Project Proposal

Max Lin & Chih-Ying Liu

2026-03-04

Problem Statement

We aim to analyze spatial temporal pattern, and evaluate the spatial equity between the communities. The inputs include geo-located traffic events, time-stamped records, community-level spatial partitions, and demographic and exposure data. By coordinating these data, we first detect the co-location traffic pattern and find out which places tend to occur the traffic events. Then we are going to take the demographic and resource data to evaluate if every community fairly take the traffic risk with proper resources. The ultimate goal is to understand the relationship between traffic patterns and community resource conditions. It can be the reference of policy making on transportation and public resource.

Significance of Problem

Rush-hour traffic often creates concentrated risks in specific locations and time periods. However, these traffic exposures are not evenly distributed across the communities. Some regions may take higher risk of congestion or the relative issues. By identifying spatial–temporal traffic patterns, we can better understand where these risks are concentrated and whether certain communities are disproportionately affected. This analysis is helpful on evaluating the spatial inequality in transportation system and provide the more effective and equitable policies and strategies for decision makers.

Challenges

The main challenges are as the following: First, averaging traffic data over time may hide important rush-hour patterns and traffic events vary significantly across both time and locations, making it difficult to accurately identify high-risk areas. Second, the different communities have the different populations, so it is necessary to perform the normalization. If directly uses the original data, it may causes the anomalous results. Lastly, detecting the high-density traffic zones and integrate the traffic data with demographic information requires more complex analysis and computations.

Proposed Approach

Our approach first detects rush-hour traffic risk patterns using geo-located traffic events. By applying a co-location technique, we identify areas where traffic events tend to cluster during morning and evening rush hours. After that, we calculate the Participation Rate for each traffic feature and use the minimum PR to derive the Participation Index (PI). Meanwhile, we also take time dimension into account, avoiding average the information an cause the information loss. For the next step, we normalize the participation index by dividing the community population, getting the new index called Equity-aware Participation Index(EPI) defined by ourselves. This index is used to evaluate whether traffic risks are disproportionately distributed relative to population. Finally, we also conduct the statistical inference and permutation tests for examining spatial equity across communities. Here, Resource Availability(RA), which represents nearby public facilities such as hospitals and clinics, is also incorporated for the analysis.

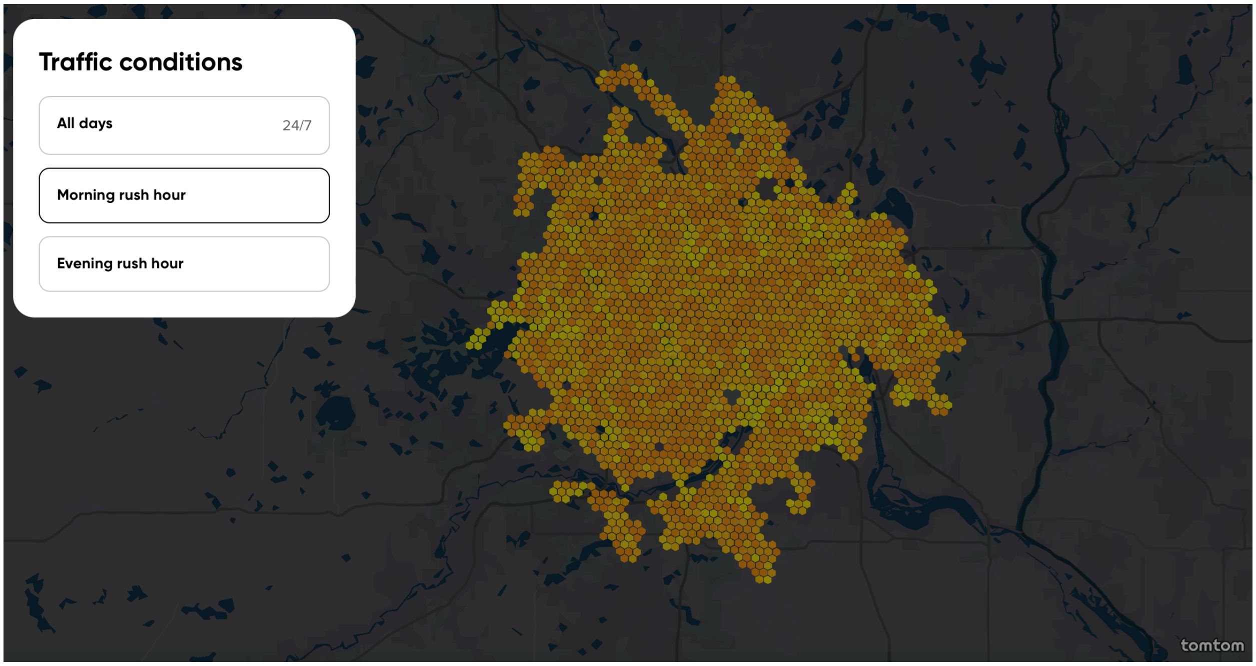

Figure 1: Morning Rush-hour

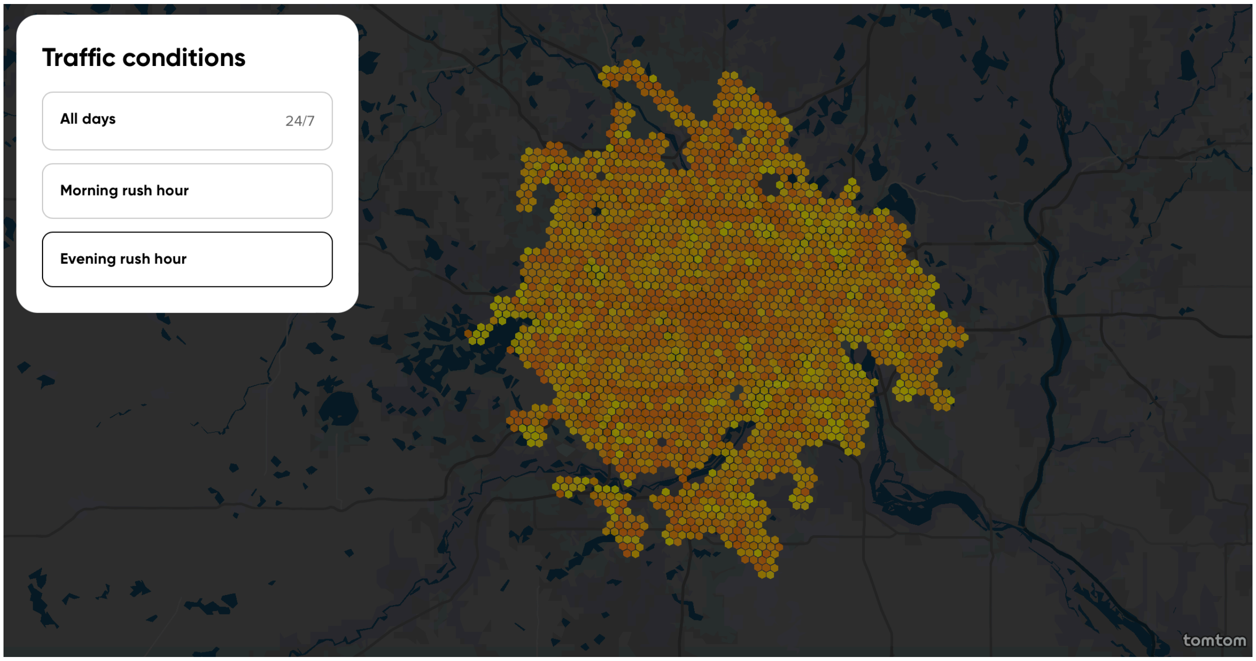

Figure 2: Evening Rush-hour

Below is a simple example demonstrates how these indices can reveal potential spatial inequality in traffic exposure and resource distribution. Also, the related formulas are displayed.

	Community A	Community B	Community C	Community D
Pop	10000	20000	15000	5000
w	0.2	0.4	0.3	0.1
PR for accident	0.3	0.2	0.2	0.4
PR for congestion	0.2	0.2	0.1	0.2
PI	0.2	0.2	0.1	0.2
EPI	1	0.5	0.33	2
F	5	12	8	1
RA	0.0005	0.0006	0.0053	0.0002
Equity Indicator(x)	7.6	6.73	6.44	9.21

T_obs	1.17
Perms	999
Count T >= T_bos	29
p-value	0.03

Spatial inequality of resources exists

9

Figure 3: Simple Example

$PR(C, f_i) = \frac{|I(C,f_i)|}{|I(f_i)|}$

$PI(C) = min_{f_i \in C} PR(C, f_i)$

$PI(C,t) = min_{f_i \in C} PR(C, f_i, t)$

$EPI(C,r,t) = \frac{PI(C,t)}{w_r}$ where $w_r = \frac{Pop_r}{Pop_{total}}$

$x_i = log\left(\frac{EPI_i}{RA_i}\right)$ where $RA_i = \frac{F_i}{Pop_i}$

$H_0 : \mathrm{Var}(x_1, x_2, \dots, x_n) = 0$

$H_1 : \mathrm{Var}(x_1, x_2, \dots, x_n) > 0$

$RA_1, RA_2, \dots, RA_n$

$x_i^{(b)} = \log\left(\frac{EPI_i}{RA_i^{(b)}}\right)$

$T^{(b)} = \mathrm{Var}\left(x_1^{(b)}, \dots, x_n^{(b)}\right)$

$b = 1, \dots, B$

$T_{obs} = \mathrm{Var}(x_1, \dots, x_n)$

$p = \frac{\#\{T^{(b)} \geq T_{obs}\}+1}{B+1}$

$p_i = \Pr\left(x_i^{perm} \geq x_i\right)$

Related Work

Spatial Co-location Pattern Mining

Spatial data mining research has widely adopted co-location pattern discovery to identify spatial features that frequently occur together with a geographic neighborhood. Spatial Co-location mining focuses on discovering sets of spatial objects which tend to appear in close spatial distance. Typical approaches include data-unaware spatial partitioning methods, such as grid-based partitioning which divide the spatial domain into predefined regions for pattern discovery. Another category is data-aware clustering methods which detect potential localities based on clusters of spatial objects. These techniques have been widely applied in many reality area like transportation analysis, urban planning, smart city applications and so on. For example, co-location pattern mining can identify spatial relationships between traffic accidents and road intersections helping researchers understand urban traffic dynamics. However, most existing co-location studies primarily focus on spatial relationships, while temporal dynamics and equity are often ignored or replaced.

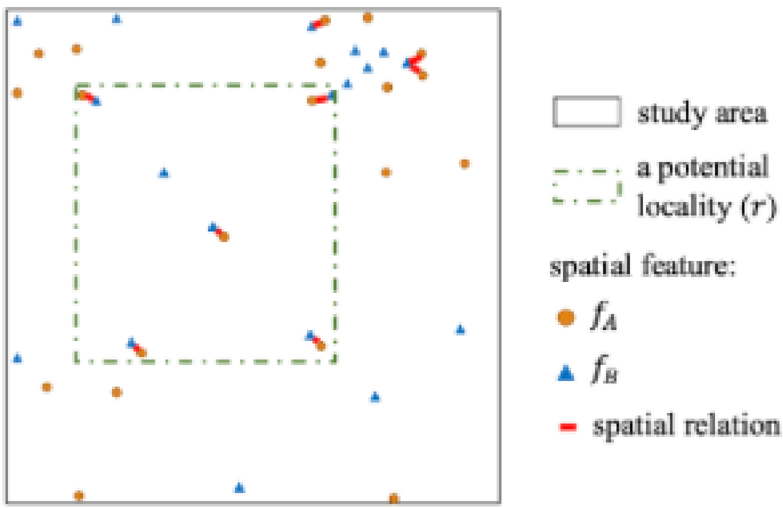

Figure 4: Yan Li, Shashi Shekhar 2018)

Traffic Hotspot Detection

Another related research area is traffic hotspot detection which analyzes traffic crash patterns using spatiotemporal clustering techniques. Previous studies have detailed crash databases containing precise geographic coordinates and timestamps of traffic incidents to identify high-risk locations on road networks. For example, studies analyzing traffic crashes on primary road networks have identified several types of hotspot dynamics, including hotspot emergence, hotspot stability, and hotspot disappearance. These patterns reveal how traffic risk evolves over time and how the change of risk that locations change across different periods. Although hotspot detection methods are effective for identifying areas with high accident density, they primarily focus on spatial concentration of traffic events, rather than interactions among different spatial features or community characteristics and equity.

Limitations of Existing Work

Despite the progress made in spatial data mining and traffic analysis, there are some several important limitations remaining.

First, existing studies rarely integrate multiple analytical components into a same framework for traffic equity evaluation. Current approaches often view spatial pattern mining, temporal analysis, demographic normalization, and public resource availability as separate research problems.

Second, many studies in spatial equity analysis primarily focus on accessibility to infrastructure, such as public facilities or transportation services. While accessibility is an important aspect of urban equity, these approaches often overlook dynamic risk exposure, such as traffic accidents or congestion during peak hours.

Table 1. Clustering results obtained using a three-year window for all traffic crashes which occurred on the Czech primary roads network between 2010 and 2018.

	10–12	11–13	12–14	13–15	14–16	15–17	16–18
No of hotspots	2242	2362	2369	2434	2482	2491	2561
Overall length [km]	266	274	283	296	303	307	323
Overall length [%]	6.8	7.0	7.2	7.5	7.7	7.8	8.2
No of TC within hotspots	7315	7775	8140	8719	9092	9417	9990
No of TC within hotspots [%]	47.7	47.9	47.3	47.2	47.3	45.6	46.5

Michal Bíl, Richard Andrášik, Jiří Sedoník (2019)

Figure 5: Michal Bíl, Richard Andrášik, Jiří Sedoník 2019)

Finally, most existing methods do not explicitly incorporate spatiotemporal interactions between traffic events and community characteristics, it may limits their ability to evaluate whether certain communities experience disproportionate traffic risks.

These limitations motivate us to creating an integrated framework that combines spatialtemporal co-location detection and equity analysis, which is the main objective of this research.

Methodology

Experimental Methodology

This study adopts an experimental methodology to detect spatial-temporal traffic patterns with rush-hour hotspots and evaluate spatial equity across communities. The proposed framework consists of four main steps: data processing, spatial-temporal co-location detection, equity normalization, and community comparison.

First, data processing is performed to ensure consistent spatial units across multiple datasets. Traffic events, demographic information, and public resource locations are integrated into a unified spatial grid structure. This spatial partitioning ensures that traffic risk, population distribution, and resource availability can be analyzed with the same spatial framework. The traffic event data are collected from transportation data sources, while public infrastructure information such as hospitals and clinics can be obtained through geographic data platforms such as the Overpass API and government websites.

Second, spatial temporal co-location pattern detection is conducted to identify traffic features that frequently occur together during rush hours. In this step, traffic events occurring with rush-hour periods are extracted from the dataset. Using co-location mining techniques, spatial features such as traffic accidents, congestion points, or intersections are analyzed to identify patterns of spatial relationship. A Participation Index (PI) is computed to measure the power of these spatial-temporal relationships. Unlike traditional participation index calculations, which average patterns over time, the time-aware PI focuses specifically on rush hour periods, enabling the detection of high risk traffic patterns that may otherwise be masked by daily averages.

Figure 6: Data Cleaning

Equity Normalization and Community Comparison

After detecting spatial-temporal traffic patterns, the next step is equity normalization. Because different communities have different population sizes and traffic or facility needy, traffic risk must be normalized to enable fairness across communities.

To address this issue, we introduce the Equity aware Participation Index (EPI). First, the population ratio for each community is computed as the proportion of the community population relative to the total population in the study area. The Participation Index is then divided by this population ratio to produce the EPI value. This normalization step allows the framework to evaluate whether certain communities experience disproportionately high traffic risk relative to their population size.

By comparing EPI values across communities, the model identifies areas where traffic risk exposure is significantly higher than expected. Communities with relatively high EPI values indicate potential spatial inequities, suggesting that residents in those areas may be exposed to higher traffic risks compared to other communities.

Evaluation Metrics

To evaluate the effectiveness of the proposed framework, several evaluation metrics are used.

First, the Participation Index (PI) measures the power of spatial-temporal co-location patterns among traffic events.

Second, the Equity-aware Participation Index (EPI) evaluates the level of traffic risk exposure after demographic normalization. Third, the variance of normalized traffic risk is calculated to measure the overall level of spatial disparity among communities.

Finally, permutation tests are applied to statistically evaluate whether the observed spatial disparities are significant. The p-values obtained from permutation testing indicate whether the distribution of traffic risks across communities differs significantly from what would be expected under random conditions.

Together, these evaluation metrics allow the proposed framework to measure both spatial traffic patterns and equity disparities across communities, providing a comprehensive evaluation of traffic risk distribution.

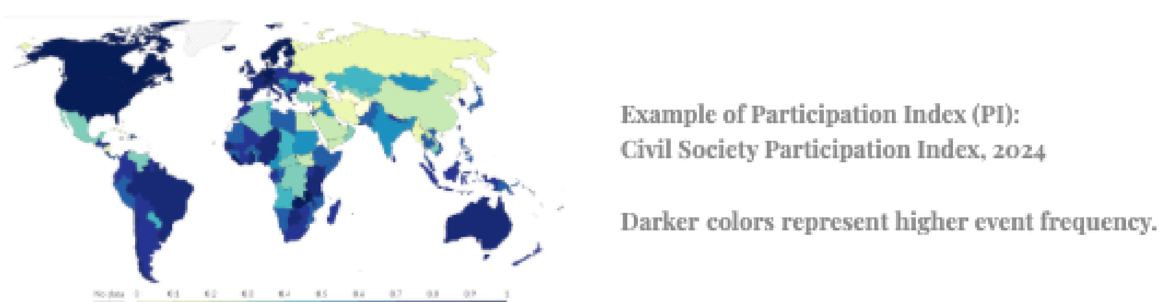

Figure 7: Example of Participation Index (PI)

Weekly Plan

Week timeline

Tasks

03/16 – 03/20

Data cleaning and initial code structure planning

03/23 – 03/27

Analyze rush-hour temporal patterns in the dataset

03/30 – 04/03

Define fairness metrics and complete the local co-location model

04/06 – 04/10

Model training

04/13 – 04/17

Analyze results and adjust the model

04/20 – 04/24

Final presentations begin