Annals of Tourism Research 73 (2018) 35–47

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

Network approach to tourist segmentation via user generated content

Juan M. Hernández[a], Andrei P. Kirilenko[b], Svetlana Stepchenkova[b,*]

[a] Institute of Tourism and Sustainable Economic Development (TIDES), University of Las Palmas de Gran Canaria, c/Saulo Torón s/n, 35017 Las Palmas, Spain

[b] Department of Tourism, Recreation & Sport Management, College of Health and Human Performance, University of Florida, USA

ARTICLE INFO

Associate editor: Josef A. Mazanec

Keywords:
Attractions
Network analysis
Social networks
Tourist segmentation
User-generated content (UGC)

ABSTRACT

The study contributes to the tourism literature by demonstrating an approach to segmenting tourists using network analysis with user-generated content. Online reviews of destination attractions are considered as a proxy for visitation data reflective of tourists' interests. The connectivity between attractions is represented with a network of links created by tourists visiting and reviewing multiple attractions. Attraction clusters are revealed by segmenting this network using network analysis tools. Two segmentation solutions are provided: a posteriori, in which only review information is taken into account, and mixed, in which tourist groups are defined a priori by their travel interests and age, and this information is combined with visitation information. The findings are validated using geovisualization and by comparing them with randomly simulated models.

Introduction

Destinations around the world strive to increase their value by delivering intelligent, customized services to tourists. Destinations' ability to provide such services is inherently connected to their capacity to collect, integrate, and analyze data from various sources and then redistribute that information to a stakeholder network of businesses, government agencies, policy-makers, as well as various organizations and activity groups (Gretzel, Werthner, Koo, & Lamsfus, 2015). Two elements, timely information and a strong network of interrelated entities, are essential in this process. The construction of tourism networks benefits the areas of learning and exchange (e.g., knowledge transfer, communications), business activities (e.g., cooperative marketing, purchasing, and production as well as enhanced cross-referrals), and community development (e.g., fostering common purpose, support for destination development, or increased sense of community) (Morrison, Lynch, & Johns, 2004). Sources of the information that is distributed via such networks have been extended from surveys and collected statistics to digital traces that tourists leave at a destination and user-generated content (UGC) on various online forums and social networks. Despite concerns related to the validity of online data (Trend, 2013), it has been successfully demonstrated that UGC provides valuable information to tourism and hospitality services (Xiang and Gretzel, 2010), e.g., hotels (Vermeulen and Seegers, 2009) and restaurants (Zhang, Ye, Law, & Li, 2010).

Destination networks have been of interest to researchers for the last three decades (e.g., Jamal, Smith, & Watson, 2008; Tremblay, 1998), and the literature on the topic is still growing. Earlier studies focus on the evolution of business networks and interorganizational relationships (Morrison et al., 2004; Pavlovich, 2003; Tinsley and Lynch, 2001), tourism policies and governance

* Corresponding author at: 312 Florida Gym, P.O. Box 118208, Gainesville, FL 32611-8208, USA.
E-mail addresses: juan.hernandez@ulpgc.es (J.M. Hernández), andrei.kirilenko@ufl.edu (A.P. Kirilenko), svetlana.step@ufl.edu (S. Stepchenkova).

https://doi.org/10.1016/j.annals.2018.09.002
Received 5 June 2018; Received in revised form 27 August 2018; Accepted 4 September 2018

agents, power structures between various destination actors and similar issues (see the review in Tran, Jeeva, & Pourabedin, 2016). A recent review of the tourism network literature (van der Zee & Vanneste, 2015) adds two more interest foci to network literature: coopeting (cooperative competition) networks with an emphasis on relationships between different stakeholders and studies interested in networks as a structure with a specific configuration. Whereas research on factors that govern network formation and functioning primarily employ a descriptive apparatus, the "network configuration" studies utilize graph theory to make conjectures based on the identified network structure. An example would be a study by Shih (2006), who focuses on how attractions and destination facilities are connected via tourist driving routes, or Liu, Huang, and Fu (2017), who describe an attraction network informed by tourist flows.

Despite the fact that tourists are the main users and primary evaluators of services at a destination, networks in tourism research are rarely investigated from the tourist's point of view (van der Zee & Vanneste, 2015). At the same time, tourists' movements at a destination between various attractions, hotels, and restaurants comprise elaborate networks that can potentially be informative for understanding tourists' interests and behavior and as such serve as a basis for tourist segmentation (Frochot & Morrison, 2000). Selecting homogeneous groups in an otherwise heterogeneous tourist market makes it possible to better tailor services, provide higher satisfaction, achieve repeat visitation, achieve more revenue for businesses, and, ultimately, create a more dynamic and vibrant destination. By being able to specialize in catering to the travel needs of a particular tourist segment, DMOs can gain an edge compared to other competing destinations (Dolnicar, 2008). "Market segmentation tends to produce depth of market position in the segments that are effectively defined and penetrated. The [organization that] employs market segmentation strives to secure one or more wedge-shaped pieces [of the market cake]" (Smith, 1956, p. 5). Thus, tourist segmentation is considered one of the most important tasks that destinations perform and one of the main subject areas of tourism research (Kirilenko & Stepchenkova, 2018; Yuan, Gretzel, & Tseng, 2015).

The present study brings together three elements discussed above. One is the need to identify sufficiently large and viable tourist segments, which would allow more effective dispensation of tourist information and customize services. Another element is building a network of interrelated attractions that tourists visit while at destination that manifests their interests and behavior and can serve as a basis for segmentation. By doing this, the study is effectively placed into the "network configuration" group of studies because it utilizes graph theory with its quantitative approach, rather than qualitative methods of network sketching. Finally, in contrast to previous studies that primarily used survey data to gain insights about places of tourists' interest at a destination, the basis for the network construction is UGC of travel reviews, which provide a source of big data information regarding tourists' actual behavior at a destination as well as their interests and personal characteristics. Thus, the main aim of this study is to investigate whether tourist segmentation can be achieved through the network analysis of attractions that tourists visit at a destination using UGC as the information source.

Network analysis

A network is a convenient way of describing connected objects such as individuals, businesses, and attraction points. In particular, a network is a set of vertices (nodes), edges linking the nodes, and their descriptors. The network is called unipartite when all nodes belong to one category. For example, in social networks, these nodes usually represent individuals/agents or groups of individuals/agents, whereas edges represent relationships among those agents (e.g., friendship among individuals). The density of edges in the network is the quotient between the observed number of edges and the number of possible edges. A bipartite, or affiliation, network includes two categories of nodes (e.g., people and events), and a relationship can be produced only between nodes from different categories (e.g., a person attending an event). Additionally, a network (both unipartite and bipartite) can be weighted; that is, every edge can have a numerical value, e.g., the number of times a person has attended an event. Social network analysis provides quantitative methods to analyze such networks and has been extensively used in social sciences since the first half of the last century. An extended introduction to social network analysis and its applications in social systems is provided by Wasserman and Faust (1994) and Scott (2012).

One of the most common research questions in the study of networks is the detection of clusters, also called communities. A map of communities reveals how the network is configured by showing the existence of parts of the graph that work to some extent autonomously. Such a map highlights the similarities or differences among nodes in terms of connectivity and, therefore, is conducive to inferring the internal forces that create the observed network configuration. The problem of clustering has existed in the analysis of social networks for several decades (Wasserman & Faust, 1994). Methodologically, community detection involves finding the best partition of nodes in groups or clusters in such a way that the density of links among nodes inside every cluster is higher than the density of edges among nodes belonging to different groups. Lately, new Big Data applications have emerged revealing e.g., scientific collaboration networks and communities in online social networks websites (see Fortunato, 2010 for a review of some applications). In tourism, Baggio (2011) analyzed communities in the collaboration network of tourism stakeholders in the island of Elba, Italy. Asero, Gozzo, and Tomaselli (2016) unveiled clusters in the network formed by origins and destinations of tourist trips in Sicily. Williams, Terras, and Warwick (2013) detected communities in electronic word-of-mouth networks for a destination. Finally, David-Negre, Hernández, and Moreno-Gil (2018) unfolded a core-periphery structure of a network of tourists and activities in a destination. Notwithstanding these studies, the application of community detection methods to tourism networks is still very limited.

Tourist segmentation

Effective market segmentation identifies tourist segments whose interests a destination can effectively serve that are sufficiently