions of users, the streams and patterns of flow and the internal structure of a network. In the construction of the graphs, we consider and test the application with all four datasets. The images in this chapter display the data of the University of Coimbra regarding the academic year of 2010-2011. This provides a contextual coherence to the entire process and demonstrates how the same information can be shown in different points of view. All scripts and applications are implemented using the *Processing* programming language and external libraries to expand the programmatic environment when required. We performed many experiments with the visualizations. We do not include previously developed models in this document due to the level of quality of those visual results, yet several examples are presented in the APPENDIXES section.

## 6.1. The Circuit of Transitions

The first application consists of the representation of the physical structure of a network and the flow of users within it. The radial layout was chosen for this purpose. A circular diagram works both in static (any frame of the animation represents the current state of the system) and dynamic (the animation shows the temporal evolution of the system) ways. Also, a circular arrangement of a network can maximize the data-ink used on the representation of nodes and edges due to its compact form and the possibility of layered distribution of information. Once we do not have any information related to the geographic distribution of APs, the circular diagram can present the network's structure in the abstract, relative and effective manner. This application loads the information about clusters from the ordered node lists [SECTION 5.1, STEP 18, PP. 80-81].

We represent the network with a *radial convergence layout*. The *clusters* are graph's nodes represented by *arc blocks* with a variable *length*, having two parts for *incoming* and *outgoing* transitions. The *length* of the cluster corresponds to the *sum* of all its transitions, what makes the node sizes vary accordingly to the accessibility. Therefore, the *length* of the whole layer of nodes corresponds to the *total amount* of existing transitions. The cluster node are distributed along a circumference. The *positioning* of nodes begins from the rightmost point of that circumference, centered in the window, and increments in the clockwise direction. The clusters' *color hue* values are assigned accordingly to the *position*, starting from zero and incrementing up to 359 (the hue range). This method results in the different colors for all clusters, creating a gradient color for the layer of nodes. The clusters are lightly spaced to provide visual distinction between adjacent nodes with similar color [FIGURE 6.1].

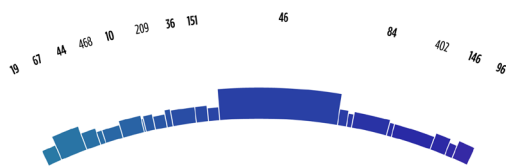


Figure 6.1. Detail of nodes in the radial imposition.

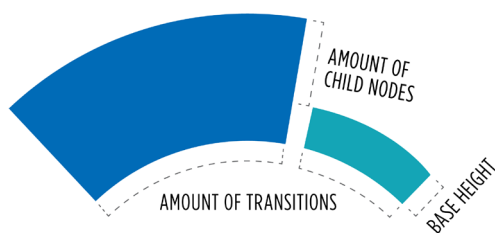


Figure 6.2. The visual structure of the graph's nodes.

The APs grouped in a cluster become its child nodes, inheriting the father's *color* with a small variation of the *hue* value, while the brightness and saturation remain the same. The single clusters (containing one child) are represented by the child node. The cluster's *height* is proportional to the number of its child nodes [FIGURE 6.2], transforming the node layer in a *circular bar chart*. The *type of imposition* of the network is *polar*. The single clusters have the smallest height, while the clusters with a bigger number of children create higher bars.

The *transitions* are edges of the graph represented with *Bézier curves*, thinner on bends and larger on extremities. The *width* of an extremity corresponds to the *count* parameter, that holds the total number of transitions between the source-target node pair, grouped in the edge. The thicker curves present a bigger quantity

of transitions, that also aids to establish a visual discrimination. A transition inherits the *color* of the *source node* to show the visual *hierarchy* between common APs and to represent the *direction* of the flow from one AP to another. The *opacity* of all edges is set to 40%. The edges displayed *inside* the main circumference connect two *distinct clusters*, while the ones *outside* the circle belong to the *same node*. Therefore, these self-node transitions do not visually affect the transitions inside the circle [FIGURE 6.3]. The *imposition* of this diagram is a *network* because any “node” element can be linked to all other “node” elements with edges that represent transitions. Transitions between two nodes X and Y can be bidirectional – from the source X to the target Y or from the source Y to the target X. These two types of transitions can have different values of the transition time. The structure of the model is shown on FIGURE 6.4.

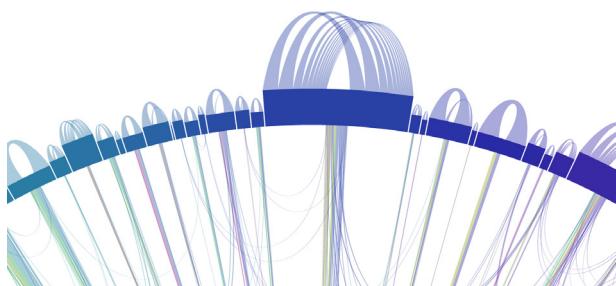


Figure 6.3. Detail of the graph's edges.

The node ordering algorithm permits a lesser intersection of edges and reduces the amount of curves stretched between the opposite sides of the graph, because all nodes are ordered by *proximity* based on *transition times* between them.

Having this, we assign the edge's *average transition time* value to the *height* of each curve: lesser transition times are represented by smaller curves, bigger transition times result in higher curves [FIGURE 6.3]. The height value is defined by a *logarithmic function* of base 10, what is not an exact representation of transition times, but it reveals the differences between curves' heights more clearly than a standard linear mapping. The edges outside and inside the circle are represented in *different height scales*: the outer curves are smaller than the inner ones, if we compare the same transition time values for both types of edges. This modification limits the excessive heights of the outer edges and prevents their rising beyond the application's window. What really matters is to *visualize routes* between pairs of nodes, the transitions from and to the same node are not so important. The *thickness* of the bend point of an edge is also proportional to the *count* parameter. This graphic element distinguishes more common transitions from the less common ones.

The incoming and outgoing transitions of a node are *organized by width* in descending order, from the center to the right and left margins of a node. With this *edge junction* arrangement, the scattered edges of a node are aggregated in its center [FIGURE 6.3], providing a higher clarity to the distribution of transitions. The following types of transitions are represented visually: cluster-to-cluster, cluster-to-node, node-to-cluster and node-to-node.

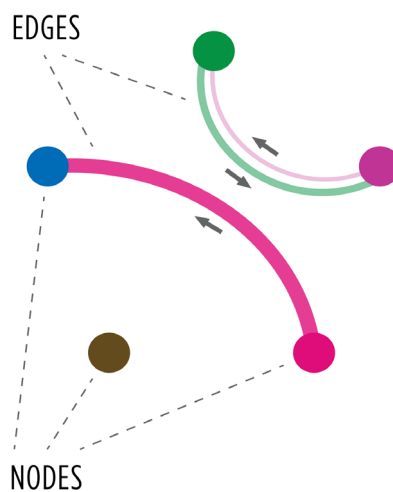


Figure 6.4. The general structure of the network layout.



Figure 6.5. The controllers of the graphic user interface of the application.

At the upper-right corner, we include the *user interface controllers*: the filtering bar, the “Show/hide the node IDs,” the “Show/hide the top 5 nodes,” and the “Save as image” buttons [FIGURE 6.5]. All interface controllers are developed by us without using any external libraries with pre-implemented elements.



Figure 6.6. The four states of a button (inactive, hover, clicked, active), distinguished by the visual appearance.

A *button* has four states of activity [FIGURE 6.6]: *inactive*, *hover* (having the mouse cursor over it), *clicked* (the mouse’s left button is pressed while the button is hovered) and *active* (the mouse’s left button is unpressed while the button is clicked). The button’s *icon* changes depending on its current state. All buttons have the same principles of interaction. The *hover* state shows the button’s description below the image and the *active* state enables the function associated to it. Counting from the left to the right, the *active* buttons from FIGURE 6.5 allow the following functionalities: the first button makes appear the IDs of the bigger nodes; the second expands the five biggest nodes and highlights their edges; and the third takes a snapshot of the displayed data and saves it in the PDF format. After activating the node IDs’ button, only the identifiers of bigger nodes are shown. This feature avoids the overlapping of the ID labels and maintains the readability of these.

The *filtering bar* is a horizontal bar with a slider box. This controller activates the *edge filtering* algorithm, implemented to remove the edges with the count parameter below a given threshold. The procedure *cleans up* the graphic information and *highlights* the important data. After all edges are filtered, the nodes with no transitions

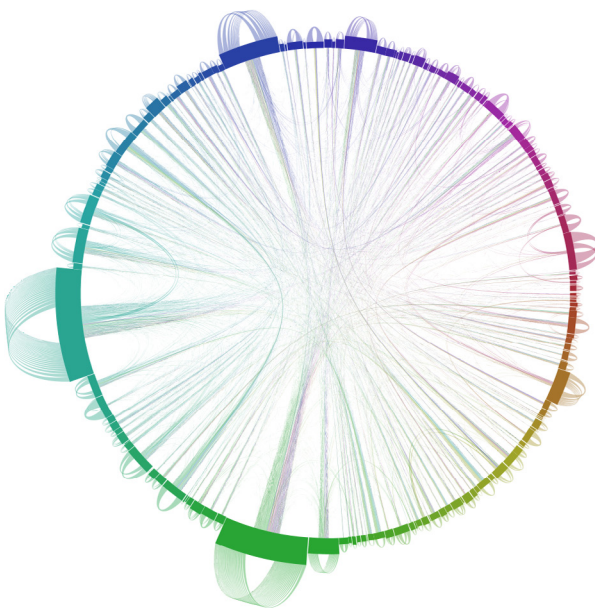


Figure 6.7. The graph with no data filtering: 2,770,912 transitions grouped in 45,049 edges, 417 nodes grouped in 169 clusters.

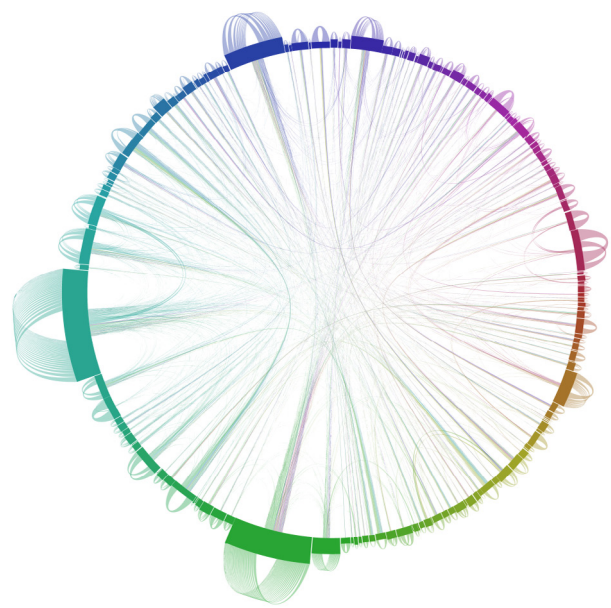


Figure 6.8. The graph with filtering below 50: 2,562,046 transitions grouped in 3,006 edges, 400 nodes grouped in 167 clusters.

UC UL UM UP  
2010-2011 2011-2012

Figure 6.10. The dataset selection controllers.

are automatically deleted (*maximization of the data ink*). The examples of the edge cleaning are shown on FIGURES 6.7, 6.8 AND 6.9. The *filtering* of the displayed data is performed by *dragging the slider* along the bar with the mouse cursor, between the minimum and the maximum defined thresholds: sliding to the right side increases the filtering limit, sliding to the left side decreases it. The mouse button must be released to enable the function and start removing edges and nodes of the graph. The current *value* of the edge filtering threshold is shown above the slider. Every time the user performs the filtering, the message “Filtering data” appears in the center of the window, then the data is restructured and, in the end of the operation, the graph is updated.

In order to provide exploration of various datasets, we include the possibility of the *dataset selection* in the graphic interface [FIGURE 6.10]. The respective button controllers are situated in the lower-right corner of the application’s window. The user can select one among the four possible universities and the intended academic year. After choosing a dataset, the available year buttons appear below. When an academic year is selected, the screen lightens and the message “Loading dataset” appears. After the data is loaded, the text message indicates the current dataset and the academic year and the graph is updated. The dataset’s designation is included in the contextual information section as a *title*. The displayed data comes from the files with the ordered clusters and the processed transitions. These files contain pre-processed transitions, grouped in edges to obtain the overall routes between clusters. As a result, we have only one transition between a unique pair of APs with all information associated to it: the source AP, the target AP, the number of performed transitions between them, the sum, the average, the variance and the standard deviation values of the transition time. The dataset selection allows the visualization and

the study of different datasets without relaunching the application. All previously and further implemented functionalities work with any dataset.

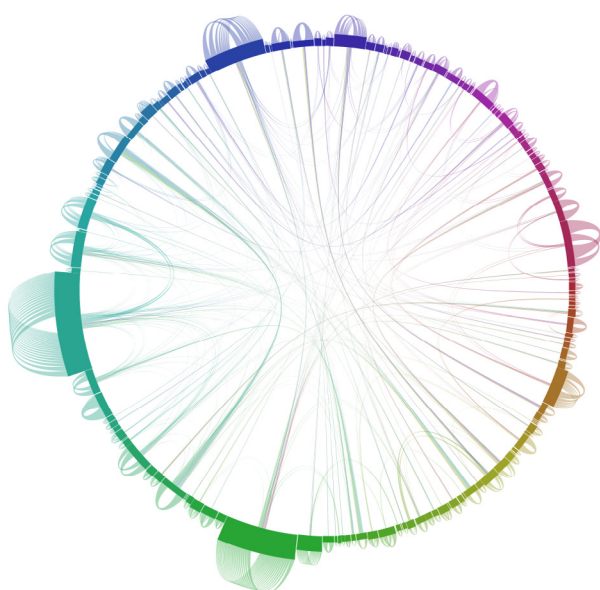


Figure 6.9. The graph with filtering below 500: 2,220,990 transitions grouped in 681 edges, 357 nodes grouped in 160 clusters.

All pre-processed datasets have a huge amount of transitions. *Bézier* curves consume a lot of computational resources. This reflects on the response time of the application: the frame rate of the graph is drastically reduced due to the number of edges. The analysis of the edge files shows that the major part of edges has small values of transition count. When displayed, these edges are almost invisible due to their low thickness, and produce visual clutter rather than the data-ink ratio. We pre-filter all datasets

in order to remove these insignificant transitions and to show only the relevant transitions, accordingly to the principle of *maximization of the data-ink*. The procedure is based on the discovery of the filtering limit that retains the big transitions and removes the small and perturbing ones. By small transitions, we consider the edges with transition count lesser than a certain threshold. For example, an edge with 50 transitions has a relatively small importance, it signifies that one AP was accessed 50 times from another AP. The amount of present clusters and the frame rate are also important aspects to consider. The determined filtering limits are different for each dataset because the numbers of transitions and active clusters vary accordingly to the data. So, we standardize all sets to show the resembling quantities of the data with the minimum filtering. The optimal ratio between the displayed data and the frame rate manifests itself with 140-160 clusters and 1,100-1,400 edges. The file loading and the response time of the application increase greatly with the filtered datasets. The visual clutter disappears and the data loss is not significant. The filtering bar adapts to the established limit when distinct files are loaded. A snapshot of the application displaying pre-filtered data is shown on FIGURE 6.11.

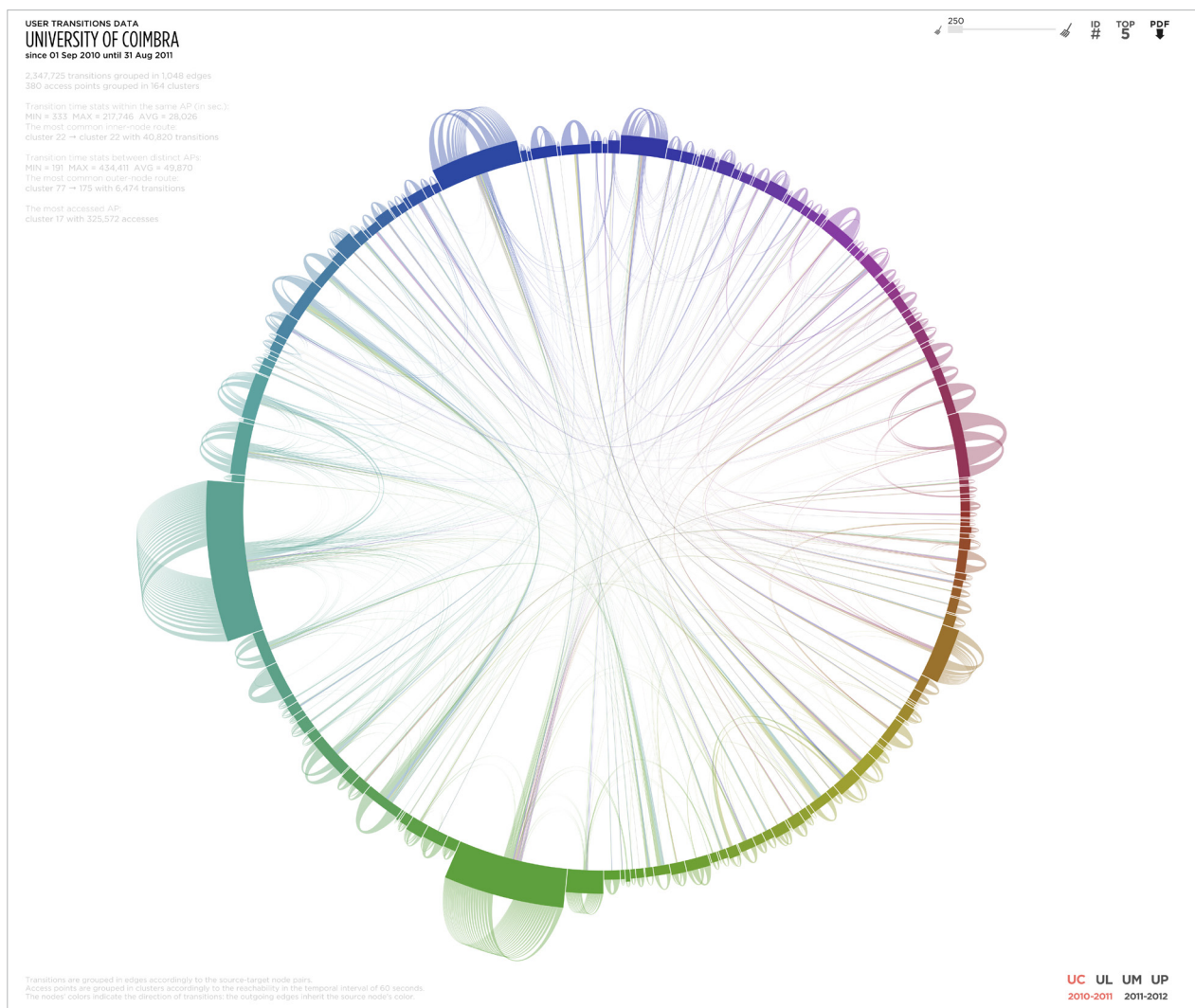


Figure 6.11. Screenshot of the application of visualization of user transitions, displaying the UC dataset of the academic year 2010-2011 with the applied pre-filtering.

We provide the *contextual information* about the visualization itself and the data [FIGURE 6.11]. It is situated at the upper-left corner of the application’s interface. The text contains the visualization’s *title* “User Transitions,” the temporal *period* covered by the displayed dataset, and the overall *amounts* of transitions, edges, APs and clusters. The transition time statistics are divided in two parts: the data about the *transitions within the same AP* (the edges outside the nodes’ circle) and the data about the *transitions between distinct APs* (the edges inside the nodes’ circle). Both parts include the minimum, the maximum and the average transition times, as well as the most common route information. At the end of the transition statistics, the information about the *most accessed AP* is shown. The explanation of the edge and node grouping and the meaning of colors is situated at the lower-left corner of the window.

Then, several *interactive possibilities* are enabled<sup>2</sup>. Putting the mouse cursor over a node makes appear its *ID number* at a certain distance. The numbers displayed in *bold* identify clusters, the IDs shown in *regular* type weight belong to APs. When the cursor points on a cluster, the *number of children* is displayed below the ID number. Also, the node’s *technical name* appears near the mouse pointer, and the node’s edges are *highlighted* with a bigger opacity. This allows the user to focus on the flow related to the selected element. The textual sections have low opacity to prevent a constant attraction of the user’s attention. The user can make the text *opaque* by pointing the cursor over it, to read the information. These are *direct selection* techniques.

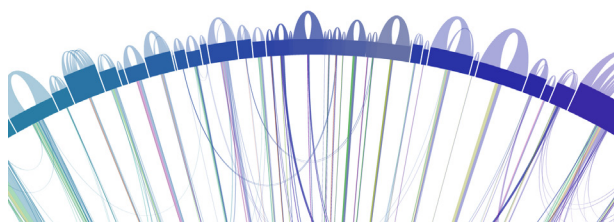


Figure 6.12. Detail of expanded and collapsed clusters, and the respective incoming and outgoing transitions.

The user may *expand* a cluster clicking on any point inside its area, and the cluster’s children appear at the position of the father node. The child nodes have no spacing between them, what makes them occupy exactly the same length that the father-cluster has. The expanding

gives control over the detail level of the visualization and allows the visualization of the transitions of individual nodes. To *collapse* the cluster, the user may click on any of its affiliates. The expanding and the collapsing processes are *animated*: the first reduces the cluster’s size and reveals the child nodes; the second makes the cluster grow and overlay its children. This is made to provide a better perception of what is happening on these actions. While collapsed, the incoming and outgoing transitions of all children of a cluster are represented by the cluster. While expanded, one can observe the transitions coming from and to the inner nodes [FIGURE 6.12]. The “Show/hide the top 5 nodes” button is useful to show the biggest APs/clusters and their transitions with a single click. When the user activates it, the *edges* linked to these nodes are *highlighted*. All these functionalities refer to the *details-on-demand* technique. Lastly, the cursor’s *icon* changes when interacting with different elements: by default, the icon is an *arrow*, becoming a *hand* while over the filtering bar, buttons or clusters.

<sup>2</sup> The video demonstrating the interactive possibilities of the application can be seen through the hyperlink: <https://vimeo.com/72983108>

This model of the visualization, the work and the generated visual results related to it were submitted to and accepted at the *ACM SIGGRAPH 2013* conference in the *poster* format<sup>3</sup>. The abstract and the images of the poster are included in the APPENDICES section [PP. 54-55].

The interactive visualization, described in this section, represents the overall user movement within the network, allowing one to perceive the flow of users and the patterns of transition between the network's APs. The visualization displays multivariate data: the *nodes* are shown along the circumference, *clustered* and *ordered* by proximity; the *total of performed transitions* within the network is depicted by the length of all node blocks, having the *total of incoming and outgoing transitions of an AP* represented by the node's arc length; the *number of clustered APs* is assigned to the node's height; the *position of a node* shows its proximity relatively to other nodes; the *quantity of transitions* grouped in a curve is determined by the thickness of the curve's extremity; the *value of the average transition time* of the edge is represented by the height of the bend point of the curve; the *origin of the transition* is indicated by the color of the edge. The textual information presents the numerical and contextual data. The graph emphasizes the most common *movement routes* in the network due to the edges' thickness. The *most accessed points* are also highlighted due to their bigger arc lengths. The *network's structure* is revealed by the clusters' grouping factor and the node arcs' thickness. Transitions within the same node/cluster prevail over the transitions between distant nodes, which occur less often. The graphic artifacts and the detailed analysis of the visual results obtained with this model of visualization of the user transition data are presented in SECTION 7.1.

## 6.2. The Matrix of Connections

The second visualization created in the context of this research displays the user connections inside a wireless network campus. With this tool, we are able to visualize when users were connected to the university's network, revealing the particular and global network's connectivity. The graphic layout is simple and precise: a dynamic two-dimensional *matrix diagram* that shows the users and the periods they were online. Since the connection datasets [SECTION 5.1, STEP 9, P. 77] contain the necessary information, we map the data directly to the layout.

The matrix represents the list of all users and the respective connections, existing in the given dataset. The evenly spaced *rows* are attributed to *users*, and their *connections* to the APs are shown on the *timeline*. The application shows two types of visualizations: the separate and the joint connections. The *separate connections'* graph is inspired by the *matrix plot* (BRUNER, 2011) representation. We represent the connection intervals of the users accordingly to the real time scale. The *time scale* is extracted from the very *first* and the very *last* connection *timesteps* of the given academic year. The *connections* of a user are represented with *red traces* with *fixed thickness*

<sup>3</sup> Ruslan Kamolov, Penousal Machado and Pedro Cruz (2013) *Visualizing the flow of users on a wireless network*, in: ACM SIGGRAPH 2013 Posters (SIGGRAPH'13). New York: ACM, Article 114, 1 page.



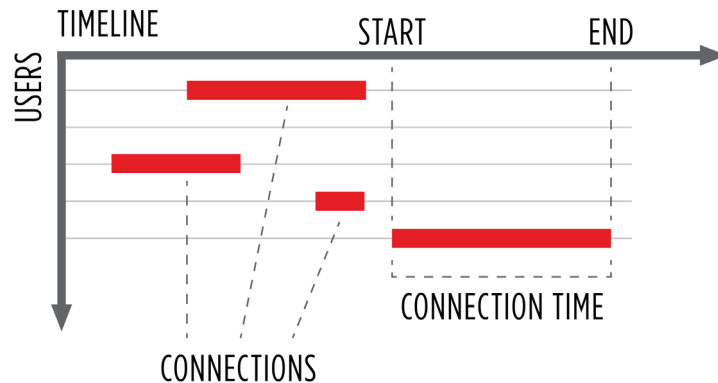


Figure 6.13. The structure of the separate connections' graph and the mapping of the connection times.

and *variable length*, mapped horizontally on the horizontal axis. Each line is drawn from the instance when the connection started to the moment when the connection stopped. Thus, the length of a line is proportional to the connection time. The red color is chosen to show connections, providing the *high contrast* between white background and black contextual information. The structure of this graph is displayed on FIGURE 6.13. On the other hand, the white spaces between the red lines represent each user's transitions, described in the previous section [FIGURE 6.14]. The *joint connections' graph* is a horizontal *bar graph* (HARRIS, 1999, PP. 37-52). That section demonstrates the total time each user spent online, merging all his connection times together [FIGURE 6.15]. The users are sorted in *descending order* regarding their *total connection times* in both graphs to provide a visual *hierarchy* in the graph. These two variations of the same graph form a *diagrammatic imposition* (two perpendicular axes with the variable of one axis being established for the variable of another axis) and belong to the *orthogonal type of imposition*, having two components: users and time. The temporal *grid ticks* appear behind the red lines of connections, and the respective *labels* are situated above and below the graph's area. In the separate connections' graph, the ticks represent real *days, weeks and months*. In the joint connections' graph, the ticks divide the timeline by days and weeks only, since week is a standard and common time unit with the biggest invariable length.

At the application's header, the *contextual information* is shown: the title "User Connections," the dataset designation, the covered temporal period, and the overall quantities of users and connections. Also, the *statistics* of connection times, such as sum, maximum, minimum and average values, are included. Similarly to the previous application, this section is almost transparent to make it less noticeable. The user can make the text opaquer by pointing the mouse cursor on it.

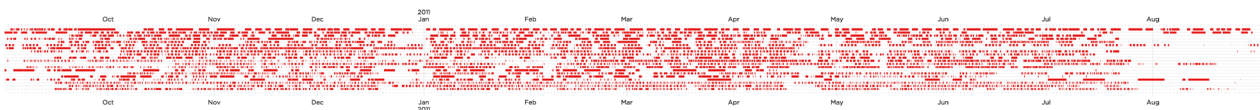


Figure 6.14. The separate connections of 20 users.



Figure 6.15. The joint connections of 20 users.



Figure 6.16. The header controllers of the application.

The header also holds the *interface controllers*, which are: the scaling bar, the “Show/hide the user IDs,” the “Show/hide the horizontal grid,” the “Separate/joint connections,” and the “Save as image” buttons [FIGURE 6.16]. The global interaction behavior of these controllers is explained in the previous section [PP. 92-93]. The controls work for both separate and joint connections’ graphs. Counting from the left to the right, the buttons in the *active* state offer the following functionalities: the first button makes appear all user identification numbers on the both sides of the graph in alternate order; the second shows horizontal grid lines along the user rows to simplify the visual localization of the single user’s data; the third toggles between the separate and the joint connections’ graphs; and the fourth takes a screenshot of the displayed information in the PDF format.



Figure 6.17. Screenshot of the application of visualization of user connections, displaying the UC 2010-2011 at the minimal zoom.

We introduce the *scaling controller* (a slider bar) to perform the *zooming* transformation of the view. The graph’s horizontal *scaling* can be made by moving the slider along the bar, between the *minimum* and the *maximum* thresholds: sliding to the right side increases the scale factor, sliding to the left side decreases it. The current value of the *scale factor* is presented above the slider. The minimum value is always equals to 1, what makes the data fit horizontally in the graph’s window, displaying the *whole period* of connectivity of users. The maximum value is defined by the number of *months* existing in the current dataset, in order to display *one month* of connectivity at the maximum amplification of the separate connections’ graph. When the user starts scaling the graph, the *scroll bar* appears at the bottom of the window, which enables the *horizontal movement* through the timeline [FIGURE 6.17].

The user must drag the slider along the bar to navigate horizontally. The slider's *position* actualizes relatively to the *range* of the actually visible data. The *size* of the slider corresponds to the current *zoom factor*: the greater zooming makes the slider proportionally smaller. Both parameters of the scroll bar, the position and the size, provide a *context map* that helps the user orient himself in the visualization. The vertical grid ticks adapt to the actual zoom – the bigger scaling factor reveals more ticks and the lesser factor hides the ticks of the smaller grid divisions. This modification allows the user to see the relevant *grid information* accordingly to the *current zoom* status and hide the unnecessary and disturbing lines, accordingly to the principle of *erasing of the non-information*. If all users stated in the dataset do not fit in the window, a *scroll bar* appears on the right, allowing the *vertical navigation* through the user lines. The size and the position of the bar's slider work likewise the horizontal bar's slider.

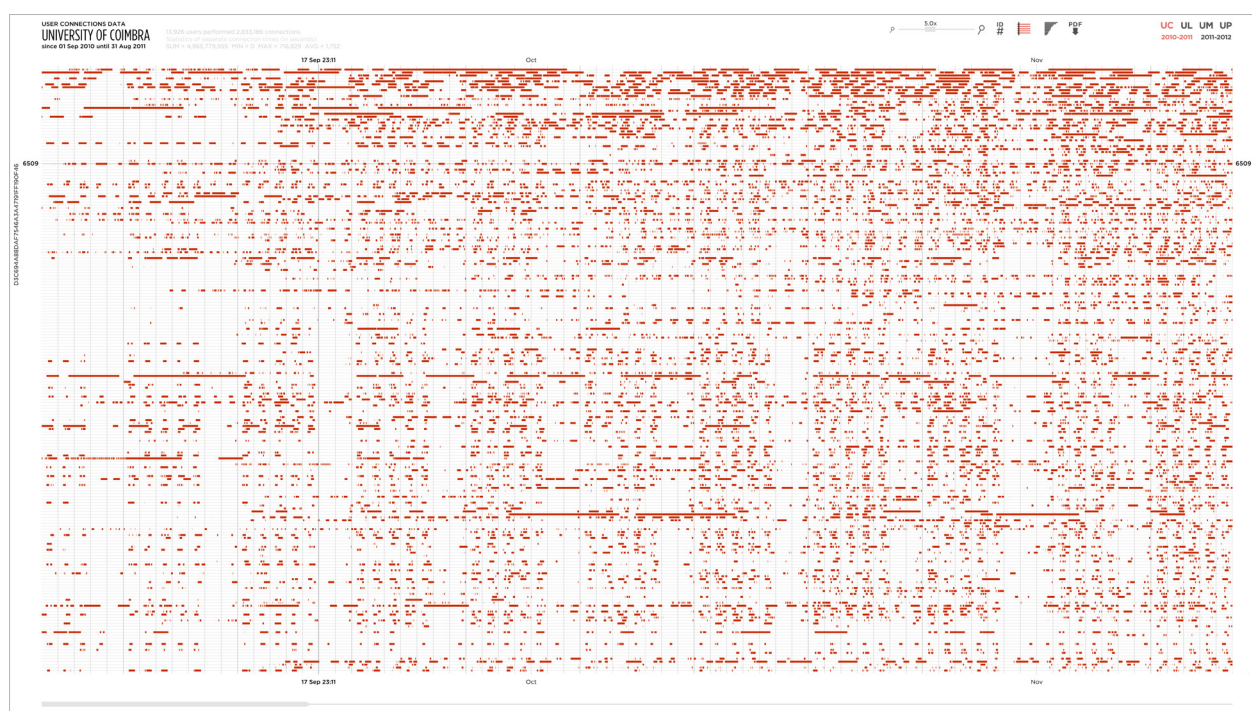


Figure 6.18. Detail of the separate connections' graph, screenshot.

The *mouse cursor* plays an important role in this visualization, offering the *exploration* of the information<sup>4</sup>. Pointing the cursor on the connections of a user highlights the *user line* and shows the user's *ID number* on both sides and the *technical name* on the left side of the graph. This feature works for both separate and joint connections' graphs. Additionally, the same action reveals the user's total connection time on the left of the bar, when the joint connections are displayed (*details-on-demand technique*). A *vertical line* is added to the actual cursor's position to facilitate the localization of the extremities of the connections on the timeline, to compare different connections situated at the same level, and to show the temporal values accordingly to the aiming position [FIGURE 6.18]. While exploring the separate connections' graph, the mouse pointer indicates the desired *date and time* at the top and bottom grid label's section. While interacting with the joint connections' graph,

<sup>4</sup> The video showing the interactive functionalities of the application can be found through the hyperlink: <https://vimeo.com/72987496>

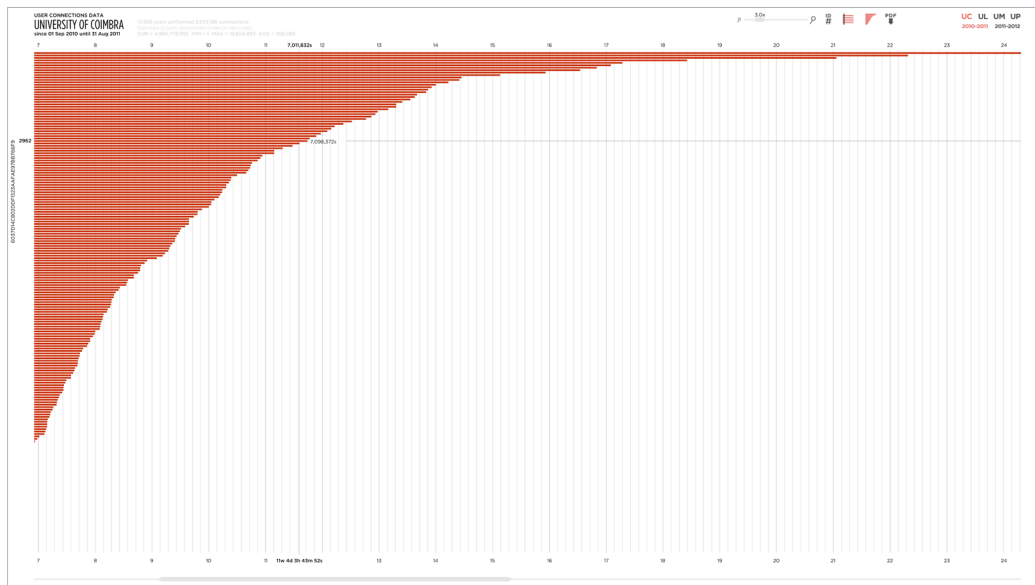


Figure 6.19. Screenshot of the amplified joint connections' graph.

the cursor shows the values of *total time*, counting from the starting point to the current position of the pointer. These values are also in the grid label's zone and follow the pointer's position as well [FIGURE 6.19]. The user can *navigate* through the graph by performing a *dragging* gesture (panning) or *scrolling* with the bars. However, both types of navigation are synchronized and can be used as alternatives. That navigation is a *camera movement technique* that changes the observer's positions relatively to the entire graph. Different types interactions are distinguished by the change of the cursor's *icon*: by default, the icon is an *arrow*; in the connections' area, the pointer is composed by *two intersecting lines*; while dragging the graph and using the scroll bars or buttons, the icon changes to a *hand*.

Similarly to the previous application, the possibility of choosing the desired dataset is present. The corresponding buttons are situated in the upper-right corner of the application's window. The user can select one of the four universities and the desired academic year. We have pre-processed the connections grouped by user ID to produce smaller files that load faster in the visualization, saving the application's response time. The chosen dataset is read from the file with processed connections, allowing the visualization, analysis and comparison of various sets in the same application. All functionalities offered by this application work for any selected dataset.

The elements of the graphic user interface, the interactive mouse cursor and all respective functionalities, described above, provide the exploration of the displayed user connections. This exploration is based on the interactivity with the user. The network's connectivity data is shown in the central area, the user can visualize separate or joint connections at a time. The information regarding the total quantity of users and connections, as well as the numerical statistics, is shown at the top of the window. The user IDs can be switched on or off, as well as the horizontal grid lines that trace the connection's line of each user. The scaling bar allows the deepening in the information, revealing its details and expanding the connections that

are visually merged without the zooming. The vertical and the horizontal scroll bars permit the navigation along the users' and the temporal axes. Lastly, a screenshot of the actually displayed information can be taken.

This section presented the developed visualization of user connections within a wireless network. The interactive graph displays the overall user connectivity, allowing the perception of temporal intervals the users spent online. The visualization displays multivariate data on two layouts: the network *users* are distributed on the vertical axis in ascending order of the total *connection time*; the *time scale* is represented on the horizontal axis; in the separate connections' graph, the *user connections* are mapped on the timeline-based matrix accordingly to the start and stop events; in the joint connections' graph, the *total times* of user connections are shown in a bar chart. The textual section describes the numerical and contextual information. The users with a *bigger connection time* are shown in the top of the graph. The graph depicts the *common patterns of connectivity* in the network regarding the density of connections. The common *connection intervals* are revealed by the vertical visual repetition of red traces and are localized by the temporal grid. The specific periods, such as rush hours, exam periods and holidays, are also clearly visible due to the level of density of mapped connections. The *zooming* and *panning* techniques allow both general and detailed views of the data. The analysis of the visual results provided by this visualization of user connections is described in SECTION 7.2.

### 6.3. The Panels of Flow Shapes

The third model of visualization is based on the revelation of patterns in the aggregated network data. We opt for the aggregation by temporal measurements in order to analyze the patterns of flow and usage in the network through time. The available connection and transition data is grouped by time units, such as months, weeks, days, hours and minutes, and the respective quantities of records are displayed. The process of aggregation is explained in SECTION 5.1 [STEP 16, P. 79].

The graphic layout is based on the *small multiples* representation, a series of adjacent panels that show the grouped data. We dispose the panels in *columns* and *rows*, accordingly to the *primary* and *secondary aggregation levels*, to create a two-dimensional *matrix diagram* [FIGURE 6.20]. Each panel, or *cell*, is an *area graph* (HARRIS, 1999, PP. 10-21) with *orthogonal imposition*. We opt for the area graph instead of the simple line graph because the visual similarities and differences between shapes are more perceptible than between lines.

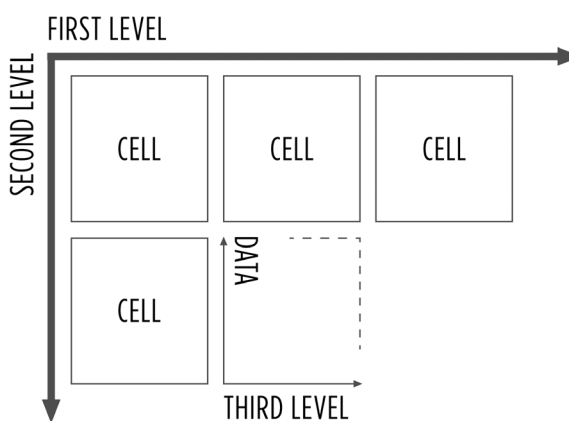


Figure 6.20. The structure of the panel matrix.

The information is mapped on two axes: *one axis* corresponds to the *third aggregation level*, and *another axis* shows the *values of data* we selected for the grouping [FIGURE 6.20]. We use a *linear scale* to map units of the aggregation level, and a *logarithmic scale* to map the network data. The choice of an exponential scale is due to the huge scale difference between different temporal fields we visualize. The minimum time varies from zero to 10 seconds, while average times vary from 1,000 to 100,000 seconds, and the total time can aggregate up to 10,000,000 seconds. Displaying the data within these extreme ranges makes the smaller values disappear, when shown on the same grid scale. With the logarithmic scale of base 10, we are able to handle the abrupt peaks in the data and show multilayered graphs within the same value range. Therefore, we obtain a *semi-log plot* for each area graph. The *curve* connecting the values is *smooth*, allowing a constant eye-following and an easier comparison of the whole shape with other panels [FIGURE 6.21]. The smooth curves avoid the harsh breaking of the graphic continuity as the segmented straight lines do.

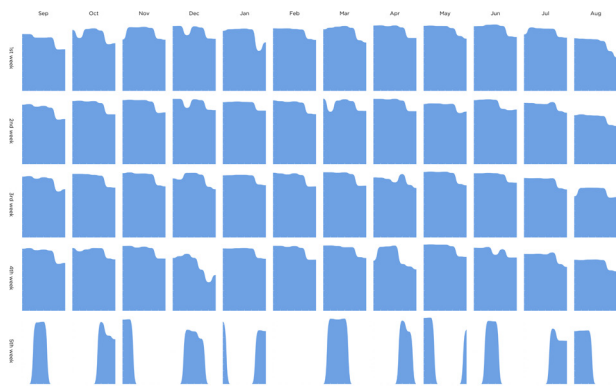


Figure 6.21. The panels' graph area with the horizontal reading view.

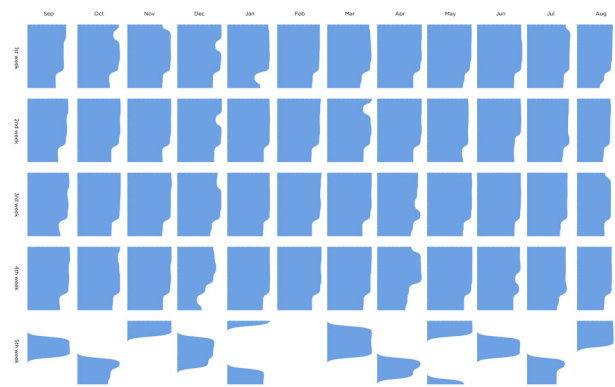


Figure 6.22. The panels' graph area with the vertical reading view.

The graph has two reading views: the horizontal and the vertical. In the *horizontal reading view*, the third level of aggregation is mapped on the horizontal axis of each cell and the data is represented on the vertical axis. There can be seen straight horizontal baselines and the variation of the heights. Each of these *height positions* displays the *value* of the data. FIGURE 6.21 shows the *quantities of users* (the vertical axis of a cell) that performed connections in seven *days* (the horizontal axis of a cell). These values are grouped by five *weeks* (the graph's rows) and twelve *months* (the graph's columns) of the given academic year. For example, the reading of the very first cell at the upper-left corner of the graph is: “on Monday of the first week of September 2010, the amount of users connected to the network was...” and check the respective height value. The *vertical reading* option is implemented to maintain the visual structure of the matrix, with the same number of rows and columns, and to rapidly switch the view to perform the vertical and continuous data reading along the columns. In this case, the weeks of a months are ordered and the *amounts of users* are mapped to the cell's *width* (the horizontal axis). The spacing between the days (the vertical axis) becomes larger due to the height of each cell and the area graphs stretch out vertically [FIGURE 6.22].

The *cells* are uniformly *spaced* to separate the individual area graphs from each other and provide a graphical distinction of cell shapes. The user can compare various panels to find differences and similarities between them. On the other hand, the horizontal and the vertical *grid unit* values are equal for both spacing and axis, what gives a visual continuity to the data from adjacent cells. This means that the *white space* unit between the panels along rows and columns is equal to the horizontal and the vertical *axes units* of a cell, respectively. This aspect is applied to both reading views.

The application's window is maximized to utilize all available space to display the panels. The visualization has a *modular grid*, which modules are *cells*. The *cell size* is equal for all panels to provide a proportional data mapping. The graph's columns and rows are *labeled* accordingly to the *aggregation levels*. The cells has a particular grid based on the currently represented values. Both axes are have small white-colored *grid ticks* showing the division units and providing the quantitative context. The data aggregation axis has a fixed amount of ticks, depending on the aggregation level, with the first and the last value aligned to the edges of the axis. The data value axis is divided in 10 ticks. Their scale varies accordingly to the current range of the chosen values. That scale is the same for all displayed cells representing the chosen dataset.

AGGREGATED USER CONNECTIONS  
UNIVERSITY OF COIMBRA  
since 01 Sep 2010 until 31 Aug 2011

PANEL MATRIX:  
Columns - aggregation by months  
Rows - aggregation by weeks

CELL:  
Vertical axis - amount  
Horizontal axis - aggregation by days

MONTHS   WEEKS   DAYS  
WEEKS   DAYS  
DAYS   HOURS

CONNECTIONS   TRANSITIONS  
TIMES   AMOUNTS  
USERS  
RECORDS



UC   UL   UM   UP  
2010-2011   2011-2012

The *user interface* area is situated at the left side of the graph [FIGURE 6.23]. At the top, the *contextual information* indicates the type of the displayed aggregated data, the name of the dataset and the covered temporal period. Then follows the explanation of the data, describing the data type shown in the columns and rows of the panel matrix, as well as the information displayed in each cell's axes.

Beneath the text, the button controllers are shown. The buttons form four blocks grouped by purpose: the first block is used for the selection of displayed data aggregation type, the second block is for the selection of the data values, the third block is for the manipulation of the visualization, and the fourth block is for the dataset picking. All buttons have the same interactive principles described in the two previous sections.

The *aggregation type* controllers consist of three columns of buttons. Each column controls one level of aggregation and has three options. The desired aggregation type can be chosen selecting one option for each level, starting from the first column. When the first option is selected,

Figure 6.23. The user interface of the application.

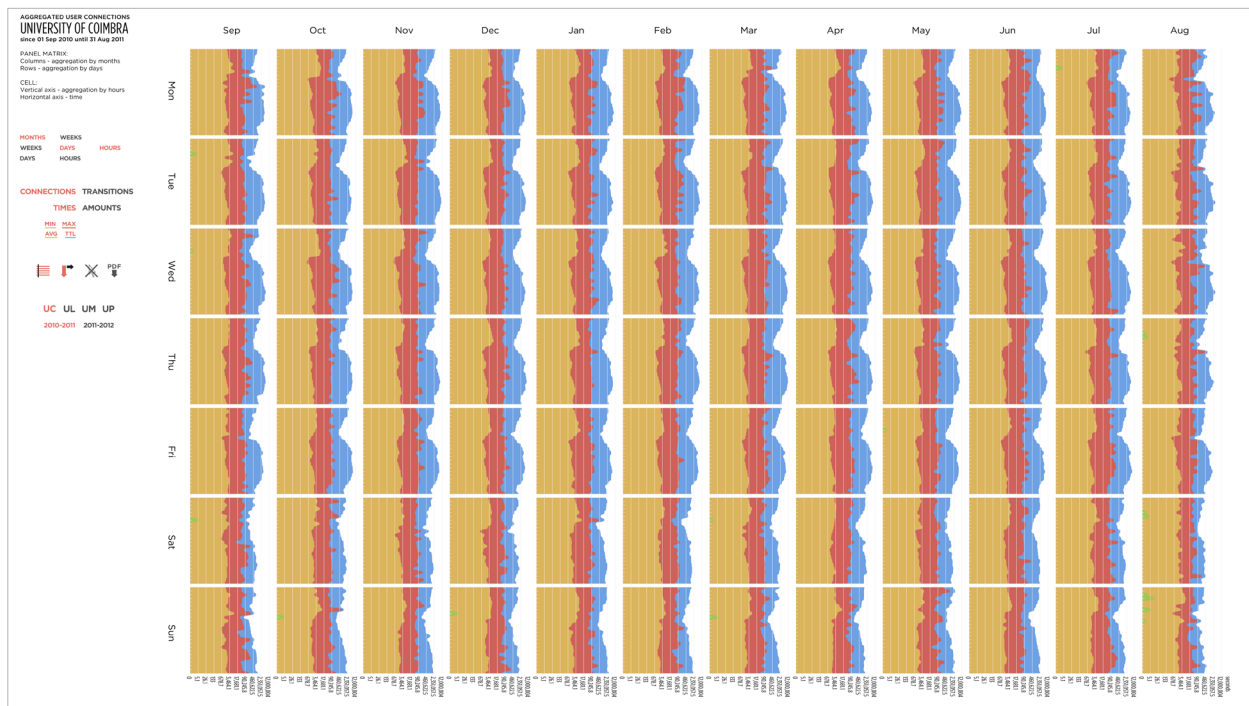


Figure 6.24. Screenshot of the small multiples showing the connection times aggregated by weeks, days and hours, with the vertical reading view.

the available types are shown in the second column. Selecting the second level of aggregation reveals the possible options of the third level. After choosing all three successive types, the respective dataset is loaded and the graph is updated.

The controllers for *data values*' selection form the next group. The selection of *data category* (the first line), has two buttons: one for connections' and another for transitions' data. The user can select only one of them. The next line refers to the *data type* with two options: times and amounts. These two data types have different scales and unit measurements: seconds for the times, records for the amounts. Thus, only one of them can be active at once. If the "Times" button is activated, the available aggregated connection (or transition, depending on the selected category) time fields appear below. There are four different *fields*: minimum, maximum, average and total time. The user can select one, two, three or four fields to be presented simultaneously. In such a case, the area graphs form *superimposed layers*. The cell's *grid scale* updates accordingly to the *biggest value* of the active fields in order to fit all data in the cell's size. The *layers are colored*: minimum values are green, average are orange, maximum are red, and total are blue [FIGURE 6.24]. The vivid colors contrast on the white background and on the layers themselves, so the data is clearly readable and understandable. The color indicator of an activated field appears below the button as a line with the respective color. The "Amounts" button presents a similar behavior, but with two options for users and records. The user amounts are blue and the record amounts are red-colored [FIGURE 6.25].

The graph's manipulation controllers are shown below the data values' selectors. This block has four distinct buttons: the "Show/hide grid," the "Vertical reading on/off," the "Deselect cells," and the "Save as image" buttons. In the *active* state, these controllers permit the following functionalities: the first button shows the grid



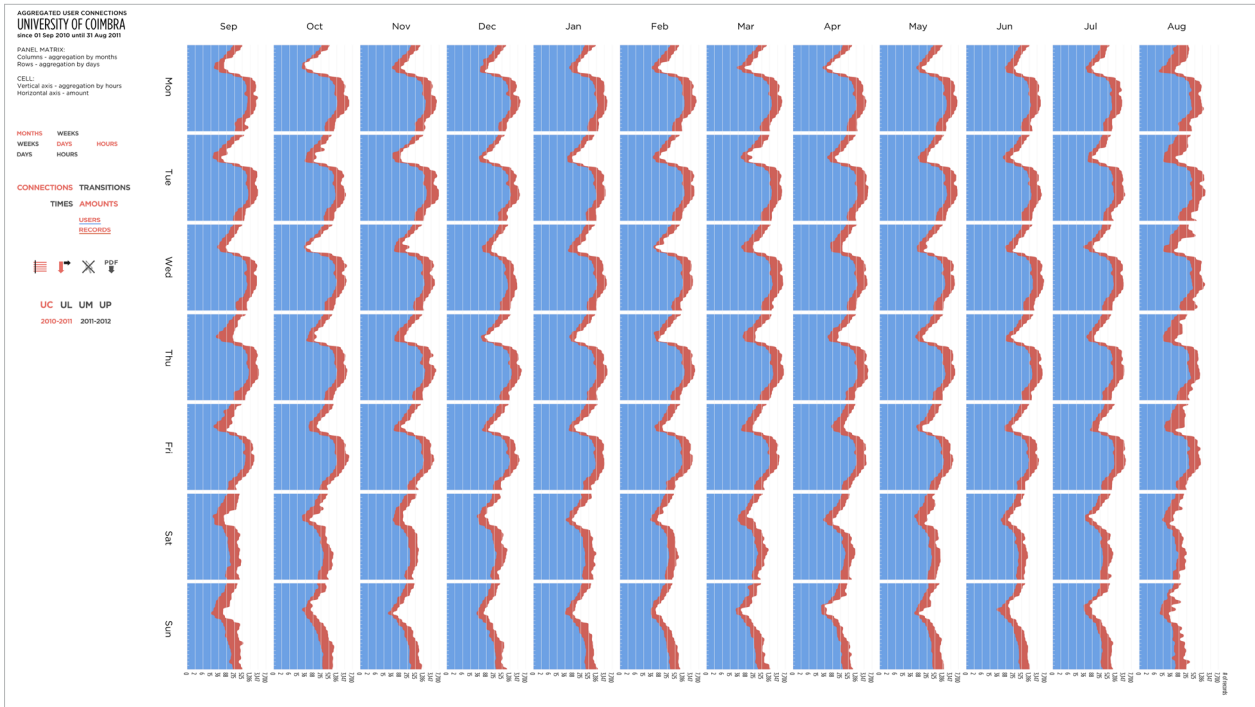


Figure 6.25. Screenshot of the panels displaying amounts of users and connections aggregated by weeks, days and hours; the vertical reading view.

lines along the user rows (in the horizontal reading view) or columns (in the vertical reading view) to provide a visual comparison of values of different panels, and the grid labels appear at the right or at the bottom of the graph, depending on the current type of imposition; the second toggles between the horizontal and the vertical data impositions; the third deselects the active cells (the description follows below); and the fourth takes a screenshot of the graph and saves it as a PDF document.

Lastly, the *dataset selection* controls are presented. When the desired dataset and the academic year are chosen, the files with the aggregated data are loaded to be visualized, analyzed and compared in the same application. From all four button groups, this last block and the aggregation type buttons make the application read different data files. The white loading screen displays the “Loading dataset” message and then shows the current designation of the dataset, when the data is ready to be visualized.

The *interaction* with the visualization is performed by the use of the *mouse cursor*<sup>5</sup>. When the user aims on any of aforementioned button groups, the respective *descriptions* appear above the controllers to explain their purpose. While interacting with the panels, the user can *highlight* any desired cell with the mouse pointer. The *color* of the area graphs becomes more *saturated*, the *labels* appear along the grid ticks, and the *data points* are shown at the mapped positions. These points are small *circles* of the same *color* that the shape these belong to, and with the *low brightness* value to make them more noticeable. The data points provide a clear localization of the values of the displayed data [FIGURE 6.27]. When the user points the mouse cursor on a data point, the respective *exact value* appears above the cursor in the numeric format (*details-on-demand technique*). If the user performs a left click on a cell, its data

<sup>5</sup> The video that demonstrates the interaction with the application can be seen through the hyperlink: <https://vimeo.com/72990693>

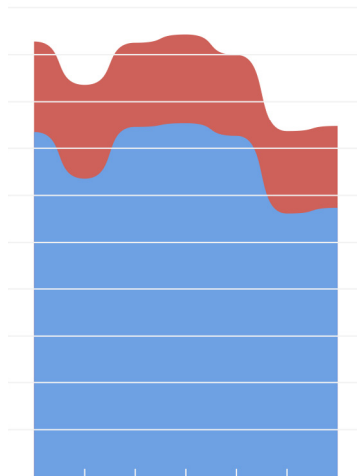


Figure 6.26. A deselected panel from the visualization.

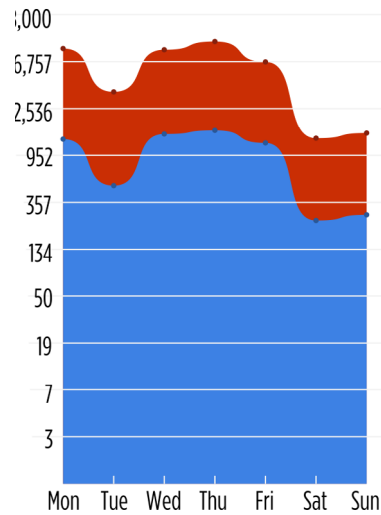


Figure 6.27. A highlighted panel.

stays highlighted. This *direct selection technique* allows a more precise comparison of the desired cells, since these stand out from the inactive panels. The cell's *deselection* is also made with the left click on it [FIGURE 6.26]. The “Deselect cells” button, described above, immediately deactivates all panels, sparing time. The cursor's *icon* changes while interacting with distinct types of elements: the default icon is an *arrow*; while over the buttons or panels, the icon changes to a *hand*.

Similarly to the previous applications, the “Save as image” button hides all interface controllers, leaving only the data visible. Since the colored underlines disappear with the buttons, the *color legend* is included in the PDF document during the saving process. It consists of a list of names of the active layers, accompanied by the small colored square on the left, situated below the contextual information [FIGURE 6.28].

The small multiples applied to the network data demonstrate peaks, falls and rhythms in the connectivity and movement of users in these wireless networks. The amounts of accesses and transitions can be interposed with the total of active users, and the different amounts of time statistics can be compared to understand the charge of a network along academic years. This is achieved due to the representation of the data aggregated by diverse temporal units. Therefore, *patterns of human behavior* can be tracked, providing visual analysis for the panel shapes of this visualization. A variety of options is offered to display data from distinct points of view.

## AGGREGATED USER CONNECTIONS UNIVERSITY OF COIMBRA since 01 Sep 2010 until 31 Aug 2011

### PANEL MATRIX:

Columns - aggregation by months  
Rows - aggregation by weeks

### CELL:

Vertical axis - amount  
Horizontal axis - aggregation by days

### Color legend:

■ Amount of users  
■ Amount of connections

Figure 6.28. The contextual information and the color legend of an exported graph.

Also, two graphic impositions can be applied to the same information in order to reveal hidden *patterns of data* and look at information in both *continuous and discontinuous ways*. The grid plays an important role in quantitative reading, localizing the panel shape's edges in a concrete value scale. The current state of the graph can be saved as an image when necessary.

This section introduced the visualization of multiple panels of the network data, created in the context of this research. The dynamic matrix presents the information aggregated by time units and divided in individual area graphs. The multivariate data is distributed by columns, rows and cells, displaying values of the same type on layers of each panel: the *primary data aggregation level* is shown in columns, the *secondary aggregation level* is shown in rows of the matrix, the *tertiary aggregation level* is presented on one axis of cells; the *data values* are mapped on the second cell's axis, accordingly to a logarithmic scale of base 10; the total of *six variables* can be selected to be displayed as superimposed area graphs in each panel. The side bar shows contextual information and user interface controllers. This application clearly shows the patterns of usage of wireless networks. The panel shapes reveal highs and lows in amounts or times of connections and transitions with smooth curved lines. The graphic results obtained with this application are analyzed in SECTION 7.3.

## 6.4. The Map of the Wireless Network

The fourth and final visualization, created within this research, portrays the wireless network as an organic system. We opt for a *force-based layout* to represent the physical structure of the network. The capabilities of spatial distribution of force-directed graphs provide a map-like visual organization of the data, creating a *network imposition*. We designate this technique as the *relative geographic approach*. Relative because the positions of the APs are not geographically precise, since we do not possess the information about their localizations. We determine these positions by placing pairs of APs and clusters at the distance that corresponds to the transition time between them.

In this application, we use the ordered node lists [SECTION 5.1, STEP 18, PP. 80-81] to load the clusters' information, the sets of processed transitions [SECTION 6.1, PP. 93-94] to determine the relative positions of APs, and the specifically created datasets with pairs of APs to represent the network's users. The *user data* comes from the files based on the connection sets ordered by user ID. We collect the unique (without duplicate entries) pairs of APs that each user has visited. Then, we create a file containing the user ID, the user name and the list of AP pairs for every user and for each dataset. We omit the transitions within the same node, since these records are not relevant for this visualization. We show the flow between distinct APs with this graph, not the flow inside the same AP. If we kept the self-node transitions, the node sizes (see the explanation below) would give a wrong idea about the overall accessibility in the network.

Similarly to *The Circuit of Transitions*, we create a system that consists of *nodes* interconnected with transitional *links*. The major difference is the implementation of the principles of Physics to simulate behaviors directed by forces. Additionally, we introduce the *users* as a part of the network system, to be visualized and explored at the same level of importance as the graph's nodes and edges. We use an external open source library called *toxiclibs* (SCHMIDT, 2007) to implement the force-based system. That library provides implemented methods for the development of an application with a physics engine.

In this visualization, the *clusters* are electrically charged *particles* represented by *white circles* with *colored stroke*. The clusters are initially positioned along a circumference and their colors are assigned accordingly to these localizations, as was explained in SECTION 6.1. The chosen vivid colors provide a *contrast* with the white background. The *perimeter* of the node's circle is proportional to the number of its *incoming and outgoing transitions*. Thus, a visual hierarchy is established between the less and the most accessed clusters. Each cluster has a *repulsion force* that activates in the nearby presence of any another cluster. That force acts on both nodes and *repels* them from each other, separating them visually [FIGURE 6.29].

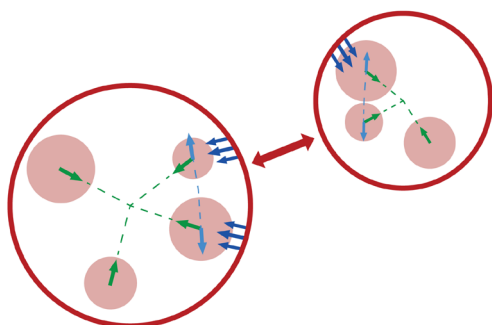


Figure 6.29. Forces affecting nodes: green – the attraction inside clusters; light blue – the repulsion between child nodes; dark blue – the repulsion from the cluster's border; red – the repulsion between clusters.

The cluster's *APs* are semi-transparent *particles*. An AP consists of a *circle* that inherits the *color* of the father node and moves *inside* its shape [FIGURE 6.29]. Both clusters and APs are *associative* visual variables. A child particle is affected by two forces: an *attraction force* with a small strength coefficient and a *repulsion force* with the strength coefficient proportional to the father's diameter. The first constantly *pulls* the particle to the cluster's *center* in order to

maintain it inside the father's area. The second is activated on collision with the father's borders or with another nodes that belong to the same cluster. It *pushes* the nodes away from each other and from the cluster's borders to provide a spacing to all children particles. If a cluster has only one child, that cluster is represented by the child node (color-filled circle). In this case, the node presents the cluster's behavior.

The *transitions* are *mechanical springs* that connect the source-target pairs of clusters. Initially, the springs are invisible, but can be shown using the respective button in the interface. While visible, an *edge* is represented by a straight *line* that inherits the *color* of the node of origin. The line's *opacity* is 60% reduced that reduces the global visual load. The *thickness* of each edge is proportional to the *quantity of transitions* grouped in that edge – thick lines represent the most common routes [FIGURE 6.30]. Bidirectional transitions are represented one above another and can be differentiated by the thickness and by the color. It is important to mention that node length and edge thickness parameters depend on the *linear scale factor* to represent the sizes

proportionally to each other: the node length is the sum of thicknesses of all edges that belong to it. That scale factor is multiplied by a small constant value to reduce the global size of elements and to avoid very large shapes.

Two main parameters control the operating principle of an edge: length and stiffness. The *length* is a *distance* between the both connected nodes. We assign the *average transition time* to the edge lengths to achieve a more accurate equilibrium between the close and far located clusters. We divide the time value on a constant number to reduce the length of very long edges to make them fit into the application's window. Also, we defined the constants for different sets, since the data varies from one university to another. The *stiffness* defines the *hardness* of the spring. The smaller values make the connection physically more rigid, the bigger values make it more flexible. Rigid springs take less time to stabilize, but have a bigger tension coefficient. This reflects on the entire system's behavior – tense springs make the highly connected nodes oscillate because the links with imbalanced lengths are in constant pressure. Soft springs take more time to stabilize and present less tension. These behave smoother but tend to expand or enlarge more easily under the effect of all other springs.

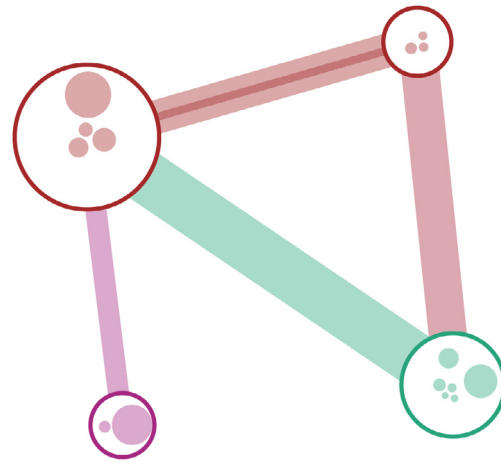


Figure 6.30. Five transitional edges interconnecting four clusters.

The main point is to find a *certain proportion* between these two properties and drive the system's force to a balanced state, minimizing the overall strain of the edges and the visual deviation of the nodes' spatial positions. Since the edge length values come from the data, we adjust the stiffness value, assigning 35% of softness to every edge in the displayed dataset. This value presents a suitable hardness/softness relationship, making the nodes move and stabilize in a fast and smooth way and preserving their relative localizations regarding every other node. When the application starts running, the positions of all clusters, firstly distributed on a circumference, are actualized accordingly to the length and the stiffness of the edges connecting them.

The network's *users* are *particles* with a particular behavior. These are displayed with *grey circles* with a *small* and fixed *size* to standardize their appearance. Each user appears near the AP/cluster he was connected to in the lastly registered connection. When reaches the target, the user starts *moving around and outside* it due to the *attraction force* of the node. This movement is granted by the implemented *collision checking method* that pulls the user back every time it collides with the node. After the shock, the user particle's *direction* is inverted and rotated by a randomly generated angle with small amplitude. This procedure allows a uniform distribution of the user particles around a cluster or an AP. As the node's position changes, the user follows it in the space [FIGURE 6.31].

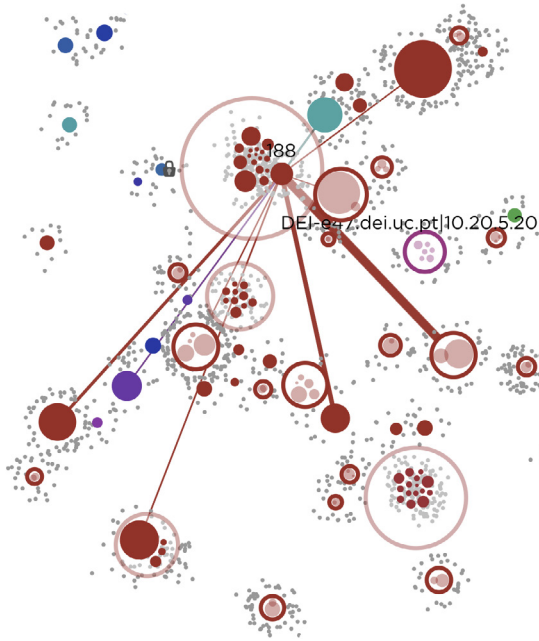


Figure 6.31. Detail of expanded clusters, APs and users.

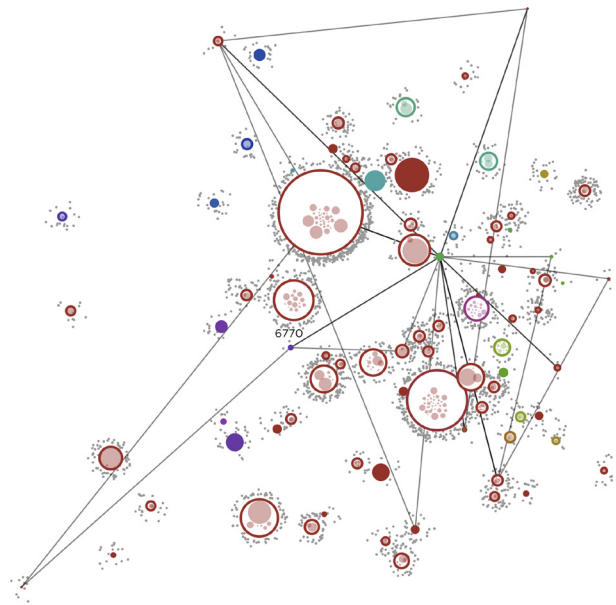


Figure 6.32. The network path of the user #6770.

The *interaction* with the visualization is performed with the *mouse cursor*<sup>6</sup>. Pointing the cursor on a cluster makes appear its *ID number* (in the bold type weight) above the circle and the *amount of children* below the mouse pointer. The incoming and outgoing *edges* of that cluster become *visible* and *highlighted* with a bigger opacity (*details-on-demand technique*). The user can *expand* a cluster double-clicking on it. The *contour* and the *background* of the expanded cluster become *semi-transparent*, while the *children* become *opaque*. The *name*, the *ID* (in the regular type weight) and the incoming and outgoing *transitions* of a child can be visualized by pointing the cursor on the child node's circle. When a cluster is *expanded*, the user particles connected to it relocate to the respective children APs [FIGURE 6.31]. Additionally, a cluster can be *affixed* in the space with a left-click on its shape. This functionality prevents the node from moving under the force of springs and allows an easier *exploration* of the cluster's children and edges. The *lock icon* appears at the left side of the stuck cluster. The *detachment* is also made with a single click, and the lock icon disappears. When the mouse cursor points on a user circle, the user's *ID number* appears above the particle and the performed *network path* is shown. That path consists of tracing semi-transparent *thin black lines* between visited pairs of APs/clusters [FIGURE 6.32]. The *darker lines* mean that the user has transited in both directions between the pair of nodes. The appearance of the mouse cursor's *icon* changes when interacting with different types of visual elements: by default, the icon is an *arrow*; while pointing on expandable or lockable nodes, on the user particles and on the filter bars or buttons, the icon changes to a *hand* image.

The header of the application contains the *contextual information* and the graphic interface controls. The information is divided in two parts. The left part shows the title “Network User’s Flow,” the designation of the current dataset and the covered time period. The right part presents the total quantities of users, of transitions

<sup>6</sup> The video presenting the interactive functionalities of the application can be watched through the hyperlink: <https://vimeo.com/72999453>



Figure 6.33. The filtering bars.



Figure 6.34. The button controllers of the application.

grouped in edges, and of APs grouped in clusters. This second section has a low opacity that makes it less noticeable. The transparency can be removed by pointing the mouse cursor over the text.

At the right side of the header, the *interface controllers* are displayed: two filtering bars, the “Show/hide node IDs,” the “Show/hide all paths,” the “Unlock all nodes,” the “Pause/resume movement,” and the “Save as image” buttons. The interaction principles with these controllers are explained in the previous sections. The first filtering bar is for the elimination of the edges, the second bar is for the removal of user particles [FIGURE 6.33]. From the left to the right, the buttons in the *active* state have the following functions: the first button displays all identification numbers above the active nodes; the second makes visible all existing routes between active APs and clusters; the third unlocks all locked nodes at once to spare the time; the fourth pauses the movement of the clusters, their child nodes and the user particles to simplify the interaction with these elements and to allow an efficient exploration of the data; the fifth button takes a snapshot of the graph along with the contextual information and saves it in the PDF format [FIGURE 6.34]. The *dataset selection* controls are also present. The interaction with all these sections is the same as in all previously described applications. In the lower-left corner of the application’s window, the contextual *length scale bar* shows the relation between the displayed length of the edges and the temporal proportion in seconds.

The datasets with transitions and users have a great number of records that do not present relevant amounts of data, when grouped. The visual elements related to that data produce a visual clutter with lines and circles. Mostly, these elements prejudice the reading of the data and drastically reduce the application’s frame rate. In order to present the significant information and to improve the performance and the response time of this interactive visualization, the transitions’ and the users’ datasets are pre-filtered. For transitions, the filtering limits are defined in such a way that the displayed quantity of edges varies from 2,000 to 4,900 grouped transitions, depending on the dataset. For example, the UL set contains 2,000,000 of records, and the filtering below the threshold of five transitions in an edge reduces the quantity of grouped edges from 24,000 to 2,000. In case of larger files, for example the UM dataset with 5,000,000 of transitions in the academic year 2011-2012, the filtering below 60 reduces the quantity of edges from 120,000 to 4,900. The same process is applied to the users’ AP pairs data. After removing the users with small amounts of visited unique AP pairs, we obtain a variation from 2,000 to 3,400 of users, depending on the dataset. Thereby, the data loss is not high, the performance is enhanced and we also *maximize the data-ink*. The respective filtering limits are shown in the filtering bars, after selecting a dataset in the application.

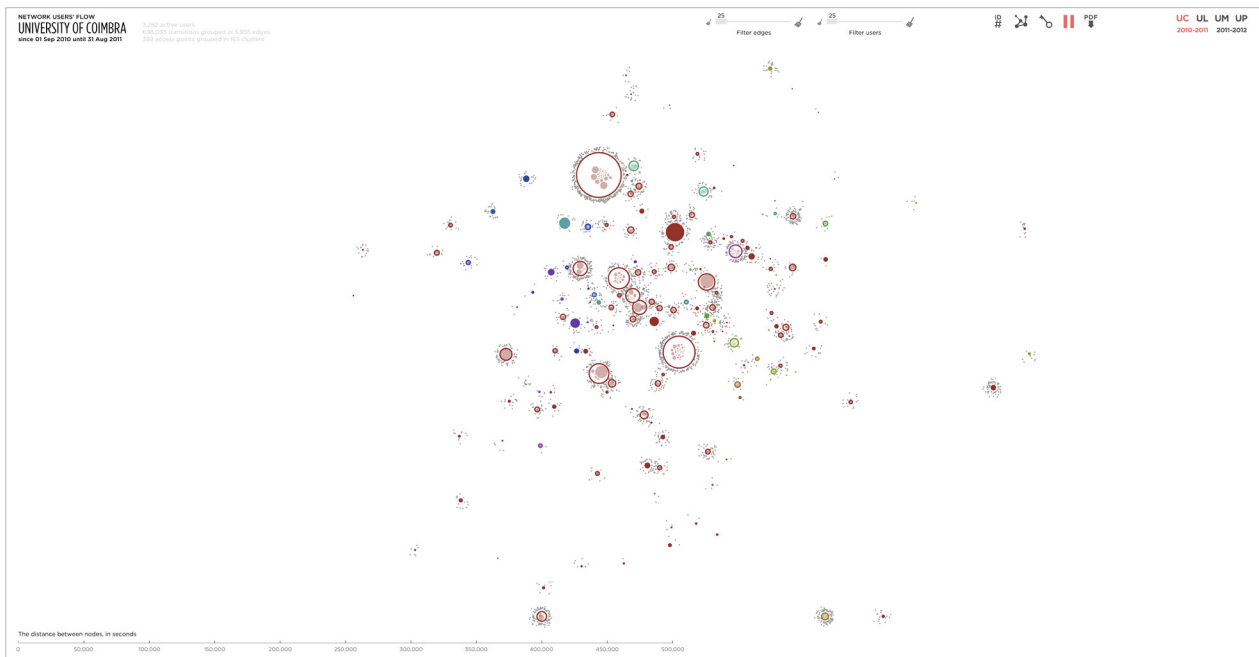


Figure 6.35. Screenshot of the application displaying the physical structure of the UC's network.

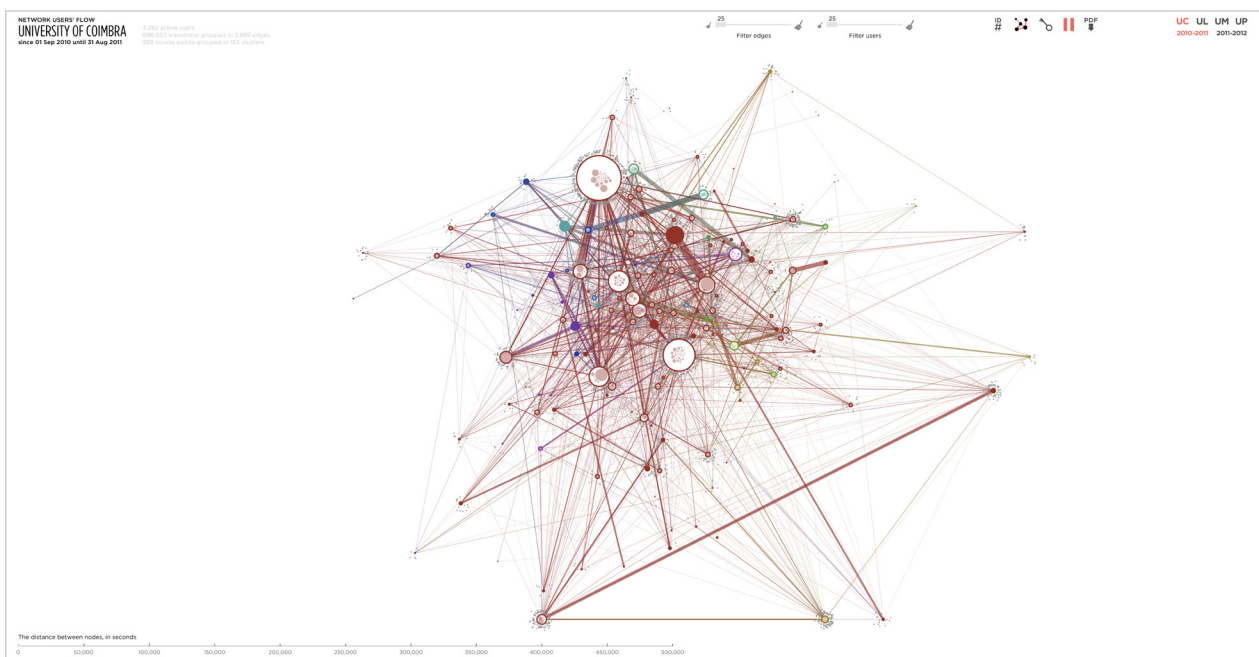


Figure 6.36. Screenshot of the application showing the transitional edges between the nodes.

The force-based visualization depicts the network as a system with behaviors inspired by the laws of Physics. Graphically, these behaviors provide an organic appearance to the whole graph and its individual elements. The elements, such as nodes, edges and users, interact with each other using forces with different effects and strengths. The clusters visually group the APs that auto-distribute themselves within the cluster's shape. The user particles are attracted by the active nodes, depicting the connection density [FIGURE 6.35]. The node sizes indicate the most accessed points, and the thickness of the transition edges demonstrates the common



and the uncommon routes between clusters and APs. The spatial proximity between clusters is achieved with the mechanical force of edges, the pairs of closer clusters represent the relative neighborhood based on the average transition time between these nodes [FIGURE 6.36]. The interactive possibilities allow the exploration of the information on both local and global levels.

The interactive force-directed graph of the network flow is described in this section. The visualization displays multivariate data: the *clusters* are distributed in space under the action of the mechanical springs that represent grouped *transitions*; the *APs* grouped in a cluster are shown inside the cluster's shape; the *total of incoming and outgoing transitions* of an AP or a cluster is represented by the length of the node's circle; the *amount of transitions* grouped in an edge is determined by the thickness of its line; the value of the *average transition time* is assigned to the length of the edge's spring; the *origin of a transition* is displayed by the color of an edge. The network *users* are represented around the lastly connected AP/cluster. The application features the contextual and numerical information at the top and the length scale at the bottom of the window. The *physical structure* of the network is shown by the clusters of APs and by the APs themselves, the *human behavior* is displayed by the user particles. The visualization highlights the *common movement routes* within the network with the thickness of edge lines. The *most accessed points* have a bigger size. The description and analysis of the obtained visual results, generated with this visualization approach, are presented in SECTION 7.4.

## 6.5. Summary

The conception of the visualizations of patterns of flow and usage in wireless networks are presented in this chapter. The transitions and connections of network users, the network's inner flow and usage and the physical structure are shown. The interactive graphs allow the perception and the exploration of the displayed data through the enabled interactive functionalities. The process of construction of each application is described in detail, revealing the complexity of the programmatic structure and the range of the included multivariate data. Four distinct visualization approaches are created, each one featuring pertinent graphic layout and properties that emphasize the specific aspect of the available data. The next chapter demonstrates and analyzes the graphic results obtained with the developed visualizations.



## CHAPTER 7.

# Experimental Results

This chapter presents the graphic artifacts obtained with the four developed models of visualization. We visualize distinct temporal intervals of the academic years of four Portuguese universities in order to find similarities and differences in patterns of flow and usage of wireless networks. Each graph is accompanied by an analysis of the displayed data and an explanation of the variation of the given result. The images demonstrated below can be seen in the APPENDIXES section, enlarged for a better reading.

### 7.1. The Transitionality in Networks

In this section, the static and animated graphs generated with *The Circuit of Transitions* are shown and described. First, we observe the sets with the complete academic year data that loads by default in the application.

The visualization on FIGURE 7.1 covers the full academic year of 2010-2011, since 1 September 2010 until 31 August 2011. 380 APs (from the possible 433) are active and grouped in 164 clusters. The majority of clusters has a relatively small number of child APs (from 2 to 10) and many clusters have only one AP. Their lengths vary, and we can observe very small nodes in every part of the graph's circumference. This variation shows that several APs/clusters were accessed more times than others. Three big clusters stand out from the rest of the nodes: the cluster #31 with 32 APs, the cluster #17 with 30 APs and the #46 with 14 APs. Their lengths indicate that these clusters also have a big amount of incoming and outgoing transitions.

2,347,725 transitions were performed in that period. It is notable that the transitions within the same node prevail over the transitions between distinct nodes. The extremities of the curves outside the circle are larger than the extremities of

USER TRANSITIONS DATA  
UNIVERSITY OF COIMBRA  
since 01 Sep 2010 until 31 Aug 2011

2,347,725 transitions grouped in 1,048 edges  
380 access points grouped in 164 clusters

Transition time stats within the same AP (in sec):  
MIN = 233 MAX = 217,746 AVG = 28,026  
The most common inner-node route:  
cluster 22 → cluster 22 with 40,820 transitions

Transition time stats between distinct APs:  
MIN = 391 MAX = 424,411 AVG = 49,950  
The most common outer-node route:  
cluster 71 → 175 with 6,474 transitions

The most accessed AP:  
cluster 17 with 525,572 accesses

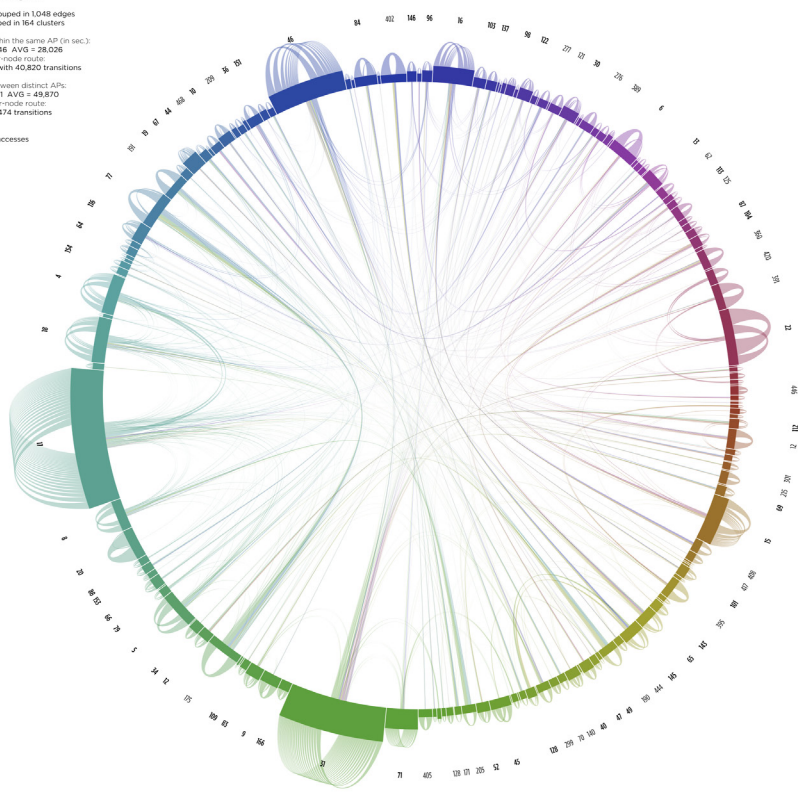


Figure 7.1. The visualization of user transitions in the wireless network of the UC regarding the academic year 2010-2011.

USER TRANSITIONS DATA  
UNIVERSITY OF COIMBRA  
since 01 Sep 2010 until 31 Aug 2011

2,347,725 transitions grouped in 1,048 edges  
380 access points grouped in 164 clusters

Transition time stats within the same AP (in sec):  
MIN = 233 MAX = 217,746 AVG = 28,026  
The most common inner-node route:  
425 → 425 with 40,820 transitions

Transition time stats between distinct APs:  
MIN = 391 MAX = 424,411 AVG = 49,950  
The most common outer-node route:  
cluster 71 → 175 with 6,474 transitions

The most accessed AP:  
cluster 18 with 111,152 accesses

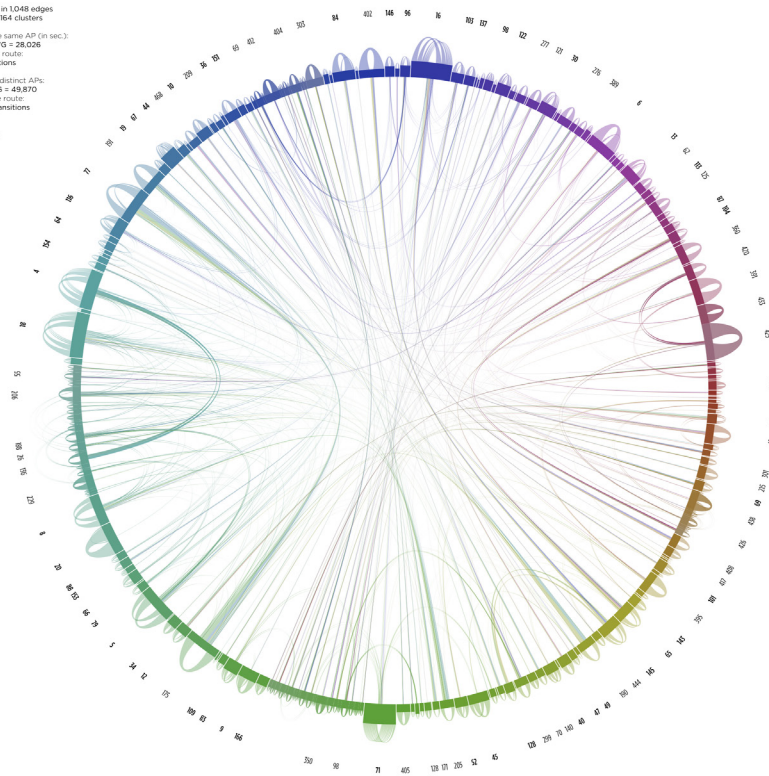


Figure 7.2. The graph with the top five clusters expanded, the UC 2010-2011 dataset.

the inner edges. The easiest way to see the difference is to look at individual nodes. The size of the bunch of edges inside the circle occupies a small fraction of the length of a node, because the rest of the length belongs to the self-node transitions. This means that the users move less in the space, while connected to the network. The graph displaying only the outer node transitions is shown on FIGURE 7.3. Outside the circle, there can be seen a hilly relief formed by the curves. The height variation tells that the average transition time of the edges varies from node to node. Inside the circumference, one can perceive the ramified edges connecting distinct nodes with different average transition times. The closely located nodes are connected with edges of lower height that indicate their relative proximity, obtained with the sorting algorithm. The most common inner-node route is clearly visible: the thick red curve inside the cluster #22. The same cannot be said about the outer-node routes, but the contextual information indicates that route: from the cluster #77 (blue color) to the AP #175 (green).

If we expand the top five biggest clusters [FIGURE 7.2], the transitions between their child APs can be observed. For example, the cluster #17 has a rich routing within the APs it groups. Also, these nodes have many transitions outwards the cluster. The gradient color of the graph's circumference indicates that the distribution of nodes is uniform and a small amount of APs is filtered. The global image of the edges is gradient, too. This means that the majority of transitions from one AP/cluster is performed to the APs/clusters located in the relative neighborhood.

USER TRANSITIONS DATA  
UNIVERSITY OF COIMBRA  
since 01 Sep 2010 until 31 Aug 2011  
445,736 transitions grouped in 630 edges  
245 access points grouped in 136 clusters  
Transition time stats:  
MIN = 101 MAX = 424,471 AVG = 49,870  
The most common route  
cluster 77 → 175 with 6,474 transitions  
The most accessed AP:  
cluster 17 with 65,984 accesses

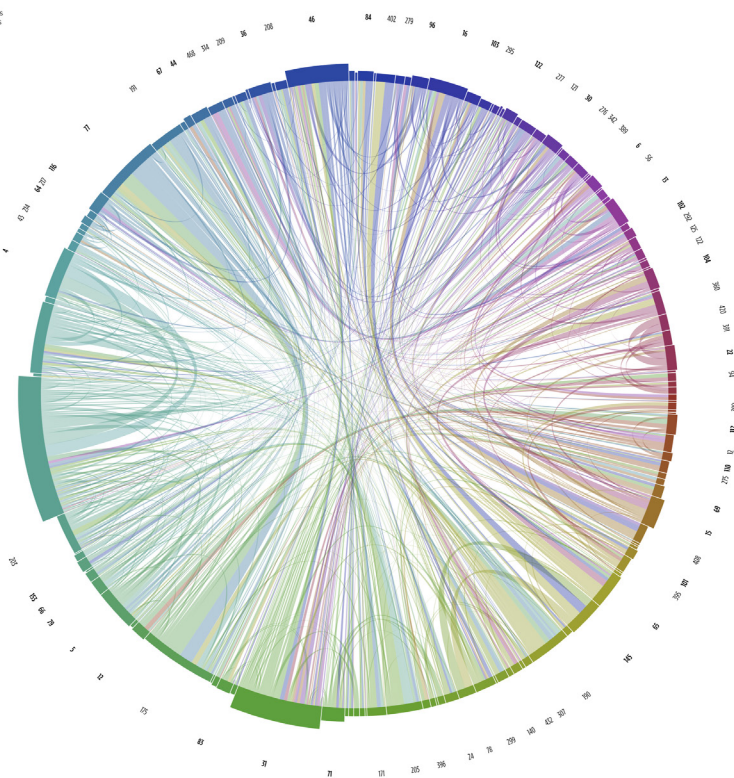


Figure 7.3. The graph showing the outer node routes only, the UC 2010-2011 dataset.

USER TRANSITIONS DATA  
**UNIVERSITY OF LISBON**  
 since 01 Sep 2011 until 31 Aug 2012

1,854,509 transitions grouped in 525 edges  
 348 access points grouped in 160 clusters

Transition time stats within the same AP (in sec):  
 MIN = 2, MAX = 266,351, AVG = 89,714  
 The most common inner-node route:  
 cluster 102 → cluster 102 with 534,356 transitions

Transition time stats between distinct APs:  
 MIN = 121, MAX = 12,701,859, AVG = 505,432  
 The most common outer-node route:  
 cluster 102 → cluster 1 with 19,542 transitions

The most accessed AP:  
 cluster 102 with 1,136,539 accesses

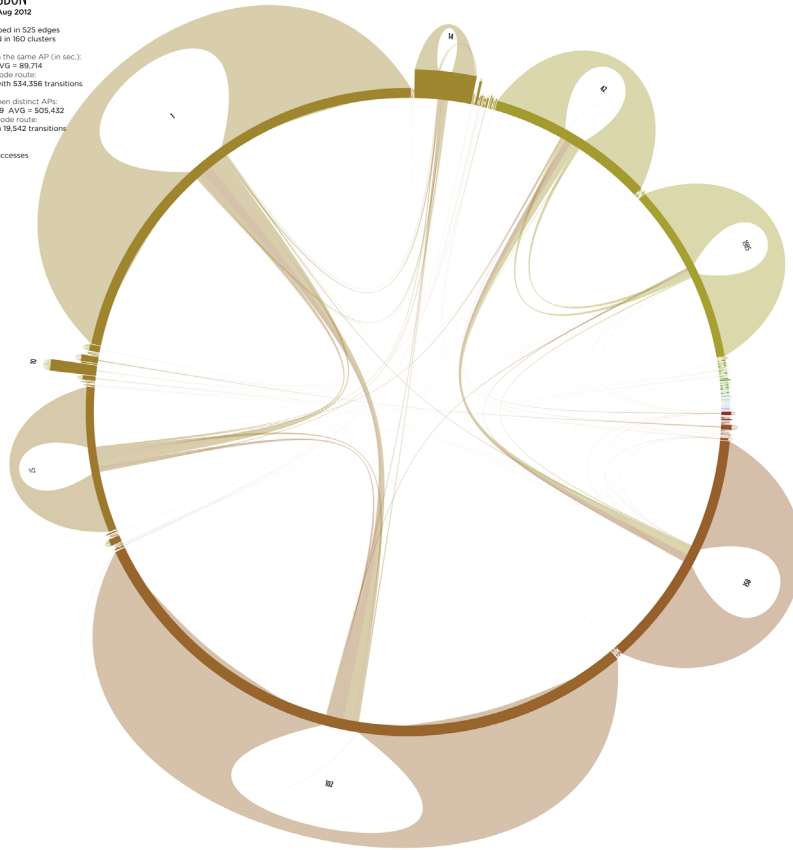


Figure 7.4. The visualization of user transitions in the wireless network of the University of Lisbon, on the academic year 2011-2012.

USER TRANSITIONS DATA  
**UNIVERSITY OF LISBON**  
 since 01 Sep 2011 until 31 Aug 2012

167,813 transitions grouped in 163 edges  
 62 access points grouped in 41 clusters

Transition time stats:  
 MIN = 121, MAX = 12,701,859, AVG = 505,432  
 The most common route:  
 2 → 93 with 19,542 transitions

The most accessed AP:  
 93 with 77,263 accesses

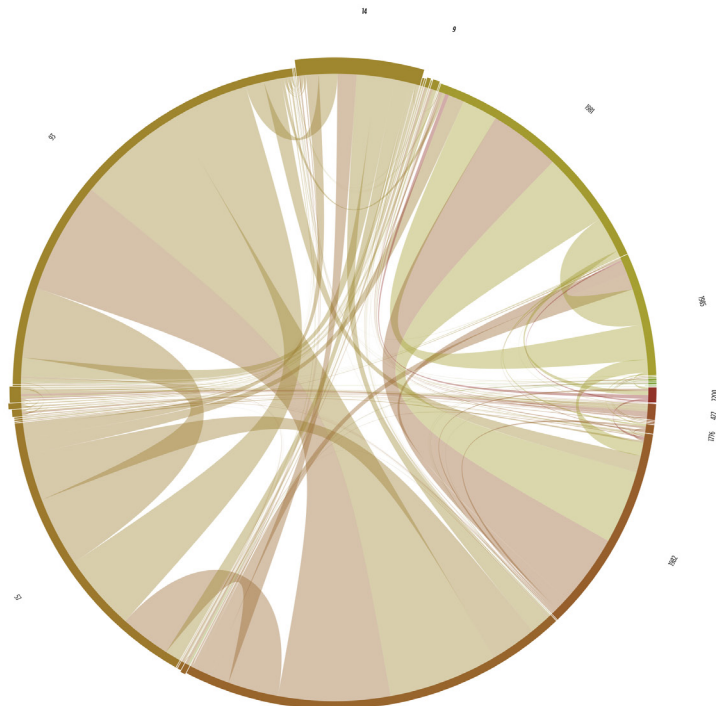


Figure 7.5. The graph displaying the outer routes only, the UL 2011-2012 dataset.

The graphic on FIGURE 7.4 shows the UL dataset on the academic year 2011-2012. The flower-like appearance presents seven nodes with enormous amounts of transitions within themselves. We can deduce that these APs/clusters are the most popular in the UL network, despite having 3,576 APs in total. The users frequently connect to these clusters. There are several transitions connecting these clusters – some users transited between them. It is curious that these highly accessed nodes have small number of APs (from one to three), only the cluster #14 groups 25 APs.

The edges with a grouped amount of transitions below 50 are filtered, and the smaller nodes are gone. We can still see the small nodes at the right, left and top parts of the visualization. For example, the cluster #10 has 47 APs, but very small amount of incoming and outgoing transitions. That cluster can represent a remote private network that belongs to the UL eduroam domain. The visualization on FIGURE 7.5 allows a better perception of the outer node transitions. One can notice that most edges have similar sizes of extremities that indicate a uniform transitionality between distinct clusters and APs.

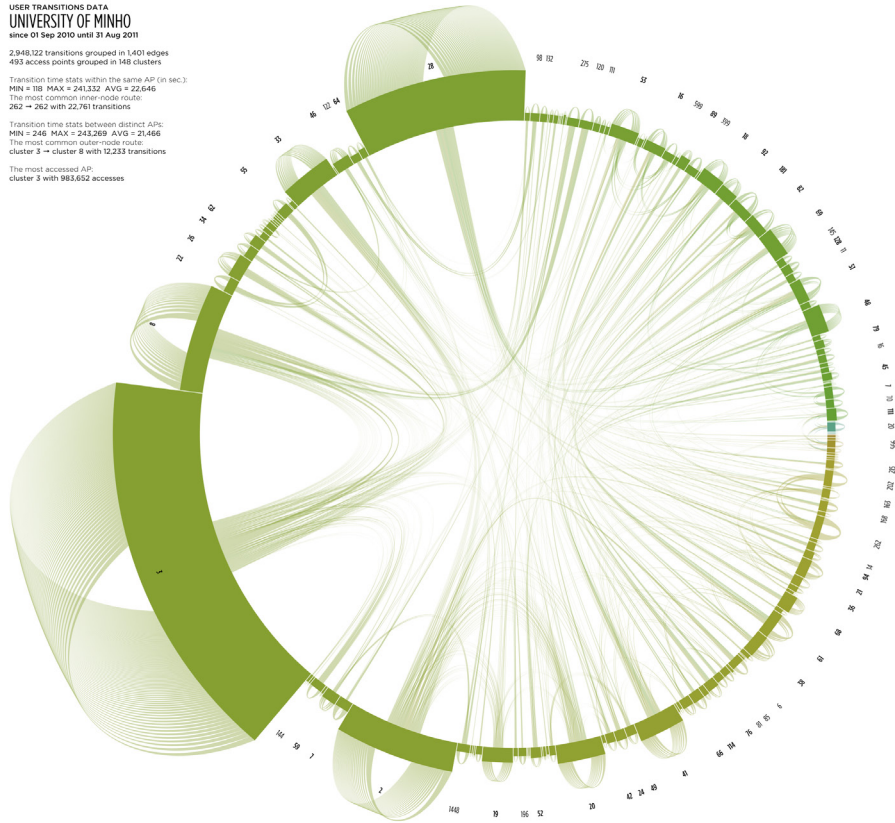


Figure 7.6. The visualization of user transitions in the wireless network of the University of Minho on the academic year 2010-2011.

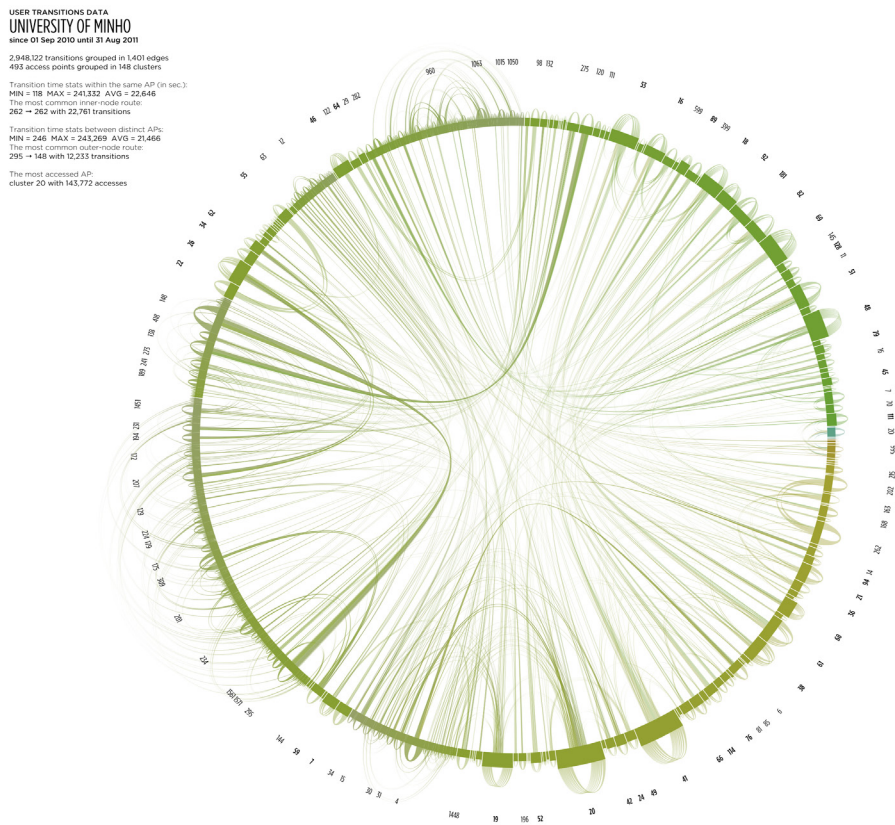


Figure 7.7. The graph displaying the UM 2010-2011 dataset, with the top five nodes expanded.



The visualization of the UM dataset, the academic year 2010-2011, shows a green-hued structure [FIGURE 7.6]. The red and blue-hued nodes, situated at the beginning and at the end of the node list sorted by proximity, have small number of accesses and were eliminated with the filtering procedure. The two huge clusters (#3 with 97 APs and #28 with 52 APs) are the preferred access points of this network. Other nodes are not so common, except the #2 and the #8 that hold the third and the fourth places in this popularity rating, respectively.

The transitions are uniform, like in the UC network. The inner-node routes clearly prevail over the outer transitions. The big clusters are deeply interconnected, while expanded [FIGURE 7.7]. One can observe transitions between the single APs inside the cluster, as well as the APs outward the cluster. This view shows a high variation of amounts and average times in edges, linking all nodes uniformly. FIGURE 7.8 depicts the routes between distinct nodes. The density of curves is very high, as well as the variation of average transition times.

USER TRANSITIONS DATA  
UNIVERSITY OF MINHO  
since 01 Sep 2010 until 31 Aug 2011  
869,062 transitions grouped in 710 edges  
306 access points grouped in 110 clusters  
Transition time stats:  
MIN = 246; MAX = 243,269; AVG = 21,466  
The most common route:  
cluster 3 → cluster 8 with 12,233 transitions  
The most accessed AP:  
cluster 3 with 171,836 accesses

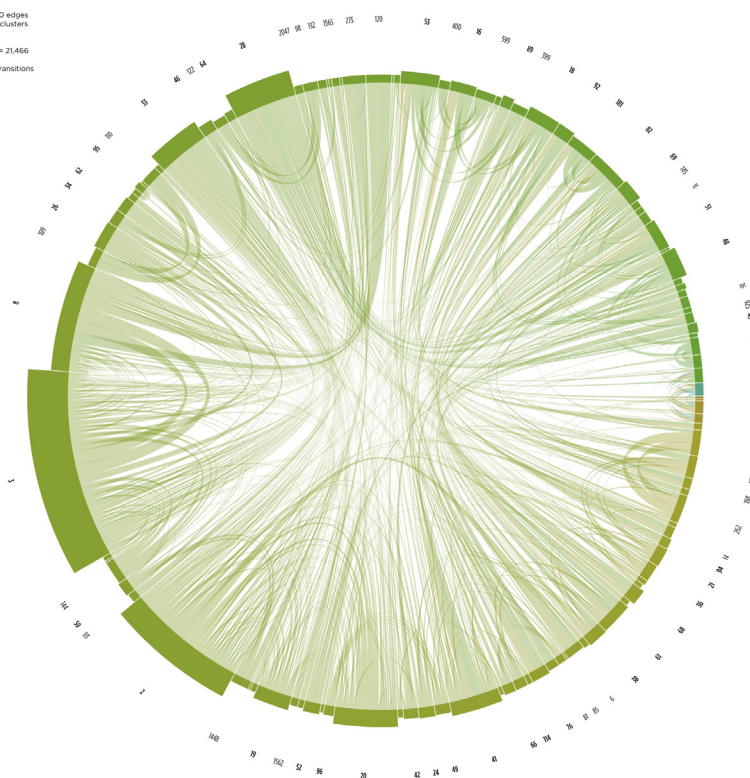


Figure 7.8. The graph portraying the outer routes, the UM 2010-2011 dataset.

**USER TRANSITIONS DATA**  
**UNIVERSITY OF PORTO**  
 since 01 Sep 2010 until 01 Jun 2011

885,292 transitions grouped in 1,253 edges  
 341 access points grouped in 157 clusters

Transition time stats within the same AP (in sec):  
 MIN = 0 MAX = 689,007 AVG = 74,083  
 The most common inner-node route:  
 cluster 86 → cluster 86 with 30,353 transitions

Transition time stats between distinct APs:  
 MIN = 437 MAX = 1,521,277 AVG = 201,141  
 The most common outer-node route:  
 cluster 86 → 26 with 7,951 transitions

The most accessed AP:  
 cluster 9 with 258,668 accesses

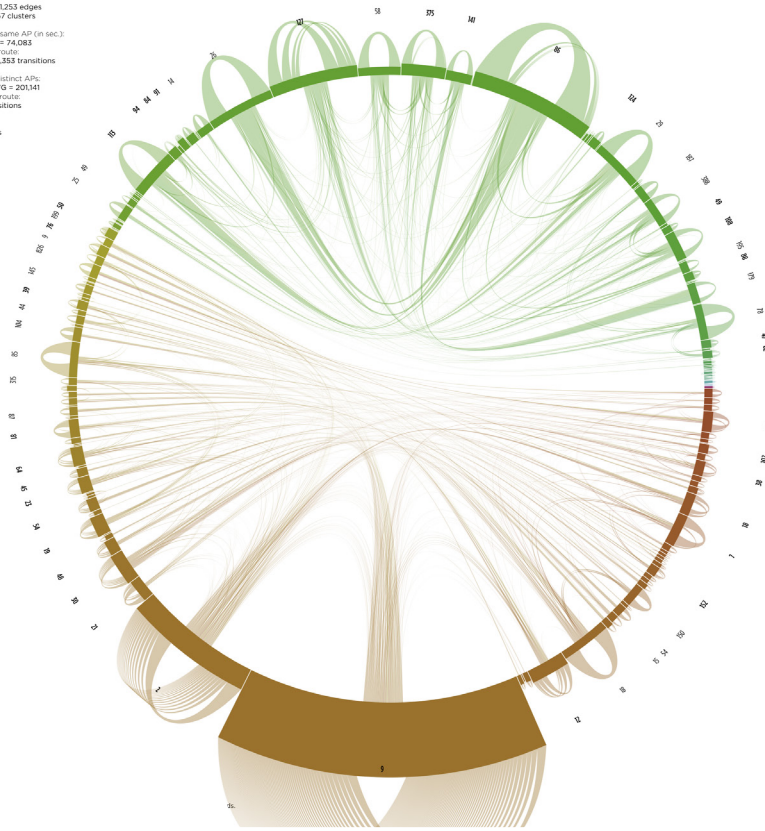


Figure 7.9. The visualization of user transitions in the wireless network of the University of Porto on the academic year 2010-2011.

**USER TRANSITIONS DATA**  
**UNIVERSITY OF PORTO**  
 since 01 Sep 2010 until 01 Jun 2011

885,292 transitions grouped in 1,253 edges  
 341 access points grouped in 157 clusters

Transition time stats within the same AP (in sec):  
 MIN = 0 MAX = 689,007 AVG = 74,083  
 The most common inner-node route:  
 32 → 32 with 30,353 transitions

Transition time stats between distinct APs:  
 MIN = 437 MAX = 1,521,277 AVG = 201,141  
 The most common outer-node route:  
 32 → 26 with 7,951 transitions

The most accessed AP:  
 26 with 63,963 accesses

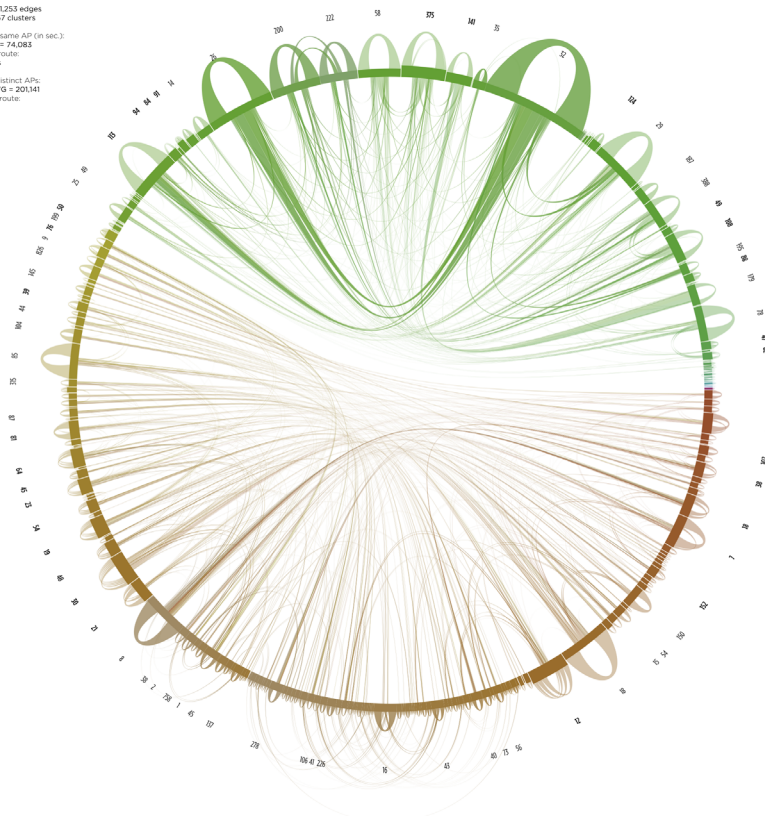


Figure 7.10. The graph displaying the UP 2010-2011 dataset, with the top five nodes expanded.

The visualization of the transitions in the network of the UP 2010-2011 dataset [FIGURE 7.9] presents a different result. The graph consists of two parts: brown and green. We presume that the UL campus has two physically separated regions: there can be seen no transition crossing the invisible border that separates these two sub-networks. The brown part has two commonly accessed clusters (#2 and #9), while the remaining nodes are not so popular. The green part have various common access points (both clusters and single APs) and a less dense distribution of outer transitions between nodes [FIGURE 7.11]. The graph with the expanded biggest clusters also presents a deep interconnection of nodes [FIGURE 7.10], more in the brown part because the green sub-network has a small amount of clustered APs.

USER TRANSITIONS DATA  
UNIVERSITY OF PORTO  
since 01 Sep 2010 until 01 Jun 2011

259,996 transitions grouped in 810 edges  
215 access points grouped in 108 clusters

Transition time stats:  
MIN = 437 MAX = 1,521,277 AVG = 201,341

The most common route:  
cluster 86 → 26 with 7,951 transitions

The most accessed AP:  
cluster 86 with 40,383 accesses

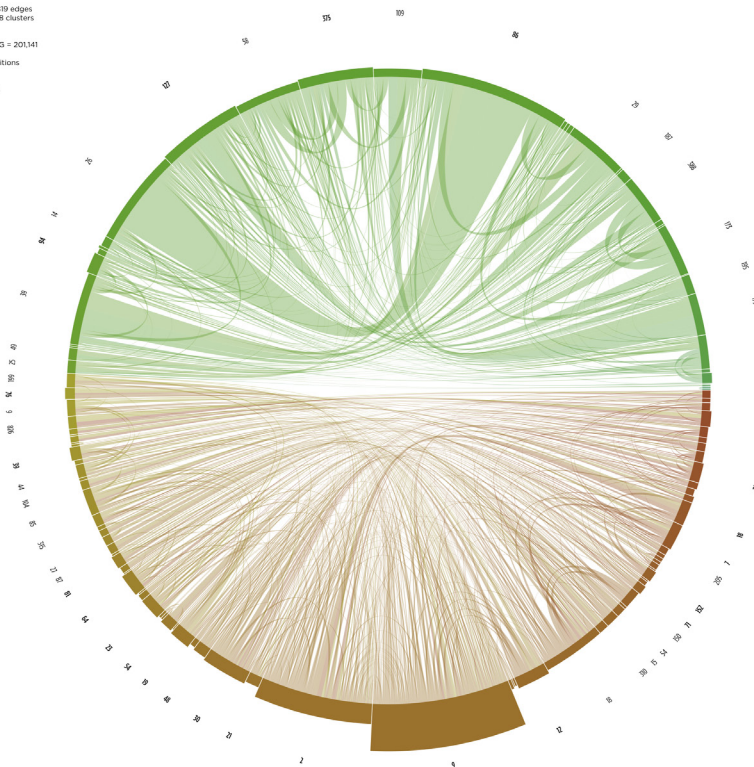


Figure 7.11. The graph depicting the outer node transitions from the UP 2010-2011 dataset.

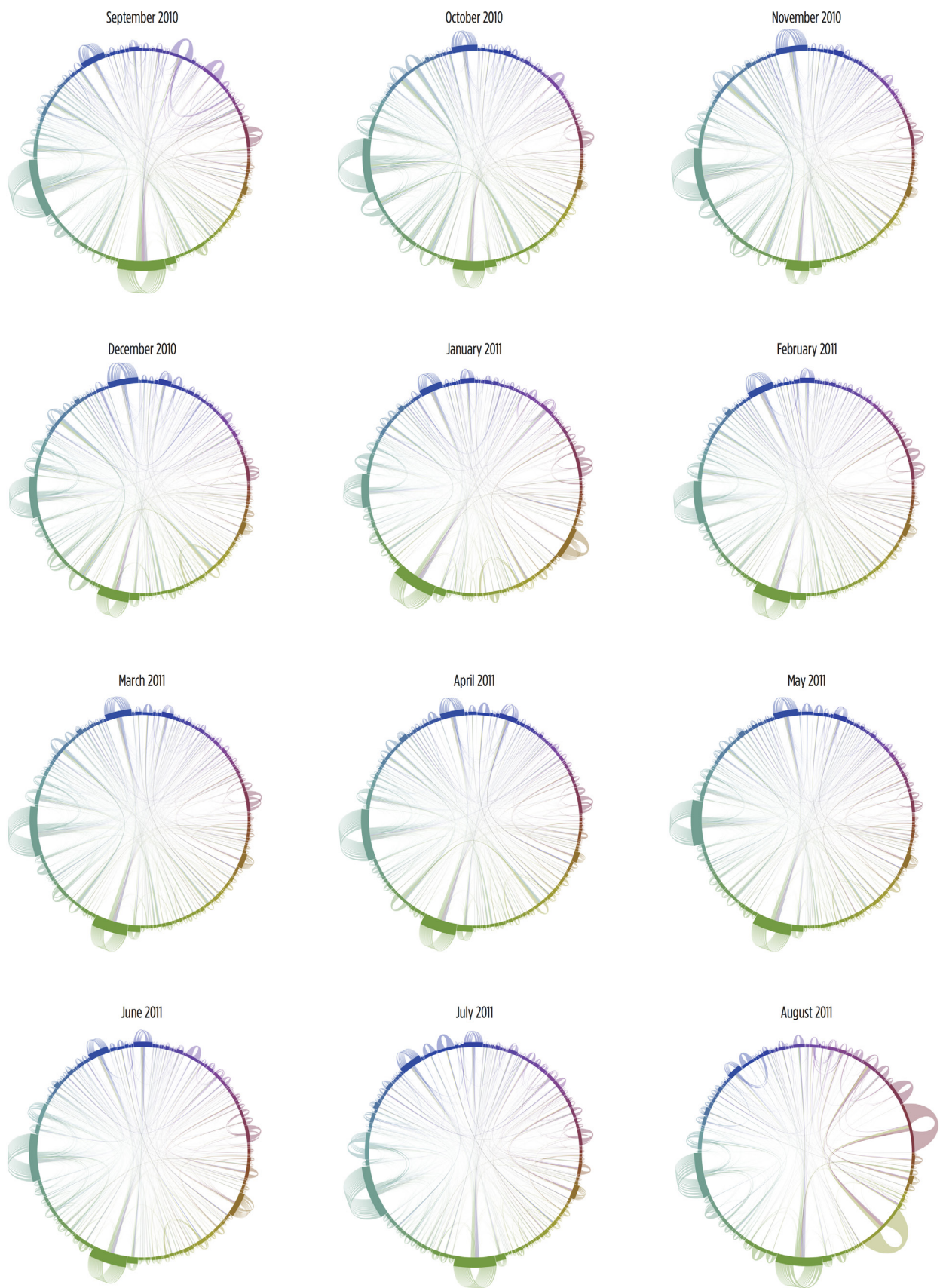


Figure 7.12. The small multiples showing monthly transitivity on the UC network on the academic year 2010-2011.

The small multiples portraying monthly flow of user transitions in the UC network are displayed in FIGURE 7.12. We created this layout to visualize how the transitionality of the wireless network changes from month to month. As a whole, the panels reveal the homogeneity through the color scheme and the general similarities, such as three big nodes with a great amount of self-transitions at the top, the left and the bottom sides of all graphs, and the uniform distribution of the edges. Analyzing each panel and comparing it to the adjacent graphs, the particular differences are evident. Many nodes grow and shrink through the months. One can easily spot these divergencies by comparing the localization of the big clusters. For example, the nodes in the lower-left quarter of the circumference gain accesses in October, if compared with the September data. The big blue cluster at the top is common in the first three months, then it starts losing accesses along the academic year, while its blue neighbors become more transitable. The month of August presents the smallest amount of transitions, along with the big lime and red nodes that are not common in the rest of the year.

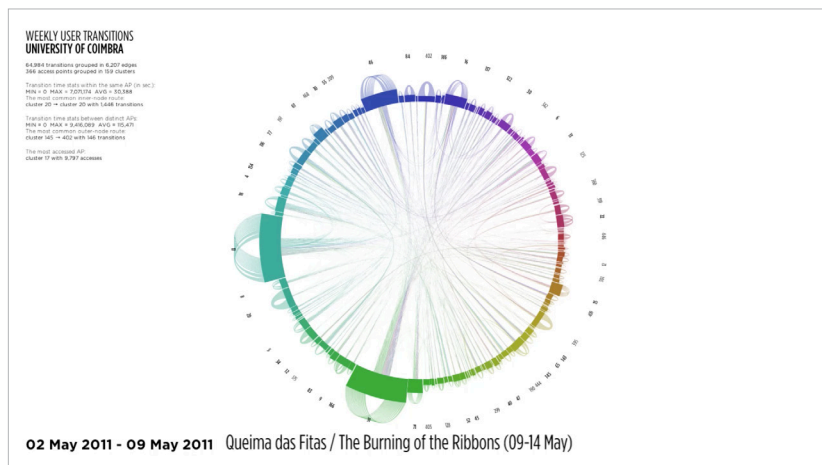
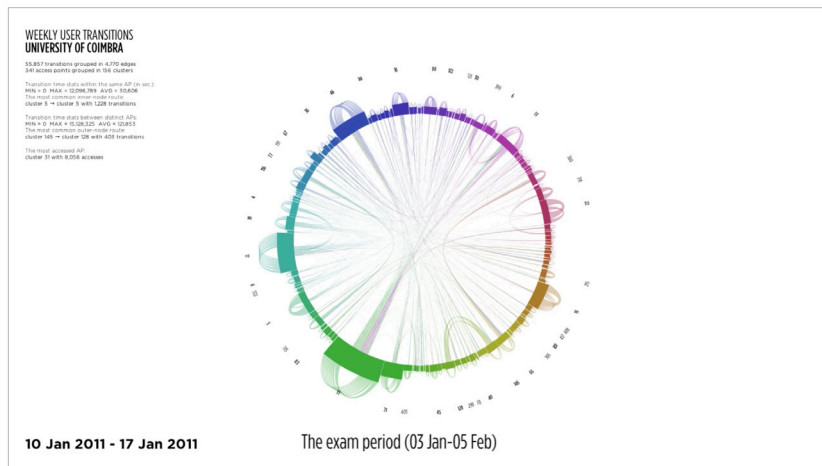
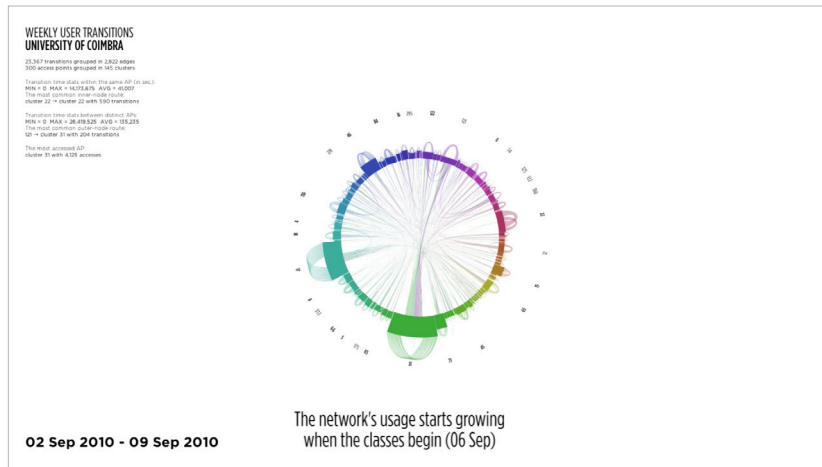


Figure 7.13. Three frames from the animation of the weekly flow of user transitions of the UC.

The first dynamic artifact allows the visualization of the weekly flow of user transitions in the network<sup>7</sup>. The animation consists of a *sliding time frame* that represents *one week* of the network's usage. The current video frame shows weekly transitions and the next frame represents the same temporal period with an advance of two hours. The diameter of the graph is proportional to the actual amount of transitions, it is mapped linearly to make the differences of the size more obvious. The temporal window, sliding upon the timeline from September 2010 to August 2011, reveals how the network's physical structure and connectivity changes over time [FIGURE 7.13]. This visual evolution allows the periodical and temporal analysis of the network.

In the beginning of the video, the textual messages contextualize the viewer, explaining what is about to happen in the visualization and how to interpret the specific graphic information. During the animation, the nodes gain incoming and outgoing transitions and the edges appear and group dynamically. As the represented temporal period is relatively small, the visual appearance of the graph does not become dense and confusing. Throughout the video, the thickness of the transition curves vary accordingly to the quantity of user accesses. As happened before, the transitions within the same nodes prevail over the transitions between distinct nodes. The textual information appears during the animation, identifying various phases of the academic year of the University of Coimbra. The purpose of these messages is to show the dependencies between the data and the academic activities. For example, in the beginning of the video, the amount of transitions starts is high because of the start of the new academic year. During holidays and academic festivities, the number of transitions lowers once the students do not attend to classes and spend less time in the university. After these epochs, the global amount of transitions grows naturally and peaks on the exam periods and at the end of the semesters.

<sup>7</sup> The animated visualization can be observed through the hyperlink: <https://vimeo.com/72670212>

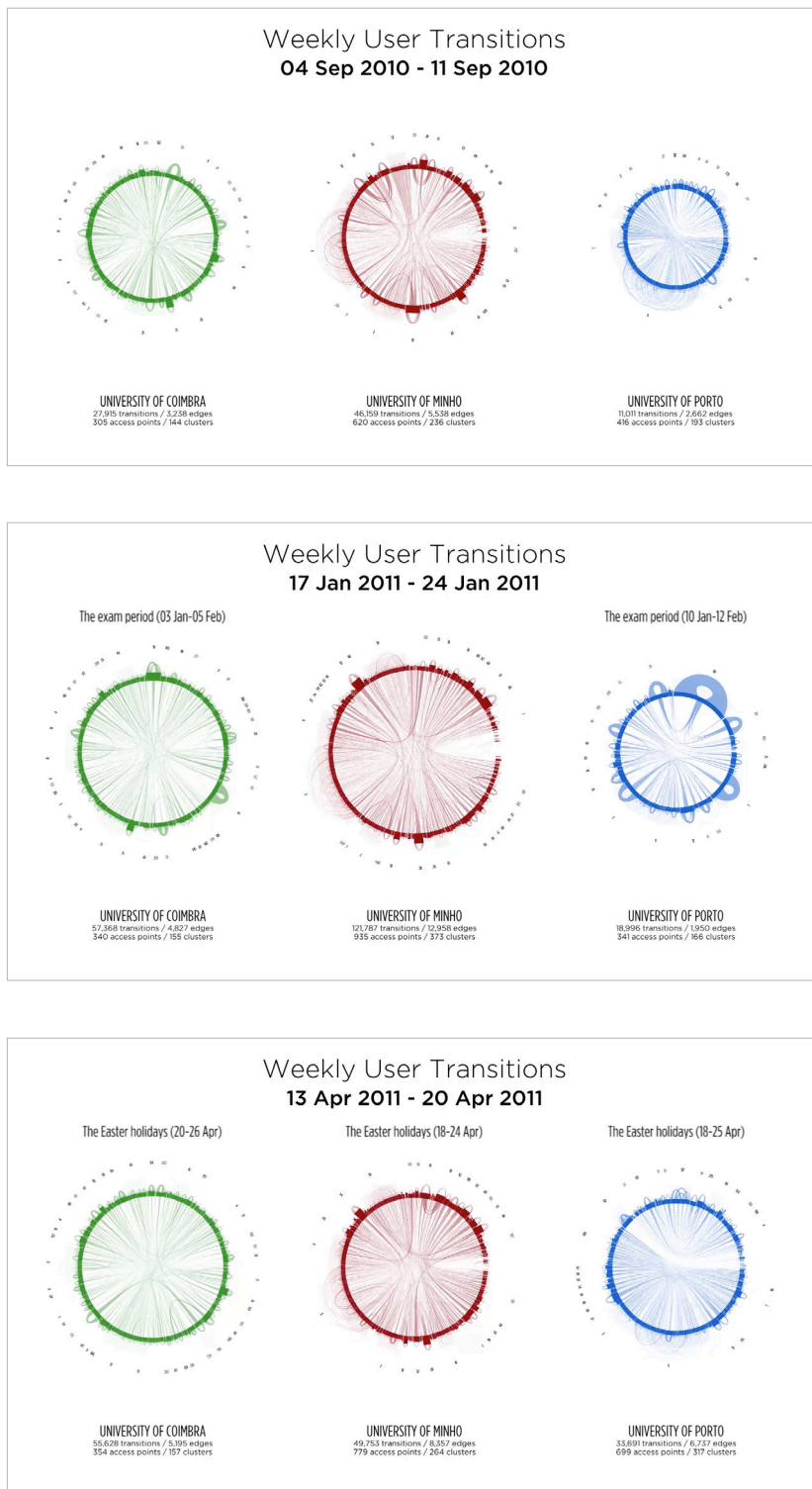


Figure 7.14. Three frames from the animation of the weekly flow of user transitions of three universities.



The second dynamic artifact displays the weekly flow of user transitions in three networks simultaneously<sup>8</sup>. Each graph has one color to avoid the visual association of nodes with the same color in distinct datasets [FIGURE 7.14]. This video also consists of a *sliding time frame of one week*, moving in two hour increments. The diameter of each graph is proportional to its current amount of transitions. The minimal and maximal amounts of weekly transitions are different for each dataset. The range of the UM dataset goes from 32,000 to 160,000 transitions, while the range of the UL goes from 4,000 to 63,000. Thus, we use the logarithmic scale of base 10 to map the diameters, preventing the graphs from abrupt size changes and spoiling the whole visualization. The temporal window reveals how the network's structure and connectivity changes over time, allowing the periodical and temporal analysis of the network. The *contextual information* is also present during the animation. The messages vary in the instants of appearance and in the presented information – distinct universities have different calendars for the same academic year.

Comparing all three universities, we can make additional observations. One can perceive that the graph of the UM is the biggest, presenting the highest average amount of transition throughout the timeline. Analyzing the quantities of students in each university, we deduce that it does not relate to the amount of weekly transitions of that university. The UC has about 20,000 of students and presents intermediate amounts of transitions, regarding two other graphs. The UM has 16,000 of students, but the connectivity in the UM network is the highest. The UL has 30,000 students (almost the double of the UM), but presents the weakest usage of the network. Since this transitions-users relation does not exist, it can rely on many another aspects, such as the existence of scientific courses that demand the use of electronic devices, the financial status of students, the habits to communicate, exchange files, search information or spend time online, the physical extension of the campus, and many other factors. In the UC network, the amount of transitions peaks before and during the exams, while in the UM it maintains or even lowers. It can happen because the students of the UM leave the university to study for the exams, and the students of the UC do exactly the opposite. During the *Burning of the Ribbons* (the traditional academic festivity), the UC users leave the campus to celebrate, while in other two universities the connectivity remains high. In addition to all previous conclusions, this visualization allows us to contemplate the intervention of the human factor.

This technique for visualization of user transitions is a dynamic graphic model. It can be applied to any of the given datasets to display the information with the same visual variables, facilitating the interpretation of information. We developed a complex visual system that shows multivariate scientific data. The complexity demands an initial study of the model to understand the context, the nature of the data and the encoded visual variables. Once that study is made, the user is able to interpret the graphic language and explore the data using the implemented interactive functionalities. The multiple radial graphs representing the different data emphasize patterns, differences and similarities in information due to the effective graphic properties of this visual system.

<sup>8</sup> The animated visualization can be observed through the hyperlink: <https://vimeo.com/72674185>

USER CONNECTIONS DATA  
UNIVERSITY OF COIMBRA

587 users performed 663,450 connections  
Statistics of separate connection times (in seconds)  
SUM = 2,165,432.16; MIN = 0; MAX = 756,829; AVG = 3,294

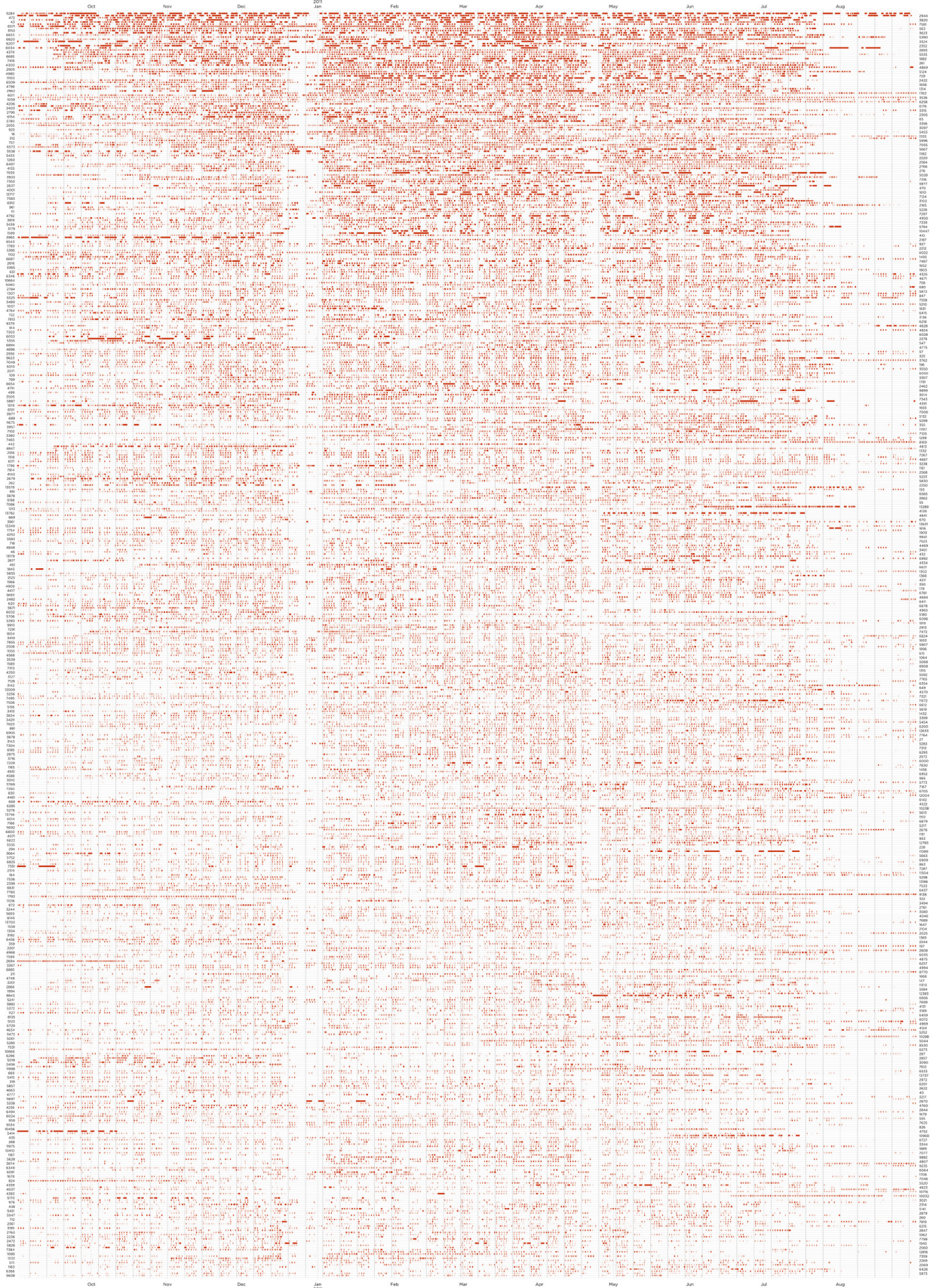


Figure 7.15. The user connections occurred on the wireless network of the UC on the academic year 2010-2011.

## 7.2. The Connectivity of Users

This section describes the visualizations of user connections produced with *The Matrix of Connections* model. The graphs with the minimal amount of zoom present visible patterns in the global and particular user connectivity.

The overall view of the matrix of user connections of the UC 2010-2011 dataset [FIGURE 7.15] portrays various patterns. As we described in SECTION 6.2, the user list is in the descending order, sorted by the total connection time. The plot shows 587 of 13,926 users that spent the greatest amount of time online. One can see the graph divided vertically in three parts by two white spaces. These connectivity pauses occurred at the end of December 2010 (the Christmas holidays) and at the end of April 2011 (the Easter holidays). There are few people that continued to perform connections, but the global connectivity is much lower than before or after these two epochs. We can see a lot of white spaces at both sides of the graph – the network has a small number of users at the beginning and at the end of that academic year. The density of the red spots in the matrix indicates the frequency of user connections on the timeline. The top of the graph is more saturated than the bottom. The users with biggest total times performed more long connections than other users. The long connections are highlighted with a vivid bars, while the short connections are thin barely visible spots. The periodicity of connections forms visible patterns. That periodical repetition is well-noticeable on groups of short connections [FIGURE 7.16]. A closer look to the lower half of the graph, where long connections are absent, reveals that the majority of the connection bars happen in the same temporal intervals during the daytime, from 10 a.m. to 4 p.m. This is the common class time in the University of Coimbra. The irregularity of the pattern happens because some users do not connect daily.

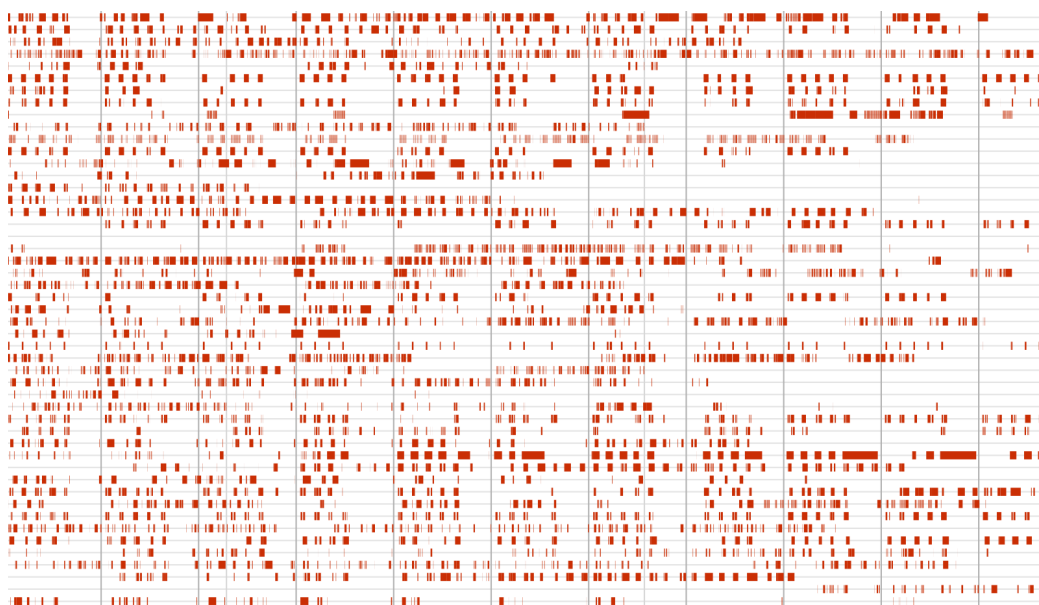


Figure 7.16. The patterns defined by the repetition of short periodic connections.

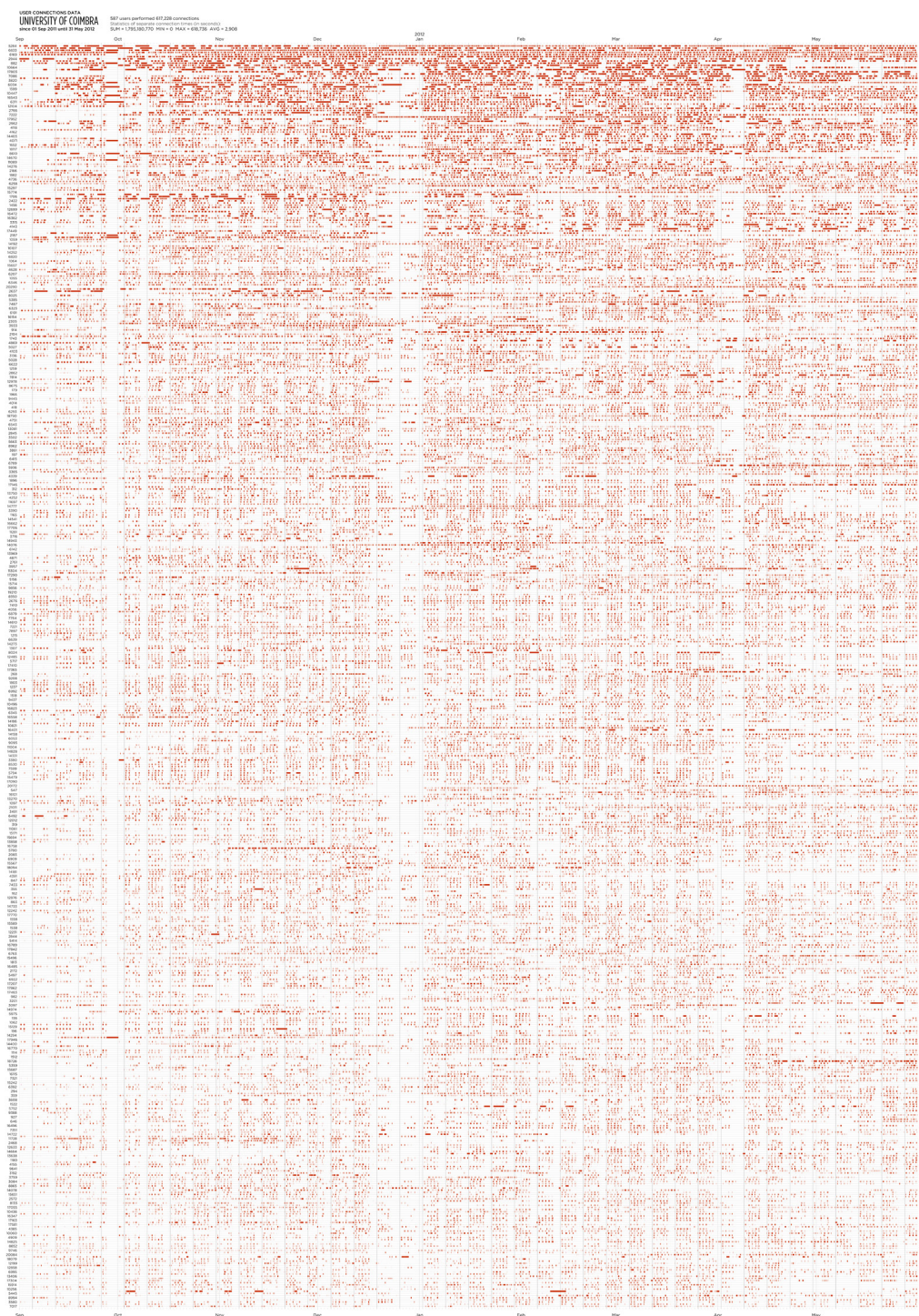


Figure 7.17. The user connections on the UC network regarding the academic year 2011-2012.

Comparing the matrices of two different universities [FIGURES 7.17 AND 7.18], we can see that the connectivity of the UC network is richer than in the UL, on the academic year 2011-2012. The UC top users also present more network activity, fading from the top down, while in the UL the visual rhythm is pretty homogeneous in all parts of the graph, along the timeline. In the UC matrix, we can see two small pauses in connectivity (at the end of September and at the beginning of April) and one larger interval (at the end of December 2011, the Christmas holidays). In the UL dataset, there are no clearly visible white spaces, except for the month of March representing a lack of data registry. Yet, the users of the UL demonstrate a more regular pattern of connection periodicity: weekdays are spent online (five consecutive

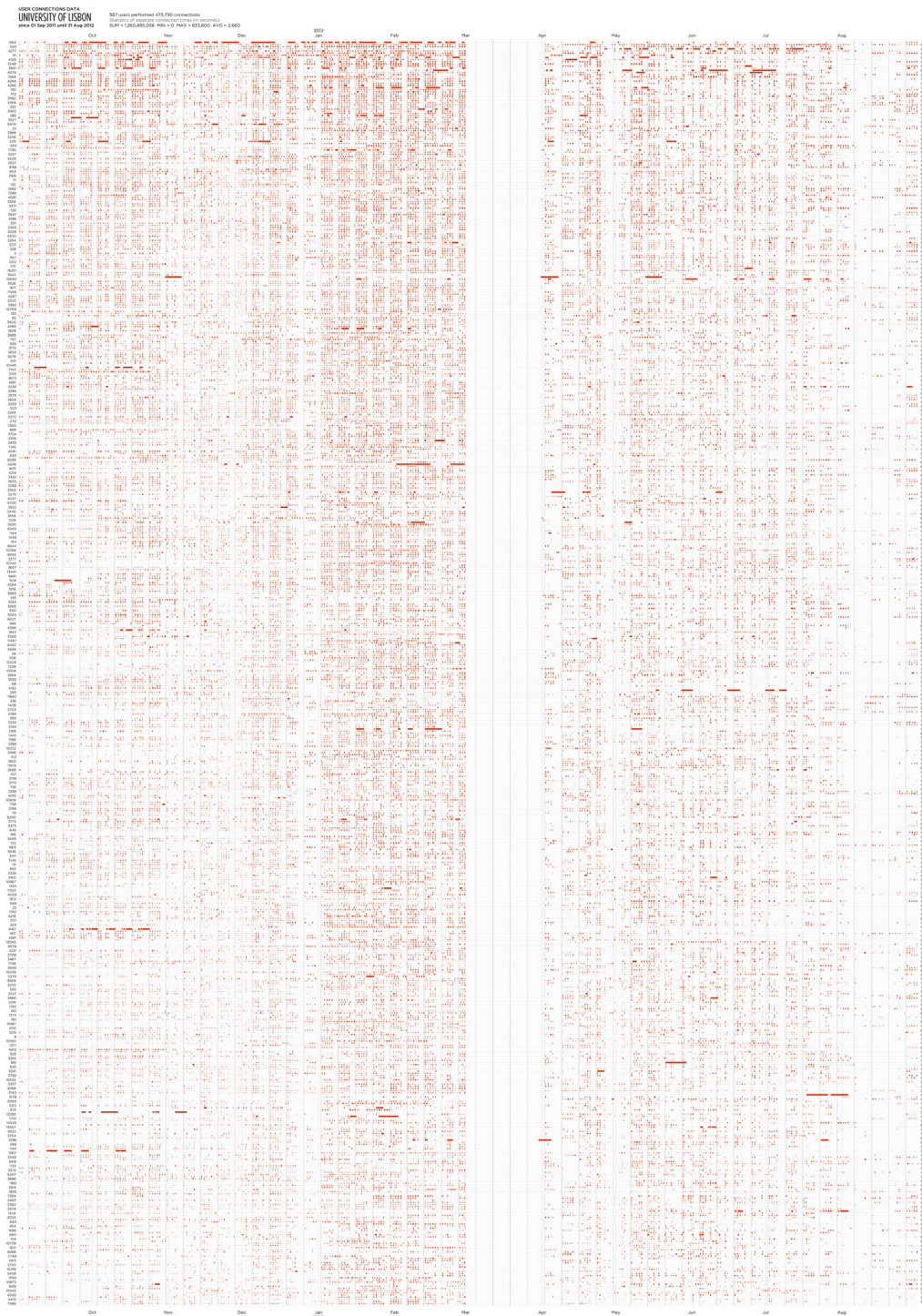


Figure 7.18. The matrix of connectivity of the UL network, the academic year 2011-2012.

red bars) and weekends are free of connections [FIGURE 7.19]. This happens in the first semester only, then the pattern becomes unstable. There are several users that perform very long connections (sometimes with a duration of two weeks), but they are not at the top of the list. For example, the user #4167 (at the bottom of the graph) performs various long connections during September/October, but he does not appear online during the rest of the year.



Figure 7.19. The pattern of the weekly connectivity in the UL network.

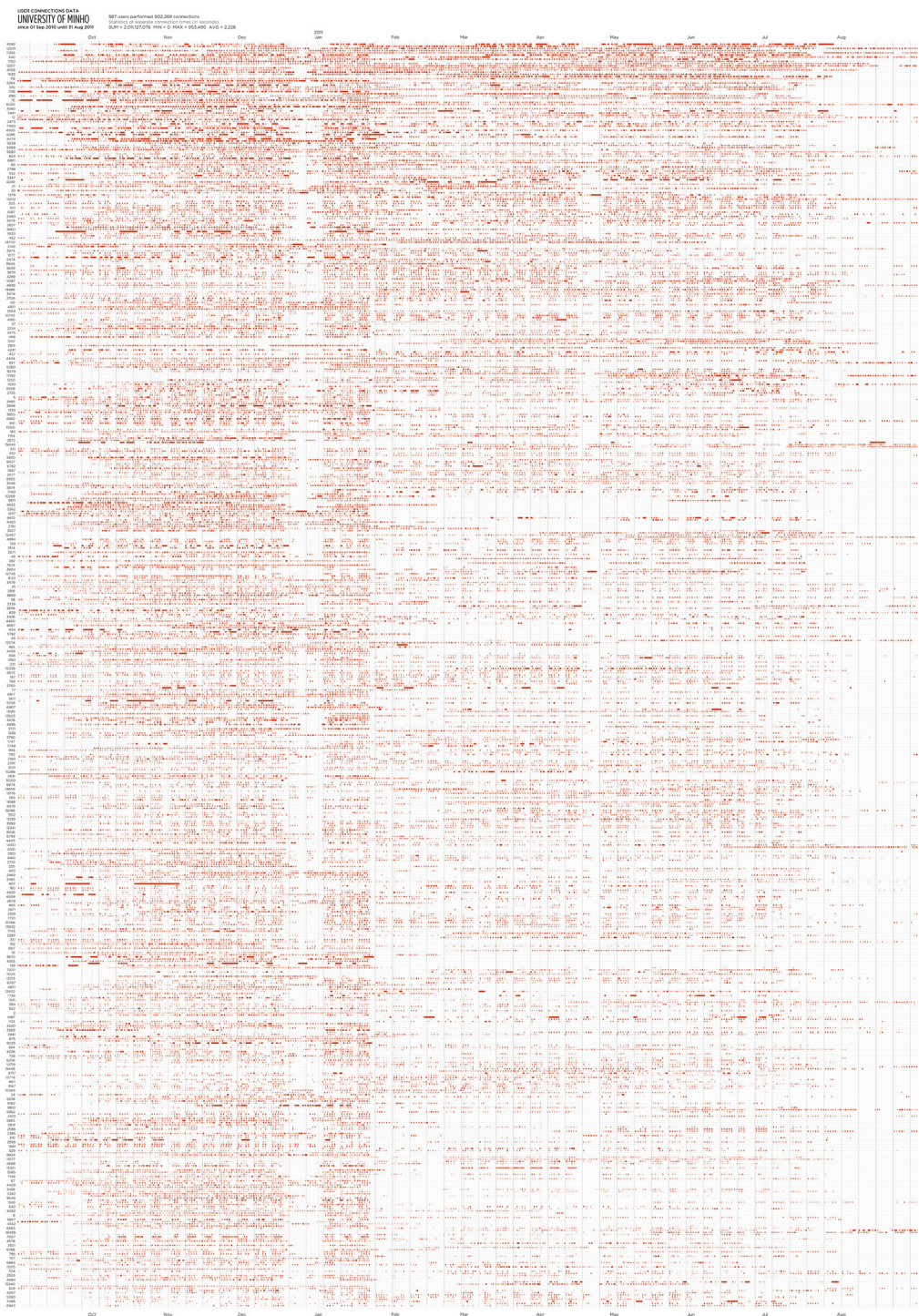


Figure 7.20. The connections of users occurred on the UM network on the academic year 2010-2011.

The University of Minho presents an uneven and diffuse distribution of user connections in the matrix. FIGURES 7.20 AND 7.21 can be seen as one because the ending instant of the first is the starting point of the second timeline. What differs in these graphs is the order of users. The first semester of 2010-2011 has a high and dense connectivity. Starting from the mid-January, the situation drastically changes, leaving far less connection traces of users – something important happened in that



Figure 7.21. The connections of users registered in the UM 2011-2012 dataset.

month. In 2011-2012, the image is inverted: the left half has lesser density of connections, and the right half is more saturated. The passage is not abrupt, like in the previous year, but smooth, creating a gradient color. The first year does not depict any evident pattern of connections, and the second has many weekly connection patterns, similarly to the UL dataset. In the upper-right corner, we evidence a large accumulation of long and frequent user connections.

USER CONNECTIONS DATA  
UNIVERSITY OF PORTO

587 users performed 198,059 connections  
Statistics of separate connection times (in seconds)  
SUM = 826,79,501 MIN = 0 MAX = 960,035 AVG = 4,371

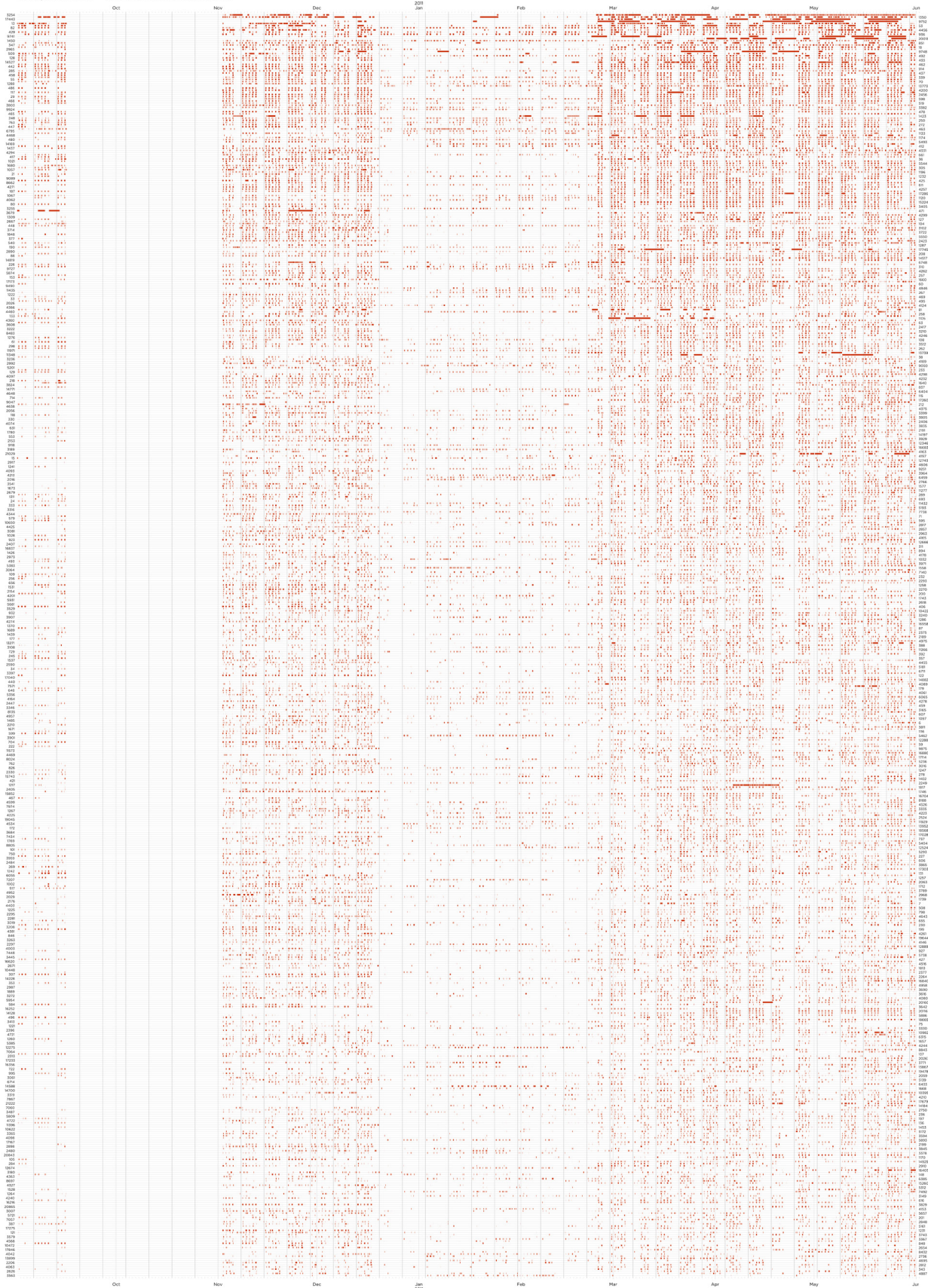


Figure 7.22. The user connections on the UP network on the academic year 2010-2011.



The visualization of connections on the University of Porto demonstrates a periodical behavior of the global connectivity [FIGURE 7.22]. The academic year starts with a low network's usage, then we observe a six and a half weeks of the lack of data. Since the beginning of November until the mid-December, the amount and frequency of connections raise. After that period, we see nine weeks of a significant reduction in connectivity. The final three months present a high density of connections again. This graph shows a uniform weekly pattern along the academic year. The week ticks (equally spaced dark gray lines) help to visualize that pattern. A column has five consecutive connections, one for each weekday, and a white space corresponding to the weekends [FIGURE 7.23]. Even in the epochs of low connectivity, one can detect this visual repetition. The eighth column has a little more white space than other weeks. That space could refer to a Friday holiday on that week.

This visualization technique representing user connections consists of a matrix plot. The correspondence between user connection intervals can be easily found on the timeline due to the graph's grid. The precise temporal instant can be consulted while using the respective interactive application, where the connections can be easily localized with the custom made mouse pointer. The two-dimensional layout and the contrasted color scheme allow the visualization of the connections' density and the periods of global activity on the overview of the graph. The patterns and the periodicity of accesses become visible at a detailed analysis. The visual simplicity of this model works effectively on the representation of connectivity in wireless networks.

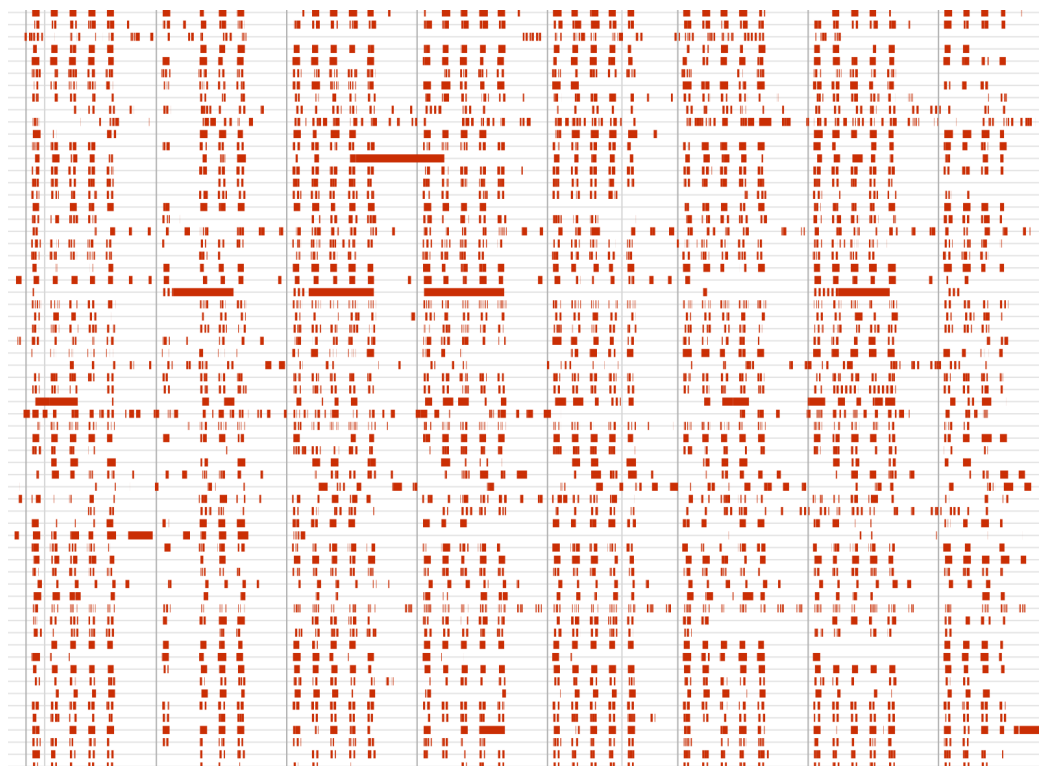


Figure 7.23. The weekly connectivity pattern in the UP network.

AGGREGATED USER TRANSITIONS  
**UNIVERSITY OF COIMBRA**  
 since 01 Sep 2010 until 31 Aug 2011

PANEL MATRIX:  
 Columns - aggregation by months  
 Rows - aggregation by days

CELL:  
 Vertical axis - amount  
 Horizontal axis - aggregation by hours

Color legend:  
 ■ Amount of users  
 ■ Amount of transitions

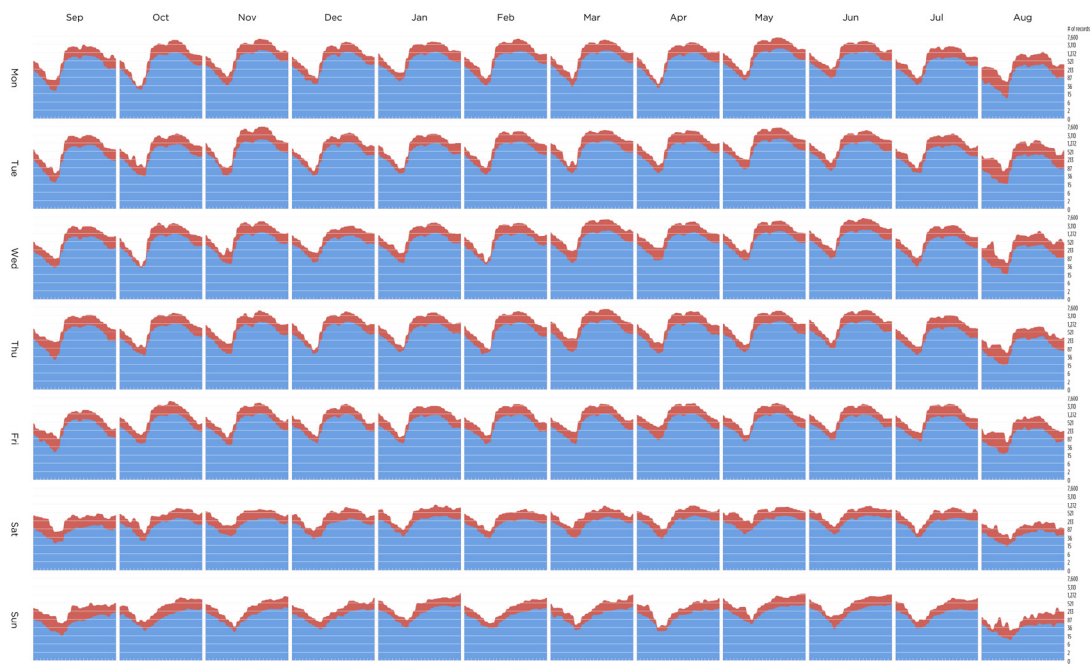


Figure 7.24. Panels of user transitions of the UC aggregated by months/days/hours, the academic year 2010-2011.

AGGREGATED USER TRANSITIONS  
**UNIVERSITY OF LISBON**  
 since 01 Sep 2011 until 31 Aug 2012

PANEL MATRIX:  
 Columns - aggregation by months  
 Rows - aggregation by days

CELL:  
 Vertical axis - amount  
 Horizontal axis - aggregation by hours

Color legend:  
 ■ Amount of users  
 ■ Amount of transitions

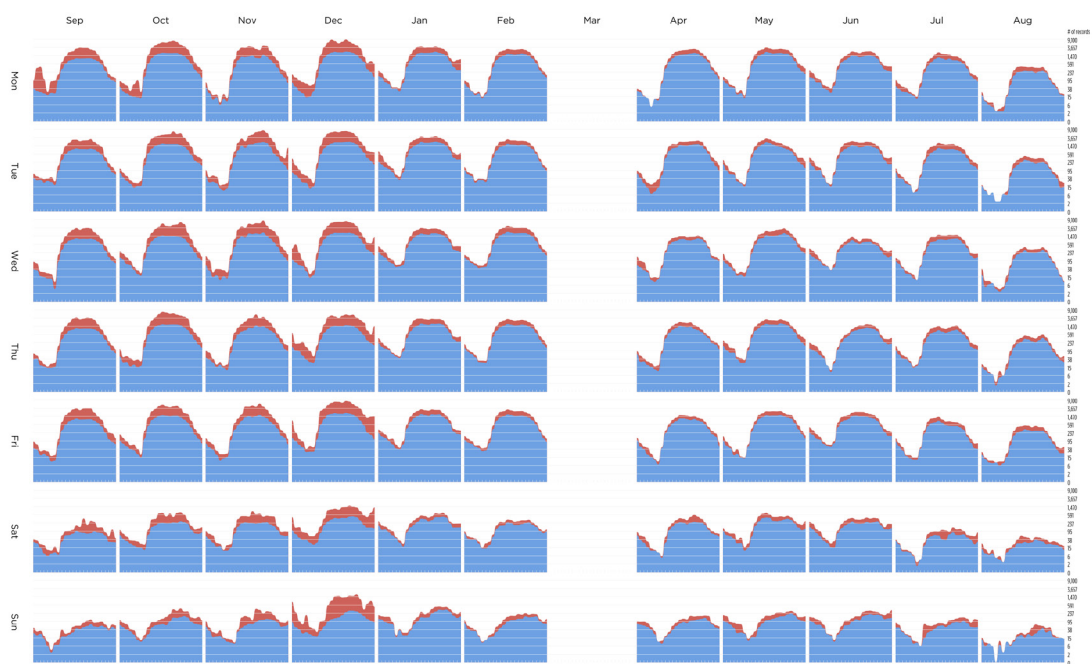


Figure 7.25. User transitions aggregated by months/days/hours, from the UL 2011-2012 dataset.

### 7.3. The Patterns of Network Usage

The graphic results obtained with *The Panels of Flow Shapes* model are presented in this section. The developed application allows the comparison of multiple panels of six types of data aggregation, regarding two categories of data combined with the two types of variables. For each academic year, we can present 24 visualizations. Since we have six years, the number of possible visualizations is at least 144. We show and describe the most interesting graphs in this section.

FIGURE 7.24 shows the amounts of users and performed transitions, aggregated by months/days/hours, in the academic year 2010-2011. All panels are visually similar, presenting minimal values from 5 to 7 a.m. and maximum values from 2 to 4 p.m. The values from the last column of the matrix, referring to August 2011, have a bigger deviation from the rest of the year. The quantity of users in the network lowers during the summer holidays, and consequently the amount of performed transitions is also small. Also, one can visualize that the numbers are low at weekends (the last two rows of the matrix), if compared to weekdays.

In the UL dataset regarding the year 2011-2012 [FIGURE 7.25], the quantity of users stays stable during the semesters, except for August (the summer holidays). The blue shapes are similar, with low values from 12 to 7 a.m. and great amounts from 9 a.m. to 6 p.m. The quantity of transitions tends to decrease through the year and during the weekdays: from 3,000 to 10,000 per hour in the first four months; from 2,500 to 5,000 in January and February 2012; and from 1,000 to 4,000 in the remaining months. We do not take into account the month of August and weekends, because these periods also demonstrate a lower data values.

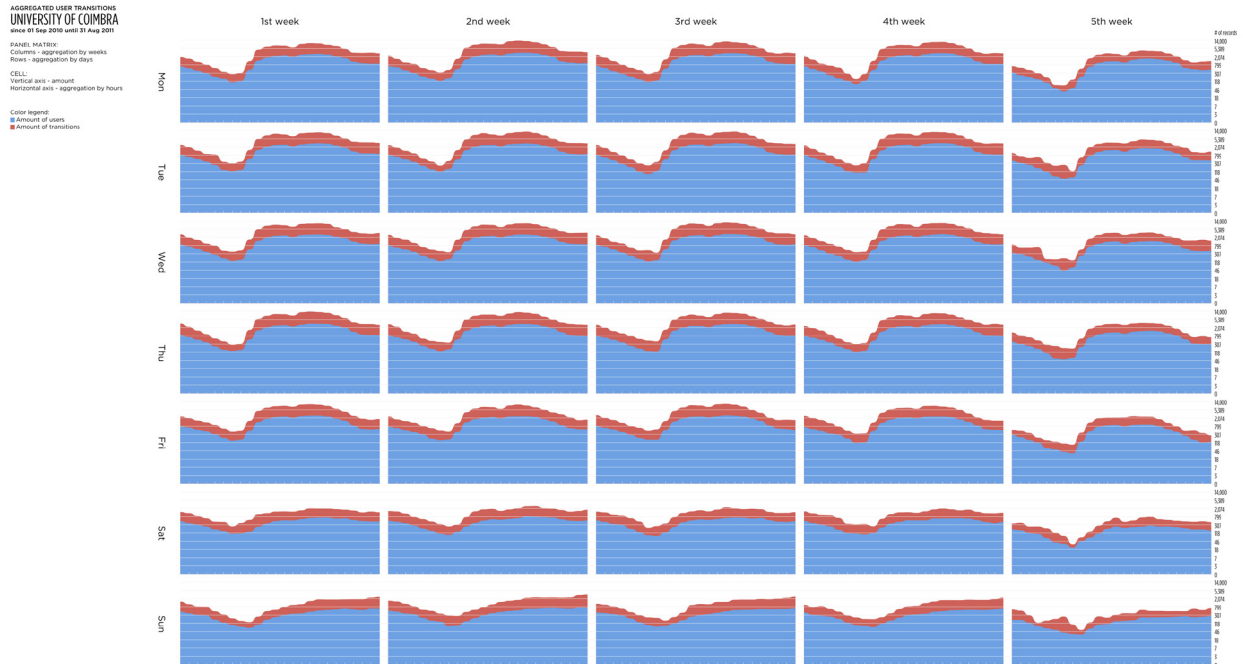


Figure 7.26. The small multiples graph of user transitions of the UC aggregated by weeks/days/hours, the academic year 2010-2011.

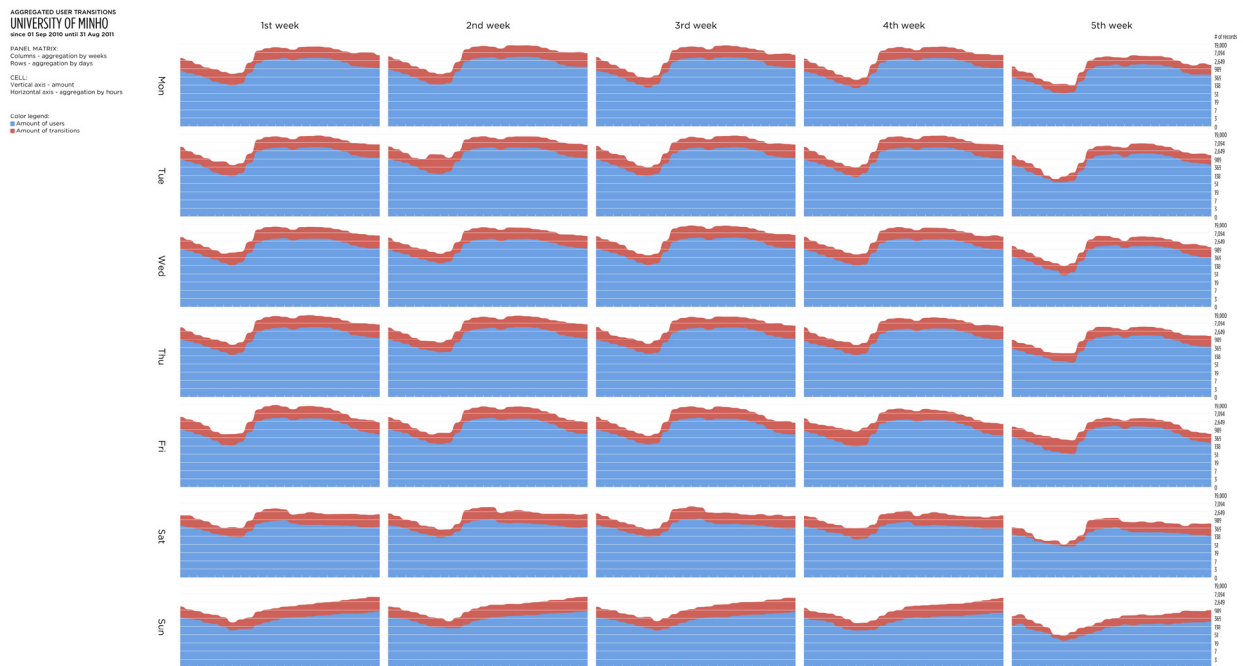


Figure 7.27. The panels showing user transitions in the UM network in 2010-2011, aggregated by weeks/days/hours.

The graphs displayed on FIGURES 7.26 AND 7.27 depict the amounts of users and transitions existent in the UC and the UM networks on the academic year 2010-2011. The columns of each matrix represent weeks of the month and the rows indicate days. Both visualizations have a similar visual pattern, the quantities of users and transitions decrease from 12 to 8 a.m., staying high on the rest of each day. On weekends, the shapes change, presenting lower values. The further exploration of the data shows that the maximum value ranges are different – 14,000 for the UC and 20,000 for the UM dataset. This means that the average daily transitionality in the UM network campus is higher than in the UC. In the UC, we can see pattern deviations on Tuesdays, Wednesdays and Sundays of the 5th week of the month. There can be seen peaks of transitions at 3 and at 6 a.m., while the transition amount shape graphically replicates the user amount shape, but with a higher value, in the remaining days and weeks.

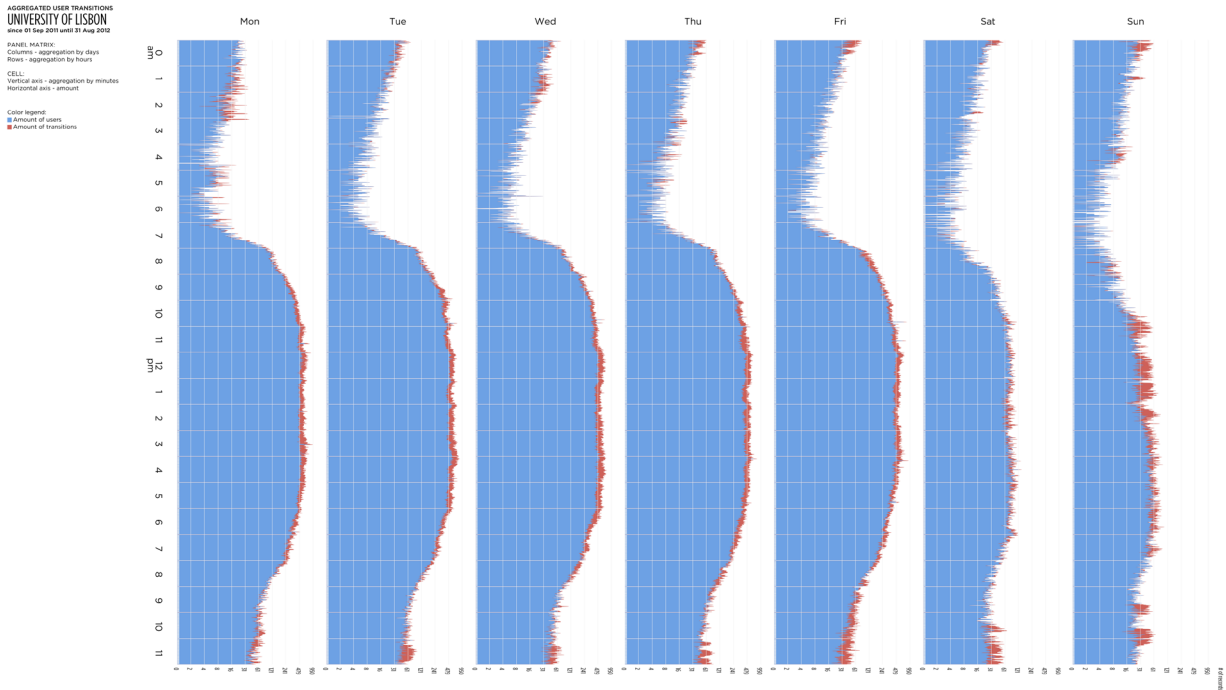


Figure 7.28. The panels of the transition data of the UL 2011-2012 dataset, aggregated by days/hours/minutes.

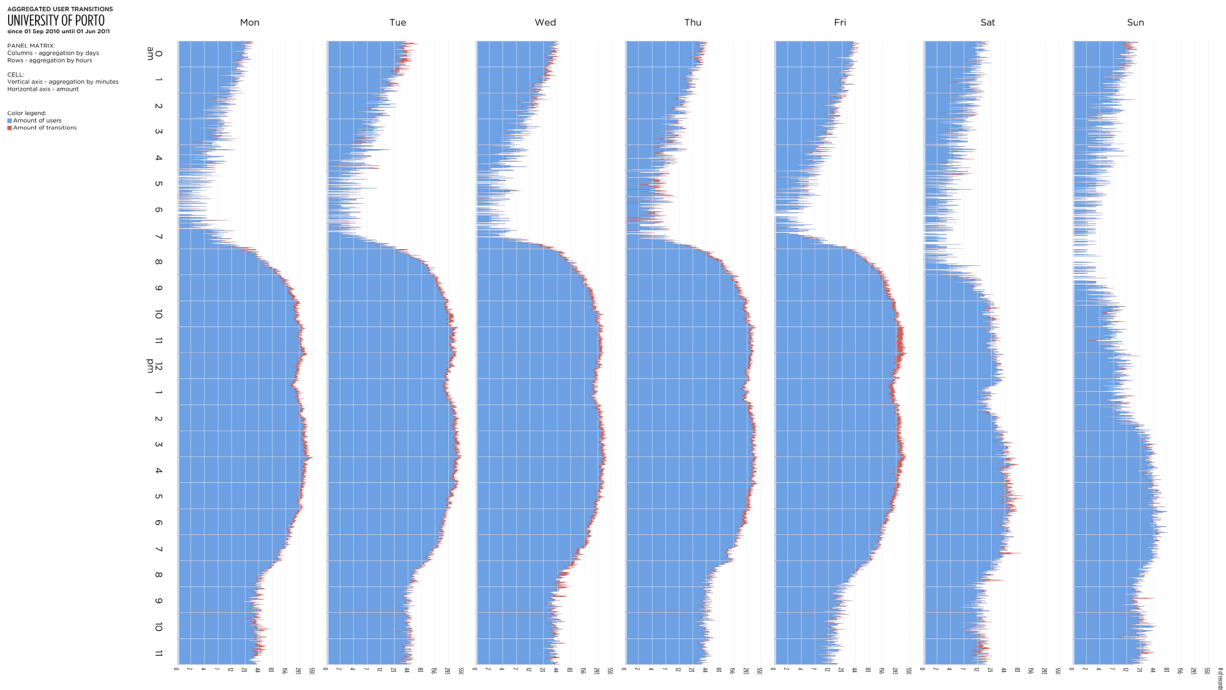


Figure 7.29. The graph of user transitions in the UP network on the academic year 2010-2011, data aggregated by days/hours/minutes.

The visualizations of the UL 2011-2012 [FIGURE 7.28] and the UP 2010-2011 [FIGURE 7.29] transitions data present the pattern and the deviation at the same time. The data shapes of the graph's columns are similar in the weekdays, having an irregular form on the weekend. The period of the low activity is from 3 to 6 a.m., then the amounts start rising and stabilizes around 10 a.m. The daytime is the period of high and uniform transitionality. By night, the network's usage reduces. The graph of the UP data displays slightly decreased values at 1 p.m. It possibly refers to the lunch break interval.

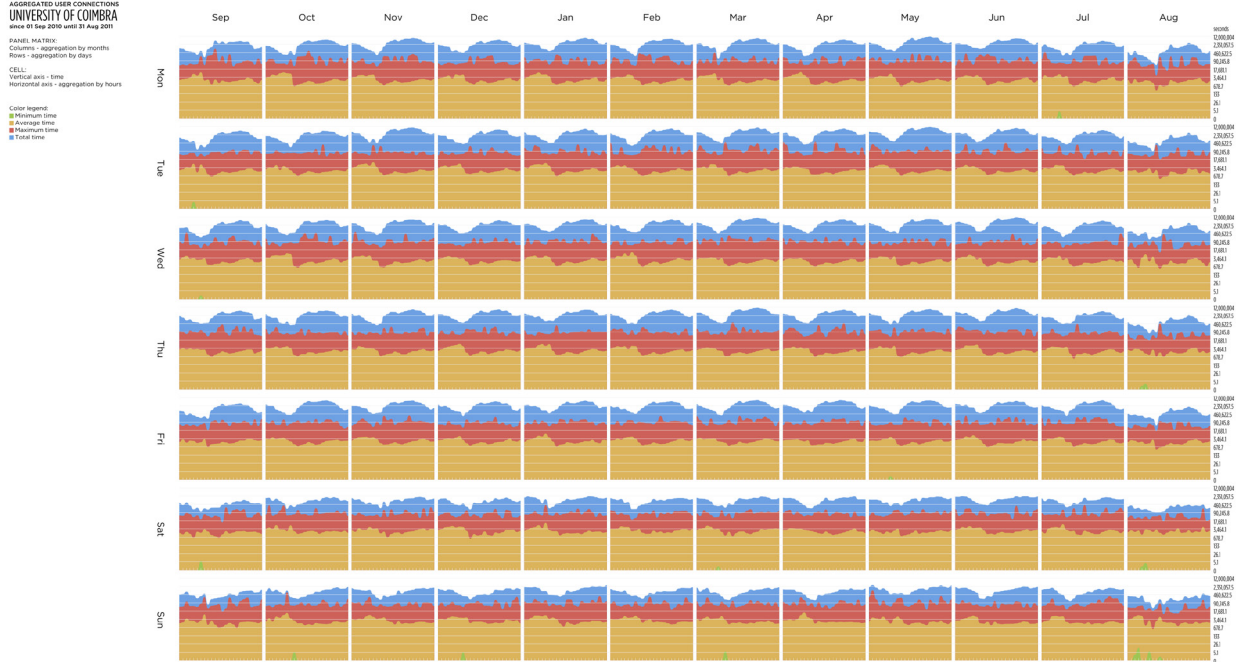


Figure 7.30. The panels of user connections in the UC network on the academic year 2010-2011, aggregation by months/days/hours.

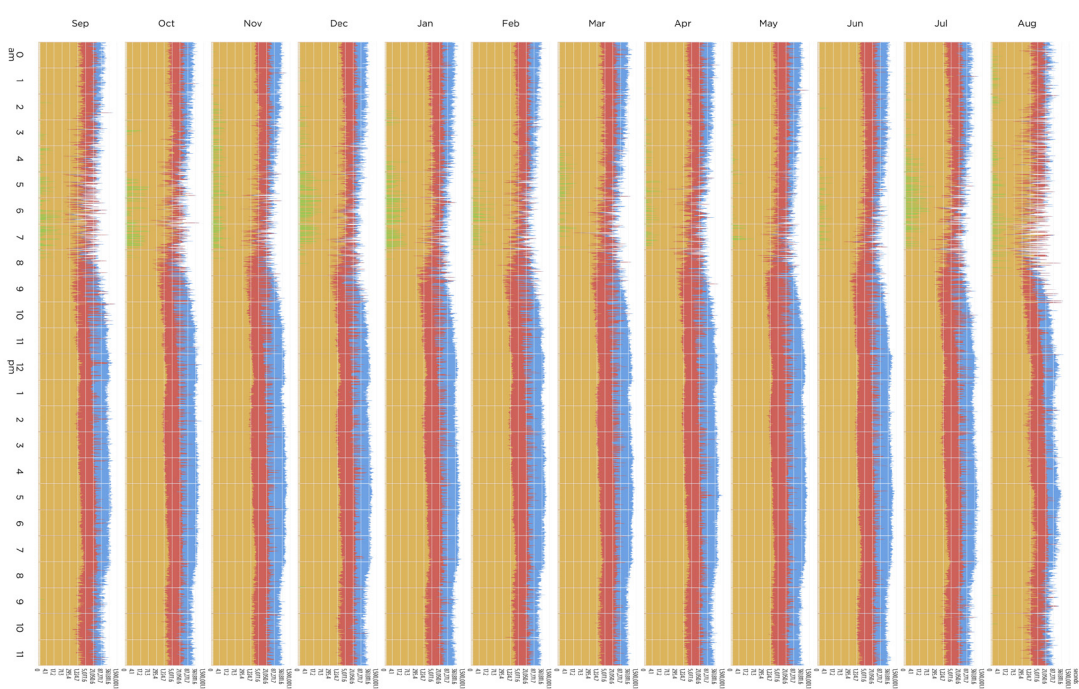


Figure 7.31. The panel graph of connection times of the UC network on the year 2010-2011 aggregated by months/hours/minutes.

In the graph showing daily connection times of the UC 2010-2011 dataset [FIGURE 7.30], four layered variables are portrayed. The minimum connection time values (green) are mostly equal to one second, peaking from 2 to 12 seconds in some days. The average times (yellow) follow the pattern of being higher from 12 to 7 a.m. and lower on the remaining hours. The maximum connection times (red) maintain at the range from 40,000 to 50,000 seconds, with various peaks and lows during the day. The total time value shapes (blue) also demonstrate approximately the same



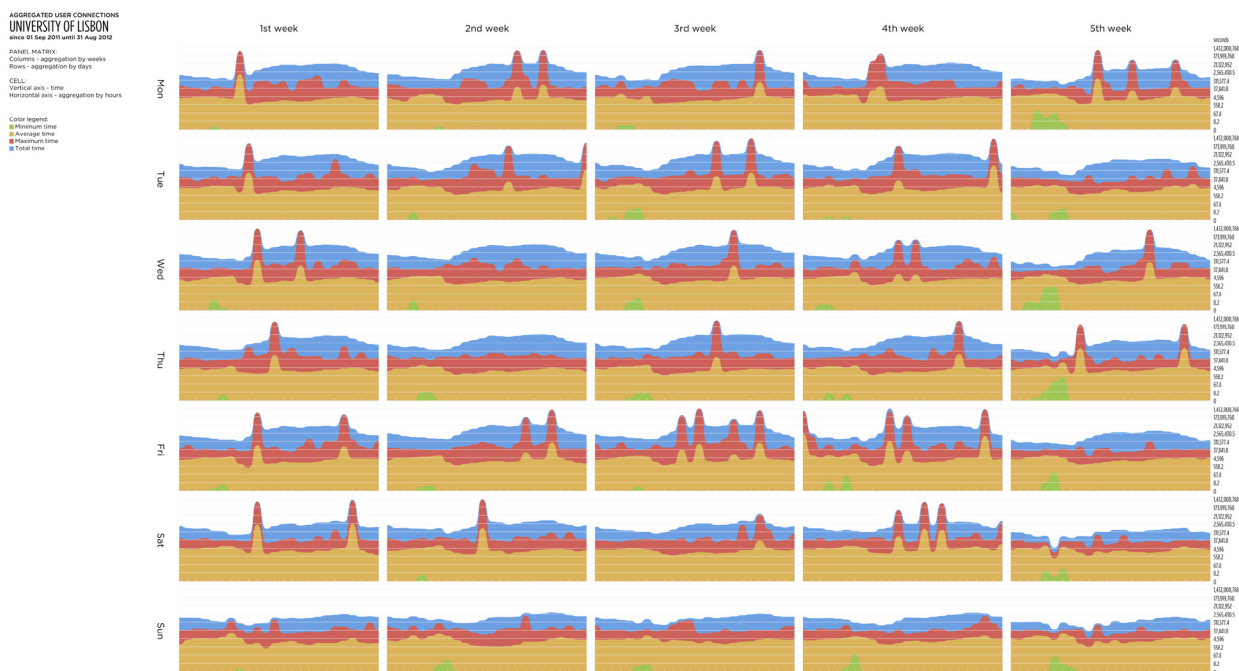


Figure 7.32. The small multiples visualization of connection times of the UL 2011-2012 dataset aggregated by weeks/days/hours.

form, decreasing in the interval from 0 to 7 a.m., then increasing from 7 to 10 a.m. and being high for the rest of the day. The scale ranges of the average and the maximum times are very different: the first varies from 500 to 12,000 seconds, while the second range goes from 38,000 up to 110,000 seconds (without considering peak values). This indicates that the common connection period is relatively short, but there are users that spend a lot of time online.

The visualization of user connection times of the UC 2010-2011 dataset, aggregated by months/hours/minutes [FIGURE 7.31], displays a quite distinct variation. While the shapes formed by the entire month data are similar (except for the column of August that differs from the rest, as expected), the curves of each hour are irregular. There are many highs and lows that create big deviations. If we concentrate on a single cell, we can visually trace a intermediate level that passes between these peaks. Yet, it is still a difficult task. The value deviations in both four variables indicate the inconstant connection intervals performed by users, and that variation of temporal interval values is common due to the human factor. We can observe that the peaks in the average time shapes vary less than in the maximum time shapes. This happens accordingly to the average count formula, but the maximum values do not tend to reach a lesser variation, even in this large dataset.

The graph on FIGURE 7.32 displays the visual properties similar to the visualization shown on FIGURE 7.30. The average and the total connection times tend to follow a visual pattern in all days. The main difference is the abnormal peaks present in these panels that modify the shape of the upper-level layer. In this case, the maximum times reach huge values (hundreds, sometimes thousands of millions of seconds) at various hours of the day. These deviations directly affect the total times of the respective hours, forcing them to grow and increase the overall value scale.

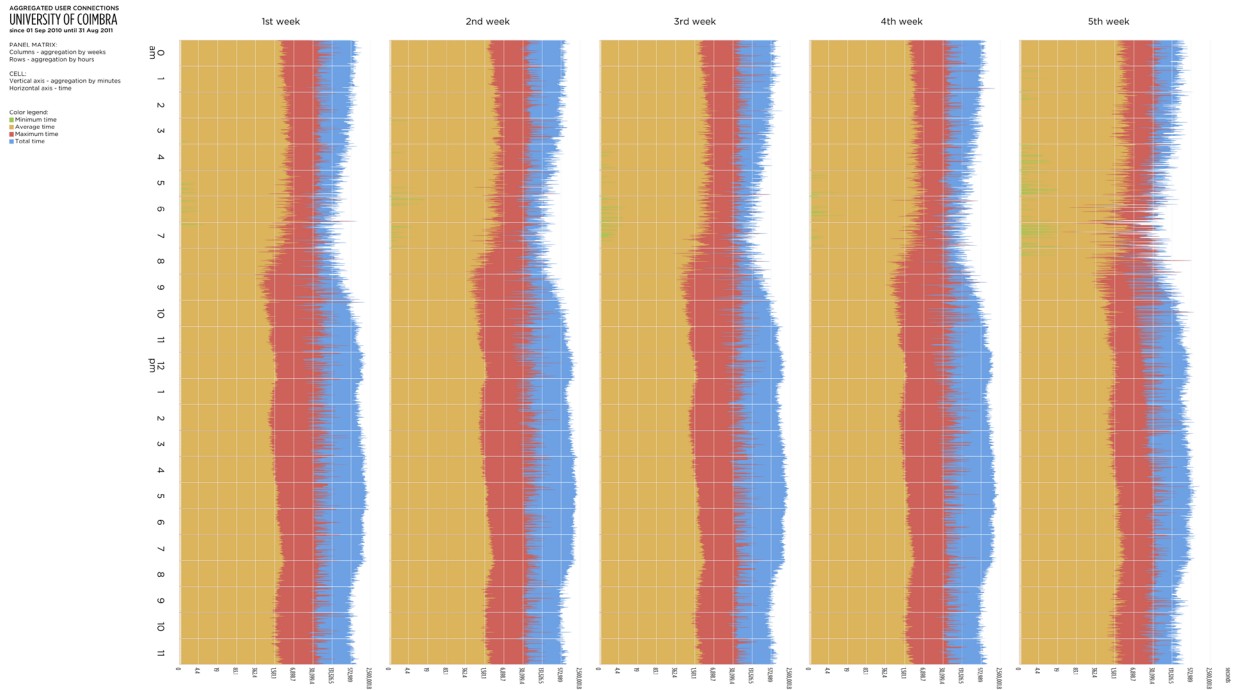


Figure 7.33. The graph of transition times of the UC 2010-2011 dataset aggregated by weeks/hours/minutes.

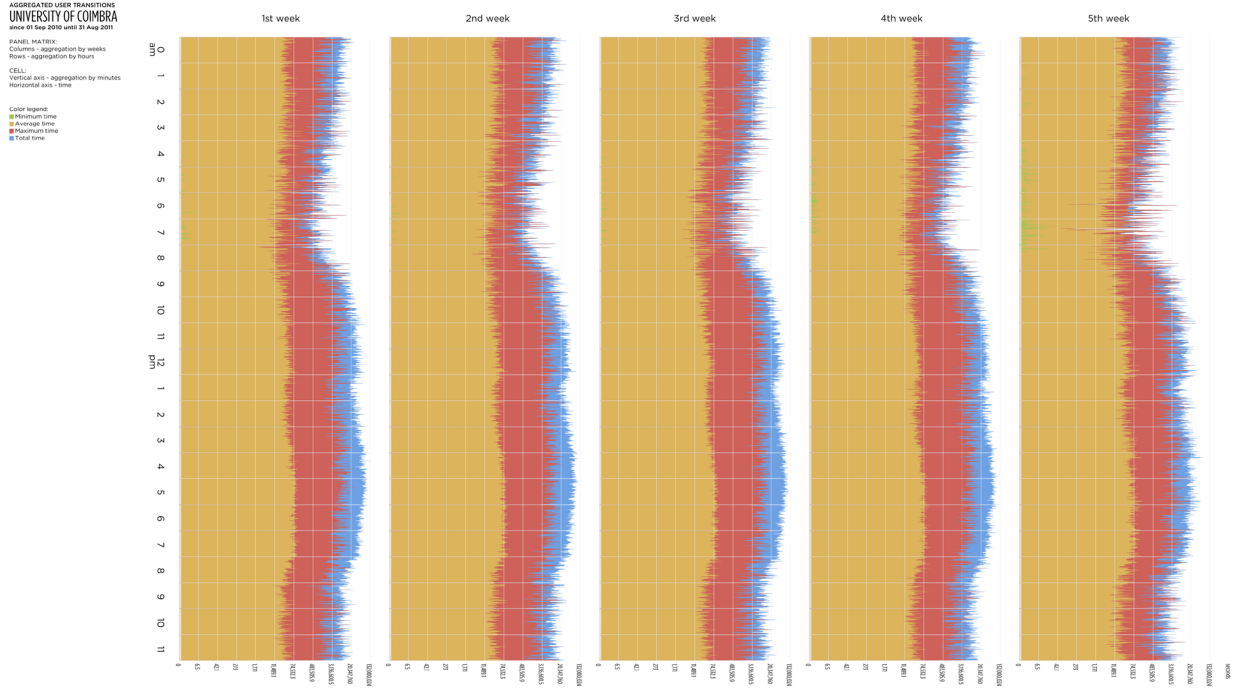


Figure 7.34. The graph of connection times of the UC 2010-2011 dataset aggregated by weeks/hours/minutes.

The visualizations depicted on FIGURES 7.33 AND 7.34 show the UC network data regarding the academic year 2010-2011. Each column shows the weekly aggregated data of the entire year. The first graph shows transition times between 1 and 2,500,000, and the second displays connections times between 1 and 132,000,000 seconds. Taking into account that the scales are different, we can focus only on the form of the shapes of four layers. The minimum times are mostly equal to 1 second. The average transition times demonstrate a continuous value consistency through the hours, varying slightly. The average connection times have a bigger variation, being higher from 0 to 7 a.m. and lower during the remaining 16 hours. The red layers are opposite: while the maximum connection times stay at the conformable range, the transition time values decrease from 0 to 7 a.m., increase and stay high until 8 p.m., then lower again until the midnight. The total times present the same appearance in both graphs, but the difference of values between the red and blue layer in the connection times graph is bigger than in the transition times graph.

The developed visualization technique of multiple panels displaying aggregated data reveals the patterns and differences in the usage of wireless networks. This is achieved with the visual similarities and variations in multiple graphs and layers of variables. Distinct temporal periods and aggregations can be analyzed and compared to understand when the similarities or deviations happened in the network's activity. The logarithmic scale minimizes the scale differences between large value ranges. The grid localizes the data values in these ranges and the particular values can be consulted using the interactive application.

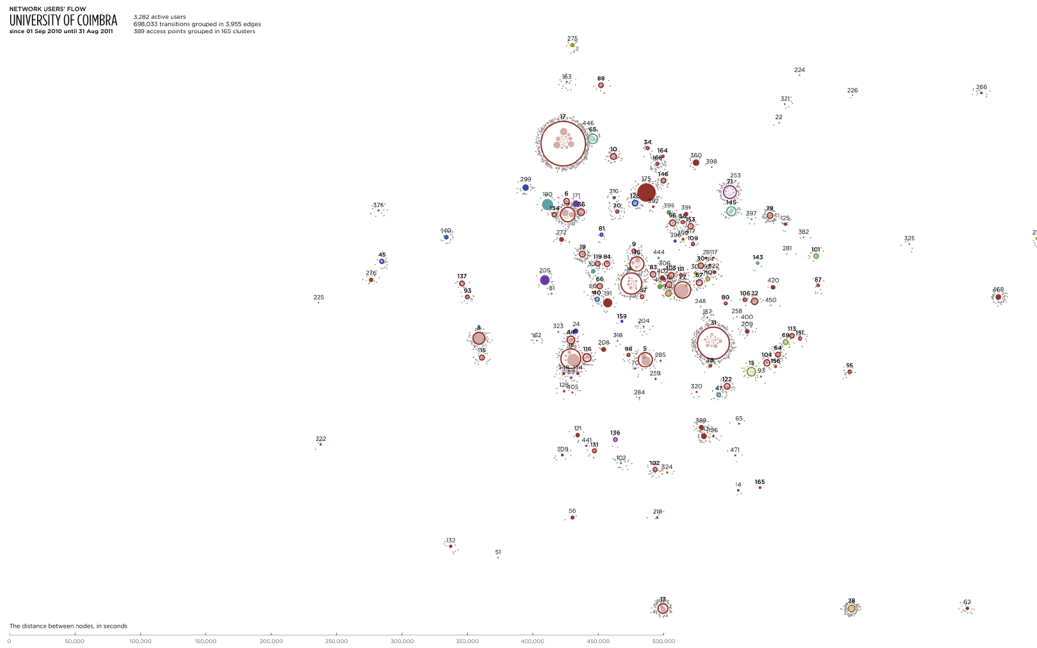


Figure 7.35. The graph showing the relative topology of the UC network on the academic year 2010-2011.

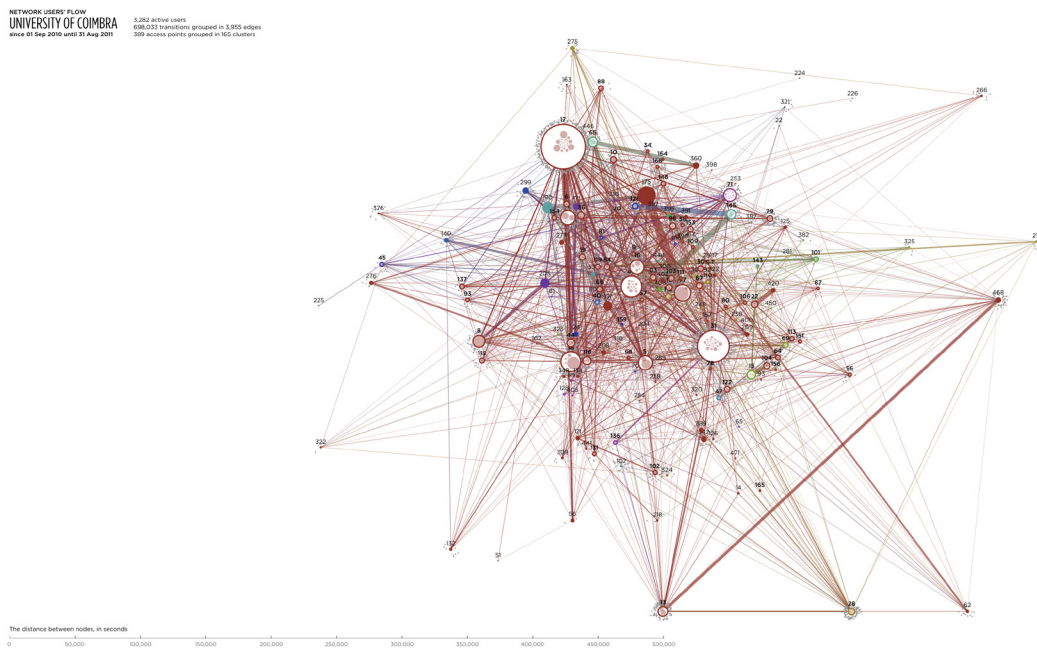


Figure 7.36. The graph of the user flow in the UC network on the academic year 2010-2011.

## 7.4. The Relative Geographic Flow of Users

This section illustrates and describes the visual results obtained with *The Map of the Wireless Network* model. Firstly, the static artifacts for each dataset and academic year are analyzed. Each graph is shown with two versions: one featuring nodes identified with numbers and the users connected to them; another displays the same elements and shows the transition paths of users between nodes. These graphs are saved in the image format after all forces come to a mechanical equilibrium. After the static graphs, a dynamic artifact is presented.

FIGURE 7.35 shows the relative map of the UC network, based on transitions of users between distinct APs in the academic year 2010-2011. One can observe two large clusters, #17 and #31 with 30 and 32 APs, respectively. Their size suggests that these are the most commonly accessed clusters in that year. These are connected to the majority of red clusters due to the relative proximity, and have many users connected to them on 31 August 2011. The remaining nodes are situated between and around these clusters. Many nodes group together [FIGURE 7.36] when the average transition times between them are small. Several nodes are pulled to the exterior of the main action by the springs with big transition values. Yet, these nodes are still accessed by various users that move around the node shapes. The whole network is deeply interconnected [FIGURE 7.37]. The major part of the transition links is thin, presenting small amount of grouped user transitions. Few edges are thick, indicating the most common routes in the network: #17 → #4, #175 → #77, #128 → #145, #145 → #402.

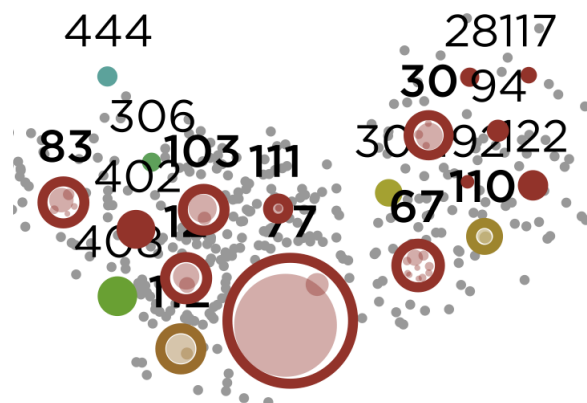


Figure 7.37. A group of nodes in the UC network.

NETWORK USER FLOW  
UNIVERSITY OF LISBON  
since 01 Sep 2011 until 31 Aug 2012

1369 active users  
194.532 transitions grouped in 2.073 edges  
136 access points grouped in 232 clusters

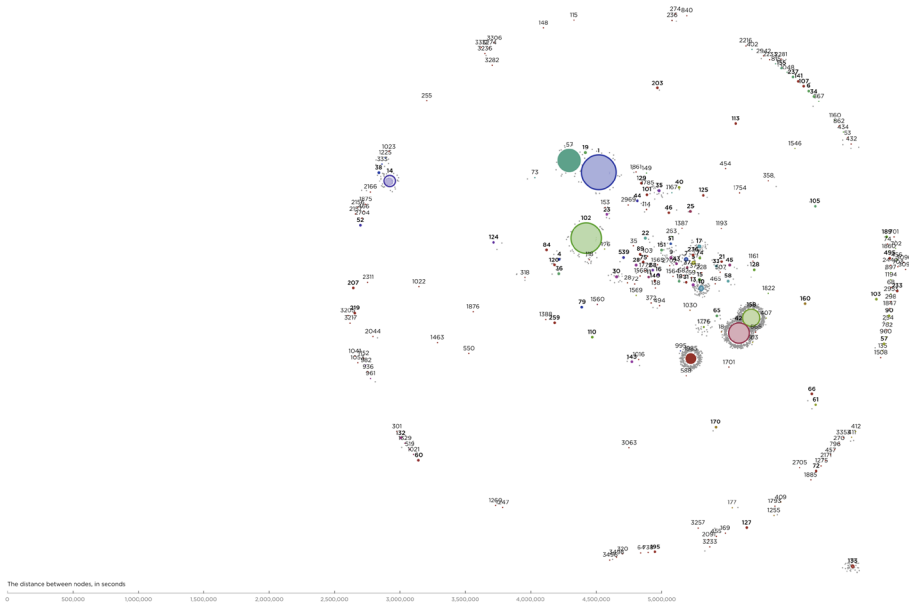


Figure 7.38. The relative topology of the UL network regarding the academic year 2011-2012.

NETWORK USER FLOW  
UNIVERSITY OF LISBON  
since 01 Sep 2011 until 31 Aug 2012

1369 active users  
194.532 transitions grouped in 2.073 edges  
136 access points grouped in 232 clusters

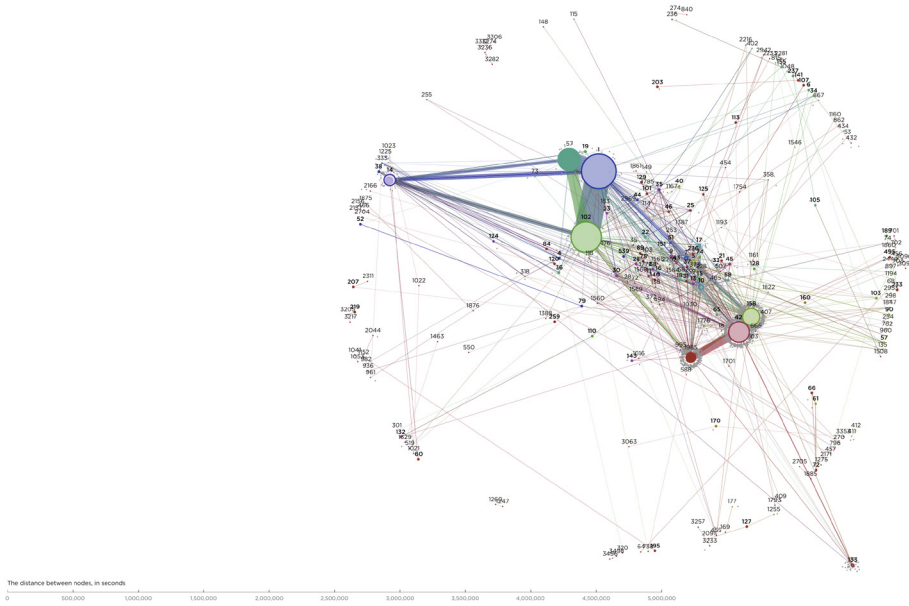


Figure 7.39. The visualization of the user flow in the UL network on the academic year 2011-2012.

In the UL network [FIGURE 7.38], the situation differs. The six most common nodes (#57, #1, #102, #1985, #42, #158) form the transitional core of the network, and the smaller intermediate nodes are attracted to them. The cluster #14 is pulled out to the periphery by common, yet long transition routes. The nodes that maintain located along a circumference have thin links with big values of the transition time to the main network, the users do not access these APs so much. The crucial routes of this network are: #57 → #1, #1 → #102, #57 → #102, #158 → #142, #142 → #1985 [FIGURE 7.39].

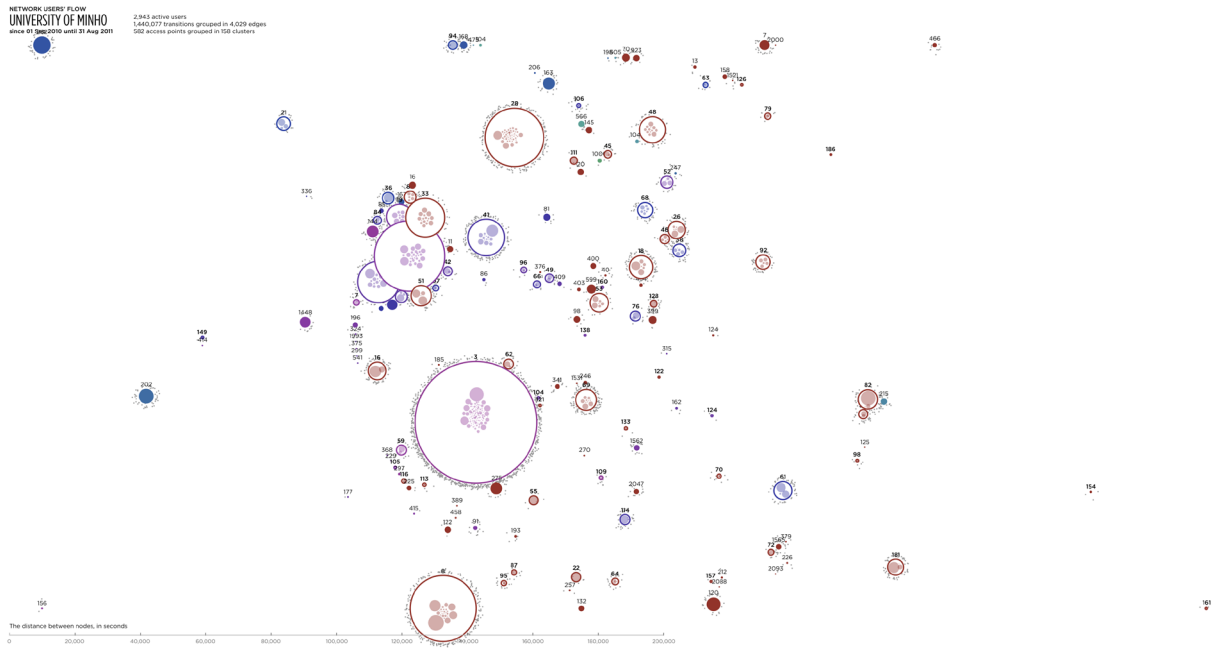


Figure 7.40. The physical structure of the UM network on the academic year 2010-2011.

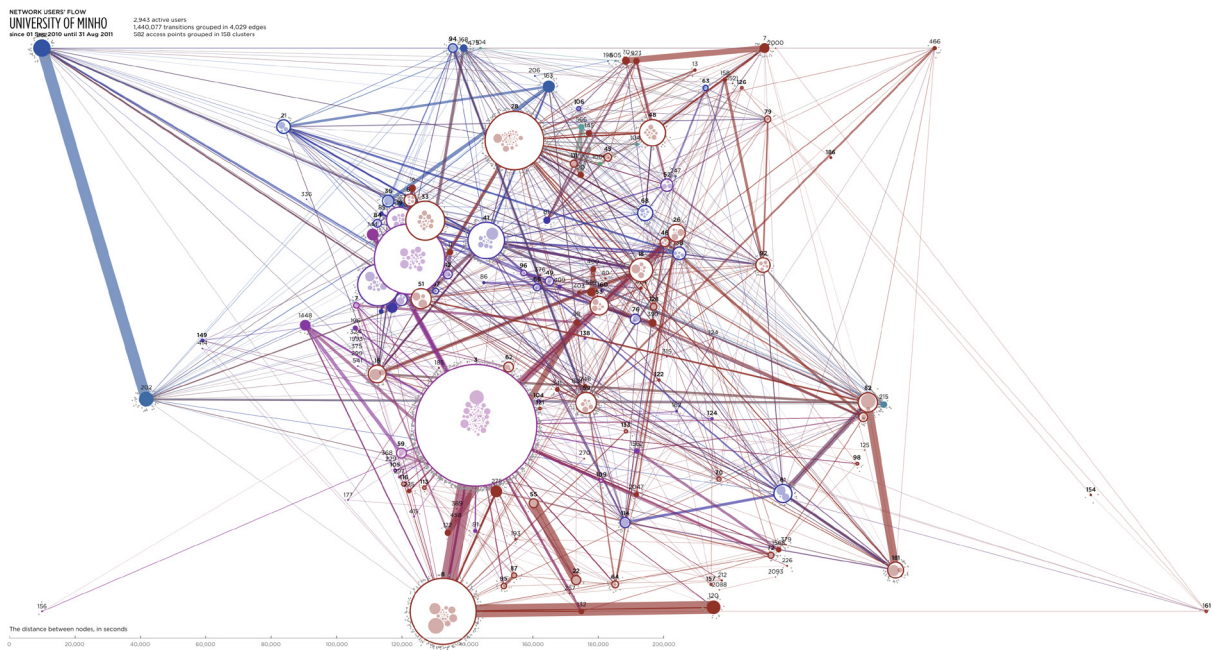


Figure 7.41. The user flow in the UM network on the academic year 2010-2011.



The visualization of the relative topology of the UM network is shown on FIGURE 7.40. There can be seen a variety of big clusters with a great quantity of incoming and outgoing transitions. Each of these clusters groups many APs, so we can deduce that the UM campus has various spots frequently accessed by users. The nodes are dispersed in space by the length of linking edges that registered many long transitions. One can observe the big group of nodes at the center of the graph [FIGURE 7.42]. The transition times between these APs and clusters are so small that joining them in an almost unique shape. All these APs could be included in one supercluster if there were performed transitions with less than 60 seconds between the individual clusters. The user flow [FIGURE 7.41] is rich, all nodes are deeply interconnected. The major routes are traced between the following pairs of distant nodes: #262 → #202, #3 → #8, #8 → #120, #82 → #181.

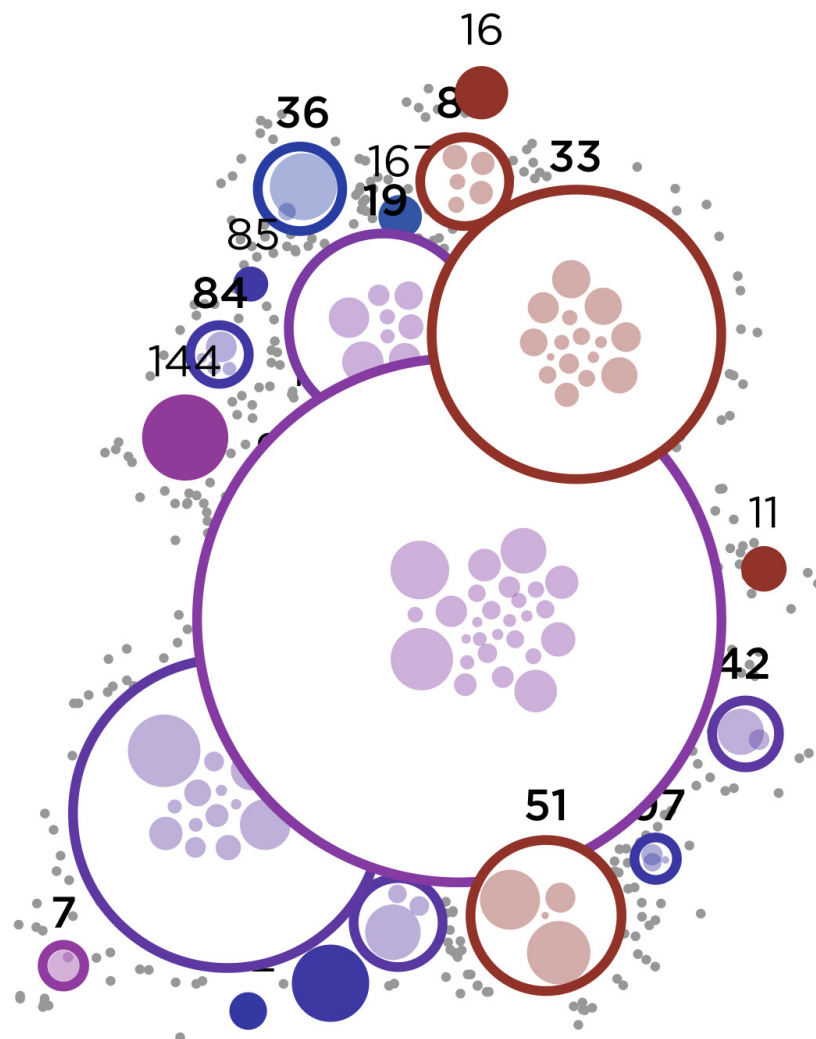


Figure 7.42. The large group of nodes in the UM network.



Figure 7.43. Three frames from the animation of the flow of users of the UC.

The dynamic artifact, illustrated on FIGURE 7.43, is the visualization of the global flow of users in the network<sup>9</sup>. The animated sequence has a *cumulative effect*, accumulating the information during the playback. The current frame displays all transitions performed by network's users in the temporal interval shown at the top of the video. The frames advance in two hour increments. When a transition happens, the user relocates to the target node, leaving a grey-colored dashed path for a very short period of time. The entire animation, starting from September 2010 and ending in August 2011, demonstrates how the network's relative topology and inner structure change with the accumulation of user transitions. The animation of the human movement allows the temporal analysis of the transitionality in the network.

Before the animation, the *textual messages* explain to the viewer how to interpret the posterior action in the visualization. When the animation starts, the nodes, initially placed along a circumference, are being positioned accordingly to the first transitions and the respective times, mapped to the length of the connecting edges. From the beginning of the video, one can notice the periodical appearance of a big quantity of the grey paths. This indicates that the majority of users transit within the network during a certain temporal interval once a day, presumably during the daytime. When an edge between two node accumulates more than 25 unique transitions, it becomes visible. During the entire animation, the nodes gain incoming and outgoing transitions, becoming bigger. The size of the grey circles around a node indicates the amount of users connected to it at that instant. The edges group dynamically, pushing or pulling the linked nodes accordingly to the transition time distance. Thus, the viewer can see the APs and clusters approaching and moving away from each other. The graph becomes more and more complex and dense as the transitions appear along the timeline. By the end of January 2011, almost all nodes are linked with edges. There are several periods of higher and lower transitionality, indicated in the dynamic artifact of *The Circuit of Transitions* [SECTION 7.1, P. 127]. These intervals are also applicable to this video. The final frame of the video shows the structure and the user flow in the network in the academic year 2010-2011.

<sup>9</sup> The animated visualization can be seen through the hyperlink: <https://vimeo.com/72561076>

This technique for visualization of the user flow in a wireless network is based on a force-directed layout. The graph allows the visualization of the relative topology of a network and the physical structure of APs and clusters extracted from the transitions' data. The spatial distribution of nodes is made with the length of edges proportional to the average transition time values. The flow of users is shown with the thickness and the color of the edges. Different network datasets present distinct structures and topologies, varying in the distribution of users by nodes and in the common routes that trace the flow lines in the network.

## 7.5. Summary

This chapter demonstrated and described the visual results obtained with the four visualization techniques created in the context of this research. Different types of artifacts are presented to show the versatility and effectiveness of the developed models. The analysis is based on the graphical properties of the visualizations, interpreting the global appearance and the visual elements of each graph. Also, a contextual analysis is performed in order to link the results with the real academic context of the universities. The next chapter resumes the whole research, draws the final conclusions and specifies the future work.





## CHAPTER 8.

# Conclusions and Future Work

This research is focused on the development of methods for visualization of flows and usage patterns in wireless networks. The process is split into a theoretical and a practical parts. In the first part, the synthesis of various layouts existing in the field of information visualization is performed [CHAPTER 2]. The respective examples are described and analyzed, and the applicability of each layout for the visualization of flows is discussed. The graphic semiology, along with the main concepts and techniques of construction and analysis of visualizations are explained, based on the work of Jacques Bertin and Edward Tufte. The variety of interactive functionalities for exploration and manipulation of graphs and data is explained. The work related to this research is analyzed in CHAPTER 3. The purpose of scientific visualization and a variety of methods for representing flows are explained, followed by the analysis of various examples of networks, flows and complex graphs. Those two chapters illustrate and synthesize the theoretical study related to this dissertation.

The goals and the methodology of the research are explained in CHAPTER 4. The main objective was to develop techniques for visualization of flows in wireless networks. That section introduces the practical part of this work. In CHAPTER 5, the preparatory process is explained. The available network data and the process of its treatment are presented. The experimental prototypes of visualization are implemented to discover the most suitable layout for representation of networks and flows. The process of construction of four interactive visualizations is described in CHAPTER 6. Each pragmatic approach is implemented from the ground up, based on the theoretical studies from CHAPTERS 2 AND 3. Different aspects of the data are represented to provide the most complete system of visualization of patterns of flow and usage in wireless networks. Lastly, the experimental results obtained with the models are described in CHAPTER 7. The static and dynamic artifacts are analyzed, as well as the adopted visual techniques, to draw conclusions about their efficiency.

The visualization technique called *The Circuit of Transitions* represents the global movement of users within a wireless network, portraying the flow and the patterns of user transitions. The access points of the network and the transition links are shown in the radial convergence diagram. The node clustering and ordering algorithms allow the grouping and sorting of the APs by relative proximity, providing visual clarity to the network's structure and displaying additional information. The graphic system permits the identification of the more common and the less common routes, the short and the long transition paths, the origin and the destination point of each transition, the most and the least accessed APs and clusters of APs.

The visualization model called *The Matrix of Connections* displays the user connections within a network. The overall and particular connectivity is revealed by duration, density and periodicity of accesses. A graphic of total connection times can also be visualized. The order of the users shows the hierarchy in information, while the timeline localizes the temporal intervals each user spent online. The model depicts the patterns of connectivity and the specific temporal periods regarding the visual density of connections. The common timeframes of access to the network are shown with the vertical repetition of red traces, located on the grid.

The technique for visualization of aggregated data called *The Panels of Flow Shapes* demonstrates peaks, lows and rhythms in the connection and in the transition of users in wireless networks. The amounts of connections/transitions can be compared with the quantities of users, and four values of connection/transition time (minimum, average, maximum and total) can be compared between themselves. The patterns of usage of the network are then revealed by choosing three time granularity levels, being shown with two impositions.

The model of visualization called *The Map of the Wireless Network* shows the structure of the network and the flow of its users. The system's behaviors are inspired by the laws of Physics, providing an organic movement to the graph's elements. All elements are affected by balanced forces with different effects and strengths. The clustering technique creates visual groups of APs. The user move around the nodes they are connected to. The pairs of nodes are distributed in space due to the mechanical force of the connecting edge with distance proportional to the average transition time. The common APs and clusters are highlighted with bigger sizes, and the common transitional routes are emphasized with bigger thicknesses.

In all four applications, the textual information identifies the current dataset, provides quantitative statistics of the data and explains important aspects of the data and of the graphing elements. The quantitative reading of the data in each application is provided by the respective grids. The graphic user interface and the implemented interactive functionalities allow the manipulation and the exploration of the displayed data. From our point of view, the constructed graphic systems are proven to be effective, responding to the established objectives and solving the proposed visualization problems. The produced visual artifacts demonstrate efficiency



in presenting understandable, interpretable and comprehensible results, allowing exploration, study and analysis of the data, and revealing the insight of the complex information. The graphic language used in our models of visualization is simple and minimalistic, yet all graphs require an initial contextualization due to the complexity of the displayed multivariate and technical data. *The Circuit of Transitions* is probably the most complex of the four visualization systems. The abstract structure of the network, presented in that model, needs to be well understood before exploring the data. The color encoding of directions of the transition edges is not straightforward, and perhaps a simple animation could solve the problem.

As expected, this research contributed with the theoretical review of the fields of information visualization and scientific visualization; the techniques and methods for visualization of flows in wireless networks; the interactive software presenting the application of the developed techniques; the results in image and video formats, and supporting materials; the respective analysis and the scientific abstract reporting the work on *The Circuit of Transitions* accepted at the *ACM SIGGRAPH 2013* conference and presented in the poster format; and the dissertation final report.

For future work for visualization of patterns of flow and usage in wireless networks, we specify two main paths. The first consists of the creation of an application that merges the four visualizations into one with one main thematic that would connect the different visualization paradigms. That thematic could be the time since all our models use temporal intervals to represent the data features. Having that, one could have a broader view of the data in one application, that could provide a more in-depth analysis of the results or extract hidden meanings from the data.

The second path would be to gather the information about real localizations of APs for each network and create another model of visualization based on that data. The comparison of that new model with *The Map of the Wireless Network* could reveal as much differences as similarities, producing interesting results and parallelisms. Also, a local-based analysis of APs and clusters will help to identify where the most accessed points are located and understand why these are frequented. Therefore, more evidences of the human behavior can be spotted. The node clustering and ordering algorithms could also improve if based on that information.

Finally, the improvement of particular aspects of the visualization techniques and of the applications are also important. Improvement of the graphic properties, optimization of the software performance, improvement of the results of the clustering and the sorting algorithms would contribute to the further work on the models. Implementation of the zooming technique in *The Map of the Wireless Network* could contribute with a higher degree of exploration of information. Also, the creation of web-based applications available to the public could contribute to better distribute the work and to explore new ideas and paths in the visualization of flows and human usage patterns in networks.



# References

- Artmedia (2012) *The New data visualization: this is not your fathers pie chart*. [Online]. Available from: <http://www.artmedia.com/blog/2012/03/the-new-data-visualization-this-is-not-your-fathers-pie-chart/> [Accessed 22 June 2013].
- Barron, R. (2007) Bringing the map to life: European satirical maps, 1845-1945, in: *6th International BIMCC Conference*, Brussels, Belgium, November 2007. Brussels: BIMCC, pp. 25-27.
- Bertin, J. (2011) Semiology of the graphic sign-system, in: *Semiology of graphics – diagrams, networks, maps*. 1st ed. Redlands: Esri.
- Big Think (2009) 377 – *Planet of the grapes*. [Online]. Available from: <http://bigthink.com/strange-maps/377-planet-of-the-grapes> [Accessed 8 December 2012].
- Bostock, M. (2010) *Gallery*. [Online]. Available from: <https://github.com/mbostock/d3/wiki/Gallery> [Accessed 16 August 2013].
- Brinton, W.C. (1939) *Graphic presentation*. 1st ed. New York: Brinton Associates.
- Britton, J. [no date] *Beck's London Underground map*. [Online]. Available from: <http://britton.disted.camosun.bc.ca/beckmap.htm> [Accessed 8 December 2012].
- Brodie, K.W., Carpenter, L.A., Earnshaw, R.A. *et al* (1992) What is Scientific Visualization?, in: *Scientific Visualization: Techniques and Applications*. Berlin: Springer-Verlag, pp. 1-3.
- Bruner, J. (2011) *Nobel laureates by country and prize*. [Online]. Available from: <http://www.forbes.com/sites/jonbruner/2011/10/05/nobel-prizes-and-american-leadership-in-science-infographic> [Accessed 8 December 2012].
- Buchin, K., Speckmann, B. and Verbeek, K. (2011) Angle-Restricted Steiner Arborescences for Flow Map Layout, in: *ISAAC'11 Proceedings of the 22nd International Conference on Algorithms and Computation*, Yokohama, Japan, December 2011. New York: Springer, pp. 250-259.
- Burke, C. (2009) Isotype representing social facts pictorially, in: *Data Designed for Decisions Conference*, Paris, France, June 2009. Vienna, Austria: OECD, p.2.
- Byron, L. (2008) *New York Times – ebb and flow at the box office*. [Online]. Available from: <http://leebyron.com/what/boxoffice/> [Accessed 25 June 2013].
- Byron, L. and Wattenberg, M. (2008) Stacked graphs – geometry & aesthetics, *IEEE Transactions on Visualization and Computer Graphics*, 14 (6), pp. 1245-1252.
- Card, S., Mackinlay, J. and Shneiderman, B. (1999) *Readings in information visualization – using vision to think*. San Francisco: Academic.
- Costigan-Eaves, P. and Macdonald-Ross, M. (1990) William Playfair (1759-1823), *Statistical Science*, 5 (3), pp. 318-326.
- Dawes, B. (2004) *Cinema Redux*. [Online]. Available from: <http://brendandawes.com/projects/cinmaredux> [Accessed 26 June 2013].
- Dayrat, B. (2003) The Roots of phylogeny: how did Haeckel build his trees?, *Systematic Biology*, 52 (4), pp. 515-527.

De Leeuw, W. and van Liere, R. (1997) Divide and conquer spot noise, in: *Proceedings of the 1997 ACM/IEEE Conference on Supercomputing*, San Jose, USA, November 1997. New York, USA: ACM, pp. 1-13.

Dias que Voam (2012) *Portugal*. [Online]. Available from: <http://diasquevoam.blogspot.pt/2012/05/portugal.html> [Accessed 8 December 2012].

Dorling, D. (1996) Area cartograms: their use and creation, *Concepts and Techniques in Modern Geography (CATMOG)*, 59.

Earnshaw, R.A. and Wiseman, N. (1992) *An Introductory guide to scientific visualization*. 1st ed. Berlin: Springer-Verlag, pp. 5-6.

Eisenberg, L. (2008) *The Great tree of life*. [Online]. Available from: <http://evogeneao.com/tree.html> [Accessed 8 December 2012].

Flow, Just Flow (2013) *Aaron Koblin*. [Online]. Available from: <http://www.flowjustflow.com/2013/01/29/aaron-koblin/> [Accessed 25 June 2013].

Fry, B. (2005) *Distellamap*. [Online]. Available from: <http://benfry.com/distellamap/> [Accessed 26 June 2013].

Fry, B. (2008) *Visualizing data*. 1st ed. Sebastopol: O'Reilly Media.

Fry, B. and Reas, C. (2004) *Processing*. [Online]. Available from: <http://processing.org> [Accessed 28 November 2012].

Galloway, A.R. (2004) *Protocol: how control exists after decentralization*. Cambridge: The MIT Press, p. 11.

Geis, I. (1939) Some highlights of the business opinion, *Survey Graphic Magazine*, September 1939. Available online from: <http://www.fulltable.com/iso/rm/am.htm> [Accessed 8 December 2012].

Gephi (2008) *Index*. [Online]. Available from: <https://gephi.org> [Accessed 25 December 2012].

Gill, C.J. and Gill, G.C. (2005) Nightingale in Scutari: her legacy reexamined, *Clinical Infectious Diseases*, 40 (12), pp. 1799-1805.

Harris, R.L. (1999) *Information graphics: a comprehensive illustrated reference*. New York: Oxford University Press.

Heinen, F. (2007) *Social network – data visualisation*. [Online]. Available from: <http://www.felixheinen.de/Social-Network-Data-Visualisation> [Accessed 26 June 2013].

Hint.fm (2009) *Flicker flow*. [Online]. Available from: <http://hint.fm/projects/flickr/> [Accessed 25 June 2013].

Historic Urban Plans (1965) *Kanton in platte grondt – plate forme de la ville de Kanton*. Ithaca: Historic Urban Plans.

Holten, D. (2006) Hierarchical edge bundles: visualization of adjacency relations in hierarchical data, *IEEE Transactions on Visualization and Computer Graphics*, 12 (5), pp. 741-748.

Iselburg, P. (1616) *Fasiculus Rariorum*. Available online from: <http://bibliodyssey.blogspot.pt/2006/12/curious-bundle.html> [Accessed 7 December 2012].

Jain, A.K. and Dubes, R.C. (1988) *Algorithms for clustering data*. New Jersey: Prentice-Hall, pp. 1-6.

Krzywinski, M. (2011) *Rational network visualization – farewell to hairballs*. [Online]. Available from: <http://www.hiveplot.net/> [Accessed 22 June 2013].

- Lau, A. and Moere, A.V. (2007) Towards a model of information aesthetics in information visualization, in: *11th International Conference on Information Visualization*, Zürich, Switzerland, July 2007. Los Alamitos, CA: IEEE Computer Society, pp. 87-92.
- Manovich, L. (2011) What is visualization?, in: *DIGAREC Keynote-lectures 2009/10*. Potsdam: University Press, pp. 116-156.
- Marey, E.J. (1885) *La Méthode graphique dans les sciences expérimentales*. Paris: G. Masson.
- Mashantucket Pequot Museum (2009) *The History of the Pequot war*. [Online]. Available from: <http://pequotwar.org/2009/10/the-pequot-war> [Accessed 7 December 2012].
- McCormick, B.H. *et al* (1987) Visualization in scientific computing, *Computer Graphics*, 21 (6).
- McLeod, K.S. (2000) Our sense of Snow: the myth of John Snow in medical geography, *Social Science & Medicine*, 50 (7-8), pp. 923-935.
- McLoughlin, T., Laramée, R.S. and Zhang, E. (2010) Constructing streak surfaces for 3D unsteady vector fields, in: *Proceedings of the 26th Spring Conference on Computer Graphics*, Budmerice, Slovakia, May 2010. New York, USA: ACM, pp. 17-26.
- Miner, D. and Kasch, N. (2008) SwarmVis: a tool for visualizing swarm systems, *UMBC Computer Science: Data Visualization*, 636.
- NASA Goddard Space Flight Center (2012) *Global sea surface currents and temperature*. [Online]. Available from: <http://svs.gsfc.nasa.gov/vis/a000000/a003900/a003912/index.html> [Accessed 20 June 2013].
- National Geographic (2008) *Fifty years of exploration*. [Online]. Available from: <http://ngm.nationalgeographic.com/map/map-day/index> [Accessed 5 June 2013].
- Norman, J. (2012a) *A Wallpainting that could be a landscape or a map (circa 6,200 BCE)*. [Online]. Available from: <http://www.historyofinformation.com/expanded.php?id=1745> [Accessed 7 December 2012].
- Norman, J. (2012b) *Perhaps the oldest map in the world (10,000 BCE)*. [Online]. Available from: <http://www.historyofinformation.com/expanded.php?id=6> [Accessed 7 December 2012].
- Oakland Museum of California [no date] *Early California: pre-1769-1840s: early explorers*. [Online]. Available from: <http://www.museumca.org/picturethis/pictures/map-western-hemisphere> [Accessed 7 December 2012].
- Phan, D. *et al* (2005) Flow map layout, in: *IEEE Symposium on Information Visualization (InfoVis 2005)*, Minneapolis, USA, October 2005. Los Alamitos, CA: IEEE Computer Society, pp. 219-224.
- Playfair, W. (1801) *The Commercial and political atlas*. 3rd ed. London: J. Wallis.
- Post, F.H. *et al* (2002) Feature extraction and visualisation of flow fields, in: *Eurographics 2002 State of the Art Reports*. Saarbrücken: Eurographics Association, pp. 69-100.
- Reynolds, C.W. (1987) Flocks, herds, and schools: a distributed behavioral model, *ACM Computer Graphics*, 21 (4), pp. 25-34.
- Rose, F.W. (1900) *John Bull and his friends - a serio-comic map of Europe*. London: GW Bacon. Available online from: <http://bibliodyssey.blogspot.pt/2009/06/satirical-maps.html> [Accessed 7 December 2012].
- Sadarjoen, A. and Post, F.H. (2000) Detection, quantification, and tracking of vortices using streamline geometry, *Computers & Graphics*, 24 (3), pp. 333-341.

- Salim-Atchekzai, M. (1974) A Pioneer of scientific observation, *The UNESCO Courier Magazine*, 27, pp. 16-17.
- Scagnetti, G. (2012) The Diagram of information visualization, *Parsons Journal for Information Mapping*, 4 (4).
- Schmidt, K. (2007) *Toxiclibs*. [Online]. Available from: <http://toxiclibs.org/about> [Accessed 24 January 2013].
- Sommerer, C., Mignonneau, L. and King, D. (2009) *Interface cultures: artistic aspects of interaction*. Bielefeld: Transcript Verlag, pp. 27-38.
- Spence, R. (2001) Presentation, in: *Information visualization*. Harlow, England: Addison-Wesley.
- Stefaner, M. (2005) *Visualizing information flow in science*. [Online]. Available from: <http://moritz.stefaner.eu/projects/eigenfactor/> [Accessed 25 June 2013].
- The History Blog (2011) *Library of Congress gets unique flat earth map*. [Online]. Available from: <http://www.thehistoryblog.com/archives/11651> [Accessed 7 December 2012].
- The Opte Project (2003) *Home*. [Online]. Available from: <http://opte.org/> [Accessed 25 June 2013].
- Tufte, E.R. (1990) *Envisioning information*. 10th ed. Cheshire: Graphics Press.
- Tufte, E.R. (2001) *The Visual display of quantitative information*. 2nd ed. Cheshire: Graphics Press.
- Tufte, E.R. (2005) *Visual explanations*. 7th ed. Cheshire: Graphics Press.
- Tufte, E.R. (2006) *Beautiful evidence*. 1st ed. Cheshire: Graphics Press.
- Turk, G. and Banks, D. (1996) Image-guided streamline placement, in: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New Orleans, USA, August 1996. New York, USA: ACM, pp. 453-460.
- Velho, B. (1568) *Cosmographia*, Bibilotèque National, Paris. Available online from: [http://en.wikipedia.org/wiki/Bartolomeu\\_Velho](http://en.wikipedia.org/wiki/Bartolomeu_Velho) [Accessed 7 December 2012].
- Viégas, F.B., Wattenberg, M. and Dave, K. (2004) Studying cooperation and conflict between authors with history flow visualizations, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria, April 2004. New York, USA: ACM, pp. 575-582.
- Viola, I. and Gröller, M.E. (2005) Smart Visibility in Visualization, in: *Proceedings EG Workshop on Computational Aesthetics*, Girona, Spain, May 2005. Wadern, Germany: Schloss Dagstuhl, pp. 209-216.
- Watkins, D. (2011) *Mapping generic terms for streams in the contiguous United States*. [Online]. Available from: <http://museinfo.cci.fsu.edu/omeka/items/show/737> [Accessed 8 December 2012].
- Wattenberg, M. (2002) Arc diagrams: visualizing structure in strings, in: *IEEE Symposium on Information Visualization*, Boston, USA, October 2002. Los Alamitos, CA: IEEE Computer Society, pp. 110-116.
- Weskamp, M. (2003) *Social circles*. [Online]. Available from: <http://marumushi.com/projects/socialcircles> [Accessed 26 June 2013].
- Wikipedia contributors (2004) *Vector field*. [Online]. Available from: [http://en.wikipedia.org/wiki/Vector\\_field](http://en.wikipedia.org/wiki/Vector_field) [Accessed 16 August 2013].



