Information Technology & Tourism, Vol. 13, pp. 215–228
Printed in the USA.

1098-3058/12 $60.00 + .00
DOI: http://dx.doi.org/10.3727/109830512X13283928066913
www.cognizantcommunication.com

DESTINATIONS AND THE WEB: A NETWORK ANALYSIS VIEW

ROLAND PIAZZI,* RODOLFO BAGGIO,† JULIA NEIDHARDT,‡ and HANNES WERTHNER‡

*Graz University of Technology, Graz, Austria
†Dondena Center for Research on Social Dynamics, Bocconi University, Milan, Italy
‡Vienna University of Technology, Vienna, Austria

Network analysis methods have gained much attention in the last few years and have provided a wealth of insights into the structural and dynamic properties of many systems. Here, we apply these methods to the study of tourism destinations' Web spaces. This exploratory analysis aims at showing how these techniques can be used and what outcomes can be obtained. After a short introduction to network analysis and a brief review of the literature, two cases are presented, namely Austria as a whole country, and a smaller destination within Italy: the island of Elba. For each case, data collection methods are described and the characteristic network parameters are calculated. The comparison between the two cases highlights both similarities and differences, which are described and interpreted. Finally, the limitations of this approach are discussed.

Key words: Web; Network analysis; Tourism destinations

Introduction

The Web has dramatically transformed the tourism industry and its landscape, and this even more as foreseen in the optimistic early days of e-tourism in the 1990s (Poon, 1993; Schertler, Schmid, Tjoa, & Werthner, 1994; Sheldon, 1997; Werthner & Klein, 1999). It is common knowledge that nearly all tourism players operate a website and that a very large proportion (approx. between 30% and 50%, dependent on region and/or reference; e.g., PhoCusWright, 2011) of the turnover generated by the sector originates from online activities. It is clear then, how strong the impact of Internet applications has been on tourism and why this industry is one of the prominent application fields in the Web, also due to its specific features. This reciprocal relationship was also already predicted in the early days of e-tourism (Werthner & Klein, 1999).

The Web-induced transformation of tourism happened on all levels (i.e., individual, enterprise related, and structural). Consumers have changed their search and booking behavior, and have benefited from a supposedly more "transparent" market, which allows them to compare different online offers and services in a convenient way. They become *actively interactive* players in the field.

Address correspondence to Rodolfo Baggio, Master in Economics and Tourism, Dondena Center for Research on Social Dynamics, Bocconi University, via Sarfatti, 25, 20136 Milan, Italy. Tel: +39 0258365437; Fax: +39 0258365439; E-mail: rodolfo.baggio@unibocconi.it

Tourism companies and organizations use the Web to directly access the electronic market, to promote themselves and to sell their products and services. A further effect of the virtualization of tourism activities is that the distinction between (types of) players blurs. Suppliers become intermediaries and vice versa, and consumers became part of the supply chain (e.g., using Web 2.0 applications and generating contents, information, and services). These phenomena point at the third level of change (i.e., the structural one). One can observe permanently new services as well as players in the field, and at the same time, a somehow opposed *winners-takes-all* (or most) trend with a high level of concentration. At this level, the Web shows a dialectic relationship of order and un-order.

But one should note that information technologies not only facilitate or force these changes, but also offer a historically unmatched possibility to observe the world and its changes. Since nearly all human activities are moving to the Web and the Web itself is a transparent medium, we have the opportunity to observe both activities and the infrastructure used—the Web becomes a mirror of the world. Data are easily available, in a cheap and fast manner. This will also lead to a change in the way economic and social studies will be performed (Hendler, Shadbolt, Hall, Berners-Lee, & Weitzner, 2008).

This second side of the coin is also of importance for the tourism sector, where up to now companies have used the Web mainly for PR activities, marketing, and sales, and not so much in the field of analysis. Also, when looking at the scientific research, this "data" power of the Web is somehow—in general terms—not well represented, with some exceptions in the field of text mining or blog mining, related to consumer behavior.

But there has been little quantitative research focusing on structural issues using, for example, network analysis. Network theoretical methods allow a structural analysis of the relationships and connections between the respective market participants, and may provide good insights, as already done in numerous other fields (da Fontoura Costa et al., 2011). The idea is to draw a global picture, which then permits to draw better conclusions than by just looking at the different actors locally. This leads to a better understanding of the configuration and the evolution of the virtual world taken as a whole. This may allow having more general and reliable information to be used in deciding strategies and tactics, which might be not easily viewable when looking only at individual actors on a local scale.

In this article, such a structural analysis is provided by examining two different cases: the Italian island of Elba and the country of Austria. But the article also deals with the question of whether it is possible to identify the market's key players by the use of such an approach. We assume that key players may not only be defined by their information/network relationship to others, but also by their economic performance, such as turnover. As it will turn out, network analysis provides a good insight into the structural issues, but it cannot paint the entire picture. The article is organized as follows. In Section 2, a short introduction to network theory is given and the concepts used in the article are described. Section 3 contains an overview of the state of the art of network analysis in tourism and e-tourism. Section 4 presents the cases and Section 5 compares and discusses the outcomes. Finally, Section 6 is dedicated to concluding remarks and possible future research in this area.

A Short Introduction to Network Analysis

Various "real-world" phenomena can be captured and modeled by using a network representation. The entities observed form a set of vertices or nodes, joined (in pairs) by edges or links, if there is a specified relationship or interaction between them. By looking at this structure, network theoretical methods enable to ascertain the global structural and dynamic characteristics of the entire system. Measurable quantities (parameters) enable one to formulate general statements and allow comparisons between different networked systems (for a thorough review of all the concepts and metrics described in this section, see da Fontoura Costa, Rodrigues, Travieso, & Boas, 2007; Newman, 2010).

A network can be formally described by a graph $G = (V, E)$. The set V is called vertices, and the set E consist of edges that connect pairs of vertices. Two vertices that are joined by an edge are called neighbors (first neighbors are nodes directly connected to the node considered). If there is an edge between each pair of vertices, G is called a complete graph.

In a directed graph, each edge has an origin and a destination. On the contrary, an undirected graph has edges with no orientation (symmetrical relationships). In a weighted graph, each edge can be assigned an additional numerical value (representing, for example, cost, speed, time, etc.).

The degree deg(v) of a vertex v in an undirected graph G is the number of first neighbors the vertex v has. The average degree of graph G is the arithmetic mean of all degrees deg(v). Obviously, in a directed graph, the in-degree and out-degree of a vertex are separately considered: in-links are connections *coming* to a vertex, out-links are those going to some other vertex.

A path in a graph is a sequence of vertices such that two consecutive vertices are joined by an edge. The number of all such edges is called the length of the path. The distance (also termed geodesic distance) $d(v, w)$ between two nodes v and w in a graph is defined as the length of the shortest path between them. The average distance in a graph is the arithmetic mean of the distances between all pairs of vertices. If there is a path from a vertex v to a vertex w, these vertices are called connected. A connected component in a graph is a set of vertices all connected among themselves. The diameter of a graph is defined as the longest possible distance existing in the network (the maximal distance between any two connected nodes). These quantities, when considering directed graphs, take into account the edges' orientations. In this case, two connected components are defined: strongly and weakly connected components. In a strongly connected component (SCC), there is a directed path from each vertex to each other. In a weakly connected component, there is one path from each vertex to each other, but the edges' orientation is ignored.

Let G be an undirected graph and n the number of its vertices, then the density ρ of G is defined as the number m of edges divided by the maximum possible number of edges (those present if G were a complete graph):

$$\rho := \frac{2m}{n(n-1)}.$$

A local density can also be defined: the clustering coefficient. Let k_i be the number of neighbors of a vertex v and let e_i be the sum of all edges between them. If each pair of neighbors of vertex v were connected by an edge, then there would be $k_i(k_i - 1)/2$ edges. Therefore, the clustering coefficient C_i of a vertex v is:

$$C_i := \frac{2e}{k_i(k_i - 1)}.$$

Hence, C_i reflects the probability that two arbitrary neighbors of v are connected by an edge. The clustering coefficient C of the entire graph G is defined as the arithmetic mean of the clustering coefficients C_i of all vertices. Looking at directed graphs, there can be two edges between each pair of vertices—one in each direction. Taking this into account, both the density ρ and the clustering coefficient C_i as defined above have to be divided by two.

Since network analysis methods should provide a better understanding of the *real-world* structure, some concepts (and the associated metrics) that facilitate a richer interpretation have been proposed. In this respect, a very important category is the class of so-called *centrality measures*. These try to formalize the idea that in many instances, some vertices (or edges) play a more important role than others, hence they should be considered as more *central*. For the study of the e-tourism networks, three commonly used centrality measures are: degree centrality, closeness centrality, and betweenness centrality.

The degree centrality $C_D(v)$ for vertex v is defined as the number of edges it is connected to: $C_D(v) = \deg(v)$.

In a directed graph, two kinds of degree centrality are usually distinguished, namely in-degree centrality and out-degree centrality.

The closeness centrality $C_C(v)$ for a vertex v is defined as the reciprocal value of the sum of all distances between v and each other vertex w:

$$C_C(v) = \frac{1}{\Sigma_{w \in V \setminus v} d(v, w)}.$$

The betweenness centrality $C_B(v)$ for vertex v is defined as follows:

$$C_B(v) = \sum_{u \neq v \in V} \sum_{w \neq v \in V} \frac{\sigma_{uw}(v)}{\sigma_{uw}}.$$

Here, σ_{uw} denotes the number of shortest paths between vertex u and vertex w, and $\sigma_{uw}(v)$ denotes the number of shortest path between those vertices that run through v. Usually the values calculated for these metrics are normalized.

The interpretation of these quantities is intuitive. According to the degree centrality, a vertex is the more important the more neighbors it has; such a vertex is able to influence many others. A vertex that lies on many shortest paths between pairs of vertices in the graph is important according to the betweenness centrality. Such a vertex is involved in the interaction between those vertices and is capable of controlling their communication. On the other hand, a vertex with high betweenness can be regarded as a bridge connecting two different areas of the network and assume the role of a bottleneck. When considering closeness centrality, vertices that have in total a small distance to all other nodes are considered to be more central or important. The vertex with the highest closeness centrality reaches all the other vertices through a minimum number of intermediaries, and is, for example, able to communicate faster with the whole network than every other vertex. Of course, the applicability and the significance of any centrality index depends on the area observed and on the questions asked. When considering directed networks, all centrality measures are calculated separately for in-links and out-links (e.g., we have in-degree and out-degree, in-closeness and out-closeness, etc.).

There are a number of properties, mostly related to the quantities described above, that many "real-world" networks have in common. One of them is the so-called power law degree distribution; that is, the degree distribution of the network's nodes can be approximated by a function of the form $p(k) = ck^{-\gamma}$, where $k \in \mathbf{N}$ denotes the degree of a vertex and $c \in \mathbf{R}$ and $\gamma \in \mathbf{R}$ are positive constants. This implies that in such a network, the majority of vertices have a very low degree while a very few members of the system have a remarkable high number of neighbors, thus acting as hubs for the network. One of the most common mechanisms for obtaining such a topology has been found in the fact that links are not added randomly, but are attached to specific vertices preferentially. Such networks are called scale free. Another common property is the so-called small world property. The term goes back to an experiment by Stanley Milgram (1967) on social networks and expresses that the average distance within such a network is relatively short.

A third property that many "real-world" networks have in common is their community structure: the network's vertices can be divided into groups within which the edges are denser than between different communities. Modularity measures can be used to qualify a particular division of a network into communities: such an index represents the ratio between the links connecting the vertices in a module and those connecting vertices belonging to different modules. The commonly used metric is called modularity index:

$$Q = \sum_i (e_{ii} - a_i)^2$$

where e_{ii} is the fraction of edges in the network between any two vertices in a given subgroup i, and a_i is the total fraction of edges with one vertex in the group. The modularity index for the whole network is the mean of those calculated for the modules present in the network. Many social and economic networks, then, see the presence of a very large connected component, which contains the vast majority of all vertices, and of other components of a smaller size with respect to the *giant* one.

It is well known that the World Wide Web exhibits all these properties. It can be modeled by a directed graph whose vertices are webpages (or websites) and whose edges are the hyperlinks between them. Because of its size and its permanent changes, its actual number of vertices and edges can only be estimated. Based on data gathered by some extensive crawlings (Kumar et al., 2000), the structure of the Web is usually described as follows.

A large SCC comprises about 28% of all webpages; associated with this SCC are one in-component and one out-component. Both of them comprise 21% of webpages. From each vertex of the in-component there is a directed path to the SCC, but not the other way round. For the out-component the situation is reversed: there is a directed path from the SCC to each of the out-component's vertices, but not vice versa. The remaining 30% of webpages cannot be reached from the SCC by a directed path nor can they reach the SCC; most of them are connected to the in-component or the out-component. In summary, there is a giant weakly connected component that comprises around 91% of all webpages. Many other subsequent studies have reconfirmed these findings and have also shown a characteristic self-similarity in this structure. In other words, this bow-tie (this is the name given to the model) structure can be found in smaller portions of the Web (Dill et al., 2002). In addition and according to the literature, the average degree of a web page is 7.2; and there is a power law degree distribution for both the webpages' in-degrees and out-degrees (with an exponent $\gamma = 2.1$ for the in-degrees and $\gamma = 2.7$ for the out-degrees). Furthermore, the World Wide Web seems to be a small world network with an average path length of 16.18.

Network Analysis in Tourism and e-Tourism: A Short Overview

As already mentioned, the network analysis line of research has been little applied in the field of tourism. In the following, we give a short overview. It must be noted, however, that these works are rather different in nature and hardly comparable. Roughly, the existing studies can divided into three categories:

- papers and studies dealing with the analysis of web network structures, where also comparisons between electronic and real-world networks are made;
- network analysis applied to physical infrastructure networks, such as roads or transportation networks; and
- works dealing with the relationships between individuals or organizations engaged in tourism (e.g., actors developing tourism policy strategies).

Real World Versus Technological Networks

This category covers studies that treat Web networks and the reflected "real" world at the same time and compares them. Real-world networks consist, in this case, of tourism businesses like hotels or other accommodations that are connected through business relations (denoted as socioeconomic networks), or just through information flows. At the other side, technological networks can be identified as networks that consist of the websites of tourism businesses that are connected through hyperlinks. Here, the most comprehensive studies were conducted by Baggio and colleagues with the analysis of the Elba tourism destination (Baggio, 2007; Baggio, Scott, & Cooper, 2010; Da Fontoura, Costa, & Baggio, 2009; Scott, Cooper, & Baggio, 2008). Both the real system and the technological network are identified and compared. Also, a comparison between different destinations was performed by studying the Web spaces of the Fiji Island and of Elba (Baggio, Scott, & Wang, 2007). Interestingly, and probably not surprisingly, both islands show similar features. Furthermore, a study of the Web network of Italian travel agencies (Baggio, 2006) showed again similar connectivity characteristics. Finally, Scott, Cooper, and Baggio, 2008 studied the networks of four differently developed Australian tourism regions. The actors of these networks were some of the key stakeholders of the tourism regions. Objects of measurement were not only quantitative information like the frequency of contact, but also qualitative data like the purpose of contact. In this study, the density, average path length, and closeness centrality were measured, as well as some modularity analysis. Besides the structural characteristics and the effects these may have on dynamic processes, such as the diffusion of information, one of the most interesting results of this series of works is the topological similarity between the real and the virtual networks, which may allow, for example, to consider the Web network of a destination as a significant sample of the real set of relationships between the destination actors.

Transportation Networks

A different approach, also using network analysis, is the one of Shih and Hsin-Yu (2006), who looked at the area of car tourism. Here, the nodes of the network are the towns and villages of the region Nantou (Taiwan) and the links are the transport facilities (roads). Unlike the studies mentioned in the previous section, no comparison was made with other networks. However, similar to "classical" social network analysis, the authors used the main centrality measures, such as degree, closeness, and betweenness. Additionally, the authors distinguish between beginning, core, and terminal nodes. This categorization arises when looking at the proportion of in-links and out-links. For example, there exists a terminal node, when it has many in-links and few out-links, because much routes end in this node. Other interpretations are that nodes with a high incloseness are easily reachable, whereas those with a high out-closeness can be called gateways, because through them many other locations can be reached in a short time. Nodes with a high betweenness value are interpreted as stopping points. The overall aim of the study was to find out where new tourism infrastructure should be built of which type they should be, and which routes should be preferred and therefore promoted. Many other studies have dealt with airlines, train, and navigation networks. The main aim, besides assessing the general characteristics of these networks, was to find possible reconfigurations of the infrastructures with aiming at optimizing travel times or reducing congestions and bottlenecks (da Fontoura Costa et al., 2011).

Actors' Networks

A different perspective was opened by Pforr (2005). The work focused on the network of actors developing tourism policy strategies (thus being only slightly related with our line of work). The main parameters measured in this study, along with the development of a tourism policy master plan, as defined by the author, are i) influence reputation, ii) cooperation, iii) information exchange, and iv) frequency of two-way communications. The objective was to find the density of the relations between the actors and, if then, to which extent key stakeholders involve their partners. Here, the relative high density factors of 0.6 for the undirected and of 0.486 for the directed network are of interest. Comparisons with other networks developing policy master plans (e.g., some general economic policy master plan activities of governments) should be made to find out possible differences due to the specific issues addressed in order to highlight peculiar features of tourism planning processes.

Two Case Studies

This section presents the results of investigations in the analysis of the Web networks of two tourism destinations: Austria (whole country) and the Italian island of Elba. The different data collection methods are described and the main results of the quantitative analysis are presented.

Austria

Piazzi (2011) analyzes the network structure of Austrian e-tourism market and tries to identify its central websites. These websites are determined by the main centrality metrics: degree, closeness, and betweenness. In addition, other different and important parameters of network analysis are calculated, such as density, average path length, average degree, and clustering coefficient, as well as the distribution of the degrees.

The object of this study is the network of the Austrian e-tourism market, which consists of the websites of the Austrian tourism players that promote and sell tourism in Austria. The hyperlinks between the websites represent the connections between these websites (i.e., nodes). It is important to note here, that the main attempt was to find a network consisting only of the tourism players in the Austrian market (according to the classification of Werthner & Klein, 1999). However, actors that are indirectly involved in the tourism industry were also considered; for example, important players in this network are social media sites, such as Facebook, Youtube, and Twitter. They play an important role as promotional channels used by tourism players.

The data gathering for the network was done with a simple Web crawler (developed by one of the authors). The tool searches the World Wide Web by starting with one or more websites, retrieves the hyperlinks, and follows them in order to uncover other websites. The Web crawler used in this study was written in Java and uses MySQL database for storing webpages and hyperlinks. The data obtained with the crawling process are transformed into a format suitable for the analysis, which was conducted by using UCINET (Borgatti, Everett, & Freeman, 1992) and Pajek (Batagelj & Mrvar, 2009).

The resulting network can be seen as a directed (asymmetric), unweighted adjacency matrix. The nodes of the network were the entire websites rather than single webpages. In this way, the central players of the Austrian tourism industry are better and more comprehensively represented. Moreover, the links were not weighted, assuming (reasonably) that multiple links do not necessary indicate a stronger relation. The network uncovered by the crawler, due to computational requirements, does not constitute the whole Austria tourism Web space. Many, mainly small, websites were not included. A deeper inspection to the most representative websites found by the crawler was conducted to check whether they belong to tourism actors (Werthner & Klein, 1999), and whether or not their products or services can be classified as tourism related. The resulting network comprises 2.395 of nodes and includes a number of general websites (such as, the social media) connected with the tourism industry. Seventy percent of these 2.395 nodes could be classified as tourism actors (including the handful of social medias) and around 16% were marked by the author as indirectly tourism-related players (comprising enterprises that sell typical products of a region and where tourists are considered as a potential target group).

In summary, the network identified can be considered as a reliable sample of the whole system, also taking into account that there are 91,600 tourism enterprises in Austria (Österreich Werbung, 2011). Figure 1 shows the network obtained, along with the cumulative degree distributions (in-degrees and out-degrees). For the sake of visibility, the figure only reports websites with a deg >30.

In this network (as one may also predict), social media sites, such as Facebook or Youtube, receive the most in-links. Their important role in the promotion of products, as well as in communication and customer relations, is well recognized, not only in tourism. Here, the central sites belonging to the tourism industry are *Tiscover.at*, *Austria.info,* and *Oebb.at*. The first one is the biggest (in terms of customers) online tourism portal of the Alps, originally starting as a destination system. The second site is operated by the Austrian tourist board, and *Oebb.at* is the portal of the Austrian railways company.

The most central websites according to their degree, closeness, and betweenness are listed in Tables 1 and 2.

The site with the highest number of out-degrees is *Austria.info*, followed by other systems that represent destinations (destination management systems). In nearly all these cases, these sites are operated by public boards, with their task to represent mainly small private tourism enterprises. They do this not only by providing information to the customer but (what is important for this study) by referencing tourism enterprises through hyperlinks on their websites. There is a difference to column in-degree with social media sites at prominent

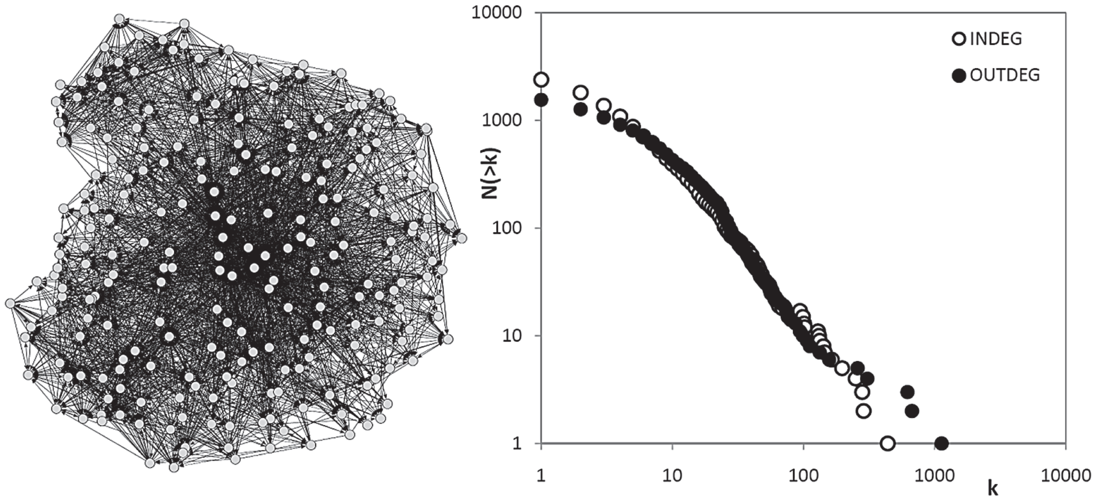

Figure 1. Network of the Austrian tourism operators (for easing visualization only nodes with degree >30 are shown), along with the in-degree and out-degree cumulative distributions [*k* is the degree and N(>*k*) the number of nodes having degree >*k*].

Table 1
Most Central Austrian Websites for In- and Out-Degree, and In-Closeness (Social Media Sites in Italic)

Website	In-Degree	Website	Out-Degree	Website	In-Closeness
facebook.com	18	*austria.info*	47	*facebook.com*	39
tiscover.at	12	tirol.at	28	tiscover.at	35
youtube.com	12	genussland.at	26	*youtube.com*	35
oebb.at	10	mobile.austria.info	13	oebb.at	34
tiscover.com	8	austriatourism.com	11	tiscover.com	31
genussland.at	7	oberoesterreich.at	7	*twitter.com*	31
twitter.com	6	vorarlberg.travel	6	austria.info	31
addthis.com	6	blog.austria.info	5	*addthis.com*	30
oberoesterreich.at	6	ausflugstipps.at	4	oberoesterreich.at	30
tirol.at	5	linz.at	4	tirol.at	30

positions. This represents the fact that these sites are used as links to joint services and/or information sources (e.g., videos, in the case of Youtube). This underlines their importance for the tourism sector, but also shows their "open" linking policy and network based business model.

The results for the in-closeness and out-closeness are similar to those of the in-degrees and the out-degrees. This seems to be evident as many in-links make a node also closer to others, so that the node can be reached by more paths and therefore faster over the whole network. This fact applies for the out-degrees and the out-closeness (see Table 2) as well. When looking at the betweenness (i.e., such a node is involved in the communication between many nodes and controls their communication) values in Table 2, it appears that, with exception of the *Austria.info*, all sites show only low values. However, social media sites are represented in this table (with its obvious interpretation) as well. But in general, the low numbers of this parameter point at only a limited number of intermediaries in the network, which in turn means that the most websites are either connected directly between them or are not connected at all. Moreover, the low betweenness values suggest a low density of the network. The most important characteristic network parameters are shown in Table 3.

The network density of 0.003 seems to be rather low, but it is quite similar to that of other similar Web networks (for more information on the values for the World Wide Web, see da Fontoura Costa et al., 2011; Piazzi, 2011). Also, the average degree can be compared with that of the World Wide Web. The average path length and the clustering coefficient of the network of the Austrian e-tourism market are respectively lower (path length) and higher (clustering coefficient), with respect to what would be expected for a Web subnetwork of the density recorded here. Examples are the studies of Adamic and Adar (2003), who compare the technological networks (consisting of homepages) of students

Table 2
Most Central Austrian Websites as for Out-Closeness and Betweenness

Website	Out-Closeness	Website	Betweenness
austria.info	71	austria.info	24
tirol.at	60	*facebook.com*	15
genussland.at	53	genussland.at	14
mobile.austria.info	53	tirol.at	14
austriatourism.com	52	tiscover.at	9
scnem.com	51	oberoesterreich.at	5
region.austria.info	51	*youtube.com*	3
oberoesterreich.at	50	*twitter.com*	2
vorarlberg.travel	49	vorarlberg.travel	2
blog.austria.info	49	tiscover.com	2

Table 3

Characteristic Properties of the Network of the Austrian e-Tourism Market

Parameter	Value
Number of nodes	2,395
Number of links	16,893
Density	0.003
Average degree	7.06
Average path length	3.33
Clustering coefficient	0.34
Key parameter of the power law distribution of the in-degrees[a]	2.47
Key parameter of the power law distribution of the out-degrees[b]	2.63

[a]In-degree $k > 12$.
[b]Out-degree $k > 29$.

and staff of the Stanford University and the Massachusetts Institute of Technology (MIT), or the network analysis of the subdomain nd.edu (also cited in Piazzi, 2011). Both investigations show a lower average path length and a higher clustering coefficient than the whole World Wide Web. But these values for the average path length and the clustering coefficient of the Austrian e-Tourism market and also of the World Wide Web are—despite being something different—similar to those of small world networks (including real world social networks), whose main characteristic is the connectedness of the nodes through short paths (average path length is around the logarithm of the number of nodes) and the formation of clusters. That is to say that the connectedness within the local neighborhood is higher than in the network as a whole.

The distribution of the degrees (Fig. 1) follows a power law, at least in the most important central area of the distribution (for in-degree > 12 and for out-degree > 29). The power law exponents reported in Table 3 were calculated following Clauset, Shalizi, and Newman (2009). Also in this case, there is a high similarity with the general values reported in the literature for the World Wide Web: the in-degree distribution exponent is slightly higher than that of the World Wide Web, while the out-degree distribution exponent is more or less equal. Apart from the obvious consideration of the difference in size, a steeper in-degree distribution, which indicates a greater difference between the number of nodes with a high degree and those with a low degree, can be a symptom of some form of higher modularization of the network.

Island of Elba

The second case described concerns the websites of an Italian tourism destination: the island of Elba. Off the coast of Tuscany, in the heart of the western Mediterranean Sea, its geographic position, temperate climate, and the variety and beauty of its landscapes, coast, and sea makes it a renowned tourist destination. The elements of the network examined are the websites belonging to the core tourism operators: accommodation (hotels, residences, camping sites, etc.), intermediaries (travel agencies and tour operators), transport, regulation bodies, and services. The whole network comprises 468 elements. In this case, the data were collected in a different way: a list of all Elba tourism operators (provided by the local tourist board) was considered and the websites belonging to them were examined. A simple crawler, complemented by a visual inspection, allowed enumerating the hyperlinks that form the network of the Elba tourism Web space. The network is shown in Figure 2, along with the cumulative distributions of the in-degrees and out-degrees (the study on the Elba network is one of Baggio, 2008).

The main measurements are reported in Table 4. As for the Austrian case, the degree distribution exponents were calculated following the procedure of Clauset et al. (2009). The network exhibits a clear and identifiable scale-free structure. Both the in-degree and out-degree distributions display an almost perfect power law decay with exponents out-degree = 1.89 and in-degree = 2.96. The in-degree exponent is higher and the out-degree exponent is lower than those typically measured for the

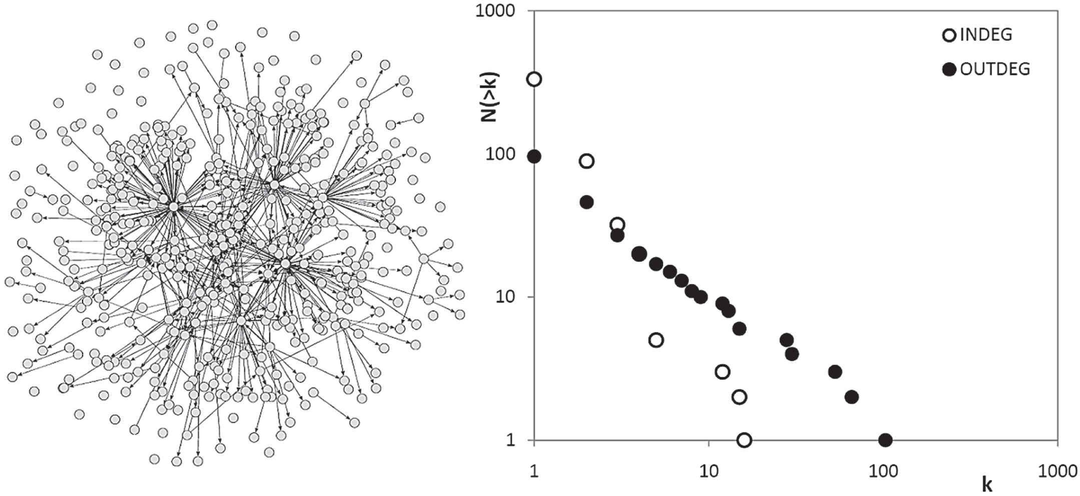

Figure 2. Elba tourism Web network and the cumulative distributions of in- and out-degrees [*k* is the degree and *N*(>*k*) the number of nodes having degree >*k*].

Web. The network is rather sparse with a very low general connectivity (link density). The path length is relatively short and global clustering coefficient quite limited. The small-world properties of the network are thus present, but not very sharply distinguishable. Overall, however, all these values are not too different from those exhibited by many other similar networks (da Fontoura Costa et al., 2011).

As Figure 2 shows, a limited number of websites seem to hold the whole network together. These hubs (see Table 5) are identified as belonging to the tourism board, to some of the attractions on the island (golf club, thermal baths, or diving), and to some initiatives of private stakeholders, which serve as virtual information points for tourists. Due to the period in which the initial study was conducted (2008), no social media websites are included in this network as their role was, at that time, negligible. The main characteristics for this network can be summarized as follows. The network shows a scale-free topology for both in-degree and out-degree distributions, which are consistent with that generally ascribed to the Web, with a very large proportion of disconnected elements. Clustering is quite limited.

Modularity Analysis of the Austrian and Elban Networks

A complex network can show a finer structure, beside the one which can be inferred by the distribution of the nodal degrees. As stated in the previous sections, the modular structure of a web network is an important element, in a tourism environment

Table 4
Characteristic Properties of the Elban Tourism Network

Parameter	Value
Number of nodes	468
Number of links	507
Density	0.002
Average degree	2.2
Average path length	4.5
Clustering coefficient (of the main connected component)	0.026
Key parameter of the power law distribution of the in-degrees	2.96
Key parameter of the power law distribution of the out-degrees	1.89

Table 5
Most Central Elban Websites for (Normalized) In- and Out-Degree

Website	In-Degree	Out-Degree
www.aptelba.it	0.0099	0.2032
www.elba-online.com	0.0039	0.1302
www.elbaexplorer.com	0.0039	0.1045
www.viagginrete-it.it	0.0020	0.0592
www.elba-capoliveri.net	0.0039	0.0552
www.elbagolfacquabona.com	0.0079	0.0256
www.termelbane.com	0.0039	0.0296
www.islepark.it	0.0316	0.0001
www.infoelba.it	0.0296	0.0001
www.divinginelba.com	0.0001	0.0237
www.elbalink.it	0.0237	0.0001

this represents the formation of cooperative groups which, as much literature suggests, plays a significant role in the development and the growth of a destination (Bramwell & Lane, 2000; Vernon, Essex, Pinder, & Curry, 2005).

Both networks were analyzed in this way. The modularity was assessed by using the Stochastic Algorithm proposed by Clauset, Newman, and Moore (2004). Essentially, the procedure tries to identify community's recursively by maximizing the modularity index Q (see Section 2). The procedure is then repeated by using, as null model, a randomized version of the networks. This randomization (a random rewiring of the existent links) is applied in a way to preserve the original degree for each node. The results are shown in Table 6 and Figure 3 shows the communities identified in both cases. As can be seen, the modularity is evident but not much higher than that calculated for the randomized versions (Austria is 9% higher and Elba 6% higher). Also, while the Austrian network has only 3 communities, the Elban network has 11, indicating a higher degree of internal fragmentation, possibly indicating a higher level of organization or cohesion of the Austrian websites.

Discussion and Comparison

The cases presented here exhibit obvious differences. The main reasons for them are in the choice of data collection methods (a semimanual method for Elba and a fully automated crawler for Austria) and in the size of the Web region examined—in the first case (Austria), it belongs to an entire country, while in the second (Elba), a local destination has been analyzed. Moreover, even if not large, the time difference between the studies (almost 2 years) might have played a role. The result can be seen mainly in the difference in density (i.e., number of links discovered). However, the two networks exhibit clear similarities in their topology, the short average path lengths and, above all, the scale-free behavior of the degree distribution are general properties that have been found in many other similar networks.

It is reasonable to assume that the different slopes of the degree distributions, although in the range of those generally found on the Web, and those in the link densities, are due to the different attitudes of the tourism operators, with a greater selfishness of the Elban companies, which tend to limit the cooperative linking practices (Baggio, 2008, 2010). This is also confirmed by looking at the clustering and modularity characteristics of the two networks.

The very limited number of studies similar to the present one does not allow relating the characteristics of the degree distribution with the evolutionary history of the destinations, even if a tendency to a more connected Web space could be attributed to a longer and better established *technological habit,* which Austria has.

One more consideration is in order here, in both cases, no major booking engines (such as Booking.com, HRS.com, Venere, and Expedia) were crawled, or no one has shown to have an important position in the network, despite the fact that they, in the *real* world, play a major role in controlling the market. The reason is that, generally speaking, these booking engines do not link to other sites, but provide

Table 6
Modularity Analysis for Austrian and Elban Tourism Web Networks (rnd is the Randomized Version of the Network, i.e., a Network Obtained by Randomly Rewiring the Existent Links While Preserving the Nodal Degrees)

Network	Q	No. of Communities
Austria	0.424	3
Austria rnd	0.384	3
Elba	0.630	11
Elba rnd	0.594	19

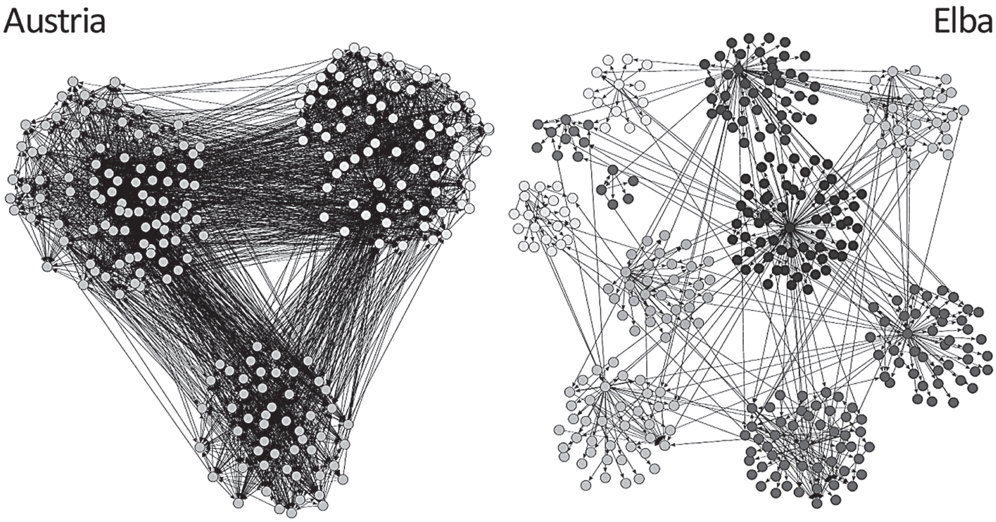

Figure 3. Modularity analysis: the communities identified for Austrian and Elban tourism Web networks (in the case of Austria, only nodes with deg >30 are shown).

only internal connections to the booking functionalities, nor do other websites (i.e., hospitality) link to them, for understandable *competitive* reasons. The only exception is, in the Austrian network, Tiscover. In this case, however, the company has its main roots in destination systems, and as such it has a higher attitude to provide its visitors with a more comprehensive view of the destination, rather than just selling bookings or packages (Pröll & Retschitzegger, 2000).

The relatively low density of both networks raises an issue when considering the possibility to find the destination's websites online and to navigate through them. New generation crawlers, in fact, built on the concepts of dynamic and focused crawling, are particularly suited to find modular well identifiable (i.e., with high link density) areas on the Web (Guimerà, Diaz-Guilera, Vega-Redondo, Cabrales, & Arenas, 2002; Liu & Menczer, 2011; Maiya & Berger-Wolf, 2011; Olston & Najork, 2010). Low densities might hinder, in the long term, the ranking of the tourism operators' websites of the destinations examined (Baggio & Antonioli Corigliano, 2009).

Conclusions

The main objective of this article was to show how network analysis methods can be used in the study of tourism destinations' Web spaces and to assess the role they can play and the insights they can provide. After a brief review of the literature and a short introduction to the techniques, we have presented two cases. One is an entire country (Austria), the second one is a smaller area (island of Elba). We have discussed the differences in the results, mainly arising from the different data collection approaches and from the timing of the analysis. We have also commented on the possible interpretation of these differences, in terms of the diverse settings, attitudes, and organizations present in the destinations. On the other hand, the studies have shown how some fundamental topological properties of the two networks are similar results, which reconfirms what the wider literature on complex networks has stated in the last years.

The results provided have shown clearly how these methods are able to uncover some structural features and how these can be used for achieving a better knowledge of the systems examined. Obviously, network analysis techniques need a strong complement in order to provide a complete picture of the objects studied. First of all, as seen, a strong qualitative knowledge is needed for a correct interpretation of the results. Secondly, network methods alone can only offer good assessments on relationships and information flows, and can be quite useful in the attempts to optimize them.

However, they are not able, by their nature, to unveil features that are not rooted in the network of online linkages, such as certain business practices and models. This is the case, for example, of those used by booking engines or other similar organizations. Moreover, further work is needed to relate other metrics (e.g., booking statistics, visits, or conversion rates) to the network position of the websites. This could greatly assist in attributing an economic value to the efforts put in place by the tourism operators, and help them in achieving higher efficiency and effectiveness in their usage of the technological environment.

The application of sound network analysis methods, although quite diffused in some disciplines, is relatively new in tourism and only a few complete works have been carried out so far. The secondary objective of this article was also to push other academics towards a wider usage of these methods. Many of the questions posed can be solved with targeted investigations, which will be conducted in the future, or by combining network analysis techniques to more traditional lines of investigation.

Biographical Notes

Roland Piazzi has a Master's degree in Business Informatics. His Master thesis deals with a network analysis of the Austrian e-Tourism. He currently works as a research fellow at the Graz University of Technology, where he is engaged in the field of information security and digital signatures.

Rodolfo Baggio has a degree in Physics and a Ph.D. in Tourism Management. He is professor at the Master in Economics and Tourism and Research Fellow at the Dondena Center for Research on Social Dynamics at Bocconi University, in Milan, Italy. He actively researches on the use of information and communication technology in tourism, and on the applications of quantitative complex systems and network analysis methods to the study of tourism destinations.

Julia Neidhardt holds a Master's degree in Mathematics from the University of Vienna. Currently, she is working on her Ph.D. thesis, and is employed as a researcher at the Institute for Software Technology and Interactive Systems at the Vienna University of Technology. Her research interests focus on Web Science and network analysis.

Hannes Werthner holds a Master and Ph.D. from the Vienna University of Technology. He was visiting professor at several universities, and was a fellow from the Austrian Schrödinger Foundation. Currently, he is also the Dean of Academic Affairs at the Informatics Faculty and the Director of the Vienna PhD School of Informatics. His research activities cover a wide range from e-Commerce and e-Tourism, Internet-based information systems, Service engineering, Recommender systems and decision support, Simulation, and Artificial Intelligence.

References

Adamic, A., & Adar, E. (2003). Friends and neighbors on the Web. *Social Networks, 25*(3), 211–230.

Baggio, R. (2006). Complex systems, information technologies, and tourism: A network point of view. *Information Technology & Tourism, 8*(1), 15–29.

Baggio, R. (2007). The web graph of a tourism system. *Physica A, 379*(2), 727–734.

Baggio, R. (2008). *Network analysis of a tourism destination*. Ph.D. thesis, School of Tourism, the University of Queensland.

Baggio, R. (2010). Collaboration and cooperation in a tourism destination: A network science approach. *Current Issues in Tourism, 14*(2), 183–189.

Baggio, R., & Antonioli Corigliano, M. (2009). Topological analysis of a tourism destination webspace: The importance of hyperlinks. *Information Technology & Tourism*, 11(4), 319–327.

Baggio, R., Scott, N., & Cooper, C. (2010). Network science: A review focused on tourism. *Annals of Tourism Research, 37*(3), 802–827.

Baggio, R., Scott, N., & Wang, Z. (2007). What network analysis of the WWW can tell us about the organization of tourism destinations. *Proceedings of the CAUTHE 2007 Conference, Sydney, Australia.*

Batagelj, V., & Mrvar, A. (2009). *Pajek: Program for large network analysis.* Retrieved May, 2009, from http://pajek.imfm.si/

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (1992). *UCINET* (Version 6.2 - 2009). Harvard, MA: Analytic Technologies.

Bramwell, B., & Lane, B. (2000). *Tourism collaboration and partnerships: Politics practice and sustainability*. Clevedon, UK: Channel View Publications.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E,* 70, 066111.

Clauset, A., Shalizi, C. R., & Newman, M. E. J (2009). Power law distributions in empirical data. *SIAM Review, 51*(4), 661–703.

da Fontoura Costa, L., & Baggio, R. (2009). The web of connections between tourism companies: Structure and dynamics. *Physica A, 388*(19), 4286–4296.

da Fontoura Costa, L., Oliveira, O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., et al. (2011). Analyzing and modeling real-world phenomena with complex networks: A survey of applications. *Advances in Physics*, 60(3), 329–412.

da Fontoura Costa, L., Rodrigues, A., Travieso, G., & Villas

Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1), 167–242.

Dill, S., Kumar, S. R., McCurley, K., Rajagopalan, S., Sivakumar, D., & Tomkins, A. (2002). Self similarity in the web. *ACM Transactions on Internet Technology (TOIT)*, *2*, 205–223.

Guimerà, R., Díaz-Guilera, A., Vega-Redondo, F., Cabrales, A., & Arenas, A. (2002). Optimal network topologies for local search with congestion. *Physical Review Letters, 89*(24), 248701.

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). Web science: An interdisciplinary approach to understanding the World Wide Web. *Communications of the ACM-Web Science, 51*(7), 60–69.

Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tompkins, A. S., & Upfal, E. (2000). The Web as a graph. *Proceedings of the 9th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 1–10), Dallas, TX, USA.

Liu, B., & Menczer, F. (2011). Web crawling. In B. Liu (Ed.), *Web data mining: Exploring hyperlinks, contents, and usage data* (2nd ed., pp. 311–362). Berlin: Springer.

Maiya, A. S., & Berger-Wolf, T. Y. (2011). *Benefits of bias: Towards better characterization of network sampling*. Paper presented at the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2011), San Diego, CA, August 21–24.

Milgram, S. (1967). The small world problem. *Psychology Today*, *1*, 60–67.

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

Olston, C., & Najork, M. (2010). Web crawling. *Foundations and Trends in Information Retrieval, 4*(3), 175–246.

Österreich Werbung. (2011) Retrieved from http://www.austriatourism.com/xxl/_site/int-de/_area/465222/_subArea/465274/_subArea2/476161/index.html, 15.04.2011

Pforr, C. (2005). Tourism policy in the making: An Australian network study. *Annals of Tourism Research, 33*(1), 87–108.

PhoCusWright. (2011). *Global online travel overview* (2nd ed.). Sherman, CT: Author.

Piazzi, R. (2011). *Analysis of the network structure of the Austrian eTourism market and identification of the central points.* Master thesis at the Vienna University of Technology.

Poon, A. (1993). *Tourism, technology and competitive strategies*. Wallingford, UK: CAB International.

Pröll, B., & Retschitzegger, W. (2000). Discovering next generation tourism information systems: A tour on TIScover. *Journal of Travel Research, 39*(2), 182–191.

Schertler, W., Schmid, B., Tjoa, A M., & Werthner, H. (1994). *Information and communication technologies in tourism*. New York: Springer/Wien.

Scott, N., Cooper, C., & Baggio R. (2008). Destination Networks: Four Australian Cases. *Annals of Tourism Research, 35*(1), 169–188.

Sheldon, P. J. (1997). *Tourism information technology*. Oxon, UK: CAB International.

Shih, H.-Y. (2006). Network characteristics of drive tourism destinations: An application of network analysis in tourism. *Tourism Management 27*(5), 1029–1039.

Vernon, J., Essex, S., Pinder, D., & Curry, K. (2005). Collaborative policymaking: Local sustainable projects. *Annals of Tourism Research*, 32(2), 325–345.

Werthner, H., & Klein, S. (1999). *Information technology and tourism: A challenging relationship*. Vienna: Springer.