Retrieval Augmented Generation for Dynamic Graph Modeling

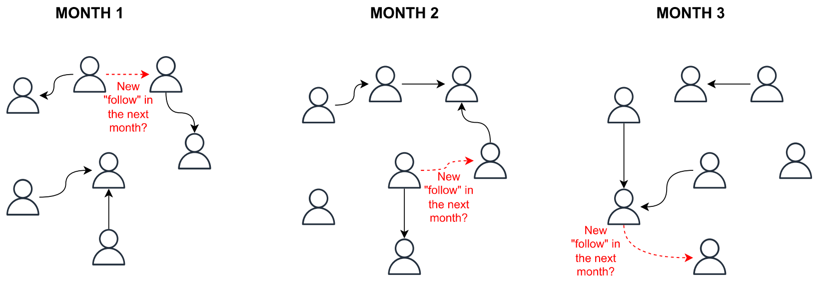
Yuxia Wu
Singapore Management University
yieshah2017@gmail.com

Lizi Liao
Singapore Management University
lzliao@smu.edu.sg

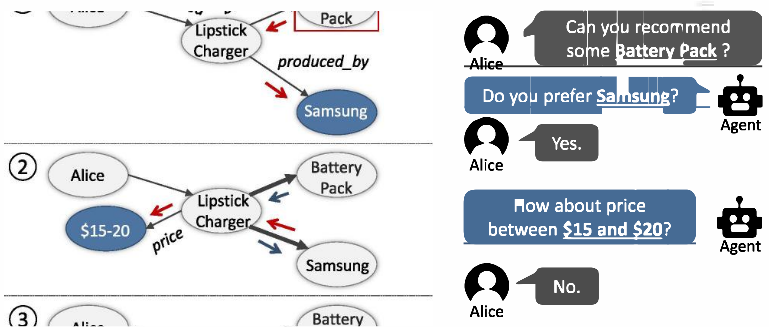
Yuan Fang
Singapore Management University
yfang@smu.edu.sg

Motivation: Dynamic Graph Modeling

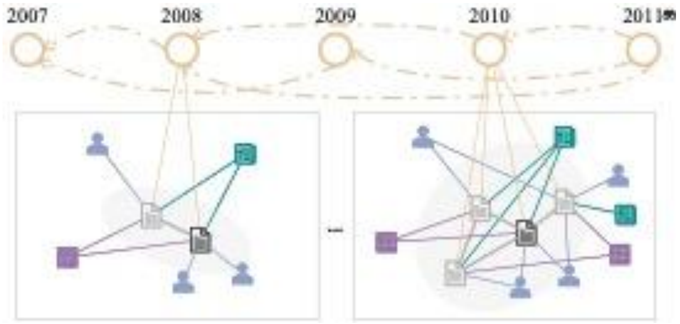
- ❑ Applications



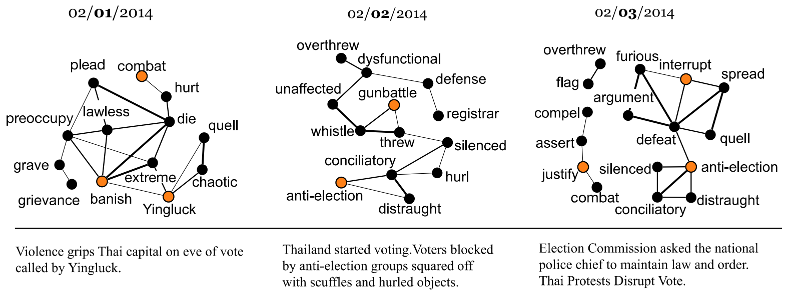
Social network



Conversation



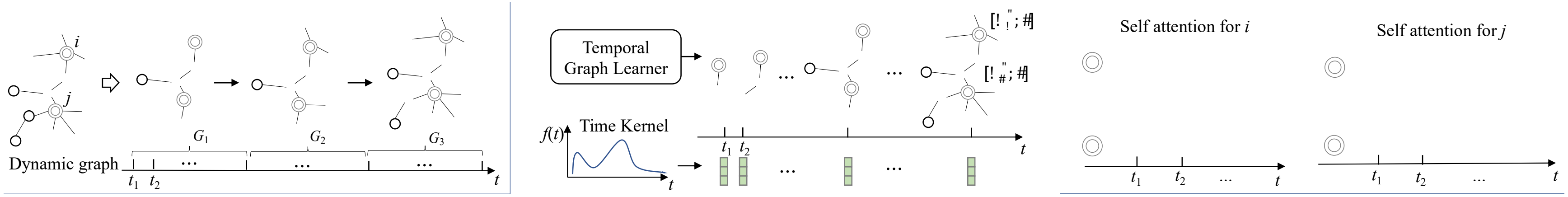
Citation network



Event graph

Motivation: Dynamic Graph Modeling

- Related works
  - GNN-based:
    - Discrete-time approaches: capture graph snapshots at specific intervals
    - Continuous-time approaches: model events as they occur, offering a more granular perspective
  - Transformer-based:
    - Capture long-range dependencies within temporal sequences



(a) Discrete-time approaches | (b) Continuous-time approaches | (c) Self-attention in Transformer

Discard the fine-grained temporal information within the snapshot

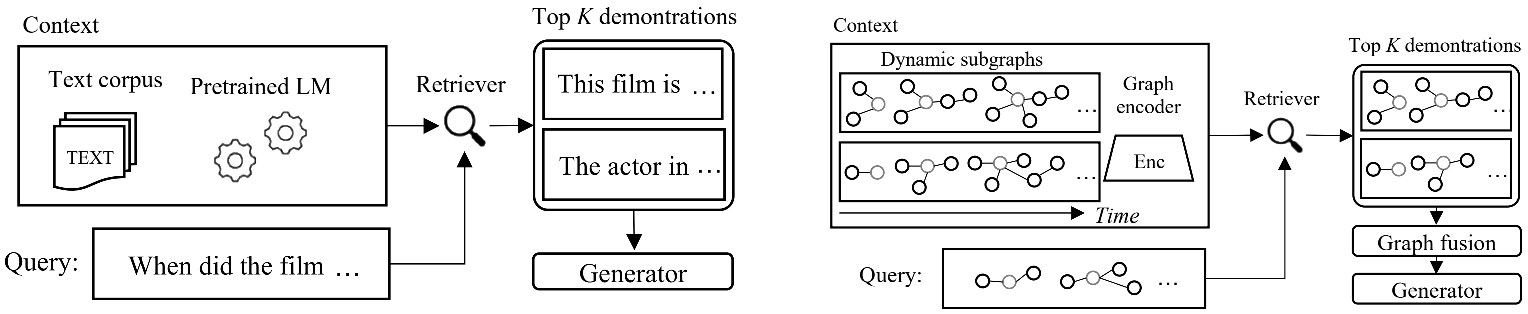
Difficult in capturing long-term dependency within historical graph data

Difficult in generalizing across different contexts and adapting to emerging patterns

Motivation: Dynamic Graph Modeling

❑ Motivation: Retrieval-Augmented Generation (RAG)

RAG has the potential to broaden the contextual understanding of dynamic graphs by retrieving and incorporating relevant examples from across the graph's temporal and contextual space



(a) RAG in NLP (b) RAG in dynamic graphs

➢ RAG in NLP:

Pre-trained LM → Encode text and retrieve related demonstrations → Concatenated to enhance the generation task

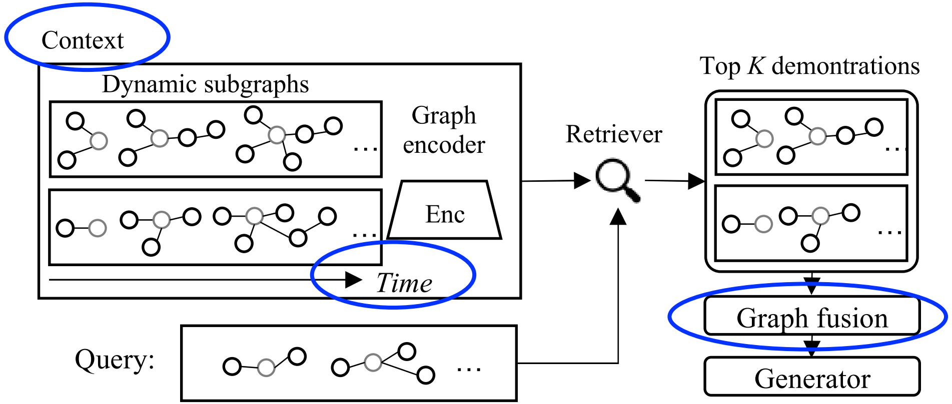
➢ RAG in graph:

Graph encoder → Encode dynamic graphs and retrieve related demonstrations → Fusion to enhance generation task

Motivation: Dynamic Graph Modeling

❑ Challenges:

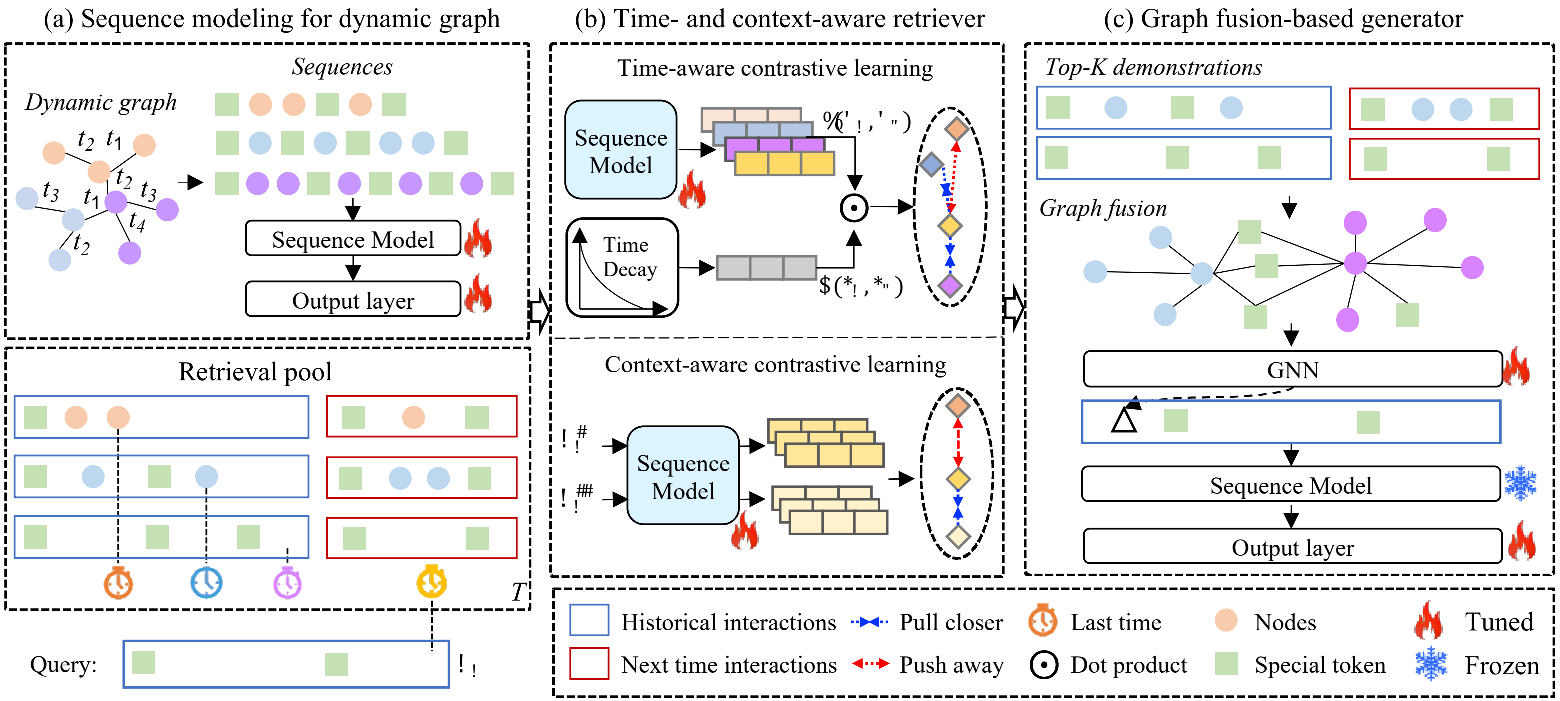
- **Selecting** high-quality demonstrations
  - Identifying **contextually and temporally** relevant demonstrations
  - Existing methods (BM25) rely on historical interactions similarities, struggling with inductive scenarios
- **Integrating** the retrieved demonstrations
  - Simply concatenating query and demonstrations → lengthy inputs and overlook structural patterns



(b) RAG in dynamic graphs

Proposed Method: RAG4DyG

❑RAG4DyG: Retrieval-Augmented Generation for Dynamic Graph Modeling



Proposed Method: Preliminaries

- **Training samples** (Retrieval pool):
  - Input sequence:

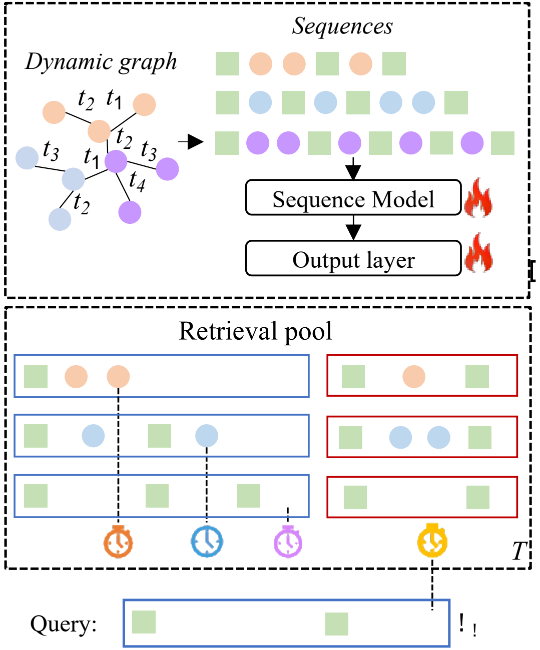
$$G_8 = [hist], E_8, [time\_1], E_8^{1,1}, E_8^{1,2}, \ldots, [time\_t], E_8^{C1}, \ldots,$$
$$[time\_T], E_8^{),1}, \ldots, [eohist],$$

  - Output sequence:

$$\sim_8 = [pred], [time\_T{+}1], E_8^{)+1,1}, \ldots, [eopred],$$

- **Problem Formulation:**

Query sequence (input sequence) → Retrieve K demonstrations → Augmented input→ Prediction



Proposed Method: Time- and Context-Aware Retriever

- Contextual similarity: $s(x_q, x_p) = f(\underbrace{x_q}_{\text{query}})^\top f(\underbrace{x_p}_{\text{candidate}}),$

- **Time-aware Contrastive Learning :**

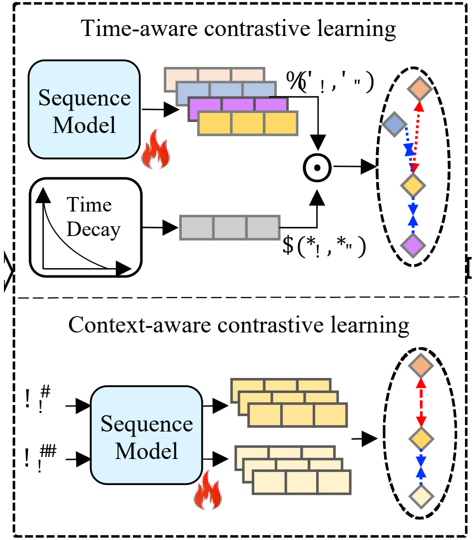
  Demonstrations closer in time to the query are more relevant than those further away.

  - Time decay function : $\mu(t_q, t_p) = \exp(-\lambda |t_q - t_p|),$
  - Reweight contextual similarity: $h(x_q, x_p) = s(x_q, x_p)\mu(t_q, t_p)$
  - Contrastive loss: $$\mathcal{L}_{\text{tcl}} = -\log \frac{\exp(h(x_q, x_p^+))/\tau}{\sum_{j=1}^{2N} \mathbb{1}_{j \neq q} \exp(h(x_q, x_j))/\tau},$$

- **Context-aware Contrastive Learning:**

  - Two types of augmentations: masking and cropping:
  - Contrastive loss: $$\mathcal{L}_{\text{ccl}} = -\log \frac{\exp(s(x_q', x_q'')/\tau)}{\sum_{j=1}^{2N} \mathbb{1}_{j \neq q} \exp s(x_q', x_j')/\tau},$$

- Training objective of retrieval: $\mathcal{L}_{\text{ret}} = \mathcal{L}_{\text{tcl}} + \alpha \mathcal{L}_{\text{ccl}},$



Proposed Method: Graph Fusion-based Generator

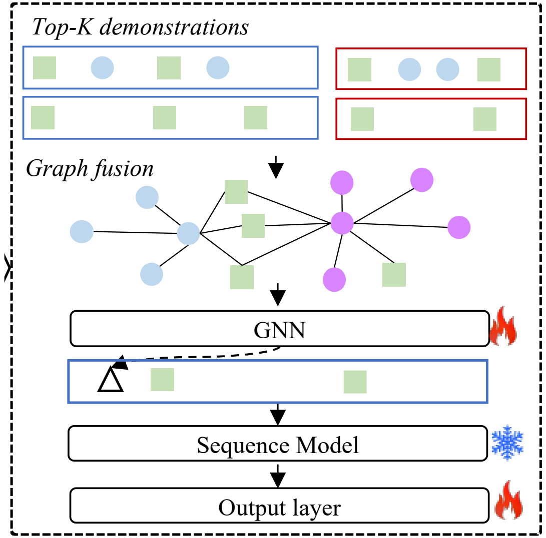
- **Fuse** top-k demonstrations with query:
  - **Concatenation:** lengthy context, neglect structural pattern ✗
  - **MLP:** map to shorter sequence, neglect structural pattern ✗
  - **Graph fusion:** fuse the demonstrations into a summary graph ✓

- **GNN Processing:**

$$e_{\text{fus}} = \texttt{MeanPooling}(\text{GCN}(G_{\text{fus}})),$$

- **Prepend** the graph readout from the GNN to the query

$$\tilde{x}_q = [e_{\text{fus}} \parallel x_q],$$



Experiments:

SMU
SINGAPORE MANAGEMENT UNIVERSITY

❑ Datasets:

Six datasets from different domains:

Table 1: Dataset statistics.

Dataset	UCI	Hepth	MMConv	Wikipedia	Enron	Reddit
Domain	Social	Citation	Conversation	Behavior	Social	Hyperlink
# Nodes	1,781	4,737	7,415	9,227	42,711	11,901
# Edges	16,743	14,831	91,986	157,474	797,907	62,919

Experiment:

Table 2: Performance comparison for dynamic link prediction with mean and standard deviation across 10 runs. Best results are bolded; runners-up are underlined. * indicates that our model significantly outperforms the best baseline based on the two-tail t-test ($p < 0.05$).

Datasets	Models	DySAT	EvolveGCN	DyRep	JODIE	TGAT	TGN	TREND	GraphMixer	IDOL	SimpleDyG	RAG4DyG
UCI	Recall@5	0.009±0.003	0.072±0.046	0.009±0.008	0.018±0.019	0.022±0.004	0.014±0.010	0.083±0.015	0.097±0.019	0.093±0.029	0.109±0.014	**0.111**±0.013
	NDCG@5	0.010±0.003	0.064±0.045	0.011±0.018	0.022±0.023	0.061±0.007	0.041±0.017	0.067±0.010	0.104±0.013	0.075±0.022	0.104±0.010	**0.122***±0.014
	Jaccard	0.010±0.001	0.032±0.026	0.010±0.005	0.012±0.009	0.020±0.002	0.011±0.003	0.039±0.020	0.042±0.005	0.014±0.002	0.092±0.014	**0.097**±0.010
Hepth	Recall@5	0.008±0.004	0.008±0.002	0.009±0.006	0.010±0.008	0.011±0.007	0.011±0.006	0.010±0.008	0.009±0.002	0.007±0.002	0.013±0.006	**0.019***±0.002
	NDCG@5	0.007±0.002	0.009±0.004	0.031±0.024	0.031±0.021	0.034±0.023	0.030±0.012	0.031±0.003	0.011±0.008	0.011±0.003	0.035±0.014	**0.045***±0.003
	Jaccard	0.005±0.001	0.007±0.002	0.010±0.006	0.011±0.008	0.011±0.006	0.008±0.001	0.010±0.002	0.010±0.003	0.006±0.001	0.013±0.006	**0.019***±0.002
MMConv	Recall@5	0.108±0.089	0.050±0.015	0.156±0.054	0.052±0.039	0.118±0.004	0.085±0.050	0.134±0.030	**0.206**±0.001	0.169±0.006	0.170±0.010	0.194±0.005
	NDCG@5	0.102±0.085	0.051±0.021	0.140±0.057	0.041±0.016	0.089±0.033	0.096±0.068	0.116±0.020	0.172±0.029	0.115±0.039	0.184±0.012	**0.208***±0.005
	Jaccard	0.095±0.080	0.032±0.017	0.067±0.025	0.032±0.022	0.058±0.021	0.066±0.038	0.060±0.018	0.085±0.016	0.015±0.002	0.169±0.010	**0.194***±0.005
Wikipedia	Recall@5	0.003±0.005	0.012±0.01	0.003±0.002	0.017±0.005	0.006±0.004	0.016±0.018	0.022±0.012	0.010±0.007	0.022±0.008	0.356±0.016	**0.369***±0.006
	NDCG@5	0.002±0.003	0.008±0.007	0.002±0.002	0.015±0.003	0.005±0.005	0.015±0.022	0.016±0.018	0.007±0.006	0.015±0.005	**0.398**±0.03	0.389±0.008
	Jaccard	0.001±0.001	0.004±0.004	0.001±0.001	0.007±0.002	0.002±0.002	0.007±0.009	0.007±0.021	0.004±0.002	0.004±0.001	0.320±0.027	**0.328**±0.007
Enron	Recall@5	0.002±0.004	0.004±0.011	0.021±0.001	0.005±0.005	0.020±0.002	0.001±0.001	0.023±0.003	0.021±0.002	0.024±0.014	0.094±0.005	**0.100***±0.003
	NDCG@5	0.001±0.002	0.007±0.020	0.036±0.002	0.061±0.039	0.036±0.001	0.003±0.001	0.027±0.001	0.037±0.001	0.025±0.011	0.114±0.005	**0.119***±0.004
	Jaccard	0.001±0.001	0.003±0.009	0.019±0.001	0.011±0.007	0.020±0.001	0.001±0.001	0.012±0.001	0.020±0.002	0.008±0.003	0.068±0.003	**0.071***±0.002
Reddit	Recall@5	0.001±0.002	0.006±0.002	0.019±0.004	0.013±0.003	0.001±0.001	0.001±0.001	0.002±0.003	0.001±0.001	0.003±0.002	0.101±0.019	**0.119***±0.006
	NDCG@5	0.001±0.002	0.012±0.003	0.020±0.004	0.015±0.002	0.001±0.001	0.002±0.001	0.003±0.002	0.003±0.001	0.005±0.003	0.134±0.012	**0.143**±0.005
	Jaccard	0.001±0.001	0.003±0.001	0.013±0.004	0.007±0.002	0.001±0.001	0.001±0.002	0.001±0.001	0.002±0.001	0.002±0.001	0.088±0.012	**0.096**±0.003

- RAG4DyG generally outperforms all baselines across different datasets under the three metrics
- RAG4DyG exhibits significant advantages in inductive scenarios such as the Hepth and Reddit datasets.

Experiment: Ablation Study

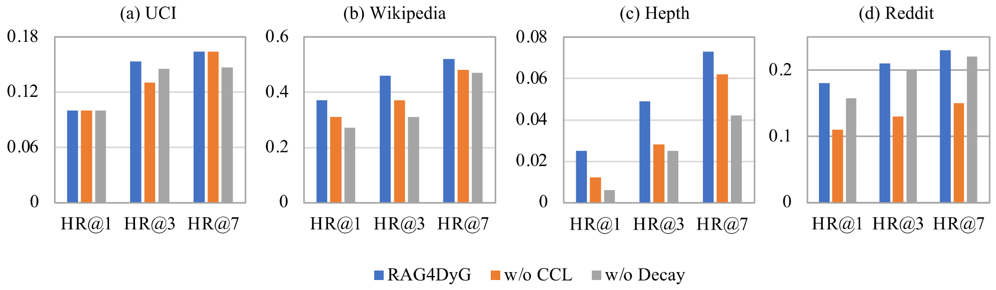


Figure 3: Ablation study for retrieval results.

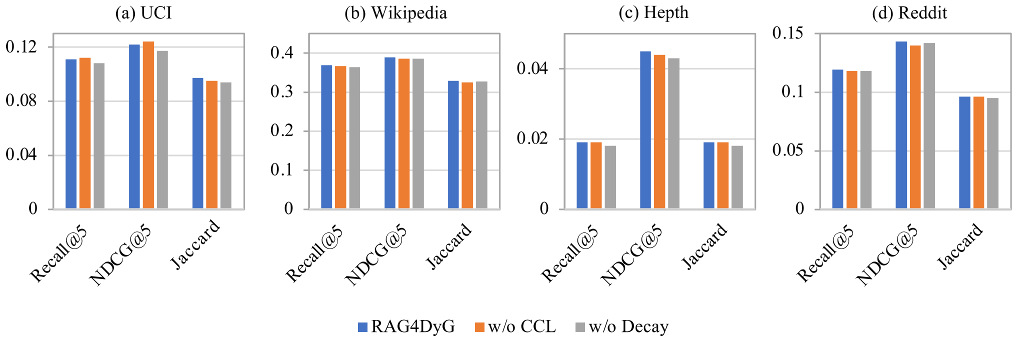


Figure 4: Ablation study for link prediction results.

- w/o CCL: exclude the context-aware contrastive learning
- *w/o Decay*: exclude the time decay

- The full model outperforms the two variants
- The w/o Decay exhibits the worst performance across both tasks, emphasizing the critical role of time decay in capturing temporal relevance

Experiment: Effect of Different Retrieval Methods

Table 3: Retrieval performance of various retrieval methods.

Method	UCI			Wikipedia			Hepth			Reddit		
	HR@1	HR@3	HR@7	HR@1	HR@3	HR@7	HR@1	HR@3	HR@7	HR@1	HR@3	HR@7
BM25	**0.100**	0.136	**0.200**	**0.369**	0.405	0.488	-	-	-	-	-	-
Jaccard	**0.100**	0.109	0.146	**0.369**	0.445	0.430	-	-	-	-	-	-
RAG4DyG	**0.100**	**0.155**	0.164	**0.369**	**0.455**	**0.523**	**0.025**	**0.049**	**0.073**	**0.180**	**0.218**	**0.228**

"-" denotes that the method is unable to perform retrieval. The reason is explained in the corresponding description of this table in Sec. 5.3.

Table 4: Generative performance of various retrieval methods.

Method	UCI			Wikipedia			Hepth			Reddit		
	Recall@5	NDCG@5	Jaccard	Recall@5	NDCG@5	Jaccard	Recall@5	NDCG@5	Jaccard	Recall@5	NDCG@5	Jaccard
BM25	**0.111**$\pm$0.007	0.121$\pm$0.009	0.093$\pm$0.004	0.368$\pm$0.01	**0.389**$\pm$0.012	0.325$\pm$0.01	-	-	-	-	-	-
Jaccard	0.104$\pm$0.009	0.113$\pm$0.011	0.088$\pm$0.010	0.368$\pm$0.013	0.388$\pm$0.014	0.321$\pm$0.011	-	-	-	-	-	-
RAG4DyG	**0.111**$\pm$0.013	**0.122**$\pm$0.014	**0.097**$\pm$0.010	**0.369**$\pm$0.006	**0.389**$\pm$0.008	**0.328**$\pm$0.007	**0.019**$\pm$0.002	**0.045**$\pm$0.003	**0.019**$\pm$0.002	**0.119**$\pm$0.006	**0.143**$\pm$0.005	**0.096**$\pm$0.003
GroundTruth	0.121$\pm$0.010	0.129$\pm$0.010	0.107$\pm$0.012	0.390$\pm$0.008	0.400$\pm$0.007	0.340$\pm$0.006	0.028$\pm$0.004	0.062$\pm$0.007	0.028$\pm$0.004	0.121$\pm$0.008	0.145$\pm$0.008	0.099$\pm$0.005

- *BM25*: calculates a relevance score (TF-IDF) between the query sequence and each candidate sequence in the retrieval pool
- *Jaccard*: calculates the set similarity by comparing the size of their intersection to the size of their union
- *"GroundTruth": an upper bound on the performance when using ground-truth retrieval results on the testing data*

- Retrieval performance: Ours shows comparable performance and can handle inductive scenarios (Hepth, Reddit)
- Generative performance: Our method performs better compared to other retrieval strategies.

Experiment: Effect of the K and Fusion Strategies

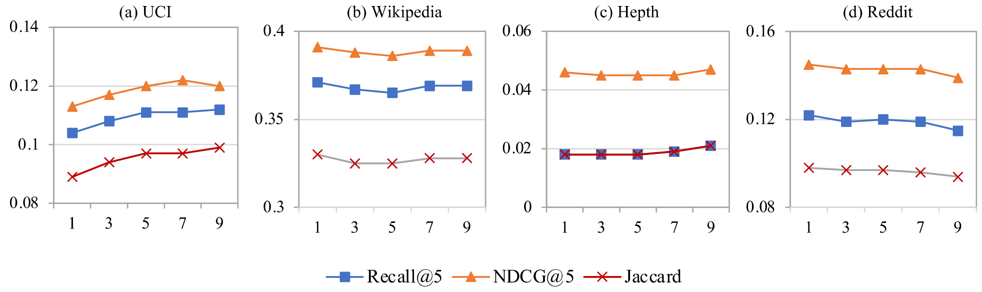


Figure 5: Effect of the number of demonstrations *K*.

- Higher *K* yields better prediction performance: more demonstrations provide richer contextual information
- Too large K may introduce more noise, which can harm the performance.

Table 5: Effect of different fusion strategies.

Fusion strategy	UCI			Wikipedia			Hepth			Reddit		
	Recall@5	NDCG@5	Jaccard	Recall@5	NDCG@5	Jaccard	Recall@5	NDCG@5	Jaccard	Recall@5	NDCG@5	Jaccard
Concatenation	0.033±0.019	0.036±0.018	0.029±0.016	0.210±0.019	0.232±0.021	0.206±0.019	0.001±0.002	0.007±0.002	0.002±0.002	0.001±0.001	0.003±0.003	0.001±0.001
MLP	0.102±0.018	0.106±0.017	0.089±0.016	0.356±0.006	0.371±0.009	0.321±0.007	0.006±0.002	0.015±0.002	0.006±0.002	0.108±0.006	0.132±0.005	0.090±0.003
GraphFusion	**0.111**±0.013	**0.122**±0.014	**0.097**±0.010	**0.369**±0.006	**0.389**±0.008	**0.328**±0.007	**0.019**±0.002	**0.045**±0.003	**0.019**±0.002	**0.119**±0.006	**0.143**±0.005	**0.096**±0.003

- **"Concatenation"** leads to lower performance compared with other strategies
- **"MLP"** maps the concatenated demonstrations into a shorter feature space, neglecting the structural pattern
- **"GraphFusion"** highlights the importance of considering both the content and the structure of the demonstrations for fusion.

Summary:

- We a novel retrieval-augmented generation approach for dynamic graph modeling
- We introduce a time- and context- aware contrastive learning module for demonstration retrieval and a graph fusion module to effectively integrate retrieved demonstrations.
- We conduct extensive experiments to validate our approach, demonstrating the effectiveness of RAG4DyG across various domains.

- Thank you for your listening!
- Q & A