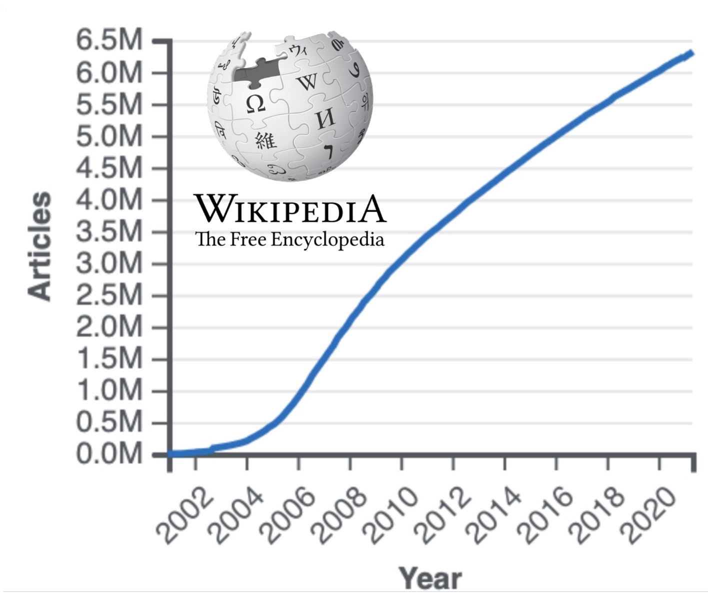
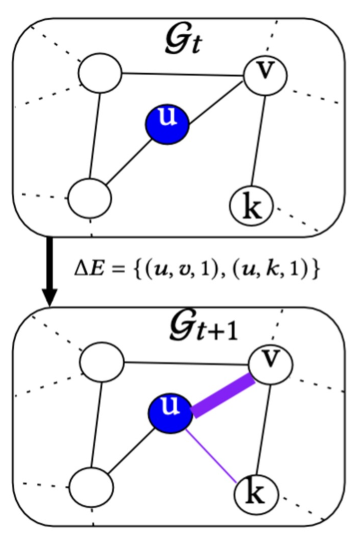
Subset Node Anomaly Tracking over Large Dynamic Graphs

Xingzhi Guo[1], Baojian Zhou[2], Steven Skiena[1]
xingzguo@cs.stonybrook.edu bjzhou@fudan.edu.cn skiena@cs.stonybrook.edu
[1] Computer Science Department, Stony Brook University, New York, USA
[2] School of Data Science, Fudan University, Shanghai, China
Stony Brook University Computer Science
FUDAN UNIVERSITY

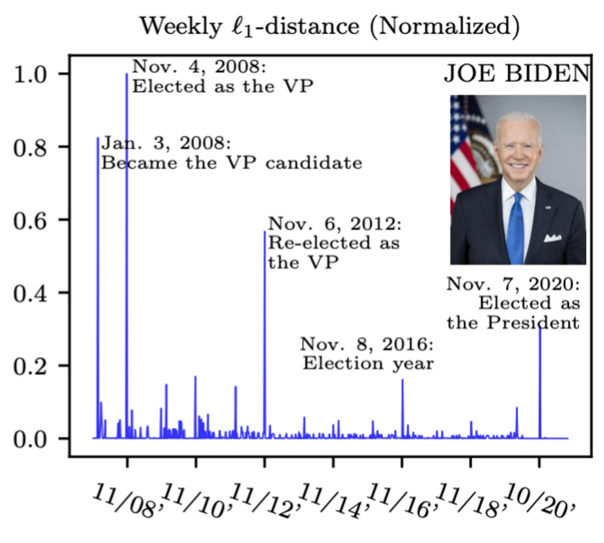
Introduction:

- Tracking a targeted subset of nodes in an evolving large graph is important for many real-world applications.
  - ❖ Node-level anomaly: Identify when specific nodes significantly changed their context.
  - ❖ Graph-level anomaly: Uncover when global changes occurred in the overall graph structure.
- These tasks become challenging as the graph size and time horizon scale up.
  - ❖ It's expensive to re-train whole node embeddings across time.
  - ❖ It's difficult to detect node anomaly in mis-aligned embeddings.





(a) Dynamic Network



(b) Application: Quantify changes of Biden

- We propose a unified framework ***DynAnom*** for subset node anomaly tracking over large dynamic graphs:
  - ❖ Efficient Personalized PageRank (PPR)-based node embeddings for a subset of nodes over dynamic graphs.
  - ❖ Flexible framework to localize both node/graph anomalies with customizable anomaly score functions for various applications.
  - ❖ Compared to others, ***DynAnom*** is 2x better and 2.3x faster.

Key Contributions

> We proposed an efficient anomaly tracking framework, ***DynAnom***, that supports node/graph-level anomaly tracking over dynamic graph.
> We generalized dynamic forward push algorithm for calculating PPR, making it suitable for weighted-graph anomaly tracking.
> Released Resources @ https://github.com/zjlxgxz/DynAnom.
- We derived a real-world graph, PERSON, for personal life changes detection.
- We released our python codebase for reproducibility.

Related Work

- Many graph-level anomaly detection cannot identify individual node changes but only uncover the global changes. Edge-level methods cannot identify node evolution when there is no direct edge event associated with the node.
- Many node embedding-based methods must calculate all embeddings even only for subset tracking. The embedding space is not well-aligned across time, making the algorithm only detects anomalies (outliers) in each static snapshot.

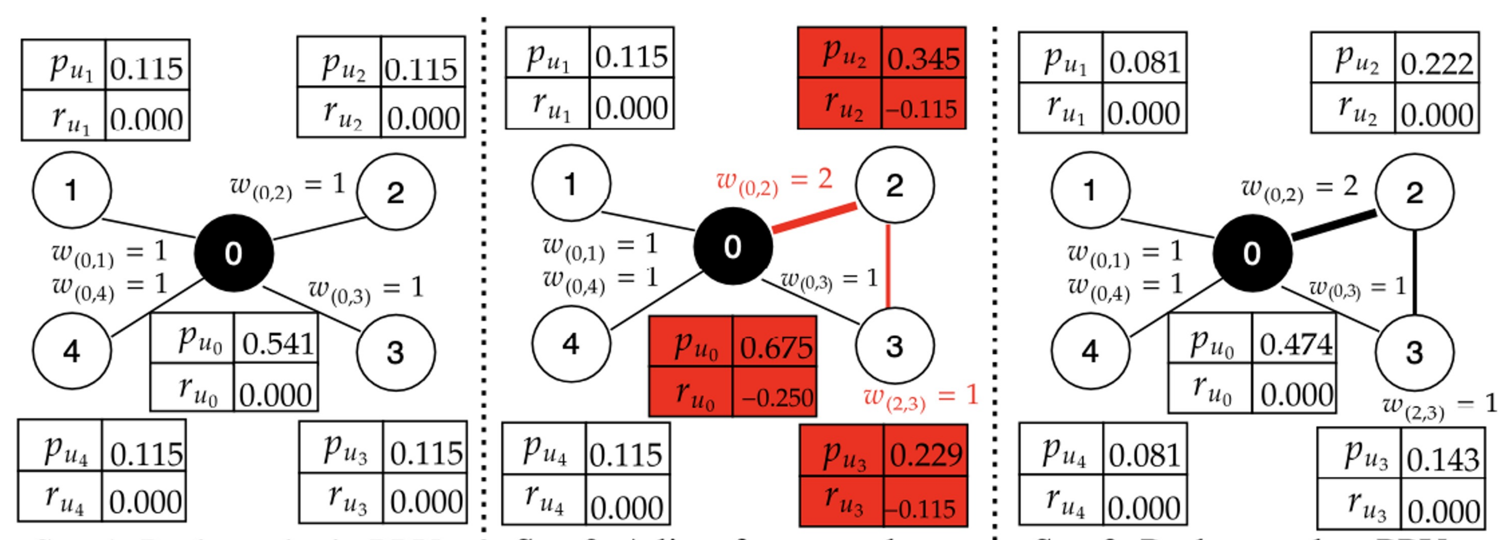
Feature / Method	Anomaly: Node level	Anomaly: Graph level	Edge Event Types: Edge Stream	Edge Event Types: Add/ Delete	Edge Event Types: Weight Adjust	Algorithm: $\lvert\mathcal{V}\rvert$ Ind.	Algorithm: Align Repr.	Algorithm: Flex. Score
SedanSpot	✗	✗	✔	✗	✗	✔	✗	✗
AnomRank	✗	✔	✗	✔	✔	✗	✗	✗
NetWalk	✔	✗	✗	✔	✗	✗	✗	✗
DynPPE	✔	✗	✔	✔	✗	✔	✔	✗
DynAnom	✔	✔	✔	✔	✔	✔	✔	✔

Proposed Algorithm: *DynAnom*

Step 1: Maintain PPR-Vector (PPV) for the target subset
- ❖ *Incremental Local Push* over dynamic weighted graphs
- ❖ Complexity is linear to #edge changes, not #nodes.

Step 2: Obtain PPR-based node embeddings in constant time
- ❖ Hash or Random Projection for Dimension Reduction

Step 3: Detect node/graph anomaly for the target subset
- ❖ Node anomaly: Compare embedding change (ℓ_1-distance).
- ❖ Graph anomaly: Aggregate node anomaly of high-degree nodes (max, median, mean).



Step1: Push to obtain PPV | Step2: Adjust for new edges | Step3: Push to update PPV

Experiments

- Node-level anomaly: Comparing to the best baseline our method ***DynAnom*** is effective (2x better) and scalable (2.3x faster) with the simplest anomaly score function.

Methods	Accuracy	Wall-time (s)
AnomRank	0.2790	905.98
AnomRank (W)	0.2652	905.98
NetWalk	OOM	OOM
DynPPE	0.1701	84.25
DynAnom	0.5425	379.33

	DARPA	PERSON
# Nodes	25.5K	609.9K
# Edge events	4.5M	8.7M
# Snapshots	1.4K	43

- Graph-level anomaly tracking: ***DynAnom*** is useful and flexible:
  - ❖ The simplest heuristic anomaly design outperforms others.
  - ❖ Easily adapt to most ML system for more sophisticated detection systems.

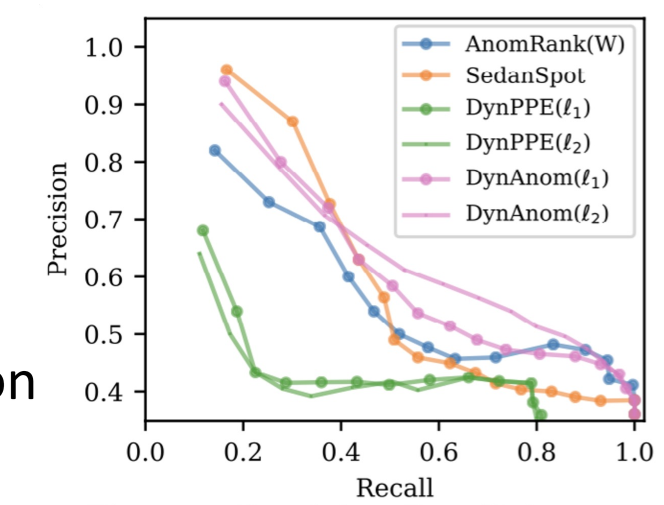
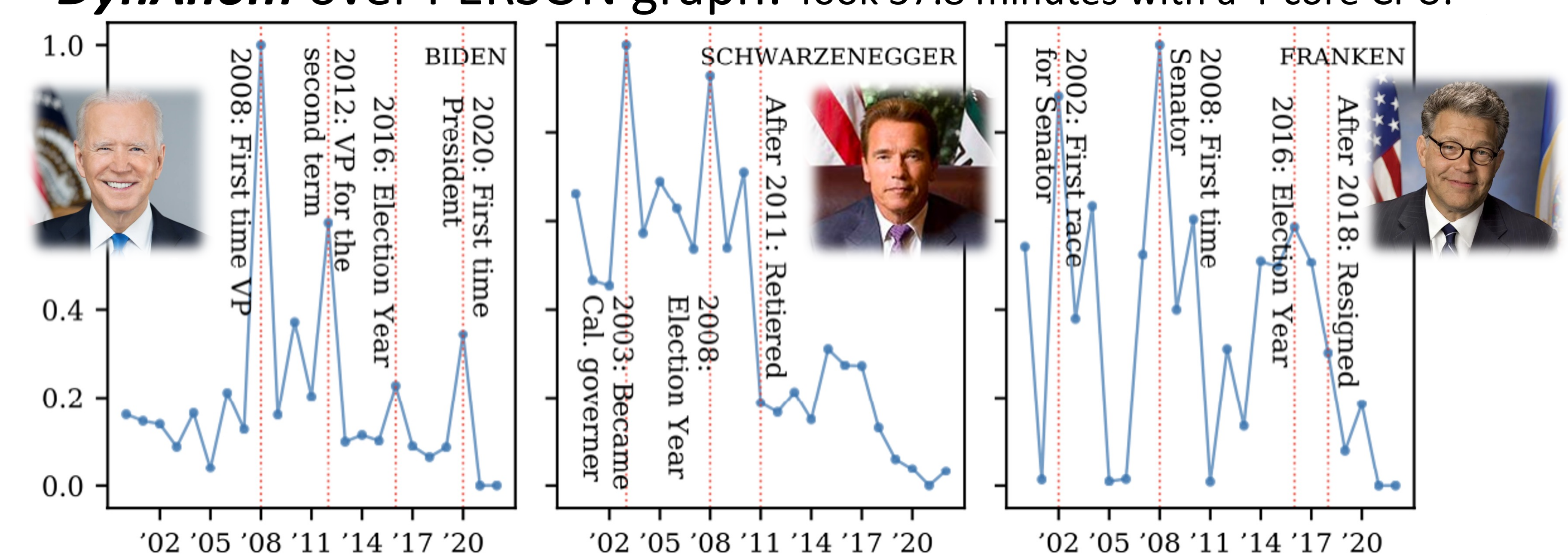


Figure 3: Precision-Recall Curve

Table 6: The precision of top-250 anomalies

Algorithms	Precision
NetWalk	OOM
AnomRankW	0.5400
SedanSpot	0.5640
DynPPE(ℓ_1)	0.4160
DynPPE(ℓ_2)	0.3920
DynAnom(ℓ_1)	0.5840
DynAnom(ℓ_2)	**0.6120**

- Tracking Real-world personal life changes over 22 years with ***DynAnom*** over PERSON graph: Took 37.8 minutes with a 4-core CPU.



Conclusions

- We propose a unified framework ***DynAnom*** for subset node anomaly tracking over large dynamic graphs.
- ***DynAnom*** can be easily applied to different graph anomaly detection tasks with a flexible score function.
- Experiments demonstrates its effectiveness and efficiency.

Xingzhi Guo, Baojian Zhou, and Steven Skiena.2022. Subset Node Anomaly Tracking over Large Dynamic Graphs.
In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Washington DC Convention Center August 14-18, 2022