



KIND Lab



Yongjia Lei Ph.D. Student



Yu Wang Ph.D. Advisor

Background

Real-world Problem-solving Needs Knowledge

What's the formula? What are they talking about?



Knowledge Retrieval is Crucial

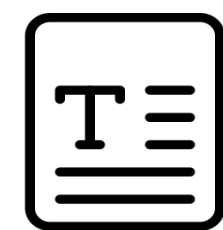
Open-book Exam



Google Search



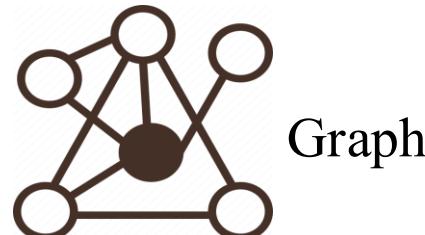
Heterogeneous Knowledge and Retrieval



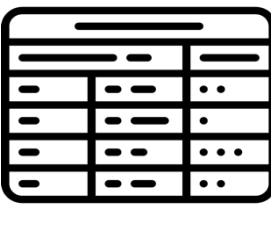
Text



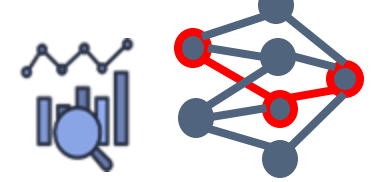
Image



Graph



Table



Motivation

Retrieval for Complex Queries

Q: Publications by Point Park University authors on stellar populations in tidal tails

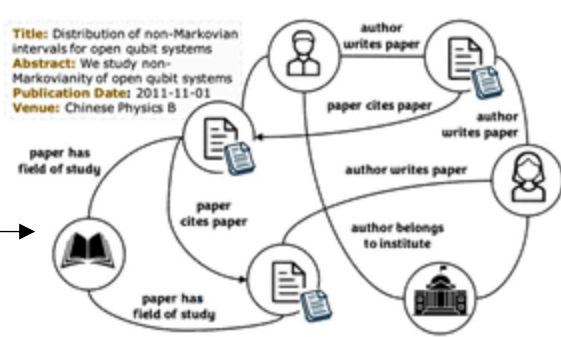


Ask

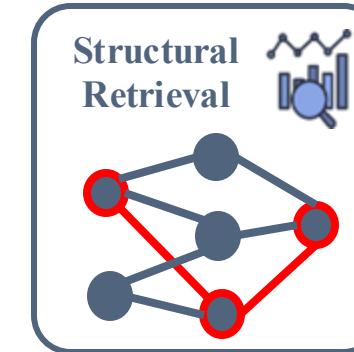
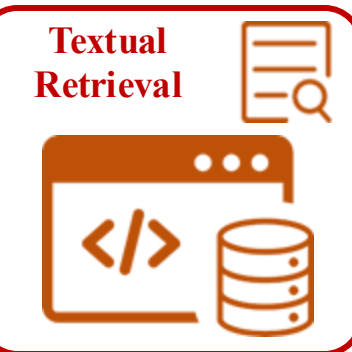
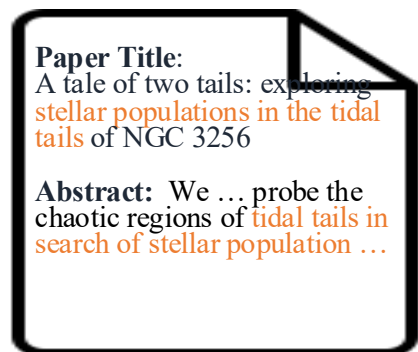


Answer

Retrieve



MAG



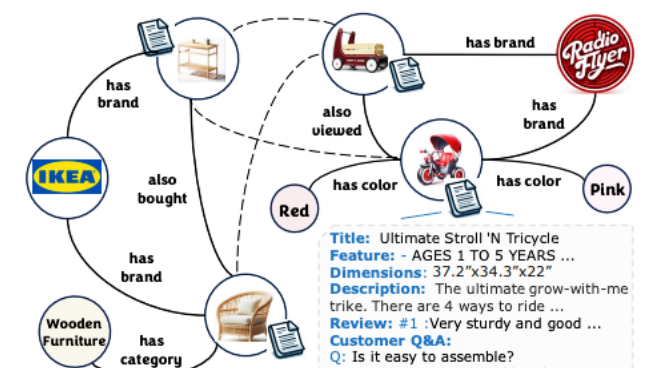
Author

Institution

Field of Study

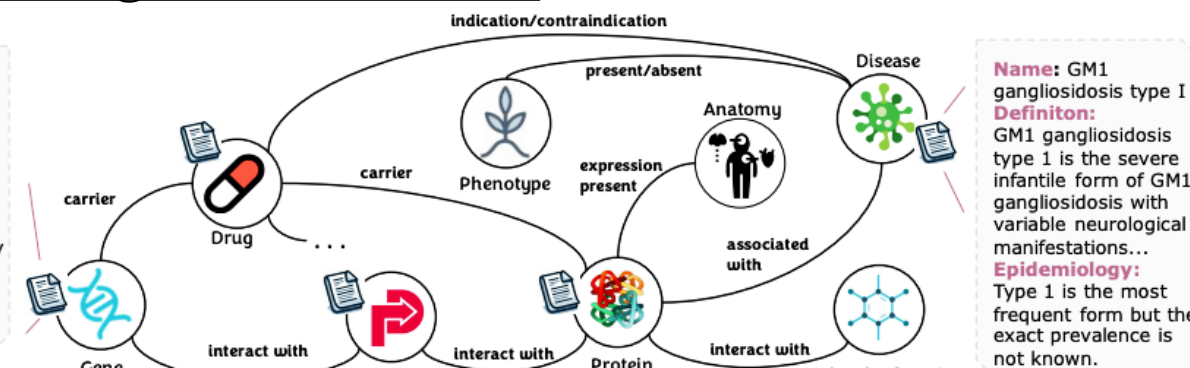
Paper

More Heterogeneous Datasets

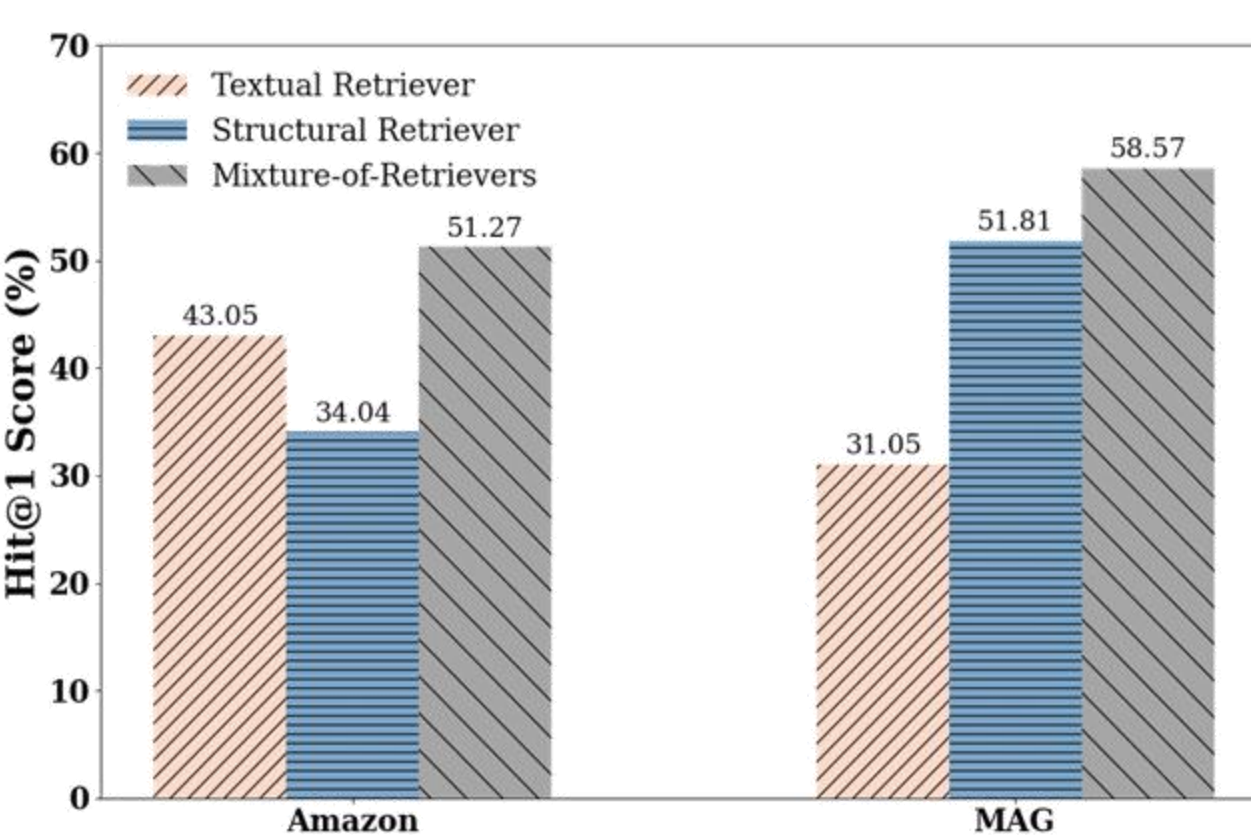


Amazon

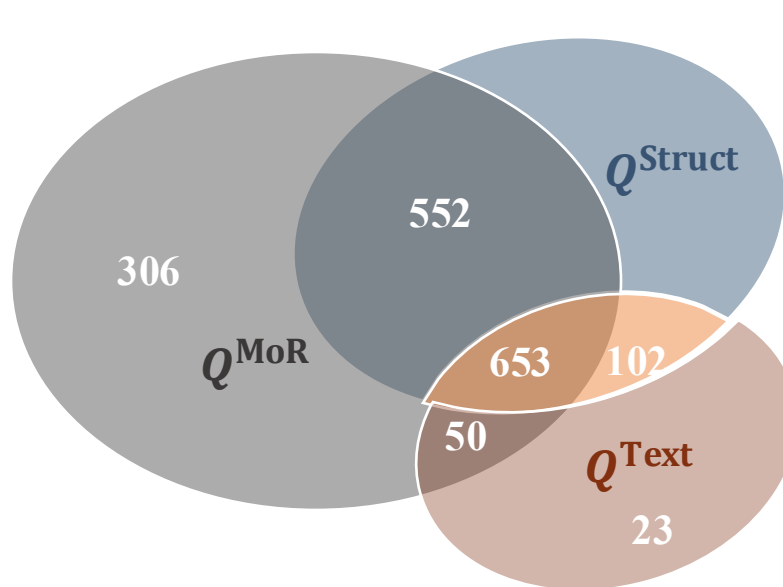
Prime



Textual vs Structural vs Mixture Retrieval



Queries in MAG

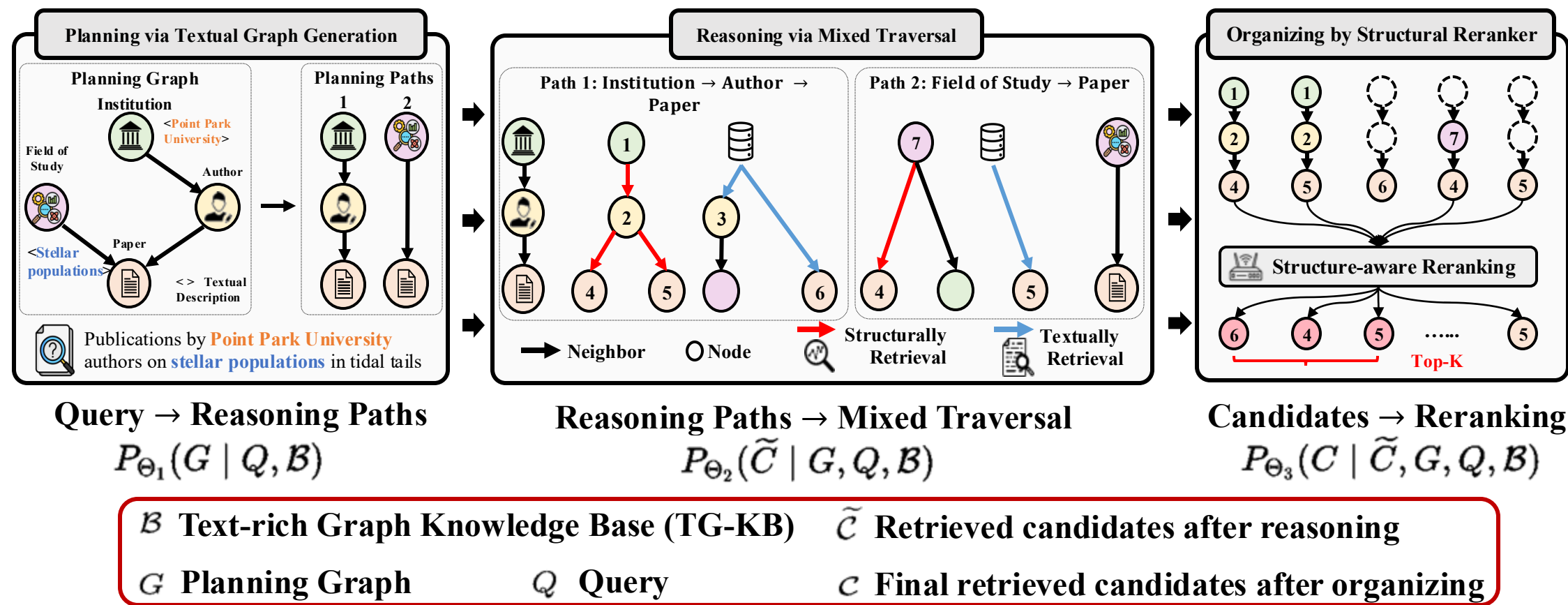


$$Q_{\text{Text}} \neq Q_{\text{Struct}} \\ |Q^{\text{MoR}}| > |Q_{\text{Text}} \cup Q_{\text{Struct}}|$$

Mixture of Structural-and-Textual Retrieval over Text-rich Graph Knowledge Bases

Method

Mixture of Structural-and-Textual Retrieval Framework (MoR)



Results

Retrieval Performance on Three TG-KBs

Category	Retrieval Baseline	AMAZON				MAG				PRIME			
		H@1	H@5	R@20	MRR	H@1	H@5	R@20	MRR	H@1	H@5	R@20	MRR
Textual	BM25 (Wu et al.)	44.94	67.42	53.77	55.30	25.85	45.25	45.69	34.91	12.75	27.92	31.25	19.84
	Ada-002 (Wu et al.)	39.16	62.73	53.29	50.35	29.08	49.61	48.36	38.62	12.63	31.49	36.00	21.41
	Multi-ada-002 (Wu et al.)	40.07	64.98	55.12	51.55	25.92	50.43	50.80	36.94	15.10	33.56	38.05	23.49
	DPR (Karpukhin et al., 2020)	15.29	47.93	44.49	30.20	10.51	35.23	42.11	21.34	4.46	21.85	30.13	12.38
Structural (KG)	QAGNN (Yasunaga et al., 2021)	26.56	50.01	52.05	37.75	12.88	39.01	46.97	29.12	8.85	21.35	29.63	14.73
	ToG (Sun et al., 2023)	-	-	-	-	13.16	16.17	11.30	14.18	6.07	15.71	13.07	10.17
Hybrid	AvaTaR (Wu et al., 2025)	49.87	69.16	60.57	58.70	44.36	59.66	50.63	51.15	18.44	36.73	39.31	26.73
	KAR (Xia et al., 2024)	54.20	68.70	57.24	61.29	50.47	69.57	60.28	58.65	30.35	49.30	50.81	39.22
	MFAR* (Li et al., 2024)	41.20	70.00	58.50	54.20	49.00	69.60	71.70	58.20	40.90	62.80	68.30	51.20
	MoR	52.19	74.65	59.92	62.24	58.19	78.34	75.01	67.14	36.41	60.01	63.48	46.92
Ablation	MoR _{w/o} R	44.21	68.87	56.50	55.28	34.33	62.55	67.55	47.40	31.59	53.48	60.74	41.81
	MoR _{w/o} RT	34.04	53.41	45.16	42.85	51.81	73.54	74.17	61.68	28.95	46.12	49.54	36.56
	MoR _{w/o} RS	43.05	69.36	57.38	54.69	31.05	51.84	50.56	40.64	22.27	38.45	39.21	29.41

MoR advances the state-of-the-art retriever by 5%

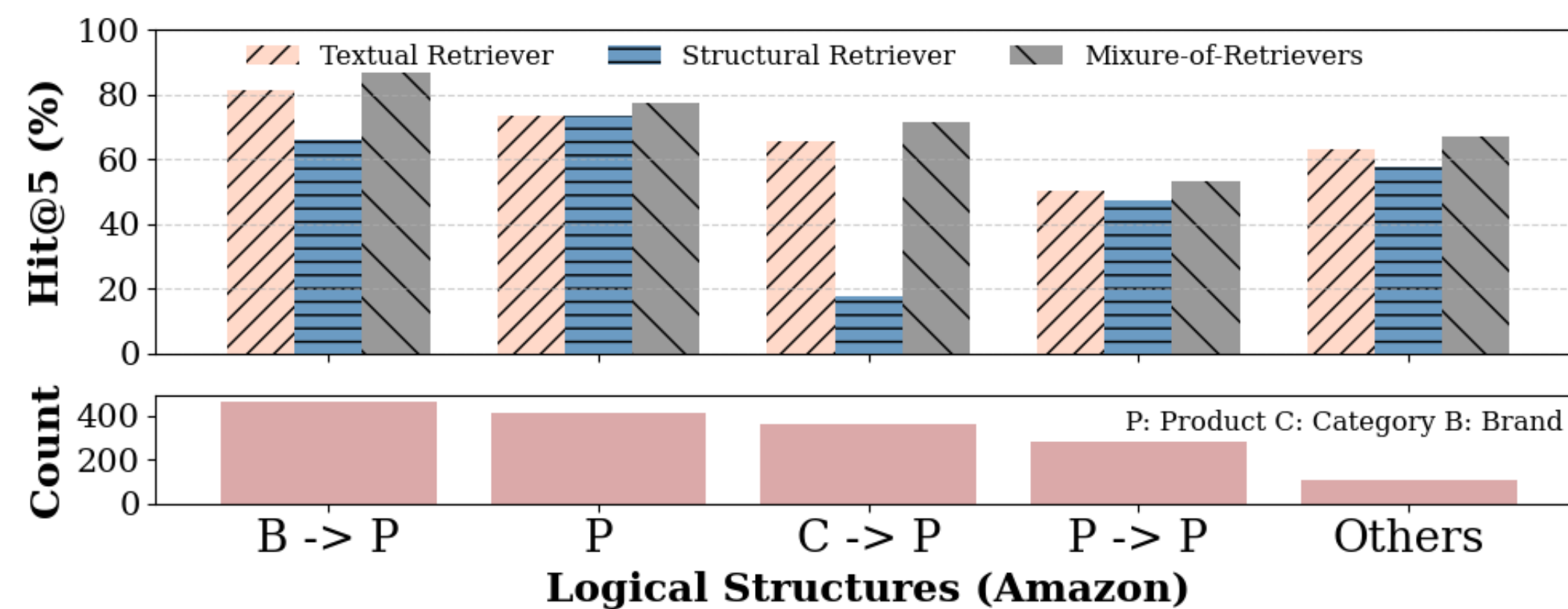
Node Feature VS. Path Feature

Feature	MAG			Amazon			Prime		
	H@1	R@20	MRR	H@1	R@20	MRR	H@1	R@20	MRR
Last Node	49.91	73.49	59.92	50.36	59.62	61.05	33.52	61.95	44.15
Trajectory	58.19	75.01	67.14	52.19	59.92	62.24	36.41	63.48	46.92

Last Node: the last node on the traversed path

Trajectory: the whole traversed path

Imbalanced Logical Structures and Performance



B → P: Brand → Product C → P: Category → Product P → P: Product → Product P: Product

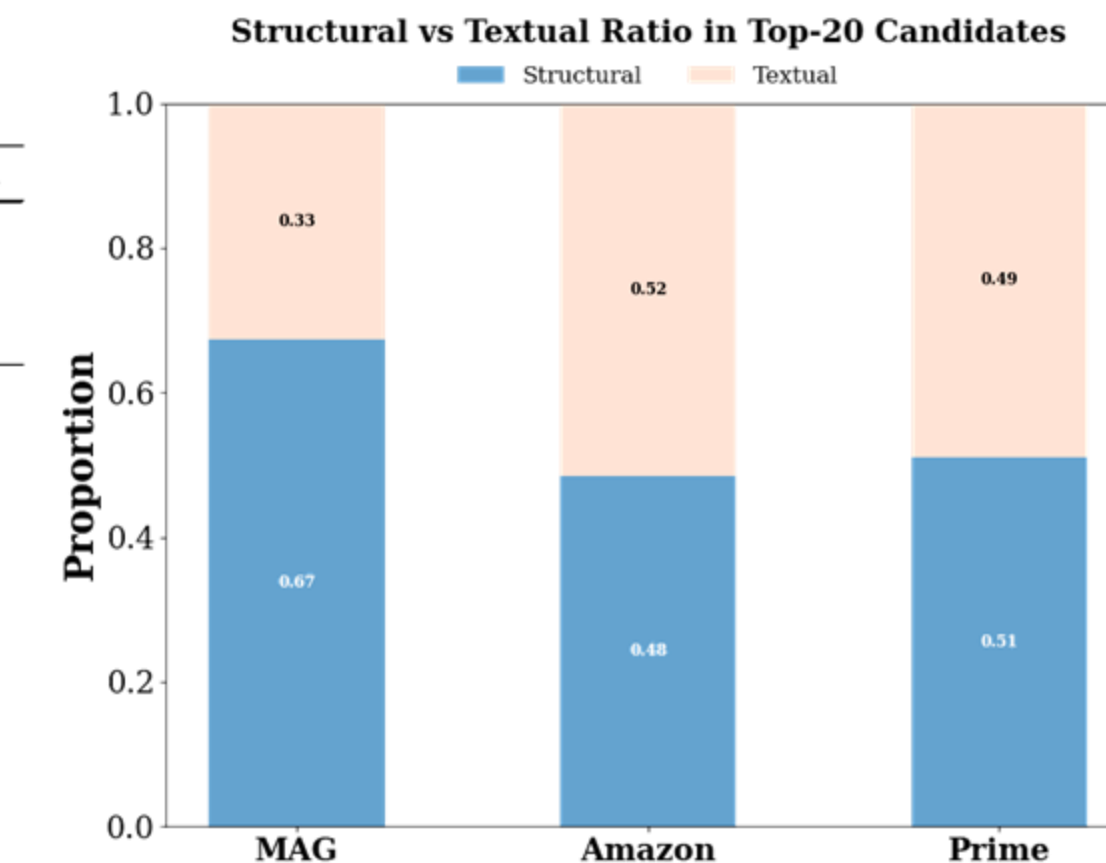


Top-20 Candidates Analysis

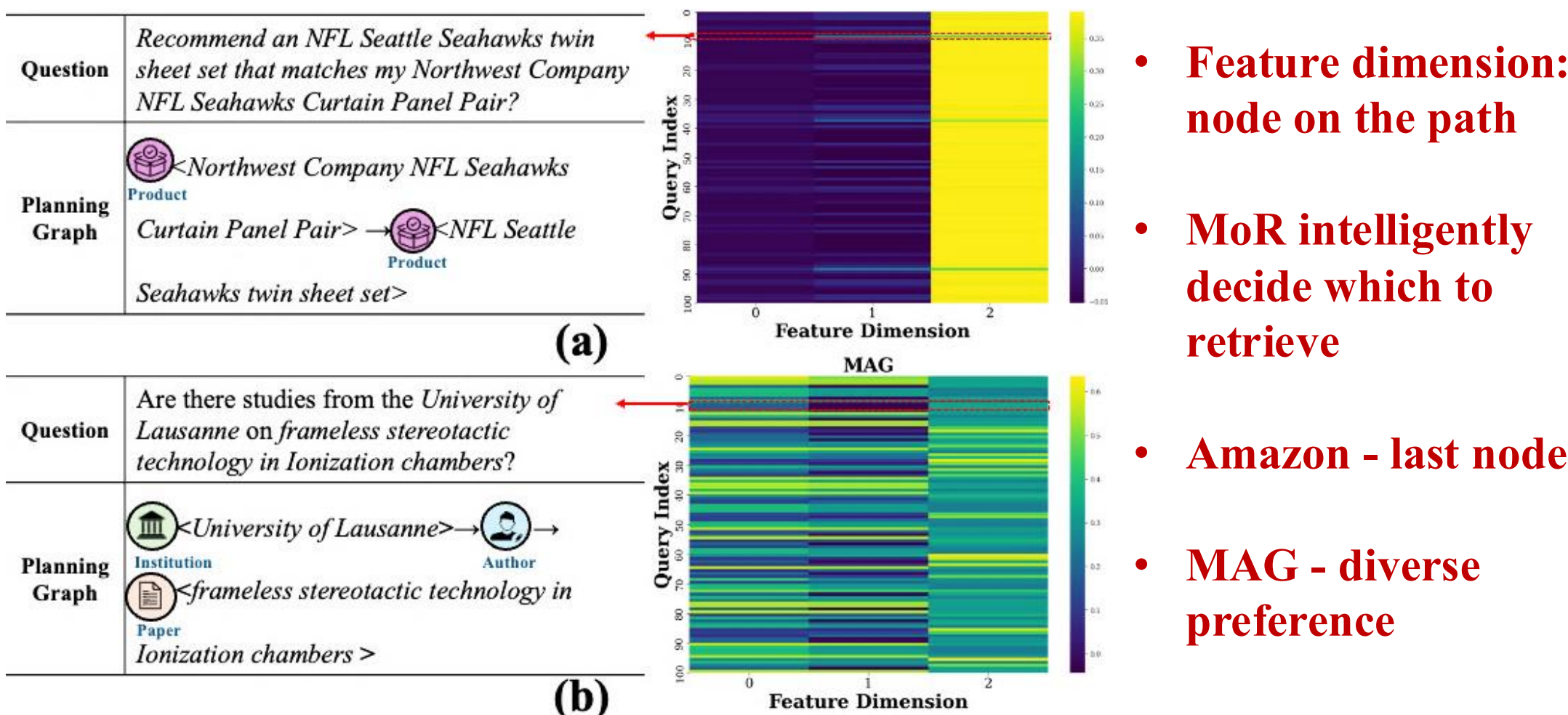
Dataset Information

Dataset	Textual Information	Structural Information
Amazon	~70%	~30%
MAG	~35%	~65%
Prime	~25%	~75%

- Prioritize retrieval strategies based on dataset properties
 - Amazon: textual knowledge
 - MAG: structural knowledge
 - Prime: domain knowledge

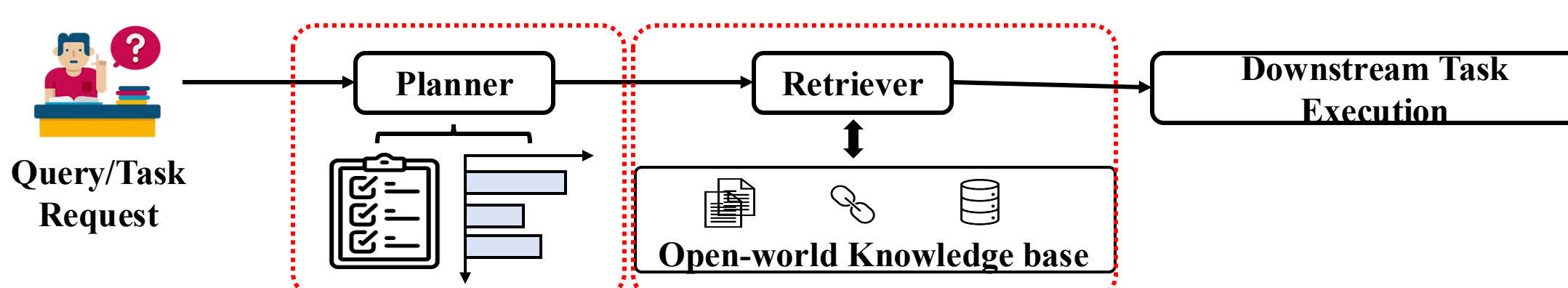


Case Study: Node importance Visualization via Saliency Map



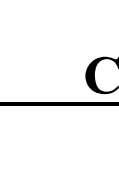
Future Work

Retrieval Under Uncertainty-aware Planning



Q: What disease should not be treated with Methylphenidate but can be managed with L-Glutamine?

Multiple Potential Plan



Methylphenidate → [not treat] → disease (0.2) ✓
Methylphenidate → [treat] → disease (0.5) ✗
L-Glutamine → [manage] → disease (0.3) ✓

- How to guarantee the correctness of generated plans?
- How to retrieve ground-truth knowledge under multiple uncertain plans?

Acknowledgements

Collaborators



Haoyu Han



Ryan A. Rossi



Franck Dernoncourt



Mahantesh M Halappanavar



Nedim Lipka



Jiliang Tang



Personal website



WeChat