Social Media Based Transportation Research: the State of the Work and the Networking

Yisheng Lv, Yuanyuan Chen, Xiqiao Zhang, Yanjie Duan, and Naiqiang Li

***Abstract*—Recently, there has been an increased interest in the use of social media data as important traffic information sources. In this paper, we review social media based transportation research with social network analysis methods. We summarize main research topics in this field, and report collaboration patterns at levels of researchers, institutions, and countries, respectively. Finally, some future research directions are identified.**

***Index Terms*—Social media, transportation, traffic information, social transportation, traffic prediction, traffic event detection.**

I. INTRODUCTION

SOCIAL media have evolved dramatically in recent years, and now is a great source of real-time user-generated contents. Social media applications like Twitter and Sina Weibo for smart phones and tablets have been widely used which allow people to publish and distribute information and opinions easily. It makes every user as a social sensor to detect the real world at any time. During the past few years, social media data have been effectively used to detect natural disasters, monitor epidemics, response crisis, analyze sentiment, and so on. Mining social media data can help people to sense the world.

Accurate and timely traffic information is the fundamental and of vital importance to the success of transportation operations [1], [2]. Many individual users and transportation agencies publish real-time traffic information like traffic jams and traffic incidents through social media platforms. This information allows traffic management centers to provide efficient and safe transportation services including traffic signal control, transit scheduling, traveler information, etc. It also can help individuals to adjust their travel schedules (trip start time, trip modes, trip routes, trip destinations) and en-route/destination choices accordingly.

Extracting traffic related information from social media platforms has attracted interests of researchers both in academics and industries. Because of the popularity of social media, the abundance of social media data is available. Social media as social traffic sensors have advantages over traditional physical traffic sensors, such as area coverage, building cost and maintenance cost. It can provide traffic information that physical sensors cannot, and it also can work as supplementary information for physical sensors [3], [4]. Mining the contents of social media can help to better understand traffic events in terms of when, where and why it happens.

To gain in-depth understanding of social media based transportation research, we use social network analysis methods to analyze recent advances in this field. The rest of the paper is organized as follows. Section II introduces the data set and the methodology. Section III presents the collaboration patterns among researchers, institutions and countries, respectively. Section IV analyzes keywords and main research topics in this field. Section V concludes this paper.

II. DATA AND METHODOLOGY

A. Data Sources

We conducted a keyword search of multiple electronic citation databases, full text databases, and search engines, to collect papers published on scientific journals, magazines and conference proceedings. Electronic databases used in this research are as follows: IEEE Xplore digital library, ACM digital library, Elsevier Science Direct, Springer, Web of Science Core Collection: Citation Indexes (including Science Citation Index Expanded, and Conference Proceedings Citation Index-Science), Engineering Village, Wanfang Database, and Google Scholar. However, using keywords alone in multiple databases can include non-transportation and duplicate papers in the search results. Therefore, we firstly did an automated filtering process to exclude duplicate search results, and then we read the titles and abstracts of the identified papers to exclude those papers without relevance to our research questions. Finally, we got 67 papers published from the year of 2011 to the year of 2015, of which 18 were journal articles, 41 were conference papers, and 8 were others.

B. Methodology

This paper adopts a social network analysis method to analyze collaborations in this filed at three levels, i.e., individual researcher level, institutional level, and country level. This

Manuscript received April 27, 2016; accepted November 8, 2016. This work was partially supported by National Natural Science Foundation of China (61533019, 71232006, 61233001) and the Early Career Development Award of SKLMCCS. Recommended by Associate Editor Fei-Yue Wang.

Citation: Y. S. Lv, Y. Y. Chen, X. Q. Zhang, Y. J. Duan, and N. Q. Li, "Social media based transportation research: the state of the work and the networking," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 1, pp. 19−26, Jan. 2017.

Y. S. Lv, Y. Y. Chen, and Y. J. Duan are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with Qingdao Academy of Intelligent Industries, Qingdao 266109, China (e-mail: yisheng.lv@ia.ac.cn; yychen5133@ia.ac.cn; duanyanjie2012@ia.ac.cn).

X. Q. Zhang is with the Department of Transportation Engineering, School of Transportation Science and Technology, Harbin Institute of Technology, Harbin 150090, China (e-mail: zxqjuly@hit.edu.cn).

N. Q. Li is with Qingdao Academy of Intelligent Industries, Qingdao 266109, China (e-mail: naiqiang.li@qaii.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2017.7510316

method has been widely used in previous studies on scientific literature review and scientific team work analysis [5]−[8].

In a researcher level coauthor network, each node represents an author and there is a link between two authors if they have coauthored at least one paper. In an institution coauthor network, a node represents an institution, and a link connects two nodes if authors from the two institutions coauthored at least one paper. In a country level coauthor network, nodes are countries, and two nodes are connected if at least one paper is coauthored via affiliations in the two countries.

We use the word cloud to illustrate keywords in the collected papers in this study. The word cloud of keywords depicts occurrence frequencies of keywords. Research topics and interests of authors can be easily seen from the word cloud.

III. Collaboration Pattern Analysis

Since social media based transportation research is a new field, and there are only 67 papers in the dataset, the coauthor networks are relatively disconnected. We analyze the topological measures on the whole networks and some meaningful sub-networks, respectively. In the dataset, 199 authors from 77 institutions in 17 countries contributed these papers. The top four publishers are IEEE, ACM, Springer, and Transportation Research Board. The papers were published on more than 40 sources. The top three publication sources are the IEEE International Conference on Intelligent Transportation Systems (ITSC) [9]−[15], Transportation Research Board Annual Meeting, and Lecture Notes in Computer Science, which have more than 25% of publications in the dataset.

Fig. 1 gives the number of publications on social media in transportation in recent 5 years, i.e., from the year of 2011 to the year of 2015. It clearly shows that there is a big leap on the number of papers published in the latest two years compared with that published in the first three years. The number of papers published in 2014 and 2015 are more than two times of that published in 2011, 2012, and 2013. These results demonstrate that the subject on social media in transportation will attract more and more researchers' interest and the number of social media based transportation publications is still increasing.

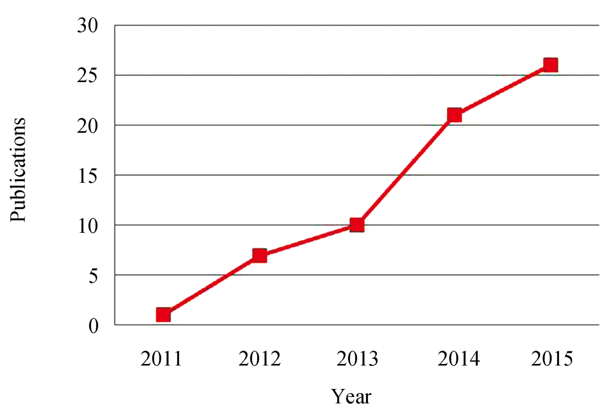


Fig. 1. The number of publications on social media in transportation over time.

A. Collaborations Among Researchers

We constructed a coauthor network to analyze collaborations among researchers based on the journal and conference papers. It is easy to understand how scientific work collaborated among researchers via a coauthor network.

At the researcher level, most of the papers are coauthored. Fig. 2 reports the distribution of the number of authors per paper. The typical number of coauthors in a paper is three and four.

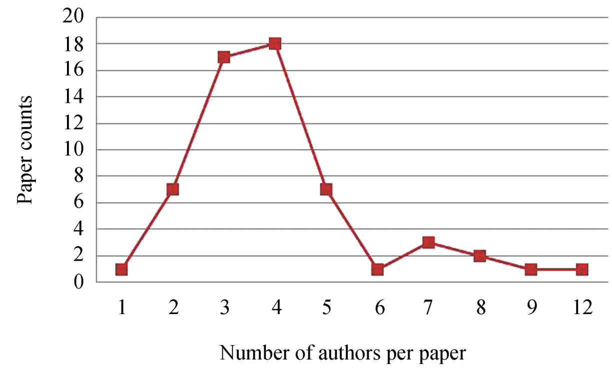


Fig. 2. Distribution of the number of authors per paper.

The researcher level coauthor network consists of 193 nodes, 470 links, and 41 connected components. The network density is 0.022. The average degree of the network is 4.28. The clustering coefficient is 0.894. Fig. 3 visualizes the coauthor network. As a new filed of social media based transportation research, the research groups are quite decentralized and relatively isolated.

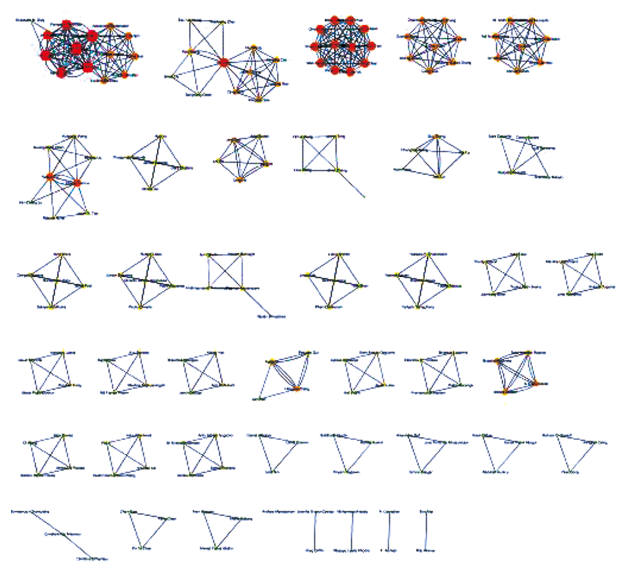
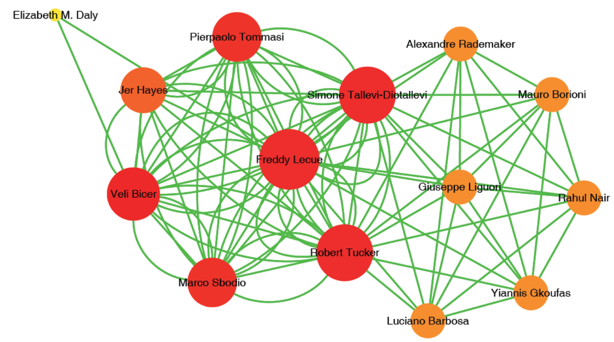


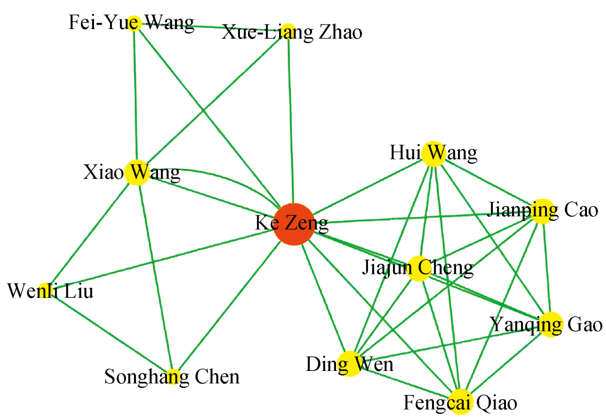
Fig. 3. Researcher level coauthor network.

The largest component (see Fig. 4 (a)) has 14 nodes. In this component, the average degree of the network is 8. The node Freddy Lecue has the largest degree which is 27. The clustering coefficient is 0.899. The network centralization is 0.449. The authors in this cluster are from IBM Dublin Research Centre, SRM-Retie Mobilita, and IBM Rio Research Centre.

The second largest cluster (see Fig. 4 (b)) represents the main collaborations in China. It has 12 nodes. The average degree of the network is 5.33. The node Ke Zeng has the largest degree which is 12. The clustering coefficient is 0.915. The network centralization is 0.618. The members in this cluster are from Institute of Automation, Chinese Academy of Sciences, Xi'an Jiaotong University and National University of Defense and Technology.



(a)



(b)

Fig. 4. Two largest components of the researcher lever coauthor network.

B. Collaborations Among Institutions

The 67 papers are produced from 77 institutions. There are 15 institutions publishing more than two papers. IBM Dublin Research Centre with 5 papers ranked the most productive institution. The second most productive institutions are Institute of Automation, Chinese Academy of Sciences, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, and Virginia Tech, all of which have 3 papers in the dataset. It is worth mentioning that some papers are produced by research branches of one institute, such as IBM (including IBM Dublin Research Centre [16]–[20], IBM Research-India [21], and IBM Rio Research Centre [16]), and Chinese Academy of Sciences (including Institute of Automation [13], [22], [23] and Institute of Geographic Sciences and Natural Resources Research [24]–[26]).

The institution level coauthor network has 77 nodes with 48 connected components. Fig. 5 visualizes the institution level coauthor network. There are 26 isolated nodes. The clustering coefficient is 0.225. The average degree of the network is 0.909. Three institutions have the degree of 4, and they are Xi'an Jiaotong University, National University of Defense and Technology, and Institute of Automation, Chinese Academy of Sciences. Virginia Tech has the degree of 3 and comes from the biggest cluster having four nodes.

C. Collaborations Among Countries

In the past five years, authors from 17 countries have published social media based transportation research papers. The top three countries are USA (21), China (15), and Ireland (6). The total number of papers published by authors from the USA, China and Ireland has dominated more than 50% of the papers in the dataset, which is shown in Fig. 6.

The country level coauthor network is constructed, which has 17 nodes and is shown in Fig. 7. The clustering coefficient is 0.176, and the average degree of the network is 0.941. It has only two connected components, and there are 8 isolated nodes which is a high number. The giant cluster has 6 nodes. The USA has the largest number of collaborative countries with degree of 5, followed by China with degree of 3. Authors from the USA and China have the strongest coauthorship, and at the same time the two countries produced most of the papers in this field.

IV. Keyword Analysis and Research Topics

A. Keyword Analysis

Keywords reflect research topics and interests of authors. By analyzing keywords, we can have a general view of one field.

Fig. 8 shows the word cloud of keywords in the collected papers in this study, which depicts the occurrence frequency of keywords. Keywords with higher occurrence frequencies have bigger sizes in Fig. 8. From Fig. 8, we can see that social media, information extraction, traffic information, traffic incident, incident detection, sentiment analysis, natural language processing, twitter, microblog, social network, and text mining, are among most frequently used keywords. Twitter and Sina Weibo are microblogging service which have spread in recent years and have a large number of users. With tremendous user generated contents, Twitter and Sina Weibo have become a new kind of real-time information sources which mainly present information in the form of texts. Natural language processing methods or text mining methods are needed to extract useful information from unstructured texts. In the field of transportation science and technology, social media are usually used to extract traffic events, traffic incidents, traffic information, and traffic sentiment.

B. Research Topics

The internet and social media have evolved dramatically in the past decade. Widespread deployment of smart mobile devices and social networks have provided a huge amount of contents generated by users, making crowdsourcing a useful source to extract real-time information in many fields like health care, public security as well as improving urban mobility [27], [28]. As a further analysis from the word cloud in Fig. 8, current research based on social media is mainly on four folds, which are traffic information extraction and visualization, traffic event detection, traffic information prediction, and traffic sentiment analysis.

Traffic information extraction and visualization from social media is one of the earliest social media based transportation research and application. At the beginning, only traffic inform-

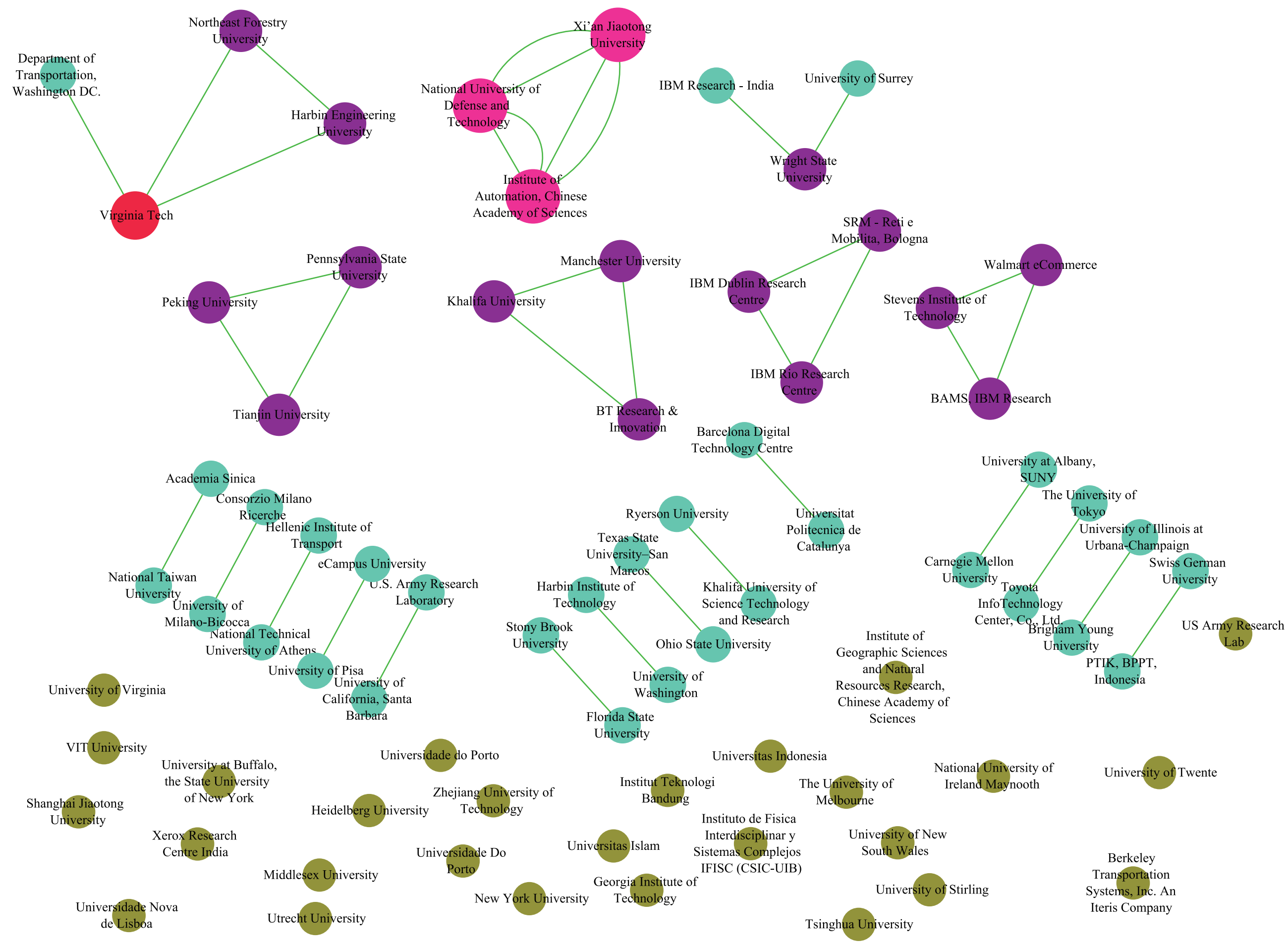


Fig. 5. Institution level coauthor network.

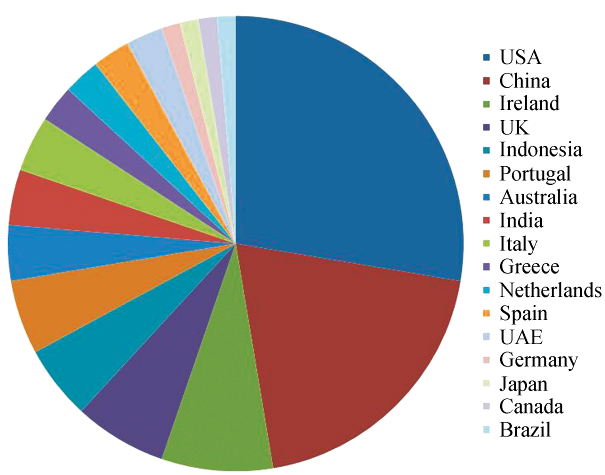


Fig. 6. Productivity over countries.

ation from public social media accounts is extracted to be visualized. Endarnoto et al. extracted traffic information from the Twitter account of TMC (traffic management center) Polda Metro Jaya and visualized the information in a map view in an Android-based mobile application [29]. Later, traffic information from other sources is also used. Freddy Lécué *et al*. presented a system named STAR-CITY supporting semantic analytics and reasoning for city traffic [16]−[19]. The STAR-CITY integrates structured and unstructured data, static and stream data. It can analyze, diagnose, explore and predict traffic scenarios such as spatio-temporal analysis of traffic status and prediction of road traffic conditions, using semantic web technologies. They reported lessons learnt from its deployment and experimentation in Dublin (Ireland), Bologna (Italy), Miami (USA) and Rio (Brazil). Singh developed a system to display traffic information extracted from sources of Twitter TFL Traffic News Profile, TFL Traffic Syndicated Feeds, and Google traffic information [30].

Social media has become one of main channels for public event announcements. Detecting traffic events from social media like Twitter data and Sina Weibo data is a hot research topic in social media based transportation research, which focuses on traffic incident detection, traffic congestion detection, etc [4], [9], [10]. We can readily get traffic related user generated contents when we search Sina Weibo with keywords like traffic jam, traffic accidents. For example, "Traffic jam again on the way to work, $\cdots$ ". D'Andrea *et al*. proposed a real-time traffic event detection system from Twitter stream analysis with text mining techniques. The traffic event detection system was deployed for monitoring several areas of the Italian road network. The authors claimed that the system can detect traffic events almost in real time and often before online traffic news web sites [31]. Xiong combined conditional random field models and regular expression to extract traffic events from Sina Weibo [32]. Cui *et al*. developed a prototype system that

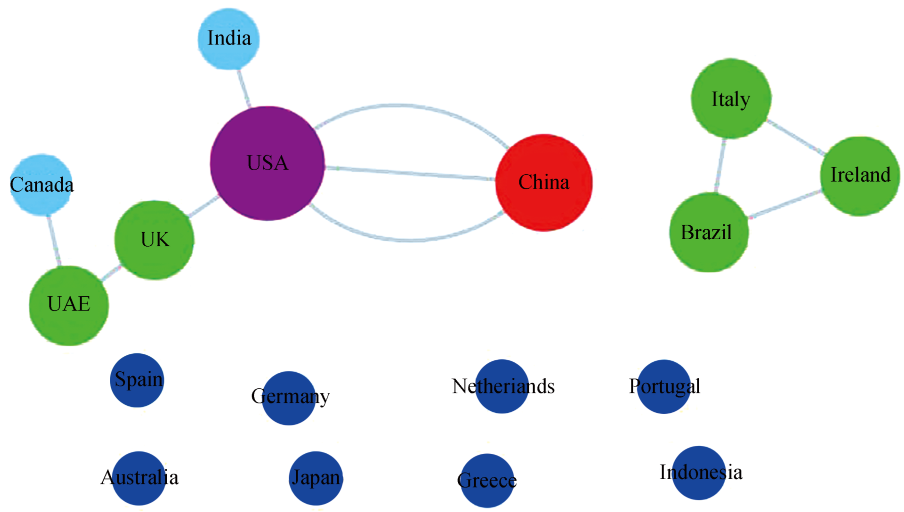


Fig. 7. Country level coauthor network.

Fig. 8. Word cloud of keywords in all papers in our dataset.

extracts traffic accidents and traffic states from Sina Weibo and published these traffic events through an Android based application [14], [33]. Gutiérrez *et al*. detected traffic related events by integrating and fusing tweet messages from traffic agencies in UK [34]. Tejaswin *et al*. used Twitter data to extract traffic related entities with background knowledge from structured data repositories, and then used this spatiotemporal data to cluster and predict traffic incidents [35]. Kurkcu *et al*. extracted traffic incident information from Twitter data and incorporated incident information into their proposed virtual sensor methodology to collect real-time traffic data from online and open sources [3], [36].

Social media can provide external information and insights on traffic prediction. Some researchers have reported that incorporating social media information can improve traffic information prediction, and there exists relationships between social media and traffic information. He *et al*. used Twitter data to predict longer-term traffic volume prediction where the forecasting horizon is beyond 1 hour [37]. They first established the correlation between traffic volume and tweet counts. And then they extracted traffic indicators based on tweet semantics. Finally, they proposed a linear regression model incorporation traffic data and Twitter data to predict traffic flow. Experimental results demonstrate the improved performance of the proposed model over the existing auto-regression based traffic flow prediction model. Ni *et al*. developed a short-term traffic flow prediction model under sport game events [38]. They incorporated the tweet rate and tweet sentiments as social media features into the prediction model. Experiments show that including social media information can improve traffic flow forecasting performances. Grosenick proposed artificial neural networks to predict traffic speed on a single road segment, in which non-recurring events like traffic accidents are considered in the model and those events are extracted from twitter data [39]. Abidin *et al*. predicted bus arrival time with a Kalman filter model incorporating traffic information from social networks [40], [41]. Zhang *et al*. first extracted traffic information from Sina Weibo, and then used a neural network based fuzzy-C-means clustering method to get road traffic conditions with highest confidence levels [24].

Sentiment analysis has developed rapidly and been widely applied to a variety of applications ranging from marketing to customer service with the spread of social media. Traffic sentiment analysis is now drawing more and more attentions of researchers and city administrators. Zeng *et al*. analyzed attentions users gave to various topics concerning Golden Week in China with topic clustering methods [22]. Cao *et al*. proposed a web-based traffic sentiment analysis system and they used the system to analyze two cases in China, i.e., the yellow light rule and the fuel price in China [23]. Semwal *et al*. developed some machine learning methods for event detection, sentiment analysis and suggestion classification using social media data along with external sources like weather and news data [42].

V. CONCLUSION

The potential of social media has been increasingly recognized by transportation researchers. This paper reviews recent advances in this field, in which we analyze researcher-, institution-, and country-level collaboration networks, and research topics. We found that the networks are relatively sparse which implies the future great development. USA and China have dominated this field. Authors from the USA and China published most of the papers in this field and have the strongest coauthorship. The current research topics based on social media focus on traffic information extraction and visualization, traffic event detection, traffic information prediction, and traffic sentiment analysis.

In addition to social media, other sources of online websites have rich traffic related information, such as official websites of traffic management and operations, web-based map service providers (like Bing map, Google map, and Baidu map), weather forecasting websites, and local events (sport games, music concerts, etc.) broadcast websites [3], [36], [43], [44]. In the future, we expect there will be a trend incorporating all of these online open sources to detect, predict and deduce traffic patterns. Online open data sources and social media based transportation research is an interdisciplinary field, which is just at the beginning stage and becoming more and more attractive. There is a lot of room to improve performance in various ways like improving event extraction accuracy and prediction accuracy. How and to what extent the impact of social media information on traffic management is a very interesting topic which we believe that it will be hot in the near future and have great potential on managing city traffic. Moreover, social media is text-rich in sentiment and emotions on traffic and transportation. Further research focus would be put on discovering people's opinions, advices, discussions on transportation planning, management, and operations to support decision making of traffic administrators.

REFERENCES

[1] N. Zhang, F.-Y. Wang, F. H. Zhu, D. B. Zhao, and S. M. Tang, "Dynacas: computational experiments and decision support for its," *IEEE Intell. Syst.*, vol. 23, no. 6, pp. 19−23, Nov.–Dec. 2008.

[2] G. Xiong, F. H. Zhu, X. S. Dong, H. S. Fan, B. Hu, Q. J. Kong, W. W. Kang, and T. Teng, "A kind of novel its based on space-air-ground big-data," *IEEE Intell. Transport. Syst. Mag.*, vol. 8, no. 1, pp. 10−22, Jan. 2016.

[3] A. Kurkcu, E. F. Morgul, and K. Ozbay, "Extended implementation method for virtual sensors: Web-based real-time transportation data collection and analysis for incident management," in *Transportation Research Board 94th Annual Meeting*, Washington DC, USA, 2015, pp. 27−37.

[4] S. Zhang, J. J. Tang, H. Wang, and Y. H. Wang, "Enhancing traffic incident detection by using spatial point pattern analysis on social media," *Transport. Res. Record: J. Transport. Res. Board*, no. 2528, pp. 69−77, 2015.

[5] Q. P. Zhang, Z. Feng, X. Li, X. L. Zheng, and L. Zhang, "25 years of collaborations in IEEE intelligent systems," *IEEE Intell. Syst.*, vol. 25, no. 6, pp. 67−75, Dec. 2010.

[6] L. J. Li, X. Li, C. J. Cheng, C. Chen, G. Y. Ke, D. D. Zeng, and W. T. Scherer, "Research collaboration and its topic evolution: 10 years at T-ITS," *IEEE Trans. Intell. Transport. Syst.*, vol. 11, no. 3, pp. 517−523, Aug. 2010.

[7] T. Wang, X. Wang, S. Tang, Y. Lin, W. Liu, Z. Liu, B. Xiu, D. Shen, X. Zhao, and Y. Gao, "Collaborations patterns and productivity analysis for IEEE T-ITS between 2010 and 2013," *IEEE Trans. Intell. Transport. Syst.*, vol. 15, no. 6, pp. 2360−2367, Dec. 2014.

[8] X. Wang, X. H. Zheng, Q. P. Zhang, T. Wang, and D. Y. Shen, "Crowdsourcing in ITS: the state of the work and the networking," *IEEE Trans. Intell. Transport. Syst.*, vol. 17, no. 6, pp. 1596−1605, Jun. 2016.

[9] S. Zhang, "Using twitter to enhance traffic incident awareness," in *Proc. 2015 IEEE 18th Int. Conf. Intelligent Transportation Systems (ITSC)*, Las Palmas, 2015, pp. 2941−2946.

[10] K. Q. Fu, C.-T. Lu, R. Nune, and J. X. Tao, "Steds: Social media based transportation event detection with text summarization," in *Proc. 2015 IEEE 18th Int. Conf. Intelligent Transportation Systems (ITSC)*, Las Palmas, 2015, pp. 1952−1957.

[11] E. Chaniotakis and C. Antoniou, "Use of geotagged social media in urban settings: Empirical evidence on its potential from twitter," in *Proc. 2015 IEEE 18th Int. Conf. Intelligent Transportation Systems (ITSC)*, Las Palmas, 2015, pp. 214−219.

[12] G. Bajaj, G. Bouloukakis, A. Pathak, P. Singh, N. Georgantas, and V. Issarny, "Toward enabling convenient urban transit through mobile crowdsensing," in *Proc. 2015 IEEE 18th Int. Conf. Intelligent Transportation Systems (ITSC)*, Las Palmas, 2015, pp. 290−295.

[13] X. Wang, K. Zeng, X. L. Zhao, and F. Y. Wang, “Using web data to enhance traffic situation awareness,” in *Proc. 2045 IEEE 17th Int. Conf. Intelligent Transportation Systems (ITSC)*, Qingdao, China, 2014, pp. 195−199.

[14] J. Cui, R. Fu, C. H. Dong, and Z. Zhang, “Extraction of traffic information from social media interactions: Methods and experiments,” in *Proc. 2014 IEEE 17th Int. Conf. Intelligent Transportation Systems (ITSC)*, Qingdao, 2014, pp. 1549−1554.

[15] A. Kocatepe, J. Lores, E. E. Ozguven, and A. Yazici, “The reach and influence of dot twitter accounts: A case study in Florida,” in *Proc. 2015 IEEE 18th Int. Conf. Intelligent Transportation Systems (ITSC)*, Las Palmas, 2015, pp. 330−335.

[16] F. Lécué, R. Tucker, S. Tallevi-Diotallevi, R. Nair, Y. Gkoufas, G. Liguori, M. Borioni, A. Rademaker, and L. Barbosa, “Semantic traffic diagnosis with star-city: Architecture and lessons learned from deployment in Dublin, Bologna, Miami and Rio,” in *The Semantic Web-ISWC 2014*, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, and C. Goble, Eds. Switzerland: Springer International Publishing, 2014, pp. 292−307.

[17] F. Lécué, R. Tucker, V. Bicer, P. Tommasi, S. Tallevi-Diotallevi, and M. Sbodio, “Predicting severity of road traffic congestion using semantic web technologies,” in *The Semantic Web: Trends and Challenges*, V. Presutti, C. dAmato, F. Gandon, M. dAquin, S. Staab, and A. Tordai, Eds. Switzerland: Springer International Publishing, 2014, pp. 611−627.

[18] F. Lécué, S. Tallevi-Diotallevi, J. Hayes, R. Tucker, V. Bicer, M. L. Sbodio, and P. Tommasi, “Star-city: semantic traffic analytics and reasoning for city,” in *Proc. the 19th Int. Conf. on Intelligent User Interfaces*, New York, NY, USA, 2014, pp. 179−188.

[19] F. Lécué, S. Tallevi-Diotallevi, J. Hayes, R. Tucker, V. Bicer, M. Sbodio, and P. Tommasi, “Smart traffic analytics in the semantic web with star-city: Scenarios, system and lessons learned in dublin city,” *Web Semant.: Sci., Serv. Agent. World Wide Web*, vol. 27−28, pp. 26−33, Aug.-Oct. 2014.

[20] E. M. Daly, F. Lécué, and V. Bicer, “Westland row why so slow? fusing social media and linked data sources for understanding real-time traffic conditions,” in *Proc. 2013 Int. Conf. Intelligent User Interfaces*, Santa Monica, California, USA, 2013, pp. 203−212.

[21] P. Anantharam and B. Srivastava, “City notifications as a data source for traffic management,” in *Proc. 20th ITS World Congr.*, Tokyo, Japan, 2013.

[22] K. Zeng, W. L. Liu, X. Wang, and S. H. Chen, “Traffic congestion and social media in China,” *IEEE Intell. Syst.*, vol. 28, no. 1, pp. 72−77, Feb. 2013.

[23] J. P. Cao, K. Zeng, H. Wang, J. J. Cheng, F. C. Qiao, D. Wen, and Y. Q. Gao, “Web-based traffic sentiment analysis: Methods and applications,” *IEEE Trans. Intell. Transport. Syst.*, vol. 15, no. 2, pp. 844−853, Apr. 2014.

[24] H. C. Zhang, F. Lu, and J. Chen, “Extracting traffic information from massive micro-blog messages,” *J. Imag. Graph.*, vol. 18, no. 1, pp. 123−129, Jan. 2013.

[25] P. Y. Qiu, H. C. Zhang, and F. Lu, “A pattern matching method for extracting road traffic information from internet texts,” *J. Geo-Inform. Sci.*, vol. 17, no. 4, pp. 416−422, Apr. 2015.

[26] H. C. Zhang, F. Lu, and P. Y. Qiu, “Extracting traffic information from micro-blog based on D-S evidence theory,” *J. Chin. Inform. Process.*, vol. 29, no. 2, pp. 170−178, Mar. 2015.

[27] T. Sakaki, M. Okazaki, and Y. Matsuo, “Tweet analysis for real-time event detection and earthquake reporting system development,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 919−931, Apr. 2013.

[28] L. Zhao, F. Chen, C.-T. Lu, and N. Ramakrishnan, “Spatiotemporal event forecasting in social media,” in *Proc. 2015 SIAM Int. Conf. Data Mining*, Vancouver, BC, Canada, 2015, pp. 963−971.

[29] S. K. Endarnoto, S. Pradipta, A. S. Nugroho, and J. Purnama, “Traffic condition information extraction & visualization from social media twitter for android mobile application,” in *Proc. 2011 Int. Conf. Electrical Engineering and Informatics (ICEEI)*, Bandung, 2011, pp. 1−4.

[30] B. Singh, “Real time prediction of road traffic condition in London via Twitter and related sources,” M.S. thesis, Middlesex University, UK, 2012.

[31] E. D’Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, “Real-time detection of traffic from Twitter stream analysis,” *IEEE Trans. Intell. Transport. Syst.*, vol. 16, no. 4, pp. 2269−2283, Aug. 2015.

[32] J. X. Xiong, “Civil transportation event extraction from Chinese microblogs based on CRF,” M.S. thesis, Shanghai Jiao Tong University, Shanghai, 2014.

[33] J. Cui, X. Feng, and Z. Zhang, “Extraction and analysis system of traffic incident based on microblog,” *Journal of Transport Information and Safety*, vol. 31, no. 6, pp. 132−135, 2013.

[34] C. Gutiérrez, P. Figuerias, P. Oliveira, R. Costa, and R. Jardim-Goncalves, “Twitter mining for traffic events detection,” in *Proc. 2015 Science and Information Conf. (SAI)*, London, 2015, pp. 371−378.

[35] P. Tejaswin, R. Kumar, and S. Gupta, “Tweeting traffic: Analyzing twitter for generating real-time city traffic insights and predictions,” in *Proc. 2nd IKDD Conf. Data Sciences*, Bangalore, India, 2015.

[36] E. Morgul, H. Yang, A. Kurkcu, K. Ozbay, B. Bartin, C. Kamga, and R. Salloum, “Virtual sensors: Web-based real-time data collection methodology for transportation operation performance analysis,” *Transport. Res. Record: J. Transport. Res. Board*, vol. 2442, pp. 106−116, 2014.

[37] J. R. He, W. Shen, P. Divakaruni, L. Wynter, and R. Lawrence, “Improving traffic prediction with tweet semantics,” in *Proc. 23rd Int. Joint Conf. Artificial Intelligence*, Beijing, China, 2013, pp. 1387−1393.

[38] M. Ni, Q. He, and J. Gao, “Using social media to predict traffic flow under special event conditions,” in *Proc. 93rd Annual Meeting of Transportation Research Board*, Washington DC, USA, 2014.

[39] S. Grosenick, “Real-time traffic prediction improvement through semantic mining of social networks,” M.S. thesis, University of Washington, Washington DC, 2012.

[40] A. F. Abidin, M. Kolberg, and A. Hussain, “Improved traffic prediction accuracy in public transport using trusted information in social networks,” in *Proc. 7th York Doctoral Symp. Computer Science & Electronics*, York, 2014.

[41] A. Abidin, M. Kolberg, and A. Hussain, “Integrating twitter traffic information with kalman filter models for public transportation vehicle arrival time prediction,” in *Big-Data Analytics and Cloud Computing*, M. Trovati, R. Hill, A. Anjum, S. Y. Zhu, and L. Liu, Eds. Switzerland: Springer, 2015, pp. 67−82.

[42] D. Semwal, S. Patil, S. Galhotra, A. Arora, and N. Unny, “Star: Real-time spatio-temporal analysis and prediction of traffic insights using social media,” in *Proc. 2nd IKDD Conf. Data Sciences*, Bangalore, India, 2015.

[43] T. Zhou, L. X. Gao, and D. H. Ni, “Road traffic prediction by incorporating online information,” in *Proc. 23rd Int. Conf. World Wide Web*, Seoul, Korea, 2014, pp. 1235−1240.

[44] F. C. Pereira, F. Rodrigues, E. Polisciuc, and M. Ben-Akiva, “Why so many people? Explaining nonhabitual transport overcrowding with internet data,” *IEEE Trans. Intell. Transport. Syst.*, vol. 16, no. 3, pp. 1370−1379, Jun. 2015.

Yisheng Lv received the B.E. and M.E. degrees in transportation engineering from Harbin Institute of Technology, in 2005 and 2007, respectively, and the Ph.D. degree in control theory and control engineering from the Chinese Academy of Sciences, Beijing, China, in 2010.

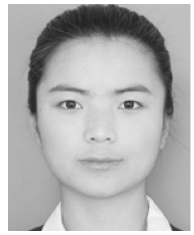
He is currently an associate professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include traffic data analysis, dynamic traffic modeling, and parallel traffic management and control systems.

Yuanyuan Chen received the B.E. degree in automation from Tongji University, Shanghai, China, in 2013. He is currently pursuing the Ph.D. degree in control theory and control engineering at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include intelligent transportation systems, machine learning and its application.

Xiqiao Zhang received the B.E. degree in traffic engineering from Harbin Institute of Technology, Harbin, China, in 2001. He received the M.E. degree in transportation planning and management from Harbin Institute of Technology in 2004. He obtained the Ph.D. degree in bridge and tunnel engineering from Harbin Institute of Technology in 2008. He is currently an assistant professor with the Department of Transportation Engineering, School of Transportation Science and Technology, Harbin Institute of Technology.

His research interests include traffic data analysis, transportation planning, emergency traffic modeling and management. Corresponding author of this paper.



Yanjie Duan received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2012. She is currently pursuing the Ph.D. degree in control theory and control engineering at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Her research interests include intelligent transportation systems, machine learning and its application.

Naiqiang Li is currently the vice-director of Institute of Smart City, Qingdao Academy of Intelligent Industries. He is qualified as the Class 1 Registered Architect in China.

He has more than ten years of working experience in smart transportation and smart city. His research interests include intelligent transportation systems and smart city.