Part II: Graph Mining with Large Language Models

Bowen Jin, Yu Zhang, Sha Li, Jiawei Han
Department of Computer Science
University of Illinois at Urbana-Champaign
Mar 4, 2024

Tutorial Website:

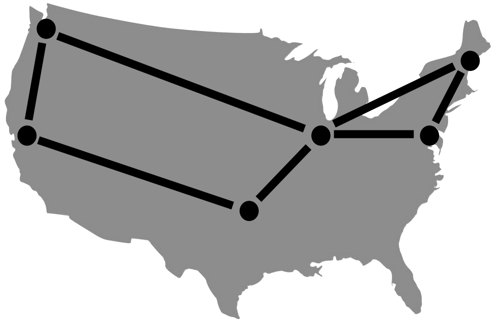
Outline

- Why Mining Graphs with Large Language Models?
  - Three Main Scenarios
- Mining Pure Graphs with Large Language Models
- Mining Text-Attributed Graphs with Large Language Models
- Mining Text-Paired Graphs with Large Language Models

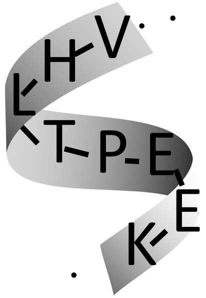
Graphs

- Graph data is ubiquitous in real world.

Traffic Graphs

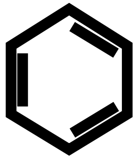


Protein Graphs

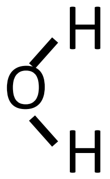


"Myoglobin holds oxygen in muscles."

Molecule Graphs

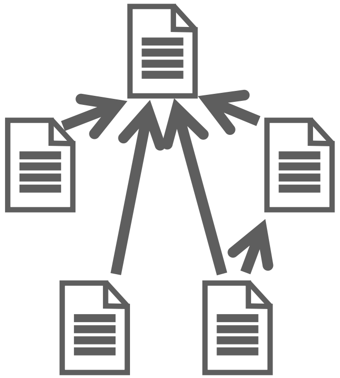


"Benzene is toxic"

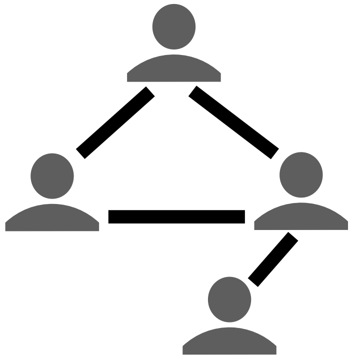


"Water is less toxic"

Academic Graphs

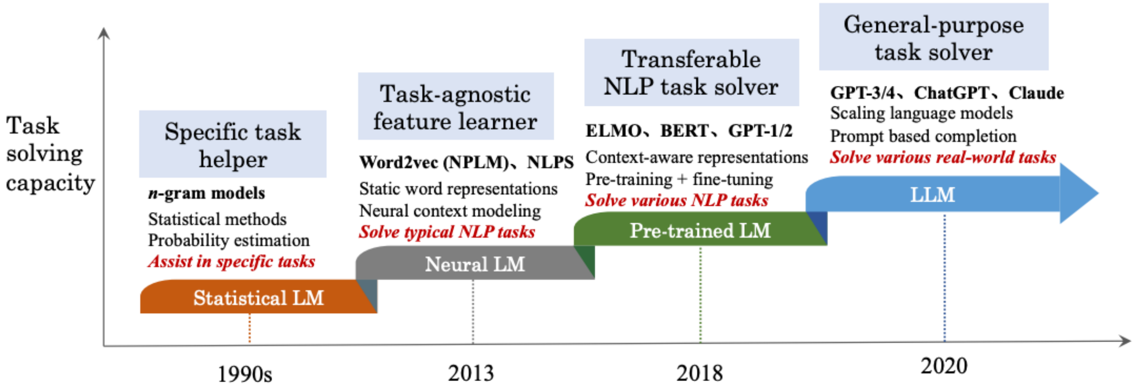


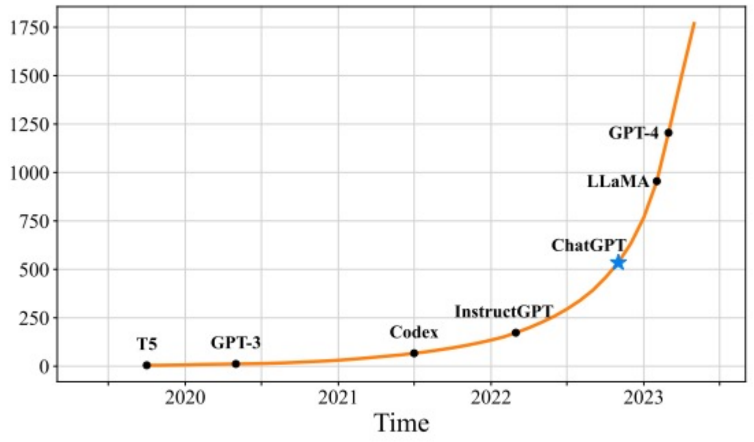
Social Graphs



Large Language Models (LLMs)

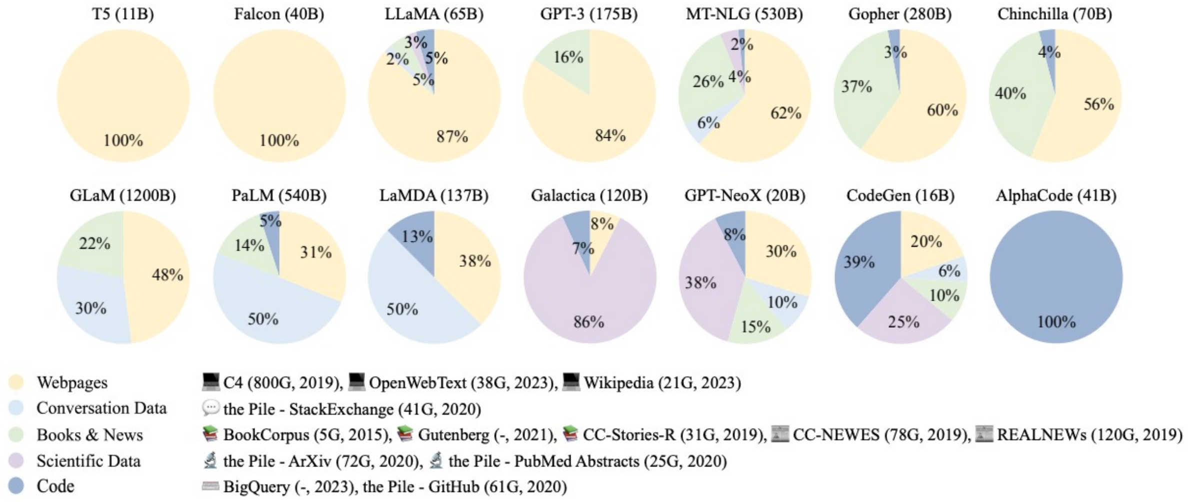
- LLMs have demonstrated their strong **text** encoding/decoding ability.





(b) Query="Large Language Model"

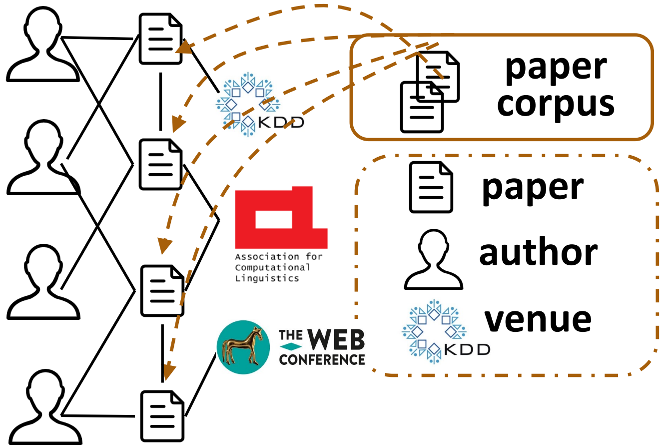
- LLMs have shown newly found emergent ability (e.g., **reasoning**).

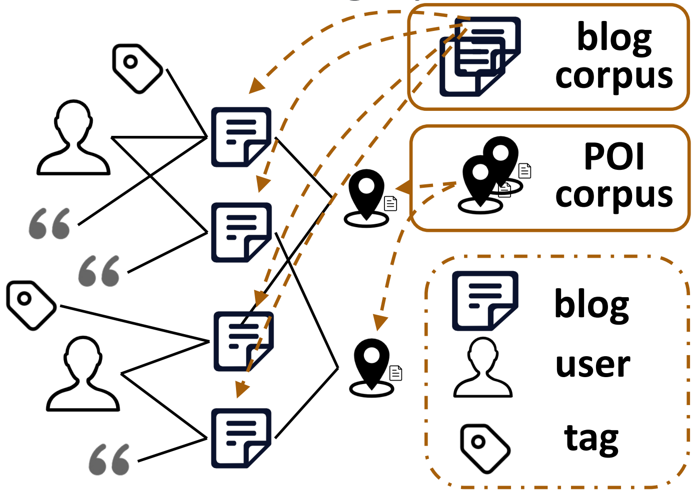


Why LLM on Graphs?

- In real world, text and graph usually appears simultaneously.
  - Text data are associated with rich structure information in the form of graphs.

Academic Network

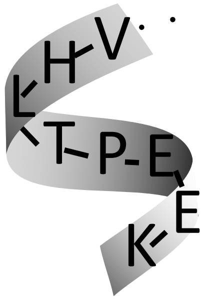




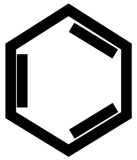
Social Media Network

  - Graph data are captioned with rich textual information.

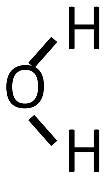
Protein Graphs



"Myoglobin holds oxygen in muscles."



"Benzene is toxic"

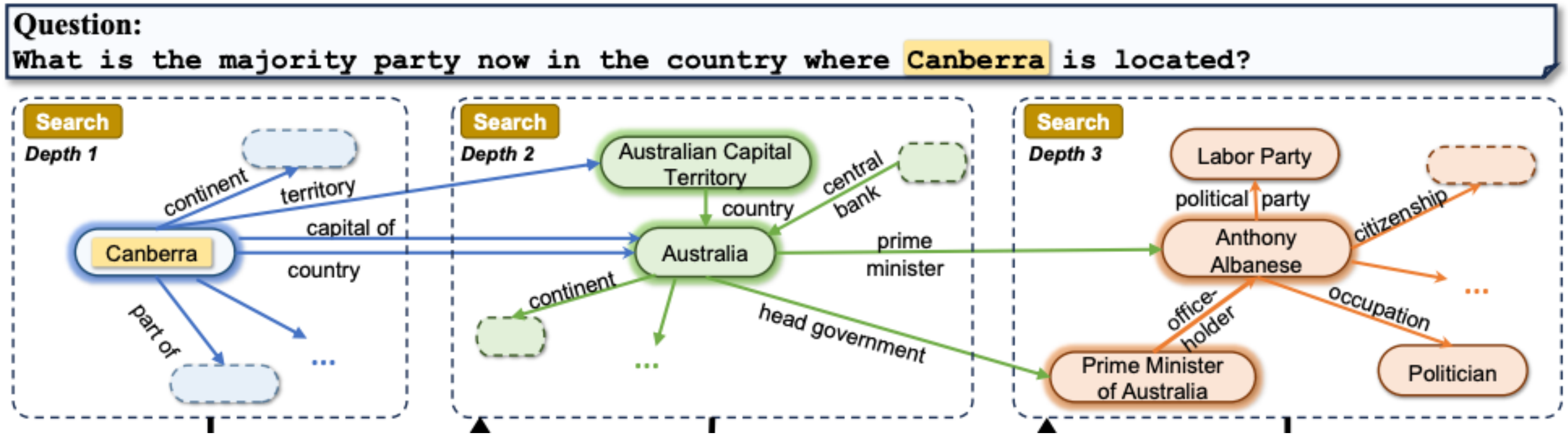


"Water is less toxic"

Molecule Graphs

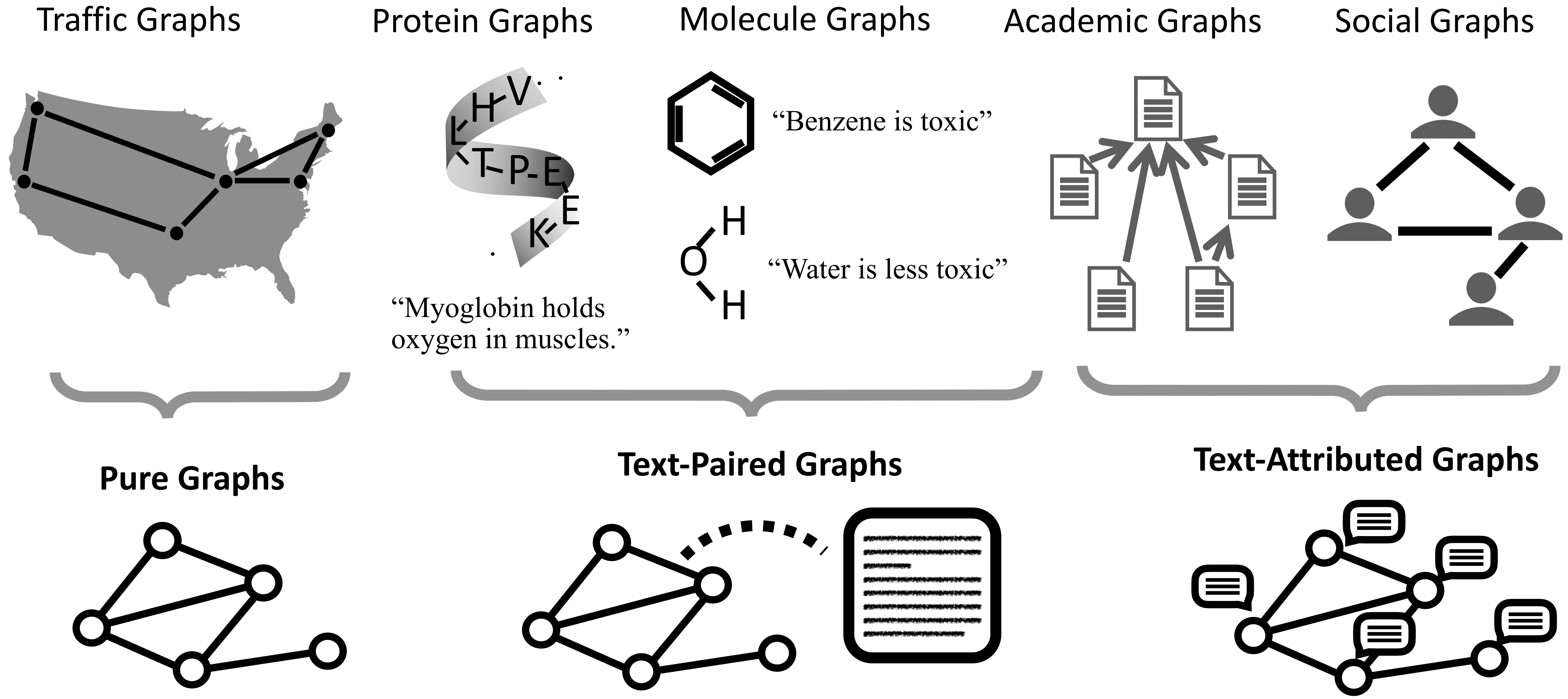
Why LLM on Graphs?

- Although LLMs have shown their pure text-based reasoning ability, it is underexplored whether such ability can be generalized to graph scenarios (i.e., graph-based reasoning)



Graphs

- Three main scenarios



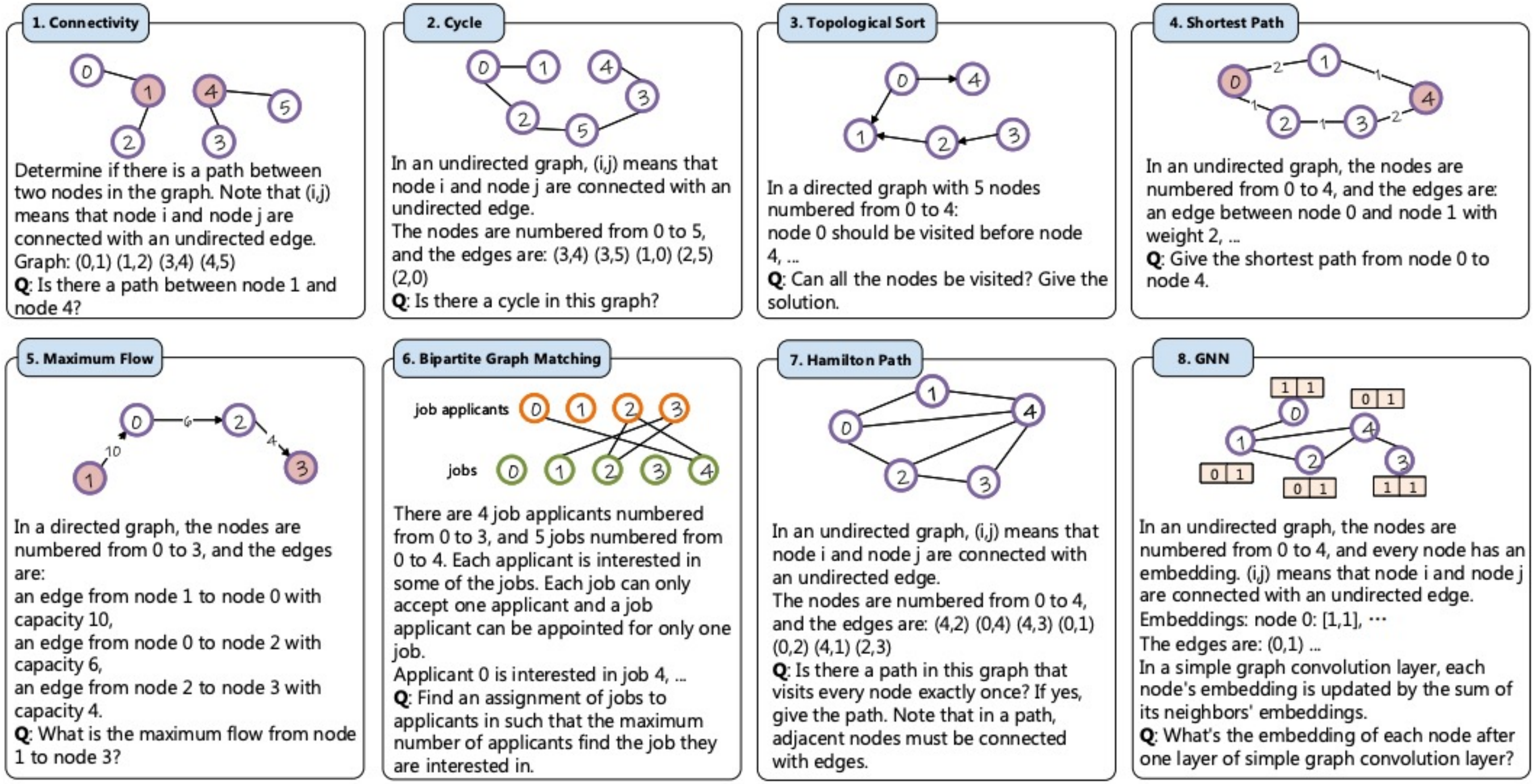
Jin, et al. Large Language Model on Graphs: A Comprehensive Survey. Arxiv. 2023.12.

Outline

- Why Mining Graphs with Large Language Models?
- Mining Pure Graphs with Large Language Models
  - Direct answering: NLGraph (NeurIPs’23)
  - Heuristic reasoning: Think-on-Graph (ICLR’24)
- Mining Text-Attributed Graphs with Large Language Models
- Mining Text-Paired Graphs with Large Language Models

NLGraph

- A comprehensive benchmark to test if LLMs on directly solve graph tasks.



Wang, et al. Can Language Models Solve Graph Problems in Natural Language? NeurIPS 2023.

NLGraph

- LLMs have preliminary graph reasoning ability.

Method	Connectivity				Cycle				Shortest Path				
	Easy	Medium	Hard	Avg.	Easy	Medium	Hard	Avg.	Easy	Hard	Easy (PC)	Hard (PC)	Avg.
RANDOM	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	6.07	6.69	14.73	13.81	17.81
ZERO-SHOT	83.81	72.75	63.38	71.31	50.00	50.00	50.00	50.00	29.40	21.00	46.00	26.76	30.79
FEW-SHOT	93.75	83.83	76.61	84.73	80.00	**70.00**	**61.00**	**70.33**	31.11	26.00	49.19	35.73	35.51
CoT	**94.32**	82.17	77.21	84.57	**84.67**	63.33	53.25	66.75	63.89	**29.50**	76.84	35.79	51.51
0-CoT	79.55	65.83	68.53	71.30	55.33	57.67	49.00	54.00	8.89	7.50	62.39	**43.95**	32.03
CoT+SC	93.18	**84.50**	**82.79**	**86.82**	82.00	63.67	53.50	66.39	**68.89**	29.00	**80.25**	38.47	**54.15**

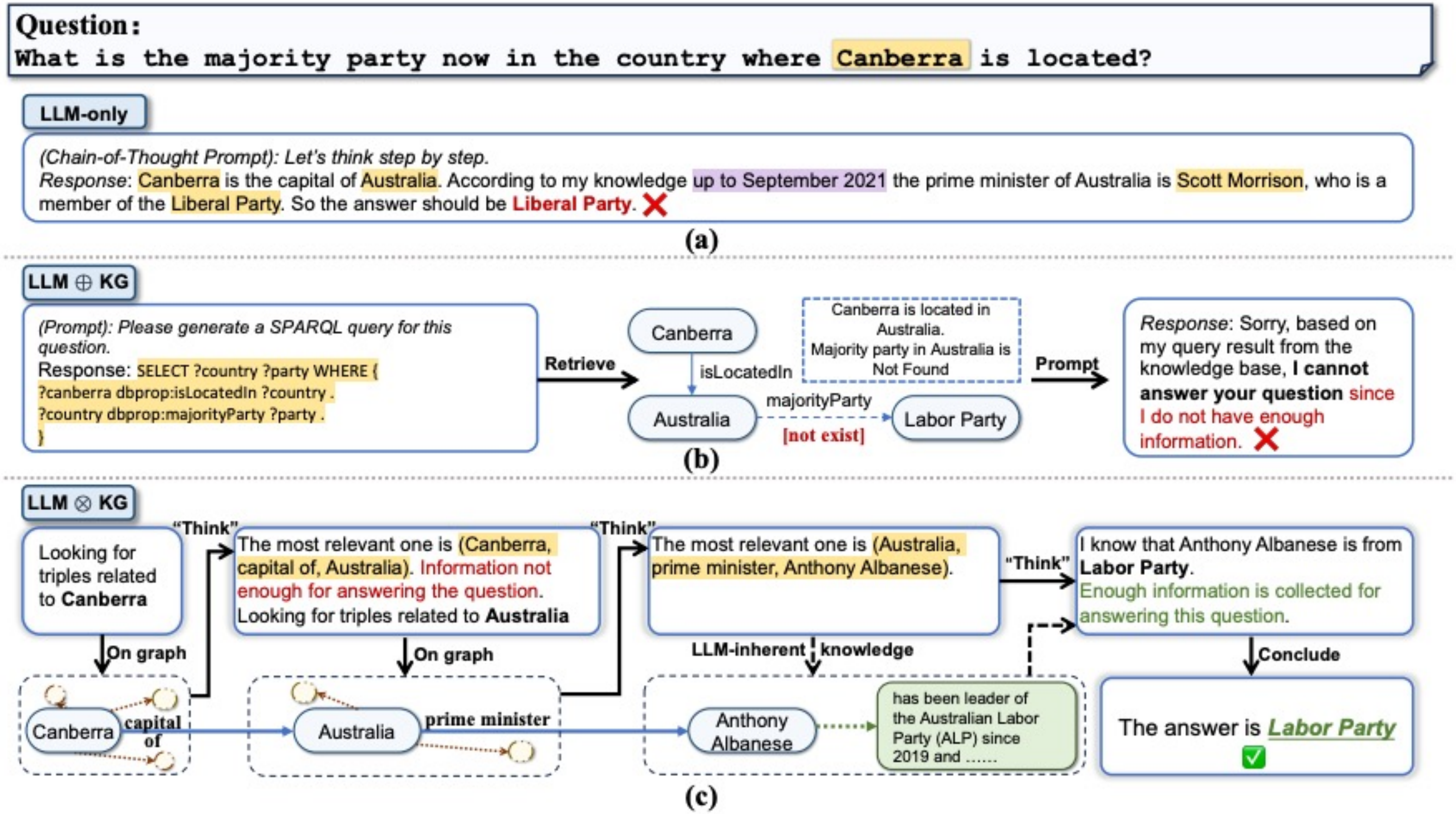
- LLMs are (un)surprisingly brittle.
  - LLMs indeed rely on spurious correlations in problem settings.

Dataset	ZERO-SHOT	FEW-SHOT	CoT	0-CoT	CoT+SC	Avg.
GENERAL	74.67	83.33	85.33	66.00	82.67	78.40
CHAIN	51.67 (-23.00)	45.00 (-35.33)	40.83 (-44.50)	92.50 (+26.50)	44.17 (-38.50)	54.83 (-23.57)
CLIQUE	60.83 (-13.84)	73.33 (-10.00)	85.00 (-0.33)	52.50 (-13.50)	83.33 (+0.66)	71.00 (-7.40)

Wang, et al. Can Language Models Solve Graph Problems in Natural Language? NeurIPS 2023.

Think-on-Graph

- Encourage LLMs to reason on graphs



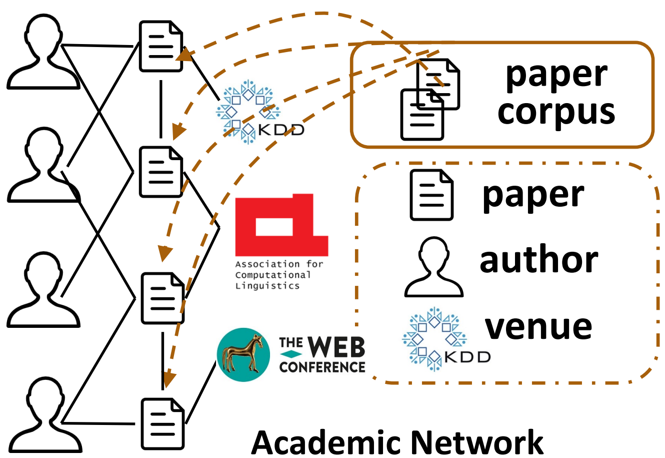
Sun, et al. Think-on-Graph: Deep and Responsible Reasoning of Large Language Model on Knowledge Graph. ICLR 2024.

Outline

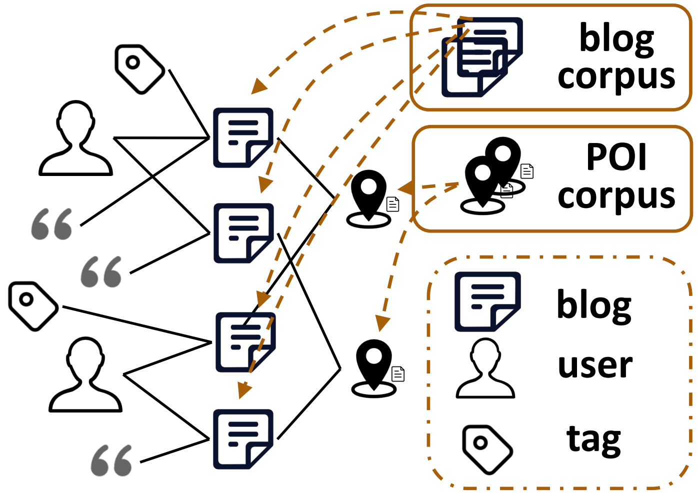
- Why Mining Graphs with Large Language Models?
- Mining Pure Graphs with Large Language Models
- Mining Text-Attributed Graphs with Large Language Models
  - Model architecture – representation learning
  - Language Model Pretraining
  - Augment LLM with Graph
- Mining Text-Paired Graphs with Large Language Models

Text-attributed Graph

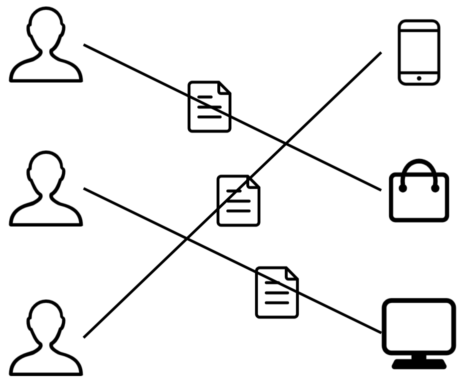
- A graph where nodes/edges are associated with rich text information.



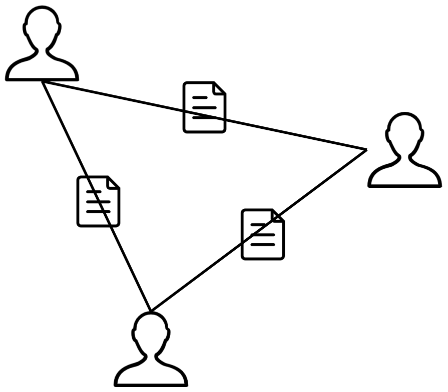
Academic Network



POI Visit Network



E-commerce Network

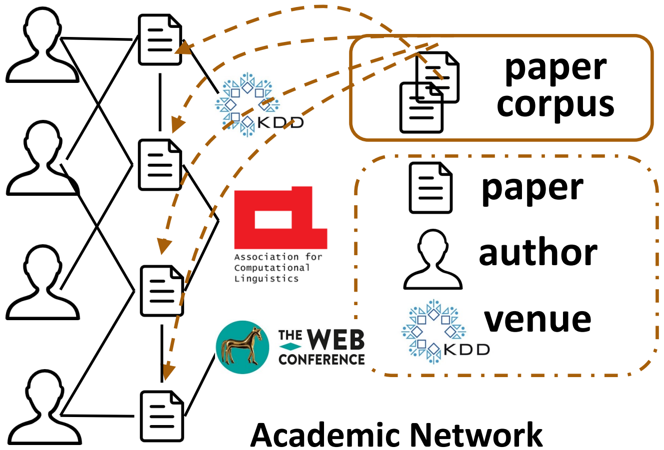


Social Network

MAPLE Benchmark

- A graph where nodes/edges are associated with rich text information.

Field	Paper Source	#Papers	#Labels	#Venues	#Authors	#References
Art	Journal	58,373	1,990	98	54,802	115,343
Philosophy	Journal	59,296	3,758	98	36,619	198,010
Geography	Journal	73,883	3,285	98	157,423	884,632
Business	Journal	84,858	2,392	97	100,525	685,034
Sociology	Journal	90,208	1,935	98	85,793	842,561
History	Journal	113,147	2,689	99	84,529	284,739
Political Science	Journal	115,291	4,990	98	93,393	480,136
Environmental Science	Journal	123,945	694	100	265,728	1,217,268
Economics	Journal	178,670	5,205	97	135,247	1,042,253
Engineering	Journal	270,006	10,683	100	430,046	1,867,276
Psychology	Journal	372,954	7,641	100	460,123	2,313,701
Computer Science	Conference	263,393	13,613	75	331,582	1,084,440
	Journal	410,603	15,540	96	634,506	2,751,996
Geology	Journal	431,834	7,883	100	471,216	1,753,762
Mathematics	Journal	490,551	14,271	98	404,066	2,150,584
Materials Science	Journal	1,337,731	6,802	99	1,904,549	5,457,773
Physics	Journal	1,369,983	16,664	91	1,392,070	3,641,761
Biology	Journal	1,588,778	64,267	100	2,730,547	7,086,131
Chemistry	Journal	1,849,956	35,538	100	2,721,253	8,637,438
Medicine	Journal	2,646,105	36,619	100	4,345,385	7,405,779



Academic Network

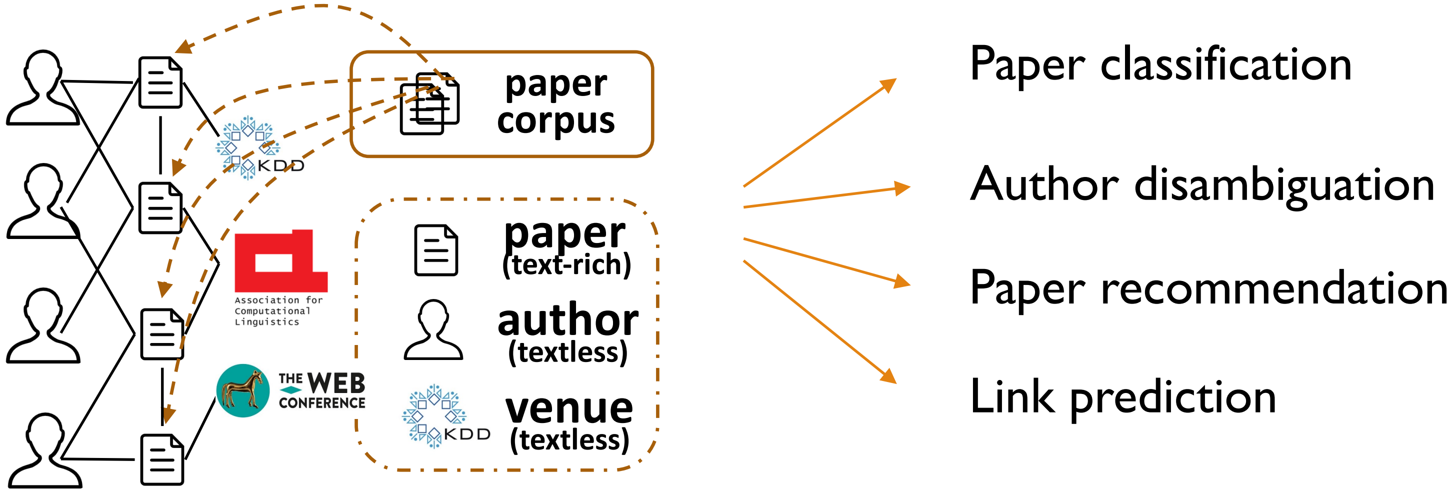
Zhang, et al. The Effect of Metadata on Scientific Literature Tagging: A Cross-Field Cross-Model Study. WWW 2023.

Outline

- Why Mining Graphs with Large Language Models?
- Mining Pure Graphs with Large Language Models
- Mining Text-Attributed Graphs with Large Language Models
  - Model architecture – representation learning
    - LLM-only: MICoL (WWW'22), METAG (arxiv'24)
    - GNN-cascaded LLM: GLEM (ICLR'23)
    - Graph-empowered LLM: GraphFormers (NeurIPs'21), Heterformer (KDD'23), Edgeformers (ICLR'23)
  - Language Model Pretraining
  - Augment LLM with Graph
- Mining Text-Paired Graphs with Large Language Models

Representation learning on text-attributed graphs

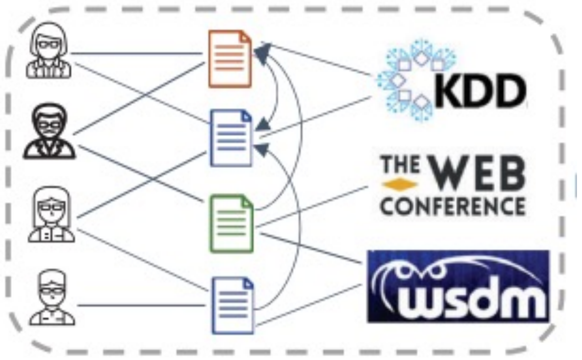
- Given a text-attributed network, people are interested in various tasks.
  - Node classification, link prediction, and node clustering.
  - E.g., academic network
    - Automatically classify each paper.
    - Find the authors of a new paper.
    - Provide paper recommendation.



Representation learning on text-attributed graphs

- Given a text-attributed network, people are interested in various tasks.
  - Node classification, link prediction, and node clustering.

- Learn representations for nodes/edges which can be utilized in various tasks.
  - Textual information & structure information

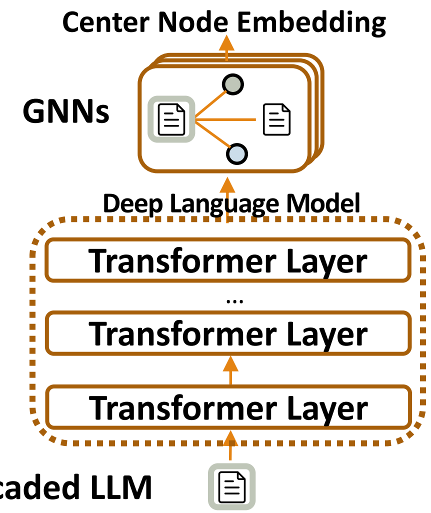


$\mathrm{f}(x)$ Model → Z Embeddings → $g(z)$ → R: Clustering, Classification, Link prediction

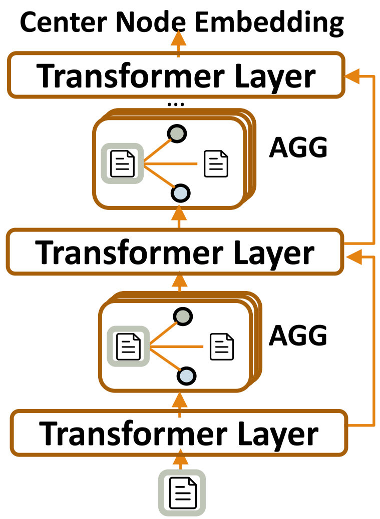
$$X \xrightarrow{\mathrm{f}(x)} Z \xrightarrow{g(z)} R$$

Model Architecture

- LLM-only
  - Finetune LLM with signal from graphs.
  - For example, MICoL (meta-path indictive), METAG (multiplex neighbors).
- GNN-cascaded LLM
  - LLM text encoding -> GNN graph encoding.
  - For example, TextGNN, GLEM.
- Graph-empowered LLM
  - Joint model for text & graph encoding.
  - For example, GraphFormers, Edgeformers, Heterformer.



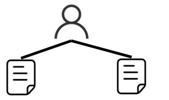
GNN-cascaded LLM



Graph-Empowered LLM

LLM-only: MICoL

- Observation
  - Although hard to know "what is what", network can provide signals on "what is similar to what"
  - E.g., papers written by the **same author** can share similar fine-grained topics



  - E.g., papers published in the **same venue** can share similar coarse-grained topics

- Meta-path
  - A meta-path is a path $\mathcal{M}$ defined on the graph $T_G = (\mathcal{T_V}, \mathcal{T_E})$, and is denoted in the form of $\mathcal{M} = V_1 \xrightarrow{E_1} V_2 \xrightarrow{E_2} \cdots \xrightarrow{E_{m-1}} V_m$ where $V_1, \dots, V_m$ are node types and $E_1, \dots, E_{m-1}$ are edge types.

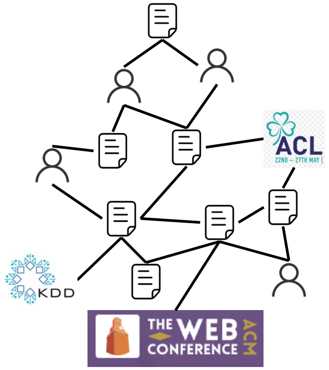
(a) meta-path: PAP

(b) meta-path: P->P<-P

- Meta-graph
  - A meta-graph is a directed acyclic graph (DAG) $\mathcal{M}$ defined on . It has single source node V_1 and a single target node V_m .

(c) meta-graph: P(AV)P

(d) meta-graph: P<-(PP)->P



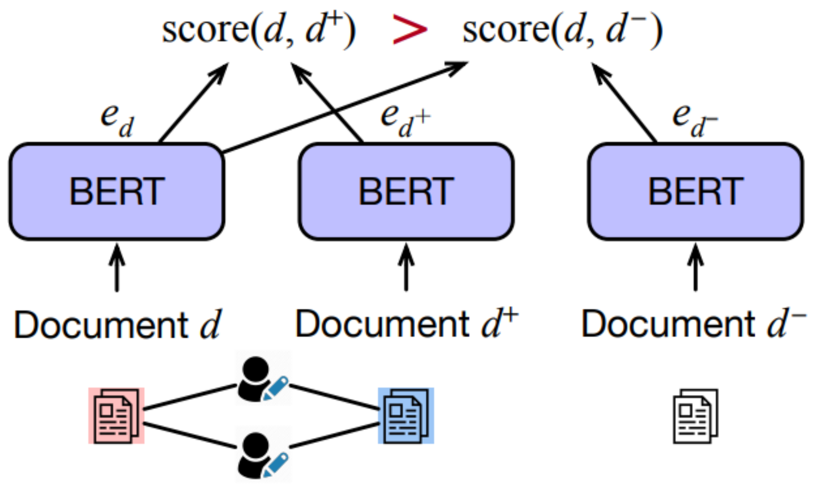
Zhang, et al. Metadata-Induced Contrastive Learning for Zero-Shot Multi-Label Text Classification. WWW 2022.

LLM-only: MICoL

- Two papers connected via a certain meta-path/meta-graph should be more similar than two randomly selected papers.

Bi-Encoder

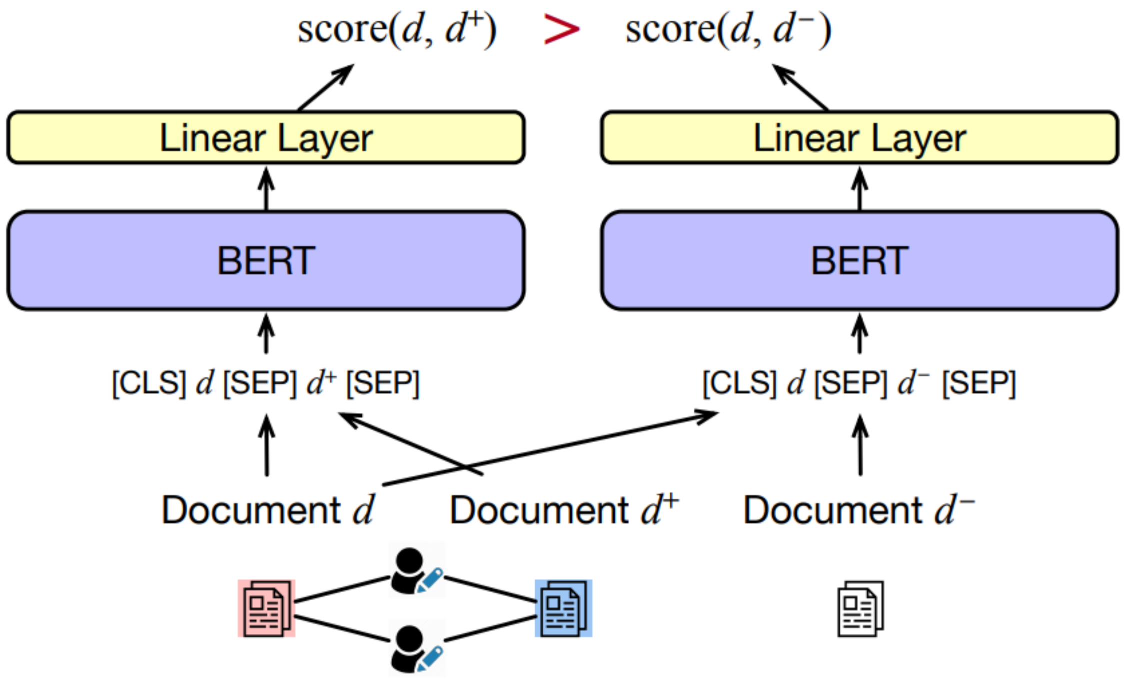
Should be larger Should be smaller



$$-\log \frac{\exp(\cos(\mathbf{e}_d, \mathbf{e}_{d^+})/\tau)}{\exp(\cos(\mathbf{e}_d, \mathbf{e}_{d^+})/\tau) + \sum_{i=1}^{N} \exp(\cos(\mathbf{e}_d, \mathbf{e}_{d_i^-})/\tau)}$$

Cross-Encoder

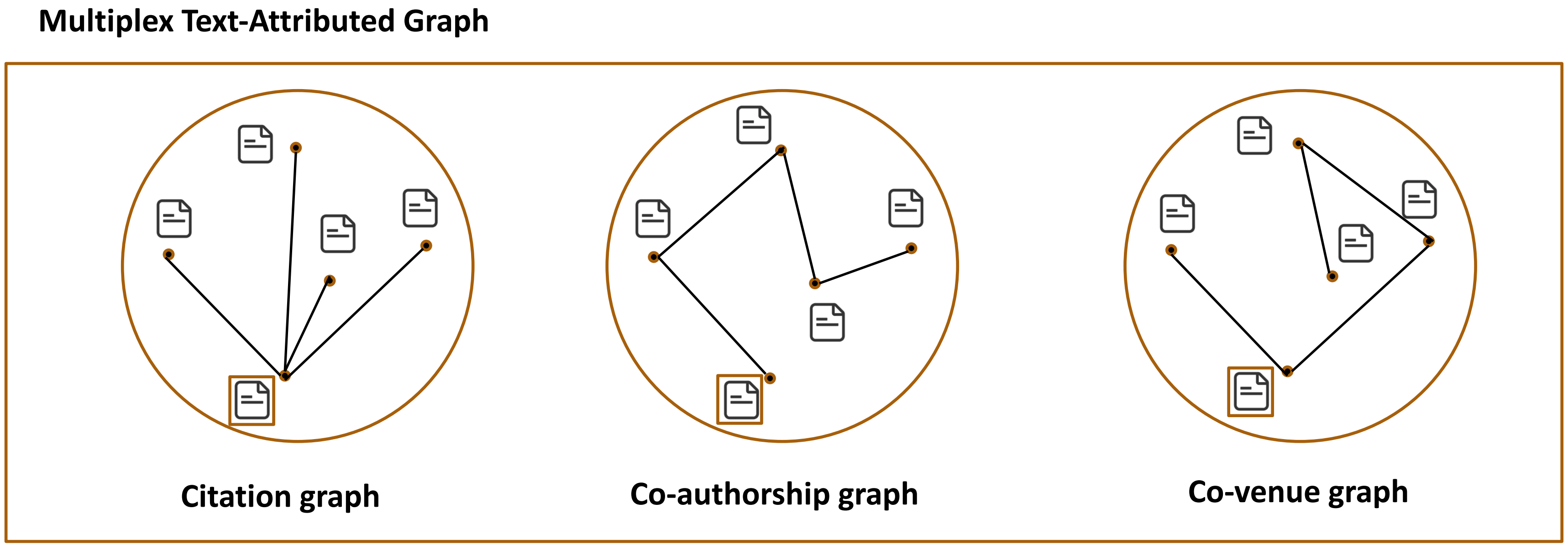
Should be larger Should be smaller



Zhang, et al. Metadata-Induced Contrastive Learning for Zero-Shot Multi-Label Text Classification. WWW 2022.

LLM-only: METAG

- Texts in the real world are often interconnected by multiple types of semantic relations
  - "same-venue" relations edges between papers -> sharing coarse-grained topics
  - "cited-by" relations edges between papers -> sharing fine-grained topics



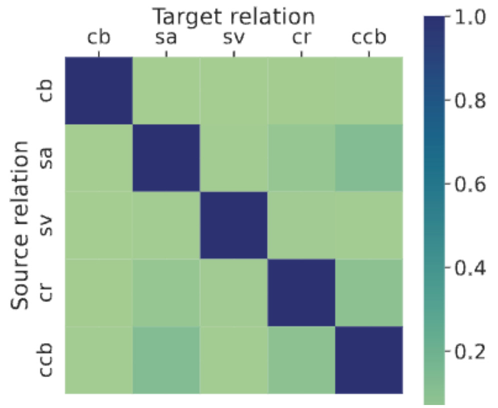
Jin, et al. Learning Multiplex Representations on Text-Attributed Graphs with one Language Model Encoder. Arxiv 2023.

LLM-only: METAG

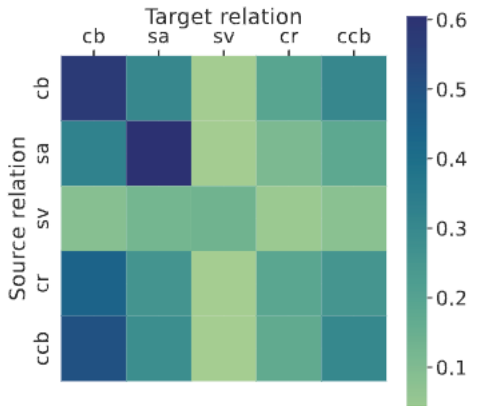
- Existing PLM-based methods: learn a single vector for each text unit
  - Assumption: the semantics of different relations between text units are largely analogous

$$P(e_{ij}|v_i, v_j) \propto \text{Sim}(\boldsymbol{h}_{v_i}, \boldsymbol{h}_{v_j}) \qquad P_{r_k}(e_{ij}|v_i, v_j) \approx P(e_{ij}|v_i, v_j) \approx P_{r_l}(e_{ij}|v_i, v_j)$$

- This assumption does not hold for multiplex text-attributed graphs
  - Semantic distribution shift exists across different relations $P_{r_k}(e_{ij}|v_i, v_j) \neq P_{r_l}(e_{ij}|v_i, v_j)$



Data distribution shift

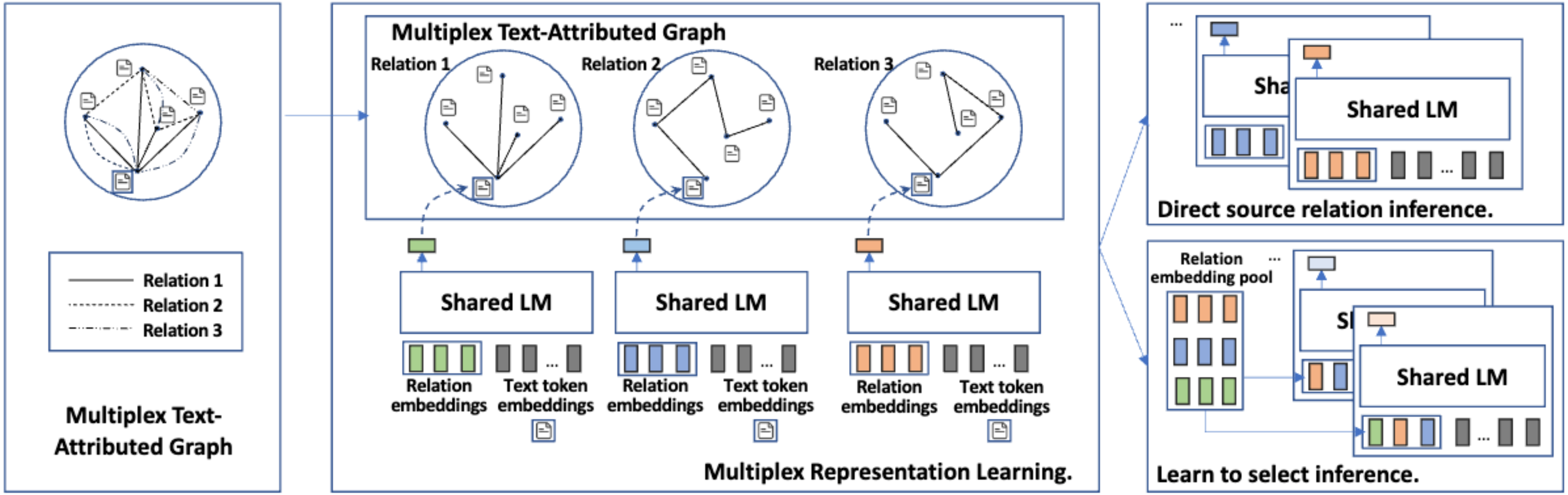


Model performance shift

Jin, et al. Learning Multiplex Representations on Text-Attributed Graphs with one Language Model Encoder. Arxiv 2023.

LLM-only: METAG

- Framework overview



Jin, et al. Learning Multiplex Representations on Text-Attributed Graphs with one Language Model Encoder. Arxiv 2023.

LLM-only: METAG

❑ Multiplex representation learning

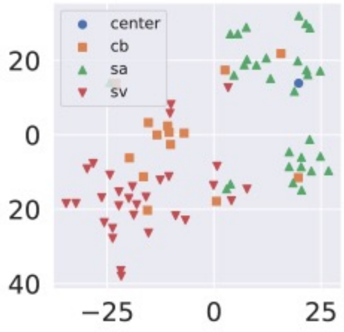
Table 1: Multiplex representation learning experiments on academic networks: Geology and Mathematics. cb, sa, sv, cr, and ccb represent "`cited-by`", "`same-author`", "`same-venue`", "`co-reference`", and "`co-cited-by`" relation respectively.

	Geology						Mathematics					
Model	cb	sa	sv	cr	ccb	Avg.	cb	sa	sv	cr	ccb	Avg.
SPECTER	12.84	12.89	1.5	5.56	9.1	8.38	28.74	23.55	2.39	15.96	25.59	19.25
SciNCL	15.91	14.3	1.57	6.41	10.4	9.72	36.14	26.41	2.83	19.82	30.69	23.18
MPNet-v2	30.87	20.94	1.94	10.36	17.16	16.25	46.12	29.92	3.11	23.60	36.42	27.83
OpenAI-ada-002	30.39	21.08	2.02	16.57	16.69	17.35	39.86	27.22	2.67	19.81	31.62	24.24
DMGI	28.99	27.79	4.91	9.86	16.32	17.58	46.55	42.62	6.11	27.80	38.87	28.85
HDMI	37.89	34.87	3.63	11.32	19.55	21.45	52.65	52.71	5.54	31.80	42.54	37.05
Vanilla FT	54.42	43.20	5.95	18.48	29.93	30.40	75.03	63.46	8.71	44.76	59.94	50.38
MTDNN	58.40	52.50	10.60	19.81	31.61	34.58	78.18	71.04	12.90	47.39	61.75	54.25
Ours	**60.33**	**55.55**	**12.30**	**20.71**	**32.92**	**36.36**	**79.40**	**72.51**	**14.03**	**47.81**	**62.24**	**55.20**

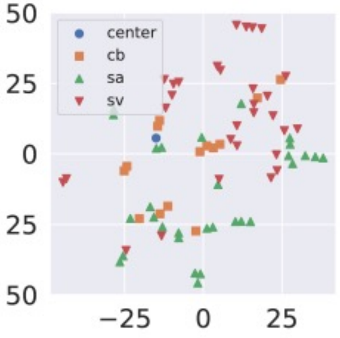
Table 2: Multiplex representation learning experiments on e-commerce networks: Clothes, Home, and Sports. cop, cov, bt, and cob represent "`co-purchased`", "`co-viewed`", "`bought-together`", and "`co-brand`" relation respectively.

	Clothes					Home					Sports				
Model	cop	cov	bt	cob	Avg.	cop	cov	bt	cob	Avg.	cop	cov	bt	cob	Avg.
MPNet-v2	55.89	60.92	59.75	39.12	53.92	52.02	61.83	62.04	38.10	53.50	41.60	64.61	49.82	40.61	49.16
OpenAI-ada-002	65.30	70.87	69.44	48.32	63.48	60.99	71.43	71.36	47.86	62.91	50.80	73.70	60.20	54.06	59.69
DMGI	56.10	52.96	58.46	30.88	49.60	48.27	52.74	57.90	48.81	51.93	41.37	46.27	41.24	31.92	40.20
HDMI	62.85	63.00	69.69	52.50	62.01	51.75	57.91	57.91	53.39	55.24	45.43	61.22	55.56	52.66	53.72
Vanilla FT	81.57	80.46	88.52	67.38	79.48	**73.72**	75.49	85.80	76.83	77.96	**68.22**	77.11	80.78	78.46	76.14
MTDNN	80.30	78.75	87.58	65.94	78.14	72.49	75.17	84.00	77.29	77.24	66.20	76.50	79.72	78.69	75.28
Ours	**82.04**	**81.18**	**88.90**	**68.34**	**80.12**	73.59	**79.06**	**86.58**	**80.07**	**79.83**	67.92	**79.85**	**81.52**	**81.54**	**77.71**

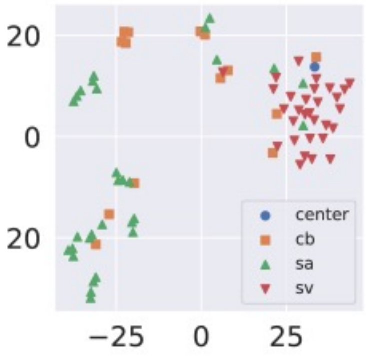
❑ Embedding visualization



(b) $\boldsymbol{h}_{v|r_{sa}}$



(a) $\boldsymbol{h}_{v|r_{cb}}$

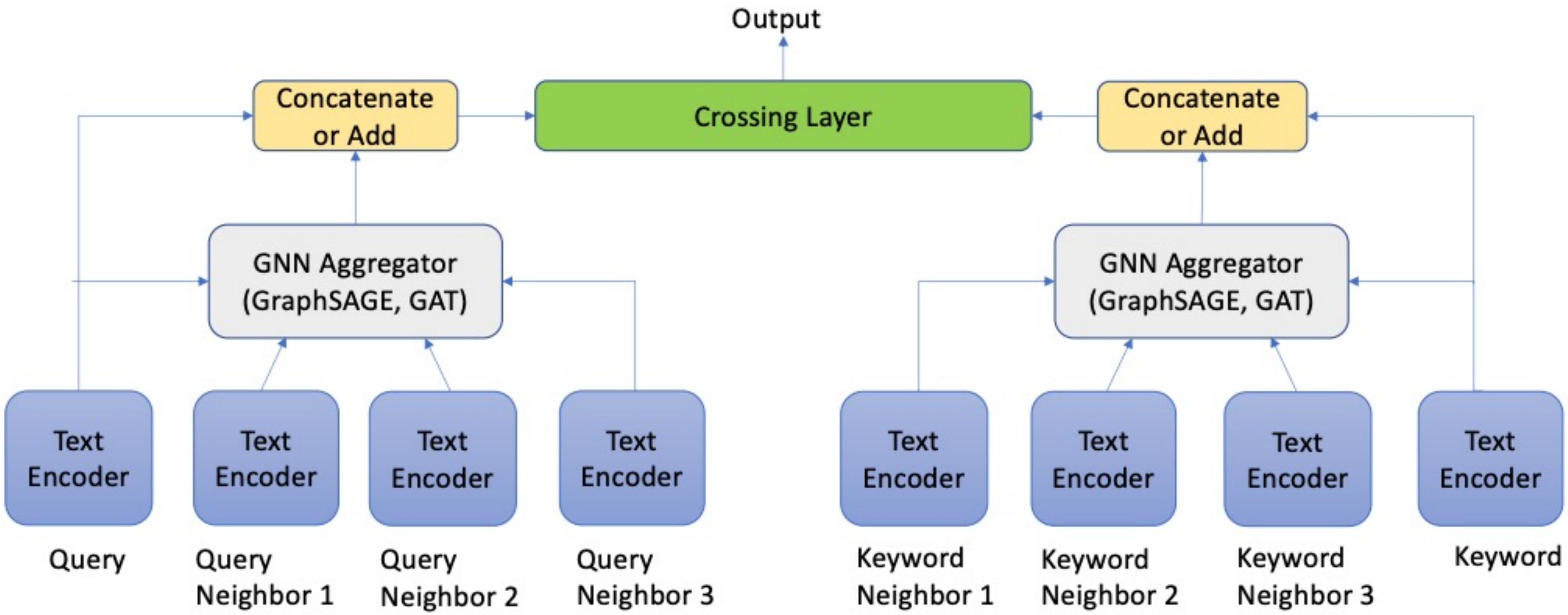


(c) $\boldsymbol{h}_{v|r_{sv}}$

Jin, et al. Learning Multiplex Representations on Text-Attributed Graphs with one Language Model Encoder. Arxiv 2023.

GNN-cascaded LLM: TextGNN

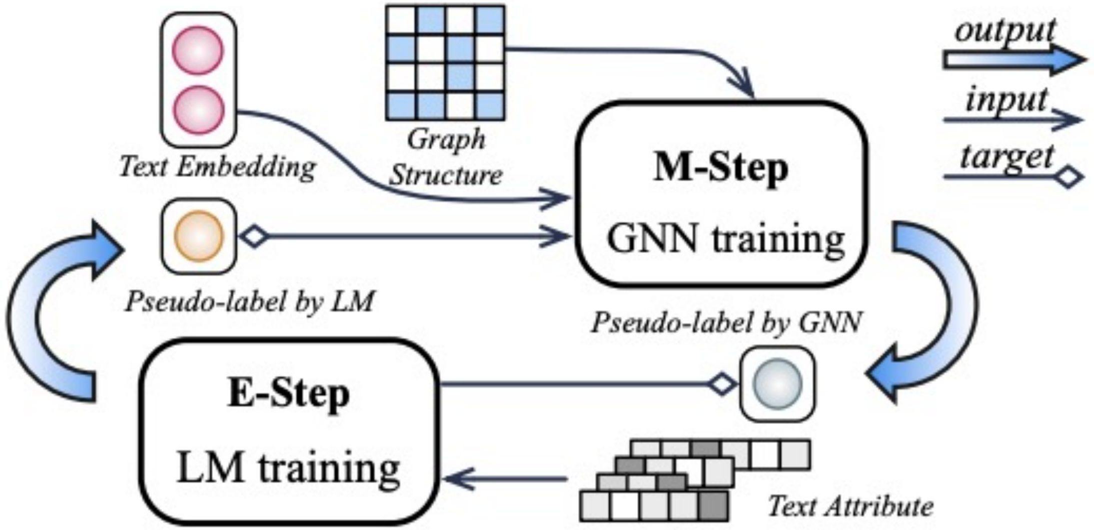
- LLM (text encoding) -> GNN (graph aggregation)
  - LLM & GNN are optimized simultaneously



Zhu, et al. TextGNN: Improving Text Encoder via Graph Neural Network in Sponsored Search. WWW 2021.

GNN-cascaded LLM: GLEM

- Iteratively optimize LLM & GNN

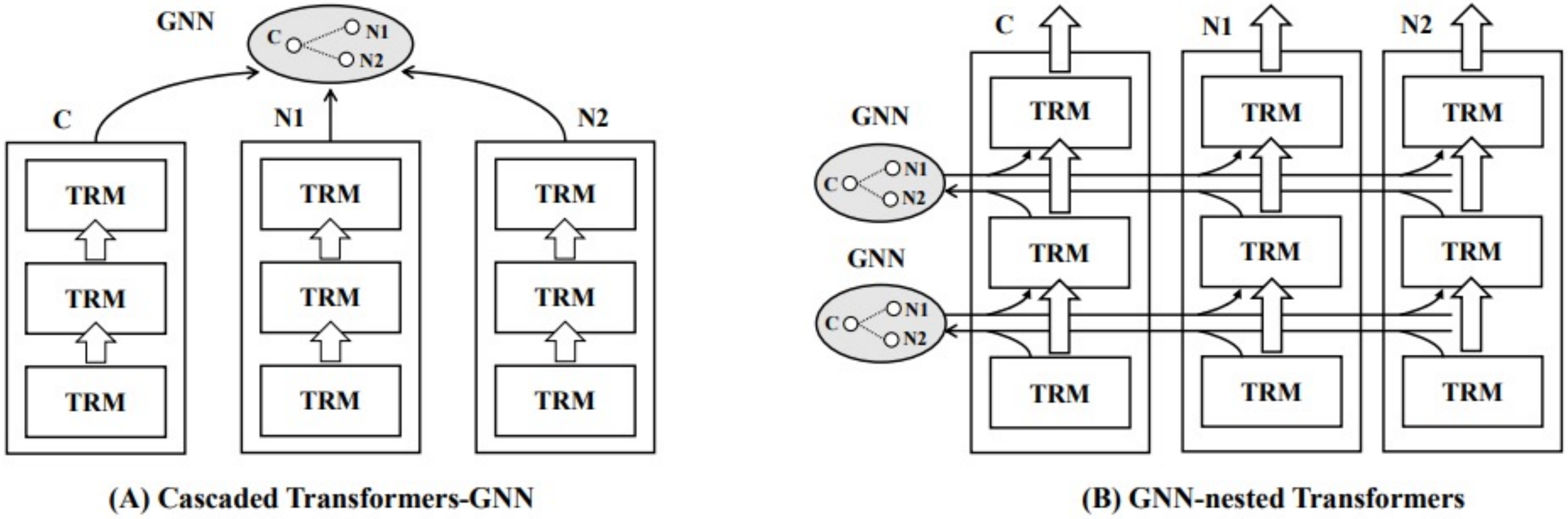


Zhao, et al. Learning on Large-scale Text-attributed Graphs via Variational Inference. ICLR 2023.

Graph-Empowered LLM: GraphFormers

- Learning on homogeneous text-attributed graphs.
  - Nodes are associated with textual information.
  - There is only one type of node and edge.

- Put GNNs in between Transformer layers



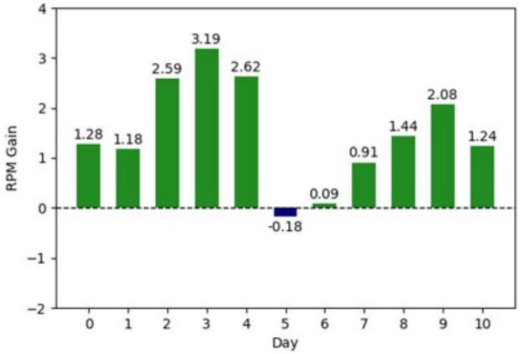
Yang, et al. GraphFormers: GNN-nested Transformers for Representation Learning on Textual Graph. NeurIPs'21.

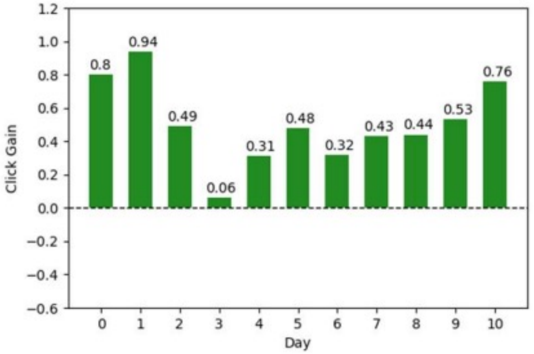
Graph-Empowered LLM: GraphFormers

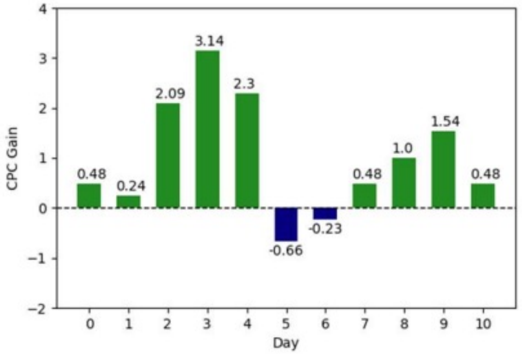
- Link prediction

	Product			DBLP			Wiki		
Methods	P@1	NDCG	MRR	P@1	NDCG	MRR	P@1	NDCG	MRR
PLM	0.6563	0.7911	0.7344	0.5673	0.7484	0.6777	0.3466	0.5799	0.4712
TNVE	0.4618	0.6204	0.5364	0.2978	0.5295	0.4163	0.1786	0.4274	0.2933
IFTN	0.5233	0.6740	0.5982	0.3691	0.5798	0.4773	0.1838	0.4276	0.2945
PLM+GAT	0.7540	0.8637	0.8232	0.6633	0.8204	0.7667	0.3006	0.5430	0.4270
PLM+Max	0.7570	0.8678	0.8280	0.6934	0.8386	0.7900	0.3712	0.6071	0.5022
PLM+Mean	0.7550	0.8671	0.8271	0.6896	0.8359	0.7866	0.3664	0.6037	0.4980
PLM+Att	0.7513	0.8652	0.8246	0.6910	0.8366	0.7875	0.3709	0.6067	0.5018
GraphFormers	**0.7786**	**0.8793**	**0.8430**	**0.7267**	**0.8565**	**0.8133**	**0.3952**	**0.6230**	**0.5220**

- Online A/B test



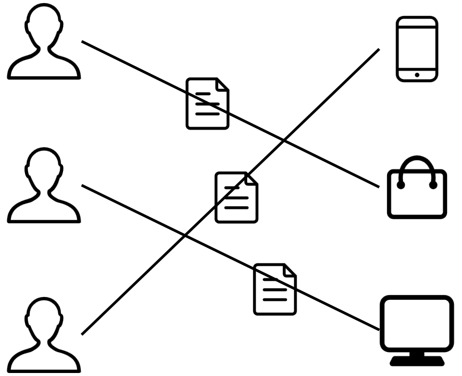




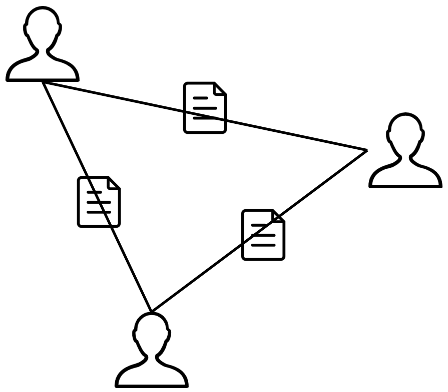
Yang, et al. GraphFormers: GNN-nested Transformers for Representation Learning on Textual Graph. NeurIPs’21.

Graph-Empowered LLM: Edgeformers

- Learning on textual-edge graphs.
  - E.g., user-review-item network, social network
  - Link prediction, edge classification, node classification, etc.



E-commerce Network



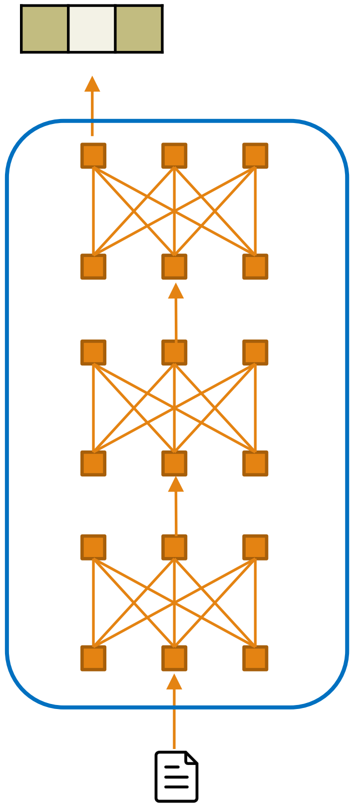
Social Network

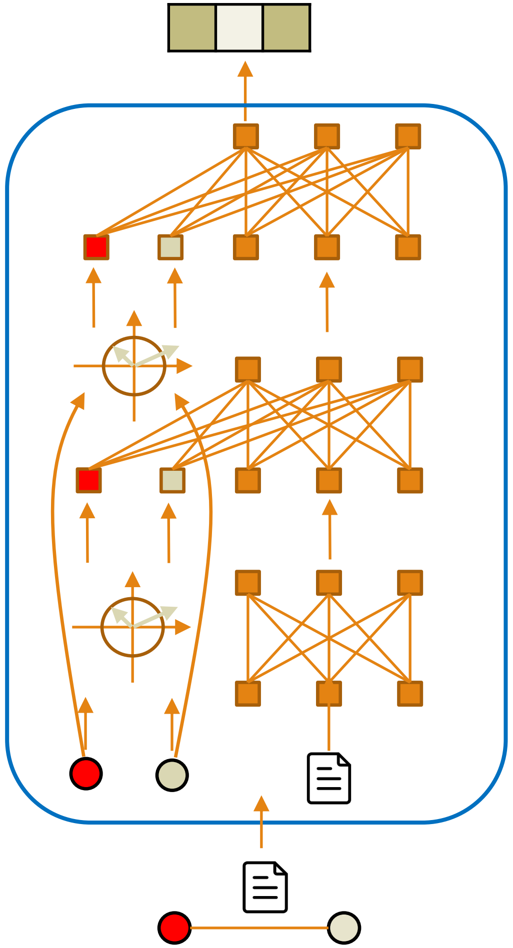
Jin, et al. Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks. ICLR'23.

Graph-Empowered LLM: Edgeformers

- Edge representation learning (Edgeformer-E)
  - Network-aware edge encoding with virtual node tokens.

Transformers



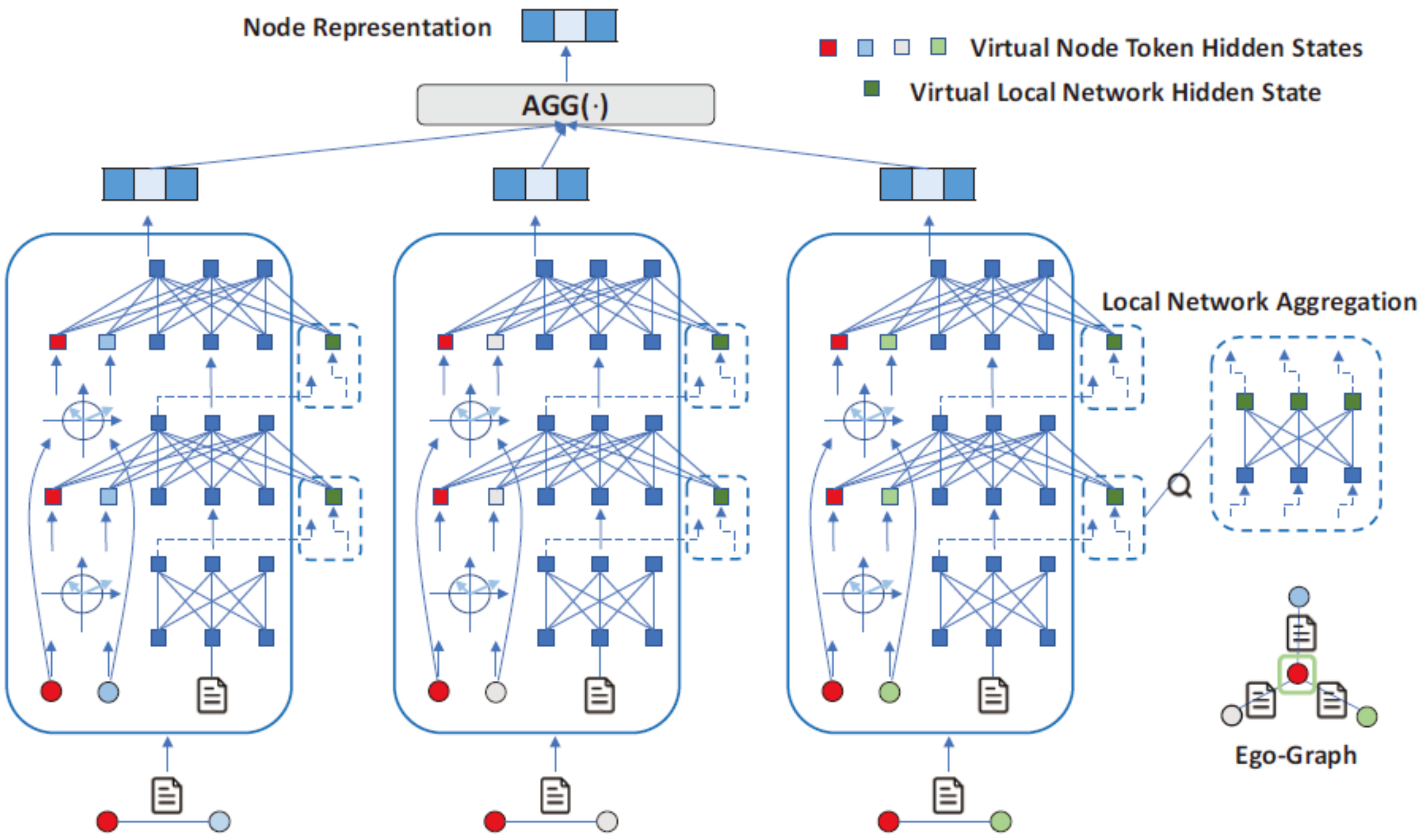


Edgeformer-E

Jin, et al. Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks. ICLR'23.

Graph-Empowered LLM: Edgeformers

- Text-aware node representation learning (Edgeformer-N)
  - Aggregate edge representations
  - Enhance edge representations with node's local network structure



Jin, et al. Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks. ICLR'23.

Graph-Empowered LLM: Edgeformers

❑ Edge classification

Table 7: Edge classification performance on Amazon-Movie, Amazon-App, Goodreads-Crime, and Goodreads-Children.

Model	Amazon-Movie Macro-F1	Amazon-Movie Micro-F1	Amazon-Apps Macro-F1	Amazon-Apps Micro-F1	Goodreads-Crime Macro-F1	Goodreads-Crime Micro-F1	Goodreads-Children Macro-F1	Goodreads-Children Micro-F1
TF-IDF	50.01	64.22	48.30	62.88	43.07	51.72	39.42	49.90
TF-IDF+nodes	53.59	66.34	50.56	65.08	49.35	57.50	47.32	56.78
EHGNN	49.90	64.04	48.20	63.63	44.49	52.30	40.01	50.23
BERT	61.38	71.36	59.11	69.27	56.41	61.29	51.57	57.72
BERT+nodes	*63.00*	*72.45*	*59.72*	*70.82*	*58.64*	*65.02*	*54.42*	*60.46*
BERT+EHGNN	61.45	70.73	58.86	70.79	56.92	61.66	52.46	57.97
BERT+MaxSAGE	61.57	70.79	58.95	70.45	57.20	61.98	52.75	58.53
GraphFormers	61.73	71.52	59.67	70.19	57.49	62.37	52.93	58.34
Edgeformer-E	**64.18**	**73.59**	**60.67**	**71.28**	**61.03**	**65.86**	**57.45**	**61.71**

❑ Link prediction

Model	Amazon-Movie MRR	Amazon-Movie NDCG	Amazon-Apps MRR	Amazon-Apps NDCG	Goodreads-Crime MRR	Goodreads-Crime NDCG	Goodreads-Children MRR	Goodreads-Children NDCG	StackOverflow MRR	StackOverflow NDCG
MF	0.2032	0.3546	0.1482	0.3052	0.1923	0.3443	0.1137	0.2716	0.1040	0.2642
MeanSAGE	0.2138	0.3657	0.1766	0.3343	0.1832	0.3368	0.1066	0.2647	0.1174	0.2768
MaxSAGE	0.2178	0.3694	0.1674	0.3258	0.1846	0.3387	0.1066	0.2647	0.1173	0.2769
GIN	0.2140	0.3648	0.1797	0.3362	0.1846	0.3374	0.1128	0.2700	0.1189	0.2778
CensNet	0.2048	0.3568	0.1894	0.3457	0.1880	0.3398	0.1157	0.2726	0.1235	0.2806
NENN	0.2565	0.4032	0.1996	0.3552	0.2173	0.3670	0.1297	0.2854	0.1257	0.2854
BERT	0.2391	0.3864	0.1790	0.3350	0.1986	0.3498	0.1274	0.2836	0.1666	0.3252
BERT+MaxSAGE	0.2780	0.4224	0.2055	0.3602	0.2193	0.3694	0.1312	0.2872	0.1681	0.3264
BERT+MeanSAGE	0.2491	0.3972	0.1983	0.3540	0.1952	0.3477	0.1223	0.2791	0.1678	0.3264
BERT+GIN	0.2573	0.4037	0.2000	0.3552	0.2007	0.3522	0.1238	0.2801	*0.1708*	*0.3279*
GraphFormers	0.2756	0.4198	0.2066	0.3607	0.2176	0.3684	0.1323	0.2887	0.1693	0.3278
BERT+CensNet	0.1919	0.3462	0.1544	0.3132	0.1437	0.3000	0.0847	0.2436	0.1173	0.2789
BERT+NENN	*0.2821*	*0.4256*	*0.2127*	*0.3666*	*0.2262*	*0.3756*	*0.1365*	*0.2925*	0.1619	0.3215
Edgeformer-N	**0.2919**	**0.4344**	**0.2239**	**0.3771**	**0.2395**	**0.3875**	**0.1446**	**0.3000**	**0.1754**	**0.3339**
+ Δ %	**3.5%**	**2.1%**	**5.3%**	**2.9%**	**5.9%**	**3.2%**	**5.9%**	**2.6%**	**2.7%**	**1.8%**

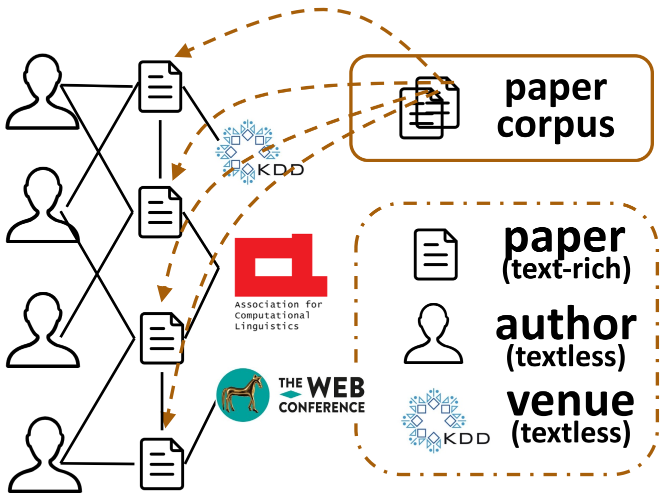
❑ Node classification

Model	Amazon-Movie Macro-F1	Amazon-Movie Micro-F1	Amazon-Movie PREC	Amazon-Apps Macro-F1	Amazon-Apps Micro-F1	Amazon-Apps PREC
MF	0.7566±0.0017	0.8234±0.0013	0.8241±0.0013	0.4647±0.0151	0.8393±0.0012	0.8462±0.0006
CensNet	0.8528±0.0010	0.8839±0.0008	0.8845±0.0007	0.2782±0.0168	0.8279±0.0006	0.8331±0.0005
NENN	0.9186±0.0008	0.9341±0.0008	0.9347±0.0007	0.3408±0.0082	0.8789±0.0019	0.8819±0.0017
BERT	0.9209±0.0005	0.9361±0.0003	0.9367±0.0003	0.7608±0.0175	0.9283±0.0015	0.9337±0.0015
BERT+CensNet	0.9032±0.0006	0.9221±0.0004	0.9227±0.0004	0.5750±0.0277	0.8692±0.0034	0.8731±0.0028
BERT+NENN	0.9247±0.0005	0.9387±0.0004	0.9393±0.0005	0.7556±0.0092	0.9306±0.0008	0.9382±0.0006
Edgeformer-N	**0.9276±0.0007**	**0.9411±0.0006**	**0.9417±0.0005**	**0.7758±0.0100**	**0.9339±0.0007**	**0.9431±0.0005**

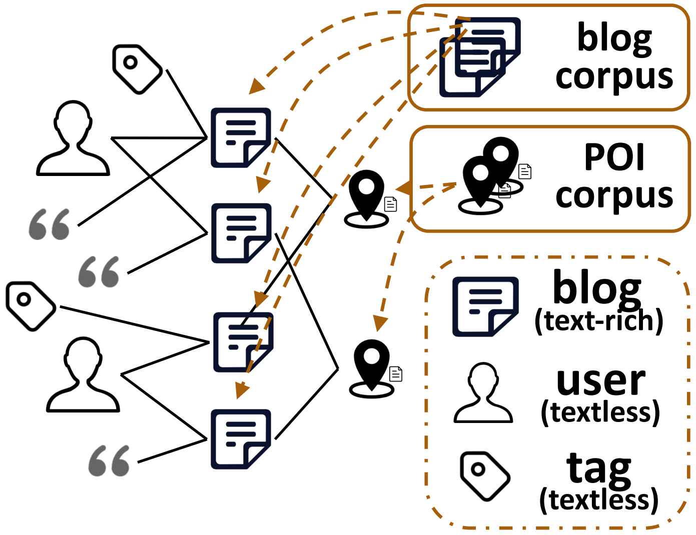
Jin, et al. Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks. ICLR'23.

Graph-Empowered LLM: Heterformer

- Learning on heterogeneous text-attributed graphs.
  - Text-attributed.
  - Heterogeneous: presence or absence of text & diversity of types.
  - E.g., Academic Networks, Social Media Networks



Academic Network

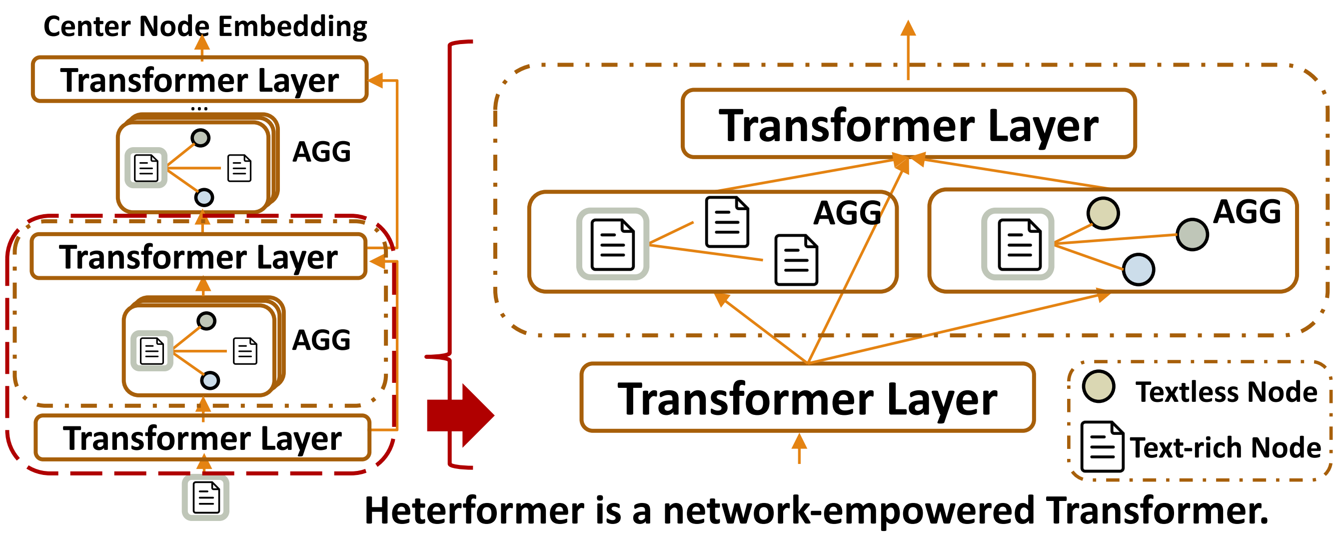


Social Media Network

Jin, et al. Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks. KDD'23.

Graph-Empowered LLM: Heterformer

- Overall framework
  - Heterformer: a graph-empowered Transformer.
  - Unifying text semantic encoding and network signal capturing.

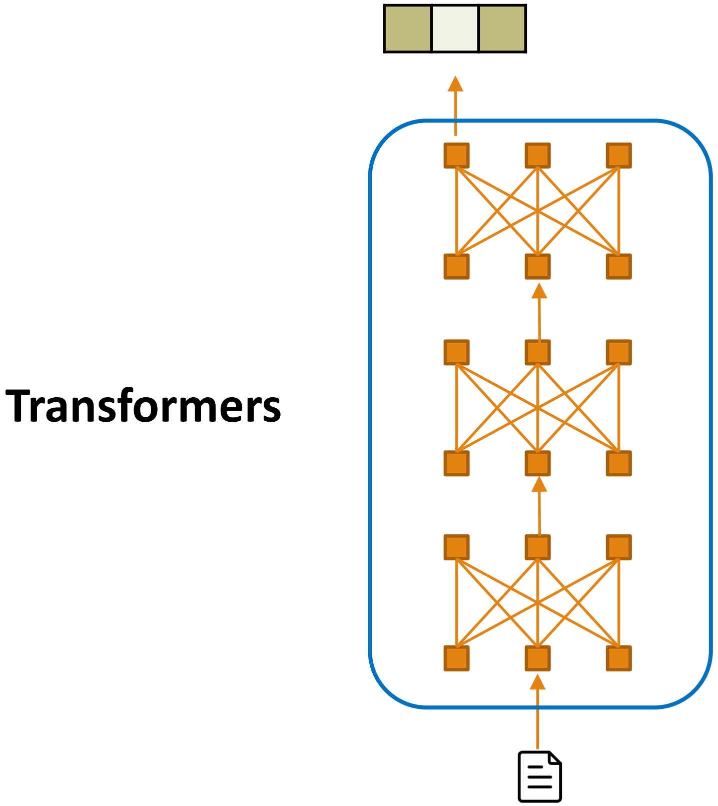


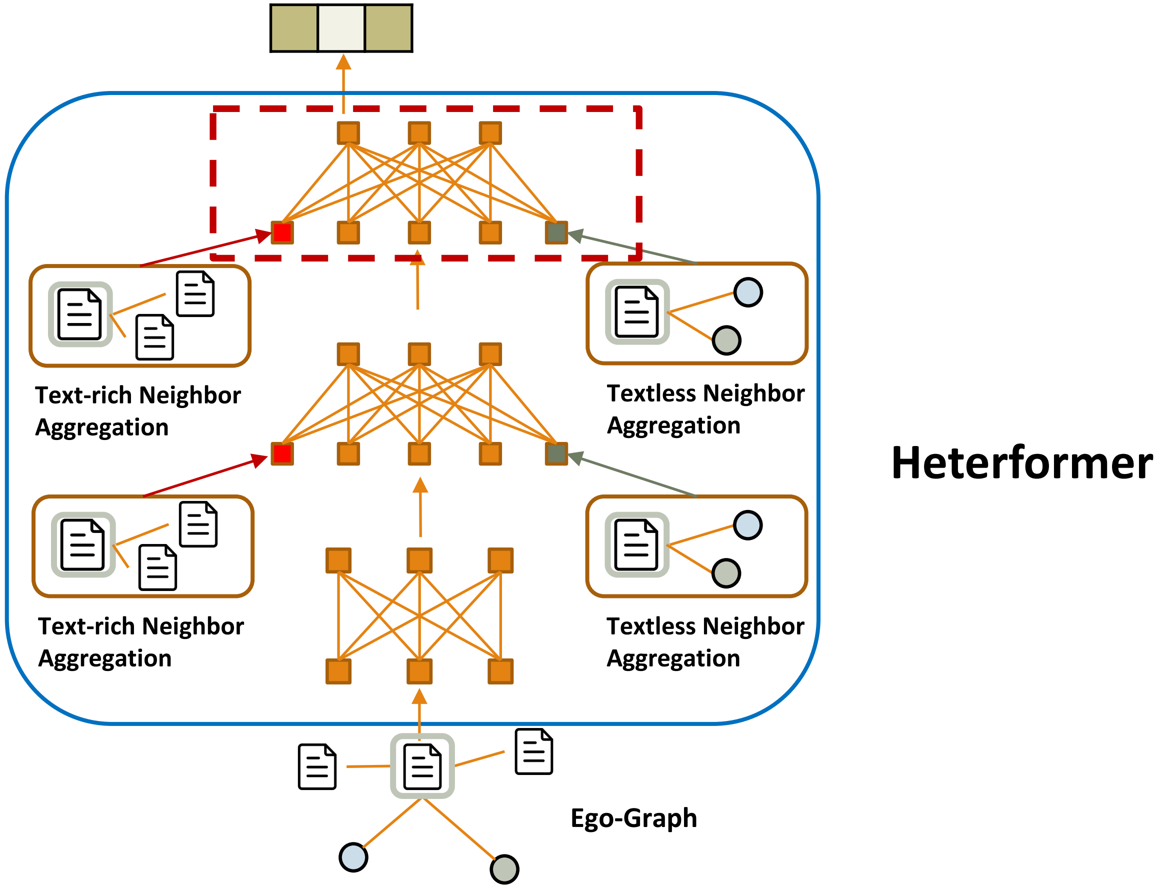
Heterformer is a network-empowered Transformer.

Jin, et al. Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks. KDD'23.

Graph-Empowered LLM: Heterformer

- Text-Rich Node Encoding
  - Network-aware node text encoding with virtual neighbor tokens.
  - Multi-head attention-based heterogeneous neighbor aggregation.





Jin, et al. Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks. KDD'23.

Graph-Empowered LLM: Heterformer

- Textless Node Encoding
  - Node type heterogeneity-based representation

$$\boldsymbol{h}_{v_p}^{(l)} = \boldsymbol{W}_{\phi_i}^{(l)} \boldsymbol{h}_{v_p}^{(0)}, \quad \text{where } \phi(v_p) = \phi_i, \;\; \phi_i \in \mathcal{A}_{\text{TL}}.$$

Node type heterogeneity

  - Textless node embedding warm up
    - A great number of textless nodes will introduce a great number of randomly initialized parameters into the model -> underfitting.
    - Warm up to give textless node embeddings good initializations.

$$\min_{\boldsymbol{h}_{v_p}^{(l)}} \mathcal{L}_w = \sum_{\substack{v_p \in \mathcal{V} \\ \phi(v_p) \in \mathcal{A}_{\text{TL}}}} \sum_{v_u \in \widehat{N}_{v_p}} -\log \frac{\exp(\bar{\boldsymbol{h}}_{v_u}^\top \boldsymbol{h}_{v_p}^{(l)})}{\exp(\bar{\boldsymbol{h}}_{v_u}^\top \boldsymbol{h}_{v_p}^{(l)}) + \sum_{v'_u} \exp(\bar{\boldsymbol{h}}_{v'_u}^\top \boldsymbol{h}_{v_p}^{(l)})},$$

Jin, et al. Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks. KDD'23.

Graph-Empowered LLM: Heterformer

❑ Link prediction

	Method	DBLP PREC	DBLP MRR	DBLP NDCG	Twitter PREC	Twitter MRR	Twitter NDCG	Goodreads PREC	Goodreads MRR	Goodreads NDCG
	MeanSAGE	0.7019	0.7964	0.8437	0.6489	0.7450	0.7991	0.6302	0.7409	0.8001
	BERT	0.7569	0.8340	0.8726	0.7179	0.7833	0.8265	0.5571	0.6668	0.7395
Homo GNN	BERT+MeanSAGE	0.8131	0.8779	0.9070	0.7201	0.7845	0.8275	0.7301	0.8167	0.8594
	BERT+MAXSAGE	0.8193	0.8825	0.9105	0.7198	0.7845	0.8276	0.7280	0.8164	0.8593
	BERT+GAT	0.8119	0.8771	0.9063	0.7231	0.7873	0.8300	0.7333	0.8170	0.8593
	GraphFormers	0.8324	0.8916	0.9175	0.7258	0.7891	0.8312	0.7444	0.8260	0.8665
Hetero GNN	BERT+RGCN	0.7979	0.8633	0.8945	0.7111	0.7764	0.8209	0.7488	0.8303	0.8699
	BERT+HAN	0.8136	0.8782	0.9072	0.7237	0.7880	0.8306	0.7329	0.8174	0.8597
	BERT+HGT	0.8170	0.8814	0.9098	0.7153	0.7800	0.8237	0.7224	0.8112	0.8552
	BERT+SHGN	0.8149	0.8785	0.9074	0.7218	0.7866	0.8295	0.7362	0.8195	0.8613
	GraphFormers++	0.8233	0.8856	0.9130	0.7159	0.7799	0.8236	0.7536	0.8328	0.8717
	Heterformer	**0.8474***	**0.9019***	**0.9255***	**0.7272***	**0.7908***	**0.8328***	**0.7633***	**0.8400***	**0.8773***

❑ Node classification

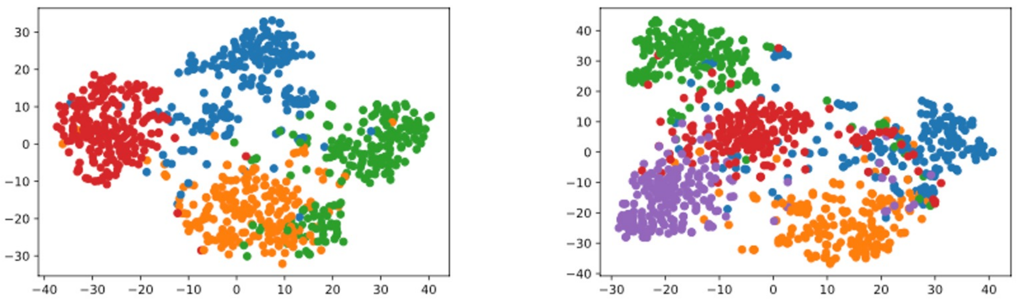
Table 3: Transductive text-rich node classification.

Method	DBLP Micro-F1	DBLP Macro-F1	Goodreads Micro-F1	Goodreads Macro-F1
BERT	0.6119	0.5476	0.8364	0.7713
BERT+MaxSAGE	0.6179	0.5511	0.8447	0.7866
BERT+MeanSAGE	0.6198	0.5522	0.8420	0.7826
BERT+GAT	0.5943	0.5175	0.8328	0.7713
GraphFormers	0.6256	0.5616	0.8388	0.7786
BERT+HAN	0.5965	0.5211	0.8351	0.7747
BERT+HGT	0.6575	0.5951	0.8474	0.7928
BERT+SHGN	0.5982	0.5214	0.8345	0.7737
GraphFormers++	0.6474	0.5790	0.8516	0.7993
Heterformer	**0.6695***	**0.6062***	**0.8578***	**0.8076***

❑ Node clustering

Method	DBLP NMI	DBLP ARI	Goodreads NMI	Goodreads ARI
BERT	0.2570	0.3349	0.2325	0.4013
BERT+MaxSAGE	0.2615	0.3490	0.2205	0.4173
BERT+MeanSAGE	0.2628	0.3488	**0.2449**	**0.4329**
BERT+GAT	0.2598	0.3419	0.2408	0.4185
GraphFormers	0.2633	0.3455	0.2362	0.4139
BERT+HAN	0.2568	0.3401	0.2391	0.4266
BERT+HGT	0.2469	0.3392	0.2427	0.4296
BERT+SHGN	0.2589	0.3431	0.2373	0.4171
GraphFormers++	0.2566	0.3432	0.2372	0.4211
Heterformer	**0.2707***	**0.3639***	0.2429	0.4199

❑ Embedding visualization



(a) DBLP (b) Goodreads

Jin, et al. Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks. KDD'23.

Outline

- Why Mining Graphs with Large Language Models?
- Mining Pure Graphs with Large Language Models
- Mining Text-Attributed Graphs with Large Language Models
  - Model architecture – representation learning
  - Language Model Pretraining: Patton (ACL'23)
  - Augment LLM with Graph
- Mining Text-Paired Graphs with Large Language Models

Why do we need language model pretraining on network?

- Given a text-rich network, people are interested in various downstream tasks
  - Document/node classification, document retrieval and link prediction

- Text-rich network contains rich unsupervised semantic information
  - Alleviate human labeling burden for downstream tasks

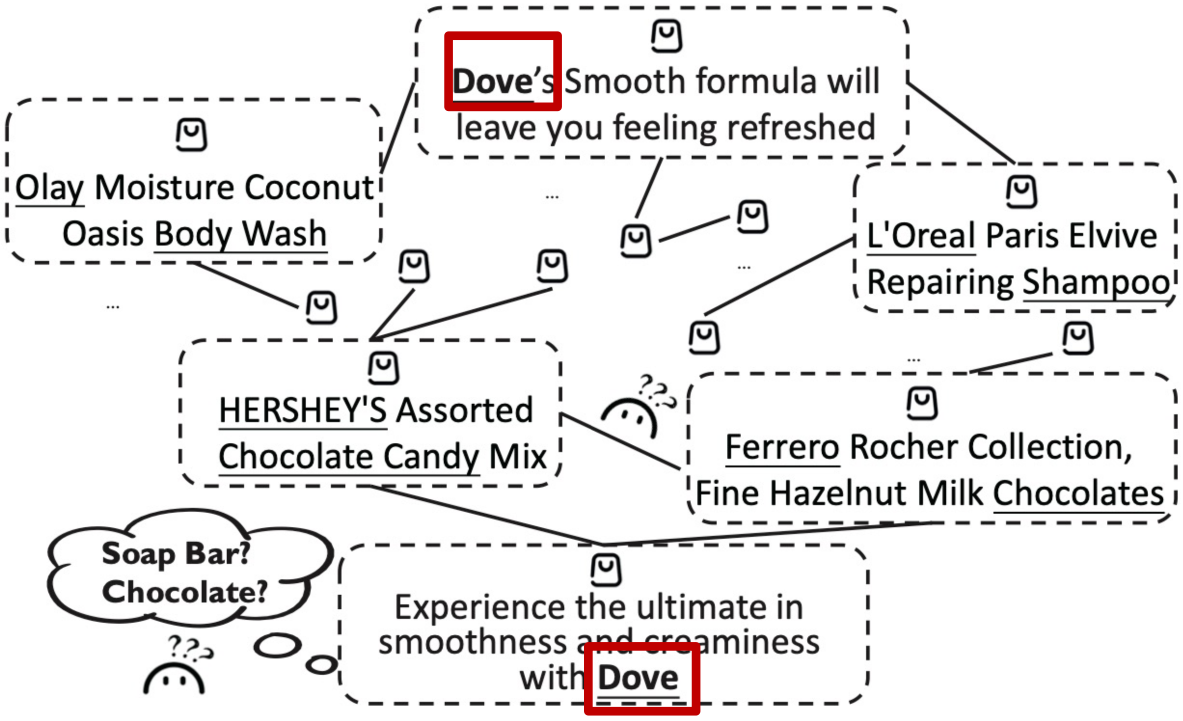
Pretraining on a Text-rich Network G

Finetuning on downstream tasks

Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL'23.

Language Model Pretraining: Patton

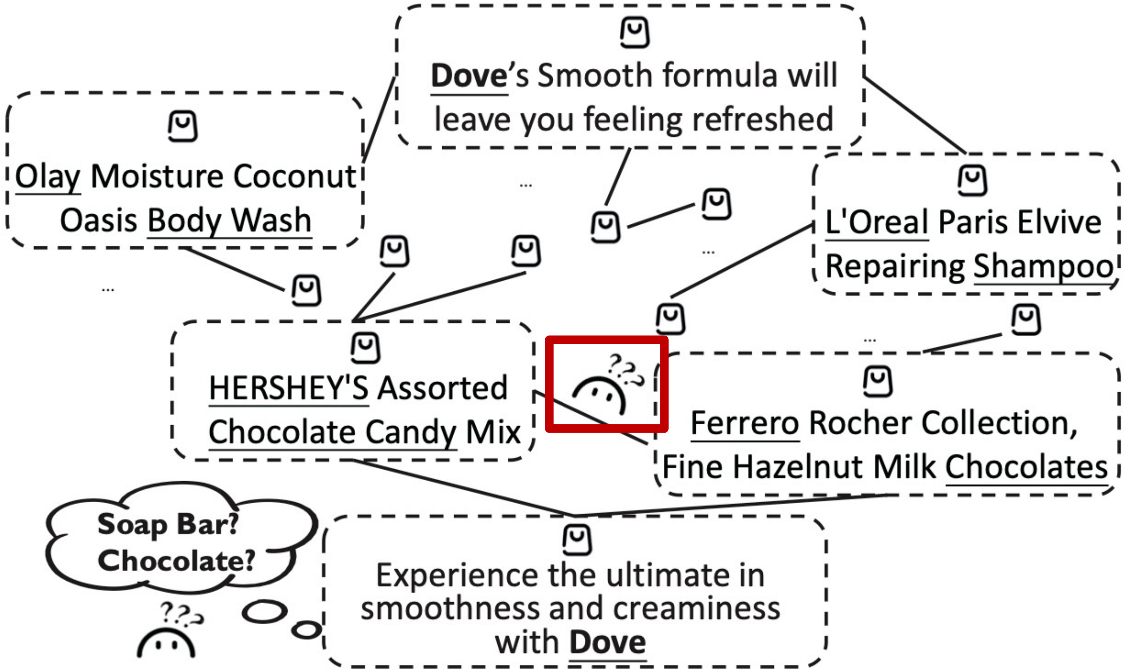
- How to design pretraining strategies to help LMs extract unsupervised semantic information from the network?
  - Motivation 1: On token-level, documents can help facilitate the understanding of tokens.



Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL'23.

Language Model Pretraining: Patton

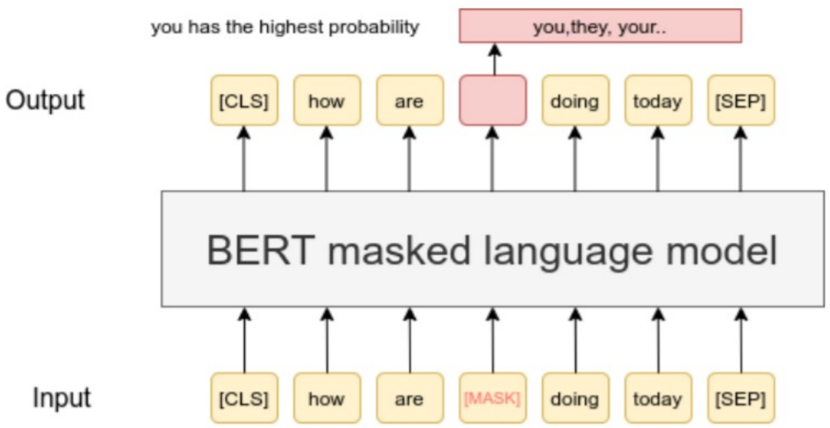
- How to design pretraining strategies to help LMs extract unsupervised semantic information from the network?
  - Motivation 2: On document-level, the two connected nodes can have quite related overall textual semantics.



Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL'23.

Language Model Pretraining: Patton

- Pretraining strategy 1: Network-contextualized masked language modeling
  - Original masked language modeling
    - BERT, domain adaptation
    - The semantics of each token can be reflected by its contexts.



$$\mathcal{L}_{\mathrm{MLM}} = -\sum_{i \in M_t} \log p(w_i | \boldsymbol{H}_i),$$

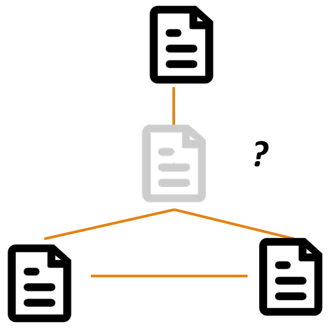
  - Ours
    - In node MLM -> Network contextualized MLM
    - Use both in-node text context and neighbor node context to conduct masked token prediction
    - Facilitate the LM to understand both in-node token correlation and network-contextualized text semantic relatedness

$$\mathcal{L}_{\mathrm{NMLM}} = -\sum_{i \in M_t} \log p(w_i | \boldsymbol{H}_x, \boldsymbol{z}_x),$$

Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL'23.

Language Model Pretraining: Patton

- Pretraining strategy 2: Masked Node Prediction
  - We dynamically hold out a subset of nodes from the network ($M_v \subseteq V$), mask them, and train the LM to predict the masked nodes based on the adjacent network structure.
  - LM will absorb document semantic hints hidden inside the network structure.



$$\mathcal{L}_{\text{MNP}} = -\sum_{v_j \in M_v} \log p(v_j | \boldsymbol{G}_{v_j})$$

  - Directly optimizing masked node prediction is computationally expensive
    - Representations for all candidates/neighboring nodes
  - We prove that masked node prediction can be theoretically transferred to a computationally cheaper pairwise link prediction task.

$$\begin{aligned}
& \prod_{v_{\texttt{[MASK]}} \in M_v} p(v_{\texttt{[MASK]}} = v_i | v_k \in N_{v_{\texttt{[MASK]}}}) \\
\propto & \prod_{v_{\texttt{[MASK]}} \in M_v} p(v_k \in N_{v_{\texttt{[MASK]}}} | v_{\texttt{[MASK]}} = v_i) \\
= & \prod_{v_{\texttt{[MASK]}} \in M_v} \prod_{v_k \in N_{v_{\texttt{[MASK]}}}} p(v_k | v_{\texttt{[MASK]}} = v_i) \\
= & \prod_{v_{\texttt{[MASK]}} \in M_v} \prod_{v_k \in N_{v_{\texttt{[MASK]}}}} p(v_k \longleftrightarrow v_i)
\end{aligned}$$

Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL'23.

Language Model Pretraining: Patton

❑ Retrieval

Table 3: Experiment results on Retrieval. We show the mean_{std} of three runs for all the methods.

Method	Mathematics R@50	Mathematics R@100	Geology R@50	Geology R@100	Economy R@50	Economy R@100	Clothes R@50	Clothes R@100	Sports R@50	Sports R@100
BM25	20.76	24.55	19.02	20.92	19.14	22.49	15.76	15.88	22.00	23.96
BERT	$16.73_{0.17}$	$22.66_{0.18}$	$18.82_{0.39}$	$25.94_{0.39}$	$23.95_{0.25}$	$31.54_{0.21}$	$40.77_{1.68}$	$50.40_{1.41}$	$32.37_{1.09}$	$43.32_{0.96}$
GraphFormers	$16.65_{0.12}$	$22.41_{0.10}$	$18,92_{0.60}$	$25.94_{0.39}$	$24.48_{0.36}$	$32.16_{0.40}$	$41.77_{2.05}$	$51.26_{2.27}$	$32.39_{0.89}$	$43.29_{1.12}$
SciBERT	$24.70_{0.17}$	$33.55_{0.31}$	$23.71_{0.89}$	$30.94_{0.95}$	$29.80_{0.66}$	$38.66_{0.52}$	-	-	-	-
SPECTER	$23.86_{0.25}$	$31.11_{0.31}$	$26.56_{1.05}$	$34.04_{1.32}$	$31.26_{0.15}$	$40.79_{0.11}$	-	-	-	-
SimCSE (unsup)	$17.91_{0.26}$	$23.19_{0.29}$	$20.45_{0.20}$	$26.82_{0.26}$	$25.83_{0.23}$	$33.42_{0.28}$	$44.90_{0.35}$	$54.76_{0.38}$	$38.81_{0.35}$	$49.30_{0.44}$
SimCSE (sup)	$20.29_{0.41}$	$26.23_{0.51}$	$22.34_{0.49}$	$29.63_{0.55}$	$28.07_{0.38}$	$36.51_{0.37}$	$44.69_{0.59}$	$54.70_{0.77}$	$40.31_{0.43}$	$50.55_{0.41}$
LinkBERT	$17.25_{0.30}$	$23.21_{0.47}$	$17.14_{0.75}$	$23.05_{0.74}$	$22.69_{0.30}$	$30.77_{0.36}$	$28.66_{2.97}$	$37.79_{3.82}$	$31.97_{0.54}$	$41.77_{0.67}$
BERT.MLM	$20.69_{0.21}$	$27.17_{0.25}$	$32.13_{0.36}$	$41.74_{0.42}$	$27.13_{0.04}$	$36.00_{0.14}$	$52.41_{1.71}$	$63.72_{1.79}$	$54.10_{0.81}$	$63.14_{0.83}$
SciBERT.MLM	$20.65_{0.21}$	$27.67_{0.32}$	$31.65_{0.71}$	$40.52_{0.76}$	$29.23_{0.67}$	$39.18_{0.73}$	-	-	-	-
SimCSE.in-domain	$24.54_{0.05}$	$31.66_{0.09}$	$33.97_{0.07}$	$44.09_{0.19}$	$28.44_{0.31}$	$37.81_{0.27}$	$61.42_{0.84}$	$72.25_{0.86}$	$53.77_{0.22}$	$63.73_{0.30}$
PATTON	$27.44_{0.15}$	$34.97_{0.21}$	$34.94_{0.23}$	$45.01_{0.28}$	$32.10_{0.51}$	$42.19_{0.62}$	$\mathbf{68.62}_{0.38}$	$\mathbf{77.54}_{0.19}$	$\mathbf{58.63}_{0.31}$	$\mathbf{68.53}_{0.55}$
SciPATTON	$\mathbf{31.40}_{0.52}$	$\mathbf{40.38}_{0.66}$	$\mathbf{40.69}_{0.52}$	$\mathbf{51.31}_{0.48}$	$\mathbf{35.82}_{0.69}$	$46.05_{0.69}$	**-**	**-**	**-**	**-**
w/o NMLM	$30.85_{0.14}$	$39.89_{0.23}$	$39.29_{0.07}$	$49.59_{0.11}$	$35.17_{0.31}$	$\mathbf{46.07}_{0.20}$	$65.60_{0.26}$	$75.19_{0.32}$	$57.05_{0.14}$	$67.22_{0.12}$
w/o MNP	$22.47_{0.07}$	$30.20_{0.15}$	$31.28_{0.89}$	$40.54_{0.97}$	$29.54_{0.36}$	$39.57_{0.57}$	$60.20_{0.73}$	$69.85_{0.52}$	$51.73_{0.41}$	$60.35_{0.78}$

❑ Classification

Table 2: Experiment results on Classification. We show the mean_{std} of three runs for all the methods.

Method	Mathematics Macro-F1	Mathematics Micro-F1	Geology Macro-F1	Geology Micro-F1	Economy Macro-F1	Economy Micro-F1	Clothes Macro-F1	Clothes Micro-F1	Sports Macro-F1	Sports Micro-F1
BERT	$18.14_{0.07}$	$22.04_{0.32}$	$21.97_{0.87}$	$29.63_{0.36}$	$14.17_{0.08}$	$19.77_{0.12}$	$45.10_{1.47}$	$68.54_{2.25}$	$31.88_{0.23}$	$34.58_{0.56}$
GraphFormers	$18.69_{0.52}$	$23.24_{0.46}$	$22.64_{0.92}$	$31.02_{1.16}$	$13.68_{1.03}$	$19.00_{1.44}$	$46.27_{1.92}$	$68.97_{2.46}$	$43.77_{0.63}$	$50.47_{0.78}$
SciBERT	$23.50_{0.64}$	$23.10_{2.23}$	$29.49_{1.25}$	$37.82_{1.89}$	$15.91_{0.48}$	$21.32_{0.66}$	-	-	-	-
SPECTER	$23.37_{0.07}$	$29.83_{0.96}$	$30.40_{0.48}$	$38.54_{0.77}$	$16.16_{0.17}$	$19.84_{0.47}$	-	-	-	-
SimCSE (unsup)	$20.12_{0.08}$	$26.11_{0.39}$	$38.78_{0.19}$	$38.55_{0.17}$	$14.54_{0.26}$	$19.07_{0.43}$	$42.70_{2.32}$	$58.72_{0.34}$	$41.91_{0.85}$	$59.19_{0.55}$
SimCSE (sup)	$20.39_{0.07}$	$25.56_{0.00}$	$25.66_{0.28}$	$33.89_{0.40}$	$15.03_{0.53}$	$18.64_{1.32}$	$52.82_{0.87}$	$75.54_{0.98}$	$46.69_{0.10}$	$59.19_{0.55}$
LinkBERT	$15.78_{0.91}$	$19.75_{1.19}$	$24.08_{0.58}$	$31.32_{0.04}$	$12.71_{0.12}$	$16.39_{0.22}$	$44.94_{2.52}$	$65.33_{4.34}$	$35.60_{0.33}$	$38.30_{0.09}$
BERT.MLM	$23.44_{0.39}$	$31.75_{0.58}$	$36.31_{0.36}$	$48.04_{0.69}$	$16.60_{0.21}$	$22.71_{1.16}$	$46.98_{0.84}$	$68.00_{0.84}$	$62.21_{0.13}$	$75.43_{0.74}$
SciBERT.MLM	$23.34_{0.42}$	$30.11_{0.97}$	$36.94_{0.28}$	$46.54_{0.40}$	$16.28_{0.38}$	$21.41_{0.81}$	-	-	-	-
SimCSE.in-domain	$25.15_{0.09}$	$29.85_{0.20}$	$38.91_{0.08}$	$48.93_{0.14}$	$18.08_{0.22}$	$23.79_{0.44}$	$57.03_{0.20}$	$80.16_{0.31}$	$65.57_{0.35}$	$75.22_{0.18}$
PATTON	$\mathbf{27.58}_{0.03}$	$\mathbf{32.82}_{0.01}$	$39.35_{0.06}$	$48.19_{0.15}$	$19.32_{0.05}$	$25.12_{0.05}$	$\mathbf{60.14}_{0.28}$	$\mathbf{84.88}_{0.09}$	$\mathbf{67.57}_{0.08}$	$\mathbf{78.60}_{0.15}$
SciPATTON	$27.35_{0.04}$	$31.70_{0.01}$	$\mathbf{39.65}_{0.10}$	$\mathbf{48.93}_{0.06}$	$\mathbf{19.91}_{0.08}$	$\mathbf{25.68}_{0.32}$	**-**	**-**	**-**	**-**
w/o NMLM	$25.91_{0.45}$	$27.79_{2.07}$	$38.78_{0.19}$	$48.48_{0.17}$	$18.86_{0.23}$	$24.25_{0.26}$	$56.68_{0.24}$	$80.27_{0.17}$	$65.83_{0.28}$	$76.24_{0.54}$
w/o MNP	$24.79_{0.65}$	$29.44_{1.50}$	$38.00_{0.73}$	$47.82_{1.06}$	$18.69_{0.59}$	$25.63_{1.44}$	$47.35_{1.20}$	$68.50_{2.60}$	$64.23_{1.53}$	$76.03_{1.67}$

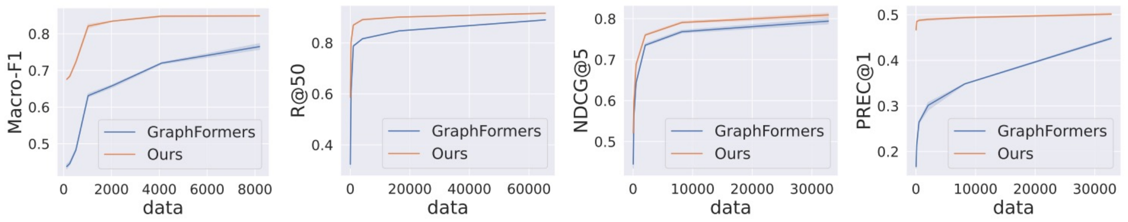
❑ Link prediction

Table 5: Experiment results on Link Prediction. We show the mean_{std} of three runs for all the methods.

Method	Mathematics PREC@1	Mathematics MRR	Geology PREC@1	Geology MRR	Economy PREC@1	Economy MRR	Clothes PREC@1	Clothes MRR	Sports PREC@1	Sports MRR
BERT	$6.60_{0.16}$	$12.96_{0.34}$	$6.24_{0.76}$	$12.96_{1.34}$	$4.12_{0.08}$	$9.23_{0.15}$	$24.17_{0.41}$	$34.20_{0.45}$	$16.48_{0.45}$	$25.35_{0.52}$
GraphFormers	$6.91_{0.29}$	$13.42_{0.34}$	$6.52_{1.17}$	$13.34_{1.81}$	$4.16_{0.21}$	$9.28_{0.28}$	$23.79_{0.69}$	$33.79_{0.66}$	$16.69_{0.36}$	$25.74_{0.48}$
SciBERT	$14.08_{0.11}$	$23.62_{0.10}$	$7.15_{0.26}$	$14.11_{0.39}$	$5.01_{1.04}$	$10.48_{1.79}$	-	-	-	-
SPECTER	$13.44_{0.5}$	$21.73_{0.65}$	$6.85_{0.22}$	$13.37_{0.34}$	$6.33_{0.29}$	$12.41_{0.33}$	-	-	-	-
SimCSE (unsup)	$9.85_{0.10}$	$16.28_{0.12}$	$7.47_{0.55}$	$14.24_{0.89}$	$5.72_{0.26}$	$11.02_{0.34}$	$30.51_{0.09}$	$40.40_{0.10}$	$22.99_{0.07}$	$32.47_{0.06}$
SimCSE (sup)	$10.35_{0.52}$	$17.01_{0.72}$	$10.10_{0.04}$	$17.80_{0.07}$	$5.72_{0.26}$	$11.02_{0.34}$	$35.42_{0.06}$	$46.07_{0.06}$	$27.07_{0.15}$	$37.44_{0.16}$
LinkBERT	$8.05_{0.14}$	$13.91_{0.09}$	$6.40_{0.14}$	$12.99_{0.17}$	$2.97_{0.08}$	$6.79_{0.15}$	$30.33_{0.56}$	$39.59_{0.64}$	$19.83_{0.09}$	$28.32_{0.04}$
BERT.MLM	$17.55_{0.25}$	$29.22_{0.26}$	$14.13_{0.19}$	$25.36_{0.20}$	$9.02_{0.09}$	$16.72_{0.15}$	$42.71_{0.31}$	$54.54_{0.35}$	$29.36_{0.09}$	$41.60_{0.05}$
SciBERT.MLM	$22.44_{0.08}$	$34.22_{0.05}$	$16.22_{0.03}$	$27.02_{0.07}$	$9.80_{0.00}$	$17.72_{0.01}$	-	-	-	-
SimCSE.in-domain	$33.55_{0.05}$	$46.07_{0.07}$	$24.56_{0.06}$	$36.89_{0.11}$	$16.77_{0.10}$	$26.93_{0.01}$	$\mathbf{60.41}_{0.03}$	$\mathbf{71.86}_{0.06}$	$49.17_{0.04}$	$63.48_{0.03}$
PATTON	$70.41_{0.11}$	$80.21_{0.04}$	$44.76_{0.05}$	$57.71_{0.04}$	$57.04_{0.05}$	$68.35_{0.04}$	$58.59_{0.12}$	$70.12_{0.12}$	$46.68_{0.09}$	$60.96_{0.23}$
SciPATTON	$\mathbf{71.22}_{0.17}$	$\mathbf{80.79}_{0.10}$	$\mathbf{44.95}_{0.24}$	$\mathbf{57.84}_{0.25}$	$\mathbf{57.36}_{0.26}$	$\mathbf{68.71}_{0.31}$	**-**	**-**	**-**	**-**
w/o NMLM	$71.04_{0.13}$	$80.60_{0.07}$	$44.33_{0.23}$	$57.29_{0.22}$	$56.64_{0.25}$	$68.12_{0.16}$	$60.30_{0.03}$	$71.67_{0.07}$	$\mathbf{49.72}_{0.06}$	$\mathbf{63.76}_{0.04}$
w/o MNP	$63.06_{0.23}$	$74.26_{0.11}$	$33.84_{0.60}$	$47.02_{0.65}$	$44.46_{0.03}$	$57.05_{0.04}$	$49.62_{0.06}$	$61.61_{0.01}$	$36.05_{0.20}$	$49.78_{0.25}$

❑ How pretraining help the model?

❑ Finetune data size study



(a) Classification (b) Retrieval (c) Reranking (d) Link Prediction

Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL’23.

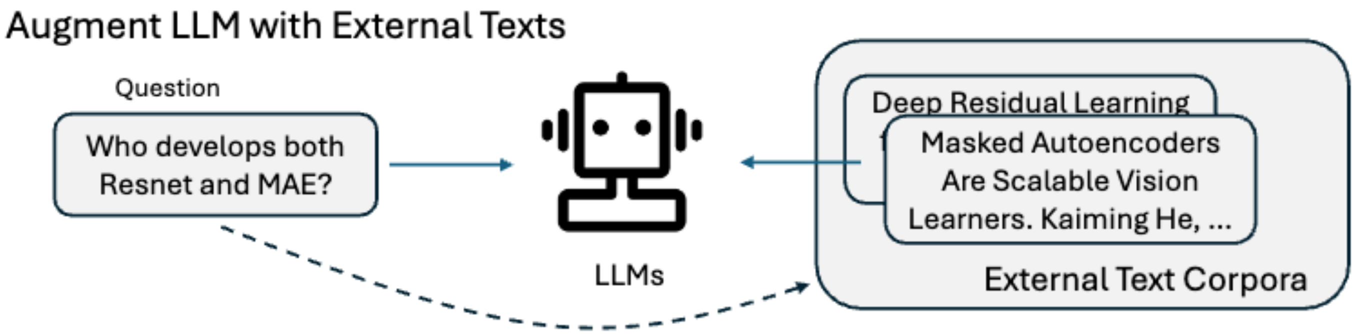
Outline

- Why Mining Graphs with Large Language Models?
- Mining Pure Graphs with Large Language Models
- Mining Text-Attributed Graphs with Large Language Models
  - Model architecture – representation learning
  - Language Model Pretraining
  - Augment LLM with Graph: Graph CoT (arxiv’24)
- Mining Text-Paired Graphs with Large Language Models

Augment LLM with Graph

- Retrieval-augmented generation (RAG)
  - Motivation
    - LLMs suffer from hallucination
    - External corpus can provide knowledge to mitigate hallucination
  - Pipeline
    - Retriever: fetch knowledge from corpus
    - LLM: inference

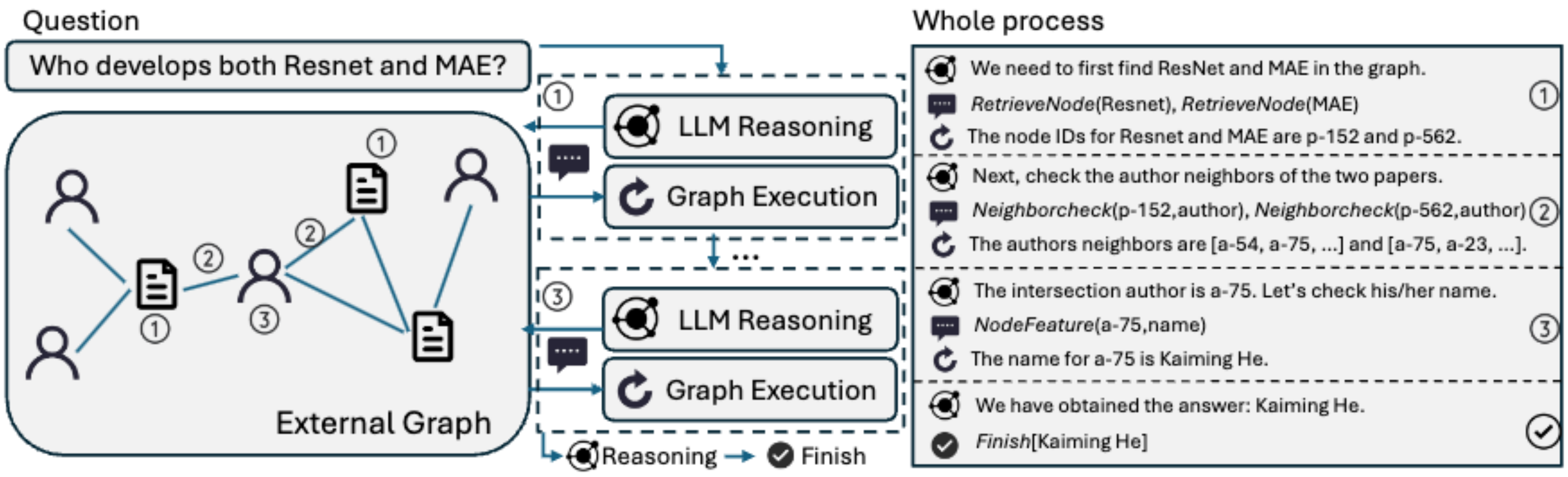
What if the text units in the corpora is linked?



Jin, et al. Graph Chain-of-Thought: Augmenting Large Language Models by Reasoning on Graphs. Arxiv'24.

Augment LLM with Graph: Graph CoT

- Graph Chain-of-Thought
  - Iteratively traverse on graph & reasoning with LLM



Jin, et al. Graph Chain-of-Thought: Augmenting Large Language Models by Reasoning on Graphs. Arxiv'24.

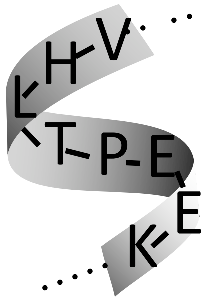
Outline

- Why Mining Graphs with Large Language Models?
- Mining Pure Graphs with Large Language Models
- Mining Text-Attributed Graphs with Large Language Models
- Mining Text-Paired Graphs with Large Language Models
  - MolT5 (EMNLP'22)

MolT5

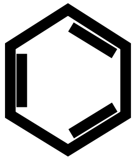
- A pretrained molecular language model

Protein Graphs

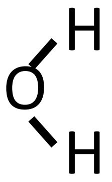


“Myoglobin holds oxygen in muscles.”

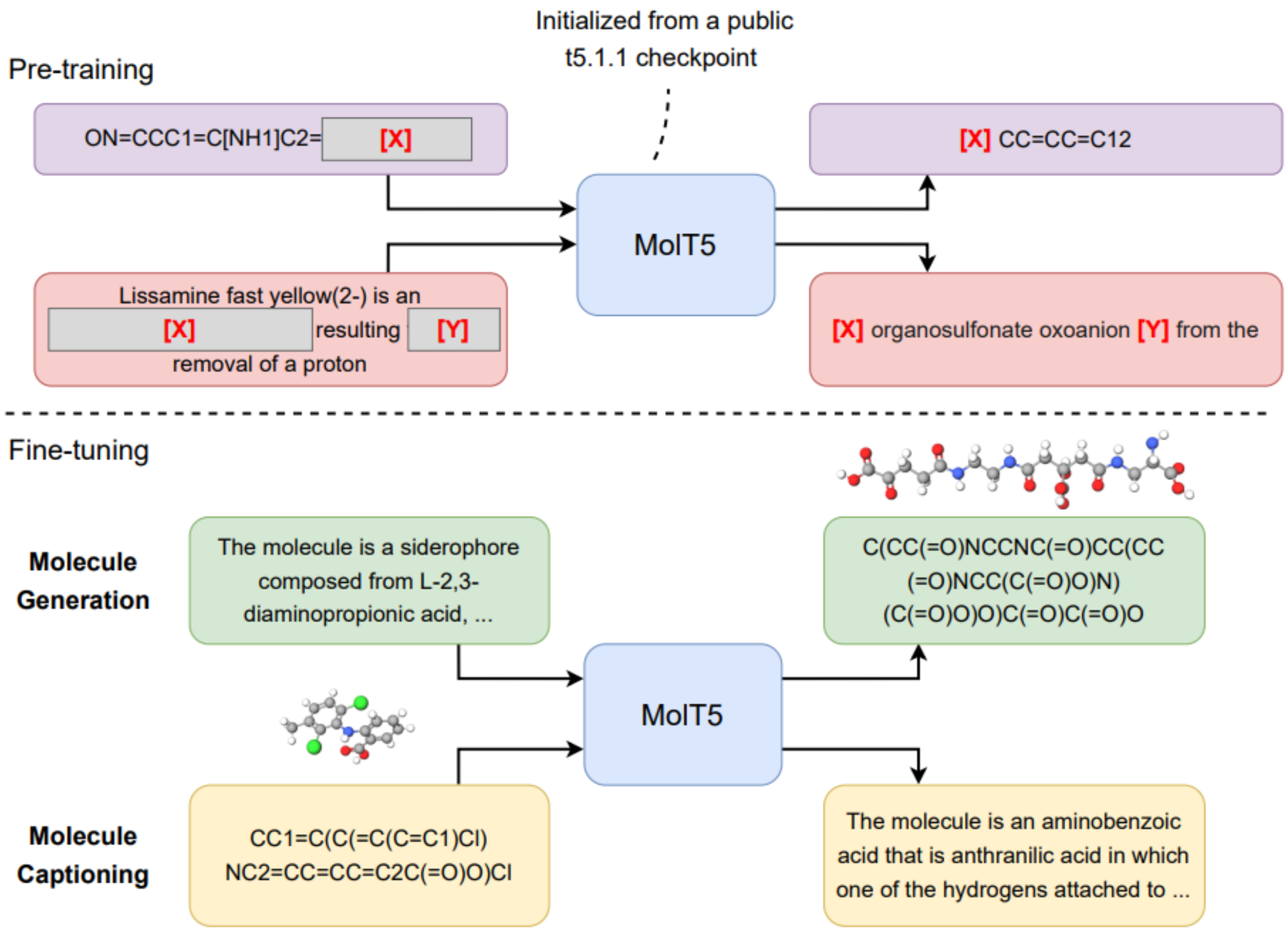
Molecule Graphs



“Benzene is toxic”



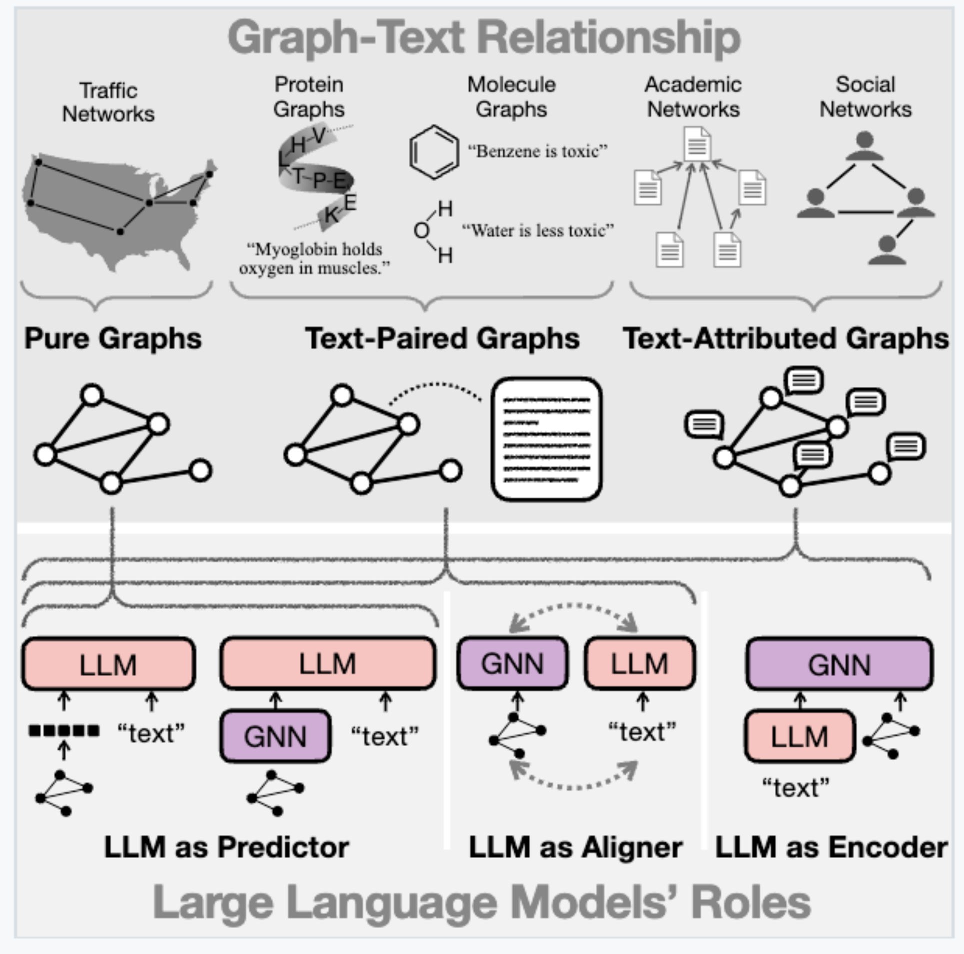
“Water is less toxic”



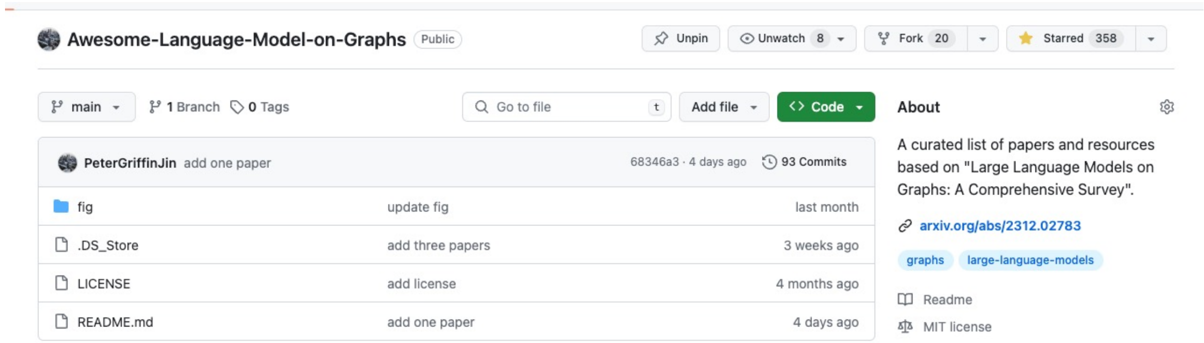
Carl, et al. Translation between Molecules and Natural Language. EMNLP’22.

Summary

- A survey paper of LLM & graphs



- A resource repo of LLM & graphs



paper

repo

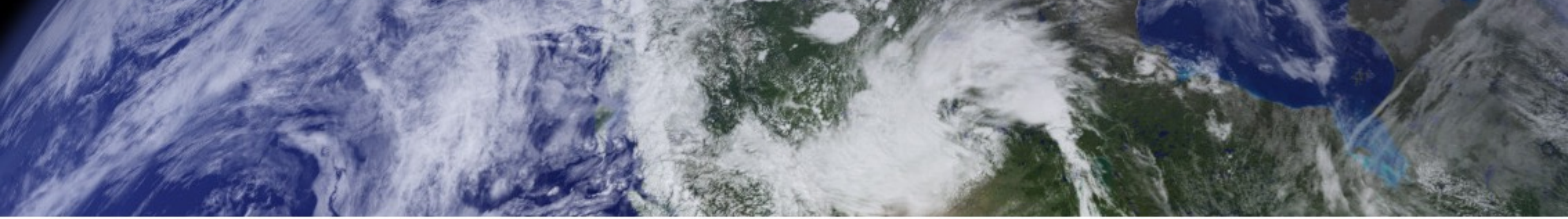
Jin, et al. Large Language Model on Graphs: A Comprehensive Survey. Arxiv. 2023.12.

References

- Jin, et al. Large Language Model on Graphs: A Comprehensive Survey. Arxiv. 2023.12.
- Wang, et al. Can Language Models Solve Graph Problems in Natural Language? NeurIPs 2023.
- Sun, et al. Think-on-Graph: Deep and Responsible Reasoning of Large Language Model on Knowledge Graph. ICLR 2024.
- Zhang, et al. The Effect of Metadata on Scientific Literature Tagging: A Cross-Field Cross-Model Study. WWW 2023.
- Zhang, et al. Metadata-Induced Contrastive Learning for Zero-Shot Multi-Label Text Classification. WWW 2022.
- Jin, et al. Learning Multiplex Representations on Text-Attributed Graphs with one Language Model Encoder. Arxiv 2023.
- Zhu, et al. TextGNN: Improving Text Encoder via Graph Neural Network in Sponsored Search. WWW 2021.
- Zhao, et al. Learning on Large-scale Text-attributed Graphs via Variational Inference. ICLR 2023.

References

- Yang, et al. GraphFormers: GNN-nested Transformers for Representation Learning on Textual Graph. NeurIPs’21.
- Jin, et al. Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks. ICLR’23.
- Jin, et al. Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks. KDD’23.
- Jin, et al. Patton: Language Model Pretraining on Text-rich Networks. ACL’23.
- Jin, et al. Graph Chain-of-Thought: Augmenting Large Language Models by Reasoning on Graphs. Arxiv’24.
- Carl, et al. Translation between Molecules and Natural Language. EMNLP’22.



Q&A