The research trend analysis on Intelligent Transportation Systems

By: Liuhui Zhao

Update: December 15, 2020

1 Introduction

By integrating advanced information and communication technologies in transportation systems, including infrastructure and vehicles, Intelligent Transportation Systems (ITS) technologies have emerged as an efficient approach to enhancing transportation safety and mobility during the past two decades [1–3]. It attracts worldwide interest in recent years, from transportation professionals, manufacturers, policy makers, as well as researchers. Evolving from mainly solving traffic congestion and safety issues, through advanced traffic management and traveller information systems such as ramp metering control, dynamic message systems, nowadays, ITS is emphasizing more on addressing mobility issues through innovative practices and technologies, such as connected and automated vehicles, data driven traffic management, integrated mobility [4]. Such trend is also revealed from Google Trends [1]: as shown in Figure 1, during past 10 years, the interest in the term ITS is slightly decreasing, whereas the interest in shared mobility and connected vehicle is steadily increasing.

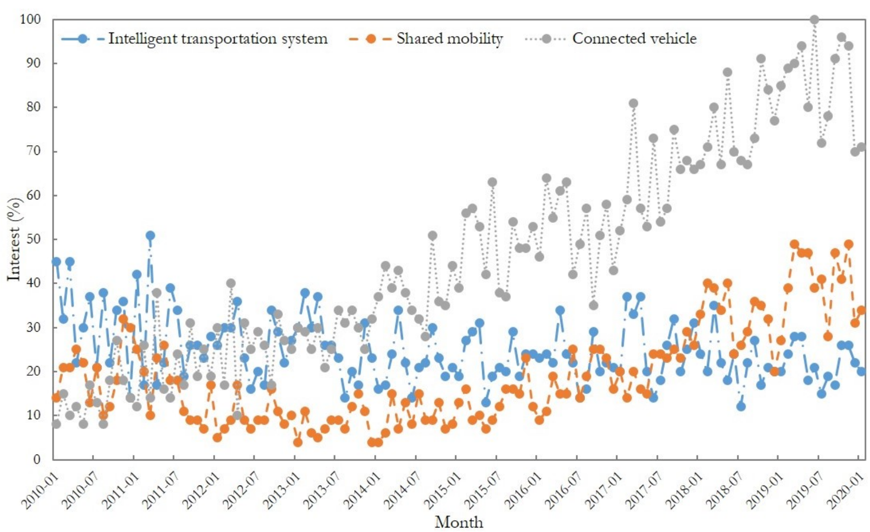

Figure 1: Trends of worldwide interest on searched items (figure adapted from Google Trends).

The nature of ITS, that is depending upon the development from many different disciplines, such as electronics, control, communication, etc., leads to increased problem complexity and requires knowledge transfer and cooperation among various communities [1]. However, it seems that for these emerging research topics, such as ITS, Connected and Automated Vehicles (CAV), or shared-use mobility, as different communities start from different angles of the problems, the majority research from one community tends to be blind from another community (e.g., difference conference presence, separate cross reference preference, etc.). For example, transportation-orientated research would be more on application and large transportation network side, mechanical engineering-orientated research maybe more on theoretical and small proof-of-concept case study side. Separate investigation may make sense for a while due

[1] Google Trends, https://trends.google.com/trends/explore?date=2010-01-01%202020-01-01q=Intelligent%20transportation%20system,shared%20mobility,connected%20vehicle.

to different approaches towards the problem. However, eventually, the efforts from different perspectives need to be combined to form a complete picture especially for the development and implementation of emerging technologies- Therefore, it would be of interest to examine the research work on the same broad topic (e.g., ITS or CAV) among different communities, and check the emphases of different research groups along the years. As a key indicator of research trends and focuses, scientific publications are of importance in order to analyze these potential similarities and differences among different research communities.

1.1 Related Work

Topic modeling, which is used for discovering hidden thematic structures that occur in a collection of documents, has gain more attention in the research area. Different algorithms have been developed since 90s, such as Latent Semantic Indexing (LSI) in [5], Probabilistic Latent Semantic Analysis (pLSA) in [6], Latent Dirichlet Allocation (LDA) in [7], and different variants of LDA as reviewed in [8]. Topic modeling has a wide range of applications in many fields, however, not until recently, few work has been done to reveal the trends and themes in transportation research area. In 2016, Das et al. applied LDA modeling technique to analyze the papers included in the compendiums of Transportation Research Board Annual Meetings, which is the most visible conference series in the transportation research area. The author affiliations, conference technical committees, paper title and abstracts were analyzed in their study to discover popular keywords, top contributors and topic trends over the years [9]. A similar study was conducted on the articles published in 22 leading transportation journals from 1990 to 2015 [10], investigating the similarities of journals and regions in terms of aggregated topic distributions in the broad transportation area. Another study [11] constructed the co-authorship network with the same data set to examine the structure of collaboration in transportation research. However, the evolution in a theme or topic of particular interest, as well as the acknowledgement and collaboration of other research communities over a particular topic within the transportation area, have not been studied in the past research body.

1.2 Problem Statement

In light of investigation needs, in this project, we would like to examine not only the temporal changes within the area of ITS, but also the research emphases and differences among different communities. Specifically, we would like to explore: 1) whether there are distinctions on the research on the same area across different disciplines, and if so, 2) is there any trend indicating more collaborations. Three journals are selected for the analysis of the project, which are generally sharing similar scopes but considered as the representatives of three different communities in the area of ITS, but a broader list of journals and/or conference proceedings can be added to the analysis within the same framework.

2 Methodology

As described in the above section, to test the hypothesis that there may be a distinction among different research groups, we will select three major journals having a specific focusing on the ITS, and collect the papers between 2010 and 2020 for analysis. In this section, we explain the data source we use to collect the data and the methodology applied to conduct the analysis in the project.

Based on the aims and scope of the three journals to be examined, including Journal of Intelligent Transportation Systems (J-ITS), Transportation Research Part C: Emerging Technologies (TR-C), and IEEE Transactions on Intelligent Transportation Systems (T-ITS), they are all devoted to the research on the ITS as well as the emerging technologies in transportation area. They are oriented from different regions of the World under different publishers, and have slightly different emphasis on the research of ITS.

Among the three journals, J-ITS focuses on the "research that leads to improved planning and operation of the transportation system through the application of new technologies" [2], and have an emphasis on social and environmental impacts of ITS. TR-C invites research that "addresses development, applications, and implications in the field of transportation of emerging technologies" [3] and have a particular interest on the impact on transportation system performance. On the other hand, T-ITS limits its scope within "the theoretical, experimental, and operational aspects of electrical and electronics engineering and information technologies as applied to ITS" [4].

[2] Journal of Intelligent Transportation Systems: https://www.tandfonline.com/action/journalInformation?journalCode=gits20

[3] Transportation Research Part C: Emerging Technologies, https://www.journals.elsevier.com/transportation-research-part-c-emerging-technologies

[4] IEEE Transactions on Intelligent Transportation Systems, https://ieeexplore.ieee.org/xpl/aboutJournal.jsp?punumber=6979

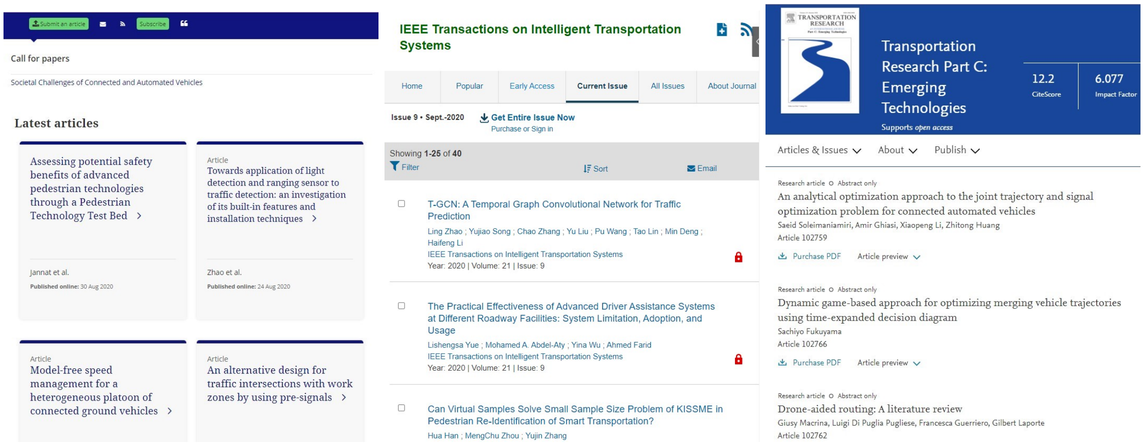

Figure 2: An overview of three major journals on ITS, left: J-ITS, middle: T-ITS, right: TR-C.

To line up with our scope, we collect the information of the published papers on the three journals, including the title, the authors' information, and the abstract. Since the abstract provides an overview of the research body of the paper, we use the abstract to analyze the research focus for each paper.

Primarily, we lay out two questions we try to answer in this project. To answer these questions, we need to analyze the data from different angles: for the first question on the differences/similarities in the journals, topic modeling technique and network analysis will be applied to identify the themes and trends in research emphasis among different communities, and for the second question on the trend of collaborations, network analysis would be applied trying to discover the communities among the published authors. Various algorithms for detecting communities in a network have been developed and employed – Louvain algorithm [12], Girvan-Newman algorithm [13], and clique percolation method [14], to name a few.

In the realm of topic modeling, LDA is one of the most popular methods to date. In this study, we will apply this algorithm to model topics in the set of collected documents. To infer the topics in a set of documents, the assumed generative process of LDA is described as follows [7]. Denote $w_i, i = \{1, 2, ..., V\}$ as an indicator of a word, belonging to a vocabulary, where V is total number of words in the vocabulary. Denote W_i as a document that consists of a sequence of words, $W_i = \{w_1, w_2, ..., w_{N_i}\}$, where N_i is the length of document i in terms of words. Denote $D = \{W_1, W_2, ..., W_M\}$ as a collection of the documents, where M is the number of documents. Denote $Z = \{z_1, z_2, ..., z_K\}$ as the set of topics in all documents, where K is the total number of topics. Then, the generative process for a set of document D can be described below. Finally, different inference algorithms can be applied to allocate topics across the documents.

- For each $i = \{1, 2, ..., M\}$, choose the distribution of a topic mixture in document i, $\theta_i \sim \mathrm{Dirichlet}(\alpha)$, where $\mathrm{Dirichlet}(\alpha)$ is a Dirichlet distribution with parameter α;
  - For each $k = \{1, 2, ..., K\}$, choose the distribution of words in topic k, $\phi_k \sim \mathrm{Dirichlet}(\beta)$, where $\mathrm{Dirichlet}(\beta)$ is a Dirichlet distribution with parameter β;
    * For each of the word in the sequence $w_n^i, n = \{1, 2, ..., N_i\}$, choose a topic $z_n^i \sim \mathrm{Multinomial}(\theta_i)$, and a word $w_n^i \sim \mathrm{Multinomial}(\phi_{z_n^i})$.

The flow chart of the study is represented as follows (Figure 3): we start with data collection from the selected list of journals, and analyze the documents through topic modeling and community analysis of authors, finally we make comparisons among the journals.

3 Case Study

3.1 Basic Information

The number of total published papers during the study period is different among the three journals: about 300 in J-ITS, but more than 2000 in TR-C or T-ITS. First of all, we extract the basic information for each of the journals, and then conduct the comparison. Overall, we observe an increase in the number of published papers over the year, and that the topic of the most cited papers in this time period is tightly related to the information and communication technologies:

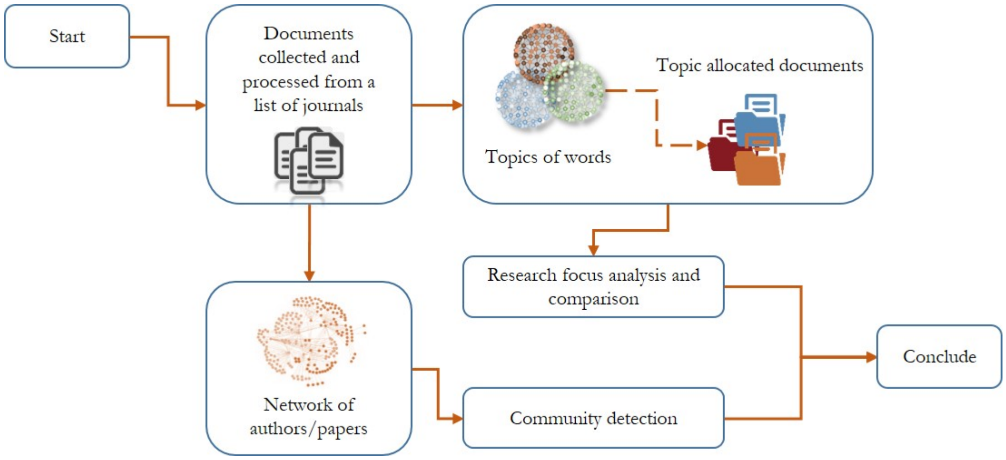

Figure 3: Flow chart.

deep learning, and vehicle connectivity technology. Individual statistics for each journal are presented in the following sections.

(a) T-ITS: In total, we identified 2706 regular papers published between 2010 and 2020 (up to date): papers of a survey/review, or erratum, or journal statistics reports are excluded (Figure 4). The most cited paper is about deep belief networks with multitask learning for traffic flow pattern, which is published in 2014.

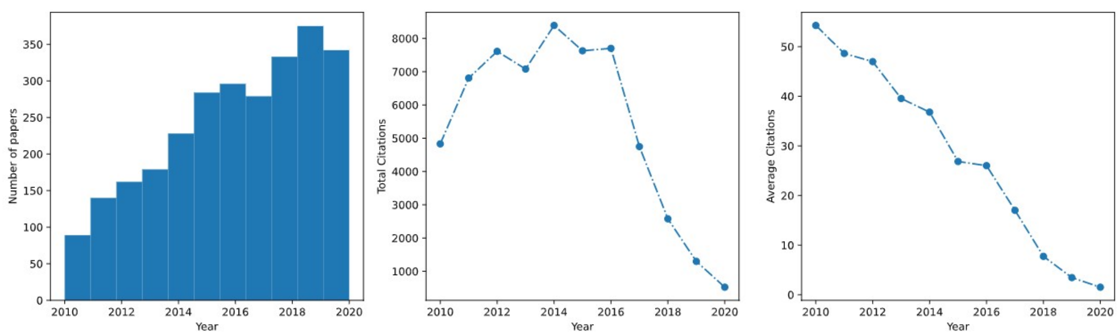

Figure 4: Number of published papers and citations per year (TITS).

(b) TR-C: In total, we identified 2029 regular papers published between 2010 and 2020 (up to date) with valid information(Figure 5). The most cited paper is on the topic of traffic speed prediction with a long short-term memory neural network, which is published in 2015.

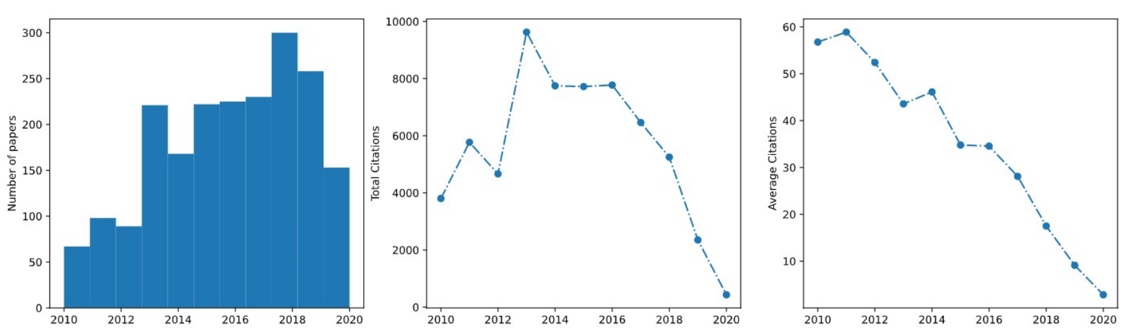

Figure 5: Number of published papers and citations per year (TRC).

(c) J-ITS: In total, we identified 54 issues published between 2010 and 2020 (up to date): a total of 306 regular papers with information (Figure 6). The most cited paper is about intersection management via vehicle connectivity, which is published in 2016.

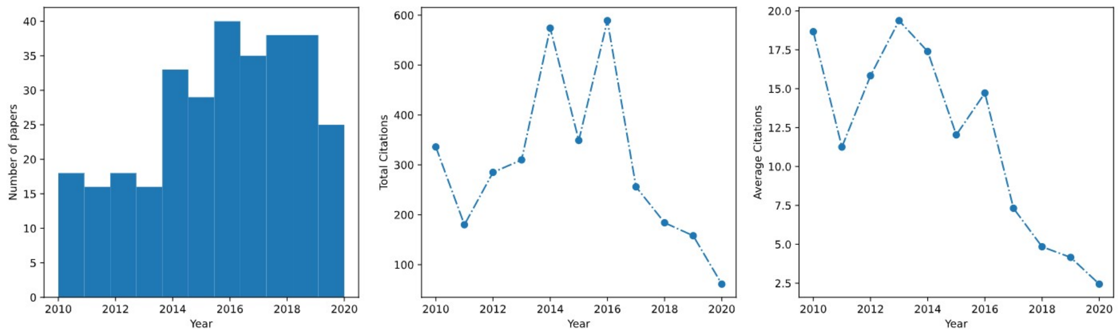

Figure 6: Number of published papers and citations per year (JITS).

To get an idea of the most concerned topics in this period, we first build a word cloud for the titles only for each journal. In the preprocessing step, we remove numbers and frequent words that do not convey meaningful topic information, including 'based', 'using', 'method', 'model', 'algorithm', 'new', 'transportation', 'traffic', 'vehicle', 'network'. From Figure 7, we can tell the similarities and differences among these three journals: 1) most of the titles in J-ITS are on the travel time estimation and traffic network information, and also cover topics across freeways and urban intersections; 2) TR-C, on the other hand, is more on the network-wide optimization and control, also touches a lot on transportation system modeling and design; 3) finally, we see that T-ITS emphasizes a lot more on vehicular or traffic network control, vehicle detection, and speed prediction. We see that the papers published during this 10-year period seem to match the scope of each journal quite well, based only on the titles.

Figure 7: Wordcloud on titles, left: JITS, center: TRC, right: TITS.

3.2 Abstract Analysis

To discover topics in the journals, we combine all the abstracts into a corpus, and apply a LDA model from *gensim* package. With the same stop words, we also build phrases within which words are often appear together, such as 'car-follow', 'deep-learning', 'lane-changing', 'short-term', 'penetration-rate', etc. Through tuning the parameters, we identify 12 topics which yields a topic coherence score of 0.50. For each topic, we select the first 30 words with the highest probabilities and draw the word cloud as in Figure 8. From the frequent words in each topic, we could make sense of the top 12 topics across the journals. For instance, most of the words in Topic #1 is related to network optimization and electric vehicles. Whereas in topic #5, we see that it is mostly related driving behavior and human factors.

3.2.1 Topic distributions by journal

After identifying top topics in the document, we try to analyze whether there is significant difference of topic distributions among the three journals. To compare, we group the corpus by journals and then average the topic distributions for each abstract that belongs to each journal. As shown in Figure 9, we see a very similar distribution among these journals, except that J-ITS has slightly different emphasis on Topic #8 - travel time estimation and route choice. Otherwise, all of the three journals has a significant amount of papers on Topics #3 - control and simulation, #2 - traffic

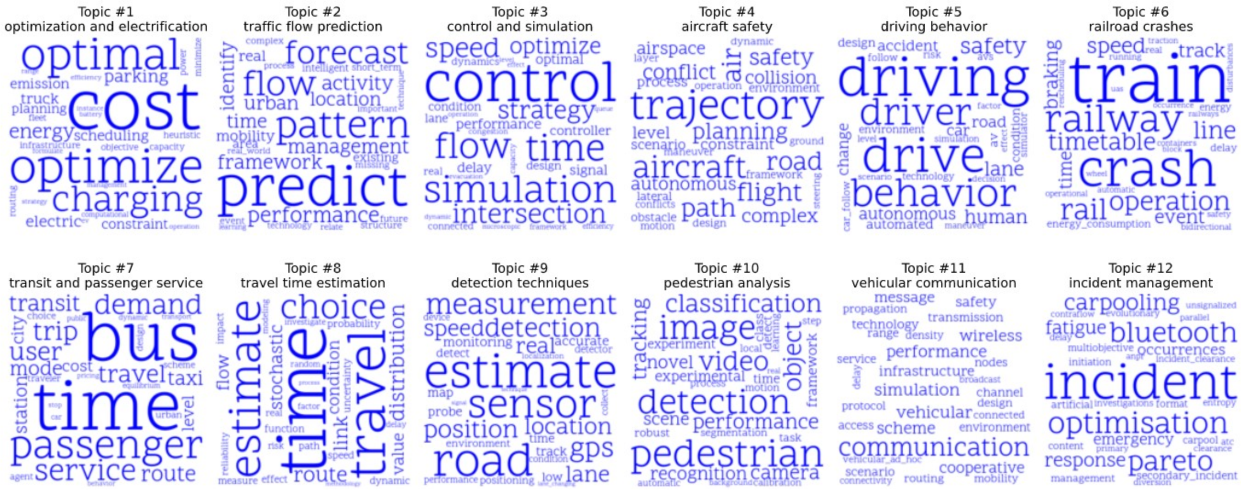

Figure 8: Word distribution in the discovered topics.

flow prediction and #9 - traffic detection techniques, T-ITS and TR-C have more papers on Topic #5 - driving behavior at the presence of automated vehicles and Topic #10 - computer vision based pedestrian analysis.

It seems that the majority of transportation studies in the area of ITS still relies on simulation-based evaluation of the applications of ITS on traffic prediction, control and management. From the topic distributions, we can see that compared to J-ITS, T-ITS and TR-C are more open to the evaluations of emerging techniques, such as the interactions of human and automated vehicles, the applications of machine learning techniques on transportation analysis, etc. Whereas J-ITS, emphasizing on travel time estimation and route choice, leans towards the traditional ITS realm.

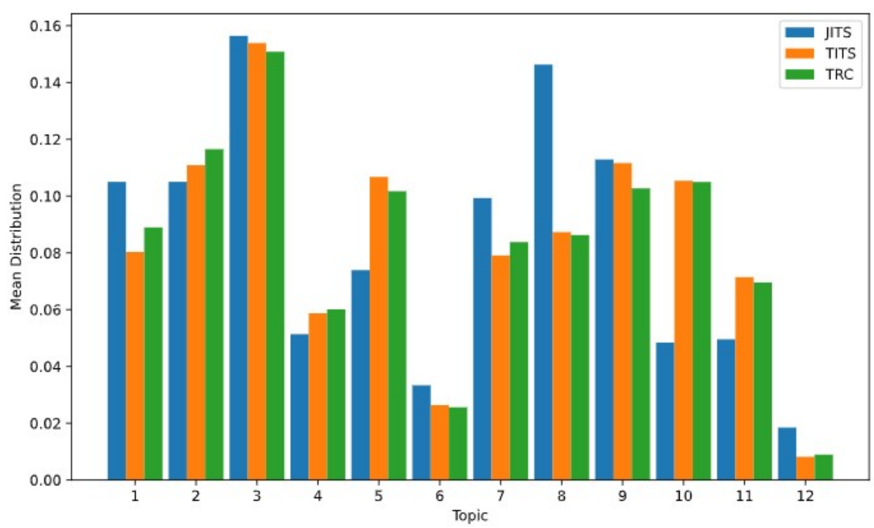

Figure 9: Topic distributions among journals.

3.2.2 Topic distributions per year

Figure 10 represents topic distributions between 2010 and 2020, mixing journals together. Generally, we can separate topics into different groups: topic group [4,6,12], including flight, train, and incidents, with constantly low presence in the across the years (left panel in Figure 10), topic group [1,5,7,11], cover optimization, bus transit, driving behavior, and vehicular communication , with constantly moderate probabilities (middle panel). In the right panel of Figure 10, we see that the probabilities of Topic #8 and Topic #9 are decreasing, whereas Topic #2 and Topic #10 are increasing, which may indicate a slow transition from traditional traffic estimation techniques to machine learning based traffic analysis.

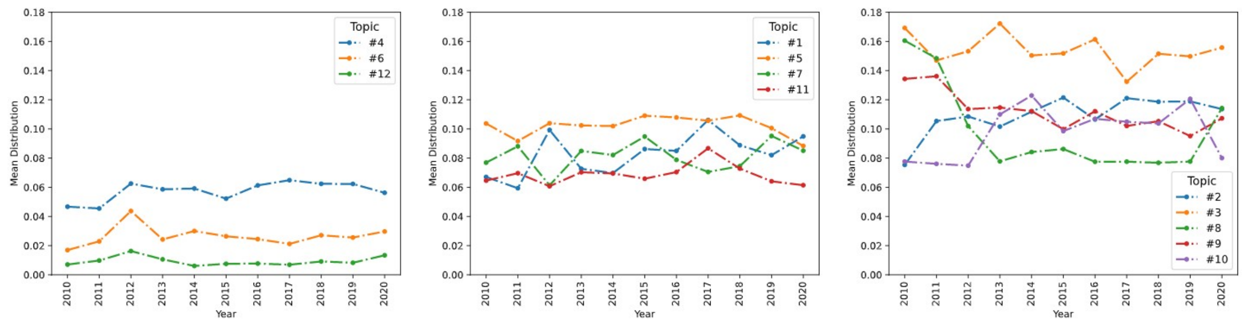

Figure 10: Topic distributions between 2010 and 2020.

3.2.3 Topic distributions per year per journal

Separate the corpus by journal and by year, we see quite different topic distributions over the year in different journals. In Figure 11, the upper panel is for J-ITS, the middle for T-ITS, and the bottom TR-C. In J-ITS, we see that previous popular Topic #8 [travel time estimation] drops significantly in recent years, while Topic #2 [traffic flow prediction], previously overlooked in J-ITS, gains attention steadily. An interesting observation is that the trends of Topic #1 [network optimization and electric vehicles] and Topic #8 seem to be negatively correlated to Topic #7 [transit service] in J-ITS: an increase in Topics #1 and #8 along with a reduction in Topic #7 over the years, and vice versa.

The distributions of different topics are quite similar to each other within the journal of T-ITS, with Topic #3 [control and simulation] featuring the most popular topic. Over the year, we see that topic group [1,2,7] (electric vehicles, traffic flow, transit) is gaining popularity while the presence of topic group [5,9] (driving behavior, vehicle detection) is gradually decreasing - which signals a transition of from seeking vehicle-emphasized solutions (i.e., sensors, drivers, vehicles) for transportation problems to network-emphasized solutions (i.e., use of renewable energy, evaluation of other travel modes).

Interestingly, we see a reverse trend in TR-C as compared to T-ITS. The presence of topic group [1,2,7] is slightly decreasing in recent years, whereas topic group [5,9] is increasing. Similar to J-ITS, a big drop in Topic #8 before year 2013 is observed.

Overall, we see that in the entire 10-year period, these three journals are similar to each other in terms of topic distribution, except for a slight different emphasis in few topics. However, the trends in different topics over the year are clearly different among the journals. Under the umbrella of ITS, the research on control and simulation (i.e., basic techniques) takes a large percentage of published work in all of the three journals. Although each journal has slightly different emphases on different topics year by year, we find that private vehicles and transit service (including shared mobility) are generally more exposed to ITS research than other modes and goods service, such as train, flight, truck. This is reflected by the under-explored topics in the past decade: topics [4, 6, 11, 12], which includes safety, flight planning, rail operation, and vehicular communication.

3.3 Author analysis

Before building a network of authors, we conduct a brief analysis of top contributing authors for each journal and inspect the overlaps among these top contributors. As a preprocessing, we remove any symbols (e.g., "-", "()") appearing in the first name, combine author names for all the journals, and check author names of the same person by looking up the author id as available in TR-C.

Shown in Figure 12 are the top contributors among journals, we see that there is not much overlap among the top contributors, except for the author 'Papageorgiou, Markos' who has presence in both T-ITS and TR-C – researchers who are familiar with the community/journal tend to publish their work in the same arena (although we see that these journals are quite similar to each other in terms of topic distributions as above-mentioned).

In Table 1, we calculate the average number of authors per paper by grouping by paper and count the number of authors for each paper and then averaging them out, and average number of papers per author for each journal by grouping by author name and count the number of papers for each author. We find that the most common number of authors per paper across the journals is 3, whereas T-ITS has the highest number of authors per paper: about 4 authors per paper, which may suggest a tendency of collaborations. Also, an average of 1.6 papers per author in TR-C and 1.5 papers per author in T-ITS seem to suggest more tight communities than J-ITS.

Figure 11: Topic distributions per journal between 2010 and 2020, upper: J-ITS, center: T-ITS, lower: TR-C.

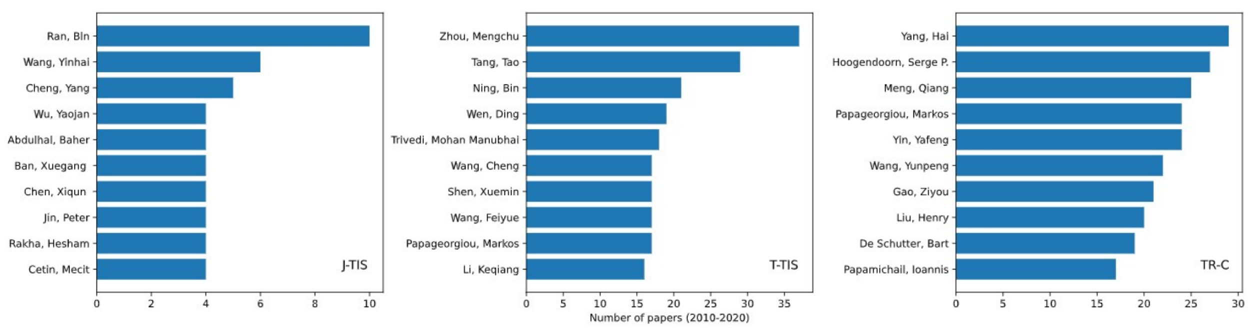

Figure 12: Top contributors per journal between 2010 and 2020, left: J-ITS, center: T-ITS, right: TR-C.

3.3.1 Statistics per region

In this section, we collect the affiliations of the authors per paper and summarize the statistics per country/region for the journals. The unique countries/regions per paper are used for the analysis purpose. Figure 13 tells that the most contributing countries/regions are China and United States in all of the three journals. We see that most of the published papers come from North America, European countries, East Asia, and Australia, where ITS could play an important role in tackling emerging transportation challenges. Also, note that the authors from United States are

Table 1: Statistics on authorship per journal.

Index	J-ITS	T-ITS	TR-C
Number of unique authors	870	7055	4367
Average authors per paper	3.43	3.89	3.46
Average papers per author	1.21	1.49	1.61
Maximum papers per author	10	37	29

significantly more than any other countries/regions in the journal of J-ITS, which may suggest that J-ITS has limited exposure to worldwide researchers.

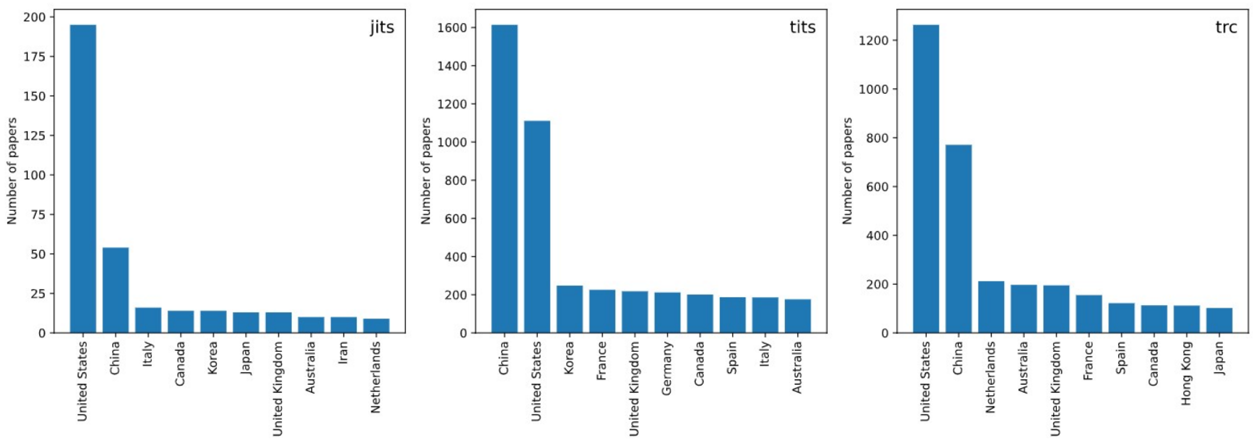

Figure 13: Top contributing countries/regions (2010 - 2020), left: J-ITS, middle: T-ITS, right: TR-C.

From the contributing countries for all the journals, we select top 10 and plot the percentage of each country in all published papers per year. Figure 14 indicates that each year, about 70% $\sim$ 80% papers in the three journals comes from these top 10 regions. Among them, China and United States take the largest proportion, and we see an increase of published papers from China in the past 10 years. Examine the percentage of each country in the top list, we find that, over the past decade, the diversity reduces when papers from the top 5 countries fills up the majority of the total publications in the three journals.

3.3.2 Collaborations

We use *NetworkX* package to analyze the collaborations among authors, and first draw a connection network among the top 50 authors with the highest number of publications in all of the three journals (as shown in Figure 15). Different colors in the figure represent different emphases on each of the three journals, while the link width stands for the number of papers. For instance, author 'Ran, Bin - 0' is colored red, meaning that this author has publications in all three journals, but the highest number of papers is with J-ITS. Observed from the figure, most collaborations among top authors happen with the same journal group - about 20% of the top authors only belongs to one journal, but we still see there are overlaps in the communities of TR-C and T-ITS (15% of the authors present in the group of TR-C and T-ITS). For J-ITS, however, the frequency of inter-journal collaborations is higher than the other two journals.

To visualize the trend of the collaborations over the year, we generate graphs with a sliding window of 3 years as shown in Figure 16. Clearly, we see an increase of community size over the past decade in the area of ITS, as well as an increased collaboration represented by a more dense core community in later years, which also matches well with the results in Google trends (i.e., the interests in different sub-topics of ITS is growing in recent years as shown in Figure 1). Additionally, with the entire author pool, we try to identify communities and draw the largest connected community in the network (Figure 17). To check whether the authors from different journals have connections to each other, we color-coded the top 10 contributors from each journal indicated in Figure 12. Shown in Figure 17, all of the top contributors of each journal shares the same core community, although some of the J-ITS authors are farther away from other researchers.

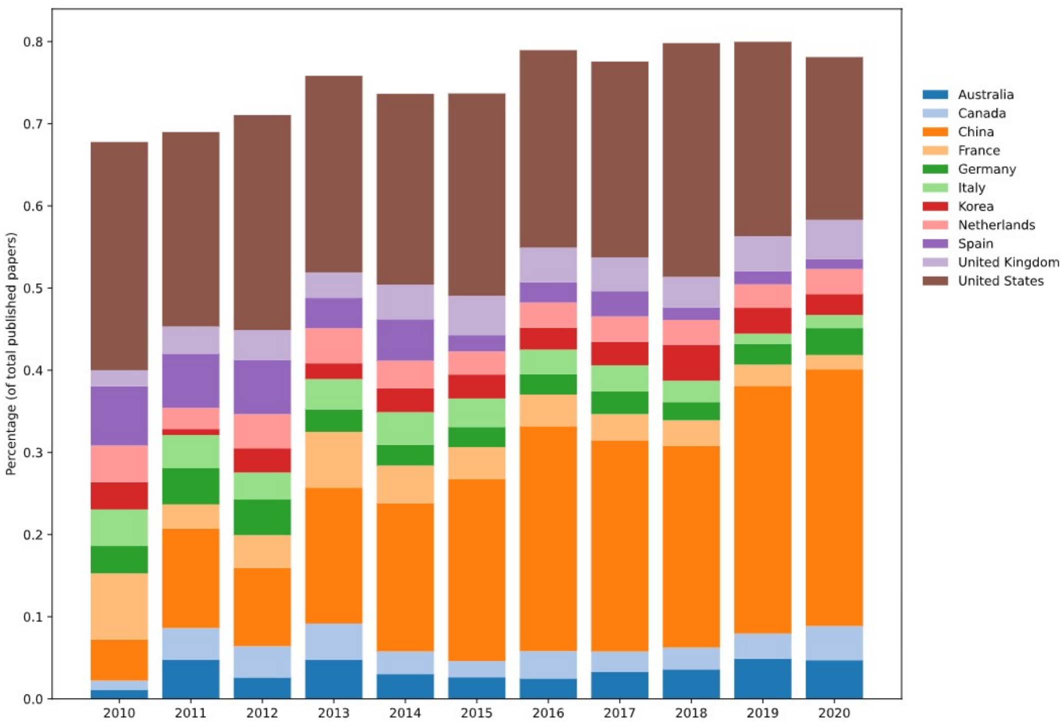

Figure 14: Top contributing countries over the years.

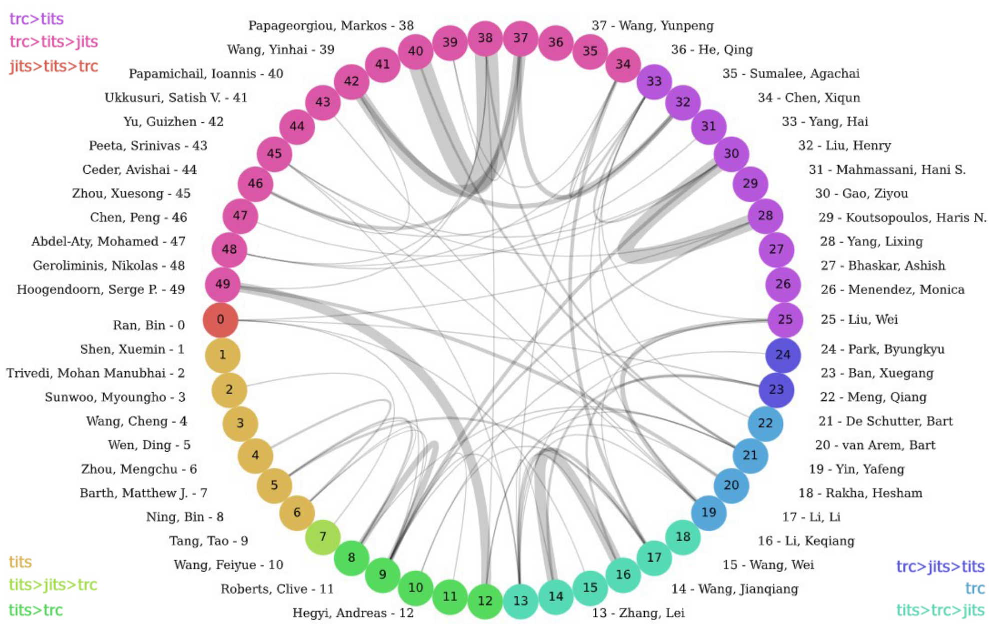

Figure 15: Collaborations among top 50 authors (2010 to 2020).

4 Conclusions

In this project, we explore the research trend of Intelligent Transportation Systems in the past decade by analyzing the abstracts and authorship in three major journals in this area: J-ITS, T-ITS, TR-C. The most studied direction under the umbrella of ITS is the control algorithms and simulation methodologies, followed by traffic flow prediction and vehicular detection techniques. Historically, we see that the topic distributions of the topi 12 identified topics among

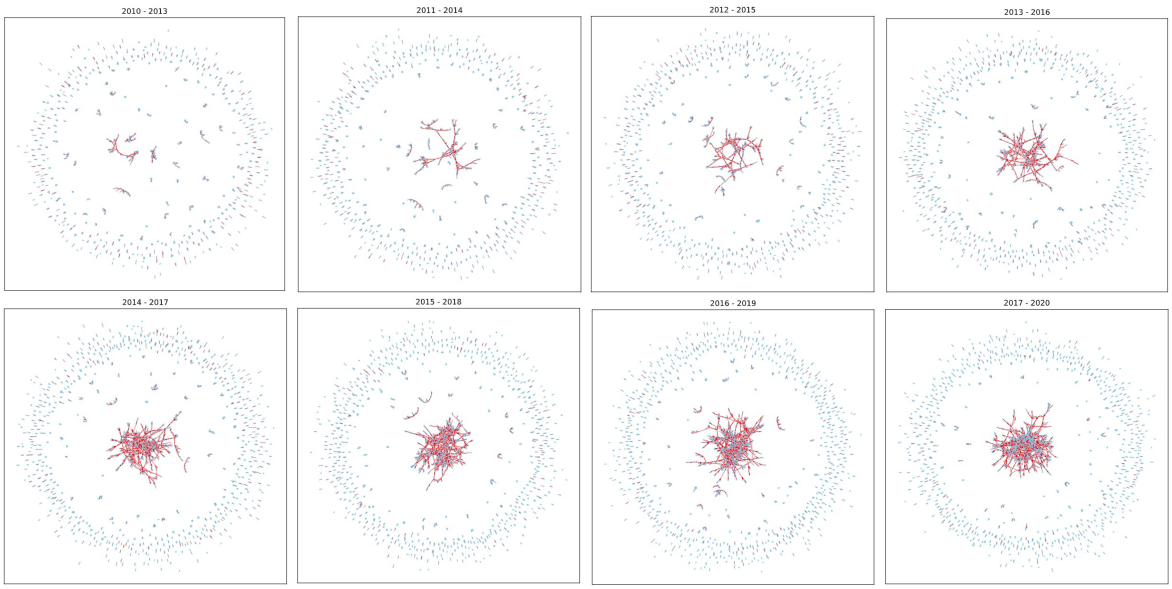

Figure 16: Trends in ITS communities (2010 to 2020).

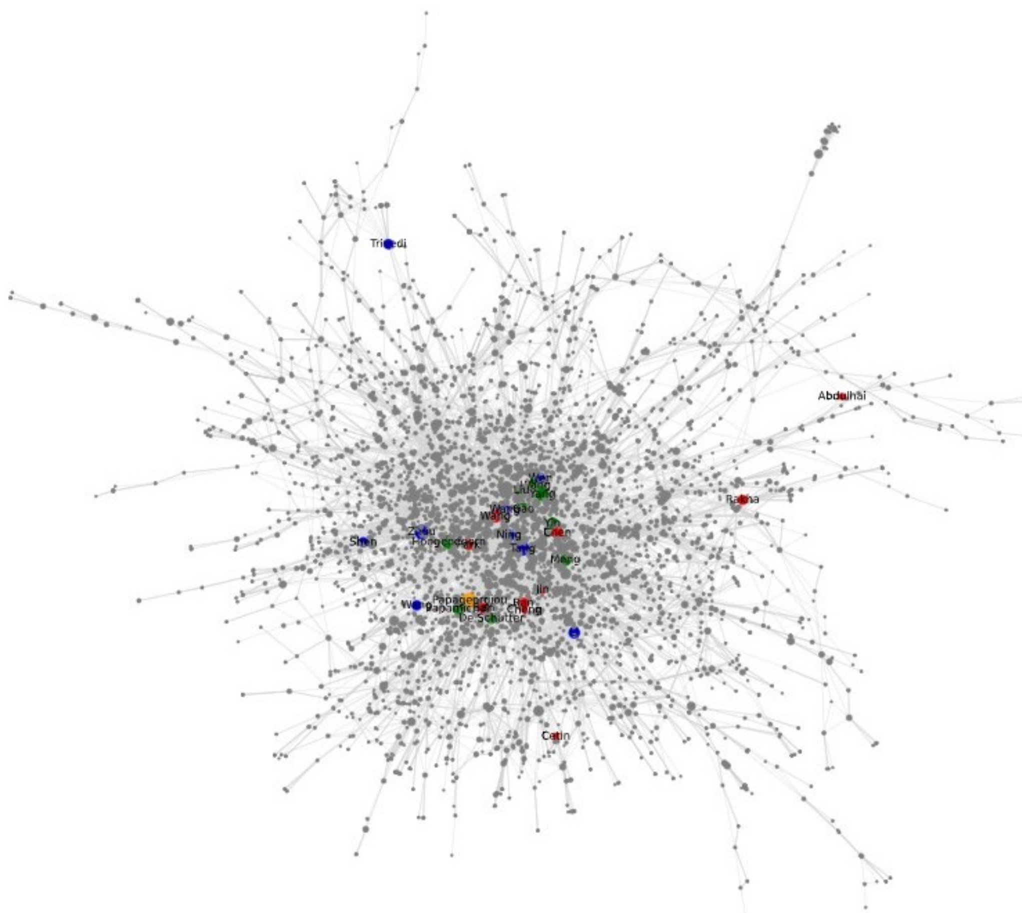

Figure 17: The largest connected community, top 10 contributors coded in colors (red: J-ITS , blue: T-ITS, green: TR-C, orange: J-ITS and T-ITS).

the three journals are quite similar to each other, indicating that these journal do share a lot in common in terms of ITS research. Over the year, the emphasis on the topics is shifting: work in electrification and transit service increases in T-ITS but decreases in TR-C, working in driving behavior and vehicular detection decreases in T-ITS but increases in TR-C. Overall, the coverage of different topics is similar to each other, except for few under-explored area such as freight transportation and incident management.

In general, the research interests in this area is growing in terms of not only number of papers, but also contributing authors and an increasing collaborations between researchers in the recent years. Researchers may have different focuses on where to publish their work, as shown by the almost-non-overlapping top contributors across the journals. Nevertheless, we do see that the researcher pools of T-ITS and TR-C have relatively high overlaps. Although J-ITS has a much lower number of studies, we still see that all of the top contributors from the three journals presents in the largest connected community, generated with the entire author pool.

Due to the scope of work, we are not able to conduct a thorough analysis on some other metrics, such as citations and references, which could potentially reveal important information in the analysis of publications and author network. Also, the time period could be longer to examine long-term changes in this specific topic. Finally, we only select three most related journals in the area of ITS, but an extended list of transportation related journals would present a complete picture of the research in this area - a filter to select those publications related to ITS is necessary in this case.

References

[1] Figueiredo, L., Jesus, I., Machado, J. T., Ferreira, J. R., and De Carvalho, J. M. Towards the development of intelligent transportation systems. In *Proceedings of 2001 IEEE Intelligent Transportation Systems*, pages 1206–1211. IEEE, 2001.

[2] Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., and Chen, C. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639, 2011.

[3] Dimitrakopoulos, G. and Demestichas, P. Intelligent transportation systems. *IEEE Vehicular Technology Magazine*, 5(1):77–84, 2010.

[4] Chan-Edmiston, S., Fischer, S., Sloan, S., and Wong, M. *Intelligent Transportation Systems (ITS) Joint Program Office: Strategic Plan 2020–2025*, 2020.

[5] Papadimitriou, C. H., Raghavan, P., Tamaki, H., and Vempala, S. Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, 61(2):217–235, 2000.

[6] Hofmann, T. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 50–57, 1999.

[7] Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3: 993–1022, 2003.

[8] Blei, D. M. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[9] Das, S., Sun, X., and Dutta, A. Text mining and topic modeling of compendiums of papers from transportation research board annual meetings. *Transportation Research Record*, 2552(1):48–56, 2016.

[10] Sun, L. and Yin, Y. Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C: Emerging Technologies*, 77:49–66, 2017.

[11] Sun, L. and Rahwan, I. Coauthorship network in transportation research. *Transportation Research Part A: Policy and Practice*, 100:135–151, 2017.

[12] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 2008.

[13] Girvan, M. and Newman, M. E. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[14] Palla, G., Derényi, I., Farkas, I., and Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.