Texas A&M Transportation Institute

Transportation Research Record Articles: A Case Study of Trend Mining

Subasish Das
Anandi Dutta
Marcus Brewer

Monday, January 13, 2020

Overview

Synopsis

Synopsis

- Transportation research is multi-faceted. Perception of trends and patterns from unstructured text data is overwhelming.
- This study used **natural language processing (NLP)** to identify trends and patterns.
- This study collected data from the titles and abstracts of the papers published in **Transportation Research Record: Journal of the Transportation Research Board, since 1978**.
- Used two NLP Tools: **Latent Dirichlet Allocation (LDA) and Structural Topic Modeling (STM).**
- Developed several interactive tools.

Study Design and Analysis

TRR Articles

Year	Number of Articles	Total Words in Titles	Total Words in Abstracts
1974	368	3,002	52,494
1975	222	1,741	31,397
1976	623	5,256	96,121
2016	875	10,847	186,506
2017	866	10,812	182,574
2018	719	9,300	153,102
2019 (partial)	584	7,461	124,637
Grand Total	**30,784**	**322,732**	**5,791,072**

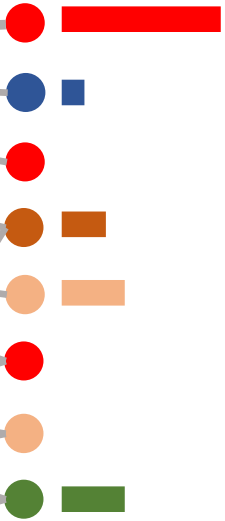
What is topic model?

Topics

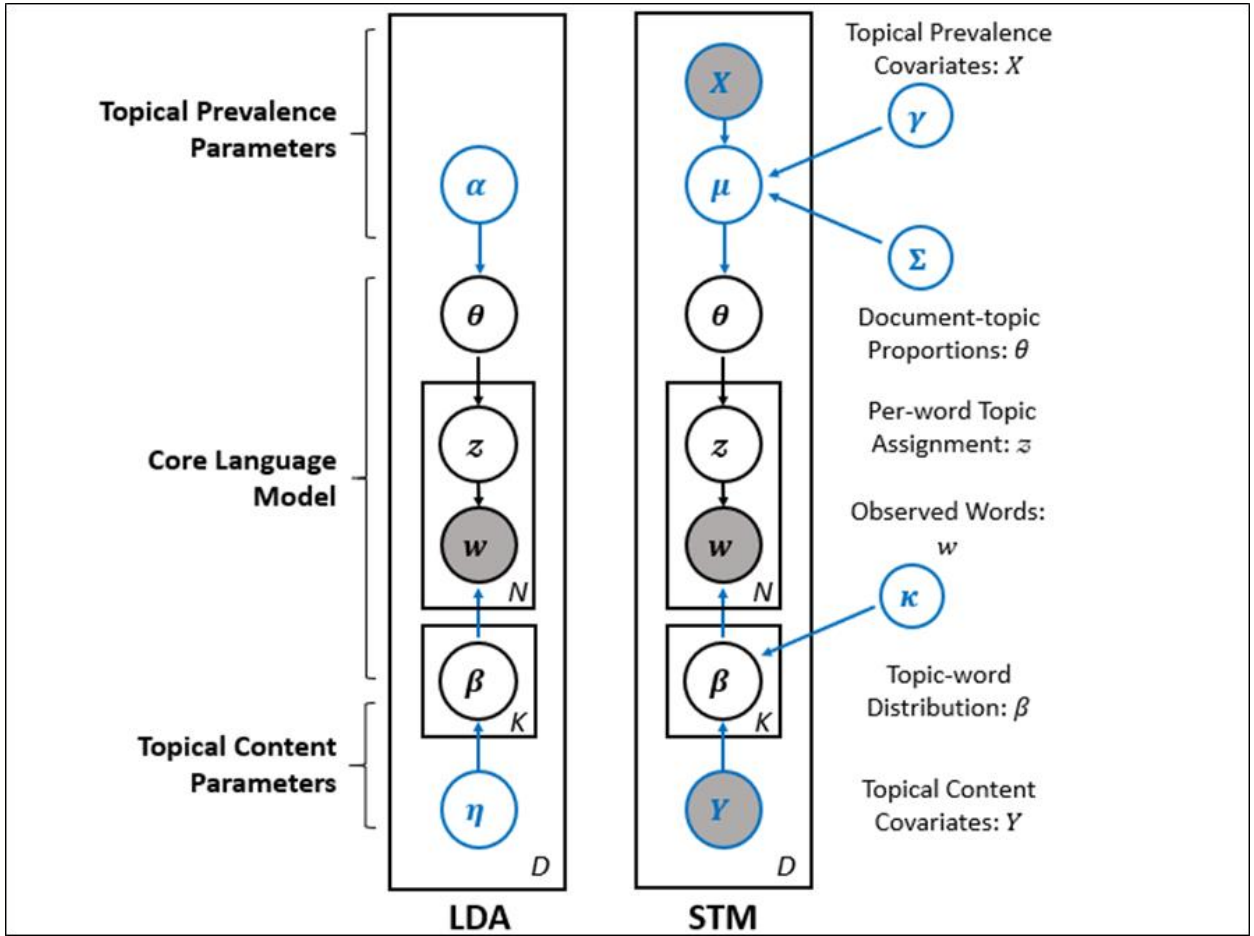
Visibility =0.5
Reduced= 0.2

Crash =0.2
Increase= 0.13

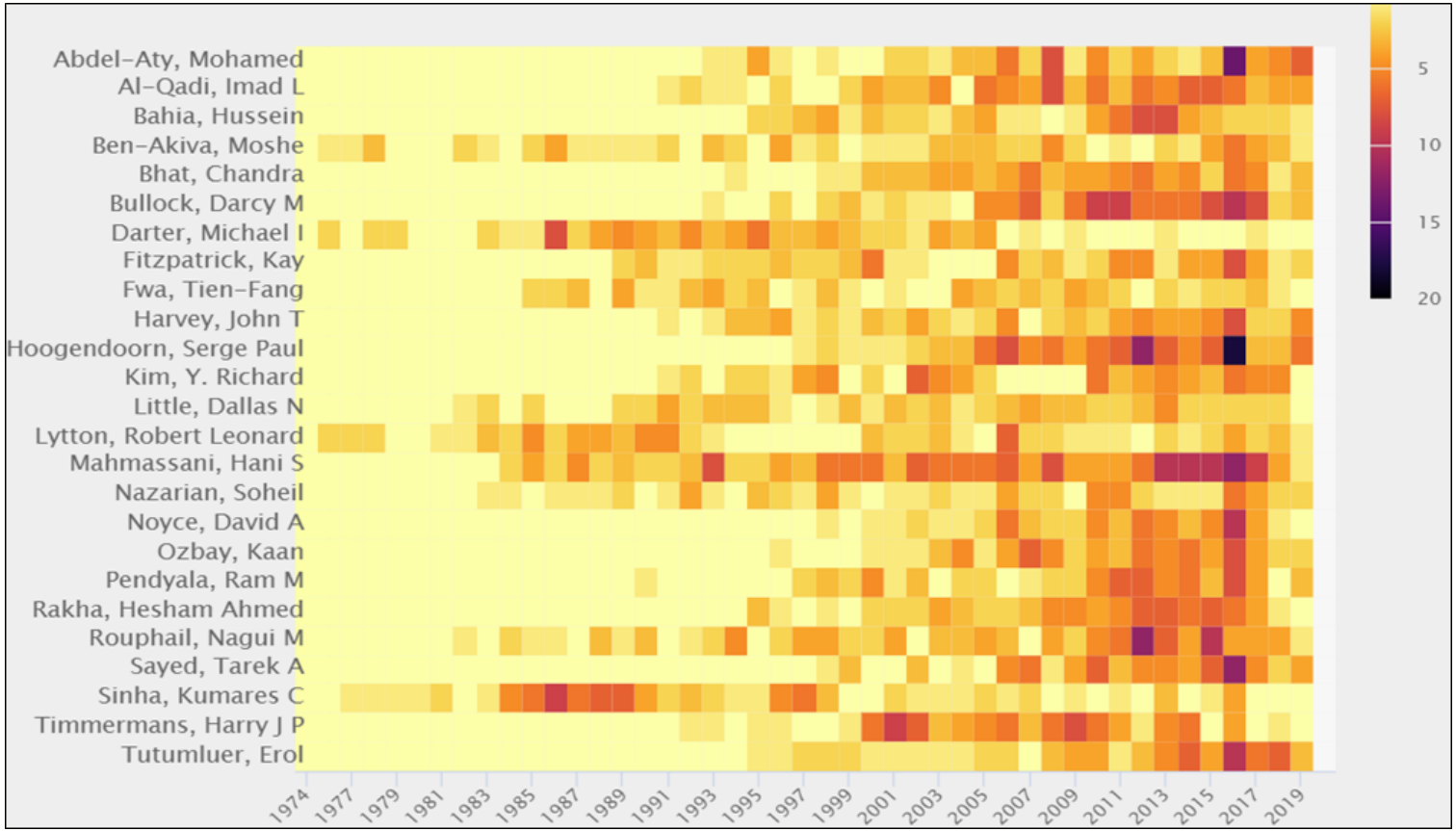
Most of the information used by drivers is acquired visually. In spite of its importance, visibility conditions at the time of a crash are often not documented at a high level of detail. A quantified investigation of the visibility level at the time of a crash was undertaken to investigate the increase in risk associated with driving during periods of reduced visibility. The study method blended data collected from the National Oceanic and Atmospheric Administration (NOAA) with reported crashes in Florida. From the thousands of logged weather events collected by NOAA, the researchers isolated time periods of normal visibility and comparable time periods of reduced visibility in a matched-pairs study. The crash data were contained in the Roadway Information Database (RID) compiled for the Strategic Highway Research Program 2 (SHRP2). The RID contains several geometric and traffic variables that allow for analyses to account for effects of factors other than visibility. The findings indicate that, as expected, the likelihood of a crash increases during periods of low visibility, despite the tendency for less traffic and lower speeds to prevail during these times.



LDA and STM

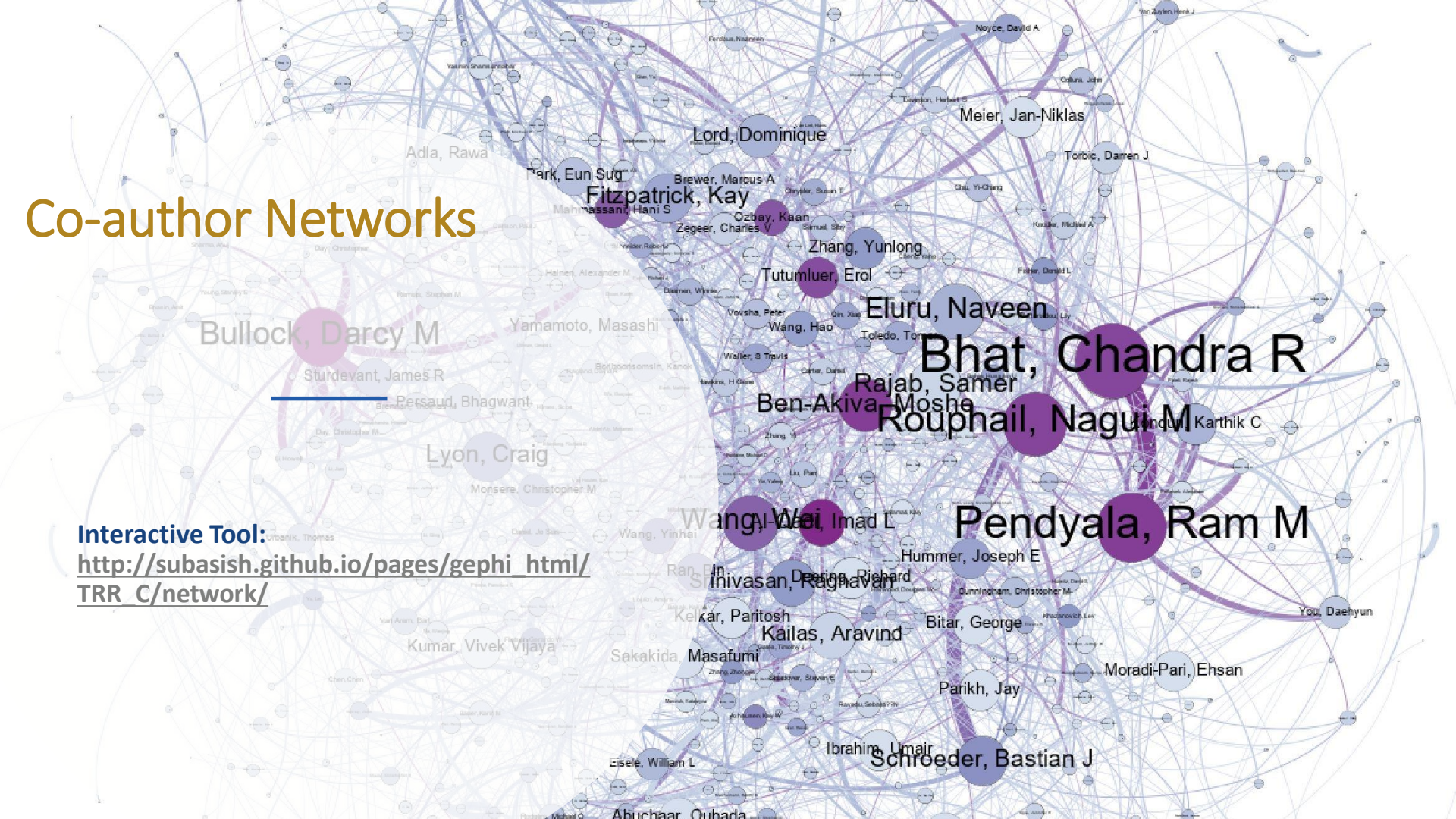

Topical Prevalence Parameters
Core Language Model
Topical Content Parameters
α
θ
z
w
N
β
K
η
D
LDA
X
μ
θ
z
w
N
β
K
Y
D
STM
Topical Prevalence Covariates: X
γ
Σ
Document-topic Proportions: θ
Per-word Topic Assignment: z
Observed Words: w
κ
Topic-word Distribution: β
Topical Content Covariates: Y

Prolific TRR Authors



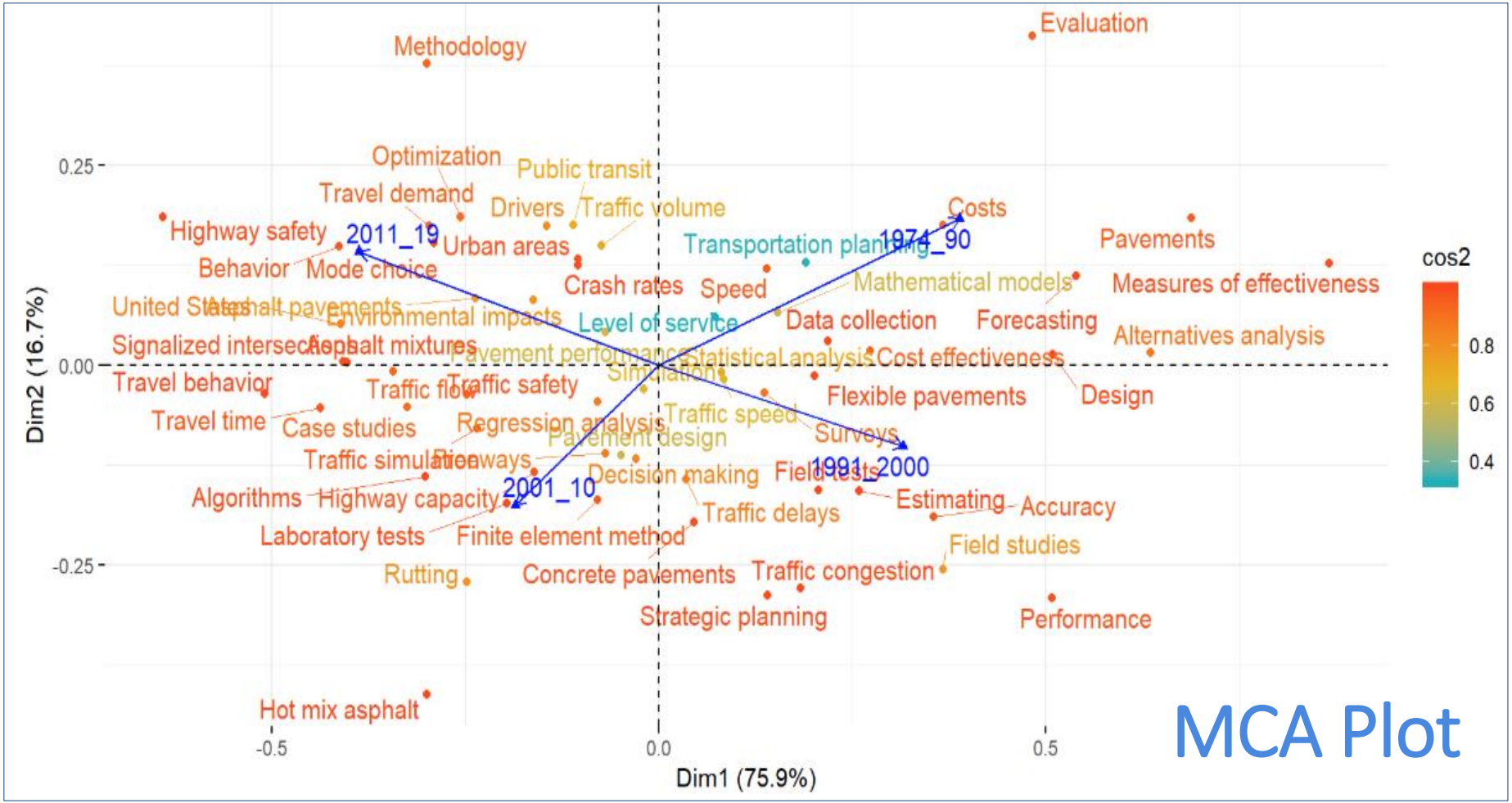
Interactive Tool: https://rpubs.com/subasish/507543

Co-author Networks



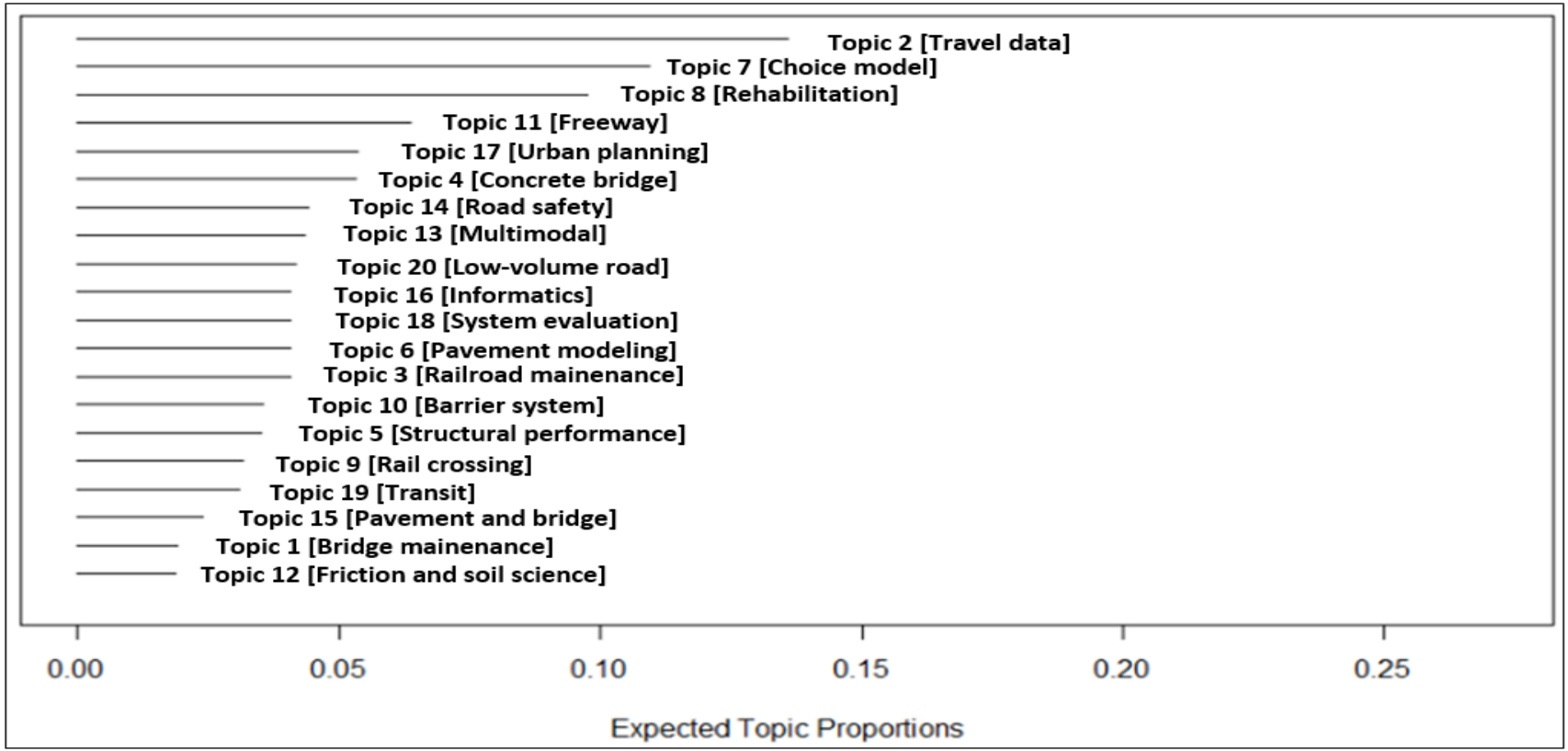
Interactive Tool:
http://subasish.github.io/pages/gephi_html/TRR_C/network/

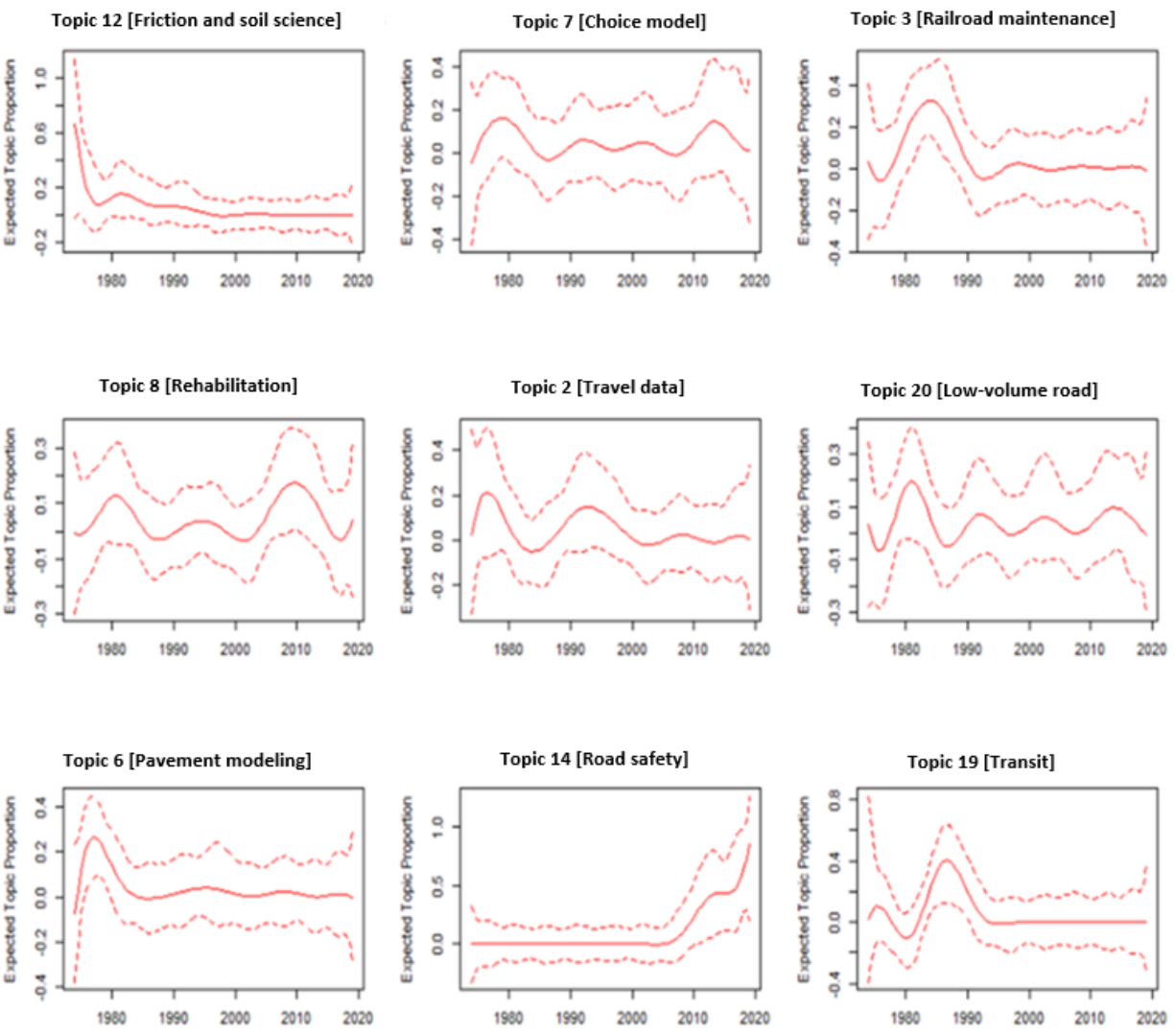
Results and Tools

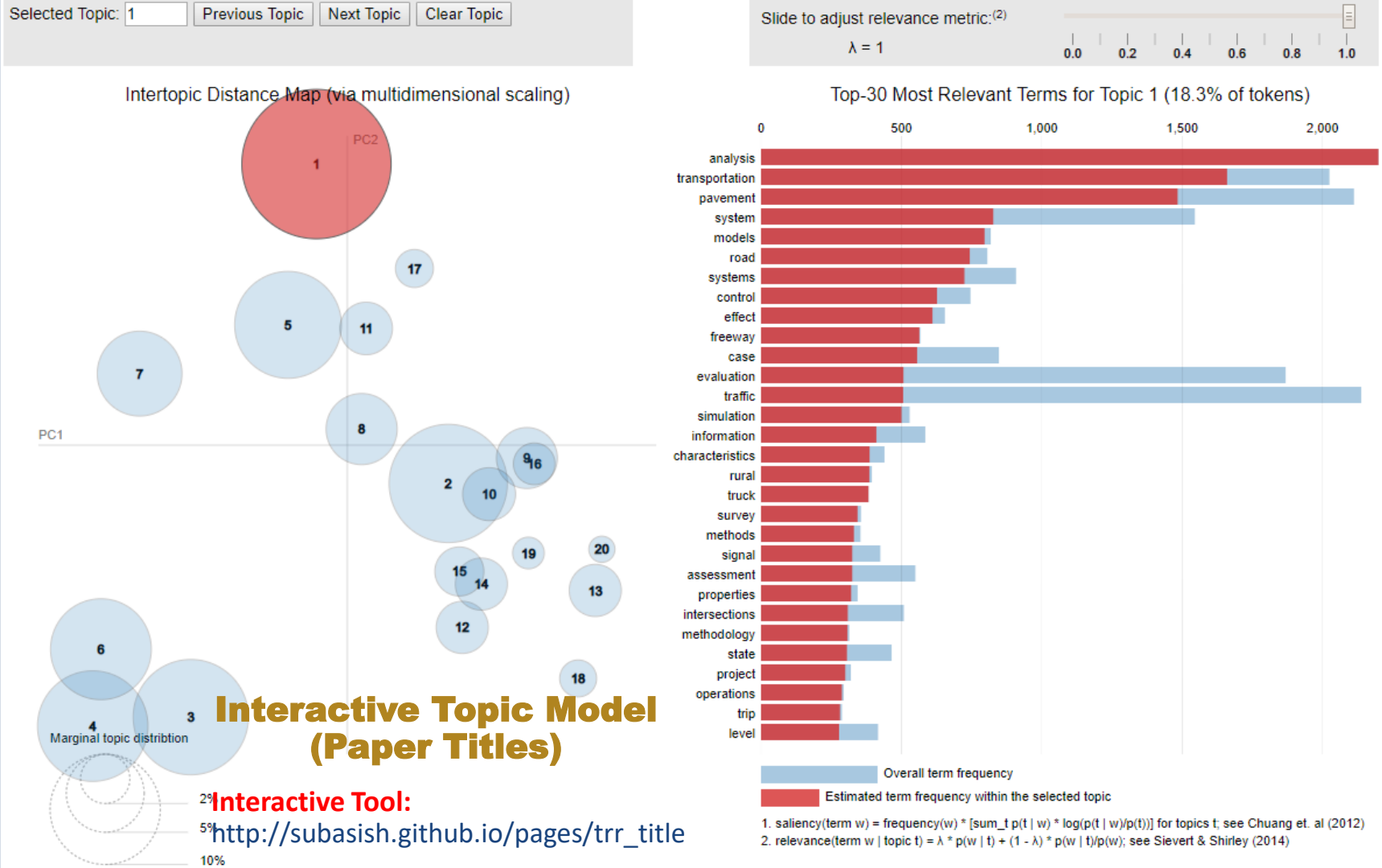

Methodology
Evaluation
Optimization
Public transit
Travel demand
Drivers
Traffic volume
Costs
Highway safety
2011_19
Urban areas
Transportation planning
1974_90
Pavements
Behavior
Mode choice
Crash rates
Speed
Mathematical models
Measures of effectiveness
United States
Asphalt pavements
Environmental impacts
Level of service
Data collection
Forecasting
Alternatives analysis
Signalized intersections
Asphalt mixtures
Pavement performance
Statistical analysis
Cost effectiveness
Travel behavior
Traffic flow
Traffic safety
Simulation
Flexible pavements
Design
Travel time
Case studies
Regression analysis
Traffic speed
Surveys
Pavement design
Traffic simulation
Freeways
Decision making
Field tests
1991_2000
Algorithms
Highway capacity
2001_10
Estimating
Accuracy
Traffic delays
Laboratory tests
Finite element method
Field studies
Rutting
Concrete pavements
Traffic congestion
Strategic planning
Performance
Hot mix asphalt
0.25
0.00
-0.25
-0.5
0.0
0.5
Dim2 (16.7%)
Dim1 (75.9%)
cos2
0.8
0.6
0.4

MCA Plot

Top 20 Topics



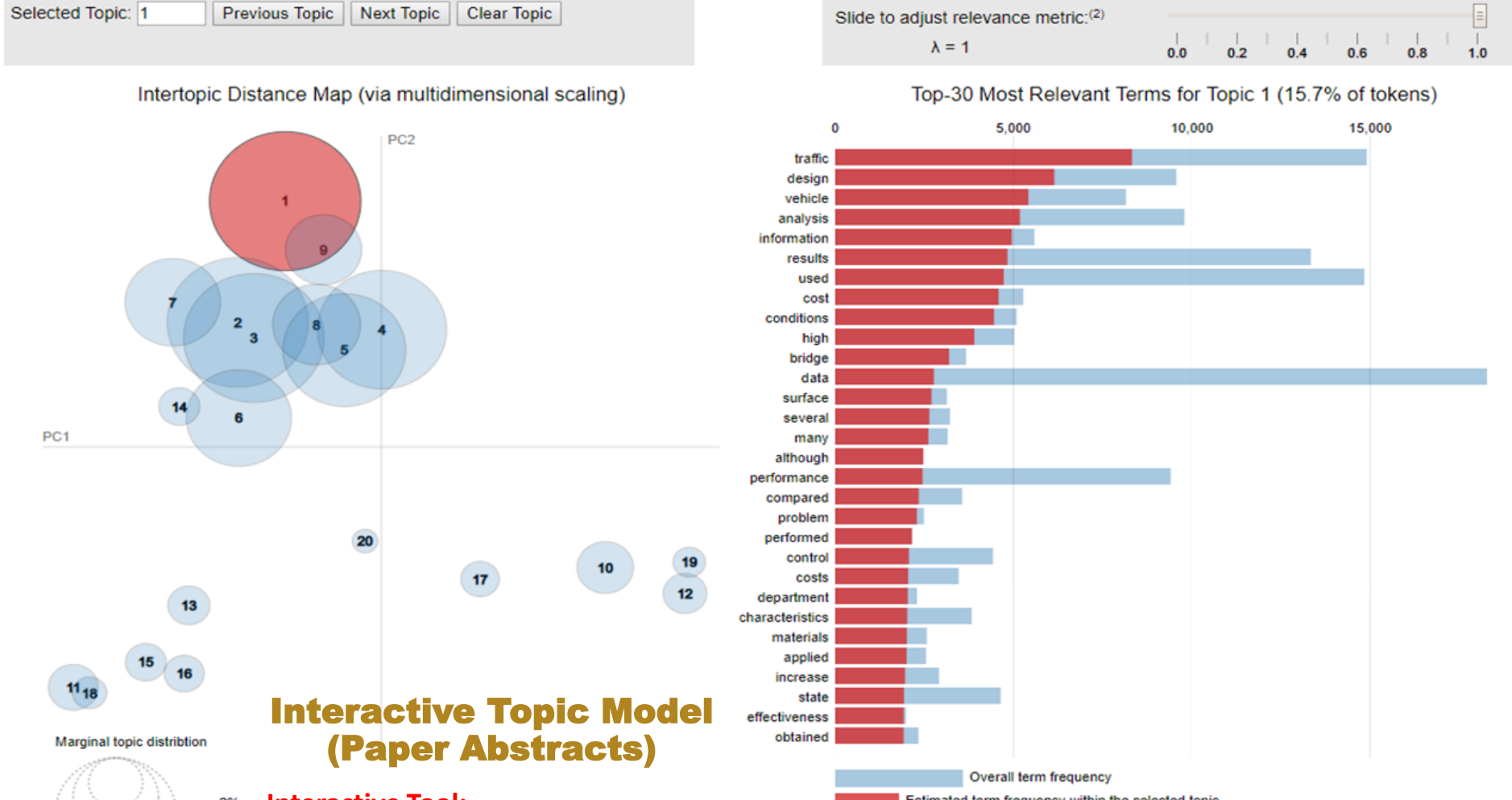

Topic 12 [Friction and soil science]
Topic 7 [Choice model]
Topic 3 [Railroad maintenance]
Topic 8 [Rehabilitation]
Topic 2 [Travel data]
Topic 20 [Low-volume road]
Topic 6 [Pavement modeling]
Topic 14 [Road safety]
Topic 19 [Transit]
Expected Topic Proportion
1980
1990
2000
2010
2020



Interactive Topic Model (Paper Titles)

Interactive Tool:

http://subasish.github.io/pages/trr_title



Interactive Topic Model (Paper Abstracts)

Interactive Tool:
http://subasish.github.io/pages/trr_abstract

Key Takeaways

- The transportation research scope has become more diverse with expanding and inter-disciplinary coverage of topics. Trends and patterns of research is rapidly changing.
- To explore more relevant patterns in the broad fields of transportation research, this study presents a unique replicable framework to probe present content and prevalence to develop a disaggregated level correlation.
- In addition, this study produced two topic model interactive tools developed separately for TRR paper abstracts and titles.


Questions?

Texas A&M Transportation Institute

Subasish Das

s-das@tti.tamu.edu

979-317-2153