Short and Long-term Pattern Discovery over Large-Scale Geo-Spatiotemporal Data

Sobhan Moosavi, Mohammad Hossein Samavatian, Arnab Nandi, Srinivasan Parthasarathy, Rajiv Ramnath

Department of Computer Science and Engineering

THE OHIO STATE UNIVERSITY

EAR-1520870

Introduction

- **Geo-spatiotemporal Data:** data that has *location* and *time*
  - Examples: Traffic events, weather events, and crime events

Traffic events

Weather events

- **Pattern Discovery:** exploring co-location, co-occurrence patterns that point to cause-and-effect
  - Example 1: *Rain ⟶ Accident ⟶ Congestion*
  - Example 2: *Bar-closing ⟶ Burglary ⟶ Harassment*
- **Importance:**
  - Traffic: useful for urban planning, traffic management, and prediction
  - Weather: useful for disaster prediction and management
  - Crime: to raise awareness and safety

Contributions

- A **new Framework** to explore two types of patterns
  - **Short-term (propagation):** short-term cause-and-effect
  - **Long-term (influential):** impact of a temporally long event on its neighborhood
- A **new Dataset** of geo-spatiotemporal events (i.e. traffic and weather events)

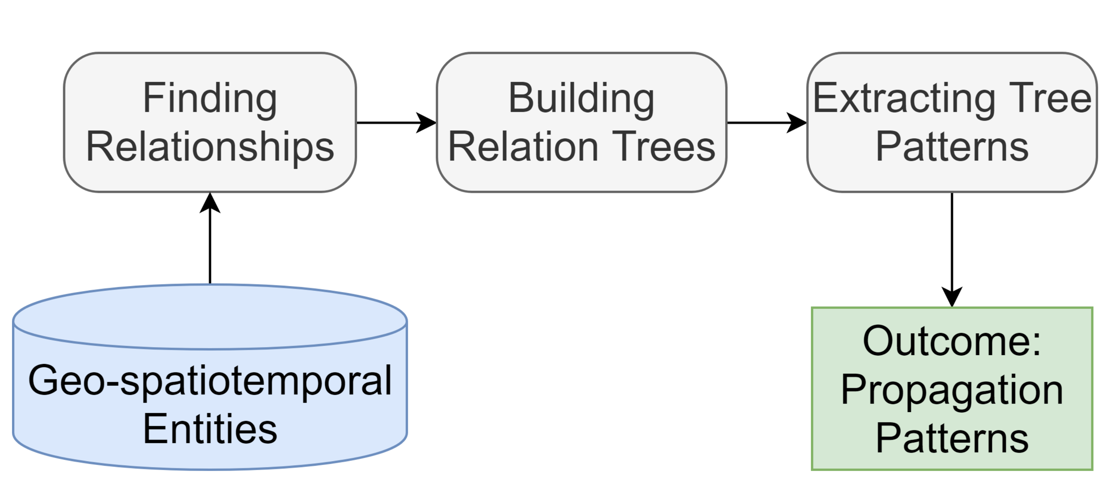
Challenges

- **Problem**: No *special purpose* framework exists for geo-spatiotemporal data
- How about existing *general-purpose* frameworks?
  - Such as: *Huan et al. 2004, Mohan et al. 2010, and Shekar et al. 2015*
  - Challenge: simplistic assumptions to achieve reasonable runtime

- Outcome: not useful for geo-spatiotemporal data

Propagation Patterns

- **Propagation or Short-term Patterns:** A group of temporally co-occurring, spatially co-located, and semantically relevant events
- **How to be discovered?** We propose a multi-step process



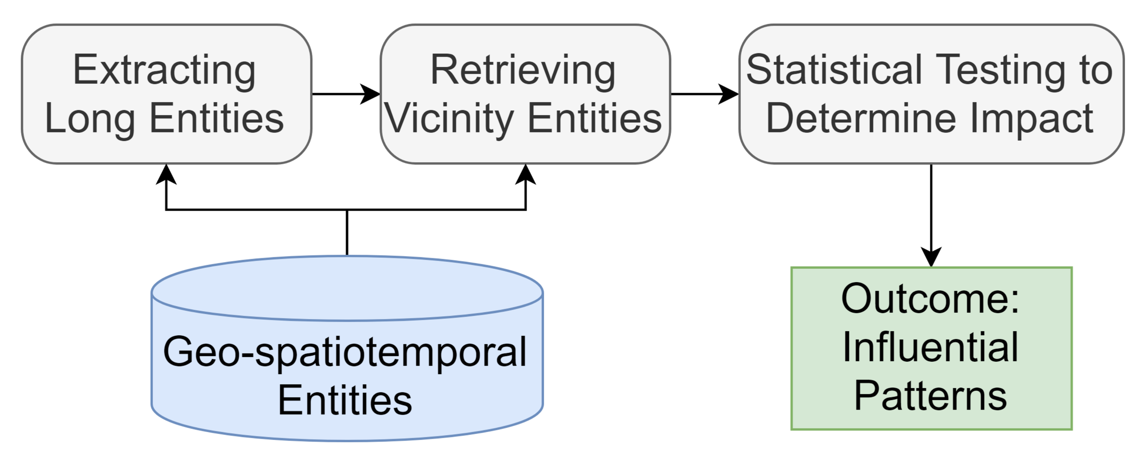
- **The Multi-Step Process**
  1. Identify **child-parent** relations
  2. Build a **forest** of relation trees
  3. Extract embedded, un-ordered frequent **tree patterns** (Zaki 2005)

Influential Patterns

- The impact of a temporally long-term event on its spatiotemporal neighborhood

Example: *major construction ⟶ more traffic jams*

- **How to be discovered?**



- Examining spatiotemporal neighborhood of long events
  - Spatial neighborhood **radius**: 14 miles
  - Comparing *current* with *before* and *after time intervals*
  - Using **two-sample t-test**
  - T-tests: positive and negative impact test
- **Outcome:**
  - **Positive** impact: existence of a long-term event ⟶ *more traffic issues*
  - **Negative** impact: existence of a long-term event ⟶ *less traffic issues*

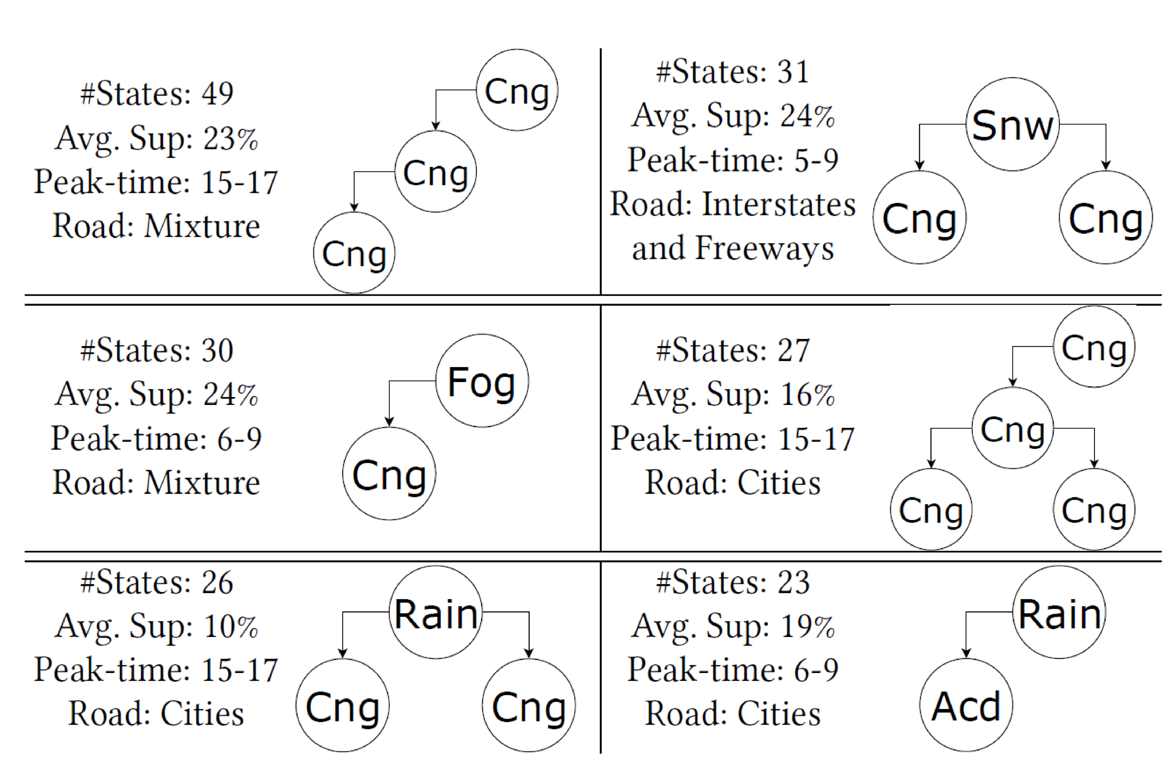
Dataset

- Large-scale Traffic and Weather Events Dataset
- Covers the contiguous United States (50 states) from August 2016 to August 2018
- Contains 15 million events, 13 million traffic and 2 million weather events

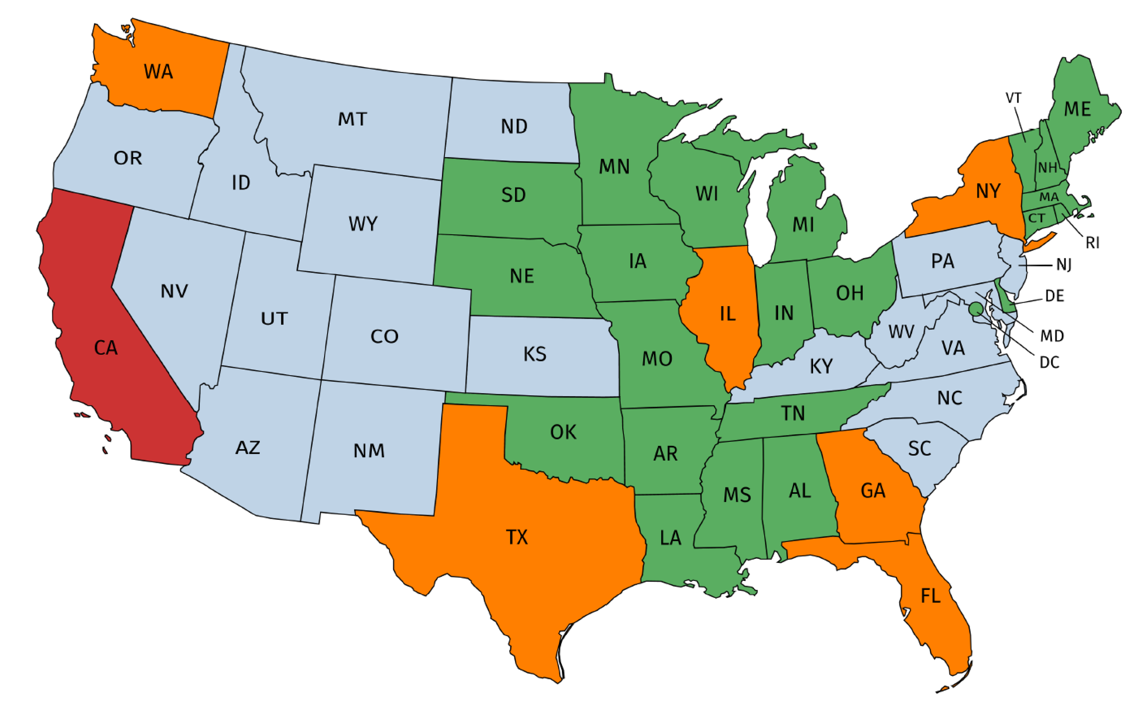
Scan to access the dataset

Results (Short-term)

- **Pattern discovery level**: Cities
  - To account for diversity
  - Assigned ***core*** patterns to corresponding state
- Extracted Relations: ~ 6 million
- Extracted Trees: ~ 1.7 million
- Extracted 708 patterns for 50 states, **90 unique** patterns

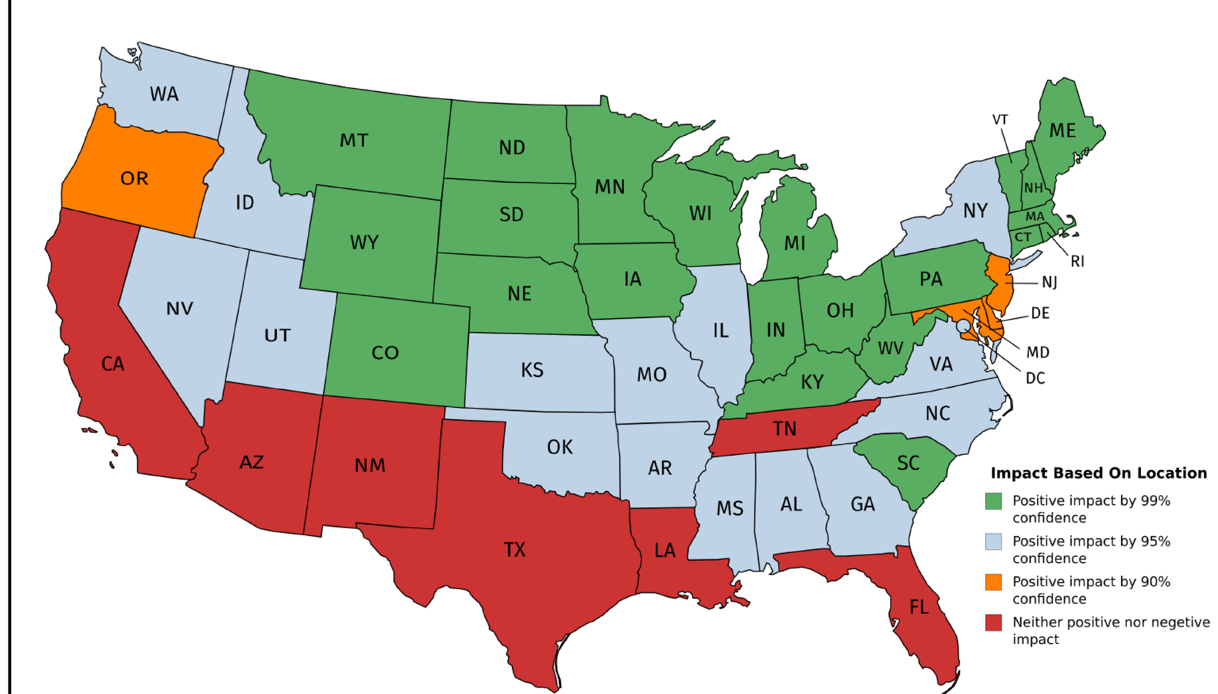


- **Clustering**: categorized states into four clusters by K-Means based on their short-term patterns

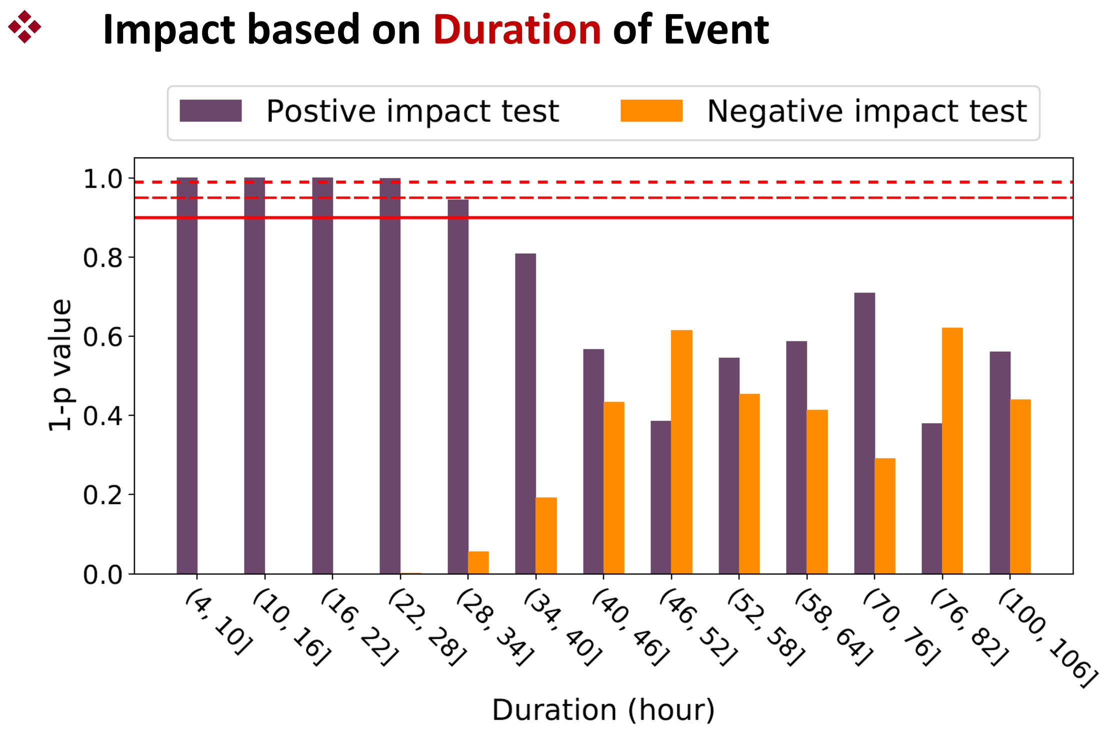


Results (long-term)

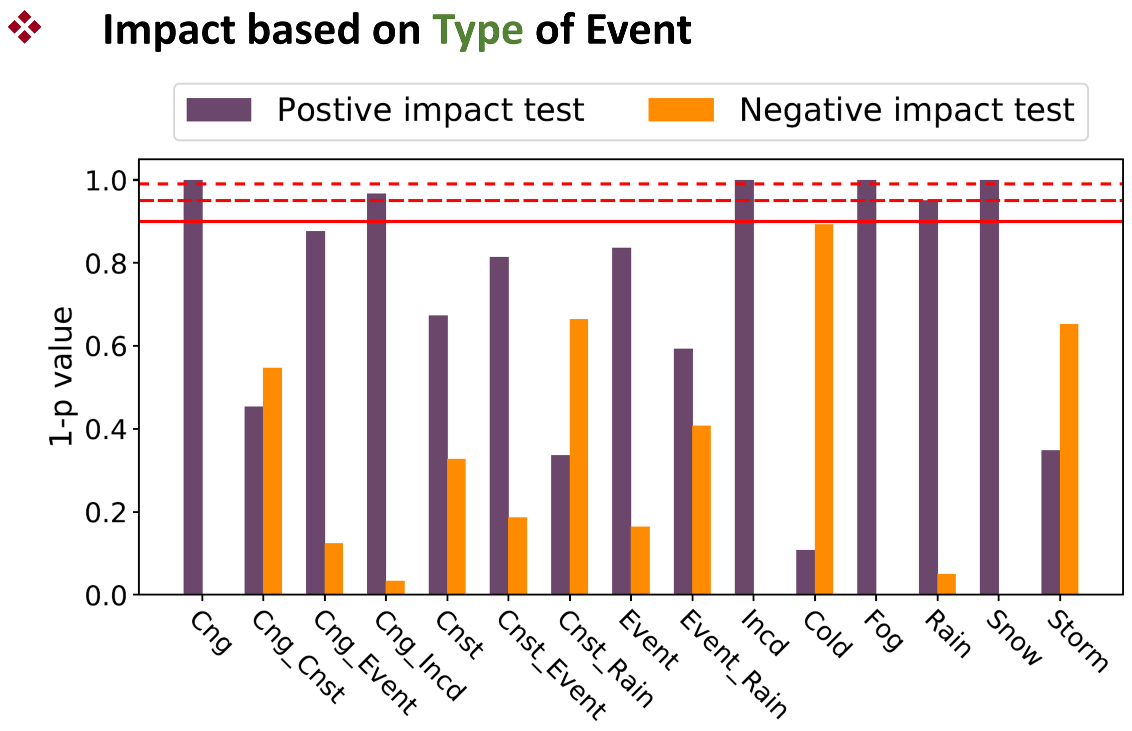
- **Long events**: 145,000 (duration > 5 hours)
  - Majority: *Rain*, *Snow*, and *Construction*
- **Impact based on Location of Event**



- **Impact based on Duration of Event**



- **Impact based on Type of Event**



Conclusion and More

- Introduced a new framework to discover *geo-spatiotemporal* patterns
- Constructed and shared a new *dataset*
- Gleaned various *insights* from these patterns

Scan for Code, Poster, and More: