The mobility network of European tourists: a longitudinal study and a comparison with geo-located Twitter data

Davide Provenzano[a], Bartosz Hawelka[b], Rodolfo Baggio[c]

[a] Department of Economics, Statistics and Business, University of Palermo, Viale delle Scienze (Parco D'Orleans), Ed. 13, 90128 Palermo, Italy, E-mail: davide.provenzano@unipa.it

[b] Doctoral College GIScience, Department of Geoinformatics-Z_GIS, University of Salzburg, Schillerstraße 30, 5020 Salzburg, Austria. E-mail: bartosz.hawelka@sbg.ac.at

[c] Master in Economics and Tourism, Bocconi University, via Röntgen 1, 20136 Milan, Italy and National Research Tomsk Polytechnic University, 30 Lenin Avenue, Tomsk 634050, Russian Federation. E-mail: rodolfo.baggio@unibocconi.it

==========

Tourism Review (2017), forthcoming

==========

Abstract

In this paper we provide a network study of the structural and dynamical characteristics of tourism flows in Europe in the period 1995-2012. Travels in Europe were studied following the network science research paradigm and focusing on the whole network of intra-European tourism destinations. Network analysis was used to map and reveal the pattern of connections between states as shaped by bilateral tourism flows.

Data were provided by the United Nations World Tourism Organization. For several countries, we updated and integrated UNWTO data with information available from the national statistical office of the individual countries.

For the year 2012, results obtained from the UNWTO record-based network were compared to geo-located Twitter data as a proxy of human mobility patterns.

Our analysis provides evidence of a shift towards an increased homogeneity in the travelling preferences of European tourists, an acquired attitude of visitors to travel shorter distances, and a tendency of mobility patterns to merge. Finally, the comparison between UNWTO and Twitter data shows a different spatial distribution of visitors.

These results provide a useful insight for policy makers involved in tourism planning.

Keywords: European tourists; Mobility patterns; Networks; UNWTO; Geo-located Twitter data

1. Introduction

The EU tourism industry has a relevant economic and social importance for its contribution to GDP, employment, and economic growth, as well as to development in rural, peripheral, or less developed areas (European Commission, 2010). Starting from the second half of 2010, EU tourism has gained momentum, consolidating the position of Europe as the most visited region in the world (see UNWTO World Tourism Barometer, January 2015). Typically, three out of four outbound trips of at least one overnight stay made by EU residents have a main destination in another member state.

Despite the large amount of data available for tourism analyses, what official statistics and literature in general fail to provide is a study, over long time intervals, of tourism flows between all European countries.

In this study we aim at filling this gap by investigating the structural properties of the relations between tourism destinations along with the dynamical (historical) evolution of the topology of intra-European tourism flows. Moreover, by comparing results from official statistics with the analysis based on geo-located Twitter data, we also contributed to the discussion about the use of social media platforms as a global source for mobility data.

The flow of intra-European overnight visitors from 1995 to 2012 is the research topic of this study, which includes all the sovereign states falling even partially under any of the common definitions of Europe, geographical or political, for a total of 54 entries. Travels in Europe were studied following the network science research paradigm (Baggio *et al*., 2010) and focusing on the whole network of intra-European tourism destinations. We used data provided by the United Nations World Tourism Organization (UNWTO, 2014) in conjunction, in several cases, with more complete and updated data found on the web page of the national statistical office of the individual countries. Network analysis was used to map and reveal the pattern of connections between states as shaped by bilateral tourism flows.

Network attributes were quantitatively measured and results displayed in network diagrams. Additionally, we used user-generated geo-located Twitter data as a proxy of human mobility patterns between selected countries. We created a similar country-to-country network and cross-validated the results obtained from Twitter data with the network based on UNWTO records for the last year of the analysis: 2012.

According to our results, tourism flows in Europe have changed during the period covered by the study, displaying a trend towards a higher homogeneity. At the same time, mobility patterns displayed a tendency to merge and an acquired attitude of visitors to travel shorter distances (the majority of trips took place among neighbouring or nearby countries).

These results provide a useful insight for policy makers involved in tourism planning. In fact, the degree of homogeneity represents an important indicator of tourism services quality and tourism management in general (Bosetti *et al*., 2007), whereas mobility is an essential strategic component for sustainable travel and tourism.

The contribution of this study is threefold. First of all, to the best of our knowledge, this is the only study that focuses on the bilateral tourism flows between all countries falling, geographically or politically, under the definition of Europe. Secondly, we provide evidence of a shift towards a greater homogeneity in the travelling patterns of European tourists. Lastly, for the first time our study provides a comparison between topological structure and bilateral mobility patterns of tourism flows, based on two different data recording methods.

In the following section we provide an overview of the existing literature dealing with the analysis of tourism flows in Europe and the use of network science in tourism studies. Section 3 describes data and methods used in this study. Results are discussed in section 4. Managerial implications, concluding remarks, limitations of the study, and possible future research topics close the paper.

2. Literature background

The issue of visitors' flows between European countries has so far been addressed only by a limited number of articles and reports.

Pearce (1987) focuses on the European country-to-country patterns of charter tourists between 1970 and 1980. The analysis of data relating to members of the European Civil Aviation Conference (ECAC) reveals that inclusive tours by charter (ITC) were responsible for much of the expansion of tourism in post-war Europe and for a more direct interaction between market and destinations in a north-south pattern of chartered traffic (first in Spain, later in Greece, Yugoslavia, Portugal, and Malta).

An early study of the scale and development of tourist areas within twelve European Community (EC) countries dates back to 1991 and can be found in a report by the Netherlands Scientific Council for Government Policy (NRIT, 1991). The study provides a picture of the overall tourism demand in the EC, together with the main origin-destination flow patterns and likely tourism trends.

In Jansen-Verbeke and Spee (1995) the analysis of inter-regional and intra-regional tourism flows reveals the geographical pattern of tourism-destination and tourism-generating areas within Europe. The results of the study indicate some relevant features of the European regional tourism market. In particular, while some regions are largely dependent on a geographically limited market area, others show a clear move towards a more international pattern of interaction. Coshall (2000) applies spectral analysis, both univariate and bivariate, to study and describe international tourism flows from 1975 to 1996. In particular, bivariate spectral analysis is applied to UK passenger travelling to the United States and Canada, and to shorter-haul passenger flows to three northwest European countries (Belgium, France, and the Netherlands). "Significant leading relationships are found between the various exchange rates and passenger movements" only for "UK air travel to the United States and Canada, and typically for UK sea travel to the European countries" (Coshall, 2000, p. 579).

Some information on the flow of European tourists in the Czech Republic, Hungary, Poland, and Slovakia during the post-1989 transition and up to their EU accession in 2004 can be found in Baláž and Williams (2005). The study reveals a "considerable continuity and path dependency in tourism in the transition period" (Baláž & Williams, 2005, p. 88). In Marrocu and Paci (2011), the flow of tourists between a sample of 199 European regions belonging to member states of the EU15, plus Switzerland and Norway, is taken as a main explanatory variable in the analysis of regional production efficiency. Over the period 2002-2004, "the [proposed] empirical models show that tourism flows generate a positive and significant effect on the regional level of production efficiency" (Marrocu & Paci, 2011, p. 756). Harjaa and Stângaciu (2013) explore the development of tourism in the EU27 in the 2007-2011 time interval, focusing in particular on the position of Romania among member states in the fifth year after EU accession. The study argues that, despite the economic crisis, tourism showed positive dynamics as a whole and in many member states except for Romania that, after accession, lost attraction for both domestic and foreign tourists.

The number and frequency of trips made by Russian citizens to EU countries are studied by Furmanov *et al*., (2012). This quantitative analysis shows a steady upward trend and a lack of evident correlations between tourism flows and some traditional statistically-measurable factors: the real money income, the relative consumer price index, and the exchange rate between the currency of the destination country and the Russian rouble.

In terms of research methods, for a long time, the literature about tourism flows has been dominated by the tourism-demand model, considered the appropriate framework for studying the tourist mobility between two or more pairs of countries (Askari, 1971; Barry and O'Hagan, 1972; Crouch, 1994; Lim, 1997; Morley, 1998; Sinclair, 1998; Witt, Witt and Wilson, 1994).

Focussing primarily on how changes in the national income, in relative prices, transportation costs, and exchange rates might affect tourism flows between origin and destination countries, the demand model provides a short run forecasting tool to estimate the demand for a destination country from its main markets (Zhang and Jensen, 2007). Yet, this model treats all destinations as undifferentiated and ignores their stage of development. That is why a different strand of literature has focused on the supply-side perspective to study tourism flows. Several supply-side related factors (such as quality, resources, destination environment, infrastructure, and value) can actually influence tourist motivation and satisfaction as well as the tourist's decision to visit a destination and her intention to return (Murphy *et al*., 2000; Melián-González and García-Falcón, 2003; Beerli and Martín, 2004; Yoon and Uysal, 2005, and references therein). However, because of the difficulties in obtaining relevant data and good proxies for supply-side factors, this investigation approach has less contributed to the literature of tourism flows, than the more traditional tourism-demand studies, especially in terms of more quantitative style studies.

By drawing on both the quantitative and qualitative paradigm, mixed-methods have proved extremely suitable to conceptualise, visualise and analyse the peculiar set of formal and informal (social) relationships that shape the tourism industry. In this sense, the use of network analytic techniques in tourism research appears logical and delivers a number of useful outcomes for the study of tourism destinations and organizations (Scott *et al*., 2008). In spite of this, tourism literature includes a limited amount of works that examine a tourism destination from a network point of view, as well as few studies using the quantitative methods of network science (see a review in van der Zee & Vanneste, 2015). In this research approach, the most important study is by Miguéns and Mendes (2008) where it is illustrated how strategic positioning benefits from market diversity and how movements among countries relate to technological and economic patterns. The authors finally provide some insights into the forces driving international travels.

All the studies about tourism mainly use official records provided by the different national statistical organizations (typically collected by UNWTO). The reliability of these data has been questioned many times (e.g. Volo, 2004) because of the poor harmonization of the data collection methods, the currency of the data and the statistical estimation procedures used (Lam & McKercher, 2013). Moreover, with the growth of multiple forms of travel and stay, many visitors or tourists go unobserved (De Cantis *et al*., 2015).

A possible way to improve data quality is resorting to the records of the innumerable trails that millions of individuals leave online using the many currently available technological platforms. Indeed, such technological footprints have recently shown to be an interesting source to assess and better measure the extent of the movements of travellers, tourists, and visitors (Lu & Stepchenkova, 2015). The idea has already been put forward by several studies that have shown how the contents produced by large masses of individuals can provide interesting insights into their beliefs, behaviours or preferences (Abbasi *et al*., 2012).

Here we use, at least partially, the data and results of Hawelka *et al*. (2014) that report a wider study of global mobility patterns as derived from the analysis of a large set of tweets recorded in 2012. In particular, we compare the network derived from geo-located Twitter messages with the one built using official UNWTO data on international tourist flows.

A few studies have warned that using social media population as a representative sample may introduce several kind of biases: sampling bias and location biases more than any other (Mislove *et al*., 2011; Hecht and Stephens, 2014; Longley *et al.*, 2015; Malik *et al.*, 2015; Pavalanathan and Eisenstein, 2015). Yet, sampling bias is likely to be prevalent for any technology that captures mobility dynamics (Yan *et al.*, (2013); Wesolowski *et al.*, 2013) in the same way as the so called official statistics exhibit problems of accuracy and comparability (Volo, 2004). Moreover, a different strand of literature has proved solid evidence that Twitter can indeed be a useful proxy for tracking and predicting human mobility patterns, within and between cities (Hawelka *et al.*, 2014; Jurdak *et al.*, 2015). In particular, Hawelka *et al.* (2014), whose dataset is used in this study, showed the usefulness of geo-located Twitter data as a proxy for country-to-country tourist/visitor flows. On the supportive results of these studies we rely our analysis.

3. Materials and Methods

Data used in this study were provided by the United Nations World Tourism Organization (UNWTO, 2014) and the national statistical offices of individual countries (data collection was completed in June 2015). We limited our review to data on visits on a country-by-country basis, including those with at least one overnight stay and omitting day trips. Moreover, we used the dataset of one full year of Twitter messages with global coverage, collected by Hawelka *et al*. (2014).

Twitter represents one of the largest contributor in the "big data phenomenon" and has become a useful data source for addressing challenging business and societal problems (Bruns and Stieglitz, 2013). In fact, the Twitter dataset is an example of a new type of human mobility data, generated bottom-up, collecting digital footprints of individual social web users, particularly the Twitter micro-blogging platform (www.twitter.com). Of all messages, those explicitly tagged with geographic coordinates (geo-tagged) were considered. For the case under study, the dataset consisted of almost a billion tweets generated by a total of 13 million users in 2012. The stream was gathered through the Twitter Streaming API. Once cleaned of consecutive locations for a single user that implied a relocating speed of over 1000 km/h, i.e. faster than a passenger plane, the dataset was further filtered out of tweeting noise, namely data generated by Twitter services as web advertising (e.g., tweetmyjob), web gaming (e.g., map-game), or web reporting (e.g., sandaysoft), which do not reflect human physical presence in either the reported place or time." In total, the refinement procedure preserved 98% of users and 95% of tweets from the initial database." (Hawelka et al. ,2014).

Most importantly for the purpose of our study, the sequence of a user's locations unlocked the possibility for us to assign each user to her country of residence. Knowing Twitter 'residencies' we could build a European country-to-country network of tourist flows, similarly as the UNWTO data.

Our list of sovereign states includes 54 entities, falling even partially under any of the common definitions of Europe, geographical or political. In addition to the current 28 EU member states, our study encompasses EFTA countries, post-soviet republics and the former states of Yugoslavia, the republics of Albania, Israel, San Marino, and Turkey, the principalities of Andorra and Monaco, and the Holy See. Due to the common practice of several national statistics to collect incoming tourists from Serbia and Montenegro up to 2005 in the same record, we considered the two states as an individual origin/destination country. No data were available for the Republic of Kosovo, nor were data found for incoming tourism in Uzbekistan and the Holy See. Data for Faeroe Islands and Gibraltar were incomplete, thus omitted in the study. These exclusions did not influence the overall results of the study (the final network layout is about 99% complete).

After cleaning the data, we selected years 1995, 2000, 2005, 2010, and 2012 for the comparisons. This choice was driven by the consideration that, with high probability, differences in patterns take some time to build, and a five-year interval would be a reasonable timeframe to highlight significant variations, if any.

Due to data availability the comparison with Twitter data was performed only for 2012.

For each selected year a tourist flow network was built in which the countries were the nodes and the links represented the flow of visitors between the different countries weighted by their number. The same was done for the Twitter data.

The resulting networks (weighted and directed) were analysed using the functions provided by the freely-available Python NetworkX (Hagberg *et al.*, 2008) and igraph (Csardi & Nepusz, 2006) libraries.

Many possible metrics can be used to describe and characterise a network; we decided to resort to the most widely employed (da Fontoura Costa *et al.*, 2007; Newman, 2010):

- *order:* total number of nodes in the network;

- *size:* total number of links in the network;
- *nodal degree and degree distribution*: number of connections, *k*, a single node has in the network (incoming and outgoing if the network is directed) and statistical distribution of the degrees. The term 'strength' is used when links are weighted;
- *density*: ratio between size of the network and maximum number of links it may have. No common definition for a weighted density exists, therefore we only report an unweighted value;
- *path:* series of consecutive links connecting any two nodes in the network (Baggio *et al*., 2010);
- *path lengths (distance) and diameter (longest shortest path):* shortest (weighted) 'itinerary' between any two nodes in the network and its globally maximum value (diameter), i.e. the largest number of nodes that must be crossed to travel from one node to another when paths which backtrack, detour, or loop are not considered;
- *assortativity:* "A network is said to be assortative when high degree nodes are, on average, connected to other nodes with high degree and low degree nodes are, on average, connected to other nodes with low degree". "Assortativity provides information about the structure of a network, but also about dynamic behaviour and robustness of the network" (Noldus & Van Mieghem, 2015, p. 1). It is defined as the Pearson correlation coefficient of degrees between pairs of linked nodes (Newman, 2002). "Assortativity is expressed as a scalar value, ρ, in the range $-1 \leq \rho \leq 1$" (Noldus & Van Mieghem, 2015, p. 1);
- *betweenness*: number of shortest paths from all nodes to all others passing through a node. High betweenness has a strong influence on the flows through the network, assuming they follow the shortest paths; high betweenness also indicates a possible bottleneck for the network flows.

For all these metrics, the mean values summarise the global properties of the network, while individual values characterise the local features.

In directed network, for any two connected nodes *i* and *j*, the link from *i* to *j* is generally different from the link from *j* to *i*. Thus, on a directed network nodes have an in- and an out-degree. The in-degree of a node is the number of links coming into it; the out-degree is the number of links going out of it. Networks may also be weighted networks when a weight (strength) is assigned to each link, measuring how good or strong a relationship is between two connected nodes.

In many cases, networked systems may exhibit some form of substructure, whose study can be central to understand the organisation and evolution of their modular structure. In these networks, the distribution of connections is not only globally, but also locally inhomogeneous, with high concentrations of links within special clusters of nodes. To some extent, communities (or modules) can be considered separate entities with their own autonomy.

Many different methods and numerical algorithms have been suggested to identify topological similarities in the local configurations of links (Fortunato, 2010). In all of them, a modularity index Q is used to gauge the clustering structure of the network. The Q is calculated as "the fraction of all links that lie within a community minus the expected value of the same quantity that could be found in a graph having nodes with the same degrees but links placed at random. The index is always smaller than one, and can be negative when the network has no community structure, or when a subgroup has less internal links than towards the other groups" (Baggio, 2011, p. 184). In order to facilitate the comparisons between networks with different number of communities, the modularity index can be normalised by the number m of modules (Du *et al.*, 2009): $Q_{norm} = mQ/(m-1)$.

For the networks studied herein modules were obtained using the algorithm proposed by Sobolevsky *et al.* (2014). The modules obtained for the networks were then compared graphically by an online applet, which allowed us to draw an alluvial diagram (Edler & Rosvall, 2013).

The alluvial diagram is a sort of flow diagram used to visualise the evolution of a network structure over time or the differences (if any) between the modular compositions of the compared networks. In an alluvial diagram, the blocks represent clusters of nodes ordered by size from bottom to top, and the height of each block reflects the volume of flow through the cluster. Stream fields reveal the changes in composition of the blocks over time, and the height of a stream field features the total size of the nodes contained in the blocks connected (Rosvall & Bergstrom, 2010).

4. Results and discussion

We studied the structural properties of the networks obtained from the official UNWTO data and their evolution through the years selected, then we compared the main metrics calculated for the two networks derived for year 2012 from the UNWTO and Twitter data, respectively.

When possible and meaningful (see Barrat *et al.*, 2004) the weighted values were derived. Individual significance values were tested running all the algorithms averaging on ten realisations of a randomly rewired version of the analysed networks (as suggested by Guimerà *et al.*, 2004).

All values were found significant at least at the 0.01 level (individual significance values are not reported to avoid cluttering data presentation).

4.1 Networks of UNWTO tourism flows

Table 1 shows the main metrics calculated for the network structures obtained from UNWTO data. The network of European bilateral tourism flows appears quite dense in all years reviewed, considering that the density values found in the literature for the social networks studied are typically of the order of 10^{-1} – 10^{-2} (Albert & Barbási, 2002; Boccaletti *et al.*, 2006; Caldarelli, 2007; Newman, 2010). This confirms the economic and social importance of tourism for the EU member states.

The increasing number of links (and density) in the studied networks over the selected years, along with a slight decrease in the average path length, well conforms to the increase of tourism reflected in the growth figures published by UNWTO for the period under analysis. Indeed, a high number of links does result in a network of connections that facilitate tourism flows. Networks also appear quite compact (low average path length and diameter). The low values of assortativity indicate a low correlation between nodal degrees, denoting no strongly ‘preferred’ paths in the flows of international tourists. All networks show a high heterogeneity of the nodal degrees with distributions characterised by clear power-law tails $P(k) \sim k^{-\alpha}$ (the exponent and its standard error were calculated following the procedure proposed by Clauset *et al.*, 2009).

Table 1. Main metrics calculated for the networks derived from UNWTO data

Network metric	Year: 1995	2000	2005	2010	2012
Order	54	54	54	54	54
Size (unweighted)	993	1495	1691	1767	1770
Density (unweighted)	0.347	0.522	0.591	0.617	0.618
Average path length (unweighted)	1.064	1.317	1.316	1.320	1.278
Diameter (unweighted)	3	3	3	3	3
Assortativity (weighted)	0.014	0.030	0.030	0.000	-0.017
Betweenness (average, weighted)	0.034	0.046	0.048	0.048	4.1E-05
Weighted degree distribution (exponent)					
in-degree	2.10±0.25	2.17±0.27	2.84±0.58	1.89±0.18	1.99±0.19
out-degree	1.88±0.17	1.74±0.13	1.66±0.10	2.00±0.19	1.97±0.19
Weighted degree heterogeneity (σ/μ)					
in-degree	2.016	1.897	1.749	1.691	1.629
out-degree	2.256	2.062	1.883	1.680	1.639
Weighted degree Gini index					
in-degree	0.773	0.747	0.711	0.708	0.692
out-degree	0.776	0.738	0.716	0.683	0.674
Modularity (weighted)					
no. of communities	5	5	5	4	4
Q	0.187	0.270	0.275	0.264	0.282
Q_{norm}	0.234	0.338	0.344	0.352	0.376

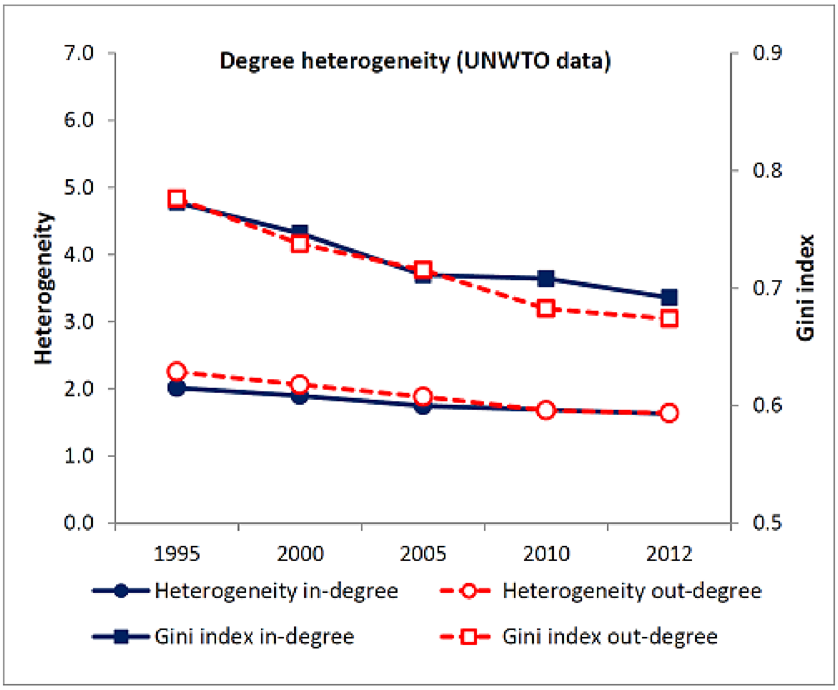

Figure 1. Heterogeneity in the degrees of the networks

The scaling behaviour shown by tourism flows is an important element to understand the self-organization of human travelling patterns, as discussed by Miguéns and Mendes (2008) with reference to worldwide tourism arrivals. More precisely, a certain asymmetry in the degree distributions proves that a few countries are responsible for the majority of movements.

Tails of both in- and out-degree distributions have a power-law functional form, but the in-degree distribution decays faster than the out-degree distribution and crosses it at a consistent nodal degree. This implies that, as link degree increases, the in-flows grow slightly faster than out-flows, indicating that countries 'export' more visitors than they receive This phenomenon, however, decreases with time and the last years show very little difference between the two distributions. If we couple this result with the decreasing heterogeneity of the networks degrees, both in and out, (Figure 1) we can conclude that, in general, there is a tendency towards a more even distribution of the travels across the countries examined. This conclusion is also confirmed by the decrease in the average betweenness, which can be read as a decrease in the probability to find bottlenecks in the overall movement of travellers.

The concentration of connections in the neighbourhood of a node was measured (Hu & Wang, 2008) by the coefficient of variation σ/μ and the Gini index (a well-known measure of statistical dispersion). Similarly to the networks degrees, local link density shows a decreasing heterogeneity over time and a very little disparity between in-degree and out-degree confirming a more homogeneous distribution of travels over the selected countries.

The (weighted) modularity analysis by the Combo algorithm points towards the same conclusions. Indeed, communities uncovered show a limited degree of separation (typically less than 0.5), namely, the presence of dense connections between the different travelling areas.

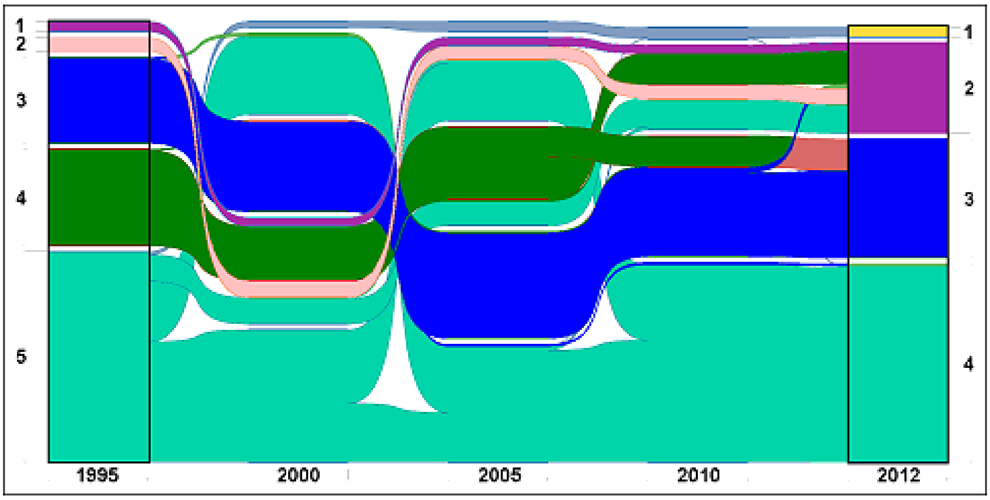

Figure 2. The evolution of communities in the UNWTO networks

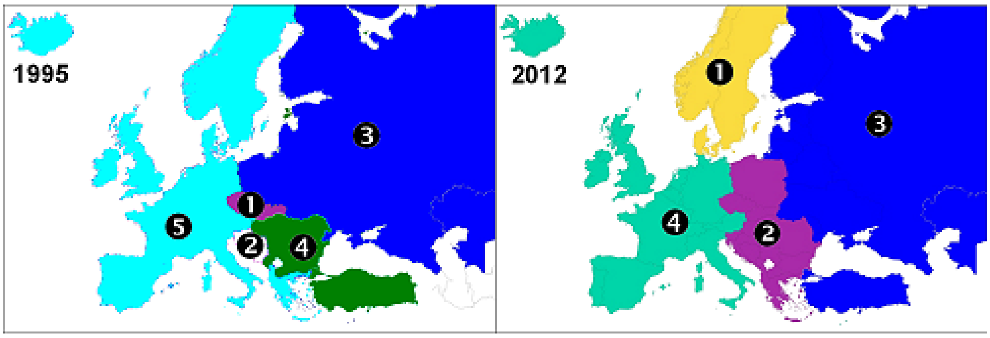

Figure 3. Clusters of mobility patterns in Europe for the years 1995 and 2012

In terms of mobility patterns, the alluvial diagram (Figure 2) clearly shows an evolution towards a higher concentration in a lower number of areas and, more interestingly, the origin in 2012 of a new travelling path (the second block) resulting from the diversion of the previous clusters of bilateral tourism flows (also shown in Figure 3).

The growing impact of modern information and communication technologies (Lu & Stepchenkova, 2015), the effects of EU enlargement (Leidner, 2007), the improvements in passenger transport infrastructures, and the increased volume of low-cost travels (Budd et al., 2014) could all be possible explanations of trends and travelling paths highlighted. However, a discussion about how these factors influence tourism is beyond the scope of this work and is left for future research.

4.2 Comparison with geo-coded Twitter data

The availability of a collection of Twitter geo-coded messages (Hawelka *et al.*, 2014) for year 2012 allowed us, for the first time, to compare the topological structures and the bilateral mobility patterns of tourism flows relying on the UNWTO and Twitter data recording methods. Table 2 shows the main results of our analysis. The comparison clearly reveals a higher number of connections and a higher density for the Twitter network structure confirming, however, the results of the analysis carried out with the UNWTO data. The average path length attests a certain density of the network.

Another metric confirmed in its sign is the assortativity coefficient. Yet the network structure based on Twitter data looks more disassortative than the UNWTO one. This, as already discussed, suggests a tendency towards no strongly ‘preferred’ paths in the flows of European tourists.

Table 2. Comparison between UNWTO and Twitter networks for 2012

	UNWTO	**Twitter**
Node count	54	54
Link count (unweighted)	1770	2183
Density (unweighted)	0.618	0.763
Average path length (unweighted)	1.278	1.253
Diameter (unweighted)	3	3
Assortativity (weighted)	-0.017	-0.075
Betweenness (average, weighted)	4.1E-05	0.029
Weighted degree distribution (exponent)		
in-degree	1.99+0.19	2.21+0.29
out-degree	1.97+0.19	1.69+0.14
Weighted degree Heterogeneity (σ/μ)		
in-degree	1.629	1.465
out-degree	1.639	1.921
Weighted degree Gini index		
in-degree	0.692	0.664
out-degree	0.674	0.768
Modularity (weighted)		
no. of communities	4	4
Q	0.282	0.162
Q_{norm}	0.376	0.215

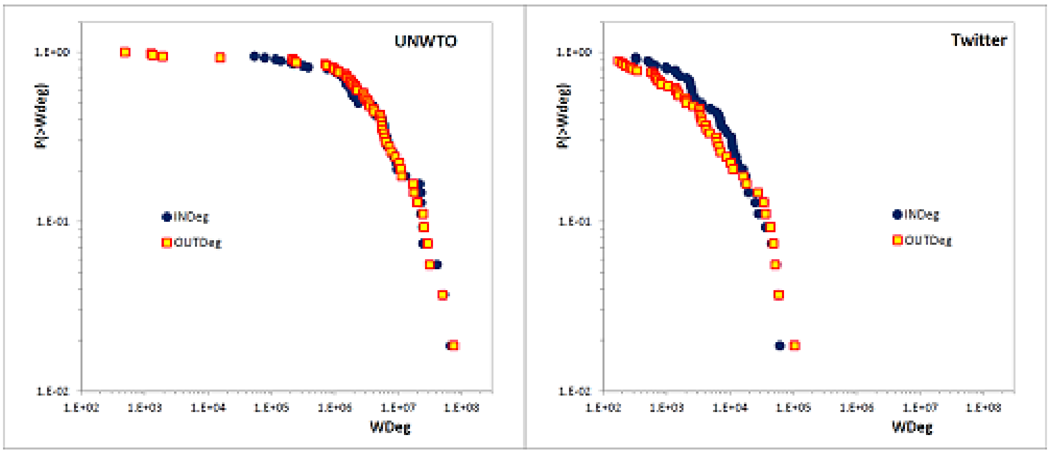

Figure 4. Cumulative degree distributions for the UNWTO and Twitter networks

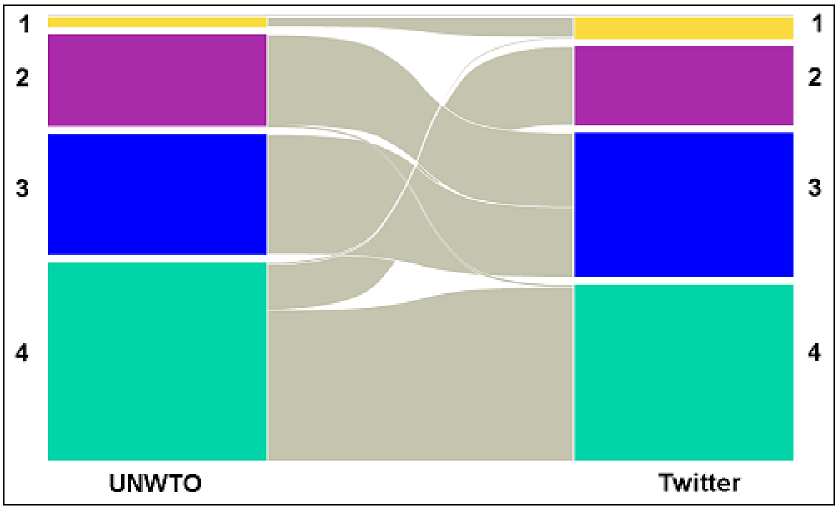

Figure 5. Communities in the UNWTO and Twitter networks

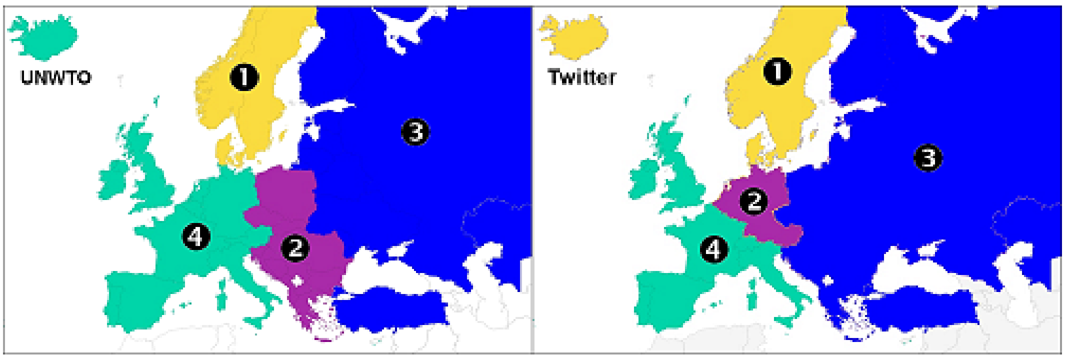

Figure 6. Comparison of mobility patterns from UNWTO and Twitter data for the year 2012

The Twitter network appears more skewed (Figure 4). This is also confirmed by the higher values of the heterogeneity measures and by the lower modularity.

The modularity analysis also shows a marked difference. Although the number of communities is the same, their composition differs, highlighting different mobility patterns and a different ‘concentration’ of tourists in the region considered (Figure 5 and Figure 6)

Surprisingly, Twitter data reveals a cluster of mobility patterns originating from the fourth cluster of UNWTO data that the UNWTO recording methods completely fail to document (block 2 of Twitter data). Figure 5 also shows that the second and third block of UNWTO data merge into a unique community of tourism destinations according to geo-locater Twitter records.

In other words the two networks, and hence the two types of observations, have a marked and clear structural dissimilarity which is confirmed by a Kolmogorov-Smirnov test that finds a highly significant ($p<10^{-4}$) difference between the two degree distributions. Therefore it is reasonable to conclude that there is a difference between what we are able to count by using traditional methods and what we may infer (although with the cautions reported above) from the *real* movements of international travellers.

At this stage it is difficult to understand what the reasons for this difference could be. With high probability one cause could be that unobserved tourism, that is those flows that go unaccounted for by the official statistics due, for example, to non-traditional choices of accommodation that has been growing substantially in the last years (see e.g. De Cantis et al., 2015).

What is clear, however, is that new methods and techniques must be devised in order to combine these different series of data for arriving at a better and more precise characterisation of the tourism phenomenon. New methods on which many are working (e.g. Aragona et al., 2016 or Struijs et al., 2014), but that still need a good consideration.

5. Managerial implications and concluding remarks

In this paper we considered European countries as a complex system of tourism destinations. A network analysis was carried out in order to understand the behaviour of Europeans in terms of their preferred tourism destinations.

Besides a general description of the structural characteristics of the network presented, we made a comparison between the network structures resulting from the flow of tourists as recorded in the UNWTO statistics and by the Twitter geo-located messages.
The analysis of UNWTO data revealed that tourism flows within Europe changed over the time horizon covered by this study displaying a trend towards a higher homogeneity. At the same time, mobility patterns displayed a tendency to merge and an acquired attitude of visitors to travel shorter distances (the majority of trips took place among neighbouring or nearby countries).
These results offer useful information to policy makers involved in the management of tourism destinations, provided that the degree of homogeneity represents an important indicator of the quality of tourism services and the effectiveness of tourism management in general (Bosetti *et al*., 2007). In practical terms, as tourism destinations attract visitors from different cultures and countries, our study emphasises the new challenge for tourism managers to offer a heterogeneous product when tourism is becoming more homogeneous. In fact, less strong travelling preferences of tourists raise the importance of product and services consumption at the tourism destination as a strategic variable for improving the holiday experience of vacationers.
Holiday destinations and travel mode choices are closely inter-related. The study of tourism mobility patterns becomes therefore an essential aspect of destination management for an efficient and sustainable tourism mobility. In this regard, our study revealed the Europeans' general preference for crossing into nearby countries or travelling in the neighbourhood of international borders. Managers should therefore implement effective strategies to promote the tourism destination image mainly in contiguous countries and consider investing in the ground transportation systems.
Finally, the comparison between the UNWTO and the Twitter data sets stressed the difference between what information can be attained from official data and what international travellers actually do. Our research showed the two measurements have a considerable overlapping degree. Yet, we obtained different results for the spatial distribution of visitors across Europe as reported by UNWTO vs. Twitter data sets. Such comparison definitely demonstrated that statistics based on official data are not able to take into account many of the features that characterise modern tourism.

Twitter is a widely used social media platform, but it is only one of many good tools people use regularly (a reasonable estimate is of about 70% of the 3 billion users reported by www.internetworldstats.com). The diffusion of smartphones running location tracking applications make these devices increasingly accurate and reliable sources for collecting data on people's actual movements. Thus, in order to depict and quantitatively analyse the actual spatial behaviour of tourists and obtain useful insights and trends for planning activities, scholars and practitioners must become familiar with the most recent data collection methods. New reliable methods for combining official and empirical measurements should also be found.

The contribution of this study is threefold. First of all, to the best of our knowledge, this is the only study that applies network analytic methods to the bilateral tourism flows between all countries falling, geographically or politically, under the definition of Europe. Secondly, we provide evidence of a shift towards a higher homogeneity in the travelling preferences of European tourists. Lastly, for the first time our study provides a comparison between topological structure and bilateral mobility patterns of tourism flows, based on two different data recording methods.

6. References

Abbasi, M.A., Chai, S.K., Liu, H., and Sagoo, K. (2012), "Real-world behavior analysis through a social media lens", in Yang, S.J., Greenberg, A.M., and Endsley, M. (Eds.), *Social Computing, Behavioral-Cultural Modeling, & Prediction proceedings of SBP12, University of Maryland, College Park, MD, 3-5 April* 2012, Heidelberg: Springer, pp. 18-26.

Albert, R., and Barabási, A. (2002), "Statistical Mechanics of Complex Networks", *Review of Modern Physics*, Vol. 74, pp. 47-91.

Aragona, B., and Zindato, D. (2016). "Counting people in the data revolution era: challenges and opportunities for population censuses", *International Review of Sociology*. Vol. 26, pp. 367-385.

Askari, H. (1971) "Demand for Package Tours", *Journal of Transport Economics and Policy*, Vol. 5, pp. 41-51.

Barry, K., and J. O'Hagan (1972) "An Econometric Study of British Tourist Expenditure in Ireland". *Economic and Social Review*, Vol. 3, pp. 143-161.

Baggio, R. (2011), "Collaboration and cooperation in a tourism destination: a network science approach", *Current Issues in Tourism*, Vol. 14 No. 2, pp. 183-189.

Baggio, R., Scott, N., and Cooper, C. (2010), "Network science - a review focused on tourism", *Annals of Tourism Research,* Vol. 37 No. 3, pp. 802-827.

Baláž, V., and Williams, A. M. (2005), "International Tourism as Bricolage: an Analysis of Central Europe on the Brink of European Union Membership", *International Journal of Tourism Research,* Vol. 7 No. 2, pp. 79-93.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004), "The architecture of complex weighted networks", *proceedings of the National Academy of Sciences of the United States of America,* doi 10.1073/pnas.0400087101, pp. 3747-3752.

Beerli, A., and Martín, J., (2004). "Factors Influencing Destination Image", *Annals of Tourism Research*, Vol 31, pp. 657-681.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006), "Complex networks: Structure and dynamics", *Physics Reports*, Vol. 424 No. 4-5, pp. 175-308.

Bosetti, V., Cassinelli, M., and Lanza, A. (2007), "Benchmarking in Tourism Destinations; Keeping in Mind the Sustainable Paradigm", in Matias, A., Nijkamp, P., and Neto, P. (Eds.), *Advances in Modern Tourism Research,* Physica-Verlag HD, pp. 165-180.

Bruns, A., and Stieglitz, S. (2013), "Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology* Vol 16, No 2, pp. 91–108. doi:10.1080/13645579.2012.756095

Budd, L., Francis, G., Humphreys, I., and Ison, S. (2014), "Grounded: Characterising the market exit of European low cost airlines" *Journal of Air Transport Management,* Vol. 34, pp. 78-85.

Caldarelli, G. (2007), *Scale-Free networks: complex webs in nature and technology,* Oxford University Press, Oxford.

Chuang, T. C., Liu, J. S., Lu, L. Y. Y., Tseng, F.-M., Lee, Y., Chang, C.-T. (2017), "The main paths of eTourism: trends of managing tourism through Internet". *Asia Pacific Journal of Tourism Research,* No 22: Vol 2, pp. 213-231.

Clauset, A., Shalizi, C., and Newman, M. (2009), "Power-law distributions in empirical data", *SIAM Review*, Vol. 51, pp. 661-703.

Coshall, J. (2000), "Spectral analysis of international tourism flows", *Annals of Tourism Research,* Vol 27 No. 3, pp. 577-589.

Crouch, G. (1994), "The Study of International Tourism Demand: A Survey of Practice", *Journal of Travel Research*, Vol. 32, No. 4, pp. 41-55.

Csardi, G., and Nepusz, T. (2006), "The igraph software package for complex network research", *InterJournal, Complex Systems,* Vol. 1695 No. 5, pp. 1-9.

da Fontoura Costa, L., Rodrigues, A., Travieso, G., and Villas Boas, P.R. (2007), "Characterization of complex networks: A survey of measurements", *Advances in Physics,* Vol. 56 No. 1, pp. 167-242.

De Cantis, S., Parroco, A.M., Ferrante, M., and Vaccina, F. (2015), "Unobserved tourism", *Annals of Tourism Research,* Vol. 50, pp. 1-18.

Du, H., White, D.R., Ren, Y., and Li, S. (2009), *A normalized and a hybrid modularity* (Draft paper): University of California, Irvine, CA, USA., available at: http://intersci.ss.uci.edu/wiki/pub/20080401drwNormalizedModularityDraft_BW.pdf (accessed 22.05.2016).

ECORYS. (2009), *Study on the Competitiveness of the EU tourism industry (for the European Commission: Directorate-General Enterprise & Industry)*, ECORYS SCS Group, Rotterdam, available at: http://ec.europa.eu/enterprise/sectors/tourism/files/studies/competitiveness/study_on_tourism_competitiveness_2009_en.pdf (accessed 11.06.2016).

Edler, D., and Rosvall, M. (2013), "The MapEquation software package", available at: http://www.mapequation.org (accessed 05.07.2016).

European Commission. (2010), *Policy paper: Europe, the world's No 1 tourist destination – a new political framework for tourism in Europe (COM/2010/352 final, 30/06/2010),* available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A7-2011-0265+0+DOC+PDF+V0//EN (accessed 02.09.2016)

Fortunato, S. (2010), "Community detection in graphs". *Physics Reports*, Vol. 486 No. 3-5, pp. 75-174.

Furmanov, K., Balaeva, O., and Predvoditeleva, M. (2012), "Tourism flows from the Russian Federation to the European Union", *Anatolia,* Vol. 23 No. 1, pp. 17-31.

Guimerà, R., Sales-Pardo, M., and Amaral, L. A. N. (2004), "Modularity from fluctuations in random graphs and complex networks", *Physical Review E,* Vol. 70 No. 2, 025101.

Hagberg, A.A., Swart, P.J., and Schult, D.A. (2008), "Exploring network structure, dynamics, and function using NetworkX", *proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA (19-24 Aug)*, pp. 11-15.

Harjaa, E., and Stângaciu, O.A. (2013), "Tourism Activity of Member States and Romania's Place within the First Five Years after EU Accession", *Procedia Economics and Finance,* Vol. 6, pp. 491-502.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., and Ratti, C. (2014), "Geo-located Twitter as proxy for global mobility patterns", *Cartography and Geographic Information Science,* Vol. 41 No. 3, pp. 260-271.

Hecht, B., and Stephens, M., (2014), “A tale of cities: Urban biases in volunteered geographic information”, in Proceedings of the International Conference on Web and Social Media (ICWSM), pp. 197-205, Menlo Park, California. AAAI Publications.

Hu, H.B., and Wang, X.F. (2008), "Unified index to quantifying heterogeneity of complex networks", *Physica A,* Vol. 387 No. 14, pp. 3769-3780.

Jansen-Verbeke, M., and Spee, R. (1995), "A regional analysis of tourist flows within Europe", *Tourism Management,* Vol. 16 No. 1, pp. 73-80.

Jurdak. R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D. (2015) “Understanding Human Mobility from Twitter”, *PLoS ONE* Vol. 10, No. 7: e0131469. doi:10.1371/journal.pone.0131469.

Lam, C., & McKercher, B. (2013), "The tourism data gap: The utility of official tourism information for the hospitality and tourism industry", *Tourism Management Perspectives,* Vol. 6, pp. 82-94.

Leidner, R. (2007), *The European tourism industry in the enlarged Community - Gaps are potentials and opportunities. Report for the Enterprise Directorate-General of the European Commission*, Office for Official Publications of the European Communities, Luxembourg.

Lim, C. (1997), "Review of International Tourism-Demand Models", *Annals of Tourism Research*, Vol. 24, pp.835-849.

Longley, P.A., Adnan, M., and Lansley, G., (2015), “The geotemporal demographics of twitter usage”, *Environment and Planning A*, Vol. 47, No. 2, pp. 465-484.

Lu, W., and Stepchenkova, S. (2015), "User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software", *Journal of Hospitality Marketing & Management,* Vol. 24, No. 2, pp. 119-154.

Marrocu, E., and Paci, R. (2011), "They arrive with new information. Tourism flows and production efficiency in the European regions", *Tourism Management,* Vol. 32 No. 4, pp. 750-758.

Melián-González, A., and García-Falcón, J. (2003),"Competitive Potential of Tourism in Destinations", *Annals of Tourism Research*, Vol. 30, pp.720-740.

Miguéns, J.I.L., and Mendes, J.F.F. (2008), "Travel and tourism: Into a complex network", *Physica A,* Vol. 387 No. 12, pp. 2963-2971.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., and Rosenquist. J.N., (2011), “Understanding the demographics of twitter users”, in Proceedings of the International Conference on Web and Social Media (ICWSM), pp. 554-557, Menlo Park, California. AAAI Publications.

Morley, C. (1998), "A Dynamic International Demand Model", *Annals of Tourism Research,* Vol. 25, pp. 70-84.

Murphy, P., Pritchard, M., and Smith, B., (2000), "The Destination Product and Its Impact on Traveller Perceptions", *Tourism Management*, Vol. 21, pp. 43-52.

Newman, M.E.J. (2010), *Networks - An introduction*, Oxford University Press, Oxford.

Newman, M.E.J. (2002), "Assortative mixing in networks", *Physical Review Letters,* Vol. 89 No. 20, 208701.

Noldus, R., and Van Mieghem, P. (2015), "Assortativity in complex networks", *Journal of Complex Networks* (first published online March 26, 2015 doi:10.1093/comnet/cnv005).

NRIT. (1991), *Trends in Tourism and Recreation in the European Community* (Working document 60). Breda, NL: Nederlands Research Instituut voor Recreatie en Toerisme. Retrieved March 2015, available at:

http://www.wrr.nl/fileadmin/nl/publicaties/DVD_WRR_publicaties_1972-2004/Trends_in_Tourism_and_Recreation_in_the_.pdf (accessed 03.04.2016).

Pavalanathan, U., Eisenstein, J. (2015), “Confounds and Consequences in Geotagged Twitter Data”, Cornell University Library. Retrieved May 2017, available at arXiv:1506.02275, 2015.

Pearce, D.G. (1987), "Spatial patterns of package tourism in Europe", *Annals of Tourism Research,* Vol. 14 No. 2, pp. 183-201.

Rosvall, M., and Bergstrom, C.T. (2010), "Mapping change in large networks", *PLoS ONE,* Vol 5 No. 1, art. e8694, doi:10.1371/journal.pone.0008694.

Scott, N., Baggio, R., and Cooper, C. (2008), *Network Analysis and Tourism: From Theory to Practice*, Channel View, Clevedon, UK.

Sinclair, T. (1998), "Tourism and Economic Development: A Survey", *The Journal of Development Studies*, Vol. 34, No 5: pp. 1-51.

Sobolevsky, S., Campari, R., Belyi, A., and Ratti, C. (2014), "General optimization technique for high-quality community detection in complex networks", *Physical Review E, Vol. 90 No.* 1, art. 012811, doi: 10.1103/PhysRevE.90.012811.

Struijs, P., Braaksma, B., and Daas, P. J. (2014). “Official statistics and big data”, *Big Data & Society*. Vol. 1, pp. 1-6.

UNWTO. (2014), *Tourism statistics - intra-EU tourism flows*. Madrid: World Tourism Organization, available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Tourism_statistics_-_intra-EU_tourism_flows (accessed 08.12.2015).

van der Zee, E., and Vanneste, D. (2015), "Tourism networks unravelled; a review of the literature on networks in tourism management studies", *Tourism Management Perspectives,* Vol. 15, pp. 46-56.

Volo, S. (2004), "A journey through tourism statistics: accuracy and comparability issues across local, regional and national levels", *proceedings of the 24th Biennial Conference on Regional and Urban Statistics: Understanding Change (SCORUS), Minneapolis, MN (21-24 May)*, pp. 210-216.

Witt, C., Witt, S., and Wilson, N. (1994), "Forecasting International Tourism Flows", *Annals of Tourism Research*, Vol. 21, pp. 612-628.

Yan, X. Y., Han, X. P., Wang, B. H., Zhou T. (2013), “Diversity of individual mobility patterns and emergence of aggregated scaling laws”, Scientific Reports, 2013; 3. doi: 10.1038/srep02678.

Yoon, Y., and Uysal, M., (2005), "An Examination of the Effects of Motivation and Satisfaction on Destination Loyalty: A Structural Model", *Tourism Management,* Vol. 26, pp. 45-56.

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., and Buckee, C. O. (2013), “The impact of biases in mobile phone ownership on estimates of human mobility”, *Journal of the Royal Society Interface*. Vol. 10, No 81, 20120986, doi: 10.1098/rsif.2012.0986