

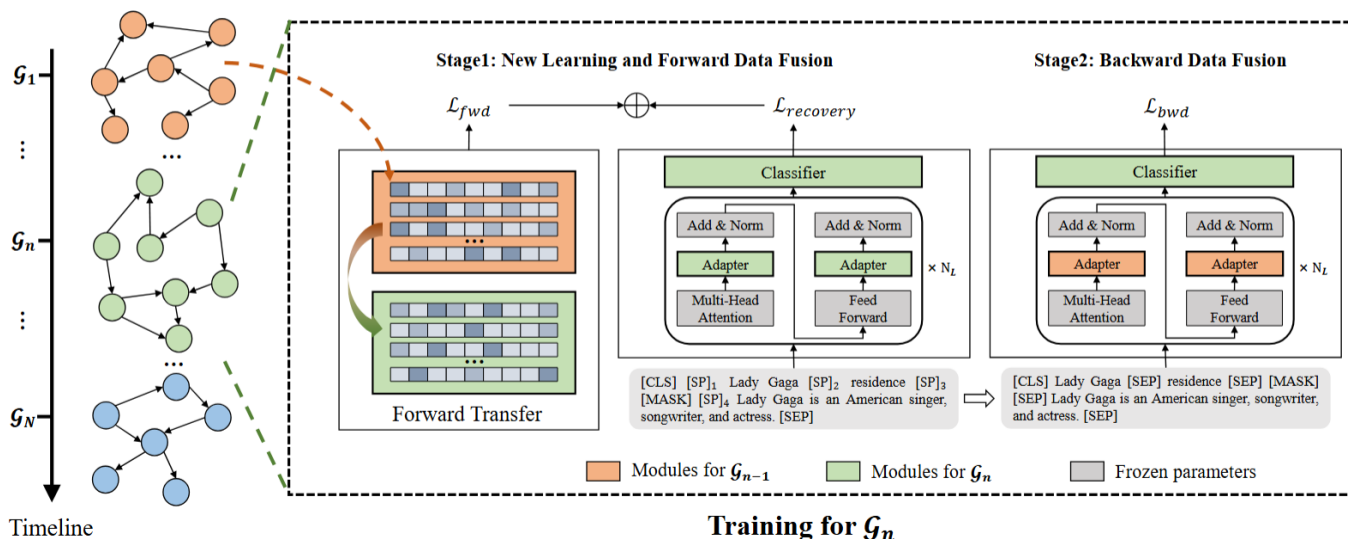
Missing Data Recovery for Heterogeneous Graphs with Incremental Multi-Source Data Fusion

Yang LIU, Xiaoxia JIANG, Yuanning CUI, Yu WANG,
Wei HU

Frontiers of Computer Science, DOI: [10.1007/s11704-025-41420-2](https://doi.org/10.1007/s11704-025-41420-2)

Problems & Ideas

- Problems of conventional missing data recovery methods
 - Existing works focus on improving a single heterogeneous graph
 - No support is provided for dynamically evolving multi-graph settings
- Ideas: A multi-source incremental fusion framework that integrates graph-specific adapters and cross-graph embedding alignment to improve missing data recovery in evolving heterogeneous graphs



Proposed SIKE framework. Left: Input heterogeneous graphs with overlapping nodes. Middle: Graph-specific adapters built on frozen PLM encode each graph independently. Right: Two-stage fusion aligns and transfers knowledge across graphs for incremental missing data recovery.

Main Contributions

- Contributions
 - A novel source-incremental missing data recovery task that leverages overlapping semantics across heterogeneous graphs for progressive enhancement
 - A unified framework with graph-specific adapters and a two-stage fusion mechanism to enable continual representation learning across graphs
 - A cross-graph embedding alignment strategy that supports efficient and robust recovery, achieving superior performance on real-world datasets

Models	DBpedia15K				Wikidata15K				YAGO15K			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
Indep. training	0.329	0.251	0.353	0.484	0.369	0.304	0.388	0.497	0.290	0.231	0.304	0.406
Re-training	0.424	0.341	0.460	0.584	0.485	0.416	0.517	0.610	0.357	0.299	0.377	0.467
Finetuning	0.286	0.194	0.318	0.465	0.382	0.309	0.412	0.522	0.313	0.248	0.336	0.433
AnyBURL	0.316	0.246	0.339	0.452	0.361	0.294	0.390	0.501	0.260	0.207	0.274	0.369
TransE	0.279	0.161	0.342	0.475	0.321	0.241	0.352	0.471	0.270	0.207	0.288	0.391
RotatE	0.295	0.243	0.310	0.394	0.328	0.271	0.343	0.437	0.253	0.197	0.269	0.360
KG-ICL	0.321	0.240	0.351	0.478	0.367	0.300	0.393	0.493	0.285	0.219	0.313	0.409
EWC	0.319	0.226	0.356	<u>0.500</u>	<u>0.404</u>	<u>0.329</u>	<u>0.437</u>	<u>0.548</u>	<u>0.320</u>	<u>0.256</u>	<u>0.342</u>	<u>0.442</u>
EMR	0.256	0.186	0.285	0.394	0.360	0.301	0.383	0.475	0.310	0.245	0.329	0.435
Adapter	<u>0.337</u>	<u>0.260</u>	<u>0.363</u>	0.488	0.359	0.306	0.375	0.461	0.283	0.227	0.299	0.390
AdapterFusion	0.333	0.254	0.361	0.486	0.356	0.300	0.374	0.460	0.281	0.227	0.299	0.385
SIKE	0.360	0.277	0.391	0.522	0.434	0.370	0.462	0.551	0.335	0.278	0.354	0.445

Results of missing recovery with DBpedia15K → Wikidata15K → YAGO15K. SIKE outperforms all competing models across both datasets on most metrics, except re-training.