Critical Analysis of “Cascading Spatio-Temporal Pattern Discovery

GROUP 8: PATRICK SKOGLUND AND XIANG LI
Given a spatio-temporal dataset with Boolean-types events over a spatio-temporal framework, a direct neighborhood relation, and a threshold for the participation index.

Find the cascading spatio-temporal patterns with participation greater than the index.

While minimizing run time.

And, making sure that all, and only, correct and statistically significant patterns are found.

This idea has many applications from crime to climate to public health.
Contributions

- A better algorithm (TKDE CSTPM) for computing the interest measure while not impacting the correctness nor completeness of the previously developed algorithm.
- Bottleneck analysis of TKDE CSTPM.
- Analysis showing TKDE CSTPM outperforms the previous algorithm.
- Experimental analysis of TKDE CSTPM
A CSTP is a partially ordered subset of events whose instances are spatially close and happen in stages.

Ex.

Legend
- Bar Closing (B)
- Assault (A)
- Drunk Driving (C)

Example CSTP

(a) (b) (c) (d) (e)
A directed neighborhood relation takes the information from a CSTP and creates an acyclic graph where the edges are dependent upon the distance between the two events in space and time.

Ex.

![Directed Neighborhood Relation Diagram]
The CPR is an estimate of the conditional probability that the CSTP occurs given an event inside the CSTP occurs.

Ex. Using the directed neighborhood graph below, the CPR for the event B and the CSTP (B→A) is 2/2 because both instances of event B participate in the CSTP (B→A)
Key Concepts: Cascade Participation Index (CPI)

- The CPI is a measure of the likelihood that an instance of an event type participates in the CSTP. It is the minimum of the CPR values of the event types in the CSTP.

- Ex. The CPI for the CSTP (B->A) in the directed neighborhood graph below is min(2/2, 4/5)=0.8
The filtering process in an apriori algorithm that does not generate possible candidates from patterns that have low CPI. Additionally, a multi-resolution filter is used to further prune the patterns before candidate generation.

Ex. Candidate generation; multi-resolution filtering
The paper uses theoretical proofs, simulations and an extended case study to show that TKDE CSTPM works. The theoretical proofs help to show that the algorithm works as proposed. The simulations, and extended case study to show that TKDE CSTPM with filtering works better than their previous algorithm and better than TKDE CSTPM without filtering in terms of run time. The authors also ran additional tests to determine the effect of dataset size, number of event types, and clumpiness on TKDE CSTPM.

The strength of doing all of these validation techniques is that they have made sure that the algorithm is correct and runs well. The only downside is that of possibly too much information on it in the paper.
Assumptions

- Boolean-Type events (may be violated for events that we care about severity).
  - If this assumption is broken, it is possible to find patterns that are not there or miss patterns that are
- Stationarity of the data (may be violated for datasets with larger timescales).
- Directed neighborhood choices premade (separate issue).
- Interest measure threshold predefined (depends on domain).
- Appropriate grid size for MST (separate issue).
- CSTP are partially (or totally) ordered (we are unable to find cyclic patterns).
Revisions- Preservation

- We would preserve almost everything.
  - The organization of topics flows fairly well.
  - The proofs are solid
  - The paper references previous papers when necessary
The things we would revise are mostly minor things

- The locations of the figures in the paper could be better. The readers are constantly flipping back and forth while trying to follow the toy example.
- There are slight annotation errors in the figures including seemingly incorrect values and the order of items in a legend.
- The algorithm for MST could be explained better, so readers can follow along.
- Related work and new contributions could be slightly more distinct.
- 10 Days seems like too large of a time scale for the real data. It might be valid, but there was no explanation as to why it was chosen.