PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction

Shichang Zhang[1], Jiani Zhang[2], Xiang Song[2], Soji Adeshina[2], Da Zheng[2], Christos Faloutsos[2,3], Yizhou Sun[1,2]

[1]University of California, Los Angeles (UCLA)
[2]Amazon [3]Carnegie Mellon University

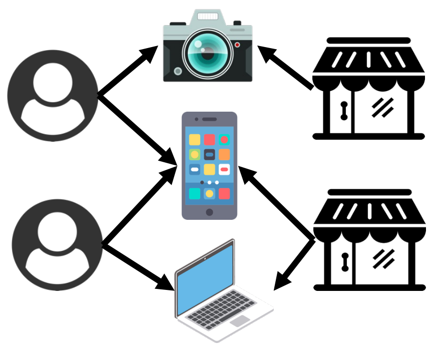
July 2023

Outline

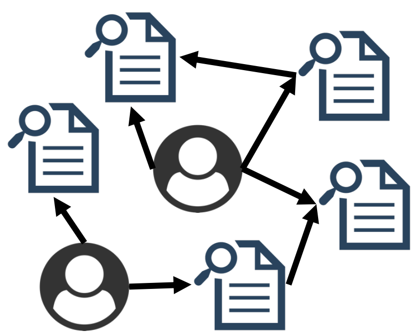
- Machine Learning on Graphs
  - Graph Neural Networks (GNNs)
  - GNNs for Link Prediction
- Model Explainability
- PaGE-Link
  - Main Idea
  - Path-Enforcing Mask
  - Experiments

Machine Learning on Graphs

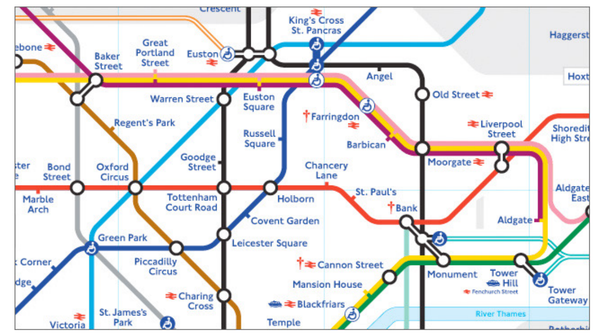
Graphs are a general language for modeling entities with relations



E-commerce graphs



Citation graphs



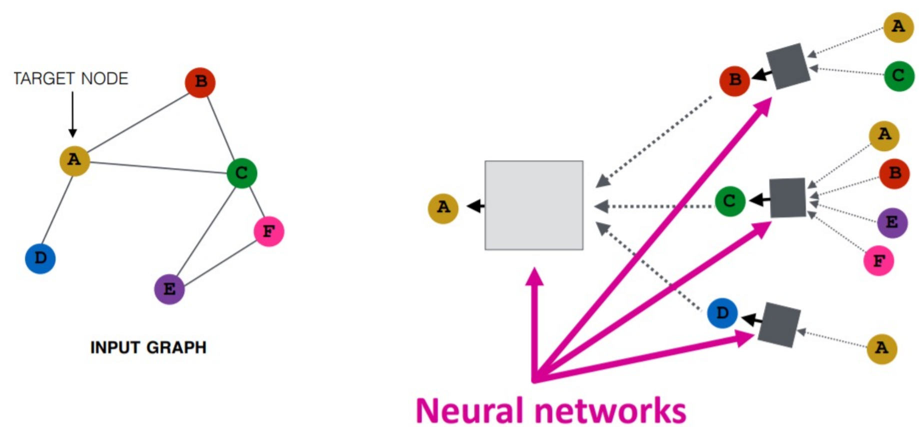
Transportation graphs

(Image Credit: www.visitlondon.com)

Molecule graphs, code graphs, scene graphs, and many more …

Graph Neural Networks

GNNs: a family of neural-network-based models for machine learning on graphs



(Hamilton, W. L., Ying, R., & Leskovec, J. 2017)

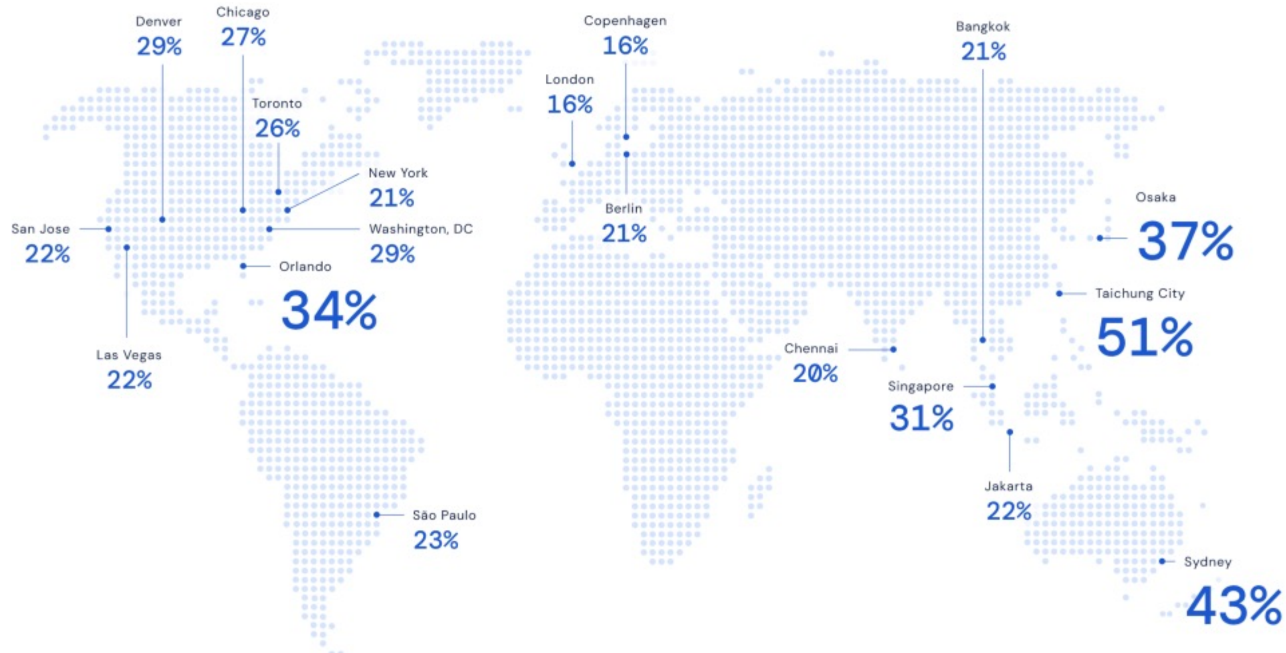
Message passing

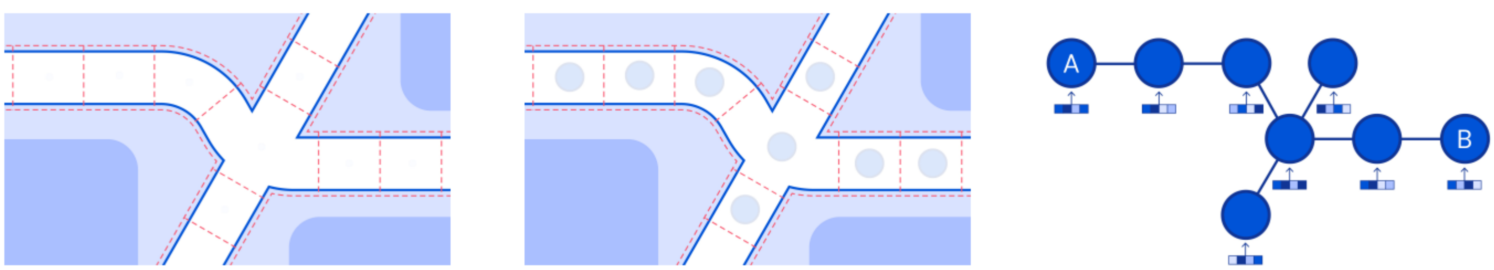
Each node aggregates messages from its neighbors and recursively extends to multi-hop neighbors

Node A with neighbors $\mathcal{N}(\mathrm{A})$ aggregates messages at step l:

$$\mathbf{h}_A^{(l)} = \textsc{aggr}(\mathbf{h}_A^{(l-1)}, \{\mathbf{h}_i^{(l-1)} | i \in \mathcal{N}(A)\})$$

GNNs for Google Map ETA Prediction

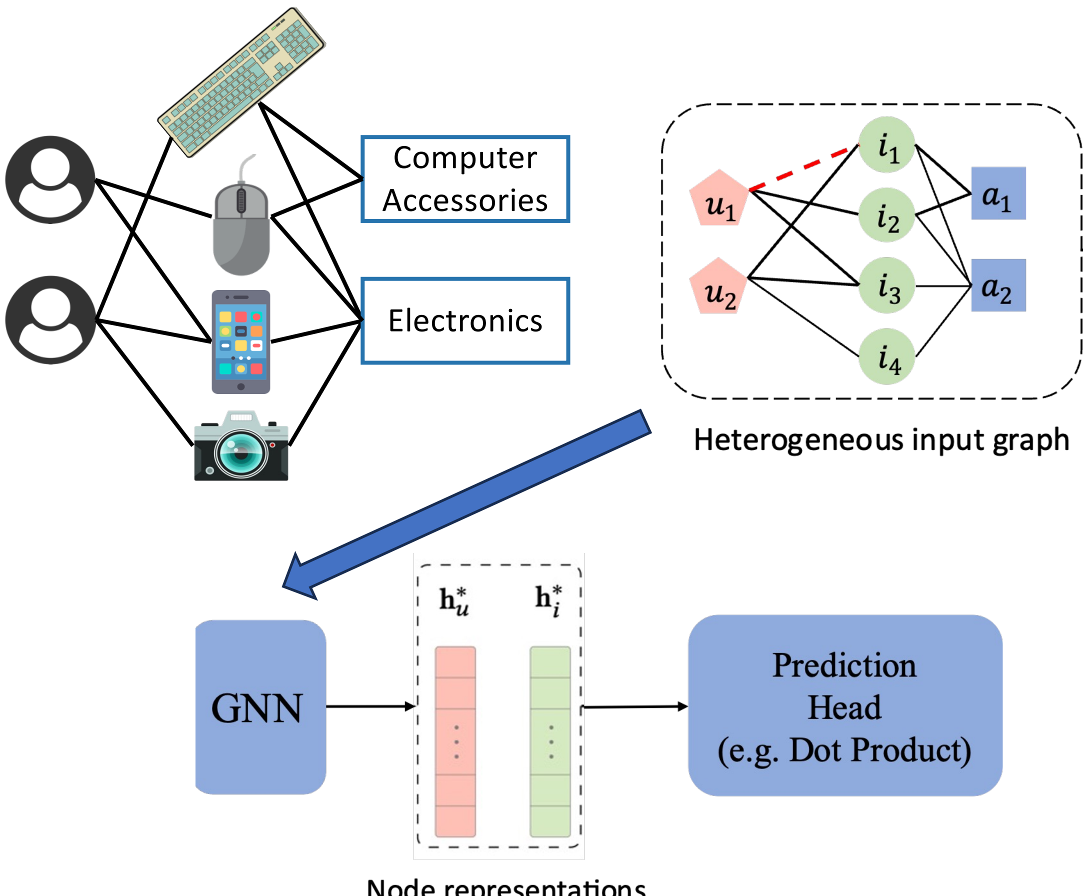




Derrow-Pinion, Austin, et al. "ETA prediction with graph neural networks in google maps." CIKM. 2021.

GNNs for Link Prediction

- Link Prediction: Recommend items to users



Traditional models (matrix factorization & shallow embedding)

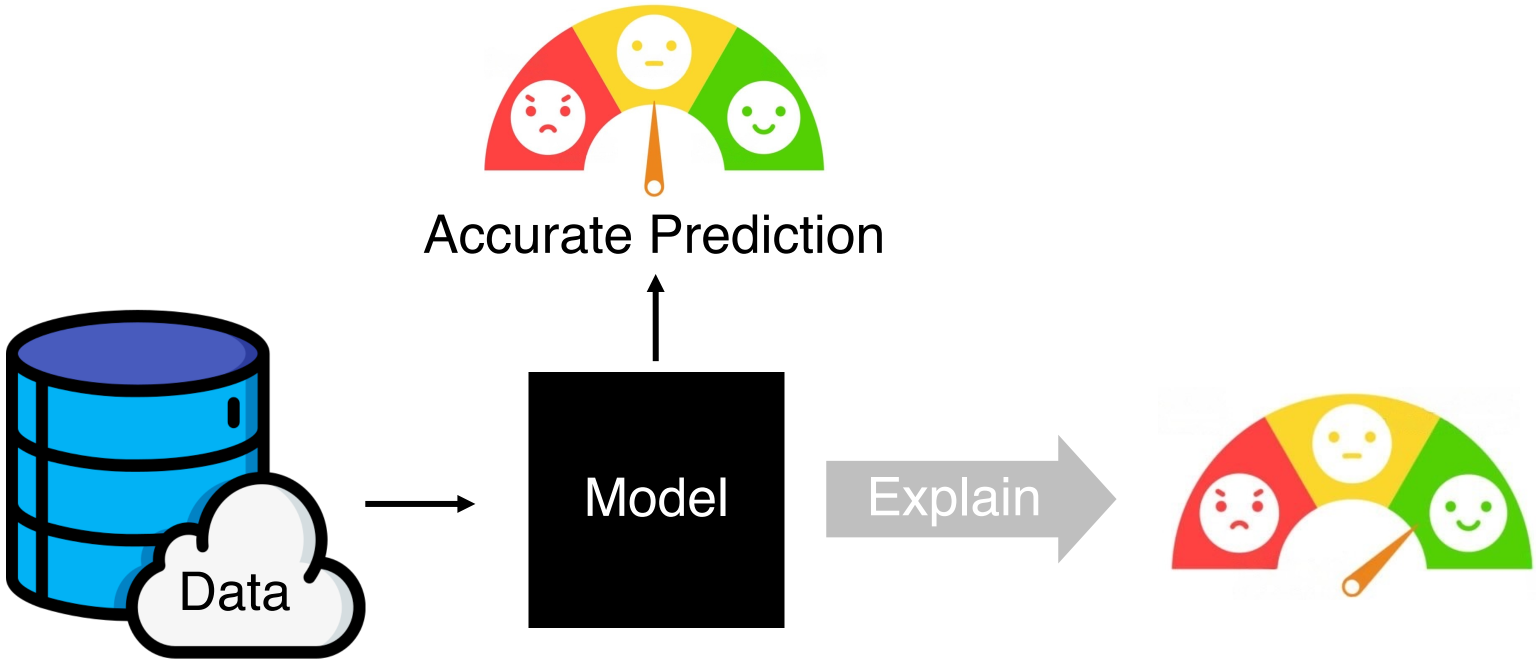
GNNs

Model	Amazon-Books	
	Recall@20	NDCG@20
MF-BPR	0.0338	0.0261
CML	0.0522	0.0428
ENMF	0.0359	0.0281
DeepWalk	0.0346	0.0264
LINE	0.0410	0.0318
Node2Vec	0.0402	0.0309
NGCF	0.0344	0.0263
NIA-GCN	0.0369	0.0287
LR-GCCF	0.0335	0.0265
LightGCN	0.0411	0.0315
DGCF	0.0422	0.0324
UltraGCN_{Base}	0.0504	0.0393
UltraGCN	**0.0681**	**0.0556**

GNN achieve SOTA link prediction/recommendation results (Mao et al. CIKM 2021).

Model Explainability

- Many start-of-the-art AI models are black boxes.
- Explainability helps to increase user satisfaction and improve model design.

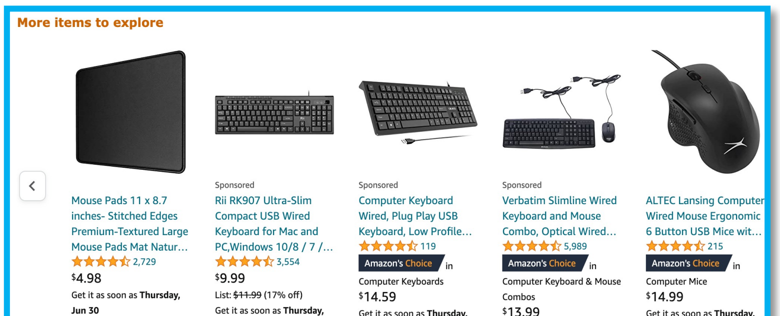


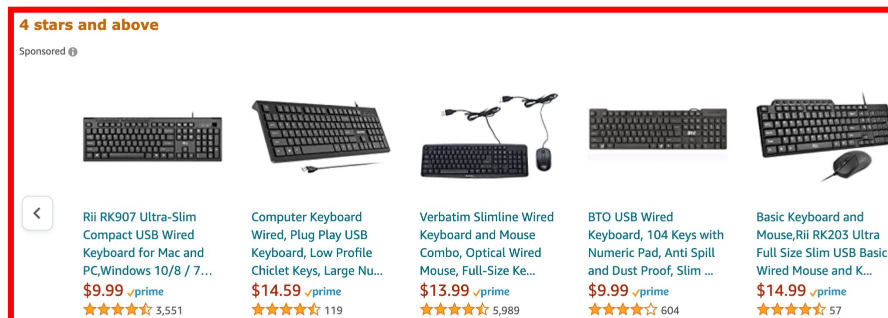
A Real Amazon Recommendation

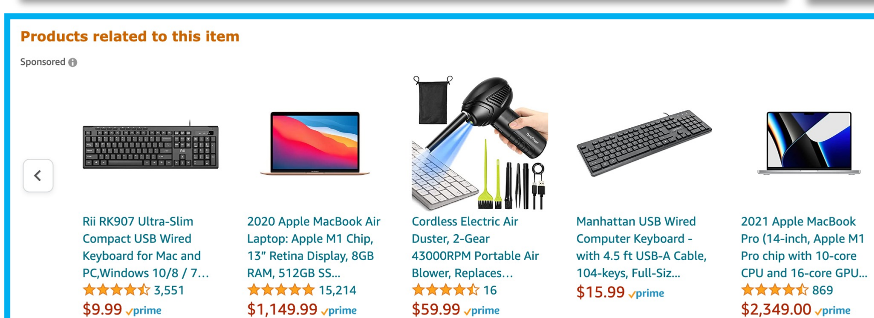
Query

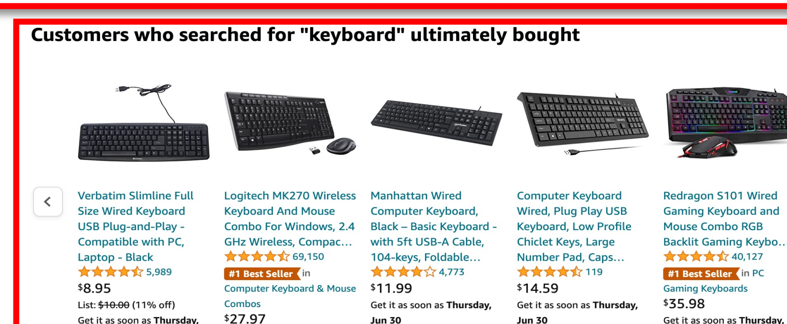
Recommendation without explanation

Recommendation with explanation



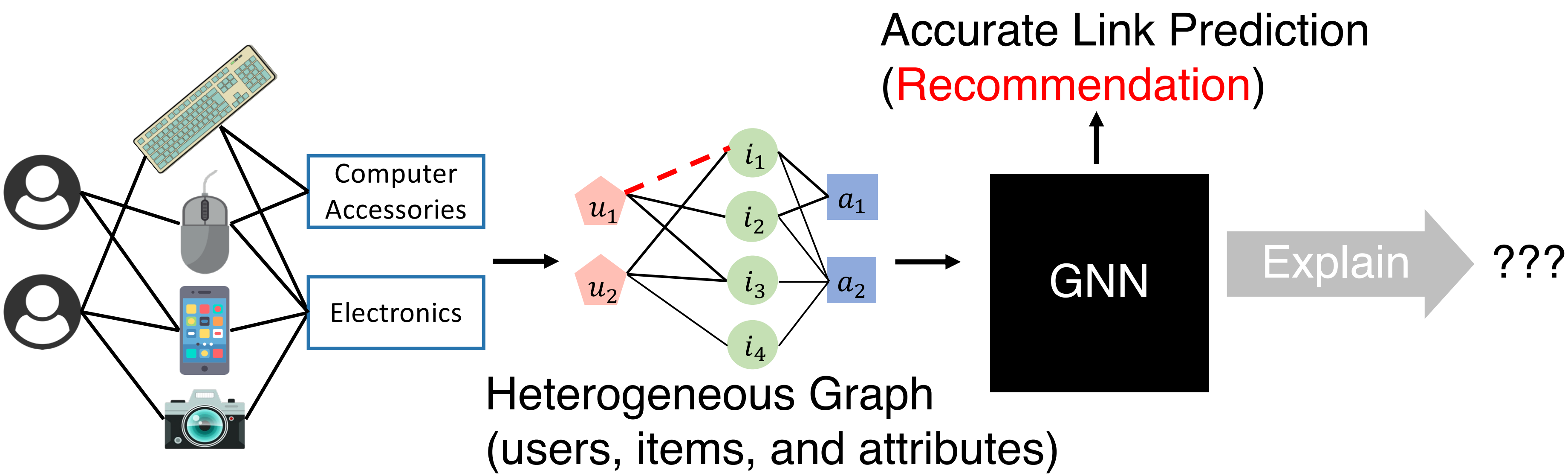






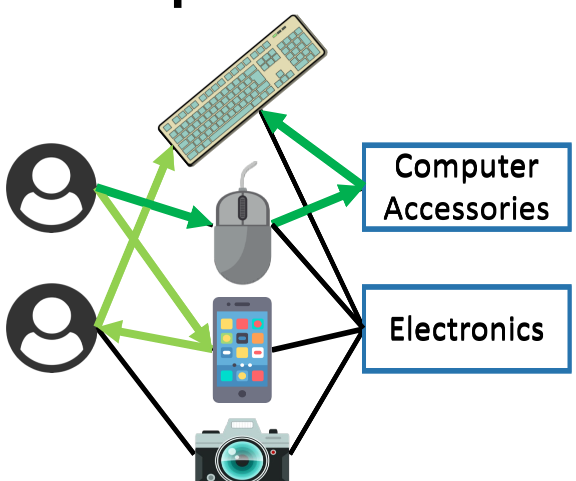
GNN Explainability

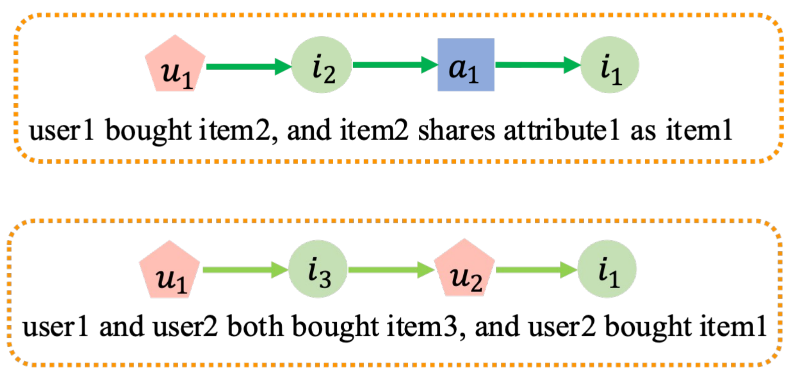
- Data: Heterogeneous graphs.
- Model: GNNs for link prediction.
- Explanation: Why recommend an item to a user? (Why predict a user-item link?)

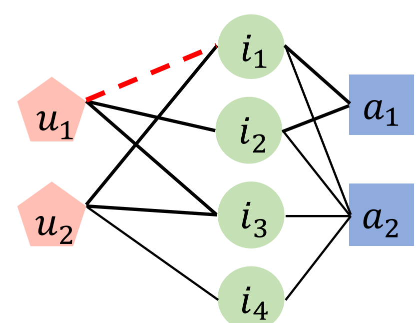


Main Idea: Paths As Explanations

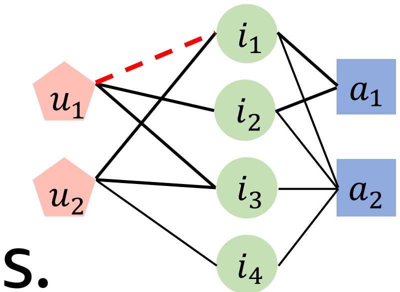
- Natural human-interpretable explanations boil down to paths.
- Paths form a much smaller search space compared to general subgraphs.
- Define explanations as *concise* and *informative* paths that are *influential to the prediction*.







PaGE-Link: Path-Enforcing Mask



- Challenges for finding good paths.
  - Many path candidates.
  - Criterion for selecting good paths.
- Learn an edge mask to select meaningful edges.
  - Edges form short paths with low-degree nodes.

$$\mathcal{L}_{path}(\mathcal{M}) = -\sum_{r\in\mathcal{R}}(\alpha \sum_{\substack{e\in\mathcal{E}_{path}\\ \tau(e)=r}} \mathcal{M}_e^r - \beta \sum_{\substack{e\in\mathcal{E}, e\notin\mathcal{E}_{path}\\ \tau(e)=r}} \mathcal{M}_e^r)$$

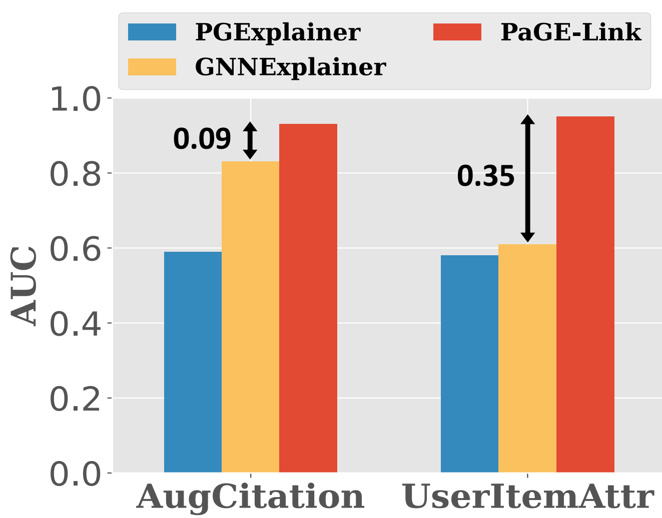
  - Edges maximize the mutual information.

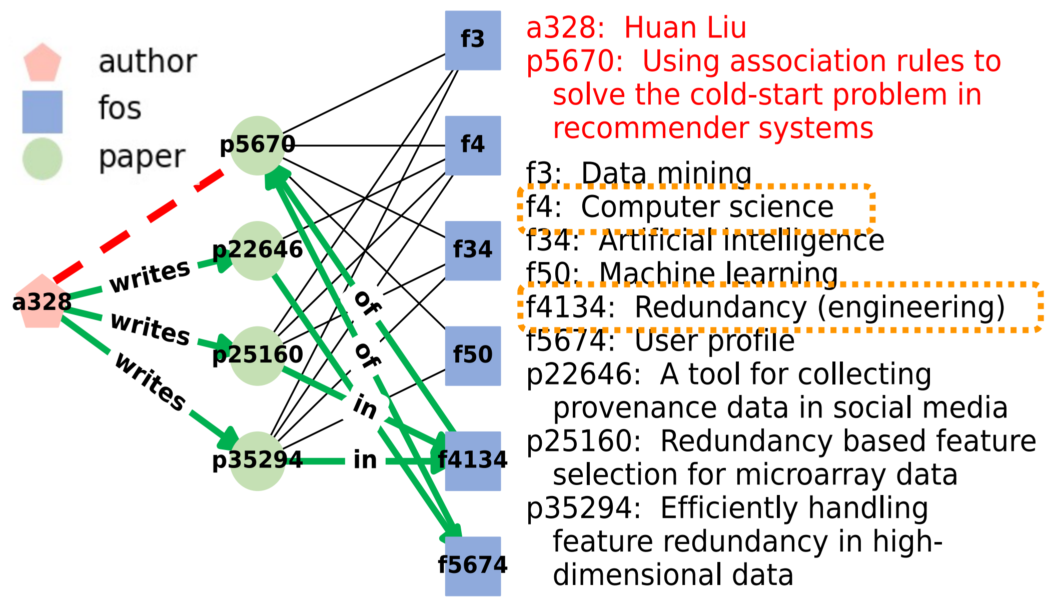
$$\mathcal{L}_{pred}(\mathcal{M}) = -\log P_\Phi(Y=1|\mathcal{G}=(\mathcal{V},\mathcal{E}\odot\sigma(\mathcal{M})),(s,t))$$

- Pruning: more informative paths and efficiency.

Experiments

- ROC-AUC: 9%-35% improvement over baselines.
- Concise paths without generic nodes.
- Human evaluation: 78.79% responses selected our method as the best compared to baselines.





Thank you!
Q & A

Paper **Code** **Contact author**

Appendix

Experiments: Dataset Generation

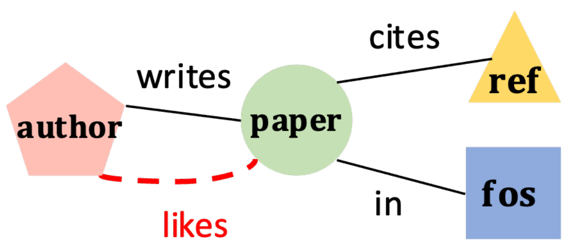
Generate datasets new evaluation

- Create a new edge s–t if they are connected by a concise and informative path p

$$\mathcal{P} = \{p|p \text{ is a } s\text{-}t \text{ path with max length } l_{max} \text{ and max node degree } D_{max}\}$$

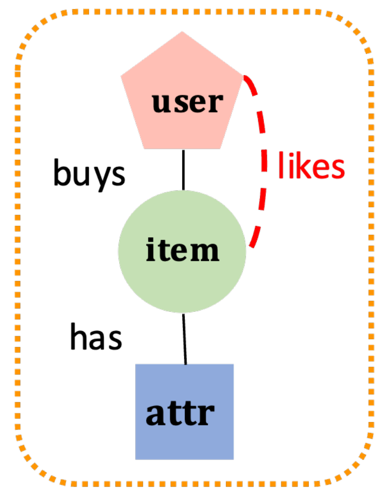
- Use p as the ground truth for evaluating the prediction of (s, t)

Citation graph



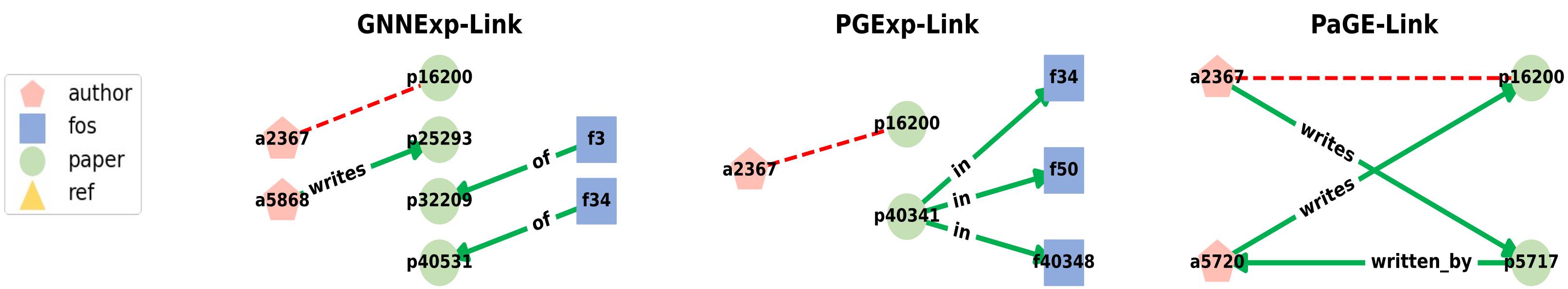
Augmented graph schema

User-item graph



Graph schema

Experiments: Visualization



a2367: Vipin Kumar
p16200: Fast and exact network trajectory similarity computation: a case-study on bicycle corridor planning

a5868: Pang-Ning Tan
f3: Data mining
f34: Artificial intelligence
p25293: Selecting the right interestingness measure for association patterns
p32209: Generalizing the notion of support
p40531: Tripoles: A New Class of Relationships in Time Series Data

f34: Artificial intelligence
f50: Machine learning
f40348: Environmental change
p40341: Incremental Dual-memory LSTM in Land Cover Prediction

a5720: Shashi Shekhar
p5717: Correlation analysis of spatial time series datasets: a filter-and-refine approach

Proposition and Theorems

- Paths form a much smaller search space

Proposition 4.1. *Let $\mathcal{G}(n,d)$ be a random graph with n nodes and density d, i.e., there are $m = d\binom{n}{2}$ edges chosen uniformly randomly from all node pairs. Let $Z_{n,d}$ be the expected number of paths between any pair of nodes. Let $S_{n,d}$ be the expected number of edge-induced subgraphs. Then $Z_{n,d} = o(S_{n,d})$, i.e.,* $\lim_{n\to\infty} \frac{Z_{n,d}}{S_{n,d}} = 0$.

- Asymptotic normality of the k-core

Theorem 5.1 (Pittel, Spencer and Wormald [28]). *Let $\mathcal{G}(n,d)$ be a random graph with m edges as in Proposition 4.1. Let $\mathcal{G}^k(n,d) = (\mathcal{V}^k(n,d), \mathcal{E}^k(n,d))$ be the nonempty k-core of $\mathcal{G}(n,d)$. Then $\mathcal{G}^k(n,d)$ will contain $\delta_{\mathcal{V}}(n,d,k)n$ nodes and $\delta_{\mathcal{E}}(n,d,k)m$ edges with high probability (w.h.p.) for large n, i.e., $|\mathcal{V}^k(n,d)|/n \xrightarrow{p} \delta_{\mathcal{V}}(n,d,k)$ and $|\mathcal{E}^k(n,d)|/m \xrightarrow{p} \delta_{\mathcal{E}}(n,d,k)$ with $\xrightarrow{p}$ stands for convergence in probability.*