# Discovering Graph Temporal Association Rules

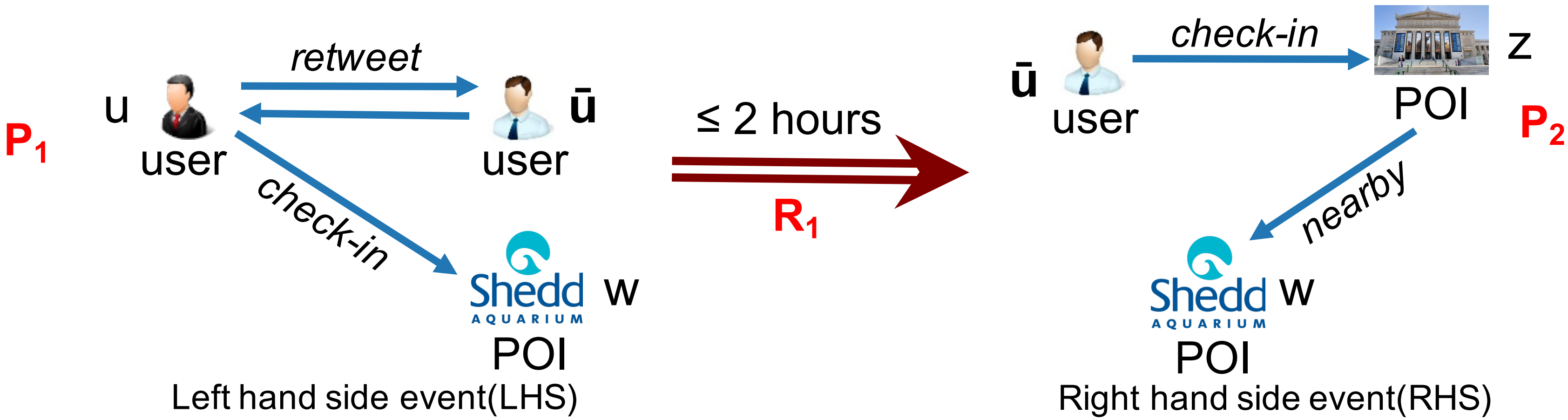
Mohammad Hossein Namaki Yinghui Wu **Qi Song**

Peng Lin Tingjian Ge*

Washington State University, *UMass Lowell

# Temporal association rules in networks

- Time-aware POI recommendation




- Shedd Aquarium in $P_2$ can be recommended as a point of interest for ū
- ū is a potential customer for Shedd Aquarium

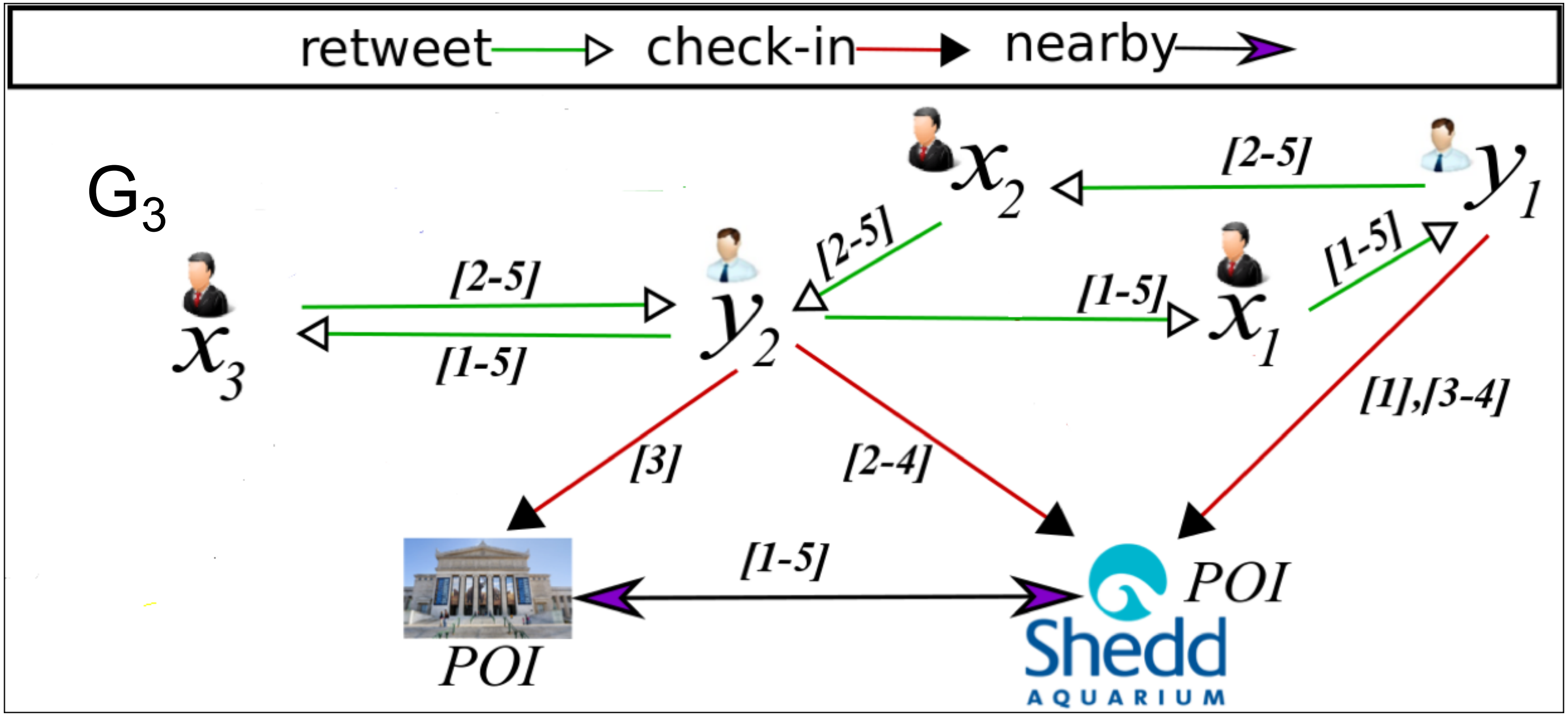
**Requirement: AR's with *topological*, *semantic* and *temporal* constraints**

# 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

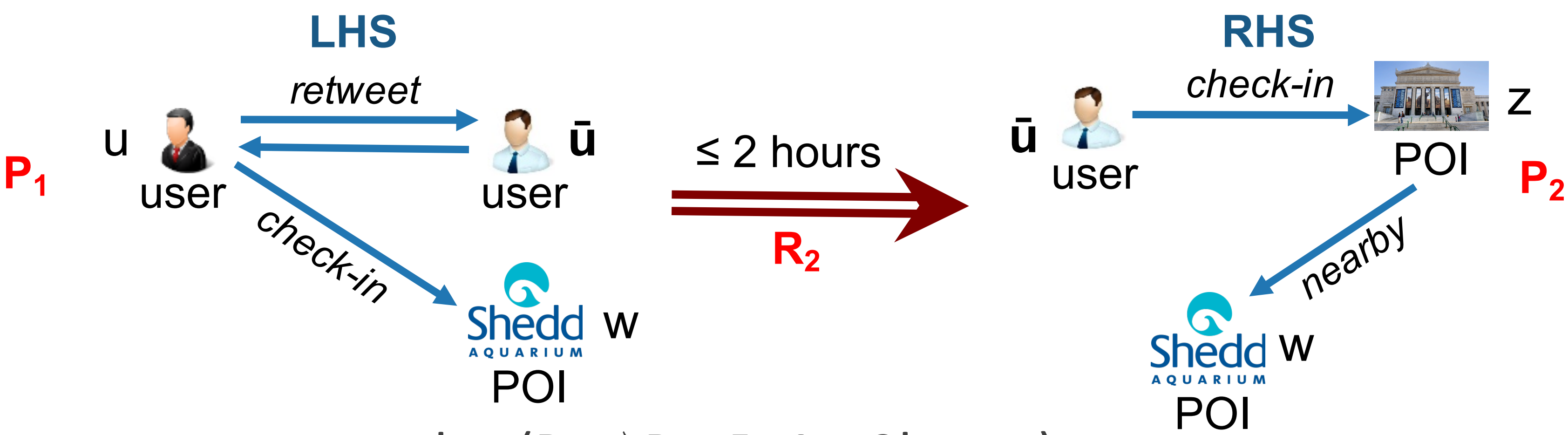
- Temporal graph $G_T(V,E,L,T)$.
- Snapshot $G_t$: induced by the set of all edges associated with time stamp t.



# Graph temporal association rules (GTAR)

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

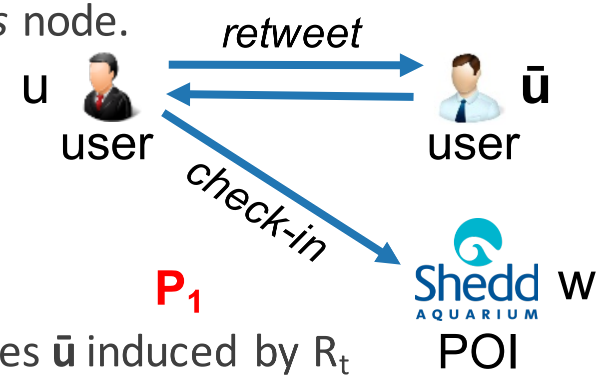
If there exists an occurrence of event $P_1$ at an entity specified by $\mathbf{\bar{u}}$ at some time t, then it is likely that an event $P_2$ occurs at the same entity, within a time window $[t, t + \Delta t]$



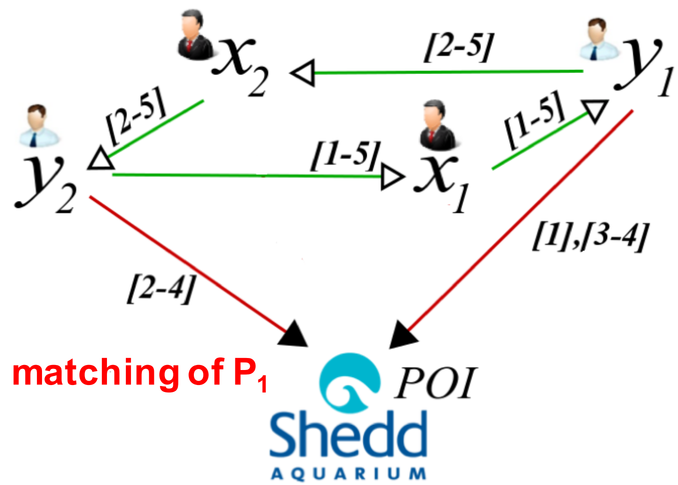

$\phi = (P_1 \Rightarrow P_2, \bar{u}, \Delta t\text{=2hours})$

# Events and Matching

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



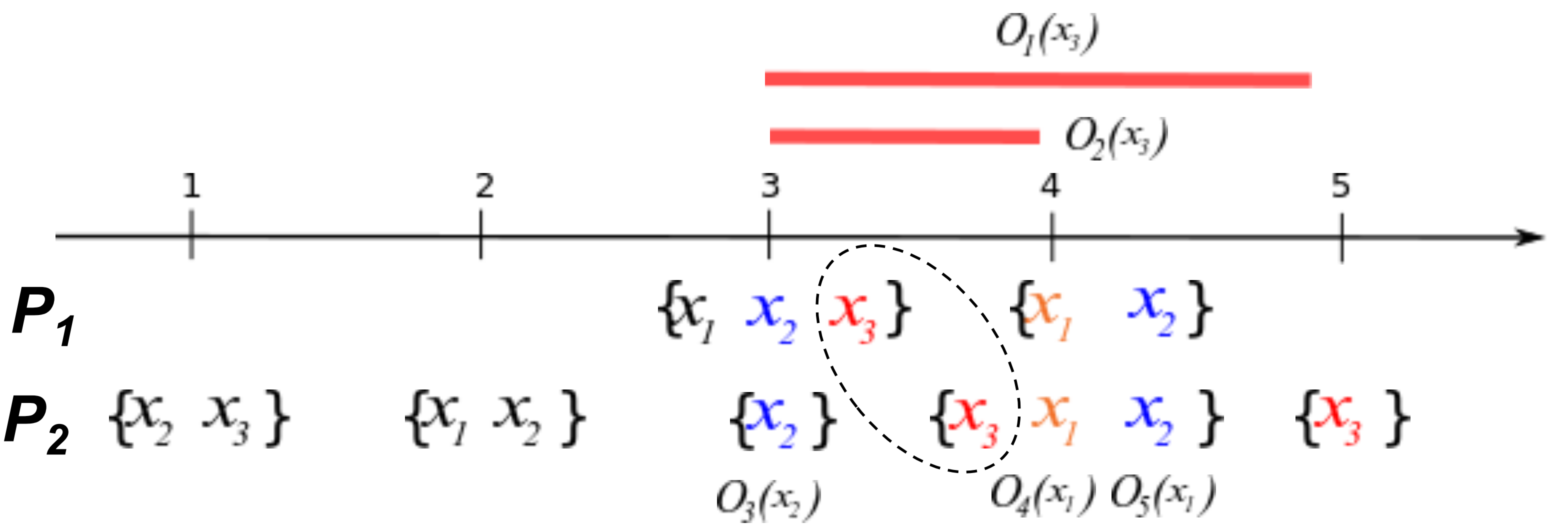

- **Event matching**
  - An event **P** occurred in $\mathbf{G_T}$ at time ***t*** if there is a *matching relation* $(R_t)$ between **P** and snapshot $\mathbf{G_t}$
  - *focus occurrence* $o(P, \bar{u}, t)$: the nodes in V that matches $\bar{u}$ induced by $R_t$
  - Example:
    - Matches of $\bar{u}$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




**One subgraph matching of $P_1$**

# 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 $\mathbf{G_T}$ supported by node v
  - There exists no $O'(v) \subset O(v)$, such that $O'(v)$ is also an occurrence



# Support and Confidence

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

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

# Occurrence of this rule

Normalizer

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

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

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

# GTAR Discovery

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

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

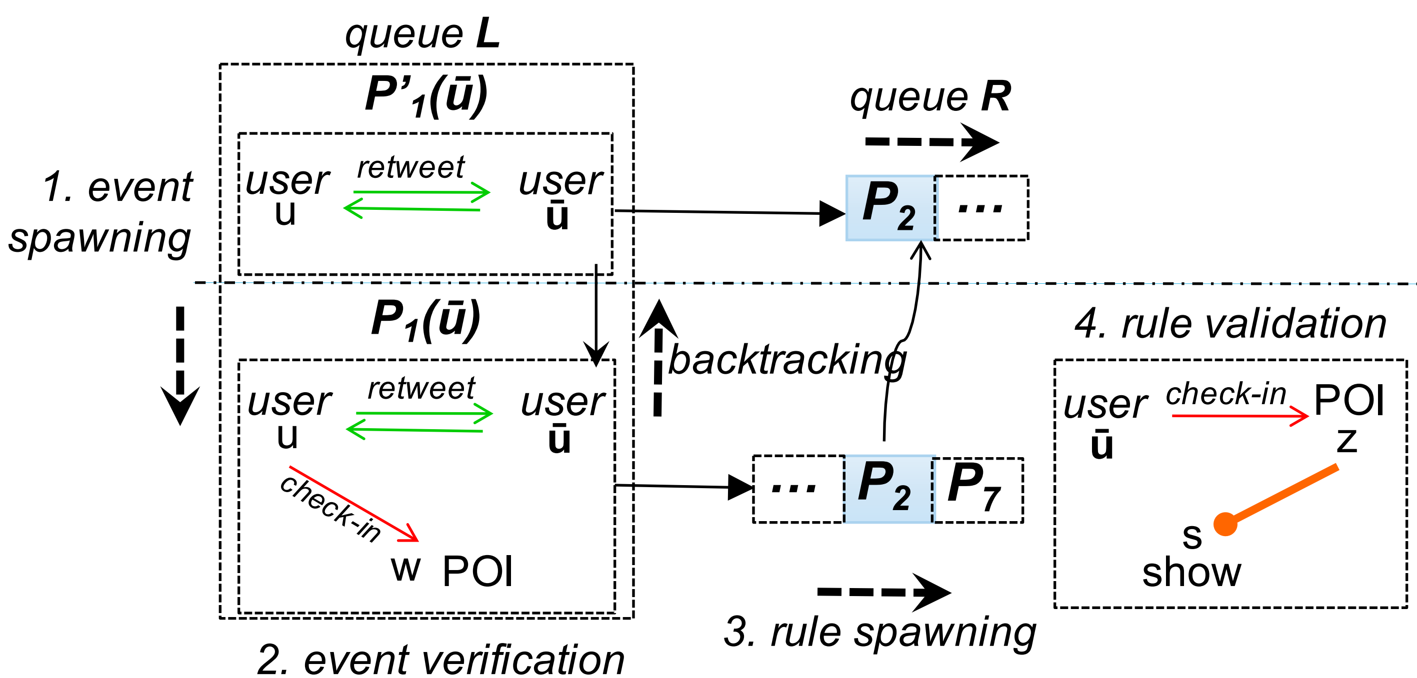
**Rule with high support**

**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

- GTAR discovery:



# 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

| | #Nodes | #Edges | #Labels | #Snapshots |
|---|---|---|---|---|
| Citation | 4.3M | 21.7M | 273 | 80 |
| Panama | 839k | 3.6M | 433 | 12k |
| Movielens | 81.5k | 10M | 21 | 1439 |

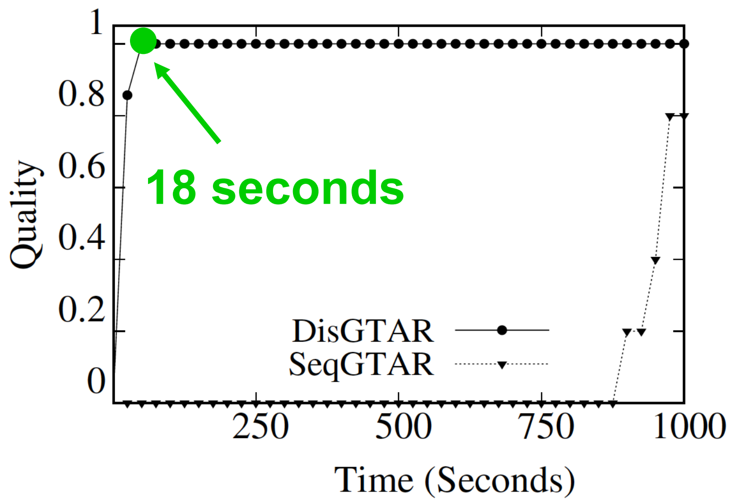
- Algorithms
  - **DisGTAR**: our integrated algorithms including all pruning rules
  - **DisGTARn**: without the pruning strategies. (**Pruning**)
  - **IsoGTAR**: isolating the snapshots and computes event matching over each snapshots one by one. (**Batch matching**)
  - **SeqGTAR:** separating event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

# Performance of GTAR discovery

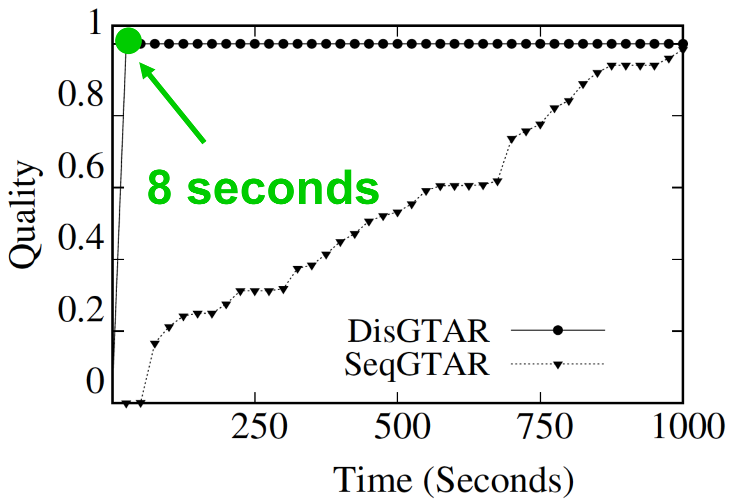
| | DisGTAR | | DisGTARn | | SeqGTAR | | IsoGTAR | |
|---|---|---|---|---|---|---|---|---|
| | Time(s) | # verif. | Time(s) | # verif. | Time(s) | # verif. | Time(s) | # verif. |
| Panama | **9** | 1,194 | 276 | 8,393 | 560 | 8,393 | N/A | |
| Citation | **22** | 157 | 994 | 12,507 | 1,621 | 12,507 | 12,721 | 11,461 |
| MovieLens | **558** | 191 | 2,432 | 1,423 | 2,445 | 1,423 | N/A | |

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

# Anytime performance




Time vs. Accuracy (Citation)

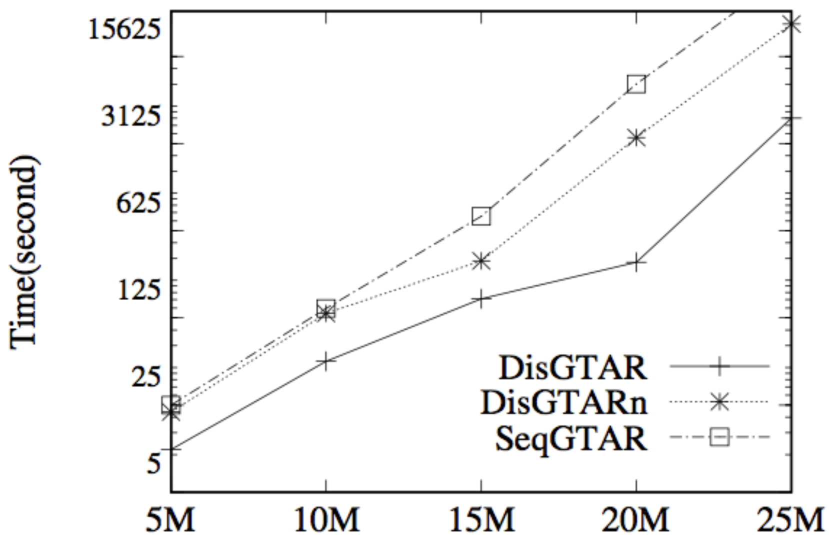



Time vs. Accuracy (Panama)

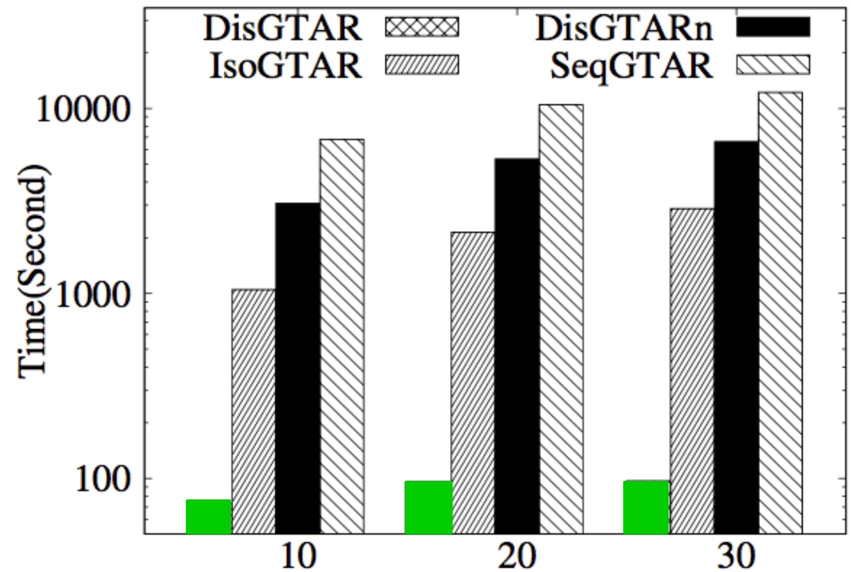
$$\text{anytime quality(t)} = \frac{\sum_{\varphi \in \Sigma^t} Conf(\varphi, G_T)}{\sum_{\varphi \in \Sigma^*} Conf(\varphi, G_T)}$$

DisGTAR converges with high quality GTARs much faster than SeqGTAR

# Scalability of DisGTAR




Varying |G|， # of edges (Synthetic)




Varying |T| (Synthetic)

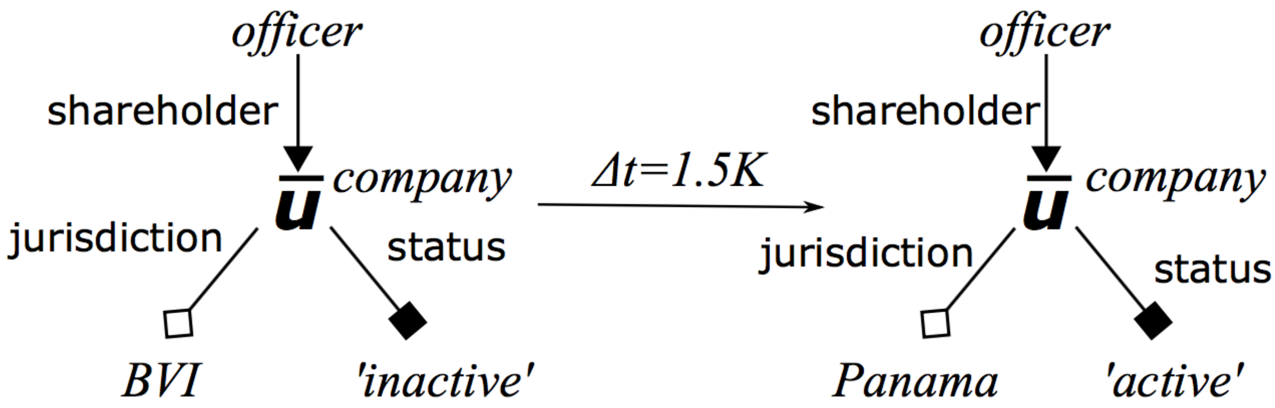
**DisGTAR** is less sensitive to |G|

Pruning rules

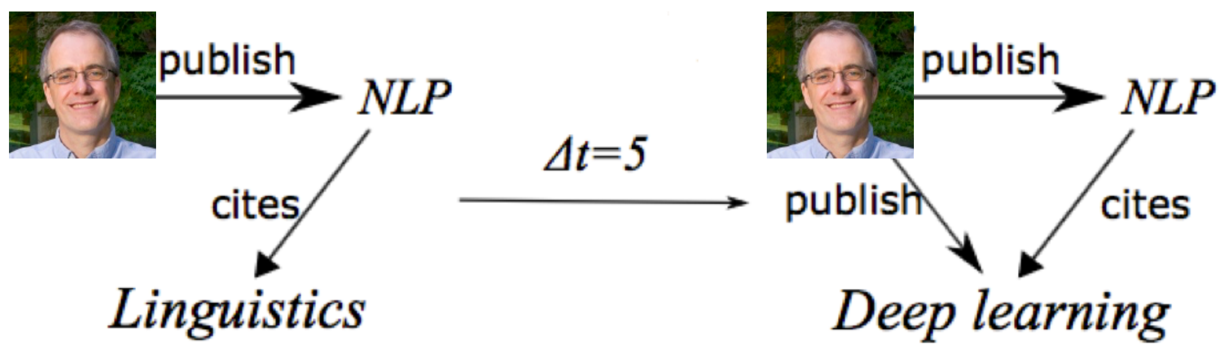
***DisGTAR*** is much less sensitive than ***IsoGTAR***

The “packing” of consecutive timestamps to time intervals

# Case Study




Matches: F.Geneve Project Management




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: NSF Google Faculty Research Awards

# 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 )