



Data Dive into *Transportation Research Record* Articles

Authors, Coauthorships, and Research Trends

SUBASISH DAS

The author is an assistant research scientist at Texas A&M Transportation Institute in San Antonio, Texas.

Transportation is vital to people's quality of life. Rapid growth in population, miles traveled, urbanization, and emerging technologies like connected and autonomous vehicles have had a substantial impact on modern living. The rising application of emerging technologies, the increasing number of peer-reviewed journals and conference proceedings, and the significant growth in interdisciplinary collaborations reflect the significance of the size and scope of transportation research. However, transportation challenges and problems have changed over time, and the transportation research scope has also become more diverse. Because of the consistent evolution of the advances in solutions and technologies, transportation research has experienced an upsurge of publications in recent decades.

Predicting future salient issues in any field of science that will dominate research is always a challenge, but as transportation research becomes more complex and cross-cutting, this challenge is critical.

Research on statistical models of co-occurrence of trending topics has led to the growth of different, useful topic models. This efficient machine-learning technique helps researchers find concealed trends inside unstructured larger textual contents. To understand the research trends in the realm of complex transportation science and engineering, an analysis of journal articles from the *Transportation Research Record: Journal of the Transportation Research Board* (TRR) may be beneficial. By applying text mining and two topic modeling techniques, this article presents an empirical analysis of 28,987 articles published in the TRR from 1977 to 2018 to identify the publication trends.

History of the TRR

The Transportation Research Board (TRB) coordinates the most comprehensive and largest annual transportation conference in the world. Established in 1920 as the National Advisory Board on Highway Research, TRB has provided a platform to convert research results into applicable

Above: An analysis of journal articles from the Transportation Research Record: Journal of the Transportation Research Board (TRR) can shed light on today's transportation research trends.

Aims and Scope of the *Transportation Research Record*

All modes of passenger and freight transportation are addressed in papers covering a wide array of disciplines, including policy, planning, administration, economics and financing, operations, construction, design, maintenance, safety, and more. The ideal TRR contribution clearly advances knowledge in its area of expertise; explicitly draws out the practical implications of the paper's results, whether for immediate or longer-term application; and is well written with an engaging style.

information about every facet of transportation. Thousands of scientists, engineers, and other transportation practitioners and researchers from the private and public sectors and academia all are included in TRB's various activities.

The year 2020 marked the 57th year of TRB's peer-reviewed journal, which debuted as the *Highway Research Record* in 1963, became the *Transportation Research Record* in 1974, and added the subtitle *Journal of the Transportation Research Board*



Photo: Risdon Photography

Emiliano Ruiz, University of Texas at El Paso, presents research at the 2020 TRB Annual Meeting. His paper, "Simulation Model to Support Security Screening Checkpoint Operations in Airport Terminals," was published in the TRR the same year.

in 2000. The technical papers in TRR have been accepted for publication through a rigorous peer-review process overseen by TRB technical committees.¹ These papers provide extensive documentation of the research activities undertaken by the transportation research community, and they provide a unique insight into the research topics that have remained active over the long term, as well as topics that have recently emerged into the forefront. To remake TRR into a timelier, more author-friendly, and more widely marketed journal, TRB transitioned the publication of TRR to SAGE Publishing in 2018.

Data Collection

As a robust, long-standing journal, TRR was chosen as the subject of this analysis of trends in keywords, topics, authors, and coauthorships. The TRR series was selected based on its long history and inclusion of a wide variety of subject matter. The TRR series was also attractive because of its rigorous review process and widespread use among both academics and practitioners. The TRID database then was used to develop the databases for this study. All information associated with TRR articles was first saved in the research information system (RIS) format and then converted into spreadsheet format. The columns in the database include the title of the paper, keywords, abstract, authors, and publication year. This analysis included 28,987

¹ ISSN: 0361-1981, Online ISSN: 2169-4052

articles published between 1977 and 2018. Publication years of the articles were extracted from TRID metadata.

As shown in Figure 1, the pace of published articles has increased over the years. The publication rate trends from 1977 to 2018 can be divided into four stages. The first stage was from 1977 to 1990. Because of its multidisciplinary nature, TRR published around 530 journal articles per year from its earliest years until 1990. The second stage was from 1991 to 2000. During this period, the number of articles published grew steadily and consistently, increasing to an average of 663 articles per year. The third stage (2001–2010) shows an average of 769 journal articles per year, and the fourth stage (2011–2018) shows an average of 903 journal articles per year.

Prolific TRR Authors

Figure 2 lists the top eight most prolific authors of TRR articles from 2009 to 2018 and illustrates the frequency of yearly publication for each author. As shown in Figure 2, in 2016, Serge Hoogendoorn published 18 TRR articles—the largest number of TRR articles by a researcher as a

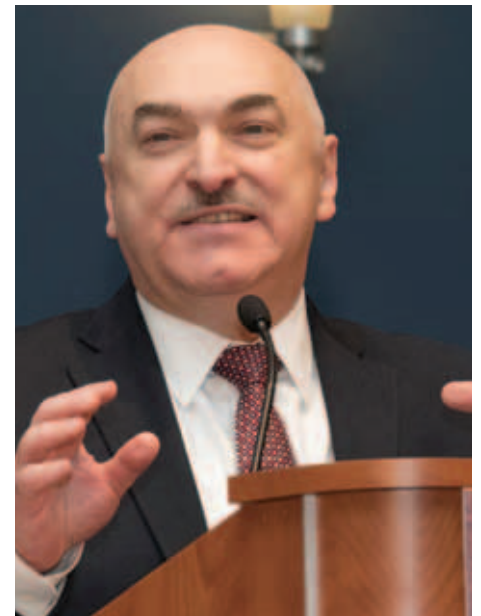


Photo: Risdon Photography

The most prolific TRR author of all time, Hani Mahmassani delivered the 2016 Thomas B. Deen Distinguished Lecture to a standing-room-only crowd.



FIGURE 1 TRR journal articles by year.

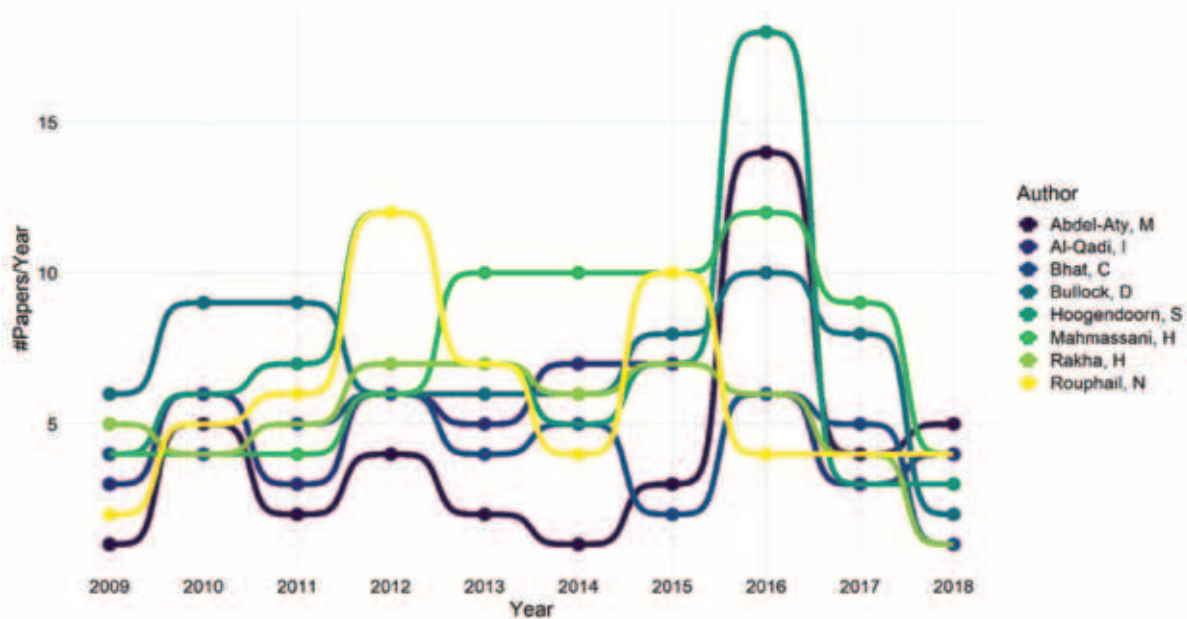


FIGURE 2 Prolific TRR authors (2009–2018).

corresponding author or coauthor in a single year. Hani Mahmassani has published 183 TRR articles since 1984, which makes him the most prolific TRR author overall.

Coauthor Network

Coauthorship networks can be used to investigate the structure of scientific collaborations. As transportation research has become increasingly cross-disciplinary, it is important to investigate author patterns and trends. This article explores a coauthor network plot for a quick understanding of the complex interdisciplinary networks among authors, but future studies can investigate the development of advanced analysis, such as author–topic models. Created using Gephi 0.9.2 software, the plot graph (Figure 3) shows the network patterns of the authors who are author or coauthor of at least one TRR article.² The complexity of this network indicates the massive number of nodes and links among TRR authors.

The node sizes are proportional to link-in counts and color by different nodes.

² <https://gephi.org/>

Interactive Tool

Developed as a part of this study, an interactive web tool is available via Github and allows users to explore the TRR coauthorship network interactively. Any TRR author can search their name to see their coauthorship network.

For more, visit http://subasish.github.io/pages/gephi_html/TRR_C/network/.

The authors Imad Al-Qadi and Darcy Bullock have the highest number of coauthor connections (127 and 123, respectively).

Clusters of Key Words by Publication Stages

Four interconnection network diagrams (Figure 4) were developed to explore the shifting research trends over the years. In the first stage (1977–1990), two main clusters formed around *pavement* and *travel models*. Two other keyword groups show higher co-occurrences: 1) *effectiveness* and *measure of effectiveness* and 2) *survey* and *data collection*. A similar trend is seen in the second stage (1991–2000);

however, in this stage more keywords are associated with the *pavement* cluster.

By contrast, by the third stage (2001–2010), the *asphalt pavement* cluster has divided into two subclusters: *asphalt pavement* and *mechanistic–empirical pavement*. Some smaller clusters of *driver behavior* and *safety* are visible in this stage. By the final stage (2011–2018), three main clusters have formed: *pavement*, *travel model*, and *highway safety*. This trend analysis clearly shows how the topical shift occurred in transportation research over the years.

Trends from Structural Topic Modeling

Because TRR journal articles cover a wide variety of topics, advanced modeling is needed to unearth research topic trends. Probabilistic topic modeling is a robust method to identify topics from complex text data. The two most common topic modeling methods are latent Dirichlet allocation (LDA) and structural topic modeling (STM). Both methods are Bayesian generative topic models that assume that each topic is a distribution over words and that each document is a mixture of corpus-specific topics (that is, a collection of texts; for example, all abstracts from the 2018 editions of TRR can be considered a corpus).

The STM algorithm identifies document-level structure information (keywords or word groups) to influence topical prevalence (for example, the proportion of topics by document frequency) and topic content (the distribution of the keywords in topics) (1–4). Several studies have applied topic modeling techniques to TRB-related journals, conference proceedings, and social media hashtags (5–7).

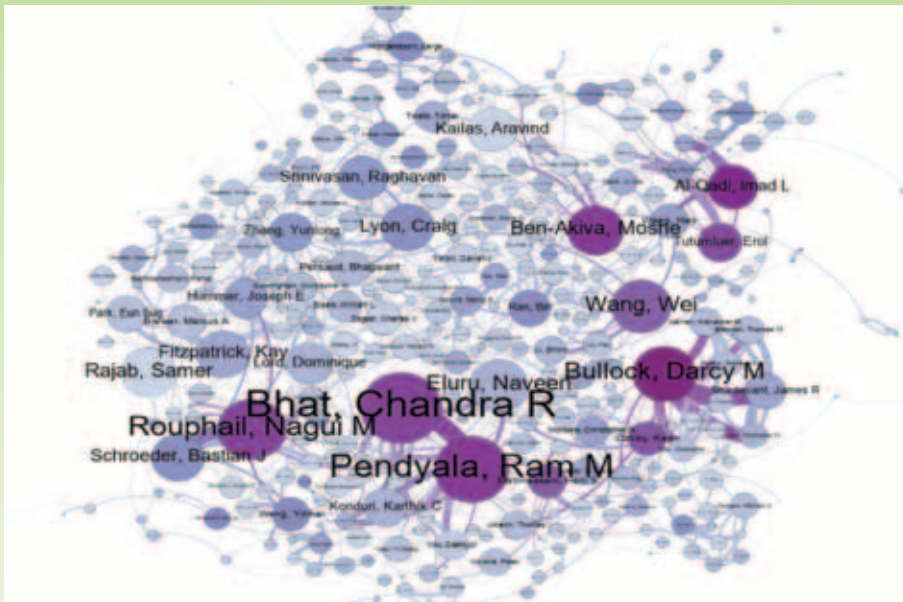
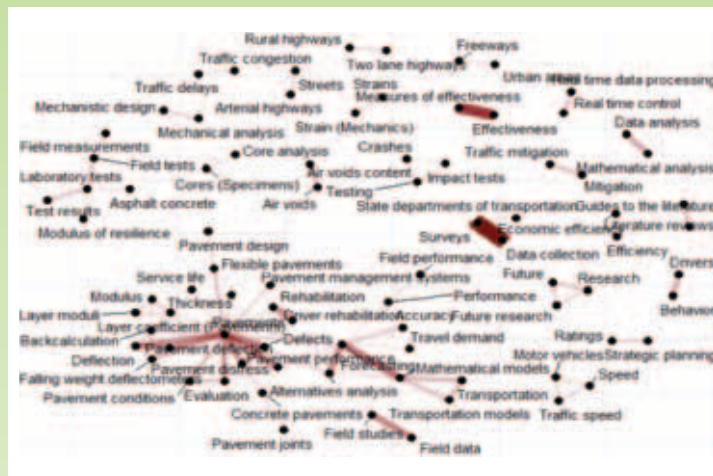


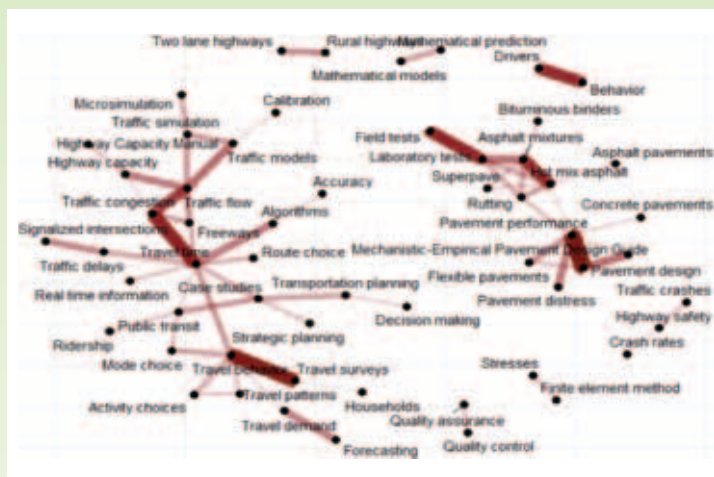
FIGURE 3 TRR coauthorship network.



1977-1990



1991-2000



2001-2010



2011-2018

FIGURE 4 Keyword networks by different year groups.

For this article, an open-source R package STM and topic model (TM) are used in the analysis (8–9). The data import process produces documents and metadata (i.e., data that provide information about other data) that is then incorporated by STM into the topic modeling framework. Figure 5 illustrates the corpus-level visualization of the top topics from a 20-topic model, showing the expected proportion of the corpus belonging to each topic. High-frequency topics include travel data

(Topic 2), choice model (Topic 7), rehabilitation (Topic 8), freeway (Topic 11), and urban planning (Topic 17).

Interactive Tools Using LDA

By using metadata, STM functions explain the trends over the years. This article examines large textual contents (papers have approximately 305,000 words, and abstracts have approximately six million words), so an interactive and comprehensive topic model is needed. The LDA

model was used primarily to visualize the output of TM fit; however, the high dimensionality of the fitted model produces challenges in creating these visualizations.

Normally, LDA is applied to thousands of documents representing topic combinations in the dozens to hundreds, which then are modeled as distributions across thousands of terms. To mitigate these challenges and create LDA visualizations (LDAvis), interactivity—a basic technique that is both compact and thorough—is

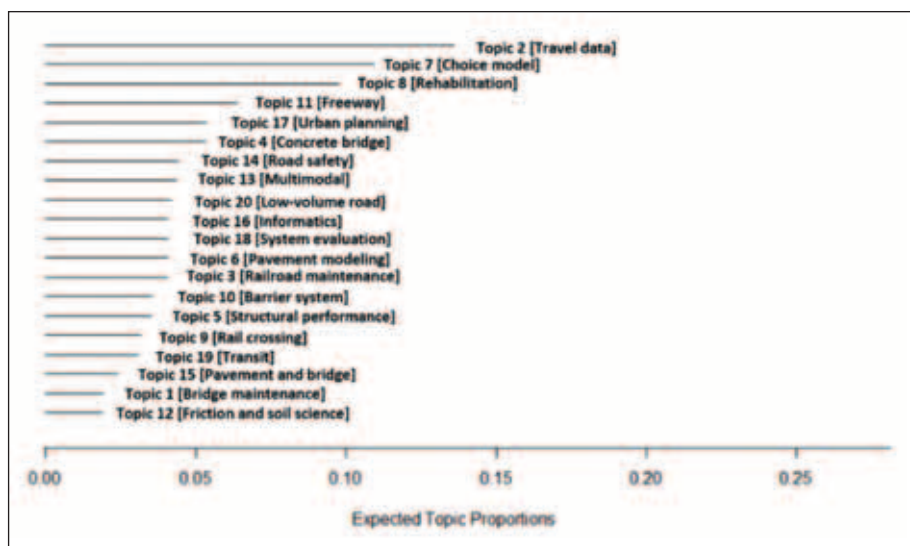


FIGURE 5 Top 20 topic models using STM.

the best technique. For this article, the LDAvis package was employed to develop interactive LDA models (10).

Figures 6a and 6b present the web interface of the interactive visualization of LDA topic models.³ The plots are composed of two sections: 1) on the left, global perspective on the topic models, and 2) on the right, a bar chart of the keywords associated with the highlighted topics.

The topics are plotted as circles in the left section. The locations of the topics are based on the measures of principal component analysis (PCA), a dimension reduction method to show relevance of topics based on the relative distances in a 2-D or 3-D space, in a way to show the general trends of the keywords by showing which topics are similar or different. By measuring the distance between topics, the centers of these topic circles are placed in the visualization. The bar plot represents the individual terms that are the most useful for interpreting the topics on the left, based on which topic is currently selected (keywords are shown horizontally). This allows users to comprehend the meaning of each topic. The overlaid bars in the plot represent both the corpus-wide

frequency of a given term and the topic-specific frequency of the term (11).

The left and right sections of the plot are interconnected. Selection of a topic circle in the left makes the bar plot on the right highlight the terms most useful to interpreting the selected topic. Additionally, selecting a term from the bar plot reveals the conditional distribution over topics for the selected term. This allows users to efficiently examine many topic-term relationships.

Conclusions

The TRR, as an international transportation journal, has provided a platform for the exchange of method, concepts, policies, and technologies to help transportation agencies to work toward a safe and secure transportation system. The findings of this study show that the pace of published articles has increased over the years. TRR is preferred by many prominent researchers, and the coauthor network shows some clusters around prominent and prolific researchers. The complex nature of the coauthor network plot indicates a wide variety of coauthorships among TRR authors.

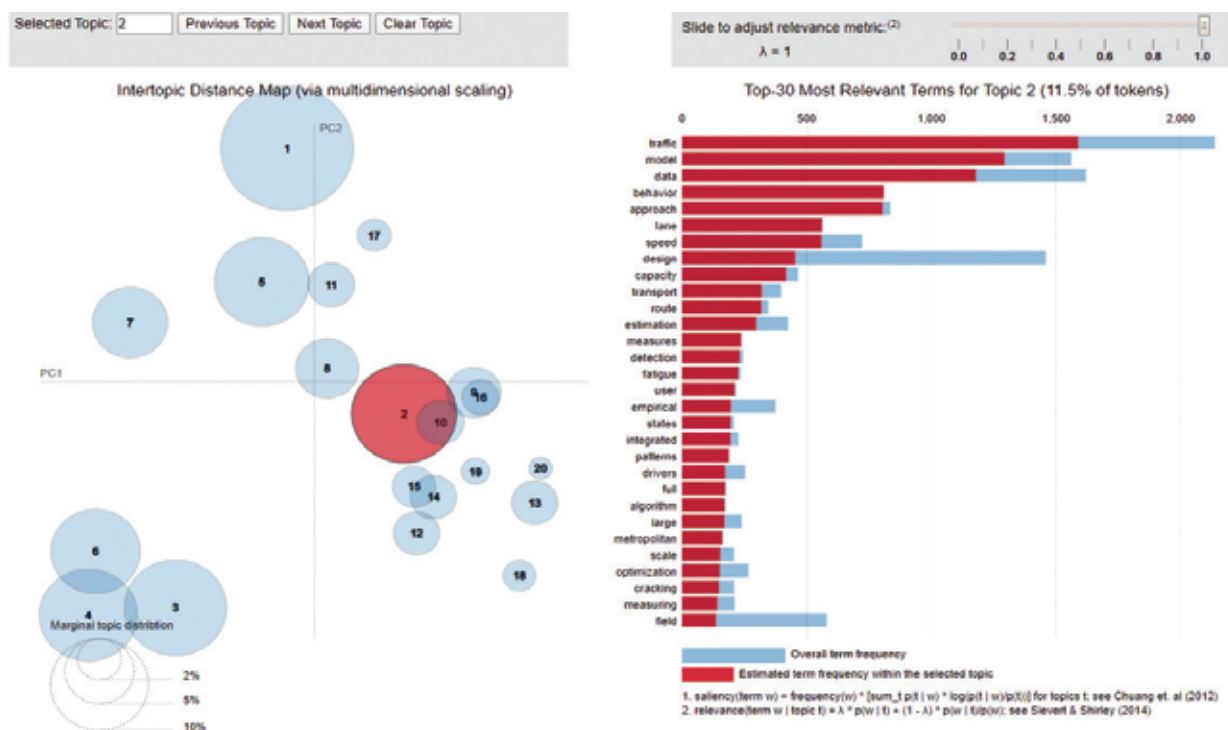
TRR has published articles covering a wide array of topics pertaining to transportation. A single topic model is not sufficient to explain the trends and patterns of the keywords over the years. To provide flexibility in understanding the topics, the

author developed two interactive LDA topic model tools to help users understand the range of topics and relevant keywords associated with the topics. It is anticipated that the TRR will continue to provide a platform for transportation science and engineering for academics, educators, researchers, enterprise leaders, and policy makers with the vision and mission to help the transportation system of the future.

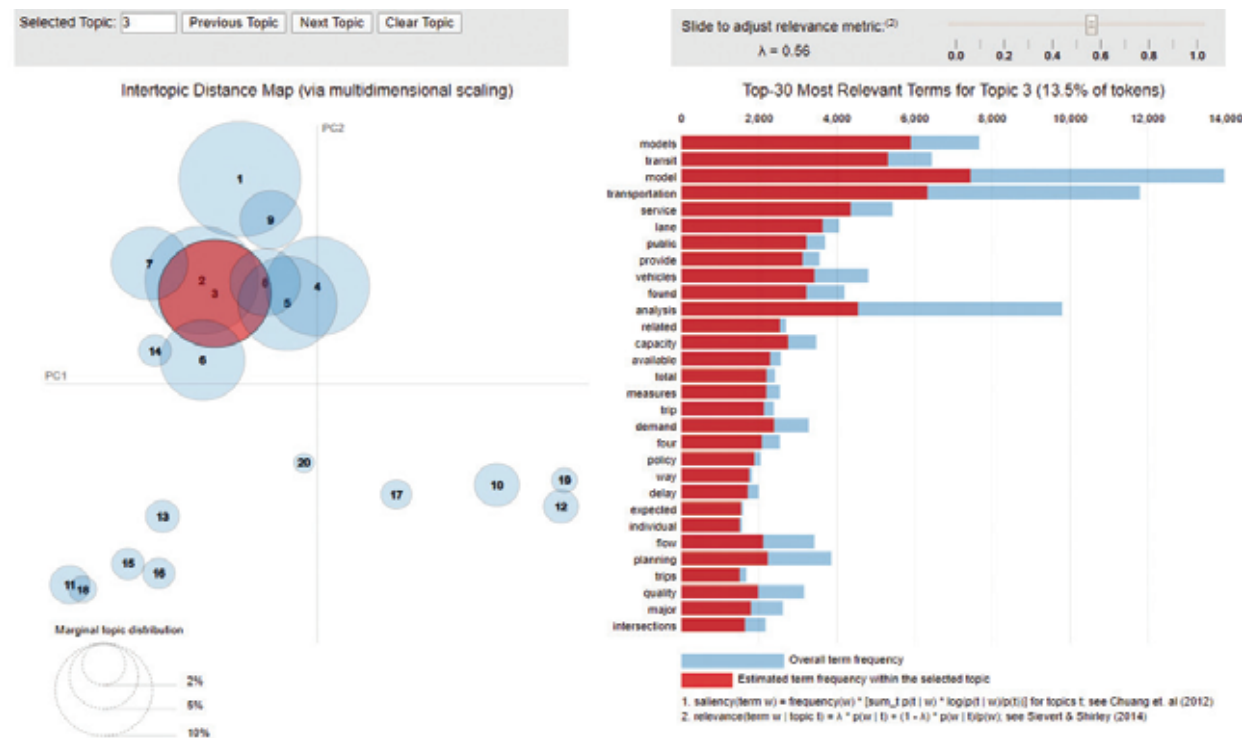
REFERENCES

1. Grimmer, J., and B. Stewart. Text as Data: The Promise and Pitfalls of Automatic Content Analysis. *Political Analysis*, Vol. 21, No. 3, 2013, pp. 267–297.
2. Chang, J., J. Boyd-Graber, S. Gerrish, C. Wang, and D. M. Blei. Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems*, Vol. 22, 2009, pp. 288–296.
3. Andrzejewski, D., and X. Zhu. Latent Dirichlet allocation with Topic-in-Set Knowledge. *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (E. Ringger, R. Haertel, and K. Tomanek, eds.), Association for Computational Linguistics, 2009.
4. Andrzejewski, D., X. Zhu, and M. Craven. Incorporating Domain Knowledge into Topic Modeling via Dirichlet Forest Priors. *ICML 09: Proceedings of the 26th Annual International Conference on Machine Learning*, Association for Computing Machinery, 2009, pp. 25–32.
5. Das, S., X. Sun, and A. Dutta. Text Mining and Topic Modeling of Compendiums of Papers from Transportation Research Board Annual Meetings. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2552, 2016, pp. 48–56.
6. Das, S., K. Dixon, X. Sun, A. Dutta, and M. Zupancich. Trends in Transportation Research: Exploring Content Analysis in Topics. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2614, 2017, pp. 27–38.
7. Das, S. #TRBAM: Social Media Interactions from the Largest Transportation Conference. *TR News*, No. 324, November–December 2019, pp. 18–23.
8. Roberts, M., B. Stewart, and D. Tingley. stm: R Package for Structural Topic Models. 2016. <http://www.structuraltopicmodel.com>. Accessed July 2016.
9. Feinerer, I., K. Hornik, and D. Meyer. Text Mining Infrastructure in R. *Journal of Statistical Software*, Vol. 25, No. 5, 2008, pp. 1–54.
10. Sievert, C., and K. Shirley. LDAvis: Interactive Visualization of Topic Models. R package version 0.3.2. <https://CRAN.R-project.org/package=LDAvis>. Accessed July 2019.
11. Das, S., A. Dutta, and M. Brewer. Case Study of Trend Mining in Transportation Research Record Articles. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2674, 2020, pp. 1–14.

³ To see the tool based on abstracts, visit http://subasish.github.io/pages/trr_abstract. To see the tool based on titles, visit http://subasish.github.io/pages/trr_title.



(a)



(b)

FIGURE 6 Interactive topic model tools using TRR journal titles (a) and abstracts (b).