A Bibliographic Analysis and Collaboration Patterns of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS Between 2000 and 2015

Xiujuan Xu, Wei Wang, Yu Liu, Xiaowei Zhao, Zhenzhen Xu, and Hongmei Zhou

Abstract—Intelligent transportation systems (ITS) has been one of the most active research fields in recent years. This paper identifies most productive authors, institutions, and countries/regions in IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS from 2000 to 2015. The results of bibliographic analysis show that the USA is the most influential country in that it not only has the most papers but also has six out of the ten most-cited papers. Meanwhile, researchers from China and Europe have published nearly half of the papers in this field. In addition, we generate three networks (including coauthorship network, keyword co-occurrence network, and author co-keyword network) to analyze collaboration patterns among authors in the field of ITS. The active keywords are investigated, and the top three are vehicles, road vehicles, and road traffic. Finally, visual pictures are presented to show topological interactions of authors' collaboration.

Index Terms—Bibliographic analysis, intelligent transportation, social network analysis, research collaboration.

I. INTRODUCTION

INTELLIGENT transportation systems (ITS) refers to a comprehensive transportation system utilizing transportation, service control, and vehicle manufacturing. Since there is a close relationship between transportation systems and our daily life, ITS has become one of the most active research fields in recent years.

IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (T-ITS) is one of the top journals in the field of intelligent transportation systems [1]. It launched in 2000 and has 16 years of publishing history. The journal focuses on the cutting-edge of the ITS field and has the highest impact factor (2.377 in 2015) in the ITS field.

In this paper, we collect and analyze the papers published in the IEEE T-ITS from the first volume in 2000 to the 15th volume in 2015. We present a framework about a bibliographic analysis and implement a collaboration analysis system for a journal called Journal-CAS (Journal Collaboration Analysis System). We identify the most productive authors, institutions and countries/regions in the IEEE T-ITS. We predict the development trend by analyzing top six keywords so as to provide a reference to other researchers. In addition, we generate three networks (including co-authorship network, keyword co-occurrence network and author co-keyword network) to analyze the collaboration patterns. Visual pictures are presented to show the relationship among the researchers. The active keywords are investigated, such as road vehicles and road traffic. The major contents of this paper are summarized as below.

1) We present a research framework for the journal analysis. First, we collect the meta data on IEEE T-ITS. Then, we propose a model for analyzing a journal's trend.

2) We analyze the bibliographic statistics for the journal. The results of the bibliographic statistics show that the USA is the most influential country in the ITS field because it has a large number of outstanding researchers and excellent institutions. In recent years, China and Europe also play significant roles in this field.

3) We construct three networks, including co-authorship network, keyword co-occurrence network and author co-keyword network. Based on the formal definitions of three networks analysis problems, we visualize the results of the three networks.

The remainder of this paper is organized as follows. Section II gives a brief review on related literature. Section III provides a research framework about journal trend analysis. Section IV shows the bibliographic analysis results. In Section V, we present collaboration. Remarks are stated in the conclusion in Section VI.

II. RELATED WORK

In recent years, intelligent transportation systems has become an important research area. In this section, we introduce related
work in this field. The majority of related work are classified into the bibliographic analysis and collaboration patterns.

A. The Bibliographic Analysis

Over the last decade, some researchers have analyzed bibliography from the data of T-ITS. In 2010, Li et al. [2] presented a bibliographic analysis of the papers published in the T-ITS and identified the most productive and high-impact authors, institutions, and countries/regions. They found that U.S. researchers and institutions dominated in the ITS field. Consequently, Li et al. [3] provided three networks (research level co-authorship, institution-level co-authorship and country level co-authorship) and analyzed ten years data of 2000–2009 from the IEEE T-ITS by using Girvan-Newman algorithm (GN algorithm) [4]. Wang et al. [5] analyzed four years data of 2010–2013 with the method of clustering. They found that some enterprises made great contribution to this field, such as Toyota in Japan. Tang et al. [6] constructed keyword networks and found popular topics, according to which they classified the papers and predicted the development trend of the topics.

Previous work also showed that h-index [7] was a parameter of author influence considering both quality and quantity. It focused on the number of influential papers that scientists have published.

B. Collaboration Patterns

The collaboration patterns of researchers in a research specialty form scientific community, which could help to understand the field-specific shaping of scientific communication practices [8]. Collaboration patterns analysis are generally classified into three classes: identifying the collaboration patterns among authors, institutions, and countries, constructing keyword co-occurrence network and author co-keyword network [3].

Co-authorship network aims to find the cooperative relationship and interpret collaboration patterns between authors [3]. Velden et al. [9] studied collaboration patterns in co-authorship networks at the mesoscopic level. They found two types of co-author-linking patterns between author clusters, which were interpreted as representing two different forms of cooperative behavior, transfer-type connections due to career migrations or one-off services rendered, and stronger, dedicated inter-group collaboration. As a consequence, Velden and Lagoze [8] described a network analytic approach that revealed the complexities of scientific communities through the examination of their publication networks in combination with insights from ethnographic field studies.

Keyword co-occurrence network is a keyword network, which helps us to find active research topics. Su and Lee [10] presented an approach to visualize a knowledge structure. Their approach created a three-dimensional “Research focused parallelship network,” a “Keyword co-occurrence network,” and a two-dimensional knowledge map to facilitate visualization of the knowledge structure created by journal papers from different perspectives.

Author co-keyword network is used to investigate authors with common interests. For example, Liu and Ram [11] identified collaboration patterns that were preferable or detrimental for data quality, thus providing insights for improving data quality in Wikipedia.

III. OVERVIEW OF ANALYZING A JOURNAL

In this section, we present two parts.

(1) We design a framework of data analysis for a journal.
(2) We give the processing model for a journal from IEEE T-ITS.

A. Framework of Data Analysis

The data analysis framework is divided into five modules: data source, data extraction, key problems, the bibliographic analysis, and collaboration patterns visualization, as shown in Fig. 1.

(1) Data source: The dataset is obtained from the website of IEEE Transactions on Intelligent Transportation Systems [1]. Each piece of data is conducted by some important information about a paper, including the paper’s title, abstract, keywords, citation numbers, authors’ names, affiliations, countries/regions.

(2) Data extraction: We extract 16 years data from the volume 1, 2000 to the issue 6, 2015 in IEEE Xplore website using a web crawler. We only keep the data with link information, and do not consider the data without link information, because they usually are review or guest editorial.

Due to the authors’ name on IEEE Xplore website are abbreviations, we extract the authors’ full name form
DBLP dataset. Each paper’s citation number is extracted from Google Scholar and Web of Science by us.

All data is heterogeneous because it includes titles, abstracts, authors, author affiliations, keywords and citation numbers.

(3) Key problems: In order to construct three networks, three problems are signed in this part, including co-authorship network, keyword co-occurrence network, author co-keyword network.

Co-authorship network: We can find the authors’ social circles by clustering collaboration relationships. Usually, authors from the same cluster have similar research direction. Collaboration pattern is also a significant indicator. Network analysis tool could help us to obtain effective information. We would identify the authors’ collaboration patterns by GN algorithm.

Keyword co-occurrence network: By clustering keywords, we can identify those keywords of papers in pairs frequently. And these keywords can be clustered into several topics. In this network, each edge represents two keywords appeared in one article.

Author co-keyword network: A keyword usually indicates a topic, so in co-keyword network each edge also means two authors focused on a similar field. In another word, each edge represents two authors used similar keywords. In other words, they have common interests.

(4) The bibliographic analysis: We present the distribution by year of IEEE T-ITS. Then we select the most productive authors, institutions and countries/regions. We investigate the frequency of all the keywords and extract 6 keywords to analyze their evolution. Finally, we find all papers’ citation numbers. In this paper, we crawl the top 12 authors’ h-index from Google Scholar in Section IV.

(5) Collaboration patterns visualization: We can use GN algorithm [4] to cluster collaboration patterns among authors. The results are clearer through visualization software. Three networks, including co-authorship network, author co-keyword network and keyword co-occurrence network, are also shown by UCINET software [12].

B. Processing of Journal-CAS

The Collaboration Analysis System for a journal (Journal-CAS) works as follows. First of all, we get data from IEEE website and clean the data. We extract authors’ full names from DBLP dataset and get citation numbers from Google Scholar and Web of Science. Then, we can identify the most productive authors, institutions, countries/regions and keywords development trend. We also crawl authors’ h-index from Google Scholar. At last, we construct three networks by GN algorithm and visualize them by UCINET NetDraw [12]. The overview of a journal analysis is shown in Fig. 2.

We use the 16 year data to analyze the situation of the ITS field. We can see the most influential authors, institutions and countries/regions. In addition, we analyze the keywords which researchers focused and identify the collaboration patterns of this field. And we hope that these results can help researchers to understand ITS better.

IV. THE BIBLIOGRAPHIC ANALYSIS

In this section, the results of the bibliographic analysis based on the experimental evaluation using IEEE T-ITS database are shown. We analyze the most productive authors, institutions and countries/regions in last 16 years. Furthermore, we investigate the keywords and identify the most cited paper.

A. Paper Statistics

In our dataset, there are 1517 papers in IEEE T-ITS. We find 4038 authors from 671 institutions in 52 countries/regions. Fig. 3 shows every year’s paper numbers.

B. Author Statistics

Table I shows the 12 most productive authors. Most of those authors are from the USA and China and have high h-index.

C. Institution Statistics

Table II shows the results of institution statistics. Many universities have multiple campuses. Each campus is often located in different areas and has different research level. In our work, we distinguish them and calculate every branch campus’s paper numbers. Compared with paper [6], the institutions in top 10 are similar and the number of Chinese universities is the most. University of California and University of Minnesota are still
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

### D. Country/Region Statistics

Table III reports the most productive countries/regions. It can be seen clearly that the USA makes up nearly a quarter. Thus this field is still dominated by the USA. However, we also notice that China is the second largest part, which indicates that China is in a very important position now. Fig. 4 shows the ratio of all countries/regions.

### E. Keyword Statistics

We investigate the frequency of all the keywords occurred in 16 years in Table IV. Indeed, we delete the keyword “Intelligent transportation systems,” because it is the journal’s name of our dataset.

### F. Trends of Top Six Keywords

We extract 6 keywords of the most frequently occurred and analyze their evolution. The trends of top six keywords are shown in Fig. 5. Fig. 5 shows that the six keywords are on the rise, but they decreased in 2003 and 2013 slightly. Especially, the trends of six keywords decreased in 2008 and 2009 from Fig. 5. The number of “Vehicles” is more than 170 in 2015, which indicates that researchers pay more attention to vehicles themselves.
TABLE V
MOST CITED PAPERS (FROM GOOGLE SCHOLAR)

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Cited</th>
<th>Country/Region</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Detection and classification of vehicles</td>
<td>852</td>
<td>USA</td>
<td>Surendra Gupta, Osama Masoud, Robert F. K. Martin, Nikolaos P. Papanikolopoulos</td>
</tr>
<tr>
<td>2</td>
<td>A review of conflict detection and resolution modeling methods</td>
<td>798</td>
<td>USA</td>
<td>James K. Kuchar, Lee C. Yang</td>
</tr>
<tr>
<td>3</td>
<td>Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation</td>
<td>692</td>
<td>USA</td>
<td>Joel C. McCall, Mohan M. Trivedi</td>
</tr>
<tr>
<td>4</td>
<td>Automatic license plate recognition</td>
<td>684</td>
<td>Taiwan</td>
<td>Shyang-Lih Chang, Li-Shien Chen, Yun-Chung Chung, Sei-Wang Chen</td>
</tr>
<tr>
<td>5</td>
<td>Challenges of intervehicle ad hoc networks</td>
<td>586</td>
<td>USA</td>
<td>Jeremy J. Blum, Azim Eskandarian, Lance J. Hoffman</td>
</tr>
<tr>
<td>6</td>
<td>Detecting stress during real-world driving tasks using physiological sensors</td>
<td>571</td>
<td>USA</td>
<td>Jennifer Healey, Rosalind W. Picard</td>
</tr>
<tr>
<td>7</td>
<td>A License Plate-Recognition Algorithm for Intelligent Transportation System Applications</td>
<td>497</td>
<td>Greece</td>
<td>Patricia Besson, Christophe Bourdin Lionel Bringoux, Erick Doussset, Christophe Maiano, Tanguy Marqueste, Daniel Mestre, Sophie Gaetan Jean-Pierre Baudry, Jean-Louis Vercher</td>
</tr>
<tr>
<td>8</td>
<td>Travel-time prediction with support vector regression</td>
<td>494</td>
<td>Taiwan</td>
<td>Horst F. Wedde, Sebastian Sengen</td>
</tr>
<tr>
<td>9</td>
<td>Stereo-and neural network-based pedestrian detection</td>
<td>452</td>
<td>USA</td>
<td>Wei Lu, Weiguo Song, Jian Ma, Zhiming Fang</td>
</tr>
<tr>
<td>10</td>
<td>Freeway ramp metering: an overview</td>
<td>450</td>
<td>Greece</td>
<td>Kichun Jo, Junsoo Kim, Myoungho Sunwoo</td>
</tr>
</tbody>
</table>

TABLE VI
MOST CITED PAPERS (FROM WEB OF SCIENCE, WITHOUT SELF-CITATION)

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Cited</th>
<th>Country/Region</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation</td>
<td>340</td>
<td>USA</td>
<td>Joel C. McCall, Mohan M. Trivedi</td>
</tr>
<tr>
<td>2</td>
<td>Detection and classification of vehicles</td>
<td>280</td>
<td>USA</td>
<td>Surendra Gupta, Osama Masoud, Robert F. K. Martin, Nikolaos P. Papanikolopoulos</td>
</tr>
<tr>
<td>3</td>
<td>A review of conflict detection and resolution modeling methods</td>
<td>254</td>
<td>USA</td>
<td>James K. Kuchar, Lee C. Yang</td>
</tr>
<tr>
<td>4</td>
<td>Automatic license plate recognition</td>
<td>247</td>
<td>Taiwan</td>
<td>Shyang-Lih Chang, Li-Shien Chen, Yun-Chung Chung, Sei-Wang Chen</td>
</tr>
<tr>
<td>5</td>
<td>Detecting stress during real-world driving tasks using physiological sensors</td>
<td>233</td>
<td>USA</td>
<td>Jennifer Healey, Rosalind W. Picard</td>
</tr>
<tr>
<td>6</td>
<td>Travel-time prediction with support vector regression</td>
<td>213</td>
<td>Taiwan</td>
<td>Horst F. Wedde, Sebastian Sengen</td>
</tr>
<tr>
<td>7</td>
<td>Challenges of intervehicle ad hoc networks</td>
<td>183</td>
<td>USA</td>
<td>Jeremy J. Blum, Azim Eskandarian, Lance J. Hoffman</td>
</tr>
<tr>
<td>8</td>
<td>Freeway ramp metering: an overview</td>
<td>181</td>
<td>Greece</td>
<td>Kichun Jo, Junsoo Kim, Myoungho Sunwoo</td>
</tr>
<tr>
<td>9</td>
<td>Research advances in intelligent collision avoidance and adaptive cruise control</td>
<td>179</td>
<td>USA</td>
<td>Ardalain Vahida, Azim Eskandarian</td>
</tr>
<tr>
<td>10</td>
<td>Real-time system for monitoring driver vigilance</td>
<td>166</td>
<td>Spain</td>
<td>Luis Miguel Bergasa, Jesús Nuevo, Miguel A. Sotelo, Rafael Barea, María Elena Lopez</td>
</tr>
</tbody>
</table>

G. Impact Analysis

Although IEEE website has citation information of papers, it is not completely accurate. In this work, we use Google Scholar to find all papers’ citation numbers (before January 8th, 2016). Citation number is related to impact. A paper has the more impact if the paper has the more citation number. In Table V, it is shown that most of them are from the USA, and their authors include some famous researchers, such as Mohan M. Trivedi and Miguel Angel Sotelo [23]. Thus the USA has more impact than other countries [13]–[22].

However, there is significant number of self-citations on Scholar for citation numbers from Google Scholar. If we delete self-citations, we could use Web of Science to find all papers’ citation numbers (before January 8th, 2016). Thus we redo the experiment to get Table VI. Similarly, Table VI demonstrates the dominance of publications coming from the USA. The most productive author Mohan M. Trivedi published a paper named “Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation” [15]. This paper is the most cited paper (340) in Table VI. The paper’s topics are “driver information systems” and “road vehicles.”

V. COLLABORATION PATTERNS

In this section, we find collaboration patterns by using GN algorithm. We construct three networks, including co-authorship network, keyword co-occurrence network and author co-keyword network.

A. Problem Definitions

Suppose there are a set of papers \( P = \{p_1, \ldots, p_n\} \) published by \( m \) authors \( A = \{a_1, \ldots, a_m\} \). A paper dataset is denoted by a bipartite graph \( G = (P, A, E) \) where authors published papers. The set of vertices has two subsets: one set \( P \) denotes all papers, and the other set \( A \) denotes all authors. The set \( E \)
of edges represents authors published papers. An edge \( e_{i,j} \) in \( E \) represents an author \( a_i \) published a paper \( p_j \).

An author relationship network \( N \) is introduced so that another relationship for co-authorship can be defined. \( N = \{ P, A, E, B, W \} \). The set \( B \) of edges represents collaboration relationships between authors. If an author \( a_x \) and an author \( a_y \) published a paper \( p_j \) together, \( B \) includes an element \( B_{xy} \).

The weight set \( W \) of edges denotes the collaborated times between two authors. Given \( | \cdot | \) denote the cardinality of a set, and \( |P| = n, |A| = m, |E| = r, |B| = s, |W| = t \).

Average degree denotes the average number of collaboration relationships in a author relationship network [3]. The formal definition of average degree is shown in Equation (1).

\[
\text{Average degree} = \frac{|B|}{|A|} = \frac{s}{m}. \tag{1}
\]

Clustering coefficients represent whether the authors tend to work alone or collaborate in groups. The formal definition of clustering coefficients is shown in Equation (2).

\[
\text{Definition 2:} \quad \text{Clustering coefficients. Given a subset } N' = \{ P', A', E', B', W' \}, \quad |P'| = n', \quad |A'| = m', \quad |E'| = r', \quad |B'| = s', \quad |W'| = t' \quad \text{as a clustering, the average degree of the clustering is:}
\]

\[
\text{Average degree} = \frac{|B'|}{|A'|} = \frac{s'}{m'}. \tag{2}
\]

Clustering coefficients represent whether the authors tend to work alone or collaborate in groups. The formal definition of clustering coefficients is shown in Equation (2).

\[
\text{Definition 2:} \quad \text{Clustering coefficients. Given a subset } N' = \{ P', A', E', B', W' \}, \quad |P'| = n', \quad |A'| = m', \quad |E'| = r', \quad |B'| = s', \quad |W'| = t' \quad \text{as a clustering, the collaboration relationship } B = \{ b_1, b_2, \ldots, b_N \}.
\]

\[
k_i \text{ is the number of nodes which are next to a node } b_i, \quad e_n \text{ is the number of edges between } k_i \text{ nodes. } c_i \text{ is the local clustering coefficients of } b_i \]

\[
c_i = \frac{e_{ii}}{\frac{k_i(k_i - 1)}{2}}. \tag{3}
\]

Then, clustering coefficients \( c \) of the clustering is:

\[
c = \frac{1}{n} \sum c_i. \tag{4}
\]

Another symbol \( \omega(A') (\forall A' \subseteq A) \) represents finding a set of authors \( A' \) in which an author \( a_i (a_i \subseteq A') \) has at least one paper co-published with another author \( a_j (a_j \subseteq A') \), find a partition of authors satisfied the two following conditions:

\[
A = A_1 \cup A_2 \cup \ldots \cup A_k \quad \tag{5}
\]\

\[
A_m \cap A_n = \emptyset (m, m \leq k, m \neq n). \tag{6}
\]

Edge betweenness is defined to be the number of geodesic shortest paths between vertex pairs that run along this edge.

\[
\text{Definition 3: Edge betweenness. Betweenness of an edge } i = \sum E_{mn}, \quad E_{mn} \text{ is an edge from one node } m \text{ to another node } n, \quad \text{and it must run along the edge } i.
\]

Here Modularity \( Q \) is a measure of a community structure presented by Girvan M. and Newman M. E. J. [4]. Assuming the current network is divided into \( k \) communities, \( e_{ii} \) represents the proportion of the edges which connect two nodes in a community \( i \). \( a_i \) is the sum of elements in each row(or each column). It represents the ratio of all edges attached to the nodes in a community \( i \).

\[
\text{Modularity } Q = \sum (e_{ii} - a_i^2). \tag{7}
\]

B. Co-Authorship Network Analysis

Co-authorship relations directly reflect collaboration between scientists [3]. In this paper, we present collaboration patterns in the ITS field at individual researcher level. Before we constructed co-authorship network, we first introduce some terms of co-authorship network [24].

\[
\text{Definition 5: Co-authorship network analysis problem. Given a relationship network } N = \{ P, A, E, B, W \}, \text{ find } \omega(A) \text{ such that Modularity } Q \text{ is maximized.}
\]

Based on the above definitions, collaboration patterns analysis algorithm could be presented as follows.

\[
\text{Algorithm 1} \quad \text{Collaboration patterns analysis algorithm.}
\]

Input: A relationship network \( N = \{ P, A, E, B, W \} \).
Output: The final status of network \( N' = \{ B', A \} \). \( B' \) is remaining relationship edges and represents network state at the time of the biggest Modularity \( Q \).

1) Calculate the betweenness for all edges in the network.
2) Remove the edge with the highest betweenness. Recalculate Modularity \( Q \) and record corresponding network structure.
3) Recalculate betweennesses for all edges affected by the removal.
4) Repeat from step 2 until no edges remain. Select the biggest Modularity \( Q \)’s network structure as ultimately divided state.

Co-authorship network has 4038 nodes and 848 components. In our experiment, the value of Modularity \( Q \) is between 0.5–0.7. The three largest components have 188, 71, and 68 nodes respectively. The average degree is 1.97. In this network, Ding Wen [25] has the largest number of collaborators (degree = 49). Figs. 6–8 show the three largest components. We use GN algorithm to cluster authors, so different colors
mean different clusters. And the size of a node represents the node’s degree. If an author has the more collaborators, the node of the author has the larger size.

The largest component has 188 nodes in Fig. 6. The average degree is 3.55 and clustering coefficient is 0.36. Ding Wen [25], Fei-Yue Wang [26], Tao Tang [27] and Bin Ning [28] are productive authors in Table I. With their collaboration relationships, the network is dramatically strong. And many authors are from China. Ding Wen is a professor in the Center for Military Computational Experiments and Parallel Systems Technology at the National University of Defense Technology [25]. As one of the most productive authors, Ding Wen plays a significant role in the largest component. Fei-Yue Wang is a professor of the Key Laboratory of Complex Systems and Intelligence Science at the Chinese Academy of Sciences [26]. His research interests include intelligent systems, social computing and complex systems. He received the 3rd high h-index in Table I.

The second largest component has 71 nodes in Fig. 7. The average degree is 5.17 and clustering coefficient is 0.71. Xiang Cheng [31] with the second largest degree (32) is in the second largest component. He is an Assistant Professor with Peking University and is currently an Associate Editor of the IEEE T-ITS. His research interests include mobile propagation channel modeling and simulation and intelligent transportation systems.

The third largest component has 68 nodes in Fig. 8. The average degree is 2.79 and clustering coefficient is 0.456.

C. Keyword Co-Occurrence Network Analysis

The keyword co-occurrence network includes has 391 keywords and 1857 links. It refers to similarity between keywords. And similar keywords represent a topic. Two keywords have closer relationship when they co-occurred more time. So we make a rule that we only record the keywords co-occurred at least ten times.

Before giving the formal definition of keyword co-occurrence network analysis problem, we first introduce some terms used in this subsection. Suppose there is a set of keywords $K = \{k_1, \ldots, k_n\}$. The set $C$ of edges represents co-occurrence relationships between papers. An edge $C_{i,j}$ represents one keyword $k_i$ and another keyword $k_j$ published in the same paper. The frequency set $F$ of edges denotes the frequency of two keywords co-occurrence.

Definition 6: Keyword Co-occurrence Network Analysis problem. Given a relationship network $NKC = \{P, K, C, F\}$, find $\omega(K)$ such that Modularity $Q$ is maximized.

In Fig. 9, we use GN algorithm to cluster keywords and identify 13 topics. Each cluster has its own color. We regard the largest degree node as the topic of this cluster. And we do not consider the word “Intelligent transportation systems.” We use “Road vehicles” to represent the red cluster.

We analyze the network in three periods of time. Firstly, the basic properties of three subnetworks are shown.

2000–2004: There are 49 keywords, 86 links and 9 topics, as shown in Fig. 10.

Alberto Broggi [30] with the largest degree of this component is 30. Both he and Mohan M. Trivedi [15] are productive authors.
There are 76 keywords, 164 links and 9 topics, as shown in Fig. 11.

2010–2015: There are 284 keywords, 1070 links and 13 topics, as shown in Fig. 12.

We find some useful things from Figs. 10–12. For example, the word “Road vehicles” is very active in the three networks. So this topic is always popular in this field. In addition, both “Air traffic control” and “Global Positioning System” are on a rise. So researchers begin to pay more attention to the two topics. And we also find that many new topics emerge, such as “Vehicular ad hoc networks,” “Fuzzy control.” Therefore, we could infer that scientists use some methods coming from computer science in recent years from Fig. 12.

2.6. Author Co-KeyWord Network Analysis

Author co-keyword network indicates authors’ common interests. To focus on outstanding authors’ common interest, we request that every author’s paper number must be over 2 and two authors’ co-keywords number must be over 20.

Before presenting the problem definition of author co-keyword network analysis, we first give the following terms. The set $\mathcal{A}K_{i,j}$ of edges represents that an author $a_i$ and another author $a_j$ used the same keyword $k_j$. The weight set $T$ of edges denotes the keyword number which the two authors used together.

Definition 7: Author Co-keyword Network Analysis problem. Given a relationship network $\mathcal{N}AC = \{P, A, K, \mathcal{A}K, T\}$, find $\omega(A)$ such that Modularity $Q$ is maximized.

2.6.1. Author Co-keyword Network Analysis

In our network, there are 218 nodes and 539 links. Authors are partitioned into common interest groups by using GN algorithm. We identify 22 topics as shown in Fig. 13. Consequently, we mainly analyze the top three clusters.

Red cluster: It is dominated by Mohan M. Trivedi [10]. He does research in many topics. This group has the most authors. They study some topics about vehicles, such as road vehicles, vehicle dynamics and intelligent vehicles.

Yellow cluster: Markos Papageorgiou [8] is the center node. Most of them are interested in some topics about optimal control, such as real-time optimal, traffic flow modeling, and traffic management.

Blue cluster: Tao Tang [20] plays a key role. They may do research in some topics about traffic control, such as train control, road traffic and railway systems. This group also includes Bin Ning [21]. They do research in traffic control together many times.

VI. CONCLUSION

In this paper, we provided a framework for data analysis of a single journal. We collected and analyzed the papers published in the IEEE T-ITS from 2000 to 2015. The major contributions of this study included two aspects as follows.

In aspect of the bibliographic analysis, we find that the USA dominates in the ITS field, having a large number of productive authors with high h-index and famous institutions. For example, Mohan M. Trivedi from University of California, San Diego [15] is the most productive author. In addition, the USA has many papers with high citation numbers and high impact. Meanwhile, researchers from China and Europe have published a lot of papers and make a great contribution to this field.

On the other hand, in aspect of collaboration patterns, we generate three networks to reveal the development of the ITS field. According to the collaboration relationships among authors, we construct co-authorship networks to analyze authors’ collaboration patterns. Then, we construct the keyword co-occurrence network to follow the tracks of keywords’ evolution. Some active keywords were identified, such as road vehicles, traffic control, Global Positioning System, road safety. Obviously, there will be more papers about them in the future. Finally, we present the author co-keyword network to show common interests of authors.
ACKNOWLEDGMENT

The authors would like to thank the anonymous referees and editors for their helpful comments and suggestions, which led to this revised version.

REFERENCES


Xiujuan Xu received the M.S. degree in 2005, the Ph.D. degree in 2008, from College of Computer Science and Technology, Jilin University, ChangChun, China. She is currently an Associate Professor of Software School in Dalian University of Technology. Her research interests include applications of data mining, intelligent transportation systems, recommender systems, and social network.

Wei Wang received the B.S. degree in 2013 from Dalian University of Technology, Dalian, China, where she is currently working toward the master’s degree with the School of Software. Her research interests include applications of big data and taxi trajectory flow.

Yu Liu received the Ph.D. degree from Xi’an Jiaotong University, Xi’an, China, in 2006. He is currently a Professor with the School of Software, Dalian University of Technology, Dalian, China. His research interests include swarm intelligence, evolutionary computation, computational intelligence, and web data mining.

Xiaowei Zhao received the M.S. degree from Queen Mary University of London, London, U.K., in 2005. She is currently working toward the doctorate degree with the Faculty of Management and Economics, Dalian University of Technology, Dalian, China. She is currently an Assistant Professor with the School of Software, Dalian University of Technology. Her research interests involve complex systems, recommender systems, and gaming theory.
Zhenzhen Xu received the Ph.D. degree from Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, in 2009. She is currently an Assistant Professor and Master’s Research/Thesis Advisor. Her research field is information processing in complex distributed system. Her current research interests include big data analysis and social computing.

Hongmei Zhou received her Ph.D. degree in Transportation and Urban Engineering from University of Connecticut, USA, in 2009. Her research interests include Traffic Safety, Intelligent Transportation Systems (ITS), and Transportation Planning and Management. She obtained significant experience from a range of projects on traffic accident data analysis, driver behavior evaluation, roadway design and traffic signal optimization. She has also served as paper reviewer for journals and conferences, including Transportation Research Record and IEEE Vehicular Technology Conference.