Network analysis

Connecting the dots to understand tourism

Rodolfo Baggio

Master in Economics and Tourism and Dondena Center for Research on Social Dynamics and Public Policy - Bocconi University, Milan, Italy
and
Tomsk Polytechnic University, Tomsk, Russia
eMail: rodolfo.baggio@unibocconi.it

Forthcoming in: Egger. R.(ed.) (2022). Applied Data Science in Tourism. Interdisciplinary Approaches, Methodologies, and Applications. Springer

Keywords

network analysis, complex systems, tourism systems,

Learning Objectives

- **Understand the fundamental concepts and methods of network analysis**
- **Explain the most important issues and become familiar with the basic notation and terminology used for network analysis**
- **Develop elementary and practical network analysis skills and be able to visualize and compute main measures**
- **Show how to apply the methods of network science to unravel real-life problems and analyze real-world networks**

Abstract

Over the last years, network analytic methods have been able to provide insights into the structural and dynamic characteristics of systems and phenomena, and they are considered a natural choice when complex systems or phenomena are involved. As a result, these methods have sparked a growing interest in both the tourism and hospitality domains. This chapter will contain an introduction to the concepts, background, and methods of network analysis. After a brief introduction in which the rationale and foundations of network analysis are highlighted, the reader will be provided with a basic series of definitions of the main metrics and with the approach that needs to be followed for a good analysis. The How-To section will contain a worked example, allowing the reader to become familiar with the operative steps of conducting an analysis and interpreting the outcomes. A full research case will then be briefly described and commented on. Lastly, the chapter will conclude with a list of the most relevant and used software packages in this area of work.

Introduction and theoretical foundations

Most known and studied systems and phenomena can be classified as complex systems. Complex, in the popular language, indicates something one has difficulty fully understanding or describing; here, however, complex refers to a specific class of elements. They are characterized by having a certain number of parts, often organized with a detectable structure (Brodu, 2009; Levin, 2003; Lewin, 1999). These elements are interconnected, and the relationships that bind them are of a non-linear nature. The system exhibits a number of peculiar features, the most relevant of which include the following:

- emergence: structures and behaviors seem to appear at a global level that cannot be easily derived from single elements;
- robustness and fragility: sudden events might be easily absorbed by the system, but some seemingly insignificant shocks might disrupt it;
- self-organization: the system seems to generate structures or hierarchies autonomously and without any central guidance;
- evolutionary dynamics (adaptiveness): systems have a continuous exchange with the environment they are embedded in, and they adapt to these evolving conditions.

The net result is a fundamental unpredictability of the detailed structural and dynamic characteristics in the long term. Examples of complex adaptive systems include the patterns of birds in flight or the interactions of various life forms in an ecosystem, the behavior of consumers in a retail environment, people and groups in a community, the economy, the stock-market, the weather, earthquakes, traffic jams, the immune system, river networks, zebra stripes, sea-shell patterns, and many others. As such, tourism and tourism systems (e.g. destinations) are, unquestionably, typical complex systems. Its basic composition (elements and relationships) naturally leads to the idea of a useful representation being that of a network. A network (graph) is an abstract model in which the elements of the system are represented as dots connected by lines. A further abstraction consists of describing the network with a matrix (a.k.a. an adjacency matrix) whose elements indicate whether two nodes are connected or not. This allows for the use of powerful methods of linear algebra to calculate a wide array of measures that provide the characterizing features of the network (Barabási, 2016; Coscia, 2021; Sayama et al., 2016).

The basic idea of network science involves mapping and analyzing the patterns of relations among the elements of a system to understand its structure and, given a strong existing link, to examine its functions and the dynamic processes that may be involved. Moreover, mathematical modeling makes it possible to employ a wide array of techniques in order to simulate phenomena in cases where a real-life experiment would not be feasible due to theoretical, ethical, or practical reasons (Baggio & Baggio, 2020).

Lastly, the topological approach, which, regardless of the nature of the elements at play, takes structural features into account, allows the consideration of ensembles of objects (networks) belonging to very diverse domains and studies the possible existence of universal features that can better help to understand specific systems via analogical means. This is done based on a strong theoretical background, that of statistical physics, from which network science borrows many techniques and methodological approaches.

Network analysis in a nutshell

Formally speaking, a graph (network) is a pair $G = (V, E)$, where V is a set of vertices (nodes), and E is a set of pairs of distinct vertices, which are members of V: $E = \{(u,v) \mid u, v \in V\}$. The elements of E are called arcs, ties, links, or edges. Links can be assigned different properties, such as direction or weight (cost, intensity, duration, etc.). The basic types shown in Figure 1 below which also contain their matrix representation (adjacency matrices).

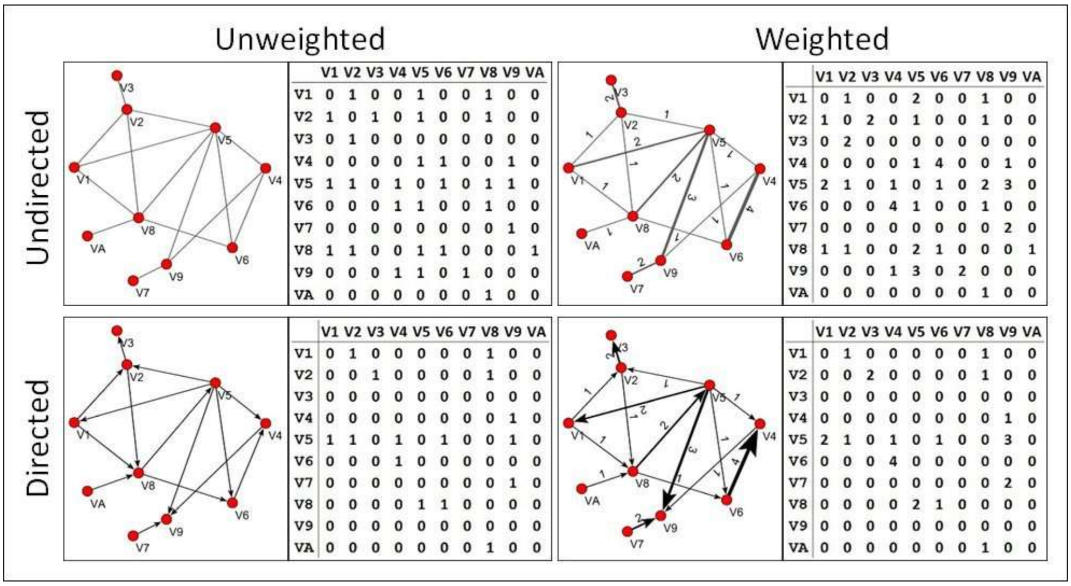


Figure 1. Network types

A special type of network is that of bipartite (a.k.a. 2-mode or affiliation network) in which the nodes can be divided into two disjoint and independent sets such that the edges only connect members of different sets. Examples may include authors-papers, affiliates-groups, topics-documents, and so on. For those interested, a complete survey of this network type can be found in the works of Pavlopoulos et al. (2018) and Guillaume and Latapy (2006).

Recent literature has provided a wealth of measures that can be used to characterize a network (Barabási, 2016; da Fontoura Costa et al., 2007). Among these, the most relevant and often used are the following:

- *density*: the portion of potential connections in a network that are actually present;
- *degree*: the number of links each node has; and *degree distribution*, the statistical distribution of the number (and sometimes the type) of the linkages among the network elements;
- *assortativity*: the correlation between the degrees of neighboring nodes;
- *average path length*: the mean distance (number of links) between any two nodes and *diameter*, the maximal shortest path connecting any two nodes;
- *closeness*: the mean weighted distance (i.e. the shortest path) between a node and all other nodes reachable from it;
- *betweenness*: the extent to which a node falls between others on the shortest paths connecting them;
- *clustering coefficient*: the concentration of connections of a node’s neighbors – it provides a measure of the heterogeneity of the local density of links;
- *eigenvector*: calculated by using the matrix representation of the network and its principal eigenvector and based on the idea that a relationship to a more interconnected node contributes to its own centrality to a greater extent than a relationship with a less efficient interconnected node. One variation of this measure is the well-known *PageRank*.

A network study starts with collecting the necessary data (i.e. nodes and links). When tourism systems (e.g. destinations) are involved, the nodes typically represent the different stakeholders of the system (hotels, restaurants, service companies, travel agencies, public bodies, etc.). The links may be collected using different methods, for instance, surveys (explicitly asking participants for and about their connections), websites' hyperlinks between companies, listings from associations or consortia, official records on co-ownership, and so on. Frequency of connections or perceived importance can be used as weights for the edges. Once a network has been built, the study makes use of suitable software packages or libraries for some programming languages in order to derive the various measures that are typically used for assessing the system's features at three levels of analysis:

- *individual (microscopic) level*: refers to the specific nodal properties such as degree, betweenness, closeness, clustering coefficient, and so on. Normalized versions of these metrics are usually known as centralities (e.g. degree centrality, betweenness centrality, etc.);
- *intermediate (mesoscopic) level*: aims at highlighting the possible modular structure of a network. These modules (communities or clusters) are formed by nodes that are more densely connected between themselves than to the rest of the network (Figure 2). The quality of the division into modules is measured by a modularity index. Several algorithms allow for the detection of these clusters (Fortunato, 2010); for instance, one of the most used and reliable algorithms is that proposed by Traag et al. (2019), also known as the *Leiden algorithm*. Built upon previous techniques, this iterative algorithm recursively assigns nodes to different groups until all elements are locally optimally assigned to a partition (i.e. the modularity index is maximized), providing communities that are guaranteed to be connected. Hierarchical structures can also be revealed by using similar algorithms;

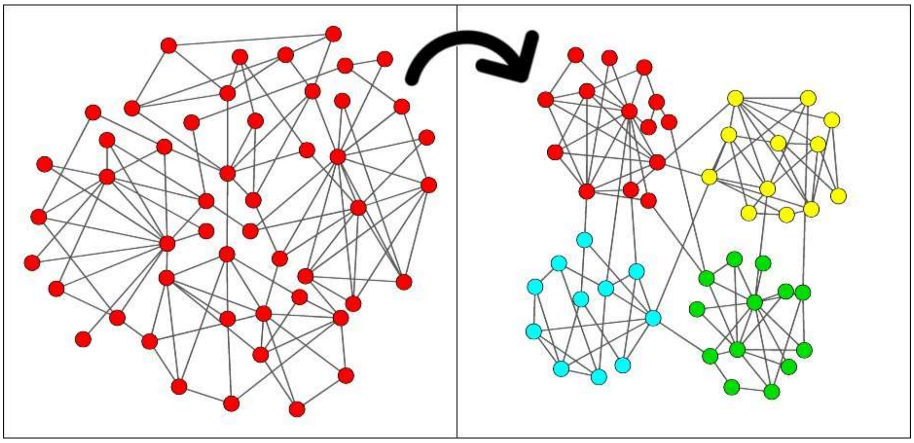


Figure 2. A network and its communities

- *global (macroscopic) level*: describes the overall structure operationalized by quantities such as density, average path length, diameter, etc. The most important and common measurement for describing the topology (structure) of a network is the probability distribution of the degrees, P(k) (degree distribution). Its mathematical form hints at the general features of the network, its complexity, and its behavior when subject to a dynamic process. For many real networks, typical degree distribution has a power-law shape ($P(k) \sim k^{-\alpha}$, see Figure 3); that is, few nodes have many connections (i.e. hubs), while many others only have a few links. The degree distribution is also an indicator of possible mechanisms for the formation and evolution of the network (Coscia, 2021; Newman et al., 2011). Other measures used to describe macroscopic characteristics are the existing correlations between the distributions of different metrics as well as the average values of the microscopic metrics over the whole network.

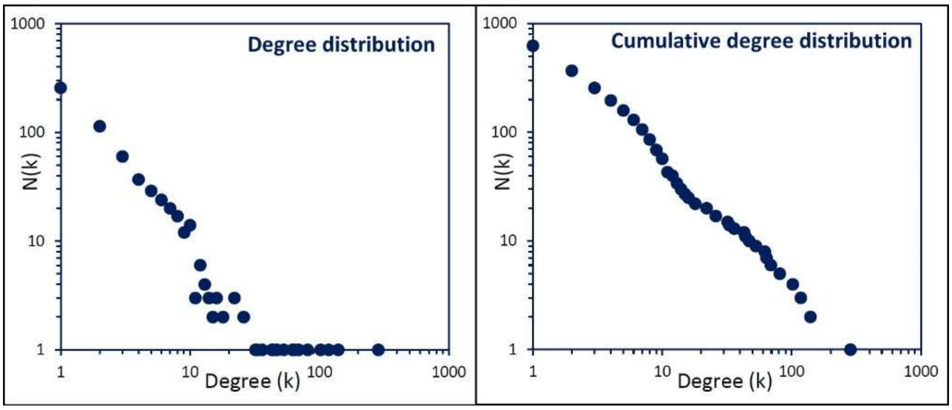


Figure 3. A degree distribution with its cumulative version (axes are logarithmic to better show the long tailed shape of the distributions)

How-to-Section

Network analysis naturally leads the researcher to adopt a broad systemic view and to focus on general issues that concern the area or the problem under investigation. In many cases, the use of these methods has resulted in outcomes that can be considered counterintuitive or not sufficiently highlighted (see e.g. Baggio, 2011). As for any other inquiry, it is important to start with a clear idea of the objectives of the work since these may affect the choices concerning the definition of the different elements, the techniques, and the metrics to be used. Normally, a good scan of the literature can provide valuable suggestions in this regard. From these preliminary explorations, a conceptual map can be drawn containing all the elements and steps needed (see e.g. Baggio & Baggio, 2020).

Once a research question has been determined, the next step consists of defining the elements of the network: the nodes and links. Usually, this process is a relatively straightforward task for what concerns the nodes; they can be, for example, the stakeholders (firms, associations, public bodies, etc.) at a destination, employees in a company, papers published, or individuals. Together with the *names* of the entities identified as nodes, it is also recommended to collect some attributes that describe them (e.g. location, size, type of business, type of entity, etc.), which can then be used for later comparisons. Links are uncovered, as previously stated above, by using a survey in which respondents are asked to indicate their main connections, public records, listings for groups, and so on. Here, too, an evaluation of the relevance (e.g. importance, cost, speed, frequency of contacts, etc.) can then be used to weigh the links, if need be. These weights can then be rendered using a suitable scale, which also allows for the use of some *qualitative* features. All in all, it is essential to try to be as complete as possible. In fact, the distributions of practically all the network's characteristics (degree, closeness, betweenness, etc.) are strongly skewed, showing long tails, and, therefore, all the *usual* sampling considerations (designed for almost-normal distributions) do not apply. Verification of this completeness can be confirmed with good knowledge of the analyzed domain and by resorting to a few interviews with some knowledgeable informants.

Ideally, all the data points should be organized into a couple of tables, as shown in Figure 4 below.

Nodes

ID	Name	Code	Type	Location	Size
1	Acme Hotels	H001	Hotel	A	Medium
2	Globex Hotel	H002	Hotel	A	Medium
3	Soylent agency	A001	Travel agency	B	Small
4	Blue Cat Hotel	H003	Hotel	B	Small
5	Umbrella consortium	G001	Association	A	Medium
6	Hooli Buses	T001	Transports	C	Small
7	Vehement association	G002	Association	C	Medium
8	Tourism board	P001	Public	A	Medium
9	Relaxicab	T002	Transports	B	Large
10	InGen	S001	Services	C	Small

Links

Source	Target	SourceID	TargetID	Type	Weight
H001	P001	1	8	Undirected	1
H002	G002	2	7	Undirected	1
H003	S001	4	10	Undirected	2
G001	H001	5	1	Undirected	1
T001	H003	6	4	Undirected	3
G002	T001	7	6	Undirected	2
P001	T002	8	9	Undirected	1
T002	G001	9	5	Undirected	1
H001	T002	1	9	Undirected	1
H001	G002	1	7	Undirected	2
A001	S001	3	10	Undirected	2
P001	T001	8	6	Undirected	3
P001	H003	8	4	Undirected	1
G002	S001	7	10	Undirected	1

Figure 4. Sample data collection table

The use of a tool such as Excel allows for the organization of the data and can easily transform the data into the format requested by the software chosen for the analysis (often, as for Gephi, a csv file).

Once the network has been obtained, a suitable software (see “Available Software”) can calculate all the desired metrics. Software packages, such as Gephi, Unicet, Pajek, etc., provide functions for basic analyses and have little (or limited) support when special network features are involved (e.g. bipartite networks, link weights, directionality of links, etc.). More advanced methods, or full treatment of special cases, as well as dynamic simulations, need to be addressed using more efficient libraries such as those implemented in Python (NetworkX, igraph, etc.), R (igraph, sna, tnet, etc.), or MATLAB.

A worked example

For this example, the software Gephi will be used, which is a good choice for an initial approach to network analysis. The reader is advised to follow the tutorials provided online in order to better understand and become familiar with the basic functioning thereof (https://gephi.org/users/)[1].

Let us consider the network of a small tourism destination in which the nodes are the companies and associations in the area. They are assigned an attribute (Biz) that codes their main business activity (Association: ASS; Hotel: HOT; Other Accommodation: OTH; Restaurant: RES; Other services: SRV; Travel Agency: TVA). The links represent any type of collaboration or relation between two entities, and the network is symmetric and unweighted. The objective is thus to analyze this network and highlight the most important items and how they cluster together in order to assess the collaborative atmosphere at the destination.

[1] The Gephi file is available at: https://github.com/DataScience-in-Tourism/Chapter-21-Social-Network-Analysis

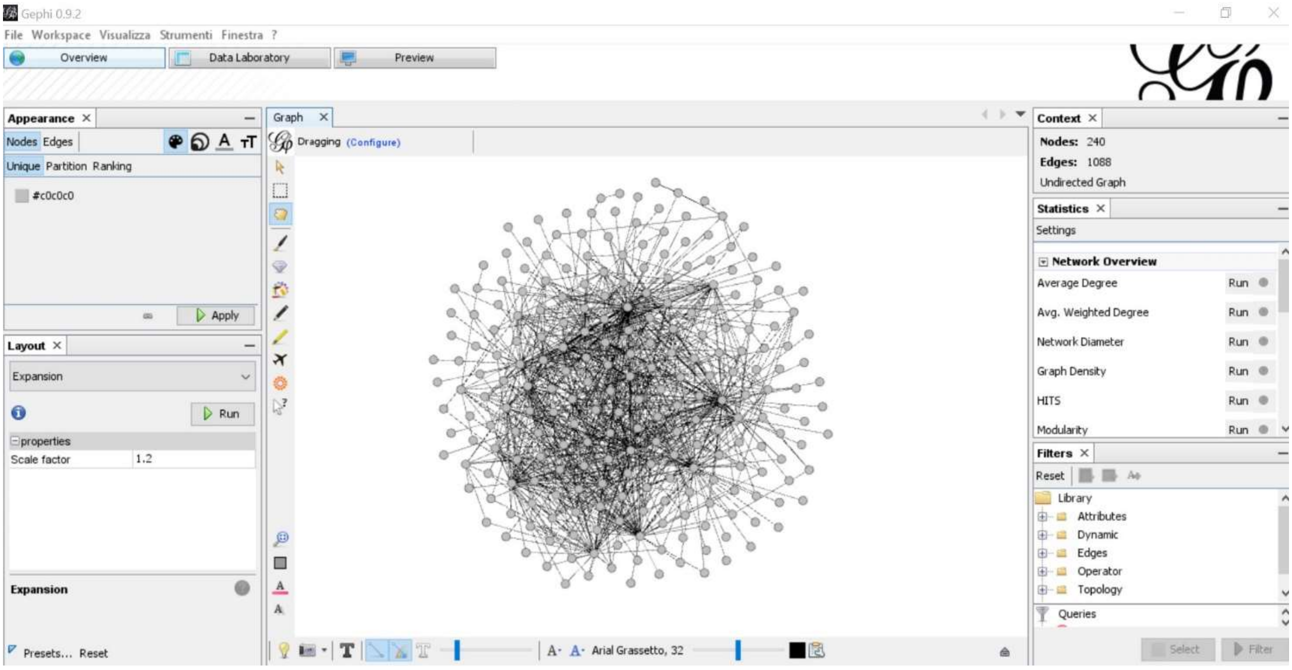


Figure 5. Gephi main screen

The network is loaded and displayed in the main Gephi panel (***Overview***), and the main working environment (Figure 5) provides the user with all the algorithms needed for analyzing the network and its visualization. In particular, the ***Statistics*** panel contains the functions needed to calculate all the metrics described above (Figure 6).

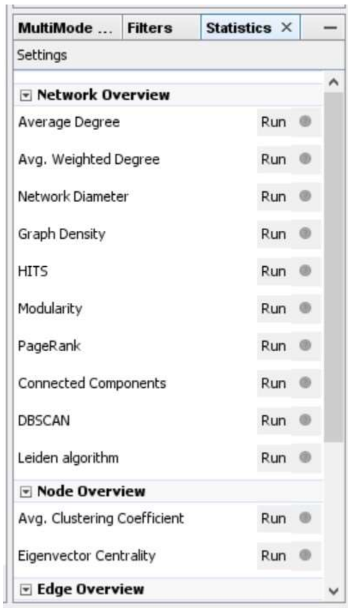
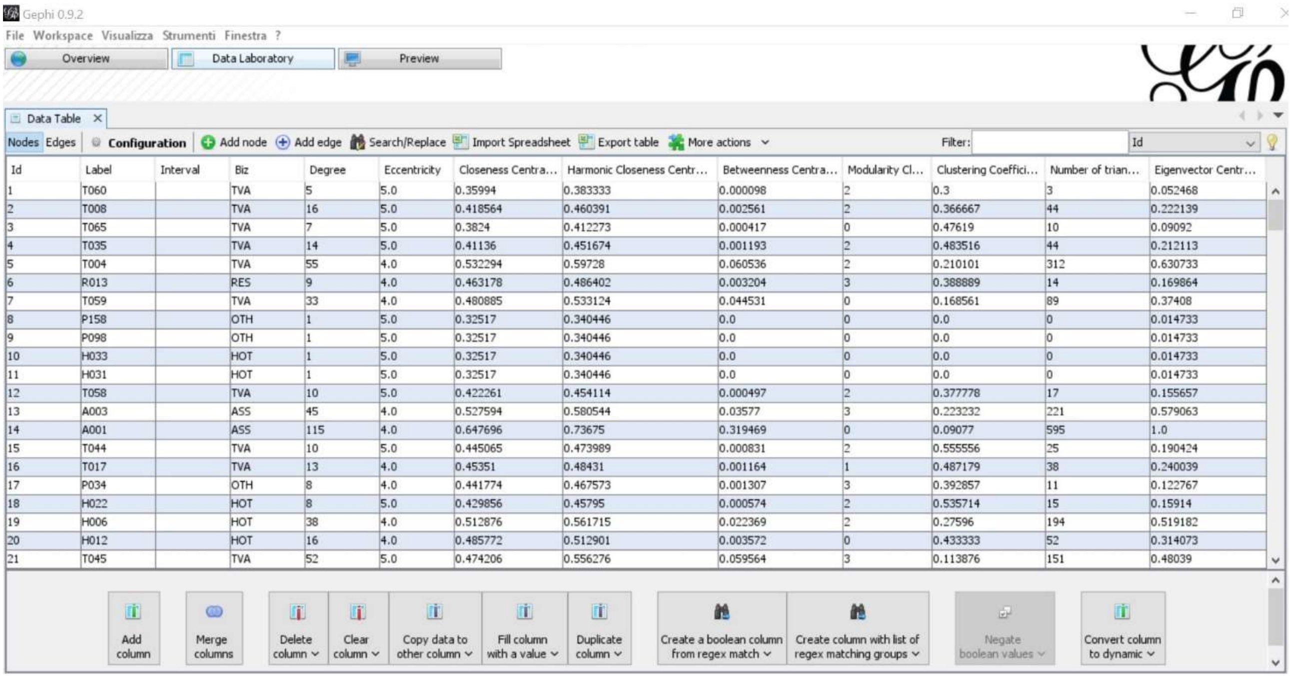


Figure 6. Gephi's Statistics functions

Once done, all the results can be found in the ***Data Laboratory*** screen (Figure 7). The table can be downloaded as a csv file, which can then be analyzed further with other tools, for example, Excel, if needed.



Id	Label	Interval	Biz	Degree	Eccentricity	Closeness Centra...	Harmonic Closeness Centr...	Betweenness Centra...	Modularity Cl...	Clustering Coeffici...	Number of trian...	Eigenvector Centr...
1	T060		TVA	5	5.0	0.35994	0.383333	0.000098	2	0.3	3	0.052468
2	T008		TVA	16	5.0	0.418564	0.460391	0.002561	2	0.366667	44	0.222139
3	T065		TVA	7	5.0	0.3824	0.412273	0.000417	0	0.47619	10	0.09092
4	T035		TVA	14	5.0	0.41136	0.451674	0.001193	2	0.483516	44	0.212113
5	T004		TVA	55	4.0	0.532294	0.59728	0.060536	2	0.210101	312	0.630733
6	R013		RES	9	4.0	0.463178	0.486402	0.003204	3	0.388889	14	0.169864
7	T059		TVA	33	4.0	0.480885	0.533124	0.044531	0	0.168561	89	0.37408
8	P158		OTH	1	5.0	0.32517	0.340446	0.0	0	0.0	0	0.014733
9	P098		OTH	1	5.0	0.32517	0.340446	0.0	0	0.0	0	0.014733
10	H033		HOT	1	5.0	0.32517	0.340446	0.0	0	0.0	0	0.014733
11	H031		HOT	1	5.0	0.32517	0.340446	0.0	0	0.0	0	0.014733
12	T058		TVA	10	5.0	0.422261	0.454114	0.000497	2	0.377778	17	0.155657
13	A003		ASS	45	4.0	0.527594	0.580544	0.03577	3	0.223232	221	0.579063
14	A001		ASS	115	4.0	0.647696	0.73675	0.319469	0	0.09077	595	1.0
15	T044		TVA	10	5.0	0.445065	0.473989	0.000831	2	0.555556	25	0.190424
16	T017		TVA	13	4.0	0.45351	0.48431	0.001164	1	0.487179	38	0.240039
17	P034		OTH	8	4.0	0.441774	0.467573	0.001307	3	0.392857	11	0.122767
18	H022		HOT	8	5.0	0.429856	0.45795	0.000574	2	0.535714	15	0.15914
19	H006		HOT	38	4.0	0.512876	0.561715	0.022369	2	0.27596	194	0.519182
20	H012		HOT	16	4.0	0.485772	0.512901	0.003572	0	0.433333	52	0.314073
21	T045		TVA	52	5.0	0.474206	0.556276	0.059564	3	0.113876	151	0.48039

Figure 7. Gephi's Data Laboratory screen with an example of the metrics computed by the program

From this window, one can easily identify the most relevant elements by sorting the different metrics. Keep in mind that, depending on the extent to which *importance* is given, different measures may apply. Here, degree signals popularity, clustering coefficient refers to collaborative relationships with immediate neighbors, and betweenness indicates a broker or a bottleneck. If necessary, a global indicator can be derived from averaging all of these values.

The analysis reveals a network with low connectivity (density = 0.038) yet good compactness for the low value of the average path length and diameter (diameter = 6, average path length = 2.539). The good average clustering coefficient (0.5) shows good capabilities and willingness from the actors' side to work with their immediate neighbors. Moreover, a modularity analysis identifies four communities (Figure 8); but relatively badly defined (Q = 0.299, a low value). If, then, we consider the composition of these communities, for example, by coloring them by business type, we see that they are all "mixed", that is, no cluster is formed by one single type of operator. It must be noted here that, unfortunately, Gephi does not provide functions for this type of node rearrangement. Therefore, the repositioning must be performed by hand, which is clearly only feasible in cases where the number of nodes is not too high. Thus, by highlighting the different communities (relatively poorly separated), a relatively good tendency to collaborate could be emphasized. Yet, considering that they are formed by different types of tourism operators, a good possibility of imagining and designing multifaceted products and services exists. Overall, all of these considerations should be validated and verified by taking deeper knowledge of the specific destination into account.

Finally, the last Gephi panel (***Preview***) shows a picture of a network that can be personalized and exported into a variety of different formats.

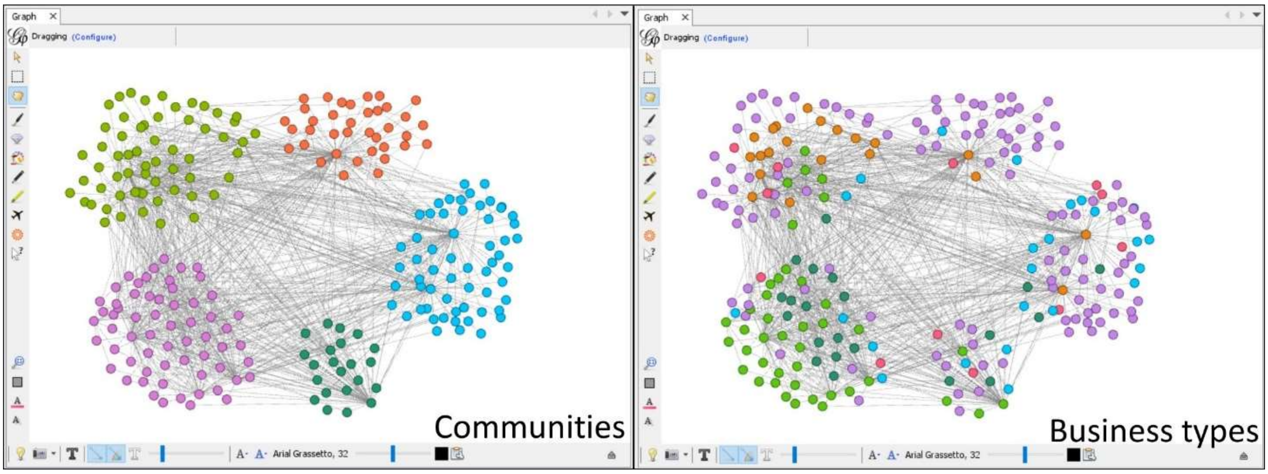


Figure 8. A network laid out and colored by the communities uncovered and by the type of business inside the communities

It is important to take into consideration that regardless of which package is selected, they all have their limitations in the fact that the metrics and the type of analyzable network (symmetric, weighted, directed, etc.) are given. Moreover, it is not always easy to understand what parameters are used for the calculations (algorithm, normalization factors, features considered, links' weights, etc.). When peculiar computations are needed, the only possibility is to use a programming language with the available libraries. The same can be said if some kind of simulation is desired or if modifications are to be implemented on the network as well as their effects assessed.

Lastly, it is worth underlining that this is a domain in which the *old* distinction between qualitative and quantitative approaches is not only meaningless but can also be viewed as dangerous. As exemplified and well-defined in Mariani and Baggio (2020), purely qualitative methods risk leading to contradictory or inaccurate results. On the other hand, good qualitative knowledge of the issues under study is of crucial importance for a correct and useful interpretation of the quantitative results. For this reason, and given the different competencies and expertise required, a meaningful study typically requires a well-mixed multidisciplinary team of researchers.

Research-Case

The case described in this section is the one discussed in Raisi et al.'s (2019) paper, "A network perspective of knowledge transfer in tourism". The objective is to investigate the characteristics of the inter-organizational knowledge transfer in Western Australian tourism by assessing the topological characteristics of the existing network at the destination. The relevance of this issue is evident since efficient, effective, and smooth flows are recognized to be of crucial importance for creating a sustained competitive advantage in both destinations and individual organizations and are a prerequisite for establishing the innovative atmosphere fundamental for maintaining a good level of competitiveness (Cooper, 2018; Hjalager, 2010).

The network was built by collecting data from companies, businesses, and organizations involved in the tourism industry in the area through an online questionnaire. Essentially, the question asked individuals to name ten entities, in order of importance, from which the participants receive information or knowledge and the (perceived) importance of these transfers for their own business. The resulting network is directed (direction being the citation of an existing relationship), and the analysis used a combination of tools including UCINET (Borgatti et al., 1992), Gephi (Bastian et al., 2009), and the Networkx Python library (Hagberg et al., 2008). These resulted in the metrics summarized in Figure 9.

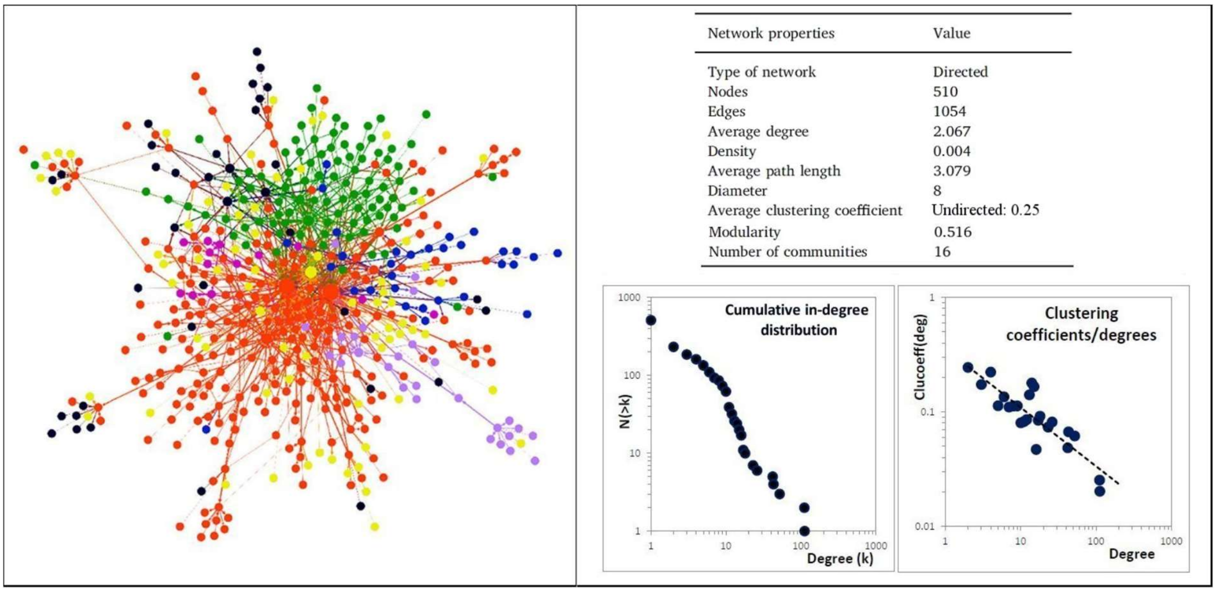


Figure 9. The Western Australian tourism network and its main characteristics (adapted from Raisi et al., 2019)

The network is rather sparse (very low density) but relatively compact (small diameter and low average path length). Furthermore, the degree distribution has a clear power-law shape, meaning that a large number of organizations receive information from a few but highly central organizations. The clustering coefficient calculated on a symmetrized version of the network is relatively high, and the modularity index is similarly high as well. This indicates that a few (16) well-defined communities exist and that the actors tend to form small, closely related collaborative groups. Drawing the relationship between the average clustering coefficient of a node and its degree, a power-law relationship is obtained that suggests a hierarchical organization of the whole system (see Ravasz & Barabási, 2003).

Rank	Importance index	Region	Sector
1	0.438	Experience Perth	Public tourism body
2	0.411	Experience Perth	Public tourism body
3	0.247	Experience Perth	Regional tourism organization
4	0.217	Australia's South West	Regional tourism organization
5	0.174	National	Public tourism body
6	0.141	Experience Perth	Tourism association
7	0.125	Australia's South West	Tourism association
8	0.121	Experience Perth	Tourism association
9	0.110	National	Public tourism body
10	0.107	Experience Perth	Tourism association
11	0.106	Experience Perth	Regional tourism organization
12	0.103	Experience Perth	Public tourism body
13	0.097	Experience Perth	Information services
14	0.092	Experience Perth	Tourism association
15	0.092	Experience Perth	Information services

Figure 10. The most relevant actors in the Western Australian tourism network (adapted from Raisi et al., 2019)

From a microscopic point of view, the authors identify the most relevant actors in the network. Given the different meanings of “importance” attributed to the various nodal measures, a workable suggestion is to use, as a global indicator, the geometric mean of the normalized values of the four main centrality measures: in-degree, closeness, betweenness, and eigenvector (Sainaghi & Baggio, 2014). The completed ranking is shown in Figure 10, which clearly renders regional public institutions and associations as important for disseminating information and knowledge.

Thus, in this particular case, the use of network analytic methods was able to provide a good picture of the situation and supply a series of outcomes that, most likely, would have remained blurred when using other methods. More importantly, however, this investigation has opened up an avenue of deeper and more interesting analyses. One possible development would be to refine the analysis by estimating, for each actor, the capabilities to absorb and transfer information and see how this modifies the overall picture. Moreover, by implementing some suitable numerical simulation, it would be possible to examine the effects different modifications of the network’s structure can have on the dynamic process of exchanging information as well as how to optimize the process.

Service-Section

Main application fields:

Practically any domain in which a “relational” aspect is considered important and in which entities with the role of nodes and relationships between any two of them can be reasonably and meaningfully defined.

Limitations & pitfalls:

The initial data collection is a delicate matter (see above) as it is the use of specific software tools that might limit the cases in which they can be used and force the researcher (usually for reasons regarding lack of awareness or know-how) to resort to unnecessary modifications of the network (e.g. symmetrizing, projecting bipartite networks, dichotomizing links’ weights, etc.), ultimately reducing the informative content of the data, without, at least, exploring the effects of various possible changes.

Similar methods and methods to combine with:

Any other method that is useful to fully answer the research questions. Often combined with some regression or correlation analysis.

Available Software / Solutions

Name:	Social network analysis software
Website:	https://en.wikipedia.org/wiki/Social_network_analysis_software
Description:	Long list of network analysis software applications and libraries

Name:	Gephi
Website:	https://gephi.org/
Description:	Visualization and calculation of basic network measures

Costs:	Free
Distribution:	Download from website. Multiplatform

Name:	Ucinet
Website:	https://sites.google.com/site/ucinetsoftware/home
Description:	Calculation of basic network measures
Costs:	Commercial
Distribution:	Download from website. Windows only

Name:	Pajek
Website:	http://mrvar.fdv.uni-lj.si/pajek/
Description:	Visualization and calculation of basic network measures
Costs:	Free
Distribution:	Download from website. Windows only (can run on Mac or Linux via Wine)

Name:	Python language and libraries (Networkx, python-igraph)
Website:	https://www.python.org/
Description:	Complete libraries for visualization, calculations, and dynamic modeling
Costs:	Free
Distribution:	Install a distribution such as Anaconda https://www.anaconda.com/, a freely available compilation of Python language and libraries that includes Networkx and all the dependencies needed. Python-igraph, if needed, must be added separately (can be installed using: "conda install -c conda-forge python-igraph"

Name:	R libraries (igraph, sna, tnet etc.)
Website:	https://www.r-project.org/
Description:	Complete libraries for visualization, calculations, and dynamic modeling
Costs:	Free
Distribution:	All libraries are available on one of the "The Comprehensive R Archive Network (CRAN)" mirrors: https://cran.r-project.org/mirrors.html

Name:	Matlab toolboxes
Website:	https://www.mathworks.com/
Description:	Complete libraries for visualization, calculations, and dynamic modeling
Costs:	Commercial (expensive); many toolboxes, however, are free.
Distribution:	Academic licenses or other facilities may exist depending on the singular institutions. Many toolboxes for network analysis exist (generally free). They can be located with a Google search (e.g. "matlab network analysis toolbox")

Further Readings & other Sources

Baggio, R. (2013). Complexity, Network Science & Tourism. Innsbruck: IFITT Education Group. Available at: http://www.iby.it/turismo/papers/rb_TourNetSci(IFITT).pdf

Barabási, A. L. (2016). Network science. Cambridge, UK: Cambridge University Press.

Caldarelli, G., & Chessa, A. (2016). Data Science and Complex Networks: Real Case Studies with Python. Oxford: Oxford University Press.

Coscia, M. (2021). The Atlas for the Aspiring Network Scientist. Copenhagen: IT University of Copenhagen.

Easley, D., & Kleinberg, J. (2010). Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge: Cambridge University Press.

Mitchell, M. (2009). Complexity: A Guided Tour. Oxford: Oxford University Press.

References

Baggio, J. A., & Baggio, R. (2020). *Modelling and Simulations for Tourism and Hospitality*. Bristol, UK: Channel View.

Baggio, R. (2011). Collaboration and cooperation in a tourism destination: a network science approach. *Current Issues in Tourism, 14*(2), 183-189.

Barabási, A. L. (2016). *Network science*. Cambridge, UK: Cambridge University Press.

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: an open source software for exploring and manipulating networks*. Paper presented at the 3rd International AAAI Conference on Weblogs and Social Media, San Jose, CA (May 17-20).

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

Brodu, N. (2009). A synthesis and a practical approach to complex systems. *Complexity, 15*(1), 36-60.

Cooper, C. (2018). Managing tourism knowledge: a review. *Tourism Review*, *73*(4), 507-520.

Coscia, M. (2021). *The Atlas for the Aspiring Network Scientist*. Copenhagen: IT University of Copenhagen.

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.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3-5), 75-174.

Guillaume, J. L., & Latapy, M. (2006). Bipartite graphs as models of complex networks. *Physica A, 371*(2), 795-813.

Hagberg, A. A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using NetworkX.* Paper presented at the 7th Python in Science Conference (SciPy2008), Pasadena, CA (19-24 Aug).

Hjalager, A. M. (2010). A review of innovation research in tourism. *Tourism management*, *31*(1), 1-12.

Levin, S. A. (2003). Complex adaptive systems: Exploring the known, the unknown and the unknowable. *Bulletin of the American Mathematical Society, 40*(1), 3-19.

Lewin, R. (1999). *Complexity, life on the edge of chaos* (2nd ed.). Chicago The University of Chicago Press.

Mariani, M., & Baggio, R. (2020). The relevance of Mixed Methods for Network Analysis in Tourism and Hospitality Research. *International Journal of Contemporary Hospitality Management, 32*(4), 1643-1673.

Newman, M., Barabási, A. L., & Watts, D. J. (Eds.). (2011). *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.

Pavlopoulos, G. A., Kontou, P. I., Pavlopoulou, A., Bouyioukos, C., Markou, E., & Bagos, P. G. (2018). Bipartite graphs in systems biology and medicine: a survey of methods and applications. *GigaScience, 7*(4), art. giy014.

Raisi, H., Baggio, R., Barratt-Pugh, L., & Willson, G. (2019). A Network Perspective of Knowledge Transfer in Tourism. *Annals of Tourism Research, 80*, art. 102817.

Ravasz, E., & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical Review E, 67*, 026112.

Sainaghi, R., & Baggio, R. (2014). Structural social capital and hotel performance: is there a link? *International Journal of Hospitality Management, 37*, 99-110.

Sayama, H., Cramer, C., Porter, M. A., Sheetz, L., & Uzzo, S. (2016). What are essential concepts about networks? *Journal of Complex Networks, 4*, 457-474.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports, 9*(1), 1-12.

Glossary

Model: a concise, actionable and predictive representation of the system or phenomenon created to meet a specific goal. Models can be broadly divided into descriptive and analytical. Descriptive models rely on simulations and are often the basis for numerical simulations. Analytical models are made up of sets of equations describing the characteristics and the behavior of a system.

Network: A system composed of interconnected elements. It can be rendered graphically (graph) by using a series of points (vertices or nodes) linked by lines (edges or links). The nodes in a network can represent simple objects (a person in a friendship network) or complex entities (a firm or a website). A link indicates some type of relationship between two nodes. This can be an information exchange, a chemical reaction, a force etc. Links can be sym¬metric or directed (a trip from a place to another) and can be assigned a weight measuring a strength, an importance or a value. many measures of individual and global features can be calculated for characterizing the different configurations. These are mostly rooted in the mathematical discipline of graph theory.

Numerical simulation: A computation, typically run as a computerized algorithm, that implements a model for a system. Numerical simulations are used to study the behavior of systems whose analytical models are too complex to provide analytical solutions, as in most nonlinear systems, or when real-life experiments are not feasible for theoretical or practical reasons as in the case of social and economic systems.

System: A conceptual or real entity made of a number of elements interacting dynamically and generating some global behavior. Systems can be simple, complicated or complex. Simple systems have few compo¬nents with linear interactions and show predictable behaviors. Complicated systems contains a large number of components, but still with linear interactions. The global behavior can be (at least in principle) analyzed and derived as a superposition of the characteristics of some smaller parts. Complex systems are characterized by nonlinear interactions and feedback loops. They display a high sensitivity to initial conditions, a dynamic behavior adaptable to the environment, and can become chaotic. Emergent, self-organizing, structures and behaviors typical of these systems cannot be derived as a com¬position of its elements' features and properties.

Topology: The study of the intrinsic properties of an object or system that are due to its structural con-figuration and that are not modified by certain types of deformations or transformations that instead may radically change the geometric characteristics.