Discovering Graph Temporal Association Rules

Mohammad Hossein Namaki  Yinghui Wu  Qi Song
Peng Lin  Tingjian Ge*

Washington State University,  *UMass Lowell
Temporal association rules in networks

- **Time-aware POI recommendation**

  \[ u \overset{\text{retweet}}{\rightarrow} \tilde{u} \quad \text{check-in} \quad \leq 2 \text{ hours} \]

  \[ u \quad \text{check-in} \quad \text{nearby} \quad \tilde{u} \]

  \[ P_1 \quad \text{POI} \quad P_2 \]

  Left hand side event (LHS)

  Right hand side event (RHS)

- In \( P_2 \) can be recommended as a point of interest for

- is a potential customer for

**Requirement:** AR’s with *topological, semantic and temporal* constraints
Outline

- Graph temporal association rules (GTARs) definition
- GTARs discovery problem formalization
- A feasible GTAR discovery algorithm
- Experiment study: verify the effectiveness of GTARs, and the efficiency of GTAR discovery algorithm.
Temporal Graph

- Temporal graph $G_T(V,E,L,T)$.
- Snapshot $G_t$: induced by the set of all edges associated with time stamp $t$. 
Graph temporal association rules (GTAR)

- GTAR $\phi = (P_1 \Rightarrow P_2, \tilde{u}, \Delta t)$
- $\tilde{u}$: common shared focus.
- $\Delta t$: a constant that specifies a time interval.

If there exists an occurrence of event $P_1$ at an entity specified by $\tilde{u}$ at some time $t$, then it is likely that an event $P_2$ occurs at the same entity, within a time window $[t, t + \Delta t]$.

$\phi = (P_1 \Rightarrow P_2, \tilde{u}, \Delta t=2\text{hours})$
Events and Matching

- **Events**
  - Connected subgraph pattern carry a designated *focus* node.

- **Event matching**
  - An event $P$ occurred in $G_t$ at time $t$ if there is a matching relation ($R_t$) between $P$ and snapshot $G_t$.
  - *Focus occurrence* $o(P, \tilde{u}, t)$: the nodes in $V$ that matches $\tilde{u}$ induced by $R_t$.
  - Example:
    - Matches of $\tilde{u}$ induced by $R_3$ in $G_3$ contains $\{(x_1,3),(x_2,3),(x_3,3)\}$
    - $o(P_1, \tilde{u}, 3)$ is $\{x_1,x_2,x_3\}$
GTAR occurrence

- Given a time window $[t_1,t_2]$, $\phi$ occurs if at least a node matches the focus of both $P_1$ and $P_2$ at $t_1$ and $t_2$, respectively.

- A time window may contain multiple occurrences of a GTAR.

- Minimal occurrence
  
  - $O(v)=[t_1,t_2]$ is an occurrence of $\phi$ in $G_T$ supported by node $v$
  
  - There exists no $O'(v) \subset O(v)$, such that $O'(v)$ is also an occurrence

---

**Diagram:**

- $O_1(x_j)$
- $O_2(x_j)$
- $P_1 = \{x_2, x_3\}$
- $P_2 = \{x_1, x_2\}$
- $O_3(x_j)$
- $O_4(x_j)$
- $O_5(x_j)$
Support and Confidence

- Based on minimal occurrences $O(\varphi, G_T)$

$$\text{Supp}(\varphi, G_T) = \frac{|O(\varphi, G_T)|}{|C(\overline{u})||T|}$$

# Occurrence of this rule

Normalizer

- Confidence: measures how likely $P_2$ occurs within $\Delta t$ time at the focus occurrence of $P_1$

$$\text{Conf}(\varphi, G_T) = \frac{\text{Supp}(\varphi, G_T)}{\text{Supp}(P_1, G_T)}$$

# Support of this rule

# Support of LHS
GTAR Discovery

Informative GTARs

- Interested in GTARs with high support and confidence
- Maximal GTARs with size bound to be more informative
- In a $b$-maximal GTAR, both LHS and RHS have at most $b$ edges.

The Discovery Problem

- **Input:** Temporal graph $G_T$, focus $\tilde{u}$, time interval $\Delta t$, size bound $b$, support threshold $\sigma$, and confidence threshold $\theta$;
- **Output:** The set of $b$-maximal GTARs $\Sigma$ pertaining to $\tilde{u}$ and $\Delta t$ such that for each GTAR $\phi \in \Sigma$, $\text{Supp}(\phi, G_T) \geq \sigma$, and $\text{Conf}(\phi, G_T) \geq \theta$. 
GTAR Discovery

- Integrate event mining and rule discovery as a single process.
- Intuition:

\[ \text{Conf}(\varphi, G_T) = \frac{\text{Supp}(\varphi, G_T)}{\text{Supp}(P_1, G_T)} \]

- Rule with high support
- LHS with low support

- LHS generation by best-first strategy.
  - Generate and verify best new LHS events.
- RHS generation given fixed LHS
  - To generate and validate new GTAR candidates by appending best RHS events to verified LHS events.
  - It prefers RHS events with high support.
GTAR Discovery

1. Event spawning

2. Event verification

3. Rule spawning

4. Rule validation

queue $L$

$P'_1(\bar{u})$

user $u$ retweet $\bar{u}$

user $\bar{u}$

user $\bar{u}$

user $\bar{u}$

check-in

check-in

$\bar{u}$ POI

$\bar{u}$ POI

queue $R$

$P_2$...

$P_2$...

$P_2$...

$P_7$

backtracking

show
Performance analysis and optimization

- **Complexity:**
  - Time: $O(|T|N(b)(b+|V|)(b+|E|)+N(b)^2|T|)$
  - Space: $O(N(b)|C(\bar{u})||T|)$
  - Size bound $b$ is small in practice and
  - Number of events $N(b)$ is significantly reduced by pruning rules

- **Optimization**
  - Pruning rules: extend (conditional) anti-monotonicity to GTARs
  - Anytime performance: returning GTARs as the events are discovered
  - Batch matching: merge snapshots to a graph and perform one matching
Experimental Study

Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Nodes</th>
<th>#Edges</th>
<th>#Labels</th>
<th>#Snapshots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation</td>
<td>4.3M</td>
<td>21.7M</td>
<td>273</td>
<td>80</td>
</tr>
<tr>
<td>Panama</td>
<td>839k</td>
<td>3.6M</td>
<td>433</td>
<td>12k</td>
</tr>
<tr>
<td>Movielens</td>
<td>81.5k</td>
<td>10M</td>
<td>21</td>
<td>1439</td>
</tr>
</tbody>
</table>

Algorithms

- **DisGTAR**: our integrated algorithms including all pruning rules
- **DisGTARn**: without the pruning strategies. (Pruning)
- **IsoGTAR**: isolating the snapshots and computes event matching over each snapshots one by one. (Batch matching)
- **SeqGTAR**: separating event mining and rule discovery to two independent processes. (Integrate mining)
## Performance of GTAR discovery

<table>
<thead>
<tr>
<th>Dataset</th>
<th>DisGTAR</th>
<th>DisGTARn</th>
<th>SeqGTAR</th>
<th>IsoGTAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time(s)</td>
<td># verif.</td>
<td>Time(s)</td>
<td># verif.</td>
</tr>
<tr>
<td>Panama</td>
<td>9</td>
<td>1,194</td>
<td>276</td>
<td>8,393</td>
</tr>
<tr>
<td>Citation</td>
<td>22</td>
<td>157</td>
<td>994</td>
<td>12,507</td>
</tr>
<tr>
<td>MovieLens</td>
<td>558</td>
<td>191</td>
<td>2,432</td>
<td>1,423</td>
</tr>
</tbody>
</table>

DisGTAR outperforms DisGTARn, SeqGTAR, and IsoGTAR by 6.28, 7.85 and 64.79 times on average.
Anytime performance

Time vs. Accuracy (Citation)

Time vs. Accuracy (Panama)

anytime quality(t) = \[ \frac{\sum_{\varphi \in \Sigma} Conf(\varphi, G_T)}{\sum_{\varphi \in \Sigma^*} Conf(\varphi, G_T)} \]

DisGTAR converges with high quality GTARs much faster than SeqGTAR
Scalability of DisGTAR

- DisGTAR is less sensitive to $|G|$.
- DisGTAR is much less sensitive than IsoGTAR.

Pruning rules

The “packing” of consecutive timestamps to time intervals.
Case Study

Matches: Prof. Christopher Manning (Stanford Univ.)
Conclusion and future work

Conclusion

- We have proposed a class of temporal association rules over graphs
- We have studied the discovery problem of GTARs
- Despite the enhanced expressive power of GTARs, it is feasible to find and apply GTARs in practice.

Future work

- Extending GTARs to multi-focus and exploring other quality metrics
- Fast online discovery of GTARs over graph streams.

Sponsored By:
Thank you!

Related Work

- Event Pattern Discovery by Keywords in Graph Streams (BigData’17)
- BEAMS: Bounded Event Detection in Graph Streams (ICDE’16) ([http://eecs.wsu.edu/~ksasani/BEAMS/Display.php](http://eecs.wsu.edu/~ksasani/BEAMS/Display.php))