Network Science and Tourism – the state of the art

Rodolfo Baggio

Master in Economics and Tourism, Bocconi University, Milan, Italy
and
National Research Tomsk Polytechnic University, Tomsk, Russia

rodolfo.baggio@unibocconi.it

Tourism Review (2017), 72(1), 120-131

Abstract

Purpose
In recent years network science has become a dynamic and promising discipline. This contribution provides a brief summary of the subject and the of the application to the tourism domain.

Design/methodology/approach
The paper is based on a non-exhaustive survey of the literature.

Findings
The state-of-the-art of network science in tourism is explored and discussed, together with possible future developments.

Research limitations/ Implications
The paper uses a limited set of works, those deemed the most significant to sketch the situation. The choice might be subjective, but the overall picture is clear. Give what accomplished so far, the methods of network science seem interesting both for their theoretical and practical outcomes. In essence they provide a better and more objective view on the structural and dynamic characteristics of the tourism phenomenon and of the different tourism systems and components.

Originality/value
The paper critically reflects on the state of network science and its application to the tourism domain. Even without claiming to be complete, the paper takes a general perspective approach rather than examining single topics or issues.

Article type: General review

Keywords: network science, network analysis, tourism, hospitality, leisure, complex systems
Introduction

There is a natural match between the study of many natural and artificial systems and that set of techniques and approaches that are today know as network science. The match comes from the idea that all these systems can be modelled as ensembles of elements tied by some kind of connection that may exist between any two of them. Once depicted in this way, then, the methods developed in the framework of network science allow examining the structural and dynamic characteristics of the system under study, and to discover its reactions to modifications and adaptations that may occur, whether due to internal or external actions.

Network analytic methods are fundamentally based on the mathematical techniques of graph theory, an old branch, and one of the few specialities for which we can set a birth date, that of the famous paper, dated 1736, by Leonhard Euler on the solution of the Königsberg seven-bridge puzzle (Euler, 1736). After that, a steady but ever growing number of works have analysed graphs, the mathematical abstractions of a network, and studied the design, analysis, and applications of algorithms that compute with and on them.

The modern, and now-classic, random graph theory was established by Erdös–Rényi in the late 1950s (Erdös and Rényi, 1959, Erdös and Rényi, 1960), and has seen an incredible growth in the last decade, mainly driven by the important technological developments that, at the end of last century, made available a large quantity of relatively accessible data on the Web, and an usable distributed computing power, needed for the treatment of the data collected. The birth of this new era can be traced back to three fundamental papers (Barabási and Albert, 1999, Faloutsos et al., 1999, Watts and Strogatz, 1998) that, through the study of the Internet and the Web, uncovered a number of interesting properties that the previous tradition of social network analysis, even if much older (the common origin is reputed to be the work of Moreno in 1934 with his sociograms), could not detect, mainly for the small size (typically a few dozen elements) of the systems studied.

Although no clear definition of network science exists, the objectives and the application domains are relatively well identified. In essence network science aims at identifying possible unifying principles that could help describing, under some generic and common rules, the structural features that are being uncovered, and modelling the resulting dynamic behaviours in order to explain what can be actually seen and experienced from the observation of the systems under study. As Newman et al. (2011: 4) state, network science is concerned with empirical as well as theoretical questions, and aims, ultimately at least, to understand networks not just as topological objects, but also as the framework upon which distributed dynamical systems are built.

Besides these basic features, the strong theoretical framework in which the recent developments of network analytic methods are embedded, statistical physics, allows drawing analogic comparisons between systems of different nature and assess, independently from the specific traits of the network elements, the similarities existing in the structural characteristics together with the influence these have on the dynamics of and on the various networks.

When it comes to tourism, it is easy to realise that the basic distinctive features, beyond the difficulties in precisely defining the phenomenon and its main actors, are the various types of relationships (social, economic, operational, informational etc.) that can be observed. It is therefore natural to think that these methods could be of help in understanding the structure and the behaviours of the whole system and of its constituents, and in recognising that its complexity can be better explored by using this approach (Baggio and Sainaghi, 2011, Baggio et al., 2010, Baggio, 2008).
Objective of this paper is not so much that of providing a complete review of what network science has accomplished in the tourism domain (for this, two good recent works are: Casanueva et al., 2016, and van der Zee and Vanneste, 2015), but rather to present the state-of-the-art of this discipline and how it has contributed to the study of tourism and its components, discussing also an agenda for possible future explorations. In the rest of this contribution, after a brief summary of what network science is, the main developments and outcomes of network science and of the applications in the tourism domain are discussed along with their theoretical and practical contributions. A survey of the open issues and of the possible future developments closes the paper.

Network science in a nutshell

Network analytic methods aim at enumerating, mapping and analysing the patterns of connections between the elements of any system, be it a natural, artificial, social or economic, that can be modelled as an ensemble of distinct elements or actors (the nodes or vertices of the network) that are connected by some kind of relationship existing between them (the links or edges), that can also carry a weight (cost, importance, etc.) or be directional. The field draws on theories and techniques that have been developed in a wide range of disciplines, from mathematics, physics, computer science, sociology, making it a splendid example of how a multidisciplinary effort can provide intriguing, stimulating and valuable outcomes that in many cases give non-intuitive and unexpected views of the object of study (for a good collection of examples in numerous different fields see: da Fontoura Costa et al., 2011, Barabási, 2016, Newman, 2010).

The main methods used for the study of a network are rooted in the mathematical techniques of graph theory (Bondy and Murty, 1982, Diestel, 2016), but have had a number of improvements, variations and expansions, mainly for what concerns the study of the dynamic characteristics of a network and of the processes that can develop when a network is the substrate. Many metrics have been defined for measuring these features, for providing rigorous quantitative assessments, and for describing dynamic statistical models for the evolution of the systems considered (see for example Goldenberg et al., 2010, Barabási, 2016, da Fontoura Costa et al., 2007, Boccaletti et al., 2006).

It must be noted here that many works have also reinforced the idea that purely qualitative or purely quantitative methods are meaningless when a complex systems is the object of study. In fact, despite the many discussions and distinctions, the two views are not only complementary, but practically mandatory if we want to fully understand the systems and be able to draw meaningful outcomes (Kelman et al., 2016, Baggio and Del Chiappa, 2016, Jørgensen, 2016). The network metaphor might seem fuzzy, but the fuzziness of this metaphor fades away when hard data are used, thus providing a complete, theoretically rigorous and elegant analysis (Thompson, 2003, Wellman, 2002).

Even from a purely quantitative point of view, no single set of metrics provides a clear idea of the main characteristics associated with a complex network. It has been shown, in fact, that the properties of a complex system are distributed, and may change, over a range of scales, and that our assessments depend on the level of detail that we use in describing them. An analysis conducted at a single level of details could turn out incomplete, and the outcomes could be dependent on the selected granularity (Marchiori and Possamai, 2015).

Three levels are typically used for a full study:

- **microscopic**: in which the properties of the single elements (nodes) are measured. The most important quantities calculated are the number of connections each node has (degree or strength,
the sum of the links’ weights if present); the density of links of the node’s ego network (that formed by its immediate neighbours) called clustering coefficient; the extent to which a node can reach any other node in the network and the distance from them (closeness); the role played as connector (bridge or bottleneck) between different regions (betweenness); or a metric derived from the eigenvectors of the matrix which represents the network (PageRank is an example), useful to better measure the overall influence of a node. The normalised version of these metrics is usually called centrality (e.g. degree centrality, betweenness centrality etc.);

- **mesoscopic**: the intermediate structure of the network, normally derived from a modularity analysis. This is a stochastic procedure that identifies areas in the network in which the nodes are more densely connected between them (modules or communities) than to other areas. Many algorithms exist for this purpose (Fortunato, 2010) that differ in terms of the features they can consider (directionality, weights) or the resolution power (capacity to identify smaller or overlapping modules). The modularity is expressed by a modularity index. Hierarchical structures can also be detected with the help of this type of analysis;

- **macroscopic**: the global topological characteristics of the network. The most common and used way to render the global topology is the statistical distribution of the degrees (degree distribution). Its shape, in fact, signals well the general properties of the network, its complex characteristics and its capacity to respond to a number of dynamic processes. Many networks have been found to have a power-law distribution of the degrees, that is they have a few nodes with many connections (the hubs) and many with few links. A power law means that there is no normal or typical degree or scale, hence the **scale-free** name given to these networks. Other measures used to describe the macroscopic characteristics are the average path length (average distance between any two nodes), the diameter (the longest distance between any two nodes), the correlations existing between the distributions of different metrics, and the average values of the microscopic metrics over the whole network.

Two more issues are important. One concerns the visualisation of networked systems. Here too many methods have been proposed that provide different renderings. Most of these methods consider the links as **springs** and let the system settle in an equilibrium state. Visualisation can be of help to spot regularities, structures and other peculiarities, but, obviously, its importance decreases when the network is large and the visual representation becomes too crammed (Tamassia, 2013).

The second concerns the collection of the data needed. When the network is an artificial or some natural system (a computer network, a transportation system or the food web in a closed ecosystem) the collection, even if not always fast, is relatively straightforward. However, when social or economic systems are concerned the task can be quite complicated. The main problem is in the fact that most of the quantities that can be used to characterise real networks are strongly non-normally distributed so that traditional sampling considerations can be applied with great difficulty. The same happens when the huge modern technological environments are used (online social networks or the Web), where a full enumeration is practically impossible. Different mixtures of methods are then advisable, that attempt to attain the best possible coverage by combining archive searches, electronic sources, interviews with knowledgeable informants or surveys (Christopoulos and Aubke, 2014, Marsden, 2011, Marsden, 1990, Baggio et al., 2010). An evaluation of the completeness of the data used and the impact this has on the network metrics is therefore mandatory (guidelines are provided by many studies such as (Wang et al., 2012, Kossinets, 2006, Frank, 2005, Rothenberg, 1995).
Finally it must be considered that one of the benefits of network modelling is that numerical simulations can be performed straightforwardly. It is thus possible to experiment different configurations or different dynamic processes when it would not be possible for theoretical or practical reasons, as it happens with a socio-economic system such as a tourism destination. The understanding gained in the system’s behaviour can then be used to draw assess its evolution, scenarios or sketch possible optimizations.

**Network science past and present developments**

The development of network science has been astonishing and quite fast. Looking back through the literature on the topic (good sources are Scopus, www.scopus.com, one of the major abstract and citation database of peer-reviewed literature, or ArXiv, arxiv.org, the principal repository for pre-prints in many disciplines), it is possible to loosely distinguish three phases, even if with some overlap, characterised by the common efforts of the scholars.

The first one is mainly dedicated to the study of the basic topological characteristics of a large set of networks in practically any field (see e.g. the Koblenz Network Collection at http://konect.uni-koblenz.de/), and to the development of metrics, tools and techniques for their analysis, together with the attempt to uncover their formation mechanisms (good reviews are: (Newman, 2003, Boccaletti et al., 2006, da Fontoura Costa et al., 2007).

The second phase is more focussed on the investigation of dynamic processes such as diffusion (ideas, innovation, but also epidemic diseases) or synchronisation (opinions, molecular processes, brain areas etc.) (see e.g. (Barrat et al., 2008, Pastor-Satorras et al., 2015, Arenas et al., 2008). Together with these we see more in-depth analyses of the mesoscopic structure of a network (modularity, hierarchies etc.) (Fortunato, 2010).

The third and most recent phase has started to analyse multidimensional networks: those made of networks of different types that interact between them. These are thought to better model certain types of systems such as a city (or a destinations), in which different physical (transport, roads, electricity, water supply), social (people, institutions, organisations) and economic (firms, shops, associations) or virtual (web, services, social media) layers can be identified and whose connections and interactions produce a rich dynamic behaviour. By using this approach we tend to improve our knowledge of complex natural systems and avoid simplifications that “may occasionally result in not fully capturing the details present in some real-life problems, leading even to incorrect descriptions of some phenomena that are taking place on real-world networks” (Boccaletti et al., 2014: 4). A second theme that is being developed in this phase is that of temporal networks. This research stream takes into account the evolving structure of edge activations and deactivations that can affect the dynamic behaviour of the system studied (Holme and Saramäki, 2012).

**Network science in Tourism: the main contributions**

Since the very beginning of the research activities in tourism it has been clear that the phenomenon is a complex one, involving a number of quite different actors, activities, connections, that make the field an ideal object of study for network science. Thus it is rather strange to see how these methods have been neglected for a long time. Although some qualitative investigations had appeared in various works (Cooper, 1981, Halme, 2001, Pearce, 1992, Selin and Beason, 1991, Stokowski, 1990, Tremblay, 1998), it is only in the last decade that a growing interest in the use of quantitative techniques has emerged.
Most of the initial applications concern the topological characterization and the identification of the structural peculiarities of a tourism destination (Bendale and Patterson, 2008, Presenza and Cipollina, 2009, Del Chiappa and Presenza, 2013, Grama and Baggio, 2014, Baggio et al., 2010). Although not many, these studies show common outcomes. All networks have a power-law degree distribution that confirms the complex nature of the systems; then we see complex structures with low density of connections, little clusterisation, and negative degree–degree correlations (i.e., highly connected nodes tend to link low-degree elements). These features have been interpreted as a symptom of the well-known tendency of tourism stakeholders to elude forms of collaboration or cooperation. The related metrics (clustering and assortativity coefficients) have thus been proposed as quantitative measurements for these characteristics (Baggio, 2007).

A better knowledge of the structural characteristics of a destination, or of a tourism supply chain, is an important result, because the identification of strategic weaknesses in the cohesiveness of the destination can be useful for policy and governance bodies (Erkuş-Öztürk and Eraydin, 2010, Timur and Getz, 2008, Tran et al., 2016). Same can be said about the creativity and innovation potential of a destination, that can be related and assessed through a quantitative examination (Baggio, 2014). The overall topological analysis of a destination has also shown how the self-organisation mechanism produce non-trivial mesoscopic structures that can be different from those usually considered, that are based on stakeholders’ typologies or geographical location (Baggio, 2011a). Fact that has an obvious interest for all those engaged in the governance, management or supervision of tourism systems. Moreover, a network representation has allowed assessing the resilience (i.e. the ability of a system to cope with externally or internally induced changes) to climate change, which can be particularly important for certain types of destinations (Luthe and Wyss, 2016).

One more issue for which network science has proved useful is the study of the formation and the effects of social capital in destinations or working environments (Sainaghi and Baggio, 2014, Beritelli and Laesser, 2011). Moreover, the analysis of a tourism network has also provided tools to identify the most relevant actors: those who are reputed to give the most important contribution to the tourism activities, or those who can better benefit from a good network position (Presenza and Cipollina, 2009, Cooper et al., 2009, Sainaghi and Baggio, 2014).

Rather obviously, the technological component of the tourism business has been approached in a number of ways. The virtual counterpart of tourism destinations, the websites of their stakeholders, have been investigated in several studies that have allowed assessing the level of utilization of advanced communication technologies in a destination and measure the extent to which they exploit (or waste) these resources (Piazz et al., 2012, Baggio and Antonioli Corigliano, 2009, Baggio, 2007, Ying et al., 2016). The structural integration between the virtual and the real components in a destination have been well recognized, giving more strength to the idea that a smart digital ecosystem needs to be fully considered when dealing with tourism activities at a destination (Del Chiappa and Baggio, 2015, Baggio and Del Chiappa, 2014b, Becheru et al., 2016). One important outcome of the study of the webspace of a destination is that there is a substantial topological similarity with the destination “real” network. This, coupled with considerations on the mechanisms with which corporate websites are interlinked, allows conjecturing that the Web can provide an efficient and effective way to gather a significant sample of the corresponding networked real system, thus easing the collection of data to be used for analyses and simulations (Baggio et al., 2010).
The dynamic behaviours of interest have received limited attention. Only a few investigations have been performed on the diffusion of knowledge and information (Baggio, 2011b, Baggio and Cooper, 2010) or the synchronization of opinions (Baggio and Del Chiappa, 2014a). The same can be said about the study of the temporal evolution of a tourism system.

The mobility of tourists is an central issue for understanding what can be offered them in terms of attractions or services, and can help substantially governance bodies and policy makers in making plans founded on empirical evidence rather than on pre-conceived ideas. The topic naturally translates into the network science language and has been addressed in a number of ways. Older (but not necessarily) works reconstruct tourists’ movements based on traditional surveys (Hwang et al., 2006, Shi, 2006, D’Agata et al., 2013), while newer approaches use (at least partially) the huge amount of digital traces left online by people (also known as big data). Researchers have used GPS (Global Positioning System) data (Taczanowska et al., 2014), social media sources (Hawelka et al., 2014, Agrzykov et al., 2017), or cell phone recordings (Baggio and Scaglione, 2017).

Other uses of network analytic methods concern the analysis of the tourism literature. Here several works have sketched the different interest areas of tourism researchers and highlighted the extent of their collaborations (Wang and Chen, 2014, Ye et al., 2013, Benckendorff and Zehrer, 2013). Also some different uses of network analysis have provided interesting outcomes as in the work of Sainaghi and Baggio (Sainaghi and Baggio, 2014, Baggio, 2014). They transform an overnight time series into a network finding confirmation of the complexity features of a destination and highlighting its dynamic behaviours by assessing the changes in the dynamics over a number of years.

**Tourism network analysis: where are we and where do we go?**

Network science in tourism is catching interest and the works in this area are growing in number and in quality. Looking at what has been produced so far, some considerations come to mind. The first one is that, in general, the methods used for the analysis are rather basic. Not many of the most progressed techniques available are used. Probably here there is an issue related to the use of tools. Standard software (Ucinet or Gephi, for example, among the most used) provide only basic measurements, while many of the most advanced techniques require a knowledge of some computer programming language (Python, Matlab or R have very good libraries for the analysis of networks) that are not very popular among the scholars in the tourism field. The same can be said about the dynamic analysis of networks and processes, or the most advanced topics such as temporal or multidimensional systems.

One more issue is that of data collection. Here, as said, usual sampling methods can be quite misleading as networks typically exhibit highly non-normal distributions of their topological features. Good completeness would be required, but this is extremely difficult to achieve mainly with the relational data. Here, as already pointed out, some help may come from the virtual side. In general, however, as some of the examples presented above show, multiple collections from different sources are needed. Big data, if used correctly, could be greatly helpful in filling the inevitable holes of manual collections (Török et al., 2016). In any case, even if not fully rigorously, an assessment of the completeness can be made with relative ease, together with an estimate of the effects the missing elements have on the measurements (see e.g. Kossinets, 2006, Lee et al., 2006, Wang et al., 2012).

The application of network science in the tourism domain is still at a relatively early stage of development, but from the growing number of works it appears clearly that the methods are able to provide interesting and useful outcomes for both theory and practice. For theory, because tourism is commonly considered...
lacking good theoretical foundations that, instead, would be highly useful for better understanding the whole phenomenon and its many facets and complexities (Farrell and Twining-Ward, 2004). For practice, because too much of the actions and activities in the field, and mainly for what concerns the most important component, the destination, are based more on anecdotal evidence, fashions or hearsay rather than on well-grounded analyses and scenarios, with inevitable failures or lost occasions (Beritelli et al., 2015).

The future of network science in tourism should be characterised by more numerous studies on full (for what possible) networks. This will better highlight common and different structural and dynamic properties and, by looking at destinations in different stages of their development, to sketch more realistic models for the evolutionary paths of these systems. Moreover, by examining the different external environments and macro political and socio-economic conditions in which destinations are embedded it would be possible to fine tune our comprehension of the effects of these conditions. The combined knowledge on these issues will let us prepare better and more accurate scenarios, so important for the governance and the functioning of our tourism systems.

Together with these aspects we need also a better characterisation of the different elements of the networks (both actors and relationships) under study and improve our ability to connect these with the topological features (as a recent example see Stienmetz and Fesenmaier, 2017). In fact, although interdisciplinarity and multidisciplinarity are commonly deemed crucial for tourism studies, practice shows that these are confined in a relatively close set of competencies, all belonging to the broad area of social sciences, while little presence of more technical skills exist (Volgger and Pechlaner, 2015). This separation is also quite relevant for the overall field of network science (Hidalgo, 2016), but in a field such as tourism it is even more detrimental. Therefore it is important to call for a more profound revision of the real needs and for improving, at least, our educational programs so that future generations of scholars might be better equipped for the enterprise.

Management, human resources, social capital are topics very intensely studied. Here network science can provide better weapons to deepen the level of understanding and prepare more efficient and effective tools for governing organizations (Aubke, 2014: 18): “After all, a network approach not only allows one to understand the problem management is facing, but also to understand the system that causes the problem.”

Investigations on the dynamic properties of networks and of the processes that use a network, or combinations of networks, as a substrate are definitely missing in tourism studies. Here the research needs to better face the issue and to provide models and simulations whether purely numerical or agent-based (Johnson et al., 2016) that can be prove useful, besides theoretically, also for projecting changes, tunings or assessments to tourism systems and components, and allow more efficient actions. This is even more valid when we look at the advances in the possibilities to uncover and control events and profound modifications (cascades in network parlance) that may occur and cause disruptions or hinder positive disseminations (Motter and Yang, 2017)

**Concluding remarks**

When looking at the development of network science in tourism the words of Galileo in his letter to Orso d’Elci (25 December 1617) come to mind (1832: 133-134): “…this is a complete art, although just born, based on principles and means that are new, but noble and commendable, and needs to be embraced,
cultivated and promoted, so that with exercise and time it will be possible to benefit from the fruits of which it has in itself the seeds and the roots.

Many good outcomes have been provided so far, and many more could come if tourism scholars will be able to refine and improve their analysis methods and tools. The main call here is to expand the competences and to foster a better and more intense collaboration across different disciplines and approaches in order to “open up for [that] methodological creativity” (Gummesson, 2007: 226) which is so needed in the tourism domain.

References


Euler, L. (1736), "Solutio problematis ad geometriam situs pertinentis", *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, Vol. 8, pp. 128-140.


 Moreno, J. L. (1934), *Who shall survive?*, Nervous and Mental Disorders Publishing Co., Washington, DC.

