Market segmentation and travelling patterns based on Tourist's paths in the Porto Region

By

Ana Ramos Silva

Dissertation in Marketing

Supervised by:

Professor Pedro Campos

Professor Carlos Ferreira

Porto, 15th of September, 2017

“Se na nossa cidade há muito quem troque o b por v, há quem pouco troque a liberdade pela servidão.”

Almeida Garret

Acknowledgments

To my supervisors – Dr. Pedro Campos and Dr. Carlos Ferreira – for all the guidance, patience and wisdom.

To my colleagues from BPI Bank, for all the support and understanding at the beginning of this journey. A special thanks to my director – Eng. Mário Fernandes – for his understanding, caring and faith in me.

To my dear best friend – Patrícia Bruno– for the unconditional support.

Index

Abstract

The increase of mobile and capture devices, like smart phones, digital cameras and tablets, are feeding the growth of online information in the past few years. For these factors, travelers are reporting their experiences through opinions, videos and photos shared on websites and online communities. Thanks to GPS technology, geographical and temporal information can be *tagged* on the photos. Information provided by such *geotagged* photos allows knowing where and when people are, providing a better understanding of tourists' travelling patterns in the destinations that can be used for marketing and strategical purposes. The aim of this dissertation is to make a market segmentation based on tourists' travelling patterns, through the analysis of photos with *geotagged* data in the city of Porto. To achieve such purpose, we applied a quantitative methodology using data from secondary sources and by performing a Sequence Mining analysis (to track the paths), a Social Network Analysis (to identify networks of sites) and a Cluster Analysis (to perform a market segmentation). The photos were collected from *Flickr.com* Social Network Site over a period of 29 months (from January 2014 to May 2016). The results can be used to support and design new marketing strategies in the Porto tourism market.

Keywords: Tourism marketing; Market segmentation; Big Data; Geotagged photos; Movement patterns; Social Networks; Frequent movements; Sequence Mining; Cluster Analysis; Flickr.com;

Resumo

O crescimento dos dispositivos móveis e equipamentos de captura (como os smartphones, câmaras digitais e tablets) estão a alimentar o aumento da informação disponível online nos últimos anos. Como causa, e consequência, os turistas partilham as suas experiências de viagem - através de opiniões, vídeos e fotografias - que partilham na web e nas comunidades online. Graças à tecnologia GPS, a informação geográfica e temporal pode ser associada ("taggada") às fotografias e a informação fornecida por estas fotos permite-nos saber quando e onde uma pessoa esteve num determinado local, o que ajuda na compreensão dos padrões de comportamento dos turistas nos destinos. O objetivo desta dissertação é realizar uma segmentação de mercado, com base nos padrões de movimento dos turistas na cidade do Porto, através de fotograficas com informação geo-temporal. Para atingir este objetivo, foi utilizada uma metodologia quantitativa, através de informação de fontes secundárias e através da realização de uma análise de Sequence Mining, de uma análise de Redes e de uma análise de Clusters. As fotografias foram extraidas da rede social Flickr.com correspondendo ao espaço temporal de 29 meses (de Janeiro de 2014 a Maio de 2016). Os resulados podem ser usados para suportar e desenvolver estratégias de marketing aplicadas ao mercado turístico do Porto.

Palavras-chave: Marketing turístico; Segmentação de mercado; Big Data; Padrões de movimento; Redes Sociais; Movimentos frequentes; Sequence Mining; Análise de custers; Análise de Redes; Flickr.com;

List of Tables

List of Figures

elaboration.

Figure 15: Social network based on the attractions photographed/visited, using the database collected from Flickr since January, 2014 to May, 2016 (Yifan Hu Proportional layout). Source: Own elaboration.

Figure 16: Social network based on the attractions photographed/visited, using the database collected from Flickr for 2014, 2015 and 2016 respectively (Furchterman layout). Source: Own elaboration

Figure 17: Network Centre and most weighted nodes based on the attractions photographed/visited, using the database collected from Flickr between January 2014 and May 2016. Source: Gephi software.

Chapter 1 | Introduction

In this chapter an overview related to this study structure will be presented.

Some introductory notes are presented on section 1.1; the complementary work is presented on section 1.2.In section 1.3, the software tools used were presented.

The chapter ends with the section 1.4 where it is presented the structure of the thesis

1.1 Introductory notes

The increase of mobile and capture devices encompasses the user-generated content phenomenon that has been created in the past few years. Thanks to GPS technology, which is incorporated in such devices, it is possible to know where and when tourists have been because individuals share their travel experience – through photos and videos- on social network sites. The photos shared by tourists have important information for tracking studies like location, time and tags. In the marketing area, user-generated content provided ability of implement market segmentation in a more effective way (Dahan and Srinivasan, 2000 *cf* Wind and Bell, 2007; Kádár and Guede, 2013; Vu *et al.*, 2015).

The theme of the present dissertation relies on a market segmentation process through the analysis of individuals' movement patterns. Market segmentation is a tool of strategic marketing that helps to understand and serves the needs of homogeneous consumer groups. For a destination manager, it is one of the most important decisions to make, in terms of long-term marketing strategies. Market segmentation can be applied by any organizations in any industry in order to make marketing mix more effective and targeted to a specific group composed by tourists with similar needs, desires and motivation (Dolnicar, 2007; Beeco and Hallo, 2014). Data generated by tourists' travelling patterns has several implications for managers engaged in strategic planning. Furthermore, tourist's behaviour patterns information can support market segmentation purposes.

Photos were collected from *Flickr.com* over a period of 29 months (from January 2014 to May 2016) and this study takes place in the city of Porto, Portugal. Based on the time and location of the photos collected, along with personal information such as origin and gender, it was possible to create a graph with tourists' activity and the corresponding tracks in space and time. While the Social Network Analysis will help us to understand the connections between the places and sites visited by tourists, the Sequence Mining Analysis will then allow the identification of the most frequent tourist patterns across the city. Finally, the Cluster Analysis will help identify tourist segments and fulfill market segmentation, based on demographic, geographic and behavioural information.

With respect to the research questions, this study aims to answer the following ones:

- Where do tourists go? What are their trips and paths, through space and time?
- How do networks, between different locations, help us define tourists' behaviour?
- How tourists' travelling patterns information, along with personal data, can be used to fulfill market segmentation?

This study is aimed at all companies related to the tourism sector, which includes, for example, hotel and transport industries, airlines, travel agencies and restaurants. The study is also addressed to political organizations and elected officials, local governments, decision makers, managers and professionals involved in the marketing area.

The Tourists' travelling patterns in urban areas are considered a complex matter mainly because bigger cities tend to be more complex (Asakura and Irvo, 2007) and therefore few studies have addressed this phenomenon (Edwards and Griffin, 2013). Despite the transient nature of tourism industry, a precise understanding of travel behaviours is critical (Edwards, Griffin, Hayllar, Dickson and Schweinsberg, 2009 *cf* Vu *et al.*, 2015) to provide insights about travel demands and travel flows, to help destinations in planning, development and management and, finally, to help planning business strategy in order to attract visitor to the destiny (Li *et al.*, 2008).

1.2 Complementary work

Some researchers used *geotagged* photos from Flickr.com in order to track tourists in cities. One of the most recent works on this subject has been made by Vu *et al.* (2015) in Hong Kong, aimed at delivering useful insights on destination development, transportation planning, and impact management. Spyrou, Sofianos and Mylonas (2015) focused on the analysis of user-generated touristic routes within urban areas. Kádár and Guedes'work (2013) was done in the city of Budapest and its purpose was to study tourists' spatial distribution. Crandall and Snavely (2011) and Ji *et al.* (2011) analyzed the relation between the location and the content of photos. Girardin *et al.* (2007) research took place in the Province of Florence and it addressed urban planners, traffic engineers, and tourism authorities.

However, none of the mentioned works aimed to fulfill market segmentation. For the present study, it is important to highlight the research done by Vu *et al.* (2015), Leung *et al.* (2012) and Xia *et al.* (2010). Xia *et al.* (2010) proposed fulfilling market segments based on the dominant movement patterns of tourists, using general log-linear models and the Expectation–Maximization algorithm. Leung *et al.* (2012) as well as Holden (2000 *cf* O'Connor *et al.*, 2005), Lau and Mckercher (2006), Mckercher, Shoval and Birenboim (2012), along with a few others, completed the insights of Xia *et al.* (2010) and Vu *et al.* (2015), by providing alternative classifications for tourist's market segments based on individuals' behaviour. However, the researchers' motivations were not exactly marketing strategies oriented, but instead they oriented towards management and planning and the authors did not used Flickr database to fulfill market segmentation process.

Regarding tourism segments characterization, several attempts of segment tourist market based on their behaviour and further segments classification, for marketing purposes, are possible to find on literature as well as the techniques used to collect tourist data and the segmentation basis. According to Xia *et al.* (2009 *cf* Leung *et al.,* 2012), despite the classification given to the movement patterns, the traveler behaviour is influenced by several factors: human 'push' factors (ex: personal motivations), physical 'pull' factors (ex: destination geomorphology and configuration) and time factors (ex: total trip duration) (Lew and Mckercher, 2006; Lau and Mckercher, 2007 *cf* Leung *et al.,* 2012). Holden (2000 *cf* O'Connor *et al., 2*005) defends a similar idea arguing that Factors such a demographic, culture, lifestyle, level of education, beliefs and attitudes can influence tourists choices and behaviour.

For instance, in 1989, Grönroos defined that tourist market could be segmented into three groups: traditional tourists (individuals travel to a specific location for an extend stay with a specific purpose), day visitors (tourists that just search for a specific service and they leave when they get satisfied) and business visitors (individuals that travel for business reasons and search for travel activities during their free time). In fact, classifying travel behaviour and segmentation became more difficulties as modern travelers combine pleasure and business because there are endless variations between this two classifications since leisure trips elements may be include on business travels and vice versa (Buhalis, 2010).

Lue *et al.* (1992 *cf* Leung *et al.,* 2012) defined five different movement patterns: single destination, enroute, base camp, regional tour and trip chaining pattern.

Flogenfeldt in 1999 *(cf* Leung *et al.,* 2012) identified four types of movement patterns based on Norwegians tourists: day trip, resort trip, based holiday and round trip.

Lau and Mckercher (2006) summarized the movement patterns into six categories: single point, base site, stopover, chaining loop, destination region loop and complex neighborhood. However, these categories are only related to interdestination movement patterns. In other words, Xia *et al.* (2009 *cf* Leung *et al., 2*012) argued that spatio-temporal movements can be classified as a macro level or micro level movements. The macro level corresponds to the interdestination movements, which means that tourists move between destination regions; the micro level is related to the intradestination ones when individuals travel from attraction to attraction or activity to activity (Leiper, 1979; Lau and Mckercher, 2007 *cf* Leung *et al.*, 2012).

Mckercher, Shoval and Birenboim (2012) argue that tourism markets can be defined by a number of dualities ,but, in a simple way, it can be classified in first and repeat visitors, business or pleasure tourists and domestic or international tourists. All these groups show differences in terms of needs, behaviours and responses to stimuli so they require different treatments in terms of product offers, marketing activities and experience provision. First and repeat visitor was also the approach presented by Lau and Mckercher (2006).

Finally, more specifically in Portugal, Coutinho (2012) studied the evolution of tourism in Porto city, its main characteristics and actual tendencies. The author classified the tourists into three different segments the city breaks standard, the city breaks upscale and the city breaks thematic. The first one corresponds to the tourists that travel using low-cost airlines companies and don't spend much money on food and sleepover. The second segment includes the tourists that choose 5 star hotels and gourmet meals (the opposite of the first segment). These tourists search for cultural activities and they have a much higher purchase power than the first segment. The third group includes tourists that fulfill thematic trips (example: musical or sportive events, fashion shows and others) so the money that they are willing to spend is related to the event choosen. In the work of Coutinho is mentioned the segmentation proposed by Portugal Tourism Department (2006), in which tourists are classified as Interactive travelers, people with more than 55 years and college students. Finally, Coutinho (2012) made a reference to the work developed by Kolb (2006) that divided the tourism market into four segments, based on the final purpose of the trip: Business travelers, visiting friends and family, day and weekend tourists and traditional vacationers.

1.4 Structure of the thesis

This thesis is structured as follows: chapter two contains the literature review about the main concepts of this study. The chapter is divided into three sections: section 2.1 approaches some main concepts of marketing and tourism areas; section 2.2 portraits the field of social media and data generation; section 2.3 presents the literature review regarding the methodology applied in this study. Chapter three portrays the methodology application. Chapter four presents the case study - methodology to study tourists' travelling patterns in Porto Region – and the results found and finally, in chapter 5, some recommendations will be presented.

Chapter 2 | Literature Review

Overview

In this chapter, the literature review about the main concepts will be presented.

The chapter is divided into three major sections: section 2.1 presents some main concepts about tourism marketing and marketing segmentation; section 2.2 describes the role of social media and social networks on tourism marketing and tourist's movement patterns; section 2.3 presents the methodology used to fulfill this study purposes.

2.1 | Marketing and Tourism

2.1.1 The appearance of tourism marketing

In the 90's, Ferrell and Lucas (1987) highlighted the importance of defining the tourism marketing concept, since more application oriented fields of marketing began to rise. Until there, tourism marketing was simply accepted as being the application of marketing in tourism management situations (Gilbert, 1989, Holloway and Plant, 1988 and Middleton 1988 *cf* Fyall and Garrod, 2005).

For Rita (1995 *cf* Coutinho, 2012), it consists of the "identification of market segments, influence the development of touristic products and delivers to the potential tourists information about those products". A few years later, Lumsdon (1997 *cf* Fyall and Garrod, 2005) defined tourism marketing as the "managerial process of anticipating and satisfying existing and potential visitors wants more effectively than competitive suppliers or destinations". Lumsdon explains that the exchange process – that is driven by profit - is largely dependent on the effective interaction between consumer and supplier and the satisfaction of the first one (see figure 1). In 2006, Kolb introduced his own concept. For the author, tourist marketing consists on the application of specific marketing concepts to plan business strategy in order to attract visitor to a destiny, which can be a city, a region or a country.

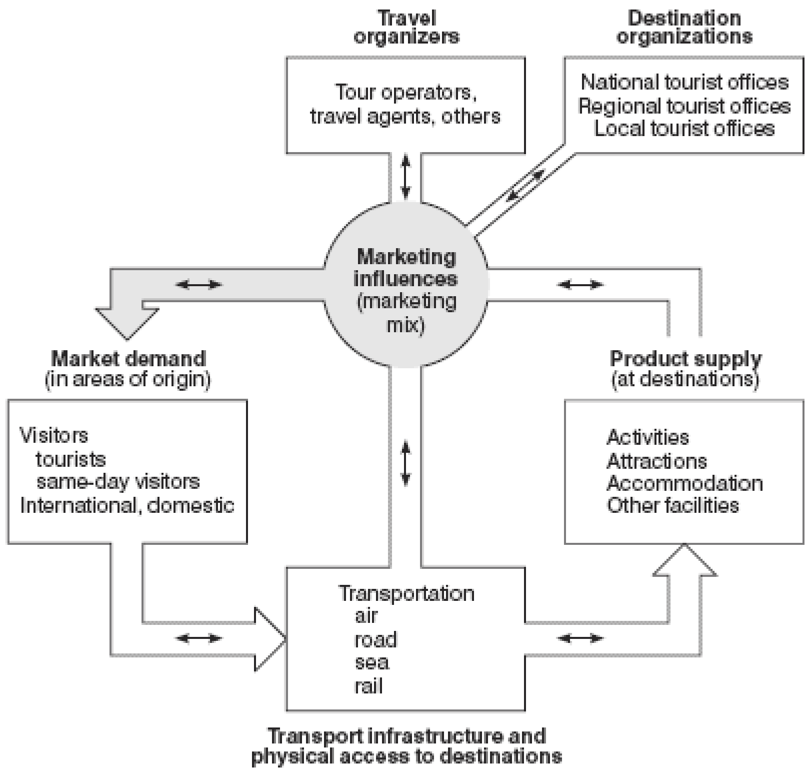


Figure 1: The link between customers and supply and the influence of marketing. Source: https://etravelweek.com/imported/meaning-marketing-travel-and-tourism.

Considering the initial marketing definition given by the American Marketing Association (AMA, 2013), tourism marketing is described as an exchange that satisfies the person (in this case the tourist) and the organization itself (in this case a service, a company or a city, for instance). The definition finally recognized that marketing is not restricted to products with prices and that not all products are exchanged for money and profit (For example, museums may be available to visitors for free or they have to pay admission charges). According to Rita (1995 *cf* Coutinho, 2012), marketing has an important management role between the offer and demand, since that consumer have the power to decide which touristic product prefer and they pay for it. Marketing tools allow the anticipation of customer needs and deliverance of value. The client – in this case the tourist – must be able to associate the tourist destination as a different, original and appealing product.

2.1.3 Tourism and tourist concept

The use of these terms is problematic and they have suffered some adaptations in the past years. Tourism is a multidisciplinary field, which means there is a range of fields studying it with different methods, which leaves us with a lack of clarity concerning the definition. Most of the terms found in the literature are reductive only to places or activities (Franklin, 2003:21; Page, 2014:10), mostly because the definitions are based on what it is not tourism. The researchers attempted to identify things that distinguish tourism from other activities such as non-ordinary places, non-ordinary activities on a daily routine and non-home environment (Franklin, 2003:28). The work of Franklin (2003) and Larsen, Urry and Axhausen (2006) give a complete overview of tourism concept evolution (See Annex 2). Internationally, The World Tourism Organization (UNWTO) defined tourism as the "activities of people travelling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes not related to the exercise of an activity remunerated from within the place visited" (Page, 2014:16).

Regarding urban tourism concept, it was even not quantifiable and ignored by researchers as a distinct field until the 80's (Edwards *et al., 2*008 *cf* Ashworth and Page, 2011; Stansfield, 1964; Ashworth and Page, 2011 *cf* Kádár and Guede, 2013) even though it is one of the most popular forms of tourism (McKercher *et al., 2*006). The lack of interest were mainly related to the lack of a simple definition for this complex phenomenon and a lack of a clear limitation about which activities can be included in this industry such complex, fragmented and heterogeneous (Selby, 2004:3; Edwards *et al., 2*008 *cf* Ashworth and Page, 2011).

Regarding the tourist concept, the use of this term is also problematic and it has suffered some adaptations in the past years. The World Tourism Organization (Law, 2002) defines tourist as "someone who moves away from home on a temporary basis for at least 24 hours, whether travelling in their own country or going to another one" , but MacCannel (1976) and Urry (1990), according to Larsen, Urry and Axhausen (2006), tourist is seen as a sightseer, consuming places by gazing and photographing. In Chadwick work (1994 *cf* Page, 2014) is highlighted the difference between tourist, visitor and traveler that it is based on their origin, the purpose of travel and time spent. In our case, the term tourists and travelers are going to be used as similar since international and domestic individuals were considered in our sample (travelers - tourists) as well as students and migrant: other travelers –not tourists (See annex 7).

Urban tourism or city tourism?

Urban tourism corresponds to all types of tourism that occur in urban areas, which are considered larger areas and "dependent on activities such as manufacturing a service and possibly large-scale mining (Law, 2002; Ashworth 2003 *cf* Ashworth and Page, 2011). However, Ashworth and Page (2011) do not agree that the adjective "urban" by itself is enough to define urban tourism activities, which highlights that the problem of a clear limitation was maintained until a few years ago. In fact, the inclusion of towns and cities into tourism definitions appeared in Law's work in 2002 and it was Pearce, in 2001 that introduced the term urban tourism in tourism lexicon.

Not all authors consider that urban tourism and city tourism are similar terms. The conflict is well described on the research of Pearce (2001). Some authors defend that the factors that characterize urban areas are the same that define cities so they can be treated as similar (Herbert and Thomas, 2013), but others argue that what should be underlined is the increase of the cities attractiveness, instead of urban tourism (Cazes, 1994). A similar problem occurs with urban and city tourists, when their existence is even questioned by Ashworth and Page (2011).

The term "city" presents numerous differences around the world and. In fact, until 2011, there was no harmonized definition of the concept "city" between European cities and the rest of the world. Despite most of the definitions in dictionaries are focused on the size, population density, economic relevance and number of facilities, the criteria that defines what it is considered a city differs from country to country (Dijkstra and Poelman, 2012). Nevertheless, cities are characterized by density and diversity whether it is of their functions, facilities, buildings, culture or people: this it is what distinguishes the urban from the rural and characterizes the urban way of life. For Kolb (2006), a city is viewed as a product for tourists' consumption and physical goods (like buildings, streets and monuments), services (for example hotels, restaurants and events) and an idea (created from the previews two factors) composes it. Those three factors are the ones that provide the tourist visiting experience. Even the core motivation to visit a certain city may be different, tourists will choose the one that has to offer the best experience. That experience is intangible, such as the services in the city (Kracauer, 1995 *cf* Kolb, 2006).

2.1.4 Market segmentation

The marketing mix model

First, marketing only started to be faced as a management discipline, instead of as an economic activity and sales support, in 1950 motivated by the customer business orientation (Drucker 1954; McKitterick 1957 *cf* Webster, 2005) which was the key to the appearance of models such as the marketing mix model and to the creation of processes such as market segmentation. The marketing management concept is well defined, by Kotler and Keller (2008), as “the art and science of choosing target markets and getting, keeping, and growing customers through creating, delivering, and communicating superior customer value” From this moment forward, marketing management perspective began to be faced as a “decision-making activity, directed at satisfying the customer at a profit by targeting a market and then making optional decisions on the marketing mix, or the "4 P's.” (McCarthy, 1960; Kotler, 1967 *cf* Vargo and Lusch, 2004).

McCarthy defined marketing mix as the combination of all the factors “at a marketing managers’ command to satisfy the target market”, giving origin the 4 P’s of product, price, promotion and place. Booms and Bitner (1980) added 3 P’s to the initial model (Participants, Physical evidence and Process) and they created the marketing mix concept for services– the 7’P. This is relevant since city services characteristics of intangibility, heterogeneity, inseparability and perishability highlight the increased role of consumers in the service process. In other words, the P of product is also the P of place which makes with the P of promotion an important factor since the tourist must be motivated to spend his time and money in the city Kolb, 2006).

From the marketing mix model to market segmentation

The marketing mix model gave birth to market segmentation process since marketing mix only make sense if companies concentrate their efforts on targeted markets and targeted markets are simply market segments after a market segmentation process. As so, in the 50’s, marketing segmentation started to be an accepted practice, as a means to improve marketing efficiency and effectiveness, essentially because it was absolutely consistent with the philosophy of customer orientation.

Companies started to base their activities on the needs and wants of customers in selected target markets (Grõnroos, 1989) and found out that moving away from mass marketing to a target marketing strategy was essential and focusing only on a particular group of customers. For that, markets had to be target and segments had to be found, in order to satisfy customers’ needs. The identification of specific customer groups to serve it is what it is called market segmentation, in which customers are aggregated into groups with similar requirements and buying characteristics.

For Smith (1956 *cf* Dolnicar, 2007), market segmentation consists of "viewing a heterogeneous market (one characterized by divergent demand) as a number of smaller homogeneous markets". Six decades later, Kotler and Keller (2011) defined it as a process of dividing the market "into well-defined slices. In other words, it is the act of dividing the market into groups of customers who share the same needs and interests". The segmentation process consists of three main elements: segmentation, targeting and positioning (Kotler, 1998 *cf* Dibb and Simkin, 1991) and it's based on four mains strategies: behaviour, demographic, psychographic and geographical ones, like it is represented on figure 2 (Larsen, 2010).

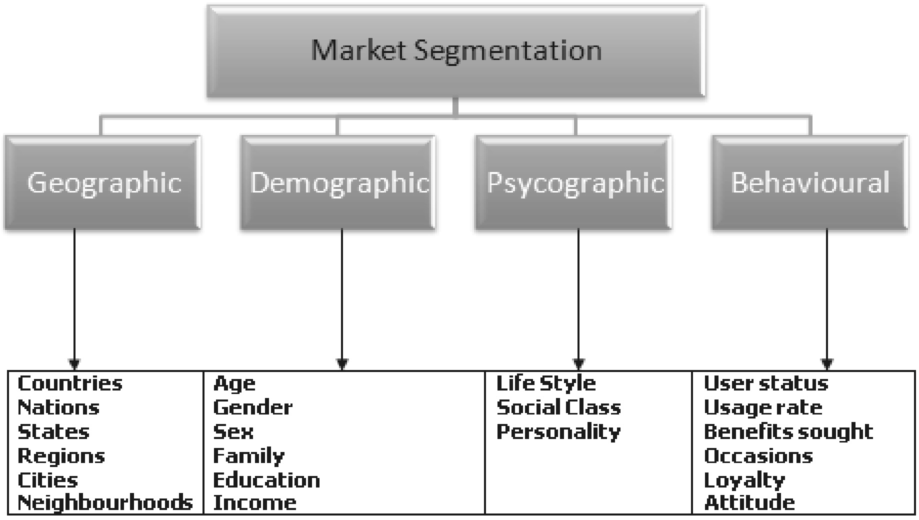


Figure 2. Market segmentation-based strategies. Source: http://www.thehomeofknowledge.com/the-importance-of-market-segmentation/

In other words, when segmenting a market, groups of individuals are developed which are similar with respect to some personal characteristic. The particular personal characteristic with respect to which similarity is explored is the segmentation criterion or segmentation base (Smith, 1956 *cf* Dolnicar, 2007). For instance, if market segmentation is supported by demographic criteria, individuals might be grouped based on age (young vs orders) or gender (man vs woman). Additional approaches are recognized in the literature (Mazanec, 2000; Wedel and Kamakura, 1998 *cf* Dolničar, 2004) like the priori segmentation one or "commonsense segmentation", which it is well recognized in tourism industry and consists on tourists groups identification based on individual's' prior knowledge. When the process is complete, companies must decide which, if any, they intend to enter. Segments should be distinct which means that members of one segment should be as similar as possible to each other and clearly different from the members of other segments, identifiable, reachable and suitable in size (Dolnicar, 2007).

In conclusion, market segmentation is the key element of modern marketing. In order take advantage of such powerful tool, companies have to be a better understanding of their targets audience in order to satisfy customers better that the competition (Dibb and Simkin, 1991; Larsen, 2010).

2.1.5 Tourism market segmentation

Tourists travelling patterns

Following Kotler and Levy (1969), business organizations recognized that customer needs and behaviour are not obvious without formal research and analysis. Tourists are different because they feel attracted by different tourist destinations, they search for different activities during their trips and they make use of different entertainment facilities (Dolnicar, 2007). Acknowledging that every tourist is unique and that tourism industry cannot possibly cater for each individual separately is the basis of market segmentation and such differences will be reflected on tourist's spatial behaviours (Weinsteins, 2004 *cf* Xia *et al.,* 2010).

Several definitions of tourists travelling patterns can be found in literature, but all related with spatial changes and sequences of movements between activities or locations (Lau and Mckercher 2007 *cf* Leung *et al.,*2012; Xia *et al.,*2010; Following Beeco and Hallo, (2014) dominant travel sequences are the patterns that are frequently used by tourists; According to Zheng, Zha and Chua (2012), tourist's behaviour patterns are sequential events of visiting different places or sites, their travel routes, and the amount of time spent at any location. According to Weinsteins (2004 *cf* Xia *et al.,* 2010) tourist behaviour is described as "a consumers' entire set of (specific psycho-physical predispositions) actions, activities and conduct, connected with making choices in given economic, social and demographic conditions.

Market segmentation process on tourism field

Following Weinsteins (2004 *cf* Xia *et al., 2*010), implementing on tourism field, market segmentation is seen as "a process of dividing a market into homogeneous subgroups: tourists in the same group are similar to each other and different from other groups, in the way they react to the market mix such as promotions or advertising which then influences theirs spatial behaviors' and it can be applied by any organizations in tourism industry: hotels, travel agencies, tourist attractions, restaurants, and local charities (Dolnicar, 2007). Considering the definitions of tourist's paths provided by Zheng, Zha and Chua (2012) and others, tourist market segmentation based on movement patterns of individuals refers to the dominant travel sequences of patterns used by tourists.

Traditionally, the demographic strategies are used with more frequency in tourism than others are, because most of the times the necessary information is easy to identify and measure (Gilaninia and Mohammadi, 2015). Tourism market segmentation was based on the identification of tourists groups from an origin perspective or based on a destination perspective and it was created different segments based on similar socio-demographic characteristics and travel modes like types of travel group, modes of transport or visit frequency (Xia *et al.,* 2010) or psychographic and travel behavioural variables (Wedel and Kamakura, 2000 *cf* Xia *et al.,* 2010). However, market segmentation strategies can be hybrid ones, which means that more than one strategy is used (Grönroos, 1989; Gunter and Furnham, 1992 and Kotler and Keller, 2009 *cf* Gilaninia and Mohammadi 2015). Moreover, the principles are the same: If tourists have different needs and desires, the promotional efforts must be targeted because the messages will appeal a certain group, but it might not appeal another one so different groups requires different promotional strategies (Beeco and Hallo, 2014).

The main benefit of market segmentations, in a tourism context lies on a tourist destination ability to specialize on a particular group needs and become the best in catering for that group. With that, destinations obtain competitive advantage marketing efforts because they can develop messages that are more effective for the segment targeted using the most effective communication channels and, consequently, will be able to communicate the message more effectively. The insights into travel behaviour have been served many purposes of destination management, product development and attraction marketing (Li, Meng, and Uysal, 2008), to redefine existing attractions and plan new ones more effectively (Lew and McKercher, 2006). It also can be useful to find bottlenecks between accommodation and attraction (Prideaux, 2000), identification of the time and space characteristics of the routes that tourists visit more frequently in order to avoid capacity overload and social, environmental and cultural impacts (Lew and McKercher, 2006). Tourist's patterns information can be very useful in terms of infrastructure management (Lew and Mckercher, 2006 *cf* Leung *et al., 2*012; Hayllar and Griffin, 2009 *cf* Edwards and Griffin, 2013) because tourists tend to be spatially concentrated rather than dispersed throughout a city (Asakura and Hato, 2004; Lew and Mckercher, 2006 *cf* Leung *et al., 2*012; Masiero and Zoltan, 2013); to measure the impact of tourist mobility on environment (Shoval al Isaacson, 2009 *cf* Versichele *et al., 2*014); for destination planning, environmental and cultural policies management (Lew and Mckercher, 2006 *cf* Leung *et al., 2*012); For urban planners, traffic engineers and tourism authorities (Wolf, 2004 *cf* Girardin *et al., 2*007); and to the planning and programming of recreation facilities and services (Manning, 2011 *cf* Beeco and Hallo, 2014).

For the marketing area in general, the study can be useful to design tourist packages, providing more attractive combinations of attractions, develop travel guidance policies, marketing services (Asakura and Iryo, 2007; Xia *et al.,* 2009 *cf* Leung *et al.,* 2012; Vu *et al.,* 2015) to perfect marketing strategies (Shoval and Isaacson, 2007) and marketing incentives (Shoval and Isaacson, 2009 *cf* Versichele *et al., 2*014). Photographs have been used in tourism advertising (Zimbardo, 1992 *cf* Li, Huang and Christianson, 2016) for instance and it has been a tool used by tourism marketers in different platforms to promote destinations and provide a virtual experience of a destination to tourists. Studies show that photographs can affect, for example, purchase decision-making (Underwood & Klein, 2002 *cf* Lo *et al.,* 2001). Finally, the information about which type of tourist visit which attractions can be used to fulfill market segmentation and to design more appropriate tour packages (Xia *et al., 2*010).

2.2 | Social Media and User Generated Content

2.2.1 Social Media

Social media, as one of the most powerful online networking tools, has been integrated into social and economic life in the real world. Social media includes social networking sites, consumer review sites, content community sites, wikis, Internet forums and location-based social media (Zeng and Gerritsen, 2014). While social media are believed to be the tools or means of communication, social networking is considered the use of "social media tools" to interact and communicate directly with people you are already connected to or with whom you wish to be connected with (Wells, 2011 *cf* Zeng and Gerritsen, 2014).

The evolution of the World Wide Web allowed the appearance of interaction and user-generated content on the online world, assuming various forms. According to Fotis, Buhalis and Rossides (2012), six types of social media have been identified in the literature: social networking websites (Facebook or Linkedin), blogs, content communities (like YouTube, Flickr), collaborative projects (Wikipedia), virtual social worlds and virtual game worlds. Other authors' work presented on Munar and Jacobsen (2014) research - such as Baym (2010) and Munar and Jacobsen, (2013) - describe Flickr as a media-sharing site like Youtube, for instance. Besides media-sharing sites, the authors classified the online revolution into "Wikis (like. Wikitravel), blogs (Travelblog), microblogs (Twitter), social network sites (e.g. Facebook, Linkedin) and review sites (TripAdvisor)".

Social media are increasingly relevant as part of Tourism practices, affecting destinations and businesses (Munar *et al.*, 2014) and play an important role as an information source. The Internet has reshaped the way the information is distributed and the way travelers plan for and consume travel (Buhalis and Law, 2008) and it is also an important platform for information exchange between the consumer and the destination marketing organizations (Werthner and Klein, 1999 *cf* Xiang *et al., 2*010).

2.2.2 Flickr and *geotagged* photos

As said before, the increase of mobile and capture devices - like smartphones, digital cameras and tablets - are feeding the user-generated content phenomenon in the past few years. These devices are built with global positioning systems such as GPS technology, which enable geographical information (Vu *et al., 2*015). In other words it is possible to know where and when people have been, which allows a better understanding about tourist's mobility travel (Kádár and Guede, 2013) and allowed the increase of tracking travel data collection methods in the past decade (Wolf, 2004 *cf* Girardin *et al., 2*007; Asakura and Irvo, 2007; Vu *et al., 2*015) because people share their travel experience like photos and videos on social media web services (Memon *et al., 2*015).

For the purpose of this study, we will be focusing our attention on the information that tourists share in their social profiles about their tourism experiment. , Flickr.com is the

Social Network used to collect tourist’s trip data. Flickr is a popular photo-sharing platform (Spyrou and Mylonas, 2014) and one of the most popular online resources for people to manage and share their travel experiment. Furthermore, Flickr is a rich data source for mining tourist travel patterns (Zheng, Zha, and Chua, 2012; Lee, Cai, and Lee, 2013 *cf* Vu *et al.,* 2015). Flickr becomes popular for being the largest collection of community collected *geotagged* photos and for offering a public API that allows anyone to access these photos and associated data, which is one of the main reasons why Flickr was chosen for this study. In 2014 there were more than 14 billion photos uploaded on Flickr and stores more than 200 million of *geotagged* images (Spyrou and Mylonas, 2014). As mentioned on chapter 1, *geotagged* photos are available for public view on Flickr, but not in a direct way: it is necessary to use Flickr's Application Programming Interface (API), a public Flickr.com tool that allows the interaction with Flickr accounts and allows the access of data associated to the photos (Shoval and Isaacson 2006).

In the last few years, individuals are contributing with *geotagged*ged photos while sharing their travel experiments. Consequently, millions of *geotagged*ged photos became available online with relevant information about people such as location, time, tags, title and weather (Zheng *et al.,* 2012; Memon *et al.,* 2015). Photos and travels are linked intrinsically. Photographs document the travel experience (Haldrup and Larsen, 2003; Larsen, 2008; Urry, 1990 *cf* Lo *et al.,* 2011) and provide an opportunity for tourists to share experiment with others (Groves & Timothy, 2001 *cf* Lo *et al.,* 2011).

For clarifying, tag or tagging consists on the addition of descriptive keywords to photos, creating metadata when individuals decide to share content (Golder and Hubberman, 2006 *cf* Nov and Ye, 2010; Spyrou and Mylonas, 2014). *Tag* is used to annotate several types of content like images, videos, bookmarks and blogs in Social Networks such as Flickr. The increase of tagging is explained by the benefits individuals gain from posting such information (Ames and Naaman, 2007; Cattuto, Loreto and Pietronero 2007 *cf* No and Ye, 2010). *Geotagging* is “the process of adding geographical identification information to several media such as *geotagged* digital photos”. Most of the time consists of latitude and longitude coordinates, entered manually by the individual or comes from a GPS receiver in the camera or in the smart phone and it can be important to characterize human behaviour in a social network environment (Crandall and Snavely, 2011). It can be referring the actual location of the place that it is being photographed or the location of the individual that is taken the photo.

2.2.3 The User Generated Content phenomenon (UGC)

The increase of mobile and capture devices, like smart phones, digital cameras and tablets, are feeding the growth of information and the user-generated content that has been created in the past few years. In other words, among the important factors that contribute for the massive grow of data, is data created by the users, in many different ways. The work developed by Luo and Zhong (2014) and previews studies developed by Gretzel (2006) Pan, MacLaurin, and Crotts (2007), Xiang and Gretzel (2010) and Green, 2007 (*cf* Leung *et al.*, 2012) revealed that the appearance of Social Network sites fostered the customer-generated content through the sharing and posting – in real – time – information about personal experiment, feelings, review and opinions by travellers. Green (2007), Chung and Buhalis (2008) O'Connor (2010) and Stringam and Gerdes (2010) supported the work developed by Leung *et al.* (2012), showing that the Internet and the new technologies fed the growth of user-generated content in tourism because travellers do not only post experiment, photos and videos, but also search opinions and reviews of other travellers in order to make decisions. Following Xiang *et al.* (2010) travellers are posting and sharing their travel-related comments, opinions, and

2.2.4 From the UGC to Big Data

The definition of Big Data are still fairly recent and it is causing a debate between the academy and the industry (De Mauro, Greco and Grimaldi, 2014) therefore, a single definition explaining this new term doesn't exist yet (Mayer-Schönberger and Cukier, 2013). This ambiguity came from the fact that because Big Data evolved so quickly and in a disorderly way, such a universally accepted formal statement denoting its meaning does not exist and because the definitions have been used to impart all kinds of concepts, turn it into a vague and amorphous concept (Moorthy *et al., 2*015).

The first person that introduced the term was Francis Diebold in 2003, by saying: "Recently much good science, whether physical, biological, or social, has been forced to confront—and has often benefited from—the Big Data phenomenon , but the key definition comes in Meta Group Report (now Gartner 2001) and the term is defined using the 3 V's - Volume, Velocity and Variety - despite the words 'Big Data having never been mentioned in the report (Ward and Barker, 2013 and Moorthy *et al., 2*015).

It was possible to identify four key areas of big data definitions: Information, Technologies, Methods and Impact (See Annex 3), expressed on De Mauro, Greco and Grimaldi (2015) research. One of the fundamental reasons why the Big Data phenomenon exists is the current extent to which information can be generated and made available (De Mauro, Greco and Grimaldi, 2015). Information is, therefore, a frequently treated topic by and among authors. As so, the nucleus of the concept is expressed by the 3 V's (see table 1): Volume, Variety and Velocity (De Mauro, Greco and Grimaldi, 2015), which are their common characteristics. Variability and complexity are considered two other features, especially by those focusing on analytic (Fang *et al.*, 2015). Snijders, Matzat and Reips (2012) for instance defends that "Big Data includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time". This means that, besides the volume, Big Data also deals with data sets, containing various formats and coming from different sources. As well as Snijders, Matzat and Reips (2012), Intel IT Centre (2012), Schroeck *et al.* (2012), Zikopoulos *et al.* (2014) and Djicks (2012), extended the definition using the 4 or the 5 V'.

V	Concept
Velocity	"The speed at which data is created, processed and analyzed" (Schroeck et al., 2012).
Variety	"Different types of data and data sources (Schroeck et al., 2012) which we collect and it can be structured, unstructured or text data, numerical, images, videos, etc (Liu, 2005).
Veracity	"Data uncertainty; the level of reliability associated with certain types of data" (Schroeck et al., 2012).

Table 1. 3 V's of Big Data definition. Source: Own elaboration

The impacts on tourist market segmentation process

The analysis of photography data related to the geographical position taken by tourists is an effective method to study tourist movement patterns in urban spaces (Kádár and Guede, 2013) because the photos with this kind of information become the digital footprints of photo takers and an implicit documentary about their spatio-temporal movements (Zheng *et al.*,2012). With reference to the market segmentation process, the Internet does not only provides huge advances in database marketing and innovative distribution approaches, but also expand the ability of implement market segmentation in a more effectively way and expand the portfolio of segmentation methods available (Dahan and Srinivasan, 2000 *cf* Wind and Bell, 2007).

Big Data affect all the sectors and industries, according to Schoenberger and Cukier (2013). No matter what sector we are talking about, the amount of data available worldwide is growing faster than ever and areas like marketing could benefit from it (George, Haas and Pentland, 2013; Akerkar, 2012; Joachimsthaler; 2013; Bucur, 2015). Lahiri and Biswas (2015) support the application and the advantages that big data can bring related with customer segmentation, product's segments, targeted campaigns, the development of personalized offers, marketing based on consumer's actual behaviours and preferences, new marketing opportunities, improved marketing strategies and optimized brand strategies as main benefits for instance; Schroeck *et al.* (2012), supports the development of recommendations and suggestions about purchases, based on the interests, past behaviour and other customers preferences and Salvador and Ikeda (2014) and Davenport (2013) refer to the Big Data's potential with respect to the marketing Mix variables, with a focus on communication and price decisions and the use of information in the marketing research (consumer understanding, consumer's attitude, interests, preferences and purchase behaviour). Finally, Regarding the tourism field, big data can provide insights that help deliver a more intelligent travel experience by analyzing structured and unstructured data (Variety factor) and as well as help to make travel more focused on travelers' preferences and needs (Schroeck *et al.* 2012; Davenport, 2013).

2.3 Sequence Mining

Sequence Mining tool is very useful to discover a range of patterns shared among objects which have between them a specific order. The knowledge of spatial-temporal individuals' movements across a destination and the personal information about them can bring important implications for marketing areas like the development of advertising campaigns given a certain target and the development of recommendation systems, given certain behaviour patterns (Zaki, 2001).

Formally speaking, let = $\{i_1, i_2... i_n\}$ be a set of items and e an event such that $e \subseteq I$. A sequence is an ordered list of events $e_1, e_2 . . . e_m$ where each $e_i \subseteq I$. Given two sequences $\alpha = a_1, a_2 . . . a_k$ and $\beta = b_1, b_2 . . . b_t$, sequence α is called a subsequence of β if there exists integers $1 \leq j_1 < j_2 < . . . < j_l \leq t$ such that $a1 \subseteq b_{j1}$, $a2 \subseteq b_{j2}$, . . . , $a_k \subseteq b_{jl}$. A sequence database is a set of tuples $< s_{id}, \alpha >$ where s_{id} is the sequence identification and α is a sequence. The count of a sequence α in a database of sequences D, denoted count (α, D), is the number of examples in D that contain the α subsequence. The support of a sequence α is the ratio between count (α, D) and the number of sequences in D. We denote support of a sequence as support (α, D). Given a sequence database D and a minimum support value λ, the problem of sequence mining is to find all subsequences in D having a support value equal or higher than the user-defined value, the λ value. Each one of the obtained sequences is also known as a frequent sequence or a sequential pattern (Peng and Liao, 2009).

There are several algorithms that can be used to find frequent sequences. In this work, SPADE algorithm (Zaki, 2001) was used to discover tourists' travelling patterns across the Porto city. This algorithm uses a vertical layout format where each sequence in the lattice is associated with the *id list*. This id list is a list of all examples containing the candidate sequence; the list contains a set of pairs where each pair consists of both the sequence id (sid) and an event identifier (typically the event time). To search for frequent patterns, SPADE also uses a candidate-generation strategy. To compute the support of each candidate pattern of level l, the id lists of sequences from level l − 1 are joined using a temporal join. The support of each candidate pattern is the number of distinct sid in the candidate id list. Like others, SPADE uses the *apriori* property to prune the search space. Five experiments were run based on percentage of frequent movement parameter, represented by the variable support in the algorithm. If we increase the support level to x, the frequency of the movements returned will increase too and the number of movements identified will decrease (See figure 3). The algorithm results returned the list of the most frequent isolated items and the list of the most frequent set of items that occur in events (Zaki, 2001). The algorithm constructed using R software is demonstrated on Annex 8.

```
> library(arules)
Loading required package: Matrix

Attaching package: 'arules'

The following objects are masked from 'package:base':

    abbreviate, write

> library(arulesSequences)
> library(Matrix)
> x <- read_baskets(con = system.file("misc", "SM_att.txt", package = "arulesSequences"), info = c("sequenceID","eventID","SIZE"))
> inspect(x)
     items sequenceID eventID SIZE
1    {71}      1          6465    1
2    {496}     1          6475    1
3    {377}     1          6485    1
4    {497}     1          6489    1
```

```
> s1 <- cspade(x, parameter = list(support = 0.4), control = list(verbose = TRUE))

parameter specification:
support : 0.4
maxsize :  10
maxlen  :  10

algorithmic control:
bfstype  : FALSE
verbose  :  TRUE
summary  : FALSE
tidLists : FALSE

preprocessing ... 1 partition(s), 0.06 MB [1.3s]
mining transactions ... 0 MB [0.82s]
reading sequences ... [0.14s]

total elapsed time: 2.25s
> inspect(s1)
   items     support
 1 <{298}> 0.4471545

> similarity(s1)
1 x 1 sparse Matrix of class "dsCMatrix"

[1,] 1
> summary(s1)
set of 1 sequences with

most frequent items:
    298 (Other)
      1       0

most frequent elements:
  {298} (Other)
      1       0



element (sequence) size distribution:
sizes
1
1

sequence length distribution:
lengths
1
1

summary of quality measures:
    support
 Min.   :0.4472
 1st Qu.:0.4472
 Median :0.4472
 Mean   :0.4472
 3rd Qu.:0.4472
 Max.   :0.4472

includes transaction ID lists: FALSE

mining info:
 data ntransactions nsequences support
    x          3655        246     0.4
> |
```

Figure 3. Frequent movements patterns, using SPADE algorithm through R programming, for a frequency level of 0,4. Source: Own elaboration.

2.4. Networks and Social Network Analysis

According to Kolaczyk and Csárdi (2014) and Dias (2015), the first reference to a network concept appears on Euler's work (1741), to the well-known Königsberg bridge problem, in which Euler proved that it was impossible to walk the seven bridges in such a way as to traverse each only once. The term graph is a product of Sylvester's research (1878) despite the first publication of this topic having been by Denes Konig (Tutte, 2001). In this work, Social Network Analysis allows us to understand the relations between places visited by tourists during their trips to Porto and it will helps us to understand tourists travelling patterns.

A Social Network represents entities and their relations as nodes and links, which form a network (see figure 2). It can be defined as a method used to map and measure relationships and flows between people, groups, organizations, and other connected information/knowledge entities. Networks can be represented using graphs in mathematical terms. Considering a graph G = (V, E) where V represents the set of vertices or nodes and E the sets of edges or links where elements of E are unordered pairs u, v of distinct vertices $u, v \in V$. In other words, E represents any kind of relation between a pair of vertices. Therefore, a graph G consists of a non-empty set V of vertices and a set E of edges. Associated to an edge may be a weight that represents the strength of the connection between two vertices. The number of vertices $N_v = |V|$ and the number of edges $Ne = |E|$ are sometimes called the order and size of the graph G, where the order of G is the total number of vertices n in V and is represented as $|V(G)| = n$; the size of G is the total number of edges m in E and is denoted as $|E(G)| = m$ (Newman, 2010; Oliveira and Gama, 2012). In our work, the nodes represent the sites and the edges are the links between the sites, created by the tourists' movements.

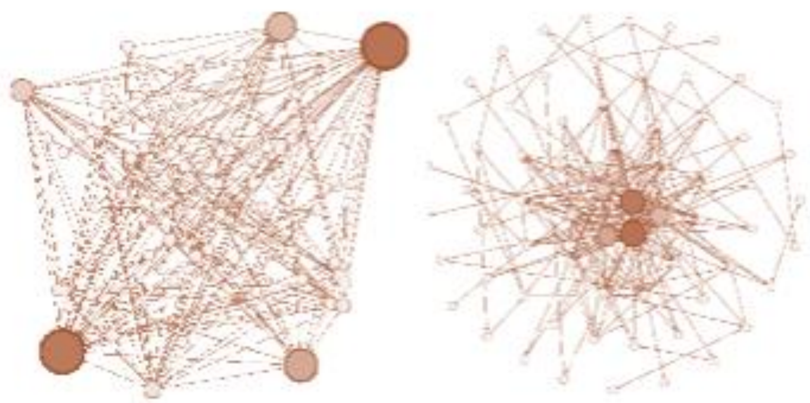


Figure 4: Network basic representation and Network representation using Fruchterman Reingold layout. Source: Own elaboration.

The network extracted from Gephi is composed of all the sets of paths made by the considered tourists chronologically organized before. Finally, centrality measures were used to quantify the importance and the influence of the vertex in the network

The Degree Centrality metric counts the number of edges that are connected to a vertex. It is reasonable to suppose that places (represented by nodes) that have a high number of connections might have more influence than those who have fewer connections (Newman, 2010 *cf* Dias, 2015). The edges may either be directed or undirected, associated with an edge may be a weight that represents the strength of the connection between two vertices.

The Weighted Degree metric has the same purpose as the first one, but considers edges' weight instead of measuring the vertex strength. The distribution of strength—sometimes called the weighted degree distribution—is defined in analogy to the ordinary degree distribution (Kolaczyk and Csárdi, 2014).

A common notion of distance between vertices on a graph is defined as the length of the shortest paths between the vertices. This distance is often referred to as geodesic distance, a similar name for shortest paths (Kolaczyk and Csárdi, 2014). The Distance Metric represents the average graph-distance between all pairs of nodes. If the nodes are connected the graph distance is equal to 1.

The diameter is the longest graph distance between any two nodes in the network (Kolaczyk and Csárdi, 2014). In other words, it shows how far apart the two most distant nodes are.

The Betweenness Centrality measures the extent to which a vertex lies on paths between other vertices (Oliveira and Gama, 2012 *cf* Dias, 2015) or the extent to which a vertex is located "between" other pairs of vertices. According to Kolaczyk and Csárdi (2014), "These centralities are based upon the perspective that importance relates to where a vertex is located with respect to the paths in the network graph". In other words, it indicates how often a node appears on shortest paths between nodes in the network. Vertices with high betweenness centrality may have considerable influence in the network (Newman, 2010 *cf* Dias, 2015) because they tend to control the flow of information between communities (Oliveira and Gama, 2012 *cf* Dias, 2015).The higher this parameter, the more influential the node is. However, the nodes which have high betweenness centrality are not necessarily the ones that have the most connections.

The Eigenvector centrality is a more complex metric than the degree centrality and "it assumes that not all neighbors of a vertex have the same importance. So, what is taking into account is not the quantity of neighbors, but the quality of those neighbours. This metric is based on the assignment of a relative score to each vertex and measures how well a given actor is connected to other well-connected actors". On a simple way, it is a measure of node importance in a network based o a node's connections so it can be the best metric that defines is what best defines the importance of a place in the network (Dias, 2015).

Finally, the Eccentricity Centrality measure indicates the extent to which a vertex lies on paths between other vertices (Oliveira and Gama, 2012 *cf* Dias, 2015). Simplifying, it calculates the distance from a given starting nod to the farthest node from in the network.

Social Networks and tourism marketing

According to Newman (2010), many issues from the real world can be represented as networks, including people and places. The real applications of graph theory and social networks in the Tourism field are strongly advocated by Leung *et al.* (2012), using the research developed by Scott (1991), Wasserman and Faust (1994) and Novelli *et al.* (2006). According to Kolaczyk and Csárdi (2014), networks have been applied in various study fields, such as computational biology, health, finance and even marketing.

Social Network Analysis (SNA) has been introduced into tourism research and it has numerous applications (Novelli *et al.*,2006); It can be a useful approach used to understand knowledge network, network clustering, research subject evolution (Scott, 1991) and to expand the understanding of the structure and constitution of Tourism (Hu and Racherla, 2008). To individual's behaviour analysis, SNA has the same function as GIS software allowing a visualization of trip diary data. However, the focus here is not the geographic movement, but rather the attractions themselves and the relationships in and among them.

2.5. Cluster Categorical data Analysis with K-Modes algorithm

Following the research of Huang (1998), "clustering is a popular approach to implementing the partitioning operation. Clustering methods partition a set of objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters (Anderberg, 1973; Jain and Dubes, 1988; Kaufman and Rousseeuw, 1990)". Following the authors, the method of "partitioning a set of objects in databases into homogeneous groups or clusters (Cormack, 1971; IBM, 1996)" can have important applications for data classification, aggregation and segmentation, serving marketing purposes. To accomplish this paper purpose, a Cluster Analysis using K-mode algorithm was applied in order to fulfill market segmentation.

There is a massive quantity of data that is categorical in the real world. For instance, variables such as gender and profession are usually defined as categorical data. Each categorical attribute is represented by a set of unique categorical values. In this study, for instance, gender attribute will be defined as {Female, Male}. Categorical values are discrete and non-ordered unlike numerical ones (Ng *et al.*, 2006).

The K-means algorithm is the most used for clustering data because of its efficiency in clustering large quantities of data, quoting. However, the algorithm only works on numeric data, which means the variables are measured on a ratio scale (Jain and Dubes, 1988). Therefore, cluster analysis for numeric data cannot use the same algorithm as the cluster analysis for categorical data. Ralambondrainy (1995) passed through the limitation presenting an algorithm extension called K-mode. His approach consists of converting multiple category attributes into binary attributes, in which 0 represents the absent and 1 the present for each category. With the binary attributes, k-means algorithm is able to read the data as a numeric one (Huang, 1998).

A range of analysis will be performed considering different attributes such as gender, origin and social state. To fulfill cluster analysis, R software will be used through K-modes {KlaR} package. Figure 5 demonstrates the results provided by through K-modes using R software for an illustration of 2 clusters.

The clustering can have important applications for data classification, aggregation and segmentation, serving marketing purposes, as it was applied in this study.

K-modes clustering with 2 clusters of sizes 168, 77

Within cluster simple-matching distance by cluster:

[1] 632 407

```
Cluster modes:
   ID Gender Country Social.State Age Profession    10   100   110    12  120   130   140   160   170   180   190    20
1   1   Male     n/a          n/a n/a        n/a FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
2   6   Male     n/a          n/a n/a        n/a FALSE FALSE FALSE FALSE TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE
    200    30   40    49    50    60    70    80     9    90
1 FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
2 FALSE FALSE TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE


Clustering vector:
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31 
  1   1   1   1   1   2   2   1   1   1   1   2   1   1   1   1   1   1   1   1   1   2   2   1   1   1   2   1   1   1   1 
 32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62 
  2   1   1   2   1   2   1   1   1   1   1   1   1   2   1   2   2   2   1   2   2   1   1   2   2   2   2   2   2   2   1 
 63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93 
  1   1   2   1   2   1   1   1   1   1   1   1   1   1   1   1   1   1   1   2   1   2   1   1   1   2   1   2   1   1   1 
 94  95  96  97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 
  1   1   1   2   1   2   2   1   1   1   1   2   1   1   2   1   1   1   2   2   2   1   1   1   1   1   2   1   2   2   1 
125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 
  1   1   1   2   2   2   2   1   1   2   1   1   1   2   1   2   1   1   1   1   1   1   2   1   2   1   1   2   1   1   1 
156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 
  2   1   2   2   1   1   1   1   1   1   1   2   1   1   2   1   1   1   1   2   1   2   1   1   1   1   2   2   2   2   1 
187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 
  2   2   1   1   2   1   1   1   2   1   1   2   1   2   1   1   1   1   1   1   1   1   2   1   1   2   1   1   1   1   2 
218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 
  1   1   1   1   1   1   1   1   2   2   2   1   2   1   1   1   1   1   2   2   2   2   1   1   1   1 
```

Figure 5. Cluster Analysis results for two clusters based on individual’s gender. Source: Own elaboration

Chapter 3 | Methodology application

This chapter presents the application of the Methodology described in chapter 2, section three, in order to study tourists' travelling patterns and fulfill market segmentation.

The chapter is divided into six steps: Step one presents the method used to collect data from Flickr.com; Steps two and three portraits how data was cleaned and organized using Excel tools; Step four explains how Sequence Mining analysis will be applied using R software; In step five it will be explained how Social Network Analysis is implemented through Gephi software;. Finally, Cluster Analysis will be approached in step six.

As presented on chapter 1, a quantitative methodology using data from secondary sources will be applied, with the help of a Social Network Analysis, a Sequence Mining Analysis and a Cluster Analysis. Photos have been collected from *Flickr* over a period of 29 months (from January 2014 to May 2016) and this study takes place in the city of Porto, Portugal. Based on the time and location of the photos collected, along with personal information such as origin, gender, it was possible to create a graph about tourists' activity and flows in space and time. The study aims to answer to the following research questions:

- Where do tourists go? What are their trips and paths, through space and time?
- How do networks, between different locations, help us define tourists' behaviour?
- How tourists' travelling patterns information, along with personal data, can be used to fulfill market segmentation?

The results will be presented on chapter 5.

The case of Porto

Porto has recently become one of the most searched destinations in Europe (Carvalho, 2013). Its origins date back to pre-Roman times, where the name *Portus Cale* appeared and, later, the name "Portugal". It is a city well-known as the *unconquered city* for people's ability to fight off the invaders of many attacks in the past. Porto is world famous for the Port wine, its historical centre - classified as World Heritage by UNESCO for its rich cultural and architectural heritage - and also by Futebol Clube do Porto, the main football club of the city (Coutinho, 2012).

The city has a lot to offer historically, culturally and architecturally, and thus has great potential for tourist attraction: "With ninety-five classified monuments (thirty-nine of these open to the public), integrated into various cultural and tourist routes and nineteen national monuments. The cultural level and leisure offer more than thirty museums open to the public, twenty-eight churches, fourteen Cellars of Port twelve markets and fairs and more than a dozen annual events with international appeal. The city has more than 60 works of contemporary architecture and many tourist sites that extend from Ribeira to Foz, Boavista and Baixa" (PENT, 2007; CCDRN, 2008; DT, 2011 *cf* Coutinho, 2012). According to Moreira (2010 *cf* Coutinho, 2012) the city's main attractions are the Port Wine (28%), the heritage (24%), the monuments and museums (22%), the cruises through the Douro River (11%), the cultural activities (10%) and others (5%). An inquiry made in 2011 by Visit Porto to tourists that bought Porto Card showed that almost 80% of them used the card to visit monuments, almost 70% to visit museums and 58% on tourist circuits (See Annex 10 and Annex 11).

Since the European Championship in 2004, there has been a positive evolution on Portuguese tourism development until current days (with exception of 2009 when the economic crises imploded and 2001 when the terrorist attack occurred) as a result of strong communication campaigns (Dias, 2015). Consequently, the number of guest in 2013 rose up in 14% comparatively to 2012 statistics (Agência Lusa, 2014) and the occupation rate had never been so high before in Portugal - from 41,1% in 2006 to 46,1% in 2015 (Aguiar, 2016). The average room income also presented significant growth, reaching the 37 Euros in 2015 (comparatively to 28 Euros in 2006). The number of sleepovers followed the same tendency with an average of 11% growth every year.

Globally speaking, in 2015 in Europe the number of sleepovers reached over 50 million and required mostly by individuals from UK, Germany and Spain. In Porto, the tourist season reached new records (Lusa, 2015). The amount of rooms booked increased significantly more in the Porto region, represented by approximately 5.4 million overnights in 2014 (OJE/Lusa, 2015). In fact, the number of sleepovers increased by more than 1 million between 2009 and 2014 and , consequently, the hospitality incomes (Correia, 2016).

Besides the football event, one of the major contributors to the growth and consolidation of Porto's brand internationally was the European Best Destination award that the city won in 2012 (Carvalho, 2013). According to the Porto City Hall statements, tourism in the city is passing through a turning point and it has been capturing public attention from people with higher purchasing power. In 2014, Porto won the same award as a result of relevant communication efforts and the city's brand consolidation. More recently, in 2017, Porto won the Best European Destination Award once again against major European cities such as Milan, Paris and Amsterdam. Porto is the only Portuguese city with such recognition and most part of the votes actually came from foreign countries.

The second award was seen as touristic offer booster which may had important consequences on the city's economic development, job creation and investment attraction (Agência Lusa, 2014a). The city hall itself received some recognition too when, for instance, Ribeira was considered one of the most beautiful streets in the world by some of the most important tourism European magazines and journals. In 2015, the city was elected as the "Best Emergence Destination" by the European Consumers Choice and the third best destination in the world by TripAdvisor (Correia, 2016).

We met in a private conversation with the Porto Offical Tourism department [1]and they provided us with the most recent data relating to the tourism flows in the city. The data are related to all tourists that went to the official-post of tourism to search for information. According to the provided data since the beginning of 2015, there has been a significant growth in information search by tourists on the official Tourism posts – from around 7000 individuals in January 2015 to more than 30 000 individuals in December 2015. In terms of country origin, 31% of the tourists that went to the post-office came from Spain and 27% from France (See Annex 12). According to Visit Porto (2015), what distinguish Porto from other destinations is not its dimensions (second biggest city in Portugal), but its human capital, the people, the knowledge and its ability to create surprise and emotion

[1] Meeting with Maria do Carmo Costa of the Porto Official Tourism Board, that belongs to the Porto City Hall, at March, 2016.

into whoever visits the city and to make them want to return.

Step 1: Data collection

As mentioned before, the data used in this database was collected from the photos available on Flickr.com, a popular photo-sharing platform where people can share photos with the option of tagging them geographically.

In 2014 there were over 14 billion photos uploaded in Flick and more than 200 million *geotagged* images stored. Related to the city of Porto, since the beginning of 2014 until May 2016, there are more than 600.000 photos with the Tag “Porto”, but only around 20% contained *geotagged* information. Of that 20%, only 55% are actually related to the city of Porto, Portugal. After data cleaning, the database only contained 8201 photos from 253 different users (See Annex 4). Given the extense list, the number of attractions are not available in annex 4 but the information is available for consulting if necessary.

Data was extracted from Flickr.com via the Flickr API (Application Program Interface). The API allows any persons to have access to these photos along with their textual *metadata*. The query used in this API encloses some parameters necessary to fulfill this study aims (See Annex 5).

The query returns the first 250 *geotagged* photos taken and uploaded since 1st January of 2014 within the tag “Porto”, given in the parameter tag. As Kádár and Guéde (2013) indicate, the maximum amount of photos in one response is 250, but it can be changed for a maximum of 500 per page. Given the limitation of 2500 photos per response (maximum of 5 pages), the algorithm was divided and ran several times, in order to overcome such limitation. Basically, the parameters *min_upload_date* and *max_upload_date* were changed in order that the API response to not overcome the 2500 results per result. The attribute *extra* was used with the value “geo” in order to simply download photos with geo information attached, as well as the information about when the photo was taken, by introducing the value “data-taken”. The attribute *accuracy* record the accuracy level of the location information and allowed to choose the precision level of the photos (in other words, photos can be related to a city view taken from a plane – lower accuracy - or related to a specific rock in the middle of the street – higher accuracy). In this study, a very precise level -between 14 and 16- will be used, which implies that data are being automatically generated by cameras, mobile devices GPS and others. In this case our photos are only the ones related to a city levels in order to avoid pictures to priced (like a picture of a rock) or to ambiguous (like pictures took from a plain).

Finally, as it was also in our interest to acquire additional and more personal information about the users a second query was used (See Annex 6). The results from this query provided insights about individual's gender, country origin, social state and profession.

Step 2: Data cleansing and filtering

Initial considerations

Three initial assumptions had to be considered to perform this study, regarding to:

- The definition of tourist;
- Trip duration;
- Information disclosure and privacy issues;

Firstly, in this study, the differences between tourists, travelers and visitors were not taken into account (see to Chadwick, 1994 in Page, 2014), so it was not necessary to distinguish these concepts them as Girardin *et al*,(2007) had done in their work (See Annex 7).

Secondly, the trip duration was calculated based on the attribute *datetaken*: two followed photos of a certain owner are considered part of a certain trip if the *datetaken* value between the pictures, do not differ more than 8 days. However, it is not possible to assume that such data portrays exactly the real individual's trip duration simply because there is nothing that can assure us that the first photo and the last photo uploaded in the network corresponds to the moment or day that tourist arrived and left the city.

Finally, we had to consider the amount of information disclosed. Sharing personal data in Flickr.com is not mandatory, so not all users provided it. From all the 253 users present in our database only 20% shared complete information about themselves. In some cases, a country was manually assigned due to spelling errors or wrong names. Additionally, the data collected only came from public publications and the real names from users were omitted, regarding privacy issues. For all the fields where information was not obtained, the value NA was assigned.

In the end, a chronologically ordered set of photos was created taking into consideration the attribute *datetaken* and the photo owner. Considering our initial purposes, we proceed to the elimination of:

- The users that only had one uploaded photo;

- The elimination of all the users and associated photos which the different between the first and the last *data-taken* value was less than 1 day;

Step 3: Data processing

Some additional fields had to be created and inserted so the data could be used for the purposes of this study.

Regarding the *owner* attribute, we create an additional one called *owner_id*. To every different owner will be signed a number N={1, …, n}. This information normalization will be necessary for the Networking Analysis.

A similar approach was used with the *title* attribute. We created an additional attribute named *title_id* with the purpose of normalizing the pictures' titles for a better and simpler analysis. For instance if title owner x = "Douro" and title owner y = "Douro River" the software will be able to recognized that both results are the *title_id* attribute. In this case, the result would be "Rio Douro". For all photos whose *title_id* was inexistent or for those photos not related to Porto attractions (for instance personal photos, photos of food, animals, etc), we applied the value *others* in *title_id* results. Therefore, these cases classified as *others*, represent 13,7% of the whole database. Figure 6 is an example of the final database.

Owner	Owner_id	Title	Title_id	Datetaken	Latitude	Longitude
12049844@N03	6	Douro at night	Douro River	2014-03-04 22:12:21	41.14007	-8.60953
12049844@N03	6	Sao bento trains	São Bento station	2014-03- 22:14:40	41.14048	-8.61319
12049844@N03	6	Praza da Ribeira	Ribeira	2014-03-04 22:30:25	41.14044	-8.61298

Owner	Owner_id	Username	Gender	Country	Social State	Profession
12049844@N03	6	Monica	Female	France	Married	Teacher

Figure 6. Representation of the final database, after data cleansing and processing steps. Source: Own elaboration.

For the initial concerns, a range of metrics were run using Excel tools in order to obtain some first insights on the data collected, such as:

- Number of tourists and photos considered;
- Number of photos taken by region/place/attraction;
- Number of tourists that photographed a certain place/region/attraction.
- General tourists' characterization, considering origin, gender, profession and age.

Step 4: Finding frequent movement patterns

SPADE algorithm was used to discover to discover tourists' travelling patterns across the Porto city (Peng and Liao, 2009) through R programming, as mentioned before (See Annex 8). Following the authors, a sequence corresponds to an ordered list of events and each even is a non-empty unordered set of items. In this case, the events are represented in each line of our database and correspond to each place visited by a certain tourist. The events are composed by a range of attributes: the first attribute is the *Sequence id* and represents which tourists have done each trip (from 1… N); the second one is the *Event id*, which is the moment in time when the visit has occurred. For that, all *datetime* values had to be transformed into number format by using excel functions; Then the place related to the *datetime* considered. For a certain *sequence id* it is created n lines and each line represents a place visited on a certain *datetime.*

For the purposes of this study, we performed five different experiments based on the frequency level attribute, represented by the variable *support* in our algorithm. The variable support allow us to define the level of frequency of the tourists travelling patterns. We considered a support level equals to 0,4 (which represents 40% or more of frequency), 0,2 (20% or higher), 0,1, 0,05 and 0,03. In other words, it was possible to identify the most frequent movement patterns of the tourists, for a frequency level of 40% or higher, 20% or higher, etc. The results can be found and are analyzed in the next chapter. Given the extend of the results of some experiences, not all of them are available in the appendix but the information is available for consult, if necessary.

Step 5: Constructing the network

In order to data be accepted by Gephi software, data had to be organized according five different attributes: Source, Target, Date start, Date end and Label attribute. The *Source* column represents the place visited by a tourist on a specific day and time during his trip to Porto and the *Target* column represents the place where the tourist chronologically went next. The new database created for Gephi software include all sets of paths made by travelers, chronologically organized before. The *date start* and *date end* fields represent the data time that tourists arrive and leave a certain place. Finally, the *Label* attribute indicates which tourist(s) has realized certain paths (see figure 7). The logical applied to Gephi is quite similar to the one used in Sequence Mining, but with different parameters. In this case, the program will return graphically the same results provided by the Sequence Mining analysis. However, Sequence Mining approache is focused on the number of frequent paths and Network Analysis provide us the connections between the different places, considering through the ins and outs of a place, considering those paths.

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions

Id	Label
0	Place A
1	Place B
2	Place C
3	Place D
4	Place E
5	Place F

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions

Source	Target	Type	Id	Label
0	1	Directed	0	Tourist x
2	3	Directed	6	Tourist x
2	0	Directed	5	Tourist x
1	2	Directed	3	Tourist x
3	0	Directed	7	Tourist x
0	5	Directed	8	Tourist x
5	4	Directed	9	Tourist x
4	3	Directed	10	Tourist x
3	5	Directed	11	Tourist x
5	1	Directed	12	Tourist x
1	0	Directed	13	Tourist x
0	2	Directed	14	Tourist x

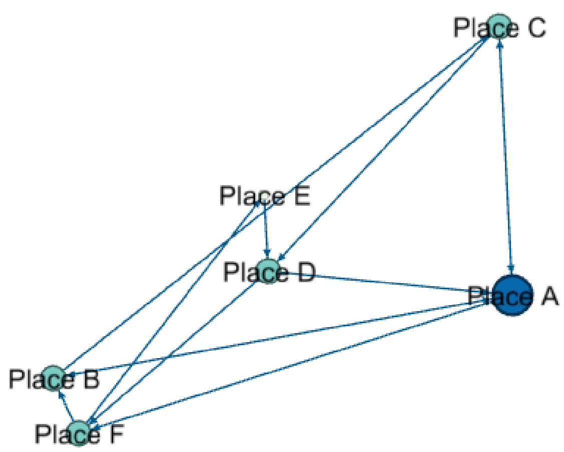


Figure 7. Example of a social network generation on Gephi programme. Source: Own elaboration.

The moment the data are introduced in the software, the Social Network is generated automatically. However, the timeline option had to been created previously, by executing the merge of the columns *date start* and *date end* on the Data Laboratory separator. Finally, using the Gephi tools, a range of statistical metrics were obtained, as mentioned before. The results were then analyzed, interpreted and discussed in the next chapter.

Step 6: Fulfilling market segmentation

Knowing our tourists, their paths and the connections between the places visited, the last question that lacks for an answer is: how can we find market segments based on tourist's movement patters information. In order to get an answer for this question, a cluster analysis was performed. Once again, R software will be used through *klaR* R package. The algorithm includes not only the individual's movement patterns, but also their personal information collected with additional query, as explained before (See Annex 9). Similar to what it is exposed on step 5, we performed a range of until we find the best number of clusters, considering a segmentation based-strategy (already approached on chapter 2). Market segmentation only makes sense if it has a purpose. In our case, our purpose is to find similarity, in terms of behaviour, among our database and make that information useful for marketing purposes, assuming that individuals that present similar paths have similar motivations so they can be grouped on a segment and organizations can apply to them the same marketing strategy. The number of clusters, in our algorithm, is represented by the variable *clx <- kmodes(na.omit (dataG), x)*. We aim to compare the results of the algorithm with the work of Grönroos (1989), Flogenfeldt (1999, *cf* Leung *et al., 2*012), Lue *et al.* (1992 *cf* Leung *et al., 2*012) and Lau and Mckercher (2006). The results and some marketing recommendations will be presented in the next chapter. Similar to the sequence mining analysis, the results were too long to be exposed on the appendix but they are available for consult, if necessary.

Chapter 4 | Results presentation

This chapter presents the results of the empirical study, in order to answer the previews research questions.

Section 4.1 contains an overview about the Porto city tourism evolution; Section 4.2 present a database characterization in terms of user's description and places visited/photographed.

In section 4.3 the most frequent movement patterns of tourist's in Porto city are presented.

Then, section 4.4 will describe the Social Network Analysis in order to understand how the relation between places visited can help us to understand tourists' behaviour.

Finally, section 4.5, it will portrait tourist's segments findings.

4.1 Database characterization

For the purpose of this study, 252 different users and 8201 photos taken by those tourists from January 2014 to May 2016 were used. On average each tourist took 33 photos.

As mentioned before, the personal information about each Flickr user is manually posted by individuals on their social network profile, which means users are not obligated to share his personal information. As a result, we only obtained complete information from 22, 5% of the users considered in the study. Almost 52% of the users have been members on Flickr for more than 5 years and only 6% became members in the past year.

Concerning the gender attribute, almost 4% of the users hid the information: 81, 3% of the users are men and the remaining ones (14, 7%) are women.

In terms of origin, it is possible to observe that almost 62% of the tourists came from Europe and almost 13% from American countries. The remain one comes from Australia (1%) and Asia (2%) Almost 22% of the tourists hide their country of origin information (Figure 8).

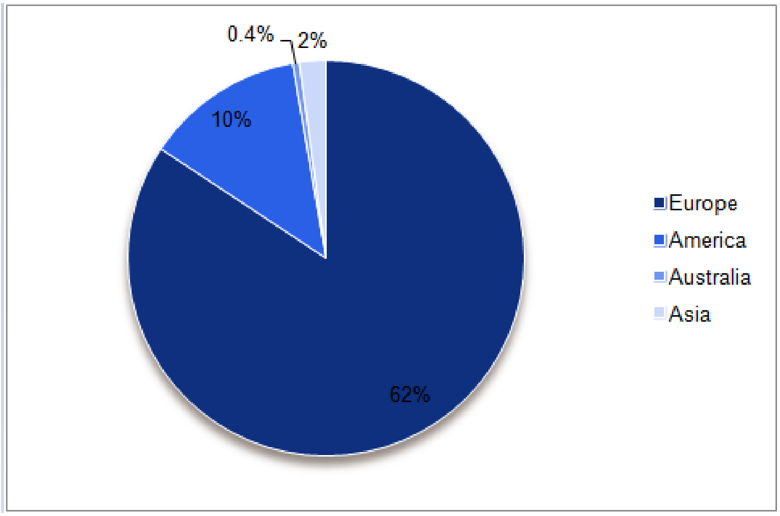


Figure 8: Tourists of the database used, between January 2014 and May 2016 (%). Source: Own elaboration.

Most of the European tourists (62% of the sample) came from Spain (21,2%), France (17,9%) and U.K. (10,3%), as it is illustrated in figure 9. Regarding the American countries (10% of the sample), most individuals came from the U.S.A (37,5%), Brazil (29,2%) and Canada (25,7%).

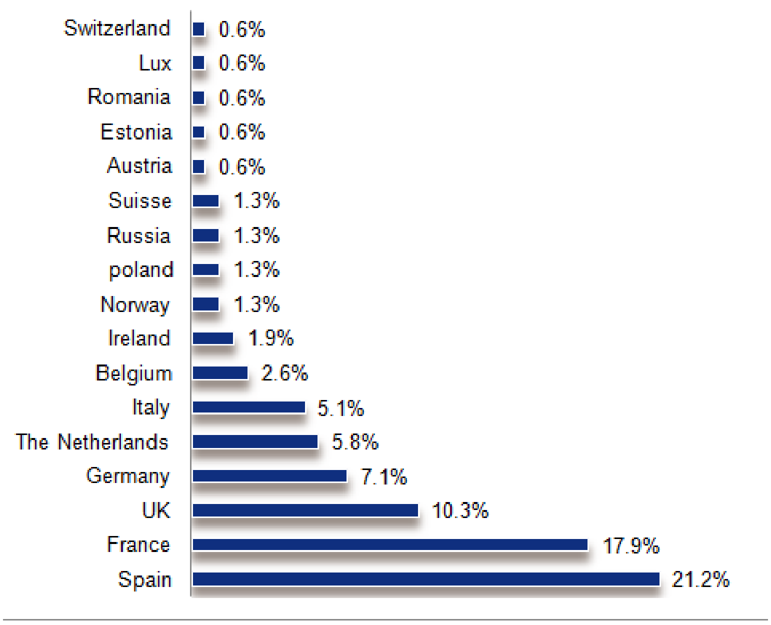


Figure 9. European tourists' origin by country, between January 2014 and May 2016, of the database used in this study. Source: Own elaboration.

With respect to individuals' social state, around 56,4% chose not to share such information, around 29% is married and almost 14,8% is single.

In the attribute "first visit or not", around 92% of the tourists said it was their first time visiting Porto. Based on our data, there has been a clear decrease, over the past three years, in trips lasting less than 4 days, as opposed to the increase in trips lasting between 4 and 8 days during the same period (Table 2).

Days/ % of tourists	2014	2015	2016
1 to 3 days	79,45%	71,11%	67,86%
4 to 8 days	15,07%	17,04%	21,43%
8+ days	5,48%	11,85%	10,71%

Table 2. Tourist's trips duration between January 2014 and May 2016. Source: Own Elaboration

Regarding the places visited and photographed, it was possible to observe that there are nine specific attractions - which represent almost 25% of the photos database – that tourists most photographed and they are represented in figure 7. It is clear the preference for the Luís I bridge, Douro River and Ribeira, as well was the city historical centre – Aliados. With respect to the most visited sites, we did the same analysis and the results are illustrated in figure 8. It is possible to conclude that Luís I bridge, Douro River Porto, Estação de São Bento, Ribeira and Clérigos Tower are the most photographed and visited attractions. However, Bolhão, Livaria Lello & Irmão are the most visited places per tourists, but they are not the ones that individuals photographed more, for instance (see figure 10 and figure 11). For the effects of this study, we are considering the approach of Girardin *et al*. (2007) that defended that a photo is a prove of the individual physical present in one place. If we continue to explore we notice that only 30 places represent 50% of the photos taken in our sample (See Annex 13). The data related to the differences among the places visited and the places photographed is quite large (because it is exposed for each attraction) but they are available to be consulted if necessary.

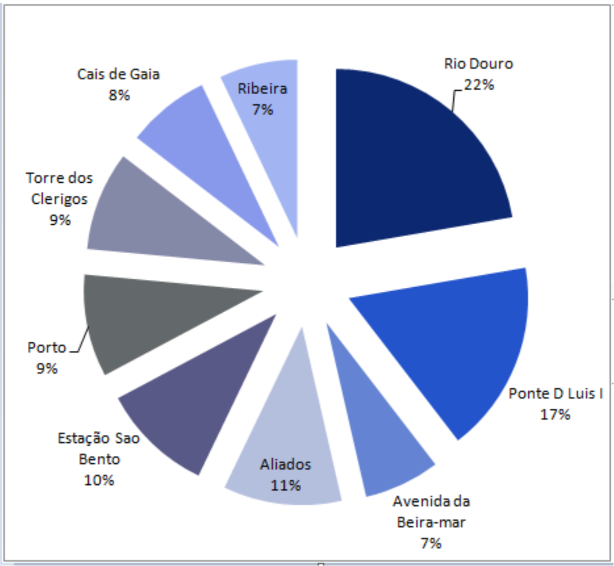


Figure 10. Most photographed attractions by tourists between January 2014 and May 2016 (%). Source: Own elaboration.

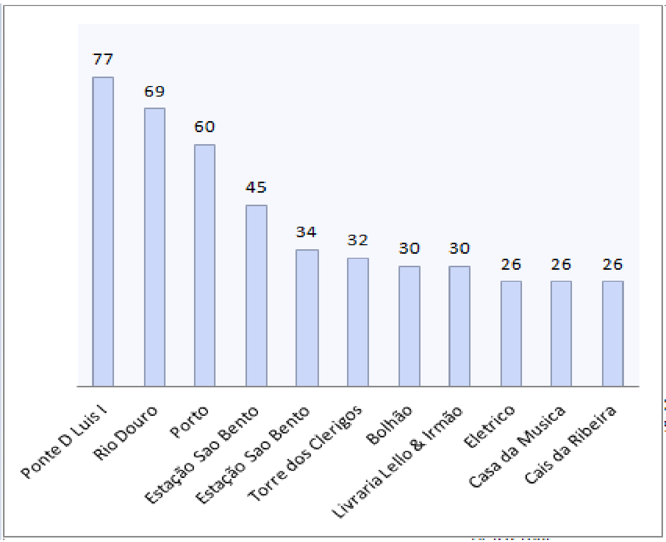


Figure 11. Most visited attractions (nº of tourists) between January 2014 and May 2016 (%). Source: Own elaboration.

If we continue to explore we notice that only 30 places represent 50% of the photos taken in our sample (See Annex 13).

Spots such as Clérigos Tower, Luís I bridge, Miradouro da Luís I bridge, Miradouro da Vitória, Cais de Gaia and Serra do Pilar are used to photograph other spots of the town, which it is perfectly reasonable considering the fantastic view that such attractions/places have to offer. Old Tram touris, Funicular dos Guindais trips, Teleférico de Gaia, Metro tours and Yellowbus trips are not used only because tourists are tired to walk (since most part of the trips are performed by walking and this facilities helps the individuals to return to the starting point), but also because tourist's enjoy and take advantage to capture amazing pictures all across the town. Such behaviour explains the differences found between where do tourists physically are and what tourists are photographing in that moment.

The results also are in line with the research of Urry (1990 *cf* Kádár & Guede, 2013), since the authors defend that cultural tourism monuments, viewpoints, and specific events are the primary objects of consumption. As we can observe, 13,2% of the photos taken are related with churches and other religious monuments (Porto city is characterized by its cultural religious and architectural aspects) and 5,2% are related to the ocean and beaches from Foz to Matosinhos and Gaia. As mentioned by the PMTD until very recently Portugal was promoted to the rest of the world as a land of sun and sea so it is natural still find a tendency and a search for the seaside. Museums and art photos related represent 2,5% of the pictures, with a special focus on the Baroque style which it is possible to find in Estação de São Bento. The PMTD call our attention to the fact that, in Porto, religious tourism is not recognized as a kind of tourism, but instead a search for different art styles that are reflected in the monuments such as churches and train stations. Porto wine photos related represent almost 2% of the sample as well as photos taken in gardens. Porto wine is one of the city *ex-libris.*

4.2 Where do tourists go? What are their trips and paths, through space and time?

With the Sequence Mining analysis, we were able to identify the most frequent trips fulfill by tourists between January 2014 and May 2016. As explained previously, five experiments will be perform based on the movement frequency rate. In other words and give the example of our first experience, a frequency level of 0,4 implicates that the algorithm will return movement patterns which frequency is equal or higher than 40%, considering all tourist's paths.

The first two experiment were not successful - for a frequency = 0.4 and 0.2, which means, we tried to find frequent paths with a level of frequency equal or hgher than 40% and 20% of frequency. The reason is that it was not possible to identify any frequent path with such levels of frequency, only frequent attractions visited (Luís I bridge with 44,7% of frequency and Douro River with 29%).

For a level of 10% (0,1) of frequency, we extract the first most frequent patterns and frequent places in our sample. It is clear that Aliados (17,1%), Estação de São Bento (18,7), Ribeira square (19,1 Clérigos Tower (19,1%) are the most frequent places/attractions, followed by Igreja dos Clérigos (12,6%), Lello & Irmão Bookstore (12,2%), Bolhão market (12,1%) and Casa da Música (10,1%).At this level of frequency, the algorithm revealed a few frequent paths, but they keep being related directly to movements near riverside

The results also revealed that the river and the attractions around it - Luís I bridge, Ribeira, Douro River and Cais de Gaia - are the core/centre of the most tourists paths, with exception of the path São Bento station to Ponte D. Luis with 10.2% of frequency.

Figure 12: Most frequent places and most frequent movement patterns to a frequency level higher than 10%. Source: Own elaboration.

As it is our intent to discover significant tourists' paths across the city, we went farther and we decrease the level of frequency to 5%. The results are expressed in Annex 14. Some places and attractions, besides the ones close to Douro River, started to appear. A significant part of the places revealed are related to religious monuments (churches most part), followed by photos taken to Palaces and photos related to Porto wine activities. Finally, pictures taken to the ocean and pictures taken through strategic spots such as bridges and Metro tours present significant frequency (See Annex 14). With respect to the most frequent movements, other attractions started to appear corresponding to the most frequent places, but still reflecting the concentration around the river, as it is possible to observe on table 3.

Path	Frequency (%)
Clérigos Tower → Douro River → Luís I bridge	6,5
Estação de São Bento → Ribeira	6,1
Estação de São Bento→ Douro River → Ponte D Luis	5,7
Cais da Ribeira → Cais de Gaia	5,3

Table 3: Frequent movement patterns, for a frequency level between 0,05 and 0,10. Souce: Own elaboration

The last experience with sequence miner performed aims to demonstrate the less frequent movement patterns, given our interest to discover with movement patterns tourist's do besides the ones related to the most popular attractions. As mentioned before, tourists' tend to disperse through the city guided by their own motivations and desires. As so, to a path frequency level equals or higher than 2% and less than 5% we run the algorithm and we found the results presented on Annex 15 (See image 13).

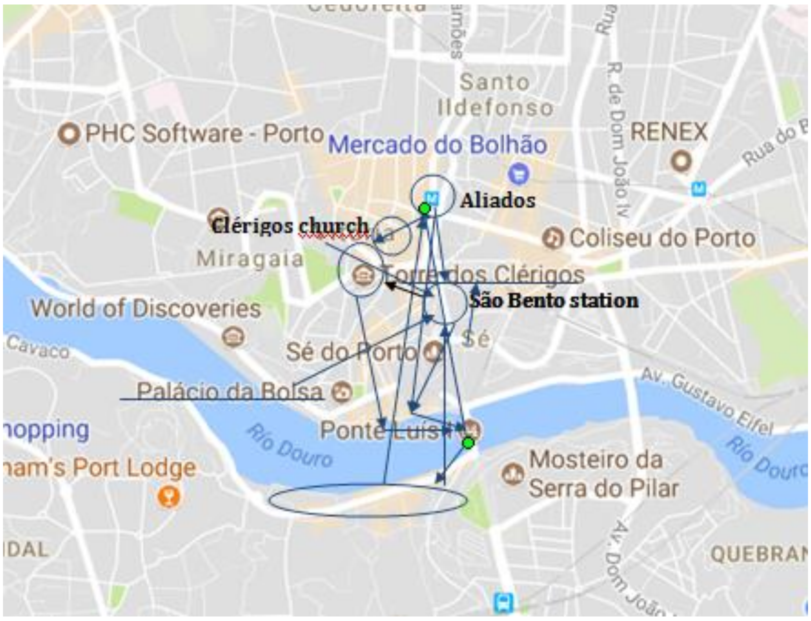


Figure 13: Most frequent places and most frequent movement patterns to a frequency level between 2% and 5%. Source: Own elaboration.

From the insights provided by this experience, we can observe that at this level there is a clear search for restaurants and art in the city (4,9%). At the top we can also find the Campanha station and the Infant D. Henrique statue as well as the majestic Serra do Pilar just across the river. Right after, some main streets – downtown (4,5%), Mouzinho da Silveira (4,1%) and Boavista (4,1%) followed by some churches and chapels such as Igreja de S Lourenço (4,5%) and Capela das Almas (4,1%). Related to Porto, Caves Calém (4,1%) and several other photos taken in Porto wine caves (3,7%) were found. The main searched beach by tourists is the one located in Foz with 4,1% of frequency. The results provided by this experience are in line with the database characterization insights. At this level, we can observe some movements between attractions that do not include the places around the river and they are expressed on table 4.

Path	**Frequency (%)**
Aliados→Estação de São Bento	4,5
Porto caves →Aliados	4,5
Estação de São Bento→Clérigos Tower	4,5
Aliados → Clérigos Tower	3,7
Estação de São Bento→Igreja dos Clérigos	3,7
Estação de São Bento ←→Praça da Liberdade	3,7
Aliados→Igreja dos Clérigos	3,7
Palácio da Justiça→Beach	3,7
Praça da Liberdade→Douro River→Luís I bridge	3,7
Clérigos Tower→Luís I bridge→Estação de São Bento	3,7
Luís I bridge→Cais de Gaia→Estação de São Bento	3,7
Palácio da Bolsa→Estação de São Bento	3,3
Igreja das Carmelitas→Estação de São Bento	3,3
Lello & Irmão Bookstore→Estação de São Bento	3,3
Terreiro da Sé→Estação de São Bento	3,3
Clérigos Tower→Cais da Ribeira→Douro River	3,3
Clérigos Tower→Terreiro da Sé→Douro River	3,3
Luís I bridge→Aliados→Estação de São Bento	3,3

Table 4. Example of some frequent movement patterns, for a frequency level between 0,03 and 0,05. Source: Own elaboration.

Results discussion

The analysis revealed significant insights about where do tourists go, and what tourists most looked for in the city. The first insights provided by this experience revealed that individuals' paths are mainly concentrated around the river (frequency level above 10%), like it is defended in Leung *et al.*, *2*012 and Masier and Zoltan (2013) research- However, other attractions - such as Clérigos Tower and Estação de São Bento – are considered a movement accelerator, which means these places present major influence in tourists' travelling patterns.

Some paths are almost mandatory given the proximity of the places, for example, Ribeira, Cais da Ribeira and Douro River: the places are so close that it is hard not to appear in a certain path when the tourist takes photos in one of those places. The same happens with Douro River and Luís I bridge and with Aliados and Liberdade Square. The tendency was observed in the work of Ashworth and Turnbridge, 1990 (*cf* Kádár, 2014; Asakura and Irvo, 2007) arguing that most part of the attractions are very close and mainly concentrated in the city historical centre, which brings serious issues in tourist tracking process.

Given the insights of the database characterization and the results provided by the sequence mining approach, we can find only very few similarities (for a level of frequency very low) with the developed by Dolnicar (2007), related to the fact that tourists are different because they feel attracted by different tourist destinations, they search for different activities during their trips and they make use of different facilities. In fact we found more than 500 places visited/photographed but even for a level of frequency of 5%, the algorithm was not able to find travelling patterns.

The results also demonstrate a downstream pattern since most trips start in highest city spots and most of them end in the lowest city spot. After that, individuals tend to cross over the river or redirect their trips in direction to Foz. Some of them decide to go back by using facilities such as Elétrico and Metro, as mentioned before. Such conclusion can be proved through the differences of frequency given a connection (of both ways) between point A and B: the frequency level from the point A to point B is higher than the frequency of the same path that starts at the point B and goes to A, when A is located on a higher spot in the city or farther away from the riverside (see table 5). Such fact was also presented in the meeting with the Porto Tourism Department[2].

Path (A to B)	Frequency (A→B)	Frequency (B→A)
Douro River → Cais de Gaia	9,8	8,9
Aliados → Luís I bridge	9,8	8,2
Clérigos Tower → Luís I bridge	8,1	5,3
Estação de São Bento→ Douro River	7,7	6,9
Estação de São Bento→ Ponte D. Luis	10,2	7,7
Estação de São Bento → Ribeira	6,1	5,3
Clérigos Tower→Ribeira	4,9	3,7

Table 5: Differences on the frequency level between two places in both directions. Source: Own elaboration.

Considering the work of Edwards *et al.* (2009 *cf* Vu *et al.*, 2015) and Li *et al.* (2008), and considering our city, the results can be useful to avoid capacity overload, social and environmental impact – in this case between Aliados and Douro river – as it is defended by Lew and McKercher (2006). For marketing areas, the design tourist packages, the design of more attractive combinations of attractions, the improvement of marketing services and the maximization of promotion and communication efforts can take advantages of the information about tourist's movement since we are aware of where do tourists go and its concentration (or not) around certain areas/attractions, as it was approached by Asakura and Iryo (2007), Xia *et al.* (2009 *cf* Leung *et al.*, 2012) and Vu *et al.* (2015).

In order to understand with more accuracy the connection between the places photographed and, considering the findings of individuals' movement across the city, a Social Network Analysis was performed and the results are presented in the next section.

[2] Meeting with Maria do Carmo Costa of the Porto Official Tourism Board that belongs to the Porto City Hall at March 2016.

4.3 How do networks between different locations help us to understand tourists' behaviour?

The Social Network Analysis, with the help of Centrality metrics analysis, aim to answer to one of our research question "How do networks, between different locations, help us define tourists' behaviour?". The Network will helps to identify connections and patterns between the different locations visited and help us to understand the existing connections between the attractions. Figures 13, 14 and 15 are the representation of the Social Network constructed based in this study database.

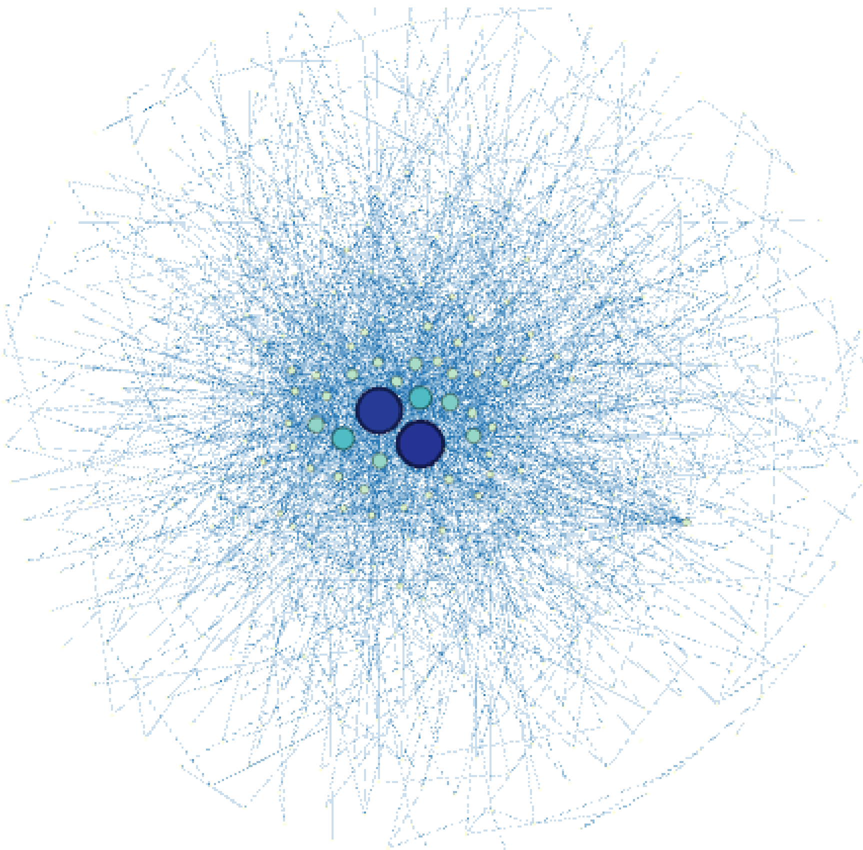


Figure 14: Social network based on the attractions photographed/visited, using the database collected from Flickr since January, 2014 to May, 2016 (Furchterman layout). Source: Own elaboration.

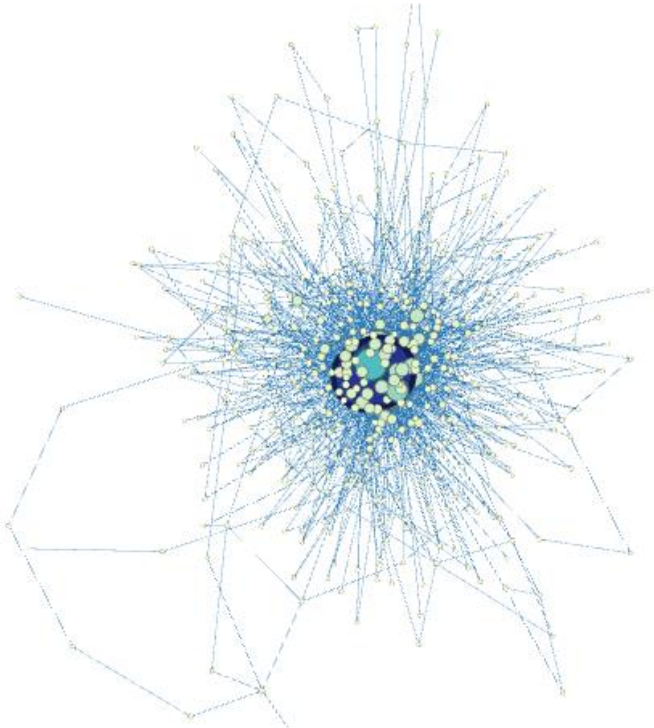
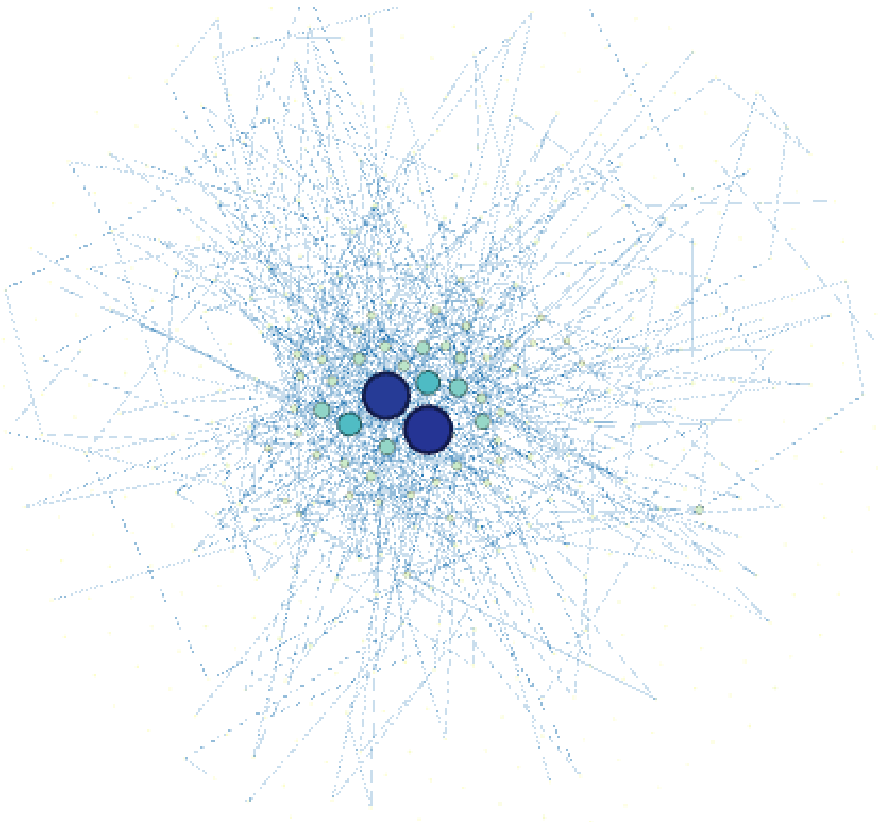


Figure 15: Social network based on the attractions photographed/visited, using the database collected from Flickr since January, 2014 to May, 2016 (Yifan Hu Proportional layout). Source: Own elaboration.



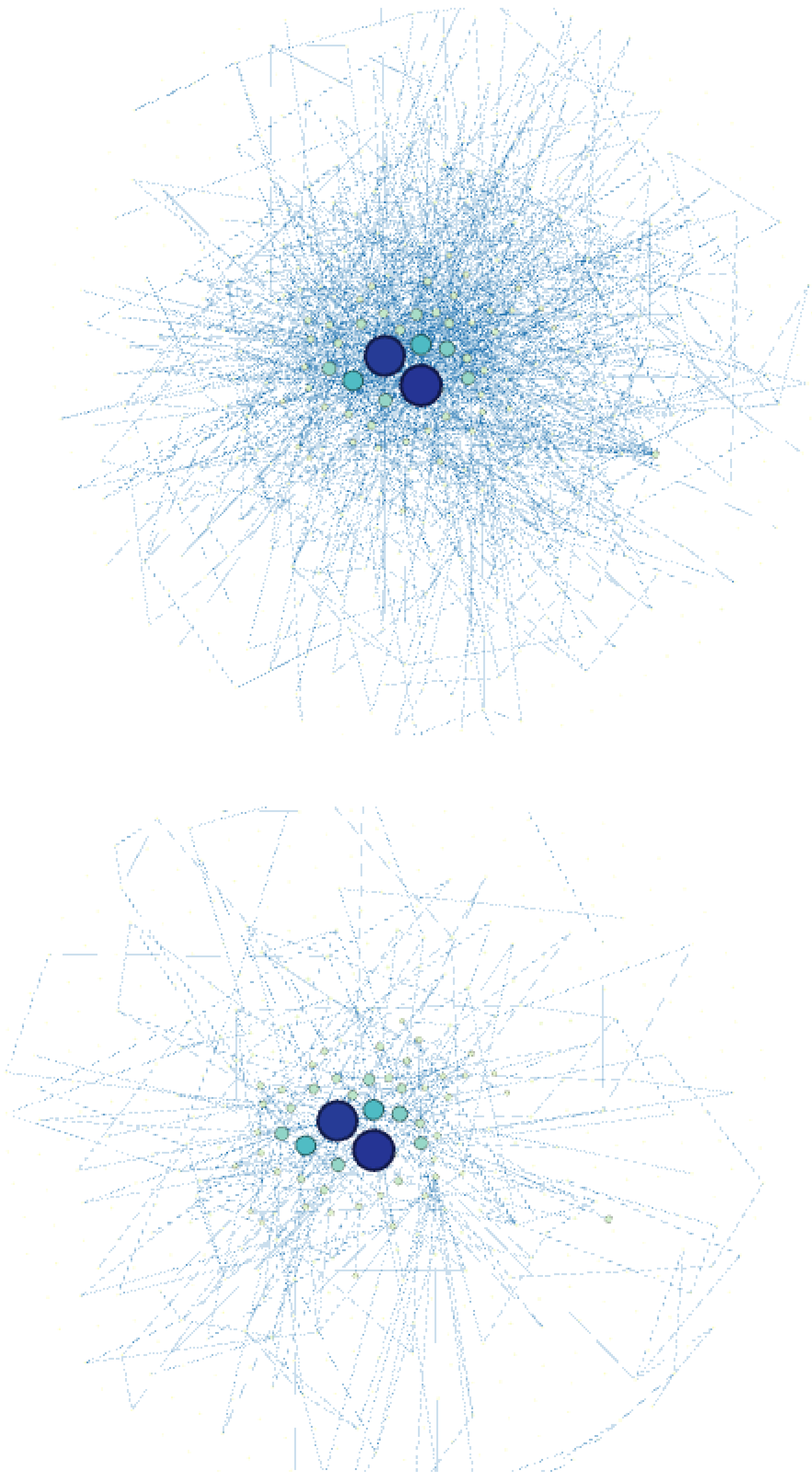


Figure 16: Social network based on the attractions photographed/visited, using the database collected from Flickr for 2014, 2015 and 2016 respectively (Furchterman layout). Source: Own elaboration.

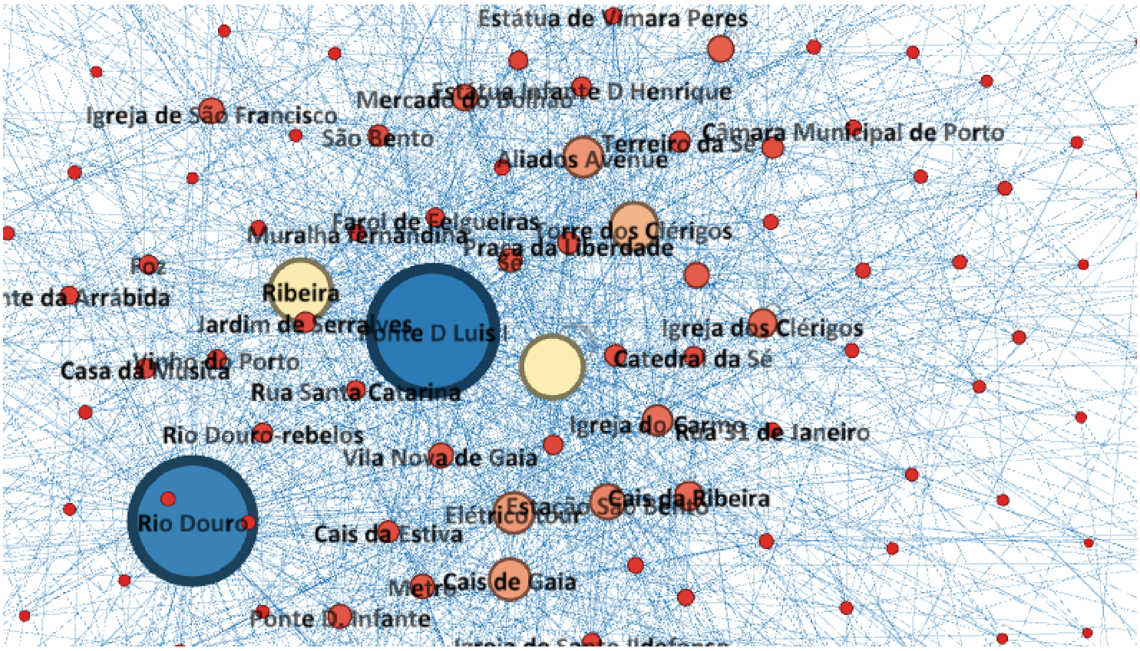


Figure 17: Network Centre and most weighted nodes based on the attractions photographed/visited, using the database collected from Flickr between January 2014 and May 2016. Source: Gephi software

The lines represent all the connections between the nodes, which are represented by the circles. Quickly we can conclude that there are eight places with higher frequency than the others, given their weight represented by the circles' size and color. A connection between two different nodes means that at least one tourist went from the first node to the second one. The size and color of the circles represent the relative frequency in the network. The bigger ones and the darker ones are the ones that appear the most in tourist's paths. At the first sight, the Network tell us that tourist's present a similar pattern across time since they are mostly concentrated on a small range of common spots (bigger nodes) and them they disperse through the city. There are a few attractions that appear in practically all paths (the blue ones) and them a range of spots that are very popular as well (the yellow and the light orange ones). In order to study the network represent above, a range of statistics provided by Gephi software was used. Given the low- level of accuracy of those places, we focused on the results provided by the metrics described in chapter 2. The results, provided by Gephi, are avaiable on Annex 16.

The Degree distribution shows that there are many places (nodes) with few edges connected and a few places with a range of connections associated, which is very clear to observe on the figures above by the connections higher density around the bigger nodes. Considering the in-degree and the out-degree results, we observe that the tendency is the same. If we recall the results from the Sequency Mining analysis, we can compare to the paths that were related to Douro River and D Luis Bridge, for instance, across the different experiences and we observed that a massive part of the paths started or ended in one of these places.

After the Degree distribution, we performed the Betweenness centrality and we can observe that there is a small range of nodes with high betweenness centrality which means that these attractions may have considerable influence in the network. Considering the Sequence Mining results and the Degree distribution results we may conclude that the nodes with higher betweeness are the nodes with more connections associated.

The Distance and Diameter metrics revealed that, on average, tourists pass through four places during their trips and the longest distance between two nodes (places) is ten places. In other words and considering a trip of 1-3 days (on average), tourists visit around four attractions during this days (on average too).

The Eigenvector Distribution reveals the node importance in a network based o a node’s connections so it can be the best metric that defines is what best defines the importance of a place in the network. The results showed that there are a few nodes (places) with significant importance in terms of node’s connection and most of them with residual relevance for the Network.

Results discussion

The network analysis revelad that tourists are spatially concentrated on a few main locations related to the historical centre and the riverside, as Hayllar and Griffin (2009 *cf* Edwards & Griffin, 2013) research also demonstrated- It is also possible to observe that most part of the individuals in our sample tend to look for a range of common spots, no matter their motivations. We can observe a profound connection between such places since they appear in most part of individuals’ trips. If we consider the results presented in Annex 14, and cross them with the results of the diameter metric, we can observe that most paths found are composed by two places so it left us to consider that tourist shared half part of the most travelling path and then they disperse through the city.

Crossing with the information provided by the Sequence Mining analysis we can easily observe that the places with higher degree level also have the highest betweenness centrality such as Douro River, Luís I bridge and Ribeira, followed by Clérigos Tower and Estação de São Bento for instance. Not only are the most photographed places (connections), but also are the most frequent items in the trips (betweenness). We also can find attractions with high degree-level, but less-betweenness centrality which can mean that there are places with a lot of connections associated, but they appear less frequently in tourists paths. From the insights provided by the last section and observing the results in Annex 15, we can conclude that those places are Bolhão, Casa da Música and Lello & Irmão Bookstore.

The results of this analysis are a complement to the insights of the sequence mining approach because we were not able to find communities among the attractions. We were expecting to observe profound connections between a certain numbers of attractions that could support the travelling paths discovered. In fact, network analysis came to prove that tourist's are spatially concentrated around a group of places, despite their preferences.

4.4 Finding market segments

Finally, a cluster analysis will be applied in order to find clusters in our sample, through the algorithm of K-modes developed for the effect, as it is explained on Chapter 2.

The segmentation-based applied to perform tourist's market segmentation in this study are hybrid and mostly based on a demographic and geographic approach, as it is exposed on section 2.3.3.

K-mode algorithm is used on cluster analysis for categorical data, since the K-means algorithm is applied on numeric data. The algorithm approach consists of converting categories' attributes into binary ones (in which 0 represents the absent and 1 the present for each one) and, with that, k-means algorithm is able to read the data as a numeric one.

In order to find the optimal number of clusters, we run several times the algorithm presented on Annex 9, by replacing the variable x on the code - *cl <- kmodes(data1, x)*- that represents the number of cluster to be found and we choose the number most appropriate, according to the results provided.

We performed three experiences, considering the personal information available on flickr.com.

Experience 1: Gender

The first experience was based on individuals' gender. In this experience, and considering the algorithm present on Annex 9, we assume that there are differences between male and female tourists, so we applied a demographic segmentation strategy. The results are illustrated on table 6. It was possible to segment the market into 10 different segmetns. It is relevant to mention that most part of the male segments have Luis I bridge, Douro River and Porto landscapes as places in common.

Cl	Variable	Nº	Characterization	Locations
1	**Female**	1	Belgium and married;	Clérigos, Posto de turismo na Sé and Terreiro da Sé;
2	**Female**	3	1/3 are Canadian and 1/3 is British; There is no more significant information;	Aliados, Gaia, Livraria Lello & Irmão, Palácio de Cristal, Luis I bridge, Praça da Liberdade, Ribeira, Douro River, beach, etc
3	**Female**	1	A married woman; graphic designer;	Praia, Porto landscapes, Porto streets, Teleférico de Gaia, Porto wine caves;
4	**Female**	29	More than 50% are European and 10% came from American countries; Amost1/3 are married or single;	NA
5	**Female**	3	European woman; 1/3 are single;	Douro River, São Bento station and Cais de Gaia;
4	**Male**	29	Almost 80% are European (50% came from France and Portugal); 25% are single and almost 40% are married;	Luis I bridge, Douro River and Cais de Gaia;
5	**Male**	40	Amot 50% are European (Portuguese, Spanish and British). All british are married; 15% came from America; All Brasilian are married too;	Porto landscapes, Douro River and Luis I bridge
6	**Male**	126	90% are European; 25% are Spanish and married as well as the Dutch; All the Canadians are single; Almost 8% are American and the remain part are Asiatic;	NA
7	**Male**	3	1/3 is single and American; 1/3 is European;	Castelo do queijo, Catedral da Sé, Old train, Aliados, São Bento station, Estátua Antonio Ferreira Borges, Estátua D. João VI, Estátua D. Pedro IV, Estátua Infante D Henrique, Fonte dos Leões, Funicular dos Guindais, Matosinhos beach, Mercado Ferreira Borges, Igreja da Sé, Igreja das Carmelitas, Museu Carro Elétrico, Igreja de São Francisco, Igreja de São Lourenço, Museu Transportes, Igreja do Carmo, Igreja dos Clérigos, Igreja Santo António dos Congregados, Palacete Correia de mello, etc
8	**Male**	2	Married men;	Aliados, Old train, Livraria Lello & Irmão, Igreja das Carmelitas, Igreja de São Lourenço, Porto landscapees, Ribeira, Douro River, Sé, Cais de Gaia, Sandeman caves
9	**Male**	1	Polish single man;	São Bento station, Dragan stadium, Houses,

				Igreja de Santo Ildefonso, Jardim do Infante D Henrique, Porto landscapes, Douro River, Alfandega, Rua das Flores, Terreiro da Sé, Torre dos Clérigos.
10	**Male**	3	2/3 are European; 1/3 is single	Caves Ferreira, São Bento, Livraria Lello & Irmão, Igreja de Santo Ildefonso, Igreja de São Francisco, Igreja dos Clérigos, Ocean, Igreja Santo António dos Congregados, Piolho, Praça Liberdade and Douro River.

Table 6: Results from the cluster analysis for the first experience. Source: Own elaboration.

Experience 2: Social State

We performed the same analysis based on individual's social state – a demographic strategy - and we were able to find some differences between the married and the single travelers, as it is possible to observe on table 7. It this case, a massive part of the married segments have Luis I bridge in common.

Cl	Variable	Nº	Characterization	Locations
1	Married	48	95% are men, mostly from Portugal (18%), Spain (14%) and from the UK (11%). Almost 30% are retired or a photographer;	NA
2	Married	1	American men;	Caves Sandeman, Aliados, Igreja de São Francisco, Igreja do carmo, Igreja dos Clérigos, Luis I bridge, Porto Cruz, Douro River, Galerias de Paris, Fernandes Tomás, Serra do Pilar, Porto wine caves, Cais de Gaia
3	Married	4	¾ of the tourists are European; ½ is retired	Luis I bridge, Passeio Alegre, Sé, Teleférico de Gaia, Ribeira, Cais de Gaia;
4	Married	10	1/3 are European woman; The men came are from Spain, Germany, Ireland and Portugal essentially.	São Bento station, Luis I bridge, Douro River, Cais de Gaia, Casa da Música;
5	Married	2	One men and one woman both photographers;	Restaurants, Gaia;
6	Married	7	French and British men;	Aliados;
7	Single	1	Single Portuguese student man;	Aliados, Porto wine caves, Igreja dos clérigos, Douro River, streets, São Bento, Terreiro da Sé, Trindade, Cais de Gaia, Capela Senhora do Além;
8	Single	23	NA	NA
9	Single	9	1/3 is from the UK; 2/9 are students; 8/9 are European and men;	São Bento station, Livraria Lello & Irmão, Ribeira, Douro River, Santa Catarina, Torre dos Clérigos, Cais de Gaia;
10	Single	4	½ are woman and ¼ are Asian;	Livraria Lello & Irmão

Table 7: Results from the cluster analysis for the second experience. Source: Own elaboration.

Experience 3: Origin

The third experience was based on individual's origin – a geographic segmentation strategy. For this experience we classified the tourists as European, American and others. The results are exposed on table 8. In this case the initial assumption is that tourists are different according their origin.

Cluster	Variable	Nº	Characterization	Locations
1	Europe	8	60% are Portuguese and French (1/3 of the French are married);	São Bento station;
2	Europe	43	Man from Spain and from the UK. More than 50% are married;	Douro River;
3	Europe	105	NA	NA
4	America	2	NA	Cultural attractions - Catedral da Sé, Vimara Peres statue and Clérigos Tower - Aliados, Flores street and Teleferico de Gaia.
5	America	10	All Brazilian and Canadian individuals are single and 50% of the American ones are retired.	Douro River;
6	South America	2	High level of studies;	Luis I bridge, Douro River, Ribeira square, São Bento station and Clérigos church;
7	America	13	NA	Luis I bridge and Aliados;

Table 8. Results from the cluster analysis for the third experience. Source: Own elaboration.

Results discussion

Considering the work developed by Xia *et al.* (2010), Leung *et al.* (2012) and Vu *et al.* (2015), we fulfill market segmentation based on tourists' dominant movement patterns in Porto region through *geotagged* photos shared on Flickr.com.

We were able to find a range of groups among our sample considering a hybrid approach – a demographic or a geographical strategy plus travelling patterns data - for the segmentation process as well as individual's movement patterns across the city. Our sample is composed mainly for married European (mostly French, Spanish and British) man so it was expect that we were able to extract more information through those variables. The three experiences revealed some details about tourists market in Porto:

- We were able to find groups that searched for art, cultural and entertainment activities;

- The married European and retired tourists search for attractions near the river and the American ones for attractions in Gaia related with wine;
- The married man tourists search for wine activities and religious monuments and the female European single tourists for the river and for the riverside;
- Behavioural differences between Spanish and German tourists by opposite to the French and British ones etc.
- The youngest individuals in this sample (the single and student ones) are mainly Spanish and British.

In general, there are groups that search for cultural and historical attractions, for attractions in the historical center and others that search essentially by the riverside, despite their origin, gender or social state.

Despite the lack of information about individual's profession, we could observe through the information available that the sample present high-level of education, which might imply a higher purchase power and a higher interest by cultural facilities as well as an average age of above 25 years old. If we consider the information provided by photos related to accommodation, activities, buildings, restaurants and nightlife categories in this sample, it was possible to observe some interesting behaviours, but it was still not enough to get significant conclusions and relations between the choices made by tourists and their purchase power and lifestyle.

In other words, we were able to segment the market of tourism in Porto, but we believe the segments are not different enough to apply specific marketing strategies on each one. The reason is that a massive part of the photos shared on Flickr.com are highly concentrated around a small group of attractions, which are the most popular ones. The remain places visited (almost 500) are not expressive enough so the algorithm could not find patterns with it so they became statistical irrelevant. This behavior is acceptable since tourists do not share every step they take on a destination, they only share the most interesting spots that will catch the attention of their followers as well.

Considering the information provided by the Visit Porto department we can observe that the places more frequent in tourists' paths correspond to the highlighted places in the city maps. Moreover, the Porto Tourism department has already packages for the different types of tourists that visit our city such as families, couples, students, retired, people that search the city for cultural/religious motives, specific holidays and events (like São João or Carmelitas week) and nature activities. Looking to the pictures present in our sample, we are able to find all the places signed in those maps. Since we do not have further information about who our tourists came to the city, it gets difficult to classify the segments found in this way.

With reference to the work of Xia *et al.* (2009 *cf* Leung *et al., 2*012) and Holden (2000 *cf* O'Connor *et al., 2*005), the traveler behaviour is defined by human 'push' factors (like personal motivations), physical 'pull' factors (ex: destination geomorphology and configuration) and time factors (ex: total trip duration). In this study we were able to identify the two last factors, as explained before. Looking at the attempts of classify tourists paths of Grönroos (1989), Lue *et al.,*(1992 *cf* Leung *et al., 2*012), Flogenfeldt in 1999 (*cf* Leung *et al., 2*012) and Lau and Mckercher (2006) into three, four, five and six groups, we could observe that these studies are mostly based on the push factors of traveler's behaviour by opposite to the ones applied in this study. However, some of the dualities presented on Mckercher, Shoval and Birenboim (2012) were found in our study (first time vs repeat, domestic, etc). So, we concluded that the previous work related to the study of tourists' market segmentation through behaviour patterns did not provided significant insights for us since our based-segmentation was related with geographical and demographical data.

Chapter 5 | Conclusions and further discussion

Final conclusion

Even in tourism, marketing efforts must be focused on a specific tourist group, which requires tourist's market segmentation. Our study targets the tourists' travelling patterns across the Porto city, based on data extracted from Flickr.com. The results were not as we expected at the beginning of this research, mostly because individuals did not share a significant part of their personal information that was necessary to define the market segments.

Starting with the first insights provided by the database characterization, the information provided by it was not enough to get significant conclusion and relations between the tourist choice and its purchase power and lifestyle. However, it was possible to observe that there is an increase of search for apartments and rental houses instead of hotels; Douro cruises are a popular activity in Porto as Tourism Porto Department data also confirm as well as walking, cycling and fishing.

Regarding the differences between *first time* and *repeat* visitors proposed by Mckercher, Shoval and Birenboim (2012), we could not get significant conclusions with our data since first time tourists represent more than 92% of Porto tourists. Our data collides with the information provided by the Visit Porto tourism in which 70% of the tourists that visited tourism post-offices were *first time* ones. In fact, most of the tourist that are classified as repeat visitors in this study are Portuguese people so we believe that the geographical proximity can be reason that explains the number of visits and the longer duration of each trip.

With respect to the places visited and photographed, we could observe that tourists are spatially concentrated around the river and around the city historical centre. We can also observe that they choose to cross over the river not only because their interest on visiting the city, but also because they intend to take pictures from there, which explains the range of photos taken from Vila Nova de Gaia regarding places or attractions located in Porto.

With respect to the most photographed attractions, the results were similar to the data provided by Porto Municipal Tourism Department as well. However, data came from all tourists that visited Porto Tourism points along with data provided by institutions that keep records of tourist's visits (such as hotels and museums), which is a very different data source compared to ours. We do not have access to the number of tourists presented in our database that visited tourism post offices in Porto. As so, we could believe indeed that individuals in our database looked for the same attractions or, at least, looked for the most popular ones whether in terms of marketing advertising, social network recommendations, other tourist's opinions, historical reasons and others.

With reference to the sequence mining analysis, the results provided significant insights regarding where do tourists go and what are trips through space and time. We expect to find longer frequent paths with higher frequency that the ones identified. However, given the high concentration of the tourists around a small group of attractions (observed on the Social Network analysis), the results are in line with tourists' behavior. Additionally, we were able to observe that a massive part of the tourists' paths ends on the river or on the bridge, presenting a downstream direction.

Considering the results provided by the network analysis, we were not able to identify significant relation between the attractions visited so it can be used to understand tourist's behavior. The results of the Social Network analysis basically completed the insights provided by the Sequence Mining analysis. It was expected to find communities among the attractions but the analysis only provided high connections between these attractions.

Regarding our last research question, the cluster analysis was able to identify tourists market segments based on their behavior and personal characteristics. However, given the lack of personal information plus the tourists travelling behavior, we believe that the segments found are not detailed enough to perform specific marketing strategies. As we can see on the database characterization, we identified more than 250 tourists, 8000 photos and more than 500 places. Looking at the sequence mining and to the network results, a massive part of these places does not appear. Their level of frequency is not enough to be inserted on a travelling pattern. We also believe that a sample with more photos will not resolve this matter since tourists share on their network the places that they believe that are the most interesting, popular or beautiful ones, which is perfectly understandable. This fact brought issues to our analysis and also revealed city image issues – as the Visit Porto also highlighted – because it makes the city looks smaller than it is.

Nevertheless, the results can be useful for marketing and management purposes: Dibb and Simkin (1991), Dolničar (2004; 2007), Larsen (2010) and Beeco and Hallo (2014) highlight the importance of market segmentation on a tourism context, which aims to support the development of tourist product, packaging, branding and promotion planning. In the case of Porto, it is clear than the historical center and the riverside suffers from overload capacity but such concentration can be benefic in terms of promotional campaigns for instance (Zimbardo, 1992 *cf* Li, Huang and Christianson, 2016). Hotels, other accommodation types, restaurants and other services can take advantage of such information to define its P of place (Asakura and Iryo, 2007; Xia et al., 2009 *cf* Leung *et al.,* 2012; Vu *et al.,* 2015) and managers can use it do add new attractions to the most frequent routes since, they are aware of the attractions that works as movement accelerators in Porto city or to programming new ones (Manning, 2011 *cf* Beeco and Hallo, 2014).

The growth of city tourist search in the past three years came from the efforts made and the increase of awareness of Porto city around the world. According to the Visit Porto Department[3], the city destination managers aim to extend the boundaries of Porto behind the city historical center through some specific attractions that work as an accelerator of the tourists' flow. Besides, Portuguese urban tourism is characterized by short trip duration - like it is presented on Coutinho (2012) research - so it was expected that the average number of days spent in the city remained low. The city managers' efforts on communication and the investment on the city image are in line with the research of Rita (1995 *cf* Coutinho, 2012), Grönroos (1989), Kolb (2006) and Gilaninia and Mohammadi (2015), since the promotional strategy in marketing cities is the most important factor in a marketing city context.

Study limitations

The first limitation of this study relies on the similarity between our data and the data provided by Visit Porto. Our sample was created based on tourists that have visited Porto city between January 2014 and May 2016, members of Flickr, and decide to share photos that they took during their visit to Porto. Despite the similarities found between our study

[3] Meeting with Maria do Carmo Costa of the Porto Official Tourism Board that belongs to the Porto City Hall at March, 2016.

and Visit Porto data, we are not aware of the total number of individuals that visit Porto during the period mentioned which are also members of Flickr so we cannot prove that we are talking about the same reality.

The second limitation is related with the lack of tourist's personal information presented on Flickr.com. We only obtained complete information from 22,5% of the users considered in the study. As Flickr users are not obligated to share their personal information, it is important to mention that a large percentage of information is not available or does not exist at all about the users considered in this study. The lack of data brought some issues when we fulfill individuals' characterization (for instance there is not any information about individual's age).

The third one is related to the trip duration definition. One of the main assumptions of this study is the individual's' trip duration. Each trip duration was calculated based on the distance that exists between the last and the first date-time photos that compose each trip, considering the same approach on the Girardin et al. (2007) work. However, as it happens with this sample, it is not possible to assume that such data portrays the real individual's trip duration.

Regarding tourists trips, we can consider for the effects of this study that a place that has been photographed represents a place that has been visited for a tourist and those photos, chronologically organized, represent a tourist path. Moreover, we cannot assure that tourist share on their Flickr accounts all the photos of their trips so we do not know if they visited another places during their staying.

Finally, the last limitation relies on the question: where do tourists go and what tourists have most photographed. In fact, we observed that sometimes the object/place/attraction that it is been photographed does not corresponds exactly to the spot that the individual is physically at the moment. However, there is a strong connection between the photos taken and the places where tourists are physically so we assume that the place photographed is the place physically visited. On a bigger perspective, this assumption could hide overload problems in some urban areas, for example, or even hide unexplored locations by business and tourism organizations.

Further recommendations

The insights provided by this kind of study can be applied on recommendation systems applications and on the development of marketing strategies, advertising and promotion.

First, the information about tourist's path on a city can be useful in terms of marketing services, marketing strategies and for business decision-making process in general (Shoval and Isaacson, 2006; Asakura and Iryo, 2007; Xia *et al.,* 2009 *cf* Leung *et al.*, 2012; Vu *et al.,* 2015). Following Modsching *et al*. (2008) research, it is essential for the tourism industry to know the places and times tourists visit. For instance, restaurants use this information to find locations that are most frequently seen by tourists. If the purpose is to open a new establishment or decide whether certain event will occur or not it is crucial to recognize where tourist go but also their personal and cultural characteristics.

Second, new applications like the recommendation systems have been developed based on tourists behaviour. The research of Spyrou and Mylonas (2014) is a perfect example of the Flick.com tourist's data applications on recommendation systems. The typical applications are focused on automatically discovering main attractions and then they allows users to decide which one to visit. Others use *geotagged* photos and data clustering in order to determine urban areas of interest or to find trends in tourist attractions in cities. Some authors went further and used this systems to not only recommends main attractions, but also tries to organize the user's' schedule and help them visit as many as they wish in a time efficient way. The systems support tourist's decision-making process when it comes to choose where to go.

Additionally, social media play a significant role in the analysis of tourists' attitudes fed by the increased purchases and recommendations to other users. Individuals' are more sincere when they share opinions in the Social Networks. Some of the shared information is even related to a company's products and services. Since the information shared by other~~s~~ tourists has a significant weight on the user decision-making process marketing can explore this opportunity to improve not only customer relationships but also tourist experience, the main factor when it comes to marketing cities. Besides that, it can support market segmentation and improve advertising and promotion. For this case, it can be used to promote and increase the number of visits of some attractions and with that decrease the excessive tourist spatial concentration in others or even to create awareness of a certain attraction that is new or secondary. Finally, if the insights provided by the Sequence Mining approach could be considered, we can introduce new places into individuals' trips and encourage them to run away from the people's flow and with that support services development. It could also guide an individual during the trip considering his or her profile, segment and, with that, improve marketing efficiency.

Finally, we believe that it is possible to fulfill a precise tourism market segmentation in Porto region using not only the information collected from Flickr.com and combine them with the information shared on other social Networks such as Tripadvisor- because it has important details about preferences, lifestyle and motivations – and Trivago – because it can provide information about trip type and purchase power.

References

Agência Lusa (2014), “Porto está num "ponto de viragem" e quer captar novos turistas”, Porto Canal: http://portocanal.sapo.pt/noticia/17812/, accessed on 15th April, 2015.

Aguiar, N. (2016), “Cinco gráficos que mostram a evolução do turismo em Portugal”, *Negocios.pt*,http://www.jornaldenegocios.pt/empresas/turismo___lazer/detalhe/cinco_graficos_que_mostram_a_evolucao_do_turismo_em_portugal.html, accessed on 27th June, 2015.

Akerkar, R. (2012), “Big Data and Tourism – To promote innovation and increase efficiency in the Tourism sector”, *TMRF White Paper.*

AMA, 2013. “About AMA – Definitions of Marketig”, https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx, accessed on 21st June, 2016.

Asakura, Y., & Hato, E. (2004). Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C: Emerging Technologies*, *12*(3), 273-291.

Asakura, Y., and T. Iryo (2007), “Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument”, Transportation Research Part A: Policy and Practice, Vol.41, pp. 684-690.

Ascensão, M.P. (2011), “Tendências Mundiais em Turismo”, O Turismo em 2011, IPDT, ed. 2011, ano V.

Ashworth, G., and S. Page (2011), “Urban tourism research: Recent progress and current paradoxes”, *Tourism Management*, 32(1), 1-15.

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Beeco, J. A., and J.C. Hallo (2014), “GPS Tracking of Visitor Use: Factors Influencing Visitor Spatial Behaviour on a Complex Trail System”, *Journal of Park and Recreation Administration*, 32(2).

Bessom, R. M., & Jackson, D. W. (1975). SERVICE RETAILING-STRATEGIC MARKETING APPROACH. *Journal of Retailing*, *51*(2), 75-84.

Blois, K. J. (1974). The marketing of services: an approach. European Journal of

Marketing, 8(2), 137-145.

Bucur, C. (2015), “Using Big Data for Intelligent Business”, *International Conference*

Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism management*, *21*(1), 97-116.

Carvalho, M. (2013), “O que o Porto tem”, *Revista Visão edição 1075*, http://visao.sapo.pt/porto-foi-eleito-o-melhor-destino-europeu-de-2014=f752439, acessed on 27th June, 2015.

Chung, J. Y., and D. Buhalis, "Information needs in online social networks." *Information Technology & Tourism*, 10 (4) pp 267-281.

Coutinho, M.D.G. (2012). “Estudo empírico exploratório sobre os turistas no Porto”. Marketing dissertation oriented by Professor Teresa Fernandes Silva, Faculty of Economics of Porto University.

Constantinides, E. (2006). The marketing mix revisited: towards the 21st century marketing. *Journal of marketing management*, *22*(3-4), 407-438.

Cormack, R. M. (1971). A review of classification. Journal of the Royal Statistical Society. Series A (General), 321-367.

Correia, J. (2016), “Acredito que a cidade tem muito por onde crescer no turismo”, *Jornalismo Porto JPN*, https://jpn.up.pt/2016/05/11/porto-acredito-cidade-crescer-no-turismo/, acessed on 29th June, 2015.

Crandall, D. and N. Snavely (2011), “Networks of Photos, Landmarks and People”, Leonardo, Volume 44(3), pp. 240-243.

Davenport, T. H. (2013), “ At the Big Data Crossroads: turning towards a smarter travel experience”, Amadeus IT Group.

De Mauro, A., Greco, M., and M. Grimaldi (2015), “What is big data? A consensual definition and a review of key research topics*”, AIP Conference Proceedings*, Vol. 1644, pp. 97-104.

Dias, R.F.L (2015), “Communities and Centralities dynamics in a stock market network”, Modelling, Data Analysis and Decision Support Systems dissertation supervised by Professor João Gama and Professor Carlos Ferreira, Faculty of Economics of Porto University.

Dibb, S., & Simkin, L. (1991). Targeting, segments and positioning. International Journal of Retail & Distribution Management, 19(3).

Dijkstra, L., and H. Poelman (2012), " Cities in Europe: the new OECD-EC definition", *Regional Focus, 1*

Dolničar, S. (2004). Beyond "commonsense segmentation": A systematics of segmentation approaches in tourism. Journal of Travel Research, 42(3), 244-250.

Dolnicar, S. (2007). 8 Market Segmentation in Tourism. Tourism Management: Analysis, Behaviour, and Strategy, 129.

Edwards, D., and T. Griffin (2013), "Understanding tourists' spatial behaviour: GPS tracking as an aid to sustainable destination management", *Journal of Sustainable Tourism* Vol. 21, pp. 580-595.

Fang, H., Z. Zhang, C. J. Wang, M. Daneshmand, C. Wang, and H. Wang (2015), "A Survey of Big Data Research", *IEEE Network.*

Ferrell, O. C. and Lucas, G. H. (1987). An Evaluation of Progress in the Development of a Definition of Marketing. *Journal of the Academy of Marketing Science, 15*(3), 12-23.

Franklin, A. (2003), "Tourism: an introduction", Sage.

Fuggle, L (2015), "6 travel trends for 2016 that will drive the global tourism industry", https://www.trekksoft.com/en/blog/travel-industry-trends-2016, accessed on 4th June, 2016.

Fyall, A., and B. Garrod (2005), "Tourism marketing: A collaborative approach (Vol. 18)", *Channel View Publications*.

Fotis, J., Buhalis, D., Rossides N., (2012). "Social media use and impact during the holiday travel planning process", Information and Communication Technologies in Tourism, pp. 13-24

Gilaninia, S. and Mohammadi, M.S.O. (2015), "Impact of Marketing on Tourism Industry", *International Journal of Business and* Behavioura*l Sciences*, Vol. 5, No. 2.

Girardin, F., Dal Fiore, F., Blat, J., and C. Ratti (2007), "Understanding of Tourist Dynamics from Explicitly Disclosed ", Vol. 58, Symposium on LBS and Telecartography.

Shostack, G. L. (1977). Breaking free from product marketing. *The Journal of Marketing*, 73-80.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. Academy of

Management Journal, 57(2), 321-326.

Gephi.org. "The Open Graph Viz Platform", https://gephi.org/features/, accessed on 23rd of July, 2017.

Goi, C. L. (2009), "A review of marketing mix: 4Ps or more", *International journal of marketing studies*, 1(1), 2.

Grõnroos, C. (1989), "Defining marketing: a market-oriented approach", *European journal of marketing*, 23(1), 52-60.

Grönroos, C. (1994). From marketing mix to relationship marketing: towards a paradigm shift in marketing. *Management decision*, *32*(2), 4-20.

Herbert, D., & Thomas, C. (2013). Cities in space: city as place. Routledge.

Huang, Z. (1998), "Extensions to the k-means algorithm for clustering large data sets with categorical values", *Data mining and knowledge discovery*, 2(3), 283-304.

Huang, J. Z. (2009), "Clustering Categorical Data with k-Modes".

Intel IT Centre (2012), "Peer Research - Big Data Analytics: Intel's IT Manager Survey on How Organizations Are Using Big Data", http://www.intel.com/content/www/us/en/big-data/data- insights-peer- research-report.html, accessed on 27th October, 2015.

Ji, R., Gao, Y., Zhong, B., Yao, H., & Tian, Q. (2011). Mining flickr landmarks by modeling reconstruction sparsity. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *7*(1), 31.

Joachimsthaler, E. (2013),"Driving New Growth Through Big Data", *Harvard Business Review Analytic Services Report*, pp 18-20.

Kádár, B. (2014), "Measuring tourist activities in cities using geotagged photography". *Tourism Geographies*, *16*(1), pp 88-104.

Kádár, B. and M. Gede (2013), "Where do tourists go? Visualizing and analysing the spatial distribution of geotagged photography", *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2), pp 78-88.

Kolaczyk, E. D. and G. Csárdi (2014). *Statistical analysis of network data with R* (pp. 1-5). New York: Springer.

Kolb, B. M. (2006), "Tourism Marketing for cities and towns: using branding and events to attract tourism"., Vol. 10, Routledge.

Kotler, P. (1967), “ Marketing Management Analysis, Planning, and Control”, 9th ed. Prentice Hall Inc., Englewood Cliffs, NJ,

Kotler, P., & Levy, S. J. (1969). Broadening the concept of marketing. *The Journal of Marketing*, 10-15

Kotler, P. (1972). A generic concept of marketing. *The Journal of Marketing*, 46-54.

Kotler, P., and K.L. Keller (2008), “Upravljanje marketingom”.

Kotler, P., and K. Keller (2011), Marketing Management 14h edition. Prentice Hall.

Laney, D. (2001).” 3D Data Management Controlling Data Volume, Velocity and Variety”, Meta Group, Application Delivery Strategies, file 949, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf, accessed on 29th October, 2015.

Lau, G., & McKercher, B. (2006). Understanding tourist movement patterns in a destination: A GIS approach. *Tourism and hospitality research*, *7*(1), 39-49.

Larsen, J., J. Urry, and K.W. Axhausen (2006). “NETWORKS AND TOURISM: Mobile Social Life”, *Annals of Tourism Research*, Vol. 34, No. 1, pp. 244–262, 2007.

Larsen N. (2010), “Market Segmentation - A framework for determining the right target customers”, *Aarhus School of Business.*

Law, C. (1993), “Urban tourism: attracting visitors to large cities*”, Mansell Publishing Limited.*

Law, C. (2002), “Urban tourism: The visitor economy and the growth of large cities”, *Cengage Learing Emea.*

Leung, X., Wang, F., Wu, B., Bai, B., Stahura, K., & Xie, Z. (2012), “ A Social Network Analysis of Overseas Tourist Movement Patterns in Beijing: The Impact of the Olympic Games”, *International Journal of Tourism Research*, Vol.4, pp. 469-484.

Li, Q., Huang, Z. J., and K. Christianson (2016), “Visual attention toward tourism photographs with text: An eye-tracking study”, *Tourism Management*, 54, 243-258.

Lo, I., Mckercher, B., Lo, A., Cheung, C., and R. Law (2011), “Tourism and online photography”, *Tourism Management*, 32(4), 7257131.

Luo, Q., and D. Zhong (2015), “Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites”, *Tourism Management*, 46, pp 274-282.

Lusa (2015), “Turismo de Portugal: época de 2015 terá novos máximos”, *Jornal Económico*, http://economico.sapo.pt/noticias/turismo-de-portugal-epoca-de-2015-tera-novos-maximos_226904.html, accessed on 27th June, 2015.

Masiero, L., & Zoltan, J. (2013). Tourists intra-destination visits and transport mode: A bivariate probit model. *Annals of Tourism Research*, *43*, 529-546.

Mayer-Schönberger, V. and K. Cukier (2013). “Big data: A revolution that will transform how we live, work, and think”. Houghton Mifflin Harcourt.

McCarthy, E. J.: Basic Marketing, A managerial approach, Richard D. irwin, Homewood Ill., 1960

Mckercher, B., Shoval, N., and A. Birenboim (2012), “ First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong”, *Tourism Geographies*, Volume 14(1), pp 147-161.

McKercher, B., Wong, C., and Lau, G. (2006), “How tourists consume a destination”, *Jornal of Business Research*, 59(5), pp. 647-652.

Memon, I., Chen, L., Majid, A., Lv, M., Hussain, I., & Chen, G. (2015). Travel recommendation using *geotagged*ged photos in social media for tourist. Wireless Personal Communications, 80(4), 1347-1362.

Microsoft Excel 2013, https://products.office.com/en-us/microsoft-excel-2013, accessed on 14th of September, 2017.

Middleton, V. T., Fyall, A., Morgan, M., & Ranchhod, A. (2009). Marketing in travel and tourism. Routledge.

Moorthy, J., R.Lahiri, N. Biswas, D. Sanyal, J. ranjan, K. Nanath and P. Ghosh (2015), “Big Data: Prospects and Challenges”, *The Journal for Decision Makers*, 40(1) pp. 74–96.

Moré, M.E. (2012). “The Evolution of Marketing”, http://morethanbranding.com/2012/04/30/the-evolution-of-marketing/, accessed on 21st June, 2016.

Munar, A.M. and J.K.S. Jacobsen (2014). “Motivations for sharing tourism experiment through social media”, Tourism Management 43 (2014) 46-54.

Newman, M. Networks: an introduction. 2010. United Slates: Oxford University Press Inc., New York, 1-2.

Niemczyk, A. (2014), “The Application of Path Modelling in the Analysis of Consumer Behaviour in the Cultural Tourism Market”, *Economics & Sociology*, *7*(1), 204.

Nov., O., and C. Ye (2010), “Why do people tag?: motivations for photo tagging”, *Communications of the ACM*, 53(7), pp. 128-131.

Ng, M. K., Li, M. J., Huang, J. Z., & He, Z. (2007). On the impact of dissimilarity measure in k-modes clustering algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(3).

Oliveira, M. and J. Gama (2012), “An overview of social network analysis”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(2) pp. 99–115.

Pacific Asia Travel Association (1999), “Finding the right balance - 21 issues & trends that will shape travel and tourism in the 21st century*”, Issues & Trends Publication.*

Page, S. J. (2014), “Tourism management”, *Routledge*.

Pearce, D. (2001), “An integrative framework for urban tourism research”, *Annals of tourism research*, 28(4), pp. 926-946.

Prideaux, B. (2000). The role of the transport system in destination development. Tourism management, 21(1), 53-63.

Ralambondrainy, H. (1995), “A conceptual version of the K-means algorithm”, *Pattern Recognition Letters*, 16(11), pp. 1147-1157.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Salvador, A. B., & Ikeda, A. A. (2014). “Big Data Usage in the Marketing Information System”. Journal of Data Analysis and Information Processing.

Schroeck, M., R. Schokley, J. Smart, D. R. Morales and P. Tufano (2012), “Analytics: The real-world use of big data - how innovative enterprises extract value from uncertain data”, *IBM Global Business Services - Business Analytics and Optimization Executive Report,* http://www-03.ibm.com/systems/hu/resources/the_real_word_use_of_big_data.pdf, accessed on 29th October, 2015.

Shoval, N., & M. Isaacson (2007), “Tracking tourists in the digital age. Annals of Tourism Research”, *Annals of Tourism Research*, 34(1) pp. 141-159.

Shoval, N., and M. Isaacson (2010), “Tourist Mobility and Advanced Tracking

Technologies”, *Advances in Tourism*, vol. 19, Routledge.

O’Connor, A., Zerger, A., and B. Itami (2005), “Geo-temporal tracking and analysis of tourist movement”, *Mathematics and Computers in Simulation*, 69(1), pp. 135-150.

Selby, M. (2004), “Understanding urban tourism: Image, culture and experience”, *IB Tauris.*

Snijders, C., Matzat, U. and U.D. Reips (2012), "'Big Data': Big gaps of knowledge in the field of Internet". *International Journal of Internet Science 7: 1–5.*

Spyrou, E. and P. Mylonas (2014), “ An overview of Flickr challenges and research opportunities”, *In Semantic and Social Media Adaptation and Personalization (SMAP*), 2014 9th International Workshop on (pp. 88-93). IEEE.

Sylvester, J. J. (1878), “On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, with three appendices”, *American Journal of Mathematics,* 1(1), pp 64-104.

Vargo, S.L. and R.F. Lusch (2004), “Evolving to a New Dominant Logic for Marketing”, *Journal of Marketing,* Vol. 68, No. 1 pp. 1-17.

Versichele, M., De Groote, L., Bouuaert, M., Neutens, T., Moerman, I., and N. Van de Weghe (2014), “ Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium”, *Tourism Management*, Vol.44, 67-81.

Van Dijk, J. (2012). The network society. Sage Publications.

Visit Porto (2015), “Investir no Porto”, http://www.visitporto.travel/Negocios/Paginas/Investir/Investir.aspx?artigo=18, acessed on 27th June, 2015.

Vu, H., Li, G., Law, R., and B. Ye (2015)” Exploring the travel behaviours of inbound tourists to Hong Kong using geotagged photos”, *Tourism Management*,46, pp. 222-232.

Tutte, W. T. (2001). Graph theory, ser. Encyclopedia of mathematics and its applications. Cambridge, UK, New York: Cambridge University Press, 21.

Xia, J., Evans, F., Spilsbury, K., Ciesielski, V., Arrowsmith, C., and G. Wright (2010), “Market segments based on the dominant movement patterns of tourists”, *Tourism*

Management, Nº 31, volume 4, pp 464-469.

Xiang, Z. and U. Gretzel (2010),"Role of social media in online travel information search", *Tourism Management*, vol 31 pp 179–188.

Ward, J.S. and A. Barker (2013), "A Survey of Big Data Definitions", *School of Computer Science University of St Andrews UK*, arXiv:1309.5821v1.

Webster Jr., F.E. (2005), "A Perspective on the Evolution of Marketing Management", *Journal of Public Policy & Marketing*, , Vol. 24, No. 1 pp. 121-126 Published by: American Marketing Association.

Webster, M. (2015), "Simple Definition of city", http://www.merriam-webster.com/dictionary/city, accessed on 7th April, 2015.

William L. Wilkie and E. S. Moore (2007), "What Does the Definition of Marketing Tell Us about Ourselves?", *Journal of Public Policy & Marketing* , Vol. 26, No. 2 pp. 269-276.

Wind, J.,and Bell, D. (2007). Market segmentation. The marketing book, 222.

WWTC (2012), "Travel & Tourism, Economic Impact 2012 – Wold", http://www.wttc.org, accessed on 21st April, 2015.

Zaki, M. J. (2001), "SPADE: An efficient algorithm for mining frequent sequences", *Machine learning*, 42(1-2), pp. 31-60

Zeng, B. and R. Gerritsen (2014), "What do we know about social media in tourism? A review", *Tourism Management Perspectives*, 10 (2014) pp. 27–36

Zheng, Y., Zha, Z., and T. Chua (2012), "Mining travel patterns from geotagged photos", *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3), 56.

Zikopoulos, P., D. deRoos, C. Bienko, R. Buglio and M. Andrews (2014), "Big Data Beyond the Hype - A Guide to Conversations for Today's Data Centre".

Appendix

Annex 1. Marketing definitions according to each business orientation era. Source: Own elaboration.

Author	Definition	Era
AMA (1935)	The performance of business activities that direct the flow of goods and services from producers to consumers."	Sales Era
Ohio State University (1965)	"The process in a society by which demand structure for economic goods and services is anticipated and enlarged and satisfied through the conception, promotion, exchange, and physical distribution of goods and services"	Marketing Company Era
AMA (1985)	"The process of planning and executing the conception, pricing, promotion, and distribution of ideas, goods, and services to create exchanges that satisfy individual and organizational objectives".	Marketing Company Era
Cobra (1992)	Marketing is not just the development of products and services, but also a compromise with the people's quality of life improvement.	Marketing Relationship Era
Kotler *et al.* (1996)	"The process used to determine what products or services may be of interest to customers and the strategy to use in sales, communications and business development" (in kbmanage.com).	Marketing Relationship Era
AMA (2004)	"The process of planning and executing the conception, pricing, promotion, and distribution of ideas, goods and services to create exchanges that satisfy individual and organizational objectives".	Marketing Relationship Era
Kotler (2005)	"A social and management process through individuals and communities get what they need and desire by products and value creation and exchange" and it is one of the most complete definitions found in the literature".	Marketing Relationship Era

Annex 2. Tourism concept evolution. Source: Own Elaboration.

Perspective	**Definition**
Based on place factor (it is seen as an escape from the ordinary life)	Cohen (1972 *cf* Larsen *et al.*, 2007), Mathieson and Wall (1982 *cf* Franklin, 2003), Buckart and Medlik (1974 *cf* Franklin, 2003) and Urry (1995 *cf* Larsen *et al.*, 2007).
Based on necessary individual's movement (through time and space)	Mckercher *et al.* (2006 *cf* Leung *et al*, 2012) and Larsen *et al.* (2012).
Based on society evolution (consumerism of modern cultures and attitude towards the world).	Meethan (2001 *cf* Franklin, 2003), Alain de Boton (2002 *cf* Franklin, 2003) and Franklin and Crang (2001 *cf* Larsen *et al.*, 2007).
Based on the relational phenomenon	Weaver and Opeerman (2000 *cf* Franklin, 2003).
Based on the migration phenomenon	Williams and Hall (2000 *cf* Larsen *et al.*, 2007; 2012).

Annex 3. Existing definitions of Big Data, considering the four major key areas defended by De Mauro et al. (2014). Source: De Mauro et al. (2014).

TABLE 1. Existing definitions of Big Data, adapted from the articles referenced in the first column. The last four columns indicate whether the definition alludes to each of the four Big Data themes identified in the first section of the paper, through the following legend: I - Information, T - Technology, M - Methods, P - Impact.

Source	Definition	I	T	M	P
(Beyer & Laney 2012)	High volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.	x		x	x
(Dijcks 2012)	The four characteristics defining big data are Volume, Velocity, Variety and Value.	x			x
(Intel 2012)	Complex, unstructured, or large amounts of data.	x			
(Suthaharan 2013)	Can be defined using three data characteristics: Cardinality, Continuity and Complexity.	x			
(Schroeck et al. 2012)	Big data is a combination of Volume, Variety, Velocity and Veracity that creates an opportunity for organizations to gain competitive advantage in today's digitized marketplace.	x			x
(NIST Big Data Public Working Group 2014)	Extensive datasets, primarily in the characteristics of volume, velocity and/or variety, that require a scalable architecture for efficient storage, manipulation, and analysis.	x	x		
(Ward & Barker 2013)	The storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning.	x	x	x	
(Microsoft 2013)	The process of applying serious computing power, the latest in machine learning and artificial intelligence, to seriously massive and often highly complex sets of information.	x	x	x	
(Dumbill 2013)	Data that exceeds the processing capacity of conventional database systems.	x	x		
(Fisher et al. 2012)	Data that cannot be handled and processed in a straightforward manner.	x		x	
(Shneiderman 2008)	A dataset that is too big to fit on a screen.	x			
(Manyika et al. 2011)	Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.	x	x	x	
(Chen et al. 2012)	The data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies.	x	x	x	
(Boyd & Crawford 2012)	A cultural, technological, and scholarly phenomenon that rests on the interplay of Technology, Analysis and Mythology.		x	x	x
(Mayer-Schönberger & Cukier 2013)	Phenomenon that brings three key shifts in the way we analyze information that transform how we understand and organize society: 1. More data, 2. Messier (incomplete) data, 3. Correlation overtakes causality.	x		x	x

Annex 4. Tourists' characterization, ordered by id variable, based on the information collected from Flickr. Source: Own elaboration.

Owner_id	Gender	Country	Social State	Profession
1	Male	Germany	Married	Student
2	Male	Spain	Married	Art historian
3	Female	USA	n/a	Assistant Editor
4	Male	UK	n/a	n/a
5	Female	The Netherlands	n/a	n/a
6	Male	n/a	single	School librarian
7	Male	Belgium	n/a	n/a
8	Male	n/a	n/a	n/a
9	Female	France	n/a	n/a
10	Male	Spain	n/a	n/a
11	Female	Portugal	n/a	n/a
12	Male	Spain	n/a	Designer
13	Male	Portugal	n/a	n/a
14	Male	Portugal	n/a	n/a
15	Female	Romania	n/a	Model
16	Male	n/a	Married	n/a
17	Female	Portugal	single	Veterinary
18	Male	Italy	single	Computer engineer
19	Female	n/a	n/a	n/a
20	Female	Portugal	single	Student
21	n/a	n/a	n/a	n/a
22	Male	France	n/a	n/a
23	Male	Spain	n/a	n/a
24	Male	Spain	single	n/a
25	Male	UK	married	n/a
26	Male	Germany	married	Photographer
27	Male	UK	single	Retired
28	Male	Portugal	Married	n/a
29	Male	Belgium	n/a	Photographer
30	Male	n/a	n/a	n/a
31	Male	Portugal	single	Unemployed
32	Male	The Netherlands	n/a	n/a
33	Male	USA	single	Tour guide
34	male	n/a	n/a	n/a
35	Male	Spain	n/a	Designer
36	Male	n/a	n/a	n/a
37	Male	Portugal	single	Student
38	Male	Portugal	single	System administrator
39	Male	n/a	n/a	n/a
40	Female	n/a	Married	Graphic Designer
41	Female	n/a	single	n/a

42	Female	Ireland	Married	n/a
43	Male	Spain	Married	Tanker truck traffic manager
44	Male	Suisse	Married	Retired
45	Male	n/a	Married	n/a
46	Male	n/a	n/a	n/a
47	Female	Japan	single	n/a
48	Male	Austria	n/a	n/a
49	Male	Spain	Married	Retired
50	Male	Italy	n/a	n/a
51	Male	n/a	n/a	n/a
52	Male	France	n/a	Photographer
53	Male	Brazil	Married	n/a
54	Male	Portugal	Married	Manager
55	Male	Spain	n/a	n/a
56	Male	n/a	Married	n/a
57	Male	Poland	n/a	n/a
58	Male	Spain	n/a	n/a
59	Male	France	Married	n/a
60	Male	The Netherlands	Married	Computer technician
61	Male	Portugal	Married	Retired
62	Female	Iceland	Married	Teacher
63	Male	Spain	single	n/a
64	Male	Norway	n/a	Retired
65	Male	Canada	single	Student
66	Female	Italy	n/a	n/a
67	Male	France	married	Retired
68	Male	The Netherlands	n/a	Art Director
69	Male	UK	married	Retired
70	Male	Portugal	married	Photographer
71	Male	France	married	Photographer
72	Male	n/a	n/a	n/a
73	Male	n/a	n/a	n/a
74	Male	France	Married	Photographer
75	Male	Suisse	single	Photographer
76	Male	Portugal	Married	n/a
77	Female	Italy	n/a	Photographer
78	Male	UK	single	n/a
79	Male	Spain	Married	Businessman
80	Female	France	single	Student
81	Male	France	n/a	n/a
82	Male	Germany	n/a	n/a
83	Male	n/a	n/a	n/a
84	Male	n/a	n/a	n/a
85	Female	n/a	Married	n/a
86	Male	France	n/a	n/a

87	Male	Portugal	n/a	n/a
88	Male	Italy	n/a	n/a
89	Male	Portugal	single	Student
90	Male	n/a	n/a	n/a
91	Male	Portugal	n/a	n/a
92	Male	Spain	n/a	Journalist
93	Male	France	Married	n/a
94	Male	Spain	Married	n/a
95	Male	UK	n/a	Translator
96	Male	Spain	single	Journalist
97	Male	Germany	n/a	n/a
98	Female	France	Married	Retired
99	Female	n/a	n/a	n/a
100	Male	UK	single	Student
101	Male	Germany	Married	n/a
102	Male	USA	single	Retired
103	Male	Russia	single	n/a
104	Male	The Netherlands	n/a	Photographer
105	Male	France	n/a	n/a
106	Male	Spain	married	n/a
107	Male	Italy	n/a	n/a
108	Male	Macau	n/a	n/a
109	Male	n/a	n/a	n/a
110	Female	Germany	n/a	n/a
111	Male	Spain	married	n/a
112	Male	France	n/a	Sculpture
113	Male	Spain	n/a	Commercial
114	Male	n/a	Married	n/a
115	Male	n/a	n/a	n/a
116	Male	Spain	n/a	Designer
117	Female	n/a	n/a	n/a
118	Male	Portugal	single	Student
119	Male	n/a	n/a	n/a
120	Male	France	n/a	n/a
121	Male	The Netherlands	Married	n/a
122	male	France	n/a	n/a
123	male	Germany	n/a	n/a
124	Male	Canada	n/a	n/a
125	Male	Spain	n/a	n/a
126	n/a	n/a	n/a	n/a
127	Male	Portugal	Married	n/a
128	Male	Spain	n/a	Freelancer
129	Male	n/a	n/a	n/a
130	Male	France	n/a	Photographer
131	Male	Australia	Married	Retired

132	Female	n/a	Married	n/a
133	n/a	n/a	n/a	n/a
134	Male	USA	Married	n/a
135	Female	n/a	n/a	Photographer
136	Male	n/a	n/a	n/a
137	Male	Poland	single	n/a
138	Male	n/a	Married	Technieker audio&multimedia
139	Male	n/a	n/a	n/a
140	Male	n/a	n/a	n/a
141	n/a	n/a	n/a	n/a
142	Male	Ireland	Married	Domestic
143	Male	n/a	n/a	n/a
144	Male	n/a	n/a	Unemployed
145	Male	UK	Married	Consultant
146	Male	Portugal	n/a	n/a
147	Male	Portugal	married	Civil Engineer
148	Female	Spain	single	n/a
149	Male	n/a	n/a	n/a
150	Male	Italy	n/a	Designer
151	Male	n/a	single	Student
152	Male	Argentina	n/a	Software Engineer
153	Male	USA	married	UX Designer & Strategist
154	Male	n/a	n/a	n/a
155	Male	USA	Married	Retired
156	Male	n/a	n/a	n/a
157	Male	Spain	n/a	n/a
158	Male	Spain	n/a	n/a
159	Male	Spain	single	Student
160	Male	Portugal	n/a	n/a
161	Male	n/a	n/a	n/a
162	Male	UK	Married	Unemployed
163	n/a	n/a	n/a	n/a
164	Male	Portugal	n/a	n/a
165	Male	UK	Married	n/a
166	Male	Portugal	Married	Retired
167	Male	Brazil	single	Student
168	Female	France	n/a	Painter
169	Male	France	Married	Retired
170	n/a	n/a	Married	n/a
171	Male	France	single	n/a
172	Male	Luxemburg	Married	n/a
173	Male	UK	n/a	Photographer
174	Male	Portugal	n/a	Architect
175	Female	Spain	n/a	n/a
176	Male	Germany	single	Student

177	Male	France	n/a	n/a
178	Female	n/a	n/a	n/a
179	n/a	Spain	n/a	n/a
180	n/a	n/a	n/a	n/a
181	Female	n/a	single	n/a
182	Male	Portugal	n/a	n/a
183	Male	n/a	n/a	n/a
184	Male	The Netherlands	Married	Photographer
185	Male	Spain	n/a	writer
186	Male	Brazil	n/a	n/a
187	Male	n/a	single	n/a
188	Male	Portugal	single	Web Marketer
189	Male	Brazil	married	Photographer
190	n/a	n/a	married	n/a
191	Male	Russia	single	n/a
192	Male	France	n/a	Graphite
193	Male	n/a	n/a	n/a
194	Female	Canada	married	n/a
195	Male	n/a	married	Photographer
196	Male	Brazil	n/a	Industrial Designer
197	Male	France	n/a	n/a
198	Female	Belgium	married	n/a
199	Male	Estonia	n/a	n/a
200	Male	n/a	n/a	n/a
201	Male	France	n/a	Architect
202	Female	Portugal	n/a	n/a
203	Male	Germany	married	Public relations
204	Male	Japan	n/a	n/a
205	Male	Spain	n/a	n/a
206	Male	UK	n/a	VFX artist
207	Male	Spain	n/a	n/a
208	Male	n/a	n/a	n/a
209	Male	Portugal	married	Civil Engineer
210	Female	n/a	married	n/a
211	Male	Ireland	n/a	Unemployed
212	Male	n/a	n/a	Retired
213	Male	n/a	married	n/a
214	Male	Spain	n/a	n/a
215	Male	Belgium	single	n/a
216	Female	France	n/a	n/a
217	Male	Norway	n/a	n/a
218	Male	UK	married	Teacher
219	Male	France	n/a	n/a
220	Male	n/a	married	n/a
221	Male	USA	n/a	n/a

222	Male	UK	n/a	Retired
223	Male	USA	married	Visual artist
224	Male	Switzerland	married	Teacher
225	Male	Portugal	n/a	n/a
226	Male	Germany	n/a	n/a
228	Male	Spain	n/a	n/a
229	Male	UK	Married	Postman
230	Female	n/a	n/a	n/a
231	Male	Spain	married	Engineer
232	Male	Brazil	married	Tourist
233	Male	Portugal	n/a	n/a
234	Female	Canada	n/a	n/a
235	Male	Canada	n/a	n/a
236	Male	The Netherlands	Married	n/a
237	n/a	n/a	n/a	n/a
238	Male	Spain	Married	Photographer
239	Male	n/a	n/a	n/a
240	Male	Brazil	Married	n/a
241	Male	France	n/a	n/a
242	Male	Spain	n/a	n/a
243	Male	Portugal	n/a	Photographer
244	Male	Italy	n/a	n/a
245	Male	Japan	n/a	n/a
246	Male	USA	single	n/a
247	Male	Canada	single	n/a
248	Female	UK	n/a	Photographer
249	Female	The Netherlands	married	Pilot
250	Male	Germany	n/a	Retired
251	n/a	n/a	n/a	n/a
252	Male	France	n/a	n/a
253	Male	n/a	n/a	n/a

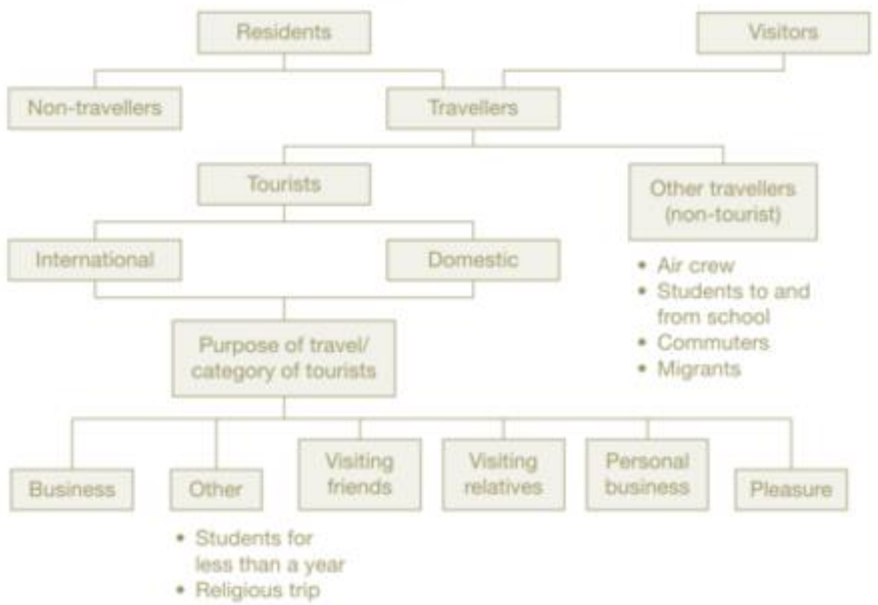
Annex 5: Query from Flickr API necessary to download textual metadata from Flickr users' photos. Source: https://www.flickr.com/services/api/.

https://api.flickr.com/services/rest/?method=flickr.photos.search&api_key=85cd5c617483fa336cdc4e5ce2911cf4&tags=Porto&min_upload_date=2014-01-01&max_upload_date=2016-05-30&accuracy=10&content_type=4&has_geo=1&extras=geo%2Cdate_taken&pe

Annex 6: Query from Flickr API necessary to obtain users personal information. Source: https://www.flickr.com/services/api/.

https://www.flickr.com/services/api/flickr.people.getInfo.html

Annex 7. Tourists' classification by Chadwick (1994). Source: Page (2014)



Annex 8. R code used to discover tourist's frequent movements. Source: Own elaboration.

```
library(arules)
library(arulesSequences)
library(Matrix)
x <- read_baskets(con = system.file("misc", "(file_name).txt", package = "arulesSequences"),
info = c("sequenceID","eventID","SIZE"))
inspect(x)
s1 <- cspade(x, parameter = list(support = 0.4), control = list(verbose = TRUE))
inspect(s1)
similarity(s1)
summary(s1)
```

Annex 9. K-mode algorithm used to perform cluster analysis. Source. Own Elaboration.

```
library(arules)
library(klaR)
library(MASS)
data2 <- read.csv("OwnersInfo.csv", header = TRUE, sep=";")
data22<- as.matrix(data2)
data <- read.csv("att.csv", header = FALSE, sep=";")
datam<-data[,c(2,5)]
colnames(datam)[1]<-"id"
dataMerge<-merge(datam,data2,by="id")
dataMerge<-dataMerge[dataMerge$Gender=="Female",c(1,2)]
dataMerge<-dataMerge[,c(1,2)]
write.table(dataMerge, file = "dataatt.csv",row.names=FALSE, na="",col.names= c("id","pl"), sep=";")
trans <-read.transactions("dataatt.csv", sep = ";", format="single", cols = c("id","pl"), rm.duplicates=TRUE)
mm <- as(trans,"matrix")
mm2<-cbind(rownames(mm),mm)
mm2[, 1]<- as.numeric(mm2[, 1])
data1 <- as.data.frame(mm2, stringsAsFactors=FALSE)
data1[, 1] <- as.numeric(data1[, 1])
data1<- data1[order(data1[, 1]), ]
colnames(data1)[1]<-"ID"
cl <- kmodes(data1[,-1], 3)
cl

inspect(trans)
cluster.output <- cbind(cl$cluster, data1)
write.csv(cluster.output, file = "kmodesclusters.csv", row.names = FALSE)
#rules <- apriori(trans, parameter=list(support=0.05, confidence=0.0))
#rules
```

Annex 10. Results of the inquiry performed by Visit Porto, in 2011, to tourists that bought Porto Card regarding number of tickets acquired in museums. Source: Visit Porto

Museus Municipais	2013	2014	2015
Museu Romântico da Quinta da Macieirinha	28.557	29.842	27.498
Casa Museu Guerra Junqueiro	25.545	27.234	23.559
Casa Museu Marta Ortigão Sampaio	9.621	11.106	7.694
Museu do Vinho do Porto	28.052	26.035	24.351
Casa Tait	13.882	8.054	8.208
Casa Oficina António Carneiro	4.524	4.791	4.753
Gabinete de Numismática	6.026	5.166	5.480
Arqueo-sítio da Rua de Dom Hugo, n.º 5	780	1.568	1.050
Visitas à Cidade*	3.419	5.580	8.287
Palacete Pinto Leite	8.285	4.174	4.261
Palacete Viscondes Balsemão	12.002	12.073	8.915
Banco de Materais	5.524	6.314	7.240
Núcleo Museológico da Casa do Infante	145.170	176.663	217.356
TOTAIS	**291.387**	**318.600**	**348.652**

Annex 11. Results of the inquiry performed by Visit Porto, in 2011, to tourists that bought Porto Card regarding places/attractions in which the card was used. Source: Visit Porto.

Annex 12. Number of tourists that went to the official-post of tourism searching for information about the city, in 2015, and its origin. Source: Visit Porto.

2015	Nº of tourists/visitors
Jan	7 071
Fev	9 046
Mar	10 562
Abr	20 021
Mai	28 269
Jun	41 906
Jul	58 973
Ago	98 017
Set	59 945
Out	54 972
Nov	31 553
Dez	31 987
TOTAL	**452 322**

2015		
Mercados	Total	%
Espanha	41323	31
França	36117	27
Portugal	9891	7
Alemanha	8197	6
Inglaterra	7286	5
Brasil	5314	4
Países Baixos (Holanda)	5161	4
Itália	4793	4
E.U.A	2786	2
Japão	1680	1

Annex 13. Top 30 of the attractions most photographed by tourists, between January 2014 and May 2016. Source: Own elaboration.

Attraction/activities	Nº of photos taken	R.F.	C.R.F
Douro River	319	11,03%	11,03%
NOS Primavera Sound	270	9,33%	20,36%
Luís I bridge	247	8,54%	29,69%
Praça Velasquez	155	5,36%	35,05%
Aliados	153	5,29%	40,34%
São Bento station	144	4,98%	45,32%
Porto	132	4,56%	49,88%
Clérigos Tower	128	4,42%	54,30%
Cais de Gaia	107	3,70%	58,00%
Ribeira	101	3,49%	61,49%
Avenida da Beira-mar	98	3,39%	64,88%
Elétrico	88	3,04%	67,92%
Dragon stadium	83	2,87%	70,79%
Bolhão	79	2,73%	73,52%
Lello & Irmão Bookstore	60	2,07%	75,60%
Sao Bento station	58	2,00%	77,60%
Rotunda da Boavista	58	2,00%	79,61%
Art	57	1,97%	81,58%
Foz	53	1,83%	83,41%
Casa da Música	52	1,80%	85,21%
Sé	52	1,80%	87,00%
Metro	52	1,80%	88,80%
Ocean	44	1,52%	90,32%
Cais da Ribeira	41	1,42%	91,74%
Igreja do Carmo	39	1,35%	93,09%
Gaia	39	1,35%	94,43%
Igreja dos Clérigos	38	1,31%	95,75%
Ponte da Arrábida	37	1,28%	97,03%
PortoAout2015	37	1,28%	98,31%
Food	36	1,24%	99,55%
Catedral da Sé	36	1,24%	100,80%

Note: R.F: Relative Frequency; C.R.F: Cumulative Relative Frequency.

Annex 14. Most frequent places and most frequent paths for a support level equals to 0,05. Source: Own elaboration.

Place	**Frequency (%)**
Igreja dos Clérigos	12,6
Praça da Liberdade Terreiro da Sé	9,3
Foz, Igreja de Santo Ildefonso, Igreja do Carmo and Sé	8,9
Palácio de Cristal and Vinho do Porto pictures related	8,5
São Bento square	7,7
Jardim da Cordoaria	7,3
Vila Nova de Gaia, Metro tours, Rua Santa Catarina and Teleférico de Gaia	6,9
Igreja de São Francisco and Rua das Flores	6,5
Caves Sandemanand Cais da Estiva	6,1
Catedral da Sé , Farol de Felgueiras, Photos taken to the ocean, Ponte D. Infante and Douro River – Rebelos	5,7
Estátua de Vímara Peres, Matosinhos, Igreja dos Congregados and Ponte da Arrábida	5,3

Path	**Frequency (%)**
Douro River → Cais de Gaia Douro River →Luís I bridge → Cais de Gaia	9,8
Aliados → Luís I bridge	9,3
Ponte D. Luis I → Cais de Gaia Cais de Gaia → Douro River	8,9
Luís I bridge → Porto landscapes Clérigos Tower → Luís I bridge Luís I bridge → Aliados	8,1
Douro River → Clérigos Tower Estação de São Bento→ Douro River Douro River → Estação de São Bento Luís I bridge → Estação de São Bento	7,7
Douro River → Luís I bridge → Douro River	7,3
Douro River → Estação de São Bento Cais da Ribeira → Douro River	6,9
Photos taken in Old Tram touris→ Luís I bridge Lello & Irmão Bookstore → Luís I bridge Clérigos Tower → Douro River → Luís I bridge	6,5
Aliados → Douro River Igreja dos Clérigos→ Douro River Estação de São Bento → Ribeira Douro River → Cais de Gaia → Luís I bridge Douro River → Igreja dos Clérigos	6,1

Douro River → Photos taken in Old Tram touris	
Clérigos Tower→ Cais de Gaia Old Tram touris→ Douro River Cais de Gaia → Douro River → Luís I bridge Porto landscapes → Douro River → Luís I bridge Estação de São Bento→ Douro River → Ponte D Luis Ponte D Luis →Ribeira → Ponte D Luis Douro River → Igreja do Carmo	5,7
Cais da Ribeira → Cais de Gaia Mosteiro Serra do Pilar → Luís I bridge Ribeira → Estação de São Bento Luís I bridge → Clérigos Tower Porto landscapes → Ribeira Luís I bridge→ Douro River → Ribeira Luís I bridge → Douro River → Cais de Gaia	5,3

Annex 15. Most frequent places and most frequent paths for a support level equals to 0,03. Source: Own elaboration.

Place	**Frequency (%)**
Estação de Campanha Serra do pilar Estátua Infante D. Henrique Photos related to food Photos related to art	4,9
Igreja de S. Lourenço Jardim do Palácio de Cristal Porto downtown	4,5
Caves Cálem Funicular dos Guindais Rua Mouzinho da Silveira Jardim do infante D. Henrique Camara Municipal do Porto Capela das Almas Avenida da Boavista Praia do Carneiro Rotunda da Boavista Rua 31 de Janeiro	4,1
Caves Vinho do Porto Rua da Alfandega Rua das Carmelitas Batalha Rua do Passeio Alegre Rua Sá da Bandeira Serralves Dragon stadium Fonte dos Leões Miradouro Ponte D. Luis I Igreja de S. Nicolau Paço Episcopal	3,7
Estátua D Pedro IV Praça da Batalha Douro River- cruzeiro Rua Campo dos Martires da Patria Rua Ribeira Negra Praça dos Leões Jardim do Morro Restaurante Miradouro da Vitoria Elevador da Vitoria	3,3

Paths	Frequency (%)
Clérigos Tower→Estação de São Bento Clérigos Tower→Porto lanscapes Clérigos Tower→Ribeira Clérigos Tower→Luís I bridge→Douro River Douro River→Foz Douro River→Praça da Liberdade Douro River→Porto lanscapes→Luís I bridge Douro River→Ribeira →Luís I bridge Douro River→Bolhão Douro River→Terreiro da Sé Igreja dos Clérigos→Douro River→Luís I bridge Aliados→Cais de Gaia Cais de Gaia→Cais da Ribeira Cais de Gaia→Luís I bridge→Douro River Porto lanscapes→Clérigos Tower Igreja de Santo Ildefonso→Douro River Praça da Liberdade→Douro River Luís I bridge→Cais de Gaia→Douro River Luís I bridge→Casa da Música Luís I bridge→Igreja dos Clérigos Ribeira →Porto lanscapes Estação de São Bento→Luís I bridge→Douro River	4,9
Lello & Irmão Bookstore→Douro River	4,8
Cais de Gaia→Estação de São Bento Igreja de Santo Ildefonso→Luís I bridge Praça da Liberdade→Luís I bridge Teleferico de Gaia→Luís I bridge Douro River→Miragaia Douro River→Lello & Irmão Bookstore Douro River→Aliados→Luís I bridge Douro River→Luís I bridge→Clérigos Tower Douro River→Estação de São Bento→Luís I bridge Ribeira –Gaia Ribeira →Luís I bridge→Ribeira Praça da Ribeira→Douro River Luís I bridge→Gaia Luís I bridge→Teleferico de Gaia Luís I bridge→Cais da Estiva Luís I bridge→Bolhão Aliados→Estação de São Bento Porto lanscapes→Luís I bridge→Porto lanscapes Clérigos Tower→Porto lanscapes→Douro River	4,5

Vinho do Porto photos related→Aliados Estação de São Bento→Clérigos Tower Estação de São Bento→Porto lanscapes Terreiro da Sé→Luís I bridge Sé→Luís I bridge	
Ribeira →Cais de Gaia Ribeira →Douro River→Luís I bridge Ribeira →Luís I bridge→Douro River Ribeira →Luís I bridge→Estação de São Bento Praça da Ribeira→Luís I bridge Estação de São Bento→Ribeira →Luís I bridge Porto lanscapes→Ribeira →Luís I bridge Porto lanscapes→Douro River→Porto lanscapes Douro River→Luís I bridge→Estação de São Bento Douro River→Cais de Gaia→Douro River→Luís I bridge Douro River→Luís I bridge→Cais da Ribeira Douro River→Luís I bridge→Ribeira Douro River→Igreja da Trindade Douro River→Porto lanscapes→Douro River Douro River→Caves Sandeman Caves Sandeman→Luís I bridge Luís I bridge→Douro River→Porto lanscapes Luís I bridge→Igreja do Carmo Aliados→Igreja dos Clérigos Aliados→Porto lanscapes Igreja de Santo Ildefonso→Porto lanscapes Igreja dos Clérigos→Ribeira Igreja das Carmelitas→Douro River Igreja das Carmelitas→Douro River Porto lanscapes→Luís I bridge→Douro River Photos taken in Old Tram touris → Porto lanscapes Porto lanscapes→Photos taken in Old Tram touris Terreiro da Sé→Cais de Gaia Terreiro da Sé→Douro River Clérigos Tower→Ribeira →Douro River Clérigos Tower→Douro River→Cais de Gaia Clérigos Tower→Luís I bridge→Cais de Gaia Clérigos Tower→Photos taken in Old Tram touris Vinho do Porto photos related→Douro River Vinho do Porto photos related→Luís I bridge Bolhão→Luís I bridge Praça da Ribeira→Luís I bridge	4,1
Estação de São Bento→Cais de Gaia Estação de São Bento→Igreja dos Clérigos Igreja das Carmelitas→Luís I bridge	3,7

Clérigos Tower→Porto lanscapes→Luís I bridge Clérigos Tower→Ribeira →Luís I bridge Serra do Pilar→Luís I bridge Douro River→Igreja dos Clérigos→Luís I bridge Douro River→Estatua Infante D. Henrique Douro River→Catedral da Sé Douro River→Luís I bridge→Porto lanscapes Douro River→Sé Douro River→Teleferico de Gaia Douro River→Luís I bridge→Aliados Douro River→Gaia Douro River→Igreja das Carmelitas Lello & Irmão Bookstore→Douro River→Luís I bridge Luís I bridge→Palácio da Justiça→Igreja dos Clérigos Luís I bridge→Igreja de S. Francisco Luís I bridge→Lello & Irmão Bookstore Luís I bridge→Caves Sandeman Luís I bridge→Cais de Gaia→Estação de São Bento Luís I bridge→Ribeira →Douro River Luís I bridge→Douro River rebelos Luís I bridge→Douro River→Clérigos Tower Luís I bridge→Vinho do Porto photos related Luís I bridge→Aliados→Ribeira Ribeira →Photos taken in Old Tram touris Cais da Ribeira→Estação de São Bento Cais de Gaia→Luís I bridge→Cais de Gaia Cais de Gaia→Cais da Ribeira→Douro River Ribeira →Clérigos Tower Cais da Ribeira→Douro River→Luís I bridge Praça da Ribeira→Estação de São Bento Clérigos Tower→Luís I bridge→Estação de São Bento Clérigos Tower→Ribeira →Estação de São Bento Clérigos Tower→Cais da Ribeira Clérigos Tower→Igreja dos Clérigos Estação de São Bento→Douro River→Ribeira Estação de São Bento→Luís I bridge→Ribeira Estação de São Bento→Praça da Liberdade Praça da Liberdade→Ribeira Aliados→Douro River→Luís I bridge Aliados→Praça da Liberdade Aliados.Clérigos Tower Aliados→Ribeira Aliados→Estação de São Bento→Luís I bridge Praça da Liberdade→Douro River→Luís I bridge Praça da Liberdade→Estação de São Bento	

Igreja do Carmo→Douro River Gaia→Luís I bridge Porto lanscapes→Cais de Gaia Porto lanscapes→Douro River→Cais de Gaia Porto lanscapes→Estação de São Bento Catedral da Sé→Douro River Clérigos→Douro River Palácio da Bolsa→Douro River Photos taken in Old Tram touris→Ribeira Lello & Irmão Bookstore→Ribeira Igreja das Carmelitas→Ribeira Catedral da Sé→Luís I bridge Casa da Música→Luís I bridge Rua Santa Catarina→Luís I bridge Palácio de Cristal→Luís I bridge Palácio da Justiça→Oceano Palácio da Bolsa→Luís I bridge	
Douro River→Cais da Ribeira→Luís I bridge Douro River→Igreja do Carmo→Luís I bridge Douro River→Igreja Santo António dos Congregados Douro River→Ponte D. Infante Douro River→Luís I bridge→Estação de São Bento→Luís I bridge Douro River→Luís I bridge→Igreja dos Clérigos Douro River→Luís I bridge→Igreja de S. Francisco Douro River→Luís I bridge→Photos taken in Old Tram touris Douro River→Ribeira →Photos taken in Old Tram touris Douro River→Rua Santa Catarina Douro River→Luís I bridge→Cais de Gaia→Douro River Douro River→Luís I bridge→Douro River→Clérigos Tower Douro River→ Rua Mouzinho da Silveira Douro River→Ribeira →Porto lanscapes Douro River→Cais de Gaia→Cais da Ribeira Luís I bridge→Oceano Luís I bridge→Igreja das Carmelitas Luís I bridge→Igreja da Trindade Luís I bridge→Foz Luís I bridge→Estatua Infante D. Henrique Luís I bridge→Douro River→Estação de São Bento Luís I bridge→Estação de Campanha Luís I bridge→Douro River→Aliados Luís I bridge→Aliados→Estação de São Bento Luís I bridge→Douro River→Photos taken in Old Tram touris Luís I bridge→Aliados→Palácio da Justiça Luís I bridge→Douro River→Cais de Gaia→Douro River Luís I bridge→Sé	3,3

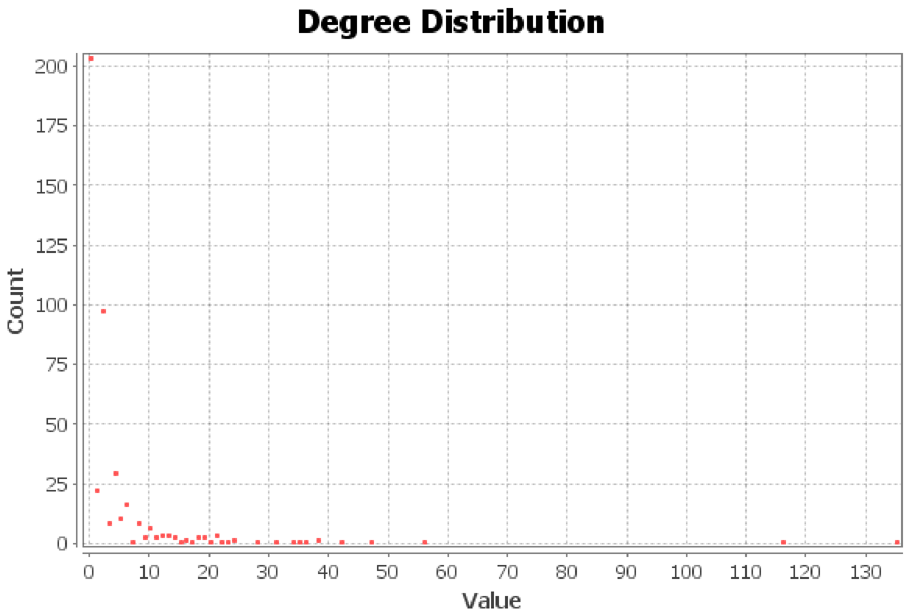
Luís I bridge→Douro River→Cais da Ribeira Palácio da Justiça→Jardim do Palácio de Cristal Palácio da Bolsa→Estação de São Bento Clérigos Tower→Porto lanscapes→Palácio da Justiça→Luís I bridge Clérigos Tower→Estação de São Bento→Luís I bridge Clérigos Tower→Igreja de S. Francisco Clérigos Tower→Cais de Gaia→Luís I bridge Clérigos Tower→Cais de Gaia→Douro River Clérigos Tower→Douro River→Luís I bridge→Douro River Clérigos Tower→Cais da Ribeira→Douro River Clérigos Tower→Terreiro da Sé→Douro River Clérigos Tower→Terreiro da Sé Palácio da Bolsa→Douro River→Luís I bridge Gaia→Ribeira Cais de Gaia→Igreja dos Clérigos Cais de Gaia→Luís I bridge→Estação de São Bento Cais de Gaia→Palácio da Justiça→Luís I bridge→Douro River Cais da Estiva→Luís I bridge Ribeira →Igreja dos Clérigos Ribeira →Cais de Gaia→Luís I bridge Praça da Ribeira→Ribeira Praça da Ribeira→Luís I bridge→Douro River Porto lanscapes→Igreja dos Clérigos Photos taken in Old Tram touris→Cais de Gaia Photos related to art→Luís I bridge Porto lanscapes→Photos taken in Old Tram touris→Douro River Photos taken in Old Tram touris→Clérigos Tower Estação de São Bento→Gaia Estação de São Bento→Ribeira →Douro River São Bento→Luís I bridge Igreja das Carmelitas→Estação de São Bento Igreja das Carmelitas→Douro River→Luís I bridge Igreja dos Clérigos→Estação de São Bento Igreja dos Clérigos→Porto lanscapes Igreja dos Clérigos→Luís I bridge→Douro River Igreja Santo António dos Congregados→Luís I bridge Igreja Santo António dos Congregados→Douro River Igreja de S. Francisco→Douro River Igreja da Trindade→Luís I bridge Igreja da Trindade→Douro River Aliados→Cais de Gaia→Luís I bridge Aliados→Vinho do Porto photos related Aliados→Luís I bridge→Douro River Lello & Irmão Bookstore→Estação de São Bento Lello & Irmão Bookstore→Porto lanscapes	

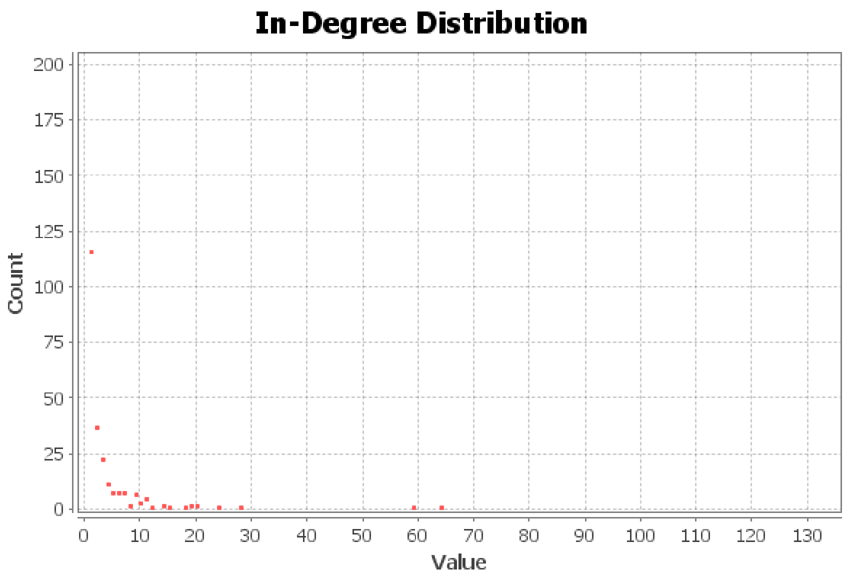
Terreiro da Sé→Estação de São Bento Terreiro da Sé→Cais da Ribeira Estátua Vimara Peres→Aliados Ocean→Porto lanscapes Ocean→Douro River Bolhão→Douro River Rua Santa Catarina→Douro River Palácio de Cristal→Douro River Palácio da Bolsa→Douro River→Luís I bridge Caves Cálem→Luís I bridge Ponte D Infante→Luís I bridge	

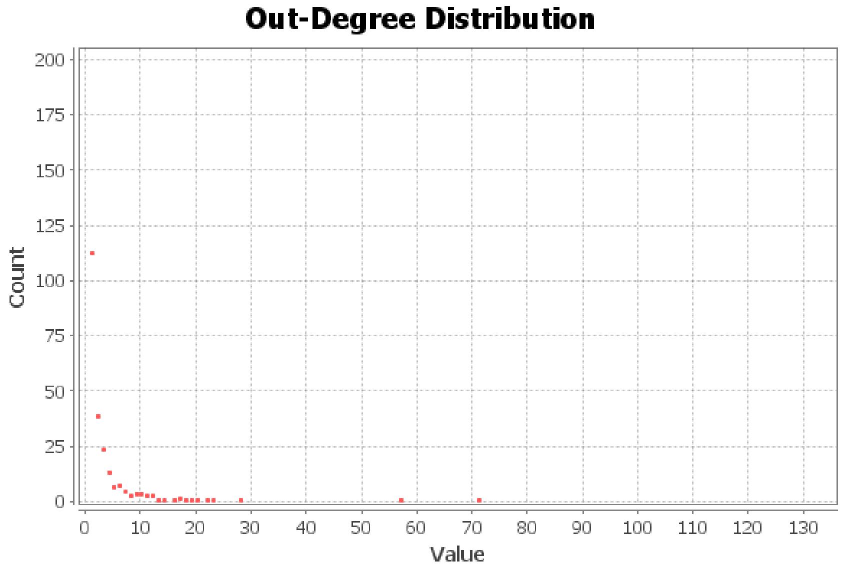
Annex 16. Central Metric results from Social Network Analysis. Source: Gephi software.

Results:

Average Degree: 4,118





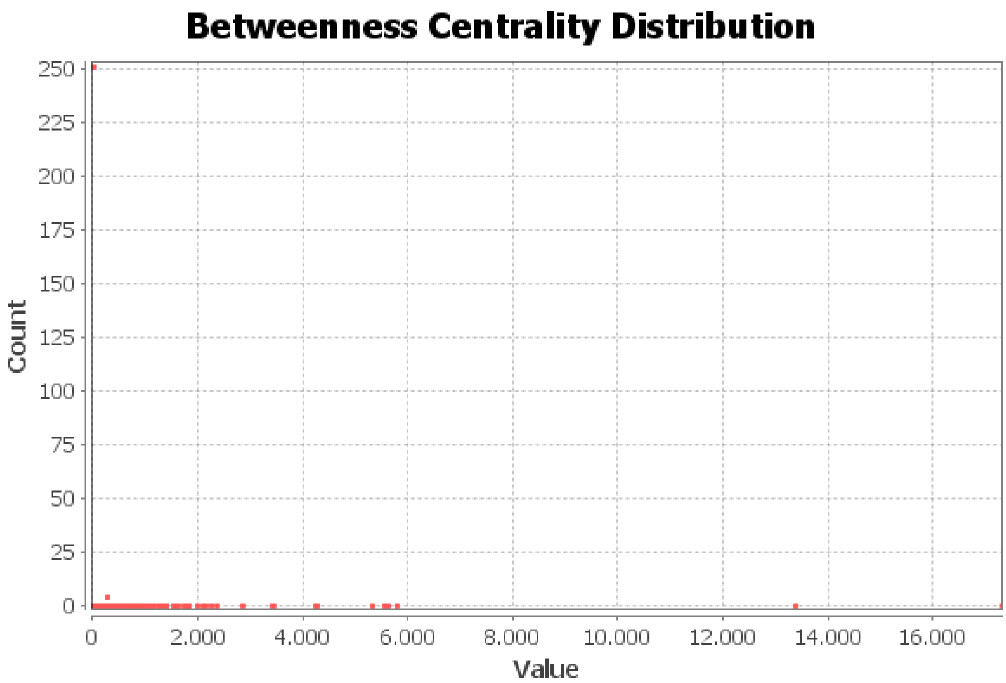


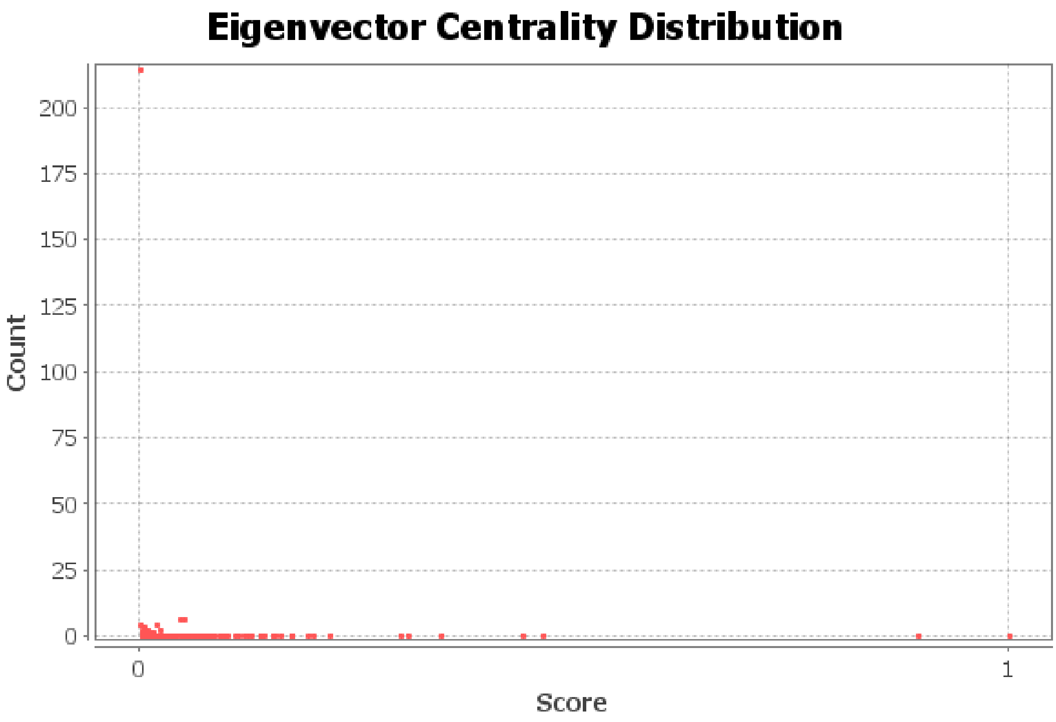
Results:

Diameter: 12

Radius: 0

Average Path length: 3.8570866845397678




Eigenvector Centrality Distribution
Count
200
175
150
125
100
75
50
25
0
0
1
Score