Tourism Management Perspectives 33 (2020) 100607

Contents lists available at ScienceDirect

Tourism Management Perspectives

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

Analyzing co-authoring communities of tourism research collaboration

Wei Fan[a,*], Gang Li[b,c], Rob Law[a]

[a] School of Hotel & Tourism Management, The Hong Kong Polytechnic University, Hong Kong
[b] School of Information Technology, Deakin University, Vic 3125, Australia
[c] Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, 830011, China

ARTICLE INFO

Keywords:
Research collaboration
Network analysis
Co-authoring community
Centrality measure
Tourism research

ABSTRACT

Research on collaboration relationships among researchers assists in understanding scientific information dissemination. Network analysis has been used in many studies on collaboration. However, understanding the co-authoring communities in collaboration networks and identifying their key researchers generally remain unexplored, although these insights can lead to an improved comprehension of collaboration relationships. The current study discovers and analyzes co-authoring communities in the collaboration networks of researchers in tourism research. Results indicate that productivities within co-authoring communities are high because of dense collaboration relationships. We also observe that location may be one of the formation reasons of co-authoring communities. Key researchers who are globally and locally important can be further identified by introducing the community centrality measure. The network robustness test demonstrates the effectiveness of this measure.

1. Introduction

Collaboration is an important and common feature in scientific research. Collaborations between different researchers and institutes allow the sharing of their resources, such as research data sets. Scientific information can be spread and exchanged by citations and collaborations among researchers from different disciplines. The collaboration relationships among researchers should be investigated to promote the progress of scientific research and dissemination of scientific information. This topic has elicited the attention of researchers. A problem on which researchers are concerned most is how researchers collaborate and which researchers are important in maintaining the collaboration relationships.

Traditional quantitative analysis of the literature has been used to investigate the collaboration of researchers. For example, the number of collaborators, frequency of collaboration, and impact and quality of research have been investigated (Katz & Hicks, 1997; Sheldon, 1991). The influence of individual researchers can be reflected by the number of citations attributed to their research. However, the collaboration paths established through researchers' collaboration relationships are not considered in these methods. Collaboration is an important means to communicate scientific information (Franceschet, 2011; Newman, 2001a). Thus, the corresponding results cannot help in understanding how research resources and scientific information spread among researchers via collaboration or how individual researchers influence this information flow (Fan, Li, & Law, 2017).

To address the aforementioned problem, the collaboration networks of researchers have been investigated using network science methods (Newman, 2001a, 2001b). In a collaboration network, researchers can be represented by the nodes of a network and collaboration relationships as edges. In the majority of previous studies, two researchers have an edge if they have co-authored at least one research article. Collaboration relationships among researchers can be substantially understood because collaboration networks can be visualized as graphs. The network structure of a collaboration network facilitates in revealing its formation pattern. Collaboration networks of tourism research have been constructed in existing studies (Fan et al., 2017; Racherla & Hu, 2010; Ye, Li, & Law, 2013). Network properties, such as degree distribution, have been studied. The changes of tourism research collaboration network over time and how individual researchers relate to the changes have been investigated (Fan et al., 2017). However, the majority of studies on collaboration networks are hindered by two major limitations.

- First, the majority of existing research presents general structural characteristics, which provide an overview of collaboration networks. These studies disregard the possibility that collaboration networks may comprise a set of co-authoring communities. Community structures have been found in social networks, such as the collaboration network of Jazz musicians (Gleiser & Danon,

* Corresponding author.
E-mail addresses: w.fan@polyu.edu.hk (W. Fan), gang.li@deakin.edu.au (G. Li), rob.law@polyu.edu.hk (R. Law).

https://doi.org/10.1016/j.tmp.2019.100607
Received 19 December 2018; Received in revised form 5 November 2019; Accepted 6 November 2019

2003) and Zachary's (1977) karate club network. In collaboration networks, the existence of communities means that researchers within the same community have more collaborations than those from different communities. In the current research, the communities in collaboration networks are called co-authoring communities. A substantial understanding of a collaboration network can be obtained by studying these co-authoring communities.

- Second, a systematic study on the identification of key researchers is lacking. Moreover, identifying key researchers is a popular topic in collaboration research and existing studies on tourism collaboration networks mainly focus on researchers with the most publications or those with the most collaborators. However, two researchers with an equal number of collaborators or publications may have different structural roles in the collaboration network. A researcher who contributes to the connectivity of a collaboration network should be important because effective resources and scientific information dissemination relies on the connectivity of a collaboration network.

Previous studies have remaining issues because of the limited understanding of community structure in social network analysis. Researchers have been proposing community detection methods to promote detection quality on the basis of a variety of quality measures and reduce computational complexity for applications in large networks (Fortunato, 2010; Fortunato & Hric, 2016). Given the advancements in community detection and its applications in different types of social networks, analyzing the underlying community structure for collaboration networks is now possible. The present study aims to solve these problems in existing research by systematically analyzing the co-authoring structure and studying key researchers in the collaboration network of tourism research. This research can answer questions on the characteristics of co-authoring communities, such as whether dense collaboration relationships within a community indicate high productivity or whether researchers in the same co-authoring community have similar features. This study also proposes a robustness test to identify researchers with important roles in the connectivity of collaboration networks. The exploration of substantial structural features of collaboration networks assists the development of tourism research. Communications within a co-authoring community may promote the research in this community, whereas that between co-authoring communities may facilitate the exchange of resources and information. The understanding of researchers' influence, particularly key researchers' influence, in co-authoring communities and the entire collaboration networks also assists scientific information spreading.

The remainder of this study is organized as follows. We review the literature on the study of collaboration in preliminary and related works. Thereafter, the methodology used is presented. The analysis results for our collaboration networks are provided. Lastly, we conclude this study and provide suggestions for future research.

2. Related works

Collaboration is a common feature in research. Investigating collaboration in research assists in understanding the dissemination of research resources and scientific information. Therefore, collaboration among researchers and institutes has been studied in the past few decades.

2.1. Bibliometric analysis of collaborations

Quantitative analysis has been applied to investigate research collaboration. Particularly studied are the number of citations, number of collaborators, and distribution. Researchers have analyzed the collaborations in tourism research. Sheldon (1991) studied tourism papers published in the 1980s. The contributions of researchers were calculated and an increasing trend of collaboration was identified. Jogaratnam, Chon, McCleary, Mena, and Yoo (2005) analyzed the repeat contributions by authors and institutions, among others. Zhao and Ritchie (2007) studied 57 prolific researchers in tourism research. These researchers' background and the number of articles that they published in different tourism journals were provided. The results show that 70% of the articles published by these researchers are collaborative research. Lastly, Hall (2011) used the percentage of international collaboration as one of the metrics to evaluate journal quality.

Traditional bibliometric methods provide statistical information on research collaboration. However, the detailed explanation of how scientific information and research resources are shared and disseminated among researchers cannot be obtained through these methods. Thus, researchers have introduced the theory of network analysis to the scientific collaboration study.

2.2. Application of network analysis in collaboration research

Network analysis has been applied in the study of numerous real networks, such as the Internet (Wang & Chen, 2003), research topic network (Chen & Zhao, 2015), and social networks (Strogatz, 2001; Watts, 2004). In the Internet network, routers or domains can be represented by a set of nodes and the physical links between them can be represented by a set of edges. The Internet can be visualized with this definition as a graph of nodes and edges connecting these nodes. For the case of social networks, individuals are represented as nodes and the relationships between them as edges. The structure of a social network can affect the dissemination of information, and its visualization facilitates the understanding of the dissemination. Collaboration between researchers or institutes is a type of social relationship and their collaboration networks are also social networks. Network analysis methods can assist in understanding the collaboration relationship among researchers. In this study, researchers are nodes in a collaboration network. An edge connects two researchers if they have published at least one article together. Collaboration networks in different research fields have been investigated by researchers (Barabási et al., 2002; Franceschet, 2011; Newman, 2001c).

Network analysis has been applied in collaboration studies and the global structural properties of collaboration networks have been investigated. Moody (2004) investigated a collaboration network in social science. Huang, Zhuang, Li, and Giles (2008) collected articles on computer and information sciences and studied the evolution of their network. Franceschet (2011) investigated collaboration networks for computer science. Key nodes (researchers) in these networks have also been studied. Newman (2004a) calculated several measures of node connectedness to identify the best-connected scientist. Franceschet (2011) assessed the contributions of the most collaborative researchers in the network using a network resilience test.

In the area of tourism research, collaboration networks have been investigated. Benckendorff (2010) analyzed the collaboration networks constructed with articles published by Australian and New Zealand researchers from 1999 to 2008. Different measures are used to identify the most productive and collaborative researchers and institutions in these networks. Although researchers with a high betweenness centrality were mentioned to be able to control communication flow, the author did not propose a method to measure such an ability. Racherla and Hu (2010) constructed a collaboration network with tourism research articles published from 1996 to 2005. They determined that researchers who connect others together are crucial in a network. The aforementioned authors concluded that additional publications would be produced if a researcher has considerable collaborations. However, Ye et al. (2013) mentioned that the sample size and time span of Racherla and Hu's (2010) data are insufficient. They alleviate this problem based on articles published from 1991 to 2010. Two types of critical researchers are defined and identified. One type of critical researchers is able to connect researchers with only one collaborator and others. The majority of the collaborators of another type of critical researchers have at least two edges. The results show that collaborations are associated with researchers' research outputs. Zhong, Wu, and Morrison (2015) provided an example to show that a researcher should be important if they perform as a “connector” in a collaboration network. Fan et al. (2017) investigated the changes in tourism collaboration networks over time. Changes in the basic properties of collaboration networks have been presented and a method to measure the impact of researchers in these changes has been introduced.

2.3. Limitations of existing works

Network analysis has been extensively applied in studies on collaborations of researchers and its effectiveness has been shown in existing research. In previous studies of collaboration networks, researchers have mainly focused on the global structure of collaboration networks and key researchers in these networks. However, the community structure of these networks in the majority of these studies is disregarded, which is a common feature in numerous social networks (Fortunato, 2010). Researchers who collaborate to publish are in a collaboration network. Researchers with similar research interests may communicate and collaborate more often with one another than with researchers with different research interests. Consequently, researchers with similar interests may have dense connections among them and form a co-authoring community. Co-authoring communities can be described by the concept of community structure in theory of network analysis. Although Newman (2006) and Rodriguez and Pepe (2008) found community structures in their collaboration networks, the sample sizes of their networks are limited. We have reviewed papers on the collaboration of researchers in tourism and found that the majority used quantitative analysis. Accordingly, community structure has not been explored in the collaboration networks of tourism research reviewed in Section 2.2. Moreover, no detailed study on the correlation between co-authoring communities and the research performance within these co-authoring communities exists.

The importance of researchers in their co-authoring communities has not been investigated. In existing studies, researchers are considered important if they are the most collaborative in the entire collaboration network. However, a researcher who contributes to the connectivity of a collaboration network should also be important because effective research resources and scientific information dissemination rely on the connectivity of collaboration networks. The majority of the existing studies did not verify whether productive researchers or researchers with the most collaborators are the ones who contribute the most to the connectivity of a collaboration network. In existing studies, the importance of researchers has been evaluated through their productivity and contribution to the connectivity of collaboration networks (Fan et al., 2017). Nevertheless, researchers' local importance in co-authoring communities was not considered. In the current study, local importance means researchers in the same community have short scientific information dissemination paths to one another. Note that local importance does not necessarily mean that a researcher is important among researchers with spatial proximity. Another possibility is that although certain researchers are not the top researchers in the global collaboration network, they remain important in their co-authoring communities. For example, a researcher may be the top in a particular country or institute but may not be a researcher with great influence on a global scale. Therefore, in this study we are interested in researchers' importance in the global collaboration network and co-authoring communities. Moreover, we aim to identify key researchers who are essential in retaining the majority connections. That is, without these researchers, the collaboration network would lose numerous connections and become fragmented. Given a collaboration network, the current study focuses on its community structure to solve the problems in the existing research. Global and local importance are considered in the investigation of key researchers. Moreover, key researchers are evaluated by their contributions to the connectivity of the collaboration network. The construction of collaboration networks is based on a set of publications. Therefore, the influence of the identified key researchers is within the scale of these publications.

3. Methodology

This section introduces the characteristics of the data set for our analysis and the preprocessing method. Thereafter, some basic terms and concepts in network analysis are introduced for an improved understanding of the following analysis. This section also provides the method of constructing collaboration networks and explains the analysis process of co-authoring communities and key researchers.

3.1. Data collection

The data collected for analysis are from *Tourism Management*, which is a first-tier journal in tourism research and has published thousands of articles. Bibliographic data for all articles (except editorials) published in the journal from 1982 to 2015 were collected from *ScienceDirect*. Although the present study is applicable to a repository of publications from more sources, we focus on one single source for two reasons. First, this journal is representative of a top tourism journal that has published articles with broad topics (Benckendorff & Zehrer, 2013; Racherla & Hu, 2010). Second, our data set has a longer time span than the ones in existing studies on tourism research collaboration networks. Our data set includes 3525 articles, which comprise research articles, research notes, book reviews, and other article types. Evidently, processing the data and visualizing collaboration networks would be difficult if multiple sources are included. The long timespan covered ensures that the structures of the constructed networks are relatively stable. Some previous studies have constructed collaboration networks with only research articles and research notes (Ye et al., 2013). Thus, we derive a sub-collection from our data set to compare with the full collection. This sub-collection has 2317 research articles and research notes.

3.2. Data preprocessing

Two collaboration networks can be built with the full collection of articles and sub-collection of research articles and research notes. To distinguish between these two collaboration networks, we will refer to the collaboration network based on all types of articles as *raw collaboration network*, while the collaboration network based on research articles and research notes as *research-article collaboration network*.

The ambiguous name problem should be considered in the process of collaboration network construction. Two cases exist, namely, a researcher with several names and several researchers with the same name. We preprocessed our data with the methods mentioned in Franceschet's (2011) work. After addressing the ambiguous name problem, 3948 authors are identified in the *raw collaboration network*, in which the number of edges is 5310. In the *research-article collaboration network*, the number of authors is 3448 with 3834 edges among them. The data set has been publicly released as “TM2015” at: http://github.com/tulip-lab/open-data.

3.3. Foundational concepts in network analysis

We provide the definitions of degree centrality and betweenness centrality, which have been used to identify key researchers in previous studies of collaboration networks. Thereafter, the concept of community structure and the community detection process are introduced. We also introduce robustness, which assists to understand the researchers' importance in maintaining network's connectivity.

3.3.1. Centrality

3.3.1.1. Degree centrality. The degree of a node is the number of its

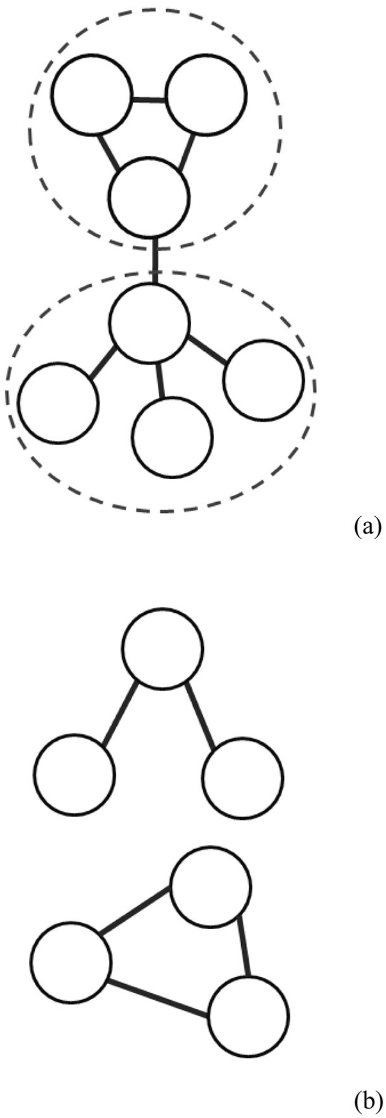


Fig. 1. Two sample networks. (a) Connected network with two communities. (b) Network with two connected components.

direct neighbors. Degree describes the number of edges connected to a node, through which scientific information can spread. Therefore, degree can reflect the importance of a node. Researchers have used degree centrality to identify key researchers (Racherla & Hu, 2010; Ye et al., 2013).

3.3.1.2. Betweenness centrality. Betweenness centrality of a node is the number of shortest paths passing through such a node. In a collaboration network, many shortest paths in this network passing through a researcher means that scientific information can disseminate rapidly through this researcher. Therefore, betweenness centrality can be used to measure the importance of researchers (Newman, 2004b; Racherla & Hu, 2010; Ye et al., 2013).

3.3.2. Community discovery

Community structure is an important structural feature of networks. Communities in a network are special clusters of nodes. The concentration of edges within these clusters is high and that between clusters is relatively low (Fortunato, 2010). For example, two communities exist in the network shown in Fig. 1a. Community structures have been found in collaboration networks. Newman (2006) found the community structure in the collaboration network of physicists. Rodriguez and Pepe (2008) compared the structural and socio-academic communities of the collaboration network for the study on sensor networks and wireless communication.

Researchers have proposed various community detection methods. Numerous quality functions are proposed to measure the quality of community detection. Although no quality function is universally accepted, many popular community detection methods are based on the quality function "modularity" (Newman & Girvan, 2004):

$$Q = \frac{1}{2m}\sum_{ij}[A_{ij} - P_{ij}]\delta(c_i, c_j) \tag{1}$$

This equation is for undirected networks with m edges. In particular, A_{ij} is an element of the adjacency matrix $\boldsymbol{A}$ of the network; $A_{ij} = 1$ if an edge exists between nodes i and j; k_i is node i's degree; $P_{ij} = k_i k_j/2m$ is the expected number of edges between nodes i and j; c_i is the index of the community where node i belongs; and $\delta(c_i, c_j) = 1$ if nodes i and j are in the same community. Thus, modularity calculates the average difference between the number of edges within communities and the expected number of such edges. A large value of modularity means a strong community structure. The modularity value for a network with strong community structure is often above 0.3 (Newman & Girvan, 2004). For example, the value of modularity for the community structure shown in Fig. 1a is $Q = 0.3571$. Many modularity-based methods are designed to seek the maximum modularity (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Newman, 2006).

In the study of community structure of collaboration networks, we observe another centrality measure, namely, "community centrality," which was defined by Newman (2006). This centrality measure can specify how central a node is in its community. To calculate community centrality, we first create a modularity matrix $\boldsymbol{B}$ with elements $B_{ij} = A_{ij}$ - $k_i k_j/2m$. $\boldsymbol{U} = (\boldsymbol{u}_1|\boldsymbol{u}_2|\cdots)$ is the matrix of the eigenvectors of $\boldsymbol{B}$, while β_i is the eigenvalue corresponding to $\boldsymbol{u}_i$. If the number of positive eigenvalues is p, then each node i has a vector $\{x_i\}$ of dimension p that $[x_i]_j = \sqrt{\beta_j}U_{ij}$. The value of community centrality of node i is equal to the vector magnitude $|x_i|$. In the majority of cases, the nodes with a high degree would also have high community centrality. However, not all low-degree nodes have low community centrality. If a node has more connections in its community than expected, then it can contribute to the modularity Q and gain high community centrality. Newman (2006) found that researchers with a high community centrality are senior researchers in their research groups. This finding indicates that community centrality can be used to measure the importance of researchers in their co-authoring communities.

3.3.3. Robustness of networks

If a path connecting any pair of nodes of an undirected network is constantly present, then this network is called connected. For example, the network in Fig. 1a is connected. Many existing studies have found that collaboration networks are not connected, thereby leading to studies on connected components (Barabási et al., 2002; Moody, 2004; Newman, 2004a). Connected components are connected sub-networks. The network in Fig. 1b has two connected components. In a connected network, information can spread to all other nodes from any node. If we remove some nodes and the edges connecting them from this network, a few paths break and the network may become unconnected. If the majority of nodes in this network remain connected after a few nodes were removed, then this network is said to be robust. Albert, Jeong, and Barabási (2000) found that scale-free networks are fragile if nodes with high degrees are removed. Franceschet (2011) obtained a similar conclusion in a collaboration network of computer scientists. We can use the robustness test to estimate the importance of researchers because the connectivity of a collaboration network can affect the dissemination of scientific information. In the current research, key researchers are important for maintaining the connectivity of a collaboration network.

3.4. Collaboration network analysis

After we have constructed two collaboration networks, community detection is performed in these networks. Furthermore, we identify key researchers in these two networks.

3.4.1. Co-authoring community detection and analysis

One of the objectives of this research is to explore whether co-authoring communities exist in collaboration networks. The current study uses the community detection method introduced in Newman's (2006) work to find co-authoring communities in our two collaboration networks. In the definition of community, nodes in the same community have dense relationships. This feature appears to indicate that if co-authoring communities exist in our collaboration networks, then researchers in the same co-authoring community have dense collaboration relationships and many published articles. To verify this claim, we need to analyze the productivity within and between co-authoring communities. For each co-authoring community, we count the articles that all their authors belong to this co-authoring community. This number is called intra-community productivity. For each co-authoring community, we also count the articles written together by researchers in this co-authoring community and researchers of other co-authoring communities. The number of such articles is called inter-community productivity. In addition, we study the publication dates of the intra-community articles for each co-authoring community to provide a temporal analysis. We explore whether the research interests and locations of researchers in the same co-authoring community are similar to obtain additional information on co-authoring communities.

3.4.2. Identification of key researchers and robustness test

After the analysis of the co-authoring communities in our collaboration networks, we design a systematic method to identify key researchers. In network analysis, the importance of nodes in a network can be reflected by their values of centrality. We calculate the values for three centrality measures for each researcher, namely, degree, betweenness, and community centralities. These centrality measures evaluate the ability of nodes to connect neighbors, be an information hub, and contribute to the formation of communities.

Suppose that scientific information disseminates though paths of a collaboration network; if this network is better connected, then the information can be spread to more researchers. Therefore, a sign of key researchers is their substantial contribution to the connectivity of a collaboration network. Thereafter, a robustness test is performed to verify the importance of key researchers in maintaining the connectivity of collaboration networks. We study the robustness of our collaboration networks by progressively removing a fraction of nodes and measuring the network's connectivity. A list of nodes that will be removed is generated prior to the simulation. At each step, approximately 1% of the nodes in our networks will be removed. We mainly focus on the largest connected components of the collaboration networks in the robustness test. Four removal strategies are applied in our experiments, namely, random, degree-based, betweenness-based, and community centrality-based removal. For the degree-, betweenness-, and community centrality-based strategies, the nodes will be removed in descending order of degree, betweenness, and community centrality, respectively. We calculate the fraction of nodes in the largest cluster to measure the connectivity of the largest connected component after removing the nodes. The experiments are repeated 20 times to avoid biases and the average values are taken to report our results. We would like to investigate which removal strategy will cause the most significant effect on the connectivity of our collaboration networks. The results also help explore which centrality measure can best reflect the importance of researchers in maintaining the connectivity of a collaboration network.

4. Findings and analysis

4.1. Results of co-authoring community detection

This section first studies the connected components of the collaboration network. In the *raw collaboration network*, 783 authors (19.8%) have no collaborators. A total of 733 connected components with at least 2 connected authors are found. The largest connected component has 857 authors (21.7%) and 2794 edges (52.6%). Other components have small sizes. The second largest connected component has only 21 authors. In the *research-article collaboration network*, 439 researchers (12.7%) have no collaborators. The size of the largest connected component is 679 (19.7%). The number of edges in this component is 1276 (33.3%). A total of 16 components with over 10 researchers are found in the *research-article collaboration network*. Thus, in these two collaboration networks, the largest connected components contribute the most in scientific information dissemination and research resource sharing. In the following sections, our analysis is based on the two largest connected components.

To verify our assumption that collaboration networks may be composed of a set of co-authoring communities, we applied the community detection method introduced by Newman (2006) to two largest connected components of our collaboration networks. The modularity values for the detection results of these two components of the *raw collaboration network* and *research-article collaboration network* are 0.6896 and 0.8450, respectively. These high values indicate that those two largest connected components of our collaboration networks have strong community structures. Figs. 2 and 3 visualize the community structures for easy understanding of the detection results. In the detection of co-authoring communities, the value of community centrality for each researcher can be obtained. In these two figures, the sizes of nodes are determined by their community centrality values. Nodes with high community centrality are central in their co-authoring communities, and they have many connections to the nodes in the same communities. Some of these nodes have also been observed to have connections to nodes in other communities, thereby indicating that these nodes act as bridges between communities. This observation means that researchers with a high community centrality are crucial in connecting their co-authoring communities. Moreover, some of these researchers can facilitate the communications between different co-authoring communities.

Fig. 2 shows the detection result of the largest connected component of the *raw collaboration network*. Twenty-eight communities are detected in this component and the largest community is composed of 84 researchers. Nineteen communities have more than 10 researchers and the average size of communities is 31. Co-authoring community assignments are presented by the colors of the nodes. Fig. 2 also illustrates that the edges within the same co-authoring community are denser than those between different co-authoring communities. The largest connected component of the *raw collaboration network* has a total of 2794 edges, while the co-authoring community at the bottom right has 84 nodes and 1374 intra-community edges because 2 articles have 52 authors. An edge constantly exists between any two of these 52 researchers. Thus, the co-authoring community to which these 52 researchers belong includes numerous intra-community edges. We calculate the intra-community productivity defined in the Methodology section for each detected community. The average intra-community productivity is 32. The highest intra-community productivity is 108, and the corresponding community has 58 researchers. In general, communities with high intra-community productivity have relatively large sizes. For example, the average intra-community productivity of 10 communities with the largest sizes is 62.

Fig. 3 shows the community structure of the largest connected component of our *research-article collaboration network*. There are 26 co-authoring communities in this component and the average size is 26. The largest community has 65 researchers. The community structure of

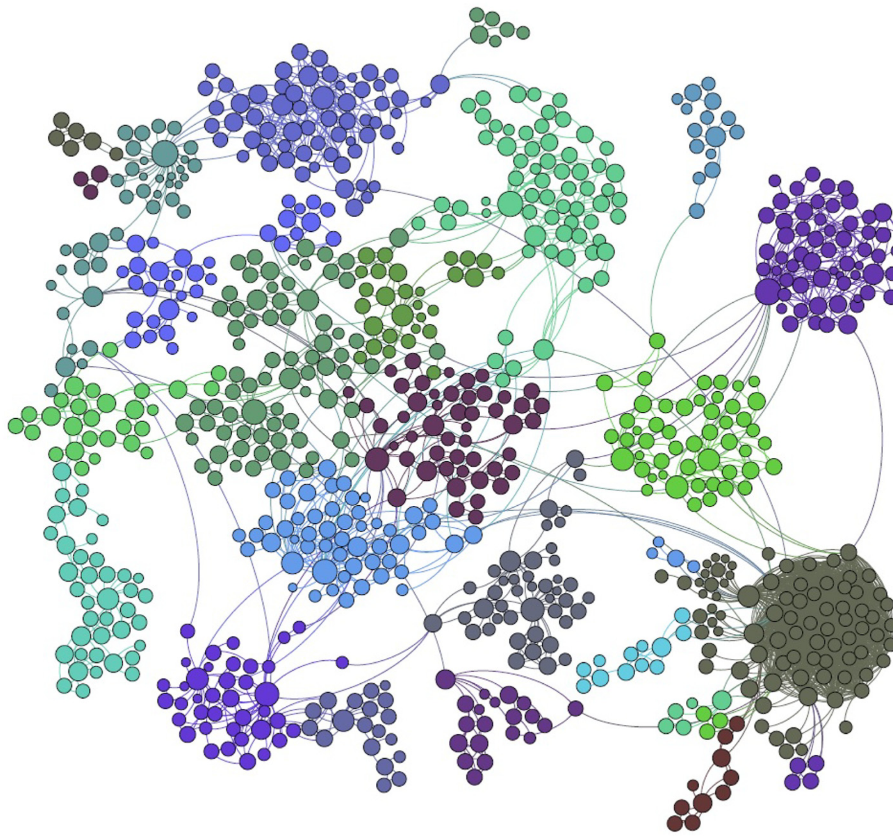


Fig. 2. Community structure of the largest connected component of the *raw collaboration network*.

the *research-article collaboration network* is clearer than the one of the *raw collaboration network*. This structure is also reflected by the modularity values in community detection. This difference indicates that researchers tend to collaborate with those in their communities while publishing research articles. The extremely dense co-authoring community in Fig. 2 does not exist in Fig. 3, because those 2 articles with 52 authors are not research articles. Intra-community productivities are also calculated for this component and the average value is 19. Similar

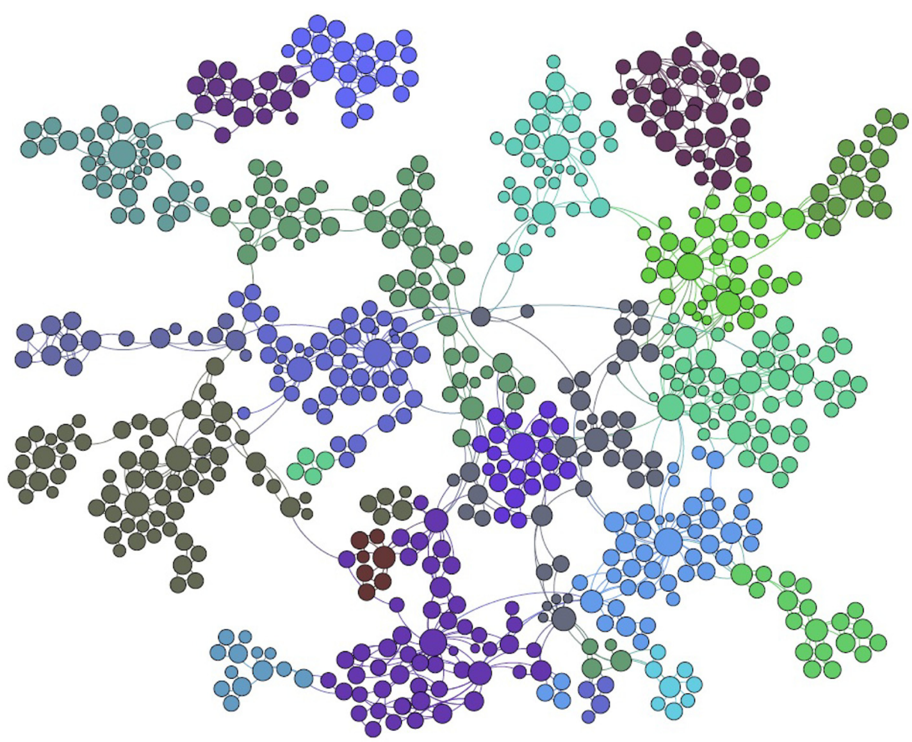
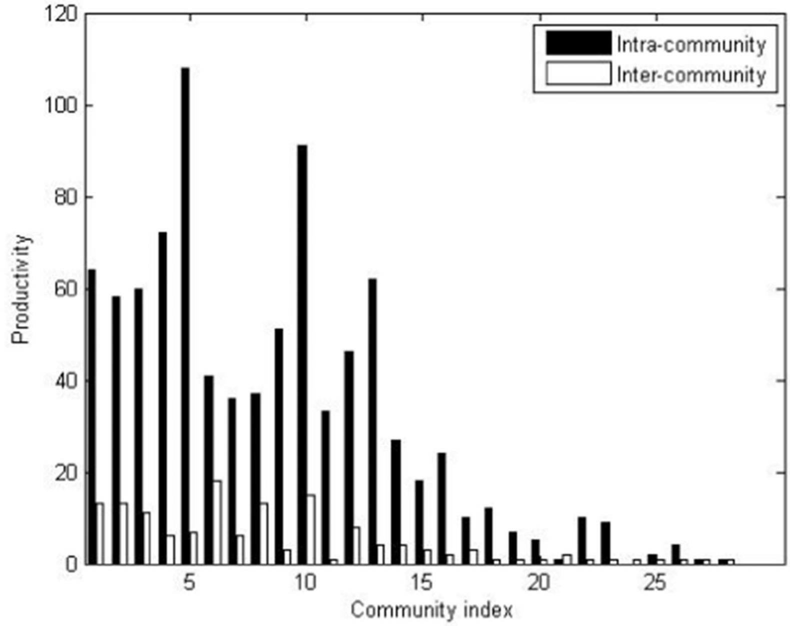
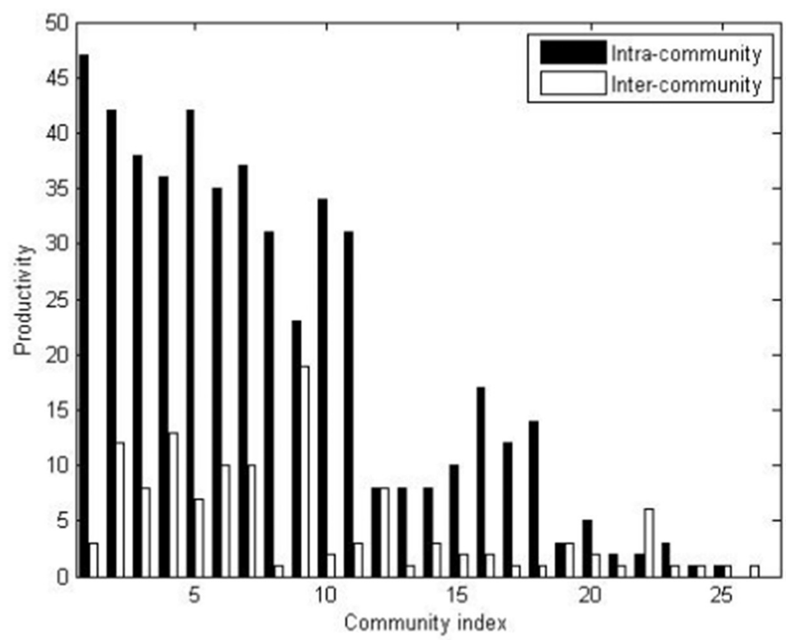


Fig. 3. Community structure of the largest connected component of the *research-article collaboration network*.



(a)



(b)

Fig. 4. Intra-community productivity and inter-community productivity for (a) *raw collaboration network* and (b) *research-article collaboration network*.

to the observation of the communities in Fig. 2, communities composed of more researchers have high probability to have higher intra-community productivities.

Figs. 2 and 3 show that researchers in the same co-authoring community have dense collaboration relationships. We calculate the intra- and the inter-community productivity defined in the Methodology section to verify our assumption that researchers in the same co-authoring community have dense collaboration relationships and numerous published articles. Fig. 4 shows the intra- and inter-community productivity for each co-authoring community in Figs. 2 and 3. In this figure, co-authoring communities are sorted in decreasing order of size. In a few co-authoring communities, inter-community productivity is higher than its intra-community productivity. The reason may be the situation that at least two researchers from different communities collaborate frequently. In this case, only one edge exists between each pair of such researchers, while the number of articles can be high. However, in the majority of co-authoring communities, the intra-community productivity is considerably higher than the inter-community productivity, particularly in co-authoring communities with large sizes. Therefore, co-authoring communities are also clusters of dense publications.

Apart from intra-community productivity, we also study the publication dates of articles with all authors in the same co-authoring community. Table 1 lists of the partial results. In the raw collaboration network, 39% of the co-authoring communities continue to publish during the time span of our study. Researchers in 21% of the co-

Table 1
Number of communities with publications in different time periods.

Time periods	Raw collaboration network	Research-article collaboration network
1982–2015	11 (39%)	6 (23%)
1990–2015	6 (21%)	9 (35%)
2000–2015	5 (18%)	6 (23%)
2010–2015	2 (7%)	2 (8%)

authoring communities published from 1990 to 2015, thereby indicating that these communities did not exist before 1990. Communities also emerged from 2000 and 2010. For the research-article collaboration network, 35% of the co-authoring communities emerged after 1990 and 23% emerged after 2000. The results imply that the community structures of collaboration networks change over time. New co-authoring communities emerged to make the collaboration networks more extensive and better connected, thereby facilitating the dissemination of scientific information.

To explore additional characteristics in co-authoring communities, we randomly select a co-authoring community from each collaboration network and collect the location information of the researchers in this co-authoring community. In the co-authoring community selected from the *raw collaboration network*, the institutes of 64% of the researchers are in Asia. Approximately 43% of the researchers are from universities in South Korea and 25% are from universities in the U.S. In the co-authoring community selected from the *research-article collaboration network*, 45% of the researchers are from universities in Hong Kong and 16% are from Australia. A total of 63% of the researchers in this co-authoring community are from universities in Asia. Co-authoring communities can seemingly aggregate researchers with the same locations. That is, our results indicate that the locations of researchers may be a possible reason for the formation of co-authoring communities.

Moreover, we explore the research interests of researchers in different co-authoring communities. For each co-authoring community, we calculate the frequency of words in titles of the articles published by the researchers in this community. In Table 2, we list the words with relatively high frequency from large communities. Note that we exclude some prepositions and articles with high frequency in this table, such as "of". It is found that some differences exist among co-authoring communities. For example, the word "ISBN" in communities 1 and 5 means that some researchers in these communities have published book reviews. In the community 8 of Table 2, some researchers have published articles related to tourism in Israel. In the community 10, some researchers are interested in hospitality management. Those observations indicate that researchers in different co-authoring communities focus on various research areas.

Table 2
Frequently appeared words in titles from different co-authoring communities.

Raw collaboration network	
1	ISBN; development; sustainable; management; price; London; China; climate.
2	Marketing; destination; case; market; management.
3	Pages; international; development; destination; case; motivation.
4	Management; Hong Kong; pages; hospitality; hotel; strategic; demand.
5	ISBN; management; destination; New Zealand; London; price.
6	Destination; model; behavior; China; case; casino; Korean.

Research-article collaboration network	
7	Service; marketing; impact; community; quality; hotel; international.
8	Destination; case; perceptions; Israel; social; risk.
9	Case; China; Korea; Korean; development; south; tour.
10	Case; international; hotel; industry; cultural; destination; outbound.
11	Hong Kong; case; demand; hotel; management; preferences.
12	Impact; Korean; behavior; case; satisfaction; casino.

4.2. Identification of key researchers

In the majority of previous studies on collaboration networks, key researchers are defined as those with a high degree of centrality or with a high productivity. However, a researcher who contributes substantially to the connectivity of the collaboration network is also important in terms of scientific information dissemination and research resources sharing. Given that co-authoring communities have been discovered in our collaboration networks, we are able to explore how the key researchers in global collaboration networks perform in their co-authoring communities. This section calculates the values of degree, community centrality, and betweenness centrality for each researcher in the largest connected components of two collaboration networks. To investigate which centrality is a sign of importance in maintaining network connectivity, we will measure the importance of researchers using a robustness test introduced in the Methodology section.

Fig. 5 shows the experiment results. The random removal strategy indicates that both networks are robust. In both sub-figures, the sizes of the largest cluster decrease gradually as the fraction of the removed nodes increases because the majority of the nodes in the collaboration networks have only a few connections. The removal of a low fraction of randomly selected nodes will not destroy the network.

Fig. 5 shows that our networks are fragile against degree-based removal. With degree-based removal strategy, the size of the largest cluster decreases considerably faster compared with the curve of random removal for each network. The fraction of the connected nodes in the largest cluster is below 10% of the original largest connected component after only 10% or 5% of the nodes are removed. Thus, the degree-based removal strategy affects the connectivity of the collaboration networks more significantly than the random removal strategy does. This result indicates that researchers with numerous collaborators are important in the global connectivity of collaboration networks.

In the *raw collaboration network*, we apply the community centrality-based and betweenness-based removal strategies and find that the curves for these two removal strategies decline considerably faster than the curve for the degree-based removal strategy. Fig. 5a shows that after 5% of the nodes with top community centrality or top betweenness are removed, the fraction of the connected nodes is only 15% or 11%. However, if 5% of the nodes with top degrees are removed, then over 85% of the nodes of the network continue to belong to the largest cluster. The curves for the community centrality-based and betweenness-based removal strategies are close to one another. However, the curve for the betweenness-based strategy declines slightly faster. If the fraction of the removed nodes is above 10%, then the degree-based removal strategy influences the connectivity of the *raw collaboration network* slightly more than the community centrality-based and betweenness-based strategies do. The experiment shows that excluding a few researchers with high betweenness or community centralities can destroy the global connectivity faster than excluding researchers with high degrees in the *raw collaboration network*.

Different observations are found in the robustness test of the *research-article collaboration network*. Fig. 5b shows that the curves for the degree-based and community centrality-based removal strategies are close, while the curve for the community centrality-based removal strategy decreases slightly faster. Betweenness-based removal strategy has a less significant influence on the *research-article collaboration network* than these two strategies do. The result indicates that researchers with high degrees or community centralities are more important in maintaining network's connectivity than researchers with high betweenness centralities in the *research-article collaboration network*.

In summary, community centrality-based and betweenness-based removal strategies for the *raw collaboration network* influence the connectivity of the network most significantly with only a low fraction of nodes removed. However, for the *research-article collaboration network*, community centrality-based and degree-based removal strategies are the most effective methods to destroy this network. Thus, community

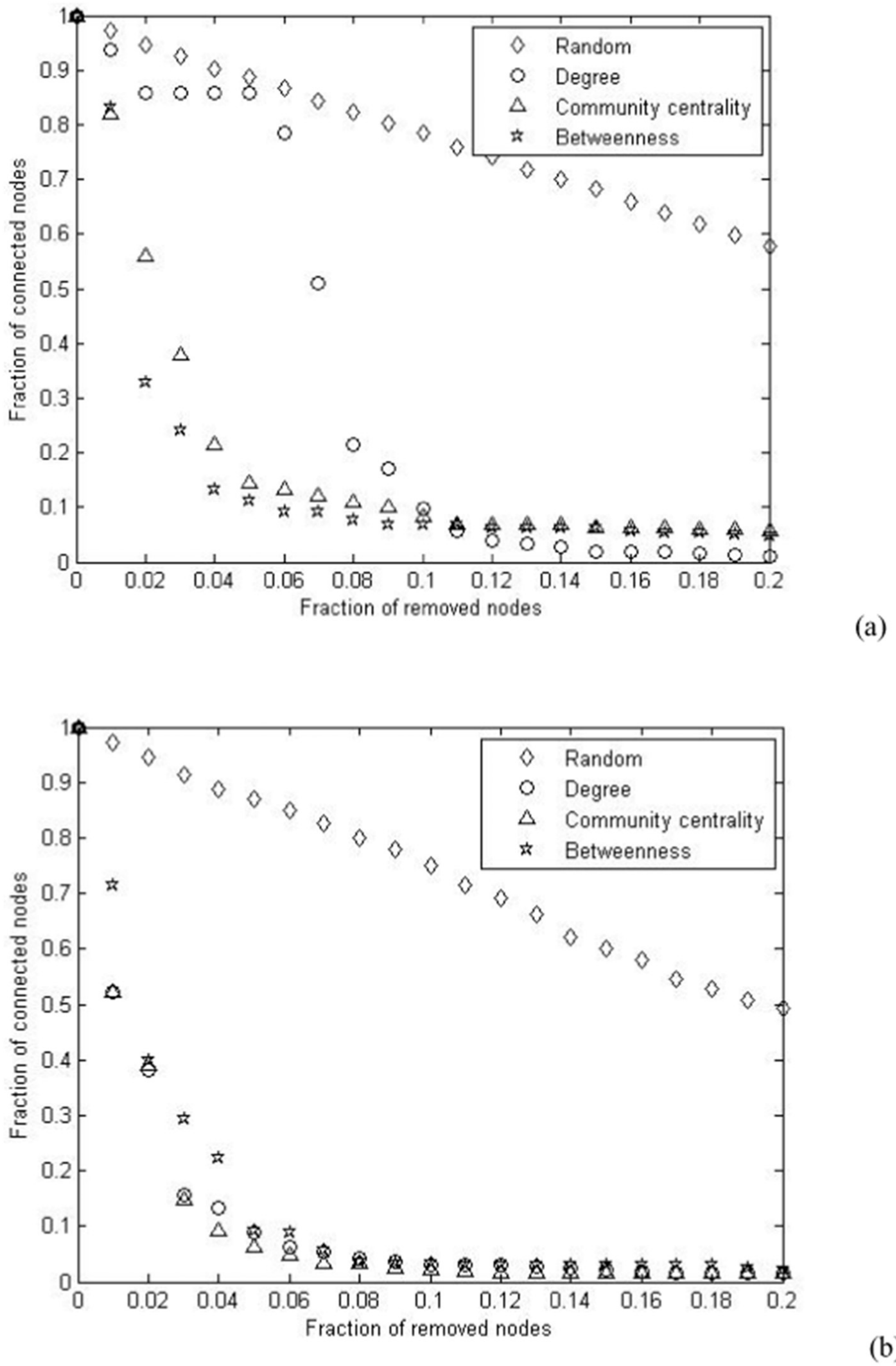


Fig. 5. Robustness test of the largest connected components of (a) *raw collaboration network* and (b) *research-article collaboration network*.

centrality is a good measure to evaluate the importance of researchers in both networks. That is, if some researchers with a high community centrality retire or stop their research, then the collaboration network will be divided into disconnected clusters, thereby leading to a negative effect on knowledge spreading. Moreover, given that researchers with a high community centrality are central in their co-authoring communities, their retirement may lead to the disappearance of these co-authoring communities.

4.3. Implications

We discovered the community structure in collaboration networks of tourism research. Researchers collaborate with those in the same co-authoring community more than with researchers in other co-authoring communities. Consequently, numerous publications are published by researchers within a co-authoring community. We also found that researchers with high community centralities have a significant influence on the connectivity of global collaboration networks. Given that the community centrality reflects researchers' importance in their co-authoring communities, this centrality measure can be used to identify researchers with global and local importance in collaboration networks. The findings and method in this study can be utilized to improve the progress in tourism research.

In this study, temporal analysis of the co-authoring communities was performed and the location information of researchers from selected co-authoring communities was investigated. In the future, additional information on co-authoring communities in collaboration networks can be collected and analyzed. For example, with research on topic analysis of published articles, the research topics that the researchers of a specific co-authoring community are interested in can be discovered. Given this information, co-authoring communities can be better understood to assist the dissemination of scientific information. Moreover, whether these researchers lead the research trends in global collaboration networks and their co-authoring communities can be investigated by analyzing the research topics and corresponding publication dates of the researchers with high community centrality. If this assumption is verified, then how these researchers lead researcher trends and which types of scientific information are important in connecting co-authoring communities and the remainder of the collaboration networks can be further studied. Our findings and method may provide an effective means to understand the trend of tourism research and research clusters in tourism research. Moreover, our method helps in identifying possible reviewers and collaborators, thereby possibly accelerating the spread of scientific information. A possible application of our findings is to implement our method in a search engine, which enables users to find researchers and information on their positions in a collaboration network by providing specific research topics. An example is the search engine developed by Chen and Zhao (2015).

Apart from the collaboration networks of researchers studied in this research, additional research networks can be investigated, such as collaboration networks of universities/institutes and citation networks of researchers/journals. The method and findings in our study can be applied in research networks in various fields to explore their respective development. For example, an analysis can be performed on whether the formation of communities of universities is caused by research interests or geographic locations and identifying key universities. Progress of research can be promoted with such results. In recent years, researchers have applied network analysis in tourism research. Our method of utilizing community centrality can be further applied to research using network analysis, particularly tourism research, to explore new discoveries.

5. Conclusions

In recent years, network analysis has become a popular approach for investigating research collaboration because such a method facilitates the provision of additional information on collaboration relationships. However, the majority of existing studies only investigate the global structure of collaboration networks. Hence, these studies cannot provide information on whether co-authoring communities exist in their networks. To fill in this research gap, the present study focuses on the discovery of co-authoring communities in collaboration networks and what these co-authoring communities mean to research collaboration.

This study constructed two collaboration networks on the basis of two collections of articles on tourism research, namely, *raw collaboration network* and *research-article collaboration network*. One motivation of this research is to determine whether co-authoring communities exist in collaboration networks. We found strong community structures in both collaboration networks. We visualized the detected community structures and revealed that researchers in the same co-authoring communities have denser collaboration relationships. Moreover, we found that in the majority of cases, many articles are produced by researchers in the same co-authoring community. Therefore, a co-authoring community is a set of densely connected researchers and a hidden collection of articles. The detected community structures were found changing over time. New co-authoring communities keep emerging in recent decades. We also studied the locations of researchers in two randomly selected co-authoring communities to learn about co-authoring communities. In each of these two co-authoring communities, over half of the researchers are from Asian universities, thereby implying that researchers in the same co-authoring community have similar characteristics. The investigation of article titles indicates that researchers in different co-authoring communities may have distinct research interests.

Furthermore, we identified key researchers in tourism research, which is another motivation of the current study. Given that co-authoring communities have been discovered in our collaboration networks, the key researchers in co-authoring communities and whether they are important in global collaboration networks could be explored. We perform robustness tests to evaluate the contribution of key researchers on the connectivity of collaboration networks. The results revealed that community centrality is a better measure to evaluate the importance of researchers than degree centrality and betweenness centrality. If researchers with high community centrality are removed from the collaboration networks, then the networks will be destroyed and become fragmented, thereby indicating that these researchers are critical in maintaining the connectivity of the collaboration networks. Given that high community centrality means the central position in a community, a researcher with high community centrality is important to their co-authoring community. That is, community centrality can measure the importance of researchers in their co-authoring community and the global collaboration network.

The present study has verified the fact that collaboration networks of tourism research are composed of co-authoring communities, although limitations exist. Firstly, the definition of "key researchers" in this study covers only a few features of researchers who play important roles in tourism research collaboration networks, and we define them as those who are important in maintaining the network connectivity. Other features of key researchers can be discussed in the future work. For example, researchers who connect different co-authoring communities may bring substantial collaborations between communities and accelerate the dissemination of research resources and scientific information. Moreover, researchers who have different research interests in a co-authoring community may introduce new knowledge to this community. The second limitation is the data source. In this study, data are from one important tourism journal. Including other tourism journals can cover more researchers and articles of broader research areas, so that bias can be reduced. Third, our approach to identifying key researchers cannot perfectly solve the problems in previous work. Not all researchers' academic activities can be covered by our data origins. For example, we aim to identify researchers with global and local importance. Although local importance does not necessarily mean the importance within a spatial community in this study, we are also interested in the important researchers in a particular country or institute. It is possible that in some countries, publishing articles in their native languages is the mainstream. Our approach in the present study cannot identify the researchers publishing with other languages.

For better understanding of collaboration networks, additional actions can be done in the future. We investigated the locations of researchers in two randomly selected co-authoring communities and observed that locations may be a reason of the formation of these detected communities. Additional information of researchers and their publications can be collected in future works to consider other formation reasons of co-authoring communities, such as research interests and location history of researchers. This current study used article titles to analyzed researchers' interests. Other methods to obtain researchers' research interests can be exploring researchers' web pages and analyzing keywords of articles. If researchers' previous publications, including those not in the area of tourism, are collected for investigation, then this situation can also assist future research. Additional journals and other sources of research articles can be included in the future. We may discover how knowledge in other research areas is introduced to the field of tourism and how researchers of other areas affect the community structures of collaboration networks by studying the research interests and publications of researchers. Given that citation is another important dissemination method of scientific information, citation networks of researchers can be investigated in the future. The findings and method in this study can be applied in the collaboration networks of other disciplines. Our findings indicate that further studies can be carried out. For example, by studying the research topics of key researchers in their co-authoring communities, research interests of researchers of different co-authoring communities can be known with less effort. Moreover, if our findings can be implemented in web search engines, then they can be helpful for recommending potential reviewers or collaborators, particularly for cross-discipline studies.

Acknowledgement

The work described in this paper was partially supported by grants from the National Natural Science Foundation of China under Grant No: 71974010. This project was also supported by a research grant funded by the Hong Kong Polytechnic University, Xinjiang Uygur Autonomous Region Research Fund, and Deakin University ASL 2019 fund. The work was completed when Gang Li was on ASL in Chinese Academy of Sciences.

References

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature, 406*(6794), 378–382. https://doi.org/10.1038/35019019.

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(3–4), 590–614. https://doi.org/10.1016/S0378-4371(02)00736-7.

Benckendorff, P. (2010). Exploring the limits of tourism research collaboration: A social network analysis of co-authorship patterns in Australian and New Zealand tourism research. *Proceedings of CAUTHE 2010: Challenge the limits*. Hobart, Australia: University of Tasmania.

Benckendorff, P., & Zehrer, A. (2013). A network analysis of tourism research. *Annals of Tourism Research, 43*, 121–149. https://doi.org/10.1016/j.annals.2013.04.005.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Chen, H., & Zhao, J. L. (2015). ISTopic: Understanding information systems research through topic models. *Proceedings of international conference on information systems (ICIS 2015)* (pp. 1415–1424). Fort Worth: Association for Information Systems (AIS).

Fan, W., Li, G., & Law, R. (2017). Temporal analysis of tourism research collaboration network. *Journal of Hospitality and Tourism Research, 41*(6), 643–672. https://doi.org/10.1177/1096348016640587.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3–5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports, 659*, 1–44. https://doi.org/10.1016/J.PHYSREP.2016.09.002.

Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology, 62*(10), 1992–2012. https://doi.org/10.1002/asi.21614.

Gleiser, P. M., & Danon, L. (2003). Community structure in jazz. *Advances in Complex Systems, 06*(04), 565–573. https://doi.org/10.1142/S0219525903001067.

Hall, C. M. (2011). Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. *Tourism Management, 32*(1), 16–27. https://doi.org/10.1016/j.tourman.2010.07.001.

Huang, J., Zhuang, Z., Li, J., & Giles, C. L. (2008). Collaboration over time: Characterizing and modeling network evolution. *Proceedings of the 2008 international conference on web search and data mining* (pp. 107–116). . https://doi.org/10.1145/1341531.1341548.

Jogaratnam, G., Chon, K., McCleary, K., Mena, M., & Yoo, J. (2005). An analysis of institutional contributors to three major academic tourism journals: 1992–2001. *Tourism Management, 26*(5), 641–648. https://doi.org/10.1016/j.tourman.2004.04.002.

Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics, 40*(3), 541–554. https://doi.org/10.1007/BF02459299.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review, 69*(2), 213–238. https://doi.org/10.1177/000312240406900204.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*(1), 16131. https://doi.org/10.1103/PhysRevE.64.016131.

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E, 64*(1), 16132. https://doi.org/10.1103/PhysRevE.64.016132.

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences, 98*(2), 404–409. https://doi.org/10.1073/pnas.98.2.404.

Newman, M. E. J. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences, 101*(Suppl. 1), 5200–5205. https://doi.org/10.1073/pnas.0307545100.

Newman, M. E. J. (2004b). Who is the best connected scientist?A study of scientific coauthorship networks. In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.). *Complex networks* (pp. 337–370). Berlin Heidelberg: Springer.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E, 74*(3), 36104. https://doi.org/10.1103/PhysRevE.74.036104.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 26113. https://doi.org/10.1103/PhysRevE.69.026113.

Racherla, P., & Hu, C. (2010). A social network perspective of tourism research collaborations. *Annals of Tourism Research, 37*(4), 1012–1034. https://doi.org/10.1016/j.annals.2010.03.008.

Rodriguez, M. A., & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics, 2*(3), 195–201. https://doi.org/10.1016/j.joi.2008.04.002.

Sheldon, P. J. (1991). An authorship analysis of tourism research. *Annals of Tourism Research, 18*(3), 473–484. https://doi.org/10.1016/0160-7383(91)90053-E.

Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*(6825), 268–276. https://doi.org/10.1038/35065725.

Wang, X. F., & Chen, G. (2003). Complex networks: Small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine, 3*(1), 6–20. https://doi.org/10.1109/MCAS.2003.1228503.

Watts, D. J. (2004). The "new" science of networks. *Annual Review of Sociology, 30*, 243–270. https://doi.org/10.1146/annurev.soc.30.020404.104342.

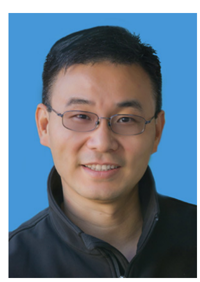
Ye, Q., Li, T., & Law, R. (2013). A Coauthorship network analysis of tourism and hospitality research collaboration. *Journal of Hospitality and Tourism Research, 37*(1), 51–76. https://doi.org/10.1177/1096348011425500.

Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research, 33*(4), 452–473.

Zhao, W., & Ritchie, J. R. B. (2007). An investigation of academic leadership in tourism research: 1985-2004. *Tourism Management, 28*(2), 476–490. https://doi.org/10.1016/j.tourman.2006.03.007.

Zhong, L., Wu, B., & Morrison, A. M. (2015). Research on China's tourism: A 35-year review and authorship analysis. *International Journal of Tourism Research, 17*(1), 25–34. https://doi.org/10.1002/jtr.1962.

Wei Fan is a research associate at the School of Hotel and Tourism Management, The Hong Kong Polytechnic University. Her research interests are in complex networks, social network analysis, big data analytics, and their applications in tourism management.



Gang Li IEEE Senior member, is an Associate Professor at the School of Information Technology, Deakin University. His research interests are in the area of data mining, machine learning, business intelligence, tourism/hospitality management and decision support systems. He has coauthored seven best paper awarded articles.

Rob Law is a Professor at the School of Hotel and Tourism Management, the Hong Kong Polytechnic University. His research interests are information management and technology applications.