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

April 2023

Outline

- Problem and Motivation
- Existing Work
- Main Idea
- Experiments

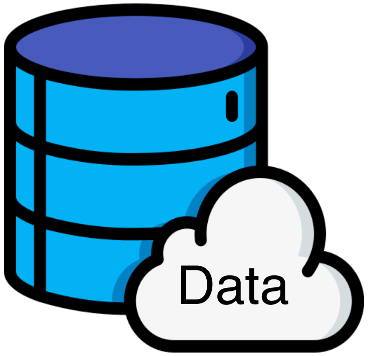
Problem: Model Explainability

Problem: Model Explainability

- Many start-of-the-art AI models are black boxes.

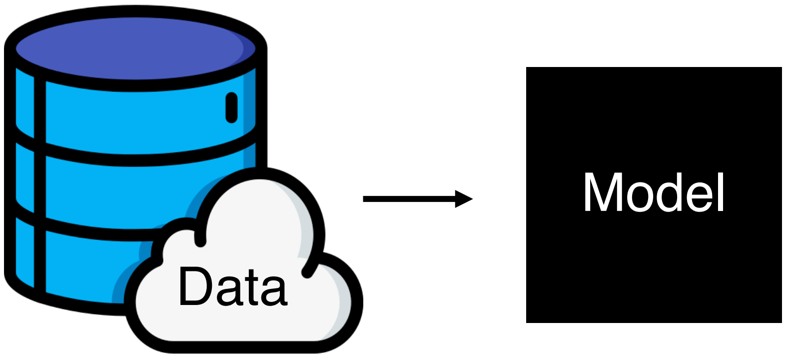
Problem: Model Explainability

- Many start-of-the-art AI models are black boxes.



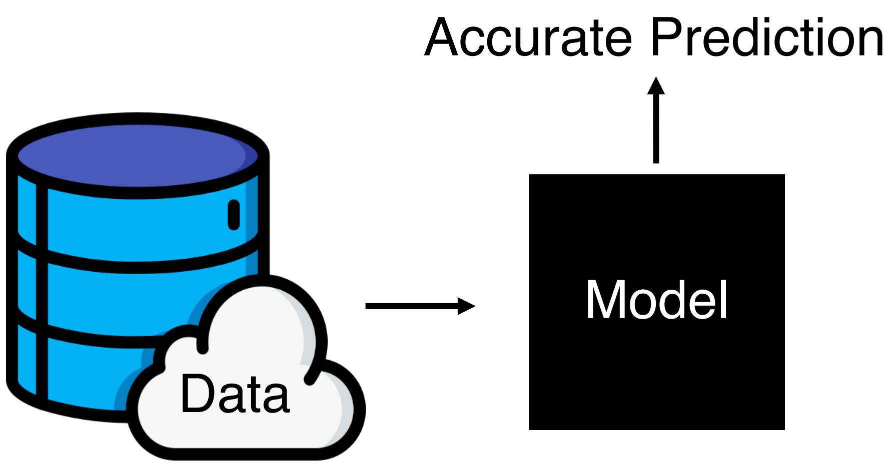
Problem: Model Explainability

- Many start-of-the-art AI models are black boxes.



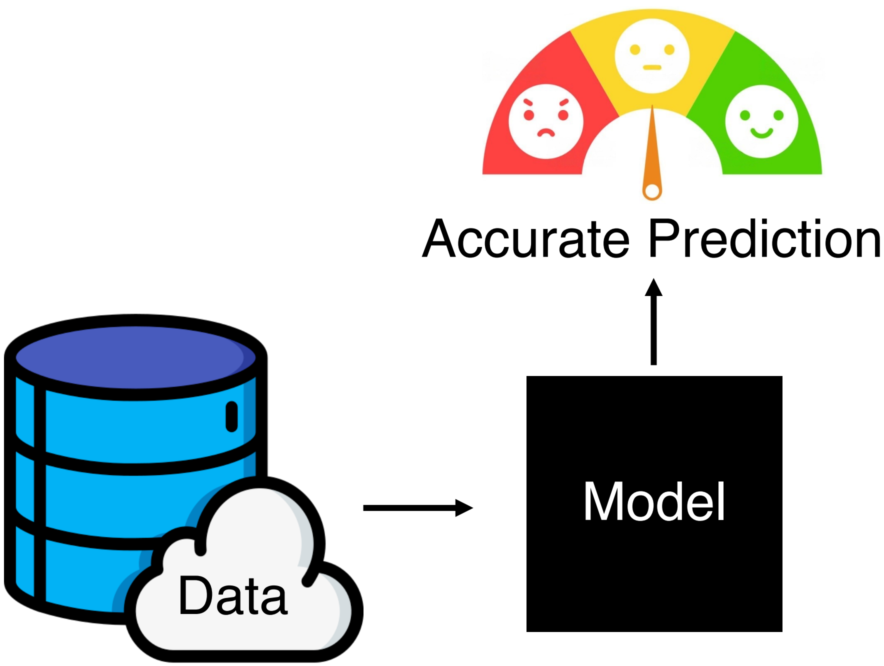
Problem: Model Explainability

- Many start-of-the-art AI models are black boxes.



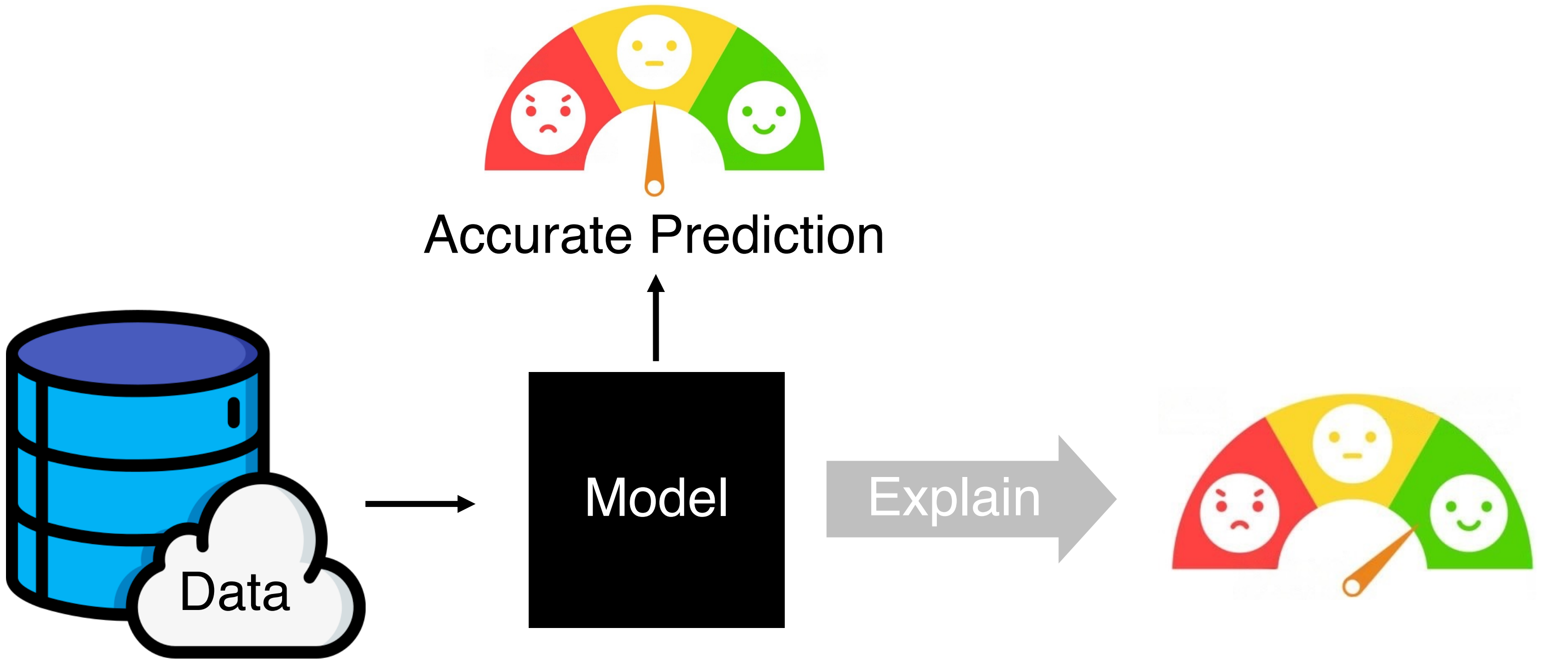
Problem: Model Explainability

- Many start-of-the-art AI models are black boxes.



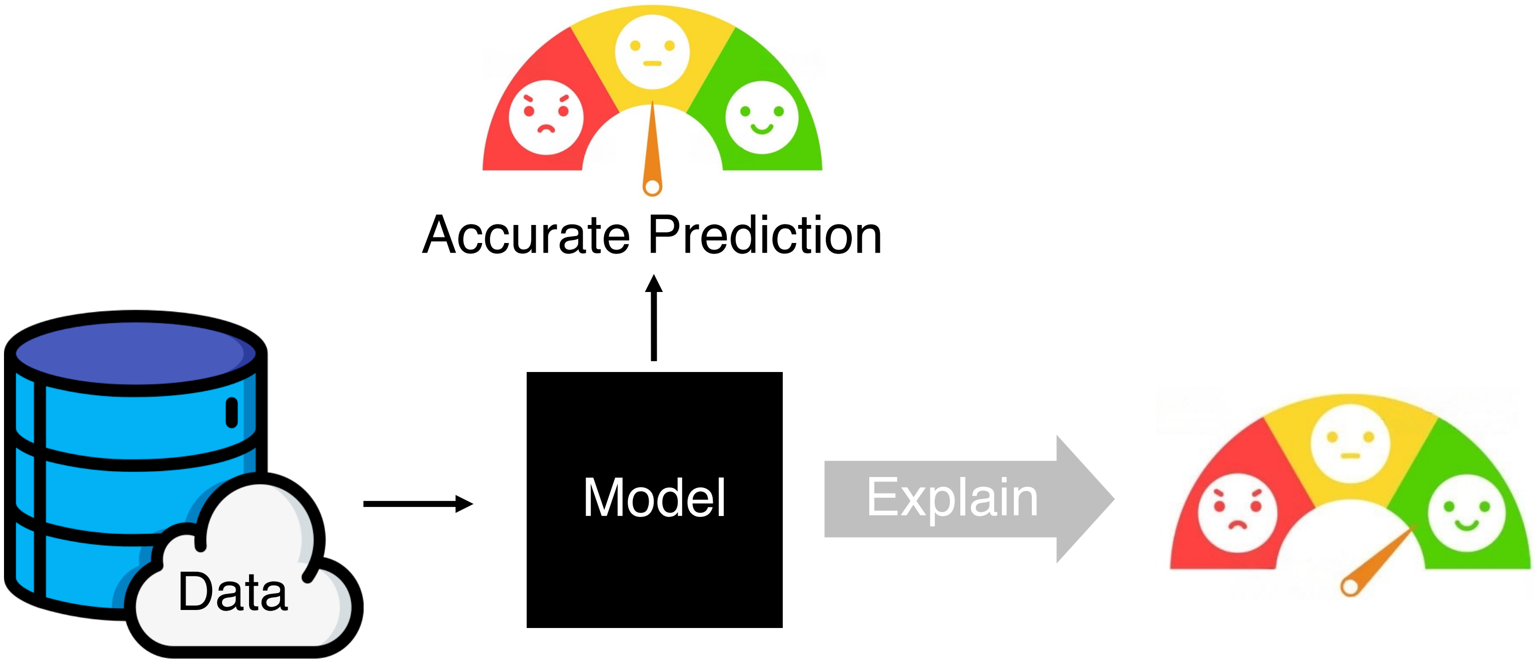
Problem: Model Explainability

- Many start-of-the-art AI models are black boxes.



Problem: Model Explainability

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



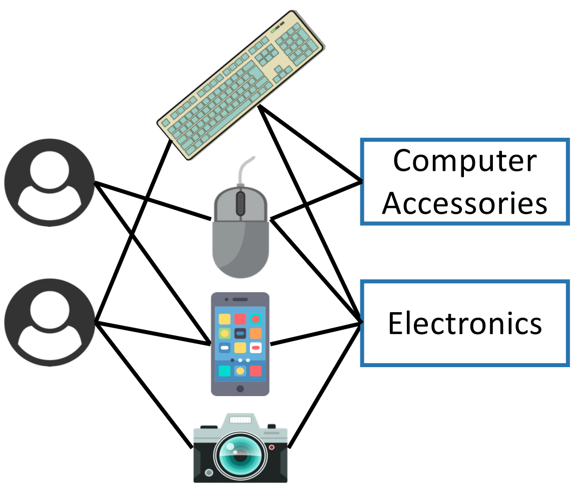
Model Explainability on The Web

Model Explainability on The Web

- Data: often as heterogeneous graphs.

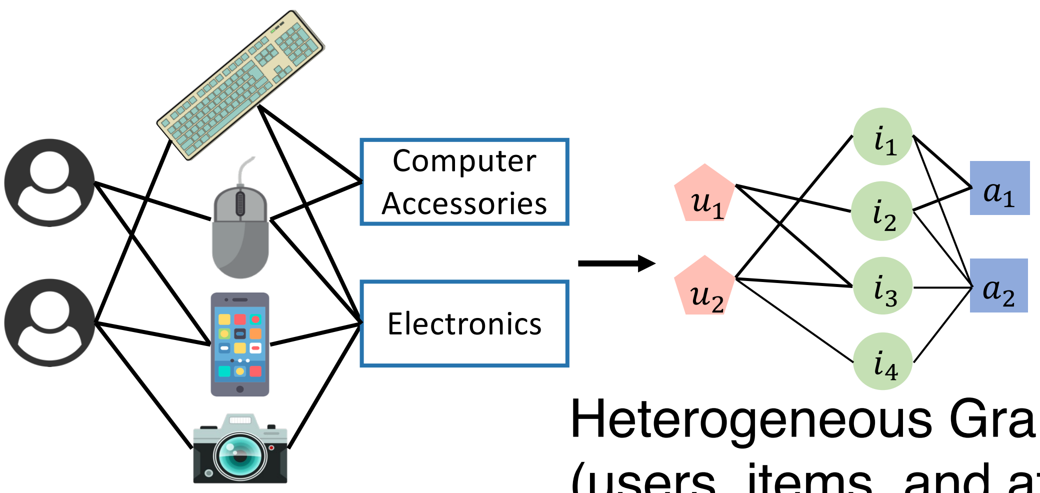
Model Explainability on The Web

- Data: often as heterogeneous graphs.



Model Explainability on The Web

- Data: often as heterogeneous graphs.

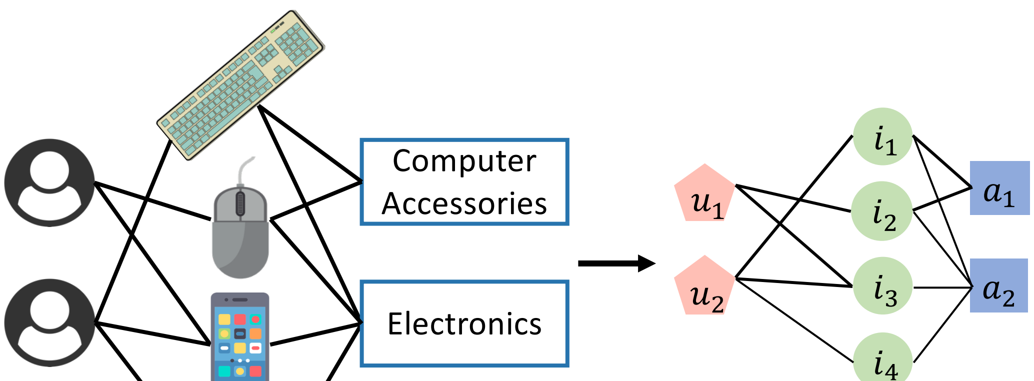


Heterogeneous Graph
(users, items, and attributes)

Zhang, Shichang, et al. "PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction" WWW 2023

Model Explainability on The Web

- Data: often as heterogeneous graphs.
- Model: Graph Neural Networks (GNNs) work well for heterogeneous link prediction.

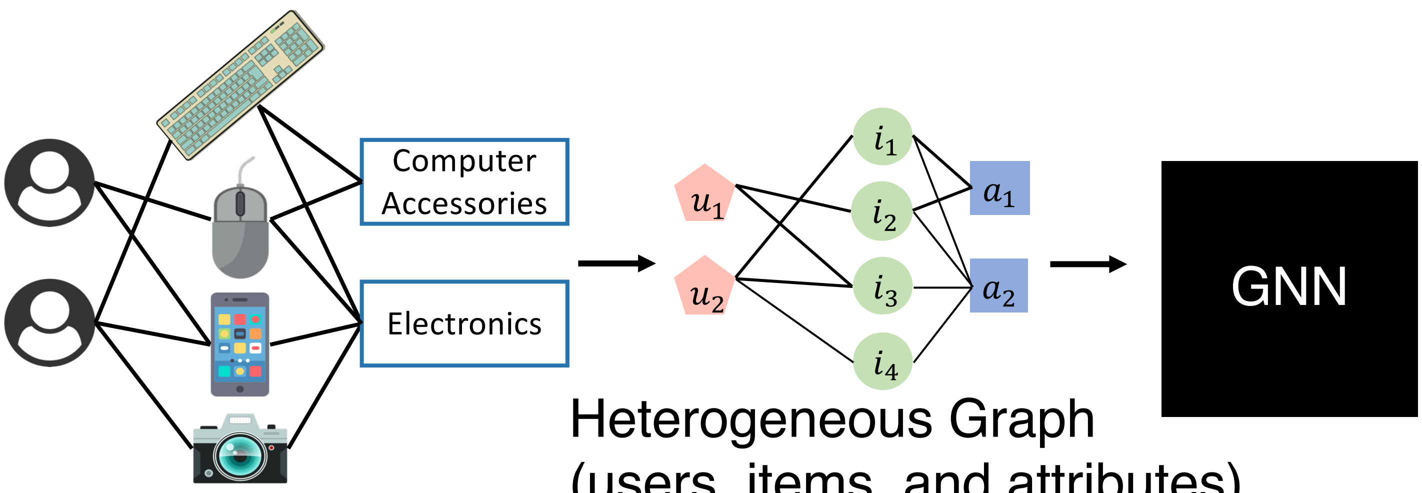


Heterogeneous Graph
(users, items, and attributes)

Zhang, Shichang, et al. "PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction" WWW 2023

Model Explainability on The Web

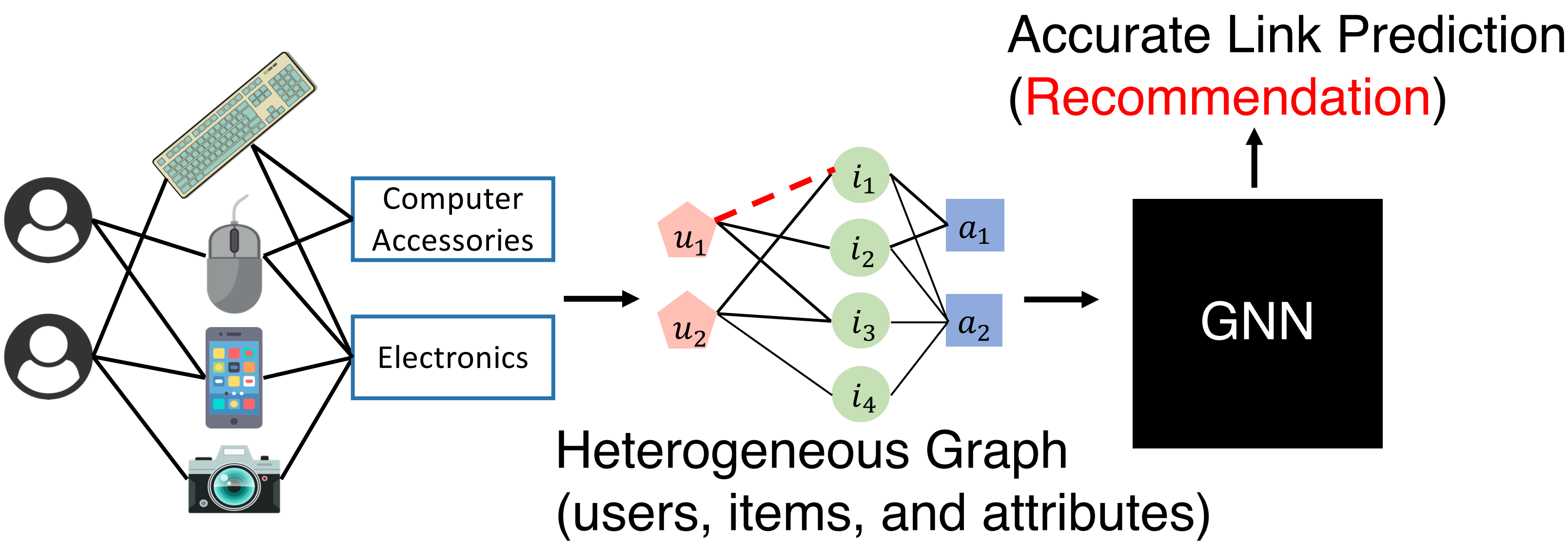
- Data: often as heterogeneous graphs.
- Model: Graph Neural Networks (GNNs) work well for heterogeneous link prediction.



Heterogeneous Graph
(users, items, and attributes)

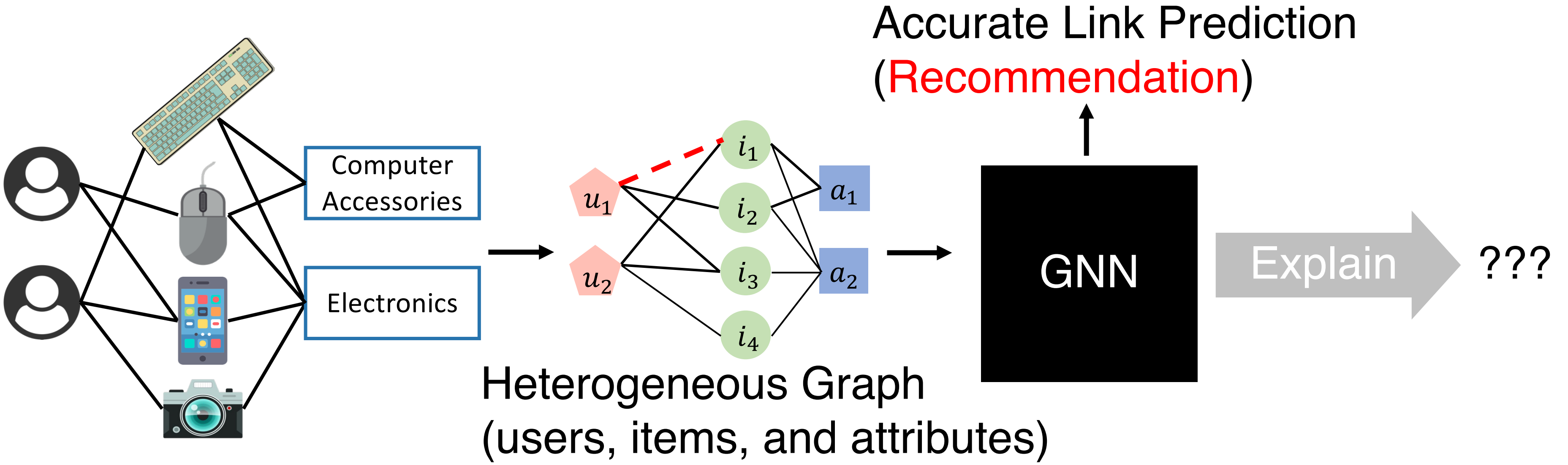
Model Explainability on The Web

- Data: often as heterogeneous graphs.
- Model: Graph Neural Networks (GNNs) work well for heterogeneous link prediction.



Model Explainability on The Web

- Data: often as heterogeneous graphs.
- Model: Graph Neural Networks (GNNs) work well for heterogeneous link prediction.
- Explanation: Why recommend an item to a user? (Why predict a user-item link?)



Zhang, Shichang, et al. "PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction" WWW 2023

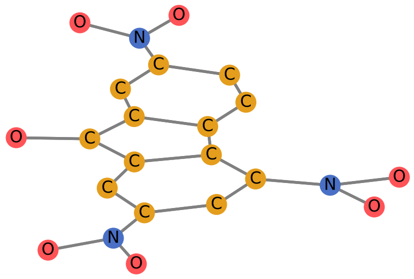
Existing Work: GNN Explanation

Existing Work: GNN Explanation

- Graph classification: property of a molecule.

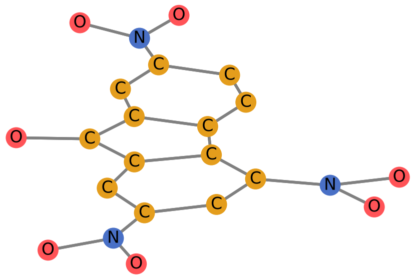
Existing Work: GNN Explanation

- Graph classification: property of a molecule.



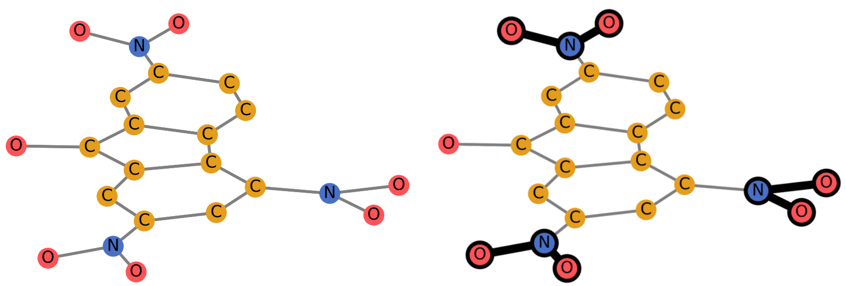
Existing Work: GNN Explanation

- Graph classification: property of a molecule.
  - Explained with general subgraphs.



Existing Work: GNN Explanation

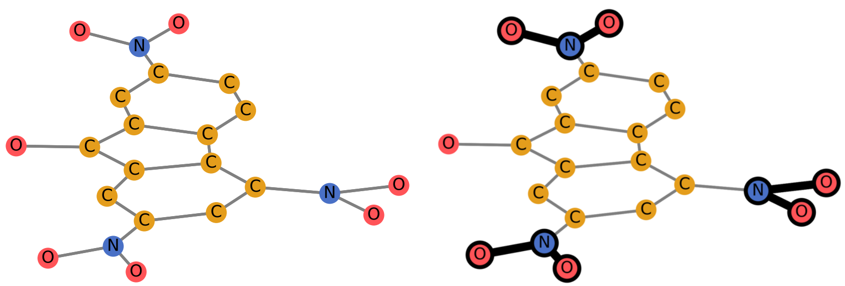
- Graph classification: property of a molecule.
  - Explained with general subgraphs.



- Explanation: subgraphs for -NO_2 motif

Existing Work: GNN Explanation

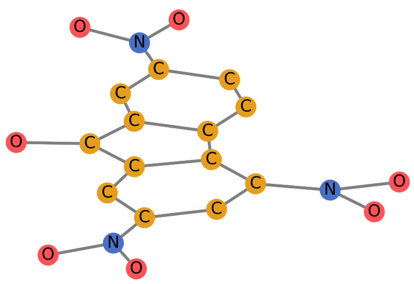
- Graph classification: property of a molecule.
  - Explained with general subgraphs.
- Heterogeneous link prediction: recommendation.

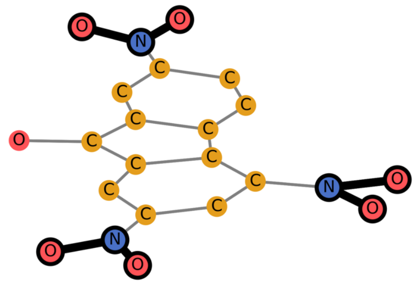


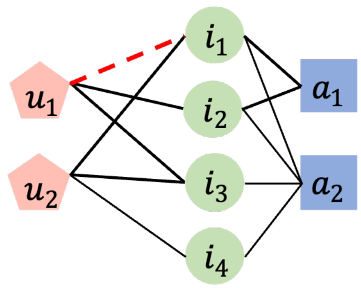
- Explanation: subgraphs for -NO_2 motif

Existing Work: GNN Explanation

- Graph classification: property of a molecule.
  - Explained with general subgraphs.
- Heterogeneous link prediction: recommendation.



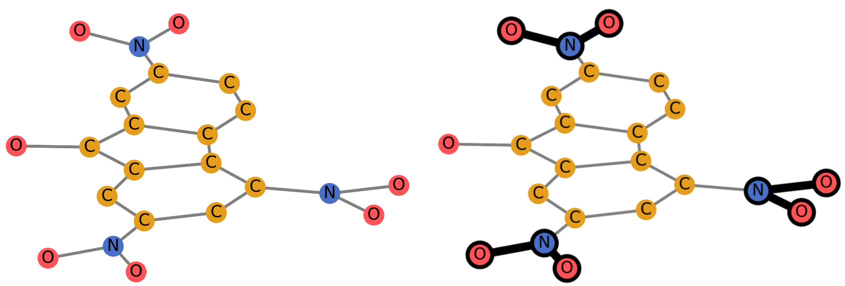


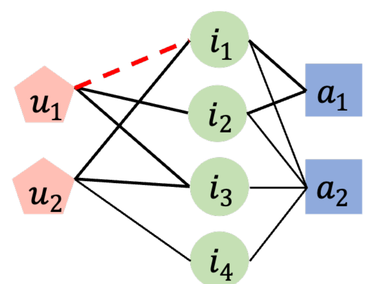


- Explanation: subgraphs for -NO_2 motif

Existing Work: GNN Explanation

- Graph classification: property of a molecule.
  - Explained with general subgraphs.
- Heterogeneous link prediction: recommendation.

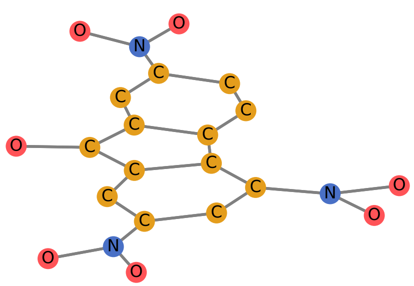


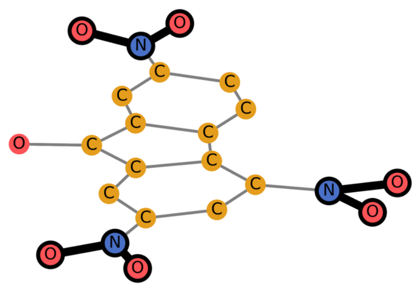


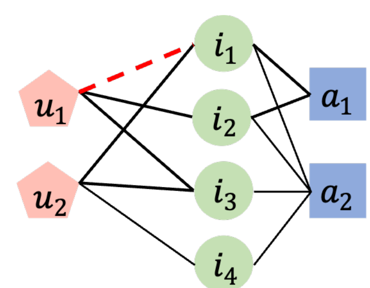
- Explanation: subgraphs for -NO_2 motif
- Explanation: ???

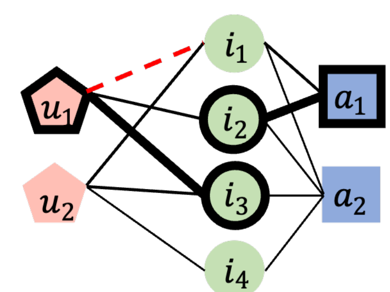
Existing Work: GNN Explanation

- Graph classification: property of a molecule.
  - Explained with general subgraphs.
- Heterogeneous link prediction: recommendation.





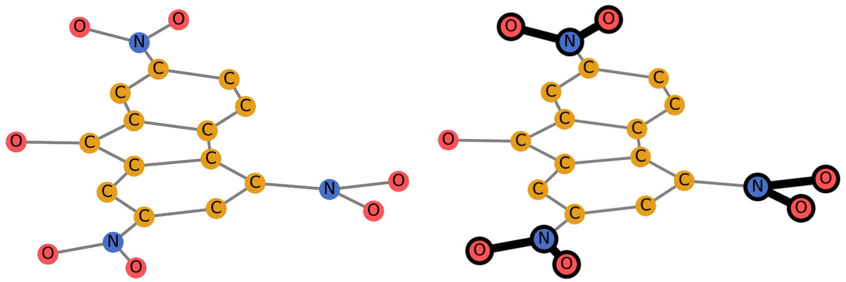


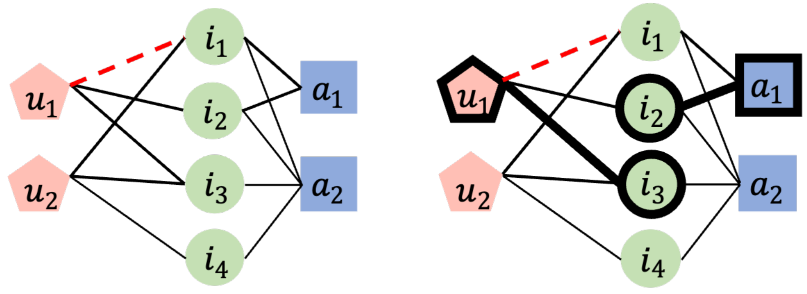


- Explanation: subgraphs for -NO_2 motif
- Explanation: ???

Existing Work: GNN Explanation

- Graph classification: property of a molecule.
  - Explained with general subgraphs.
- Heterogeneous link prediction: recommendation.
  - An ideal explanation should capture the connection between the source and the target.

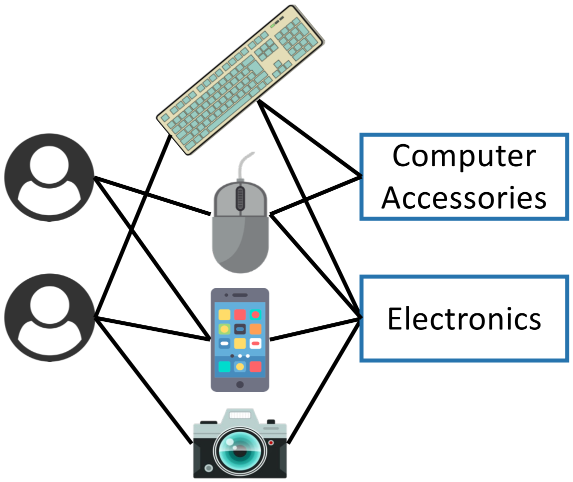




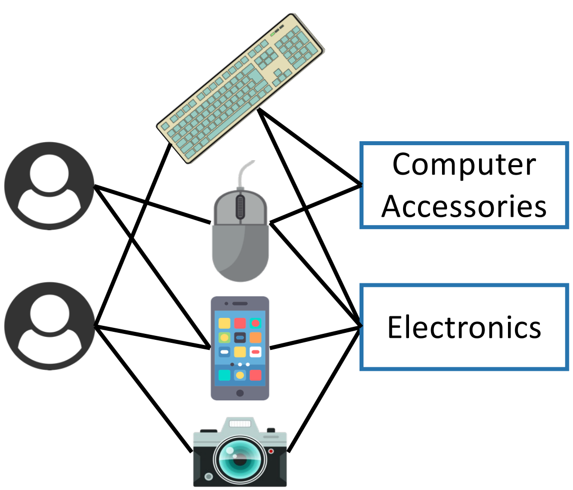
- Explanation: subgraphs for -NO_2 motif
- Explanation: ???

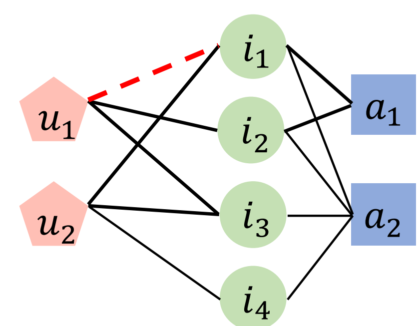
Main Idea: Paths As Explanations

Main Idea: Paths As Explanations



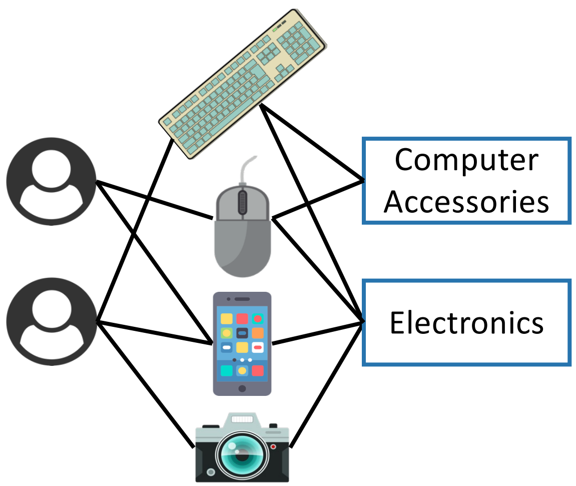
Main Idea: Paths As Explanations

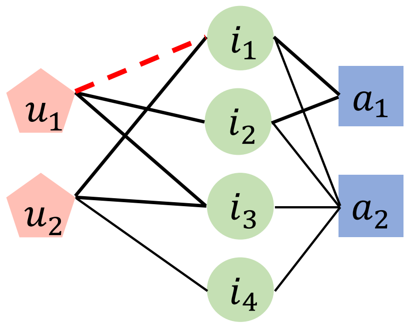




Main Idea: Paths As Explanations

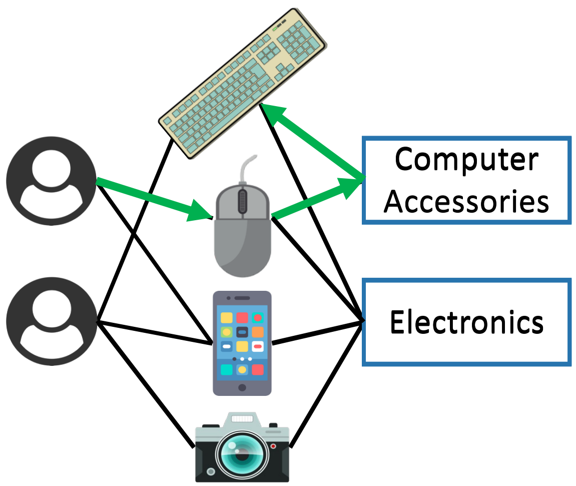
- Natural human-interpretable explanations boil down to paths.

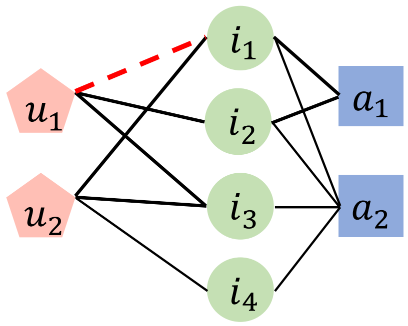




Main Idea: Paths As Explanations

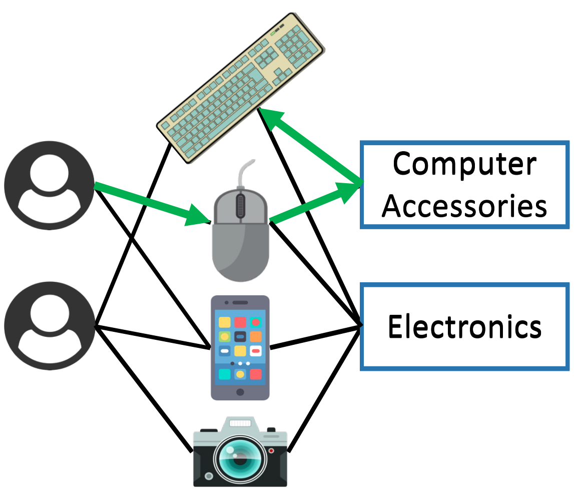
- Natural human-interpretable explanations boil down to paths.

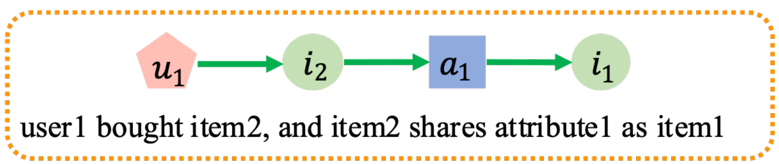


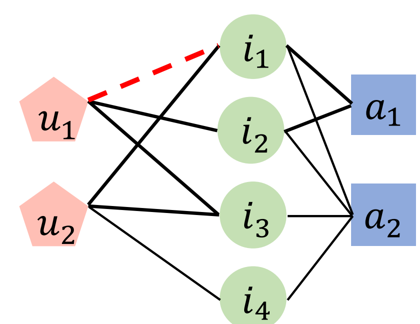


Main Idea: Paths As Explanations

- Natural human-interpretable explanations boil down to paths.



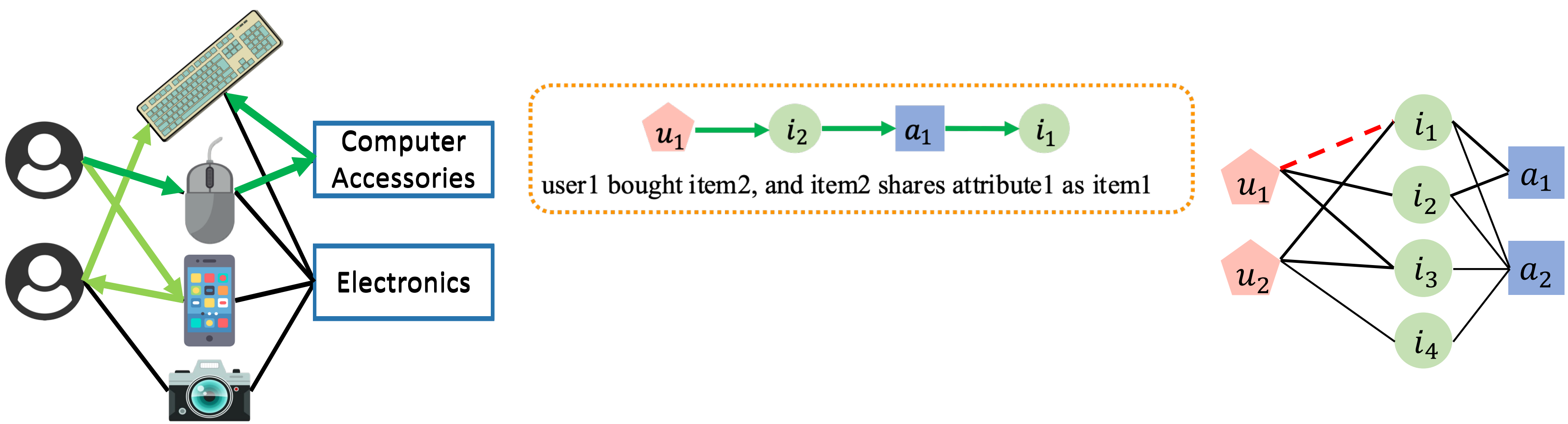




Zhang, Shichang, et al. “PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction” WWW 2023

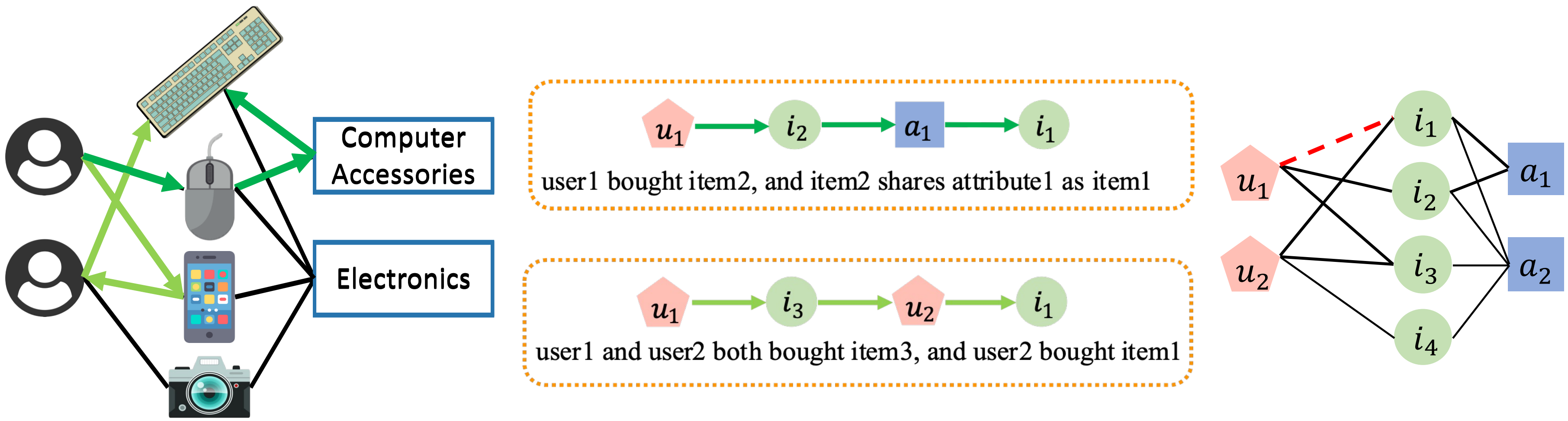
Main Idea: Paths As Explanations

- Natural human-interpretable explanations boil down to paths.



Main Idea: Paths As Explanations

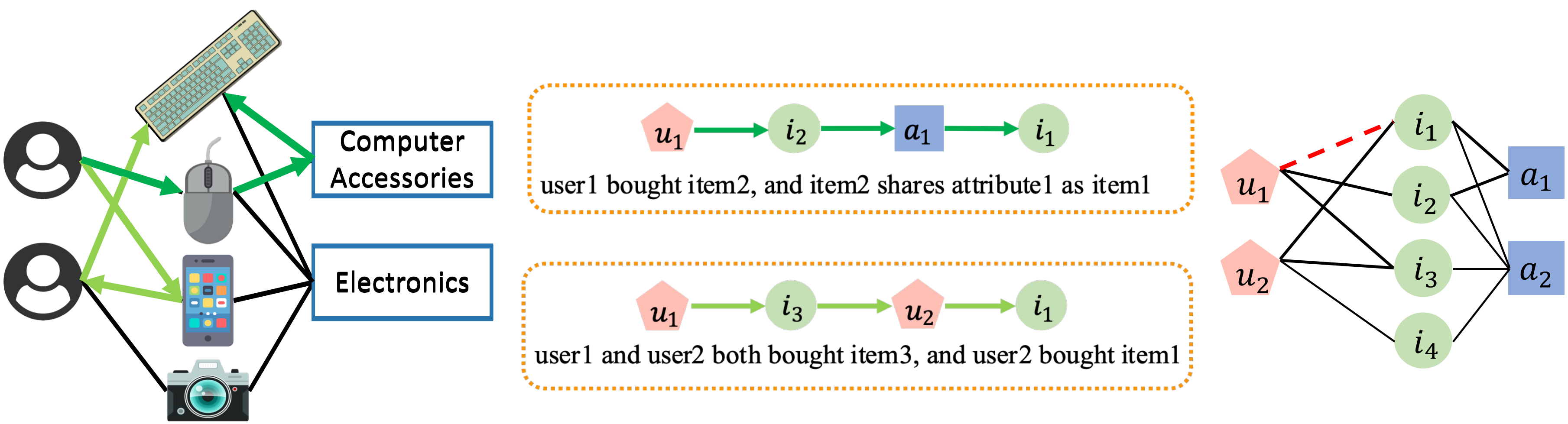
- Natural human-interpretable explanations boil down to paths.



Zhang, Shichang, et al. "PaGE-Link: Path-based Graph Neural Network Explanation for Heterogeneous Link Prediction" WWW 2023

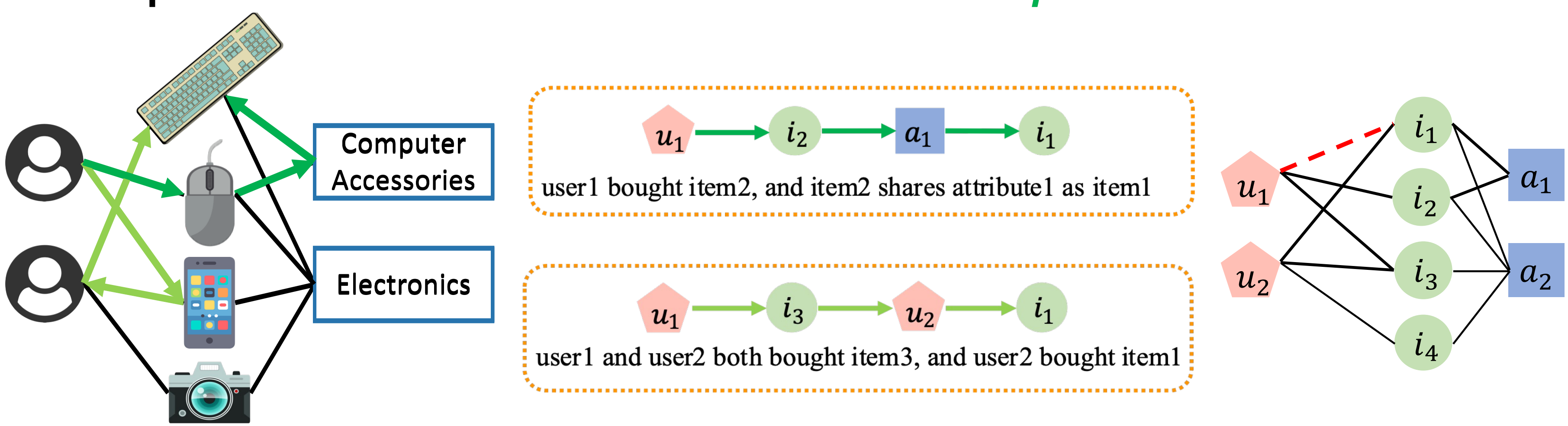
Main Idea: Paths As Explanations

- Natural human-interpretable explanations boil down to paths.
- Paths form a much smaller search space compared to general subgraphs.



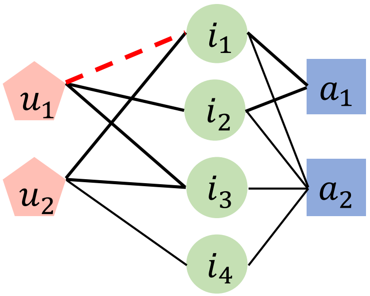
Main Idea: Paths As Explanations

- 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*.



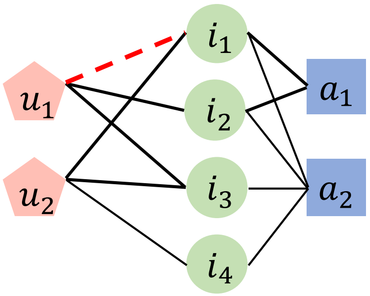
PaGE-Link: Path-Enforcing Mask

PaGE-Link: Path-Enforcing Mask



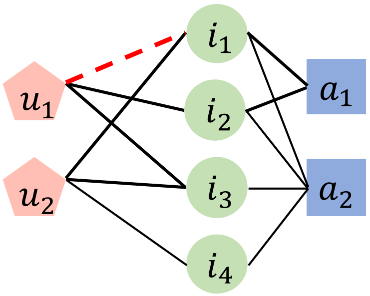
PaGE-Link: Path-Enforcing Mask

- Challenges for finding good paths.



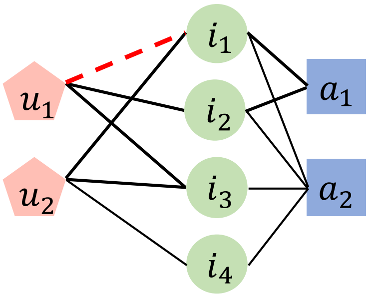
PaGE-Link: Path-Enforcing Mask

- Challenges for finding good paths.
  - Many path candidates.



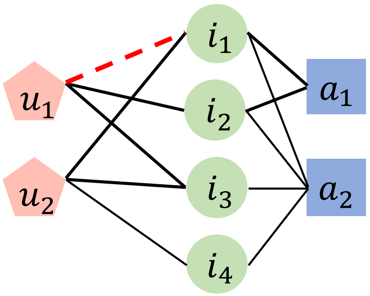
PaGE-Link: Path-Enforcing Mask

- Challenges for finding good paths.
  - Many path candidates.
  - Criterion for selecting good paths.



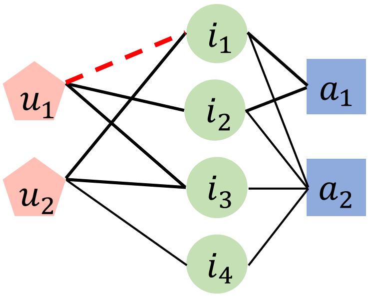
PaGE-Link: Path-Enforcing Mask

- Challenges for finding good paths.
  - Many path candidates.
  - Criterion for selecting good paths.
- Learn an edge mask to select meaningful edges.



UCLA

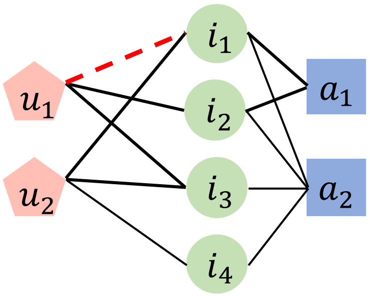
PaGE-Link: Path-Enforcing Mask



- 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.

PaGE-Link: Path-Enforcing Mask

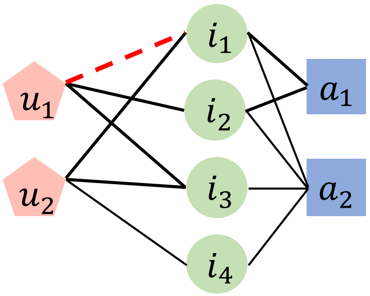
- Challenges for finding good paths.
  - Many path candidates.
  - Criterion for selecting good paths.



- Learn an edge mask to select meaningful edges.
  - Edges form short paths with low-degree nodes.

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

PaGE-Link: Path-Enforcing Mask

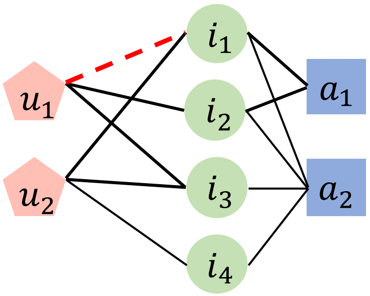


- Challenges for finding good paths.
  - Many path candidates.
  - Criterion for selecting good paths.
- Learn an edge mask to select meaningful edges.
  - Edges form short paths with low-degree nodes.

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

  - Edges maximize the mutual information.

PaGE-Link: Path-Enforcing Mask



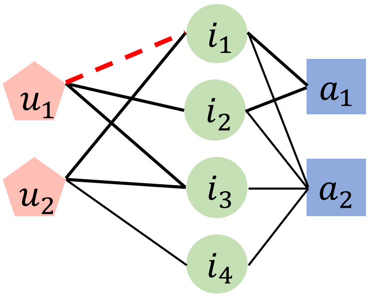
- Challenges for finding good paths.
  - Many path candidates.
  - Criterion for selecting good paths.
- Learn an edge mask to select meaningful edges.
  - Edges form short paths with low-degree nodes.

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

  - Edges maximize the mutual information.

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

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.

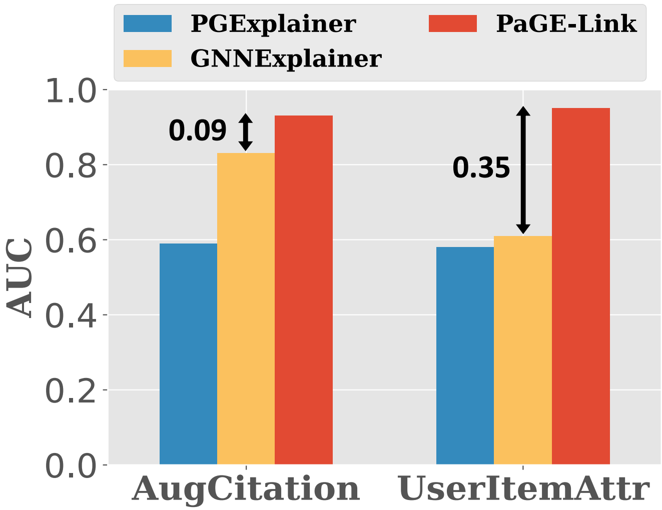
Experiments

Experiments

- ROC-AUC: 9%-35% improvement over baselines.

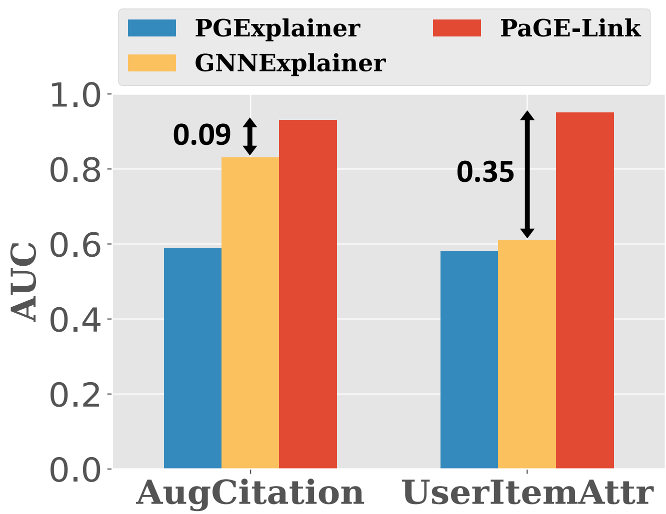
Experiments

- ROC-AUC: 9%-35% improvement over baselines.



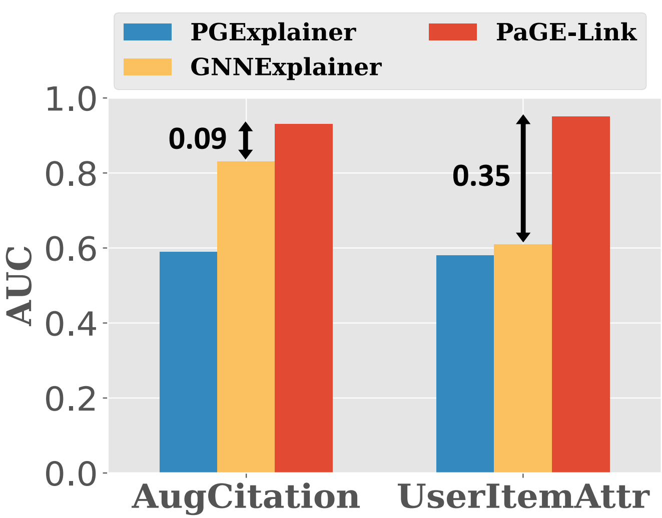
Experiments

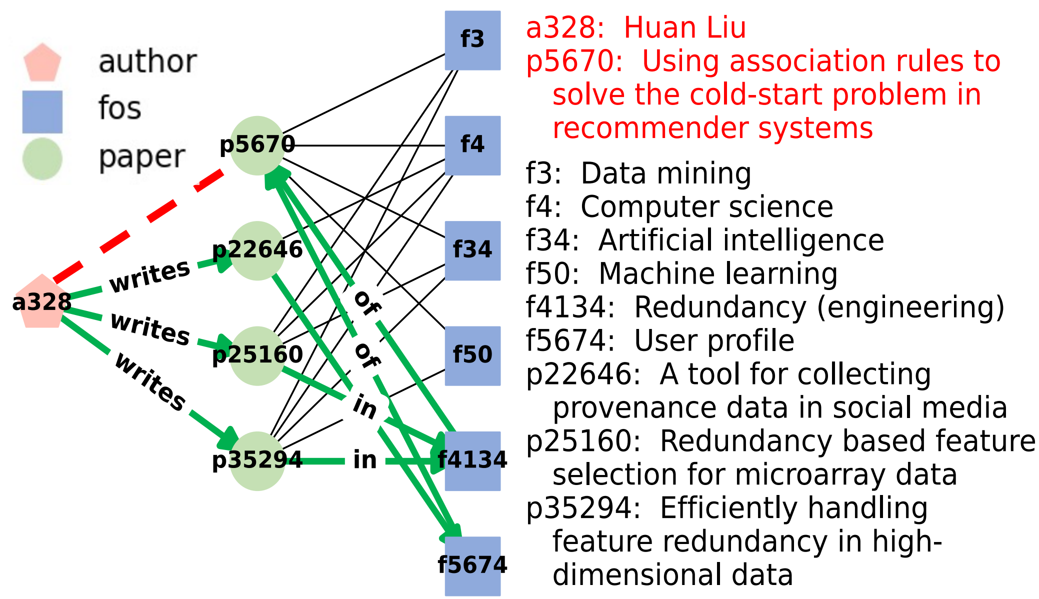
- ROC-AUC: 9%-35% improvement over baselines.
- Concise paths without generic nodes.



Experiments

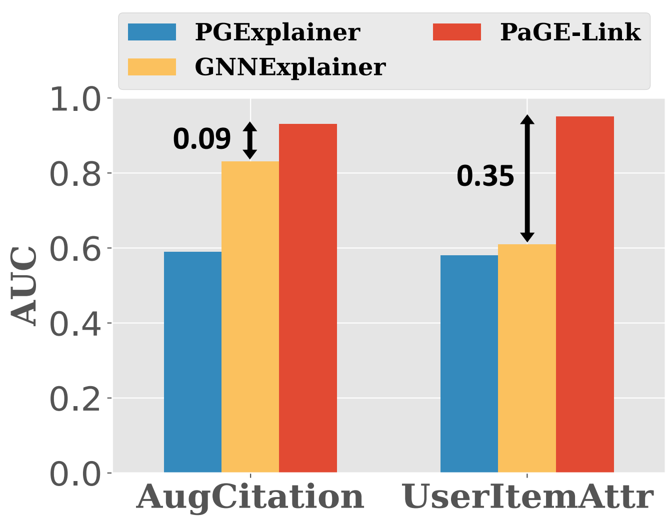
- ROC-AUC: 9%-35% improvement over baselines.
- Concise paths without generic nodes.

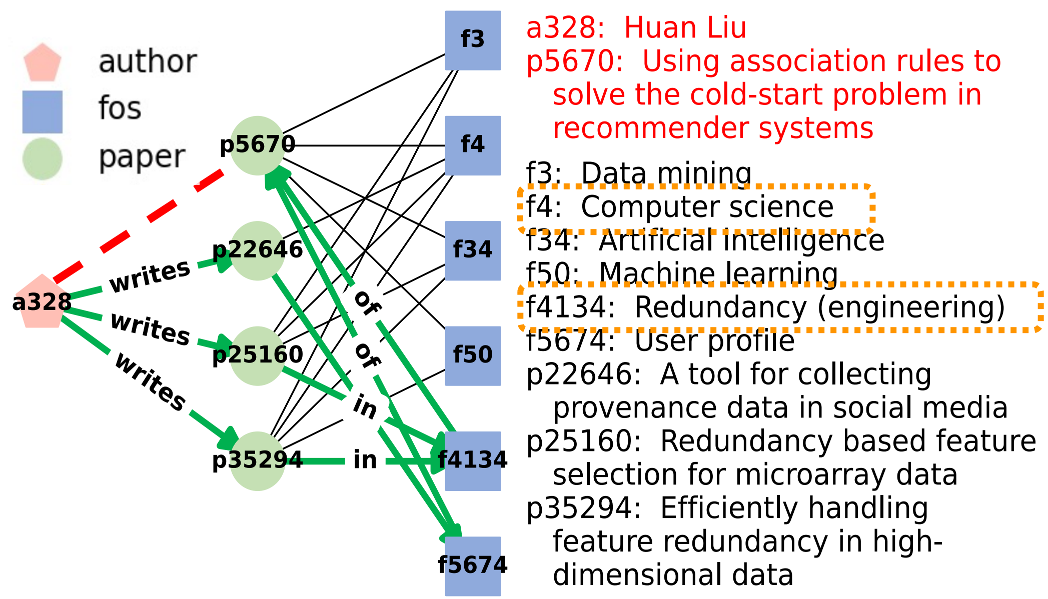




Experiments

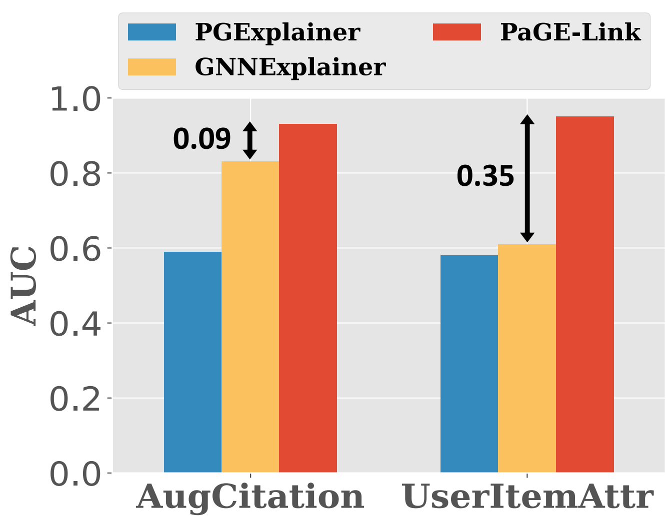
- ROC-AUC: 9%-35% improvement over baselines.
- Concise paths without generic nodes.

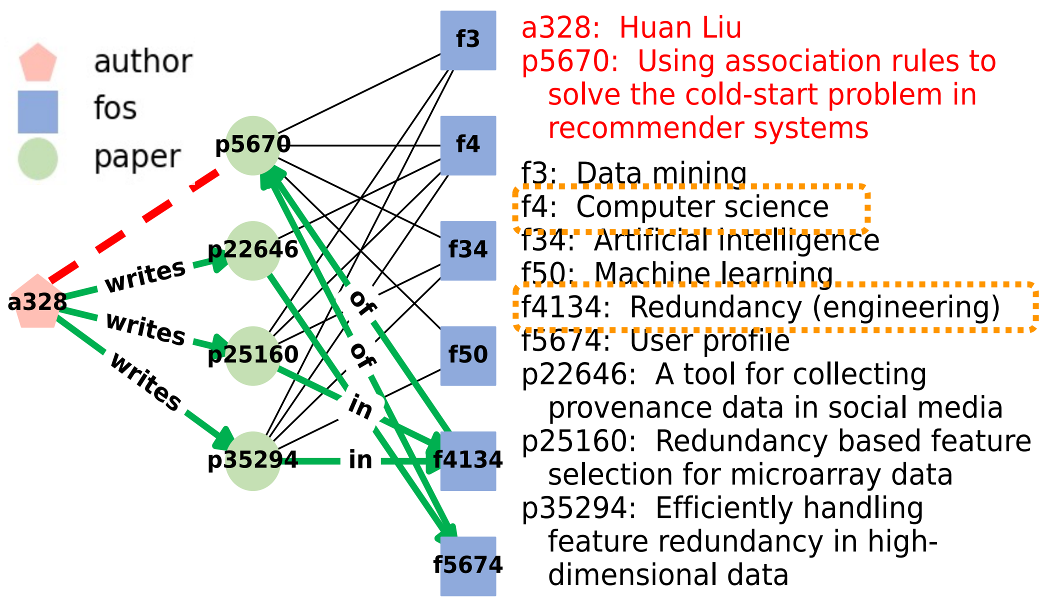




Experiments

- 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

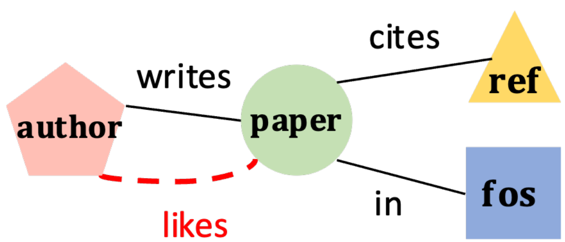
Generate datasets new evaluation

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

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

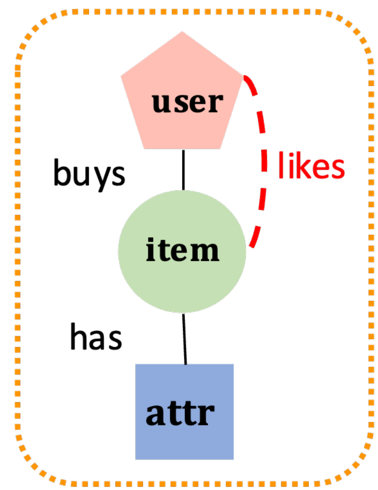
- Use p as the ground truth for evaluating the prediction of (s, t)

Citation graph



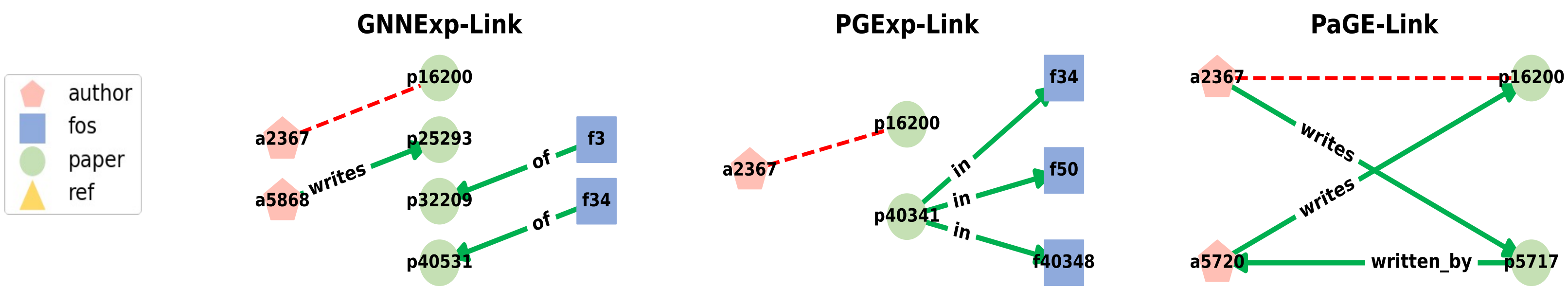
Augmented graph schema

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.*