http://arxiv.org/abs/physics/0606043
online: 5 Jun 2006

The Web Graph of a Tourism System

Rodolfo Baggio

Master in Economics and Tourism, Bocconi University, Milan, Italy

School of Tourism and Leisure Management, The University of Queensland, Australia

Present address:

Master in Economics and Tourism - Bocconi University

via Sarfatti, 25

20136 Milan, Italy

Phone: +39 02 5836 5437

Fax: +39 02 5836 5439

Email: rodolfo.baggio@unibocconi.it

Abstract

The website network of a tourism destination is examined. The main statistical characteristics of the underlying graph are calculated. The topology of the network is similar to the one characterizing similar systems. However, some differences are found, mainly due to the relatively poor connectivity among the vertices of the network.

PACS: 89.20.Hh; 89.75.Da; 89.75.Fb; 89.75.Hc

Keywords: web, internet, scale-free networks, tourism systems

1. Introduction

The vast catalogue of studies on complex networks which has been compiled in the last few years is missing an important component: the tourism industry (see the reviews [1] [2] [3] [4]). This paper aims at filling this gap and presenting an investigation on the websites network of a tourism destination.

In the second part of last century, tourism has become probably the largest economic sector of the

World economy. In broad terms, according to the last estimates of the World Travel and Tourism Council [5], it is expected to total 10.6% of GDP and more than 200 million jobs. And the growth is thought to continue for the next years at a rate close to 5% per year.

The boundaries of the tourism and travel industry are fairly indefinite. It brings together, and strongly influences, segments from a number of different activities with a wide variety of products and services exhibiting very little homogeneity.

There is probably no other economic sector with such a diversity and this has raised the question of whether tourism and travel should even be classified as an industry by itself, in the traditional sense of manufacturing or trade.

In the last decade, then, tourism has become an extremely dynamic system [6]. In the last years, the globalisation enabled by technology development and by less expensive travel costs has greatly increased competition. The intensified marketing efforts of all tourism organisations has led to a more effective approach: the destination management approach [7].

The spectrum of definitions describing a destination is extremely broad, and there are many difficulties in setting clear boundaries to a Tourism Destination (TD). In general, every place for a holiday, every place to visit may be considered a destination. In broad terms, a tourism destination may be intended as a geographical area that offers the tourist the opportunity of exploiting a variety of attractions and services [8].

The supply is provided by a more or less definite set of private and public organizations and companies that, in the ideal case, collaborate and coordinate their efforts in order to maximize their profits and to assure a balanced and sustainable progress of the local resources, avoiding any possible threat to people and environment [9].

Apart from the definition problems, a TD is the archetype of a complex adaptive system (CAS). It shares many (if not all) of the characteristics usually associated with a CAS: non-linear relationships among the system components, self-organization of the structures, robustness to external shocks [10].

The intangibility of a tourism product stresses its information component so that it is always described as an “information intensive” one [11]. Therefore it is not surprising that the relationship between tourism and information technology is very strict. This sort of genetic tie originates at the dawn of the electronic computer history, at the end of the 1960s with the deployment of the first computerized reservation systems. The evolution of the two industries has almost always been parallel and, not unexpectedly, the Internet sees in travel and tourism organizations its most numerous and important component.

The Internet age has allowed the development of new ways for producing and distributing tourism services. Web-based approaches and technologies are helping tourism suppliers and agencies reduce service costs and attract customers [9]. A website looks to be a major (and, probably, it will be the only one in the future) tool to conduct business in the tourism field. According to PhoCusWright's estimates [12], for example, online sales in the U.S. will be 35% of the travel market in 2005 and more than 50% in 2006.

2. The web space of a tourism destination

The websites of a tourism destination have been analyzed. The destination is the island of Elba, off the coast of Tuscany, Italy, in the heart of the western Mediterranean Sea. It is an important environmental resource; its geographic position, temperate climate, and the variety and beauty of its landscapes, coast and sea, make it a tourist destination famous all over the world.

As for many other destinations, the Web has become, in the last years, an important means of promotion and of commercialization for the whole community of local tourism operators.

The elements of the network examined are the websites belonging to the "classical" tourism operators: accommodation (hotels, residences, camping sites etc.), intermediaries (travel agencies and tour operators), transport, regulation bodies, services etc. The whole size of the network is not huge, it comprises 468 elements, but of a size which can be considered sufficient to show meaningfully the graph's statistical properties [2] [13].

The websites have been analyzed considering them as the nodes of a complex network. Links among the websites have been counted by using a simple crawler, complementing the data obtained with a visual inspection of the websites. Besides that, links connecting the Elban websites, in both directions, to the rest of the Web have been identified. In what follows an Elban website is one belonging to a tourism operator based on Elba island, E denotes the graph defined by the edges connecting only Elban websites, W is the set of links between Elban websites and the rest of the World Wide Web.

All links are considered of directed nature. Figure 1 (drawn with Pajek [14]) gives a graphical representation of the E network thus obtained.

<< Figure 1 >>

3. The results of the statistical analysis

The E network is rather sparse, its link density is $d = 0.002$ and almost 21% of the websites have no connection whatsoever with other sites. The diameter is $D = 11$, the average distance $L = 4.5$ and the global clustering coefficient C = 0.003.

<< Figure 2 >>

<< Figure 3 >>

Key parameters characterizing the structure of a directed network are the in-degree (k_{in}) and out-degree (k_{out}) distributions. Both the E and W networks, as depicted in fig. 2 and 3, display an almost perfect power law decay $P(k) \sim k^{-\gamma}$. The cumulative degree distributions are shown in Figure 2 and Figure 3. The exponents calculated for the networks are listed in table 1

<<Table 1 >>

All the exponents (with the exception of γ_{in} for the E network) are lower than those typically measured for the Web ($\gamma_{in} \sim 2.1$ and $\gamma_{out} \sim 2.7$ [15]) showing thus a more skewed and "sparse" distribution which may be seen as a very low propensity to reference the external world..

A spectral analysis confirms the main topological characteristics of the E network.

The shape of the spectral density $\rho(\kappa)$ of a graph is known to be an indicator of the topological properties of a network [16] [17]. For random graphs with a giant connected component it converges to a semicircle following Wigner's law [18]. All other cases see different distributions: a highly skewed multipeaked structure for a small-world network and a triangular shape for scale-free graphs. As Figure 4a shows, the power law behaviour of the degree distribution for the E network is evident.

The $\rho(\lambda)$ spectral distribution of the Laplacian matrix associated to the E network is shown in Figure 4b. It can be noted that a high number of the Laplacian eigenvalues are null. This is an indication [19] of the scarce connectedness of the network. The multiplicity of the null eigenvalue, in fact, corresponds to the number of the connected components of the network.

The general topological properties of the World Wide Web have been studied by a number of authors. In particular, it has been possible to highlight a complex structure of its components (web pages or websites). According to Broder et al. [20] the structure has a bow-tie shape, in which it is possible to recognise a number of components characterised by their connectivity characteristics. The model, widely accepted, sees a strongly connected component (SCC), formed by all pages mutually connected by a directed link, an in-component (IN) and an out-component (OUT), formed by nodes connected to the SCC in a monodirectional way plus a series of secondary structures such as TENDRILS, containing pages that cannot reach the SCC and cannot be reached from it, TUBES, directly linking the IN and OUT parts without crossing the SCC and some disconnected elements (DCC), similar to isolated islands, with no connection at all to the other components.

This structure has also been identified in several subnetworks of the whole Web, thus confirming the hypothesis of a self-similar configuration for the Web [21].

The Elban websites network, besides the general low connectivity among its websites, still exhibits a bow-tie structure. Table 2 displays the estimated proportions for the bow-tie components along with the values accepted for the whole Web.

<< Table 2 >>

Local and global efficiency [22] are: $E_{loc} = 0.0145$ and $E_{glob} = 0.16981$. These values are sensibly lower than those found for similar systems. The local efficiency value confirms the poor clusterisation of the network.

If we think of the efficiency as a measure of how well information is exchanged over the network [22], both at a local and at a global scale, this result induces some apprehension with regard to the capability of achieving an effective cooperation among the various elements of the tourism network.

Thinking of a web space as a virtual counterpart of an important economic and social system, an interesting quantity to measure is the assortativity coefficient. Many other social networks studied so far are characterized by assortative mixing, i.e. the edges preferably connect vertices with similar degrees, and this is usually interpreted as a sign of collaboration among the actors of the network [3] [23] [24] [25]. The coefficient can be calculated as the Pearson correlation coefficient between the degrees of adjacent vertices in the network [24]. For the Elban websites network (E network) the coefficient is: -0.101 ± 0.094 (standard error is computed as in [24]). This quite low value can be seen, again, as an index of a certain reluctance to form collaborative groups among the tourism operators (at least on the web), fact that can also be confirmed by the relatively poor general connectivity of the system.

4. Summary

The network formed by the websites of a tourism destination has been analysed, as part of a larger project on the structure of the relationships existing among the stakeholders of such systems. The statistical mechanics tools developed in the last years for this purpose [4] have been used to derive the main topological characteristics of this network.

The results show a general agreement with similar results [20] [21] [26] [27], obtained by studying the Web and websites configurations. This may reinforce the idea of a substantial self-similarity in the structure of the Web space [21]. Some of the values, though, show different characteristics: basically a lower connectivity and a higher sparseness.

It is difficult to explain these differences as the history of the system's evolution is not known.

Nevertheless, if we accept the idea that, given the wide-reaching diffusion of the Internet, a web network is a close representation of the social network formed by the publishers of the websites [28], then we see in the results presented here a faithful representation of the relationships among the actors of a typical tourism destination. The fragmented nature of the tourism industry, for the diversity of activities and organisations involved, is one of its natural traits [29], and the lack of cooperation among the actors of these networks is equally well known. On the other hand, this much sought-after capacity to team up is considered to be an important element for the success of a destination [30].

The outcomes presented here show how the statistical analysis of networks can render quite faithfully the structure of a peculiar social network. This is, at author's knowledge, one of the very first attempts to use these techniques in this field. Further work is under way to explain the dynamical evolution of a tourism system and to relate it with the methods traditionally used for its analysis. Moreover, the application of simulation algorithms can suggest modifications to the structure of the network in order to optimise its features and behaviours.

5. References

[1] R. Albert, and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev, and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] S. Boccaletti *et al.*, Phys. Rep. **424**, 175 (2006).
[5] WTTC, *The 2005 Travel & Tourism Economic Research* (World Travel & Tourism Council, London, 2005).
[6] B. H. Farrell, and L. Twining-Ward, Annals of Tourism Research **31**, 274 (2004).
[7] J. R. B. Ritchie, and G. I. Crouch, *The Competitive Destination: A Sustainable Tourism Perspective* (CABI Publishing, Oxon, UK, 2003).
[8] J. Jafari, *Encyclopedia of Tourism* (Routledge, London, 2000).
[9] D. Buhalis, *eTourism: Information technology for strategic tourism management* (Pearson/Prentice-Hall, Harlow, UK, 2003).
[10] J. H. Holland, *Hidden Order: How Adaptation Builds Complexity* (Helix Books, Reading, MA, 1995).
[11] A. Poon, *Tourism, technology and competitive strategies* (CAB International, Oxon, UK, 1993).
[12] PhoCusWright, *Online Travel Overview: Market Size and Forecasts 2004-2006* (PhoCusWright, Sherman, CT, 2005).
[13] R. V. Solé *et al.*, Complexity **8**, 20 (2000).
[14] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek* (Cambridge University Press, Cambridge, 2005).
[15] R. Pastor-Satorras, and A. Vespignani, *Evolution and structure of the Internet - A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2004).
[16] I. J. Farkas *et al.*, Phys. Rev. E **64**, 026704 (2001).
[17] K.-I. Goh, B. Kahn, and D. Kim, Phys. Rev. E **64**, 051903 (2001).
[18] E. P. Wigner, Ann. Math. **67**, 325 (1958).
[19] B. Mohar, in *Graph Theory, Combinatorics, and Applications*, edited by Y. Alavi *et al.* (Wiley, New York, 1991), pp. 871.

[20] A. Z. Broder *et al.*, Computer Networks **33**, 309 (2000).
[21] S. Dill *et al.*, ACM TOIT **2**, 205 (2002).
[22] V. Latora, and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).
[23] M. Catanzaro, G. Caldarelli, and L. Pietronero, Phys. Rev. E **70**, 037101 (2004).
[24] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2002).
[25] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[26] J. M. Kleinberg *et al.*, in *Lecture Notes in Computer Science*, edited by T. Asano *et al.* (Springer, Berlin, 1999), pp. 1.
[27] R. Kumar *et al.*, in 9th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems Dallas, TX, 2000), pp. 1.
[28] B. Wellman, Science **293**, 2031 (2001).
[29] S. Williams, *Tourism Geography* (Routledge, London, 1998).
[30] B. Bramwell, and B. Lane, *Tourism Collaboration and Partnerships: Politics Practice and Sustainability* (Channel View Publications, Clevedon, UK, 2000).

Tables:

Table 1. Degree distribution exponents

Network	Degree distribution exponent	
E	γ_{out}	1.89
	γ_{in}	2.96
W	γ_{out}	1.86
	γ_{in}	1.72

Table 2. Relative size of the components for the E network and the Web ([20] [21]) in the hypothesis of a bow-tie structure

	E Network (%)	Web (%)
SCC	3.4	28.0
IN	2.1	21.0
OUT	52.4	21.0
TENDRILS	15.6	21.0
TUBE	1.3	9.0
DCC	25.2	

Figures:

Figure 1 : The network of the Elban tourism websites (E network)

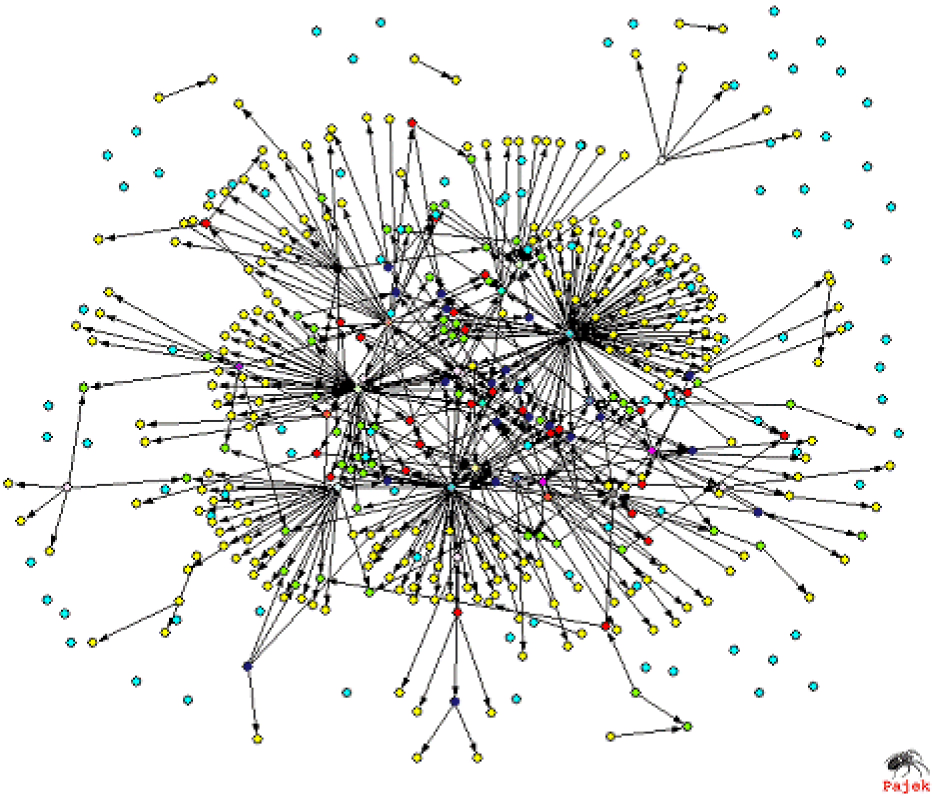

Figure 2: Cumulative in-degree (k_{in}) and out-degree (k_{out}) distributions for the E network of the tourism websites of the Elba island

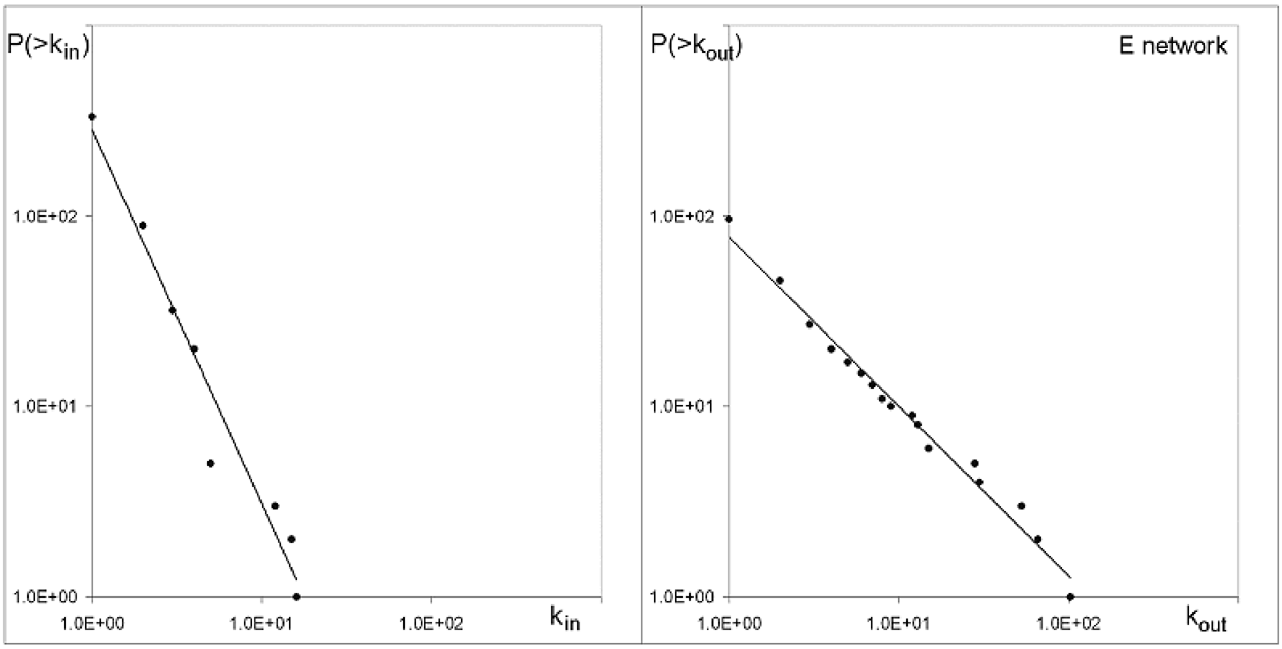

Figure 3: Cumulative in-degree (k_{in}) and out-degree (k_{out}) distributions for the W network of the tourism websites of the Elba island

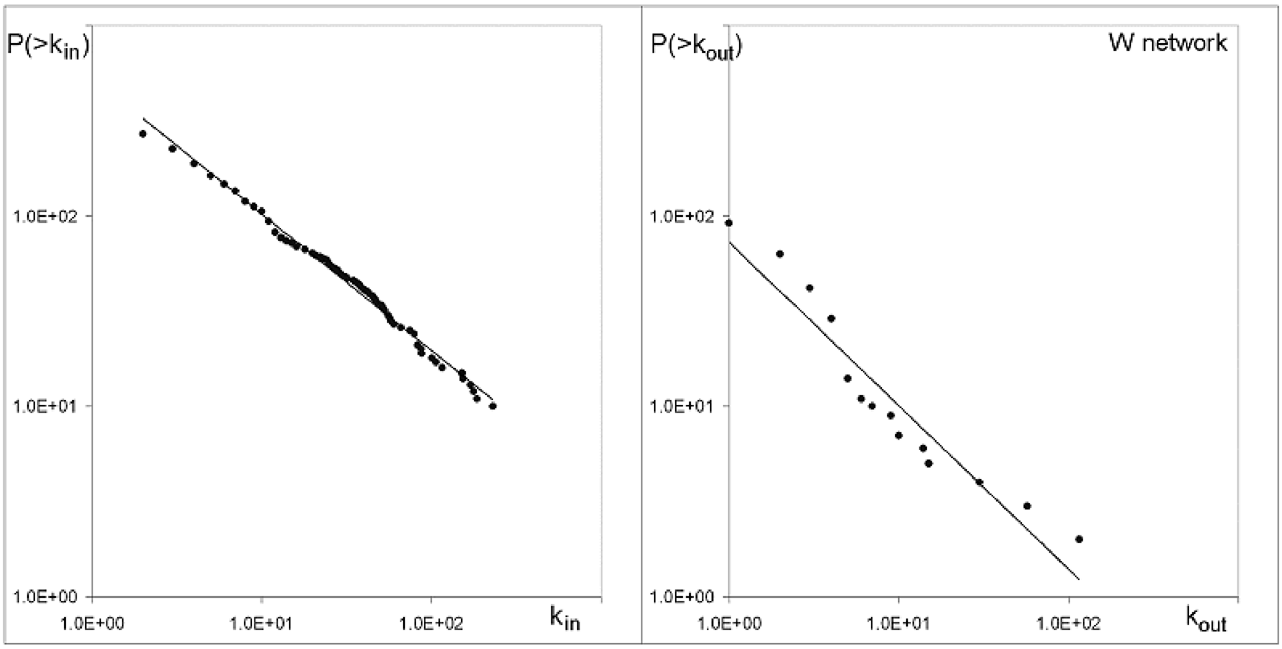

Figure 4: Spectral density of the adjacency matrix (a) and of Laplacian matrix (b) for the E network of the tourism websites of the Elba island