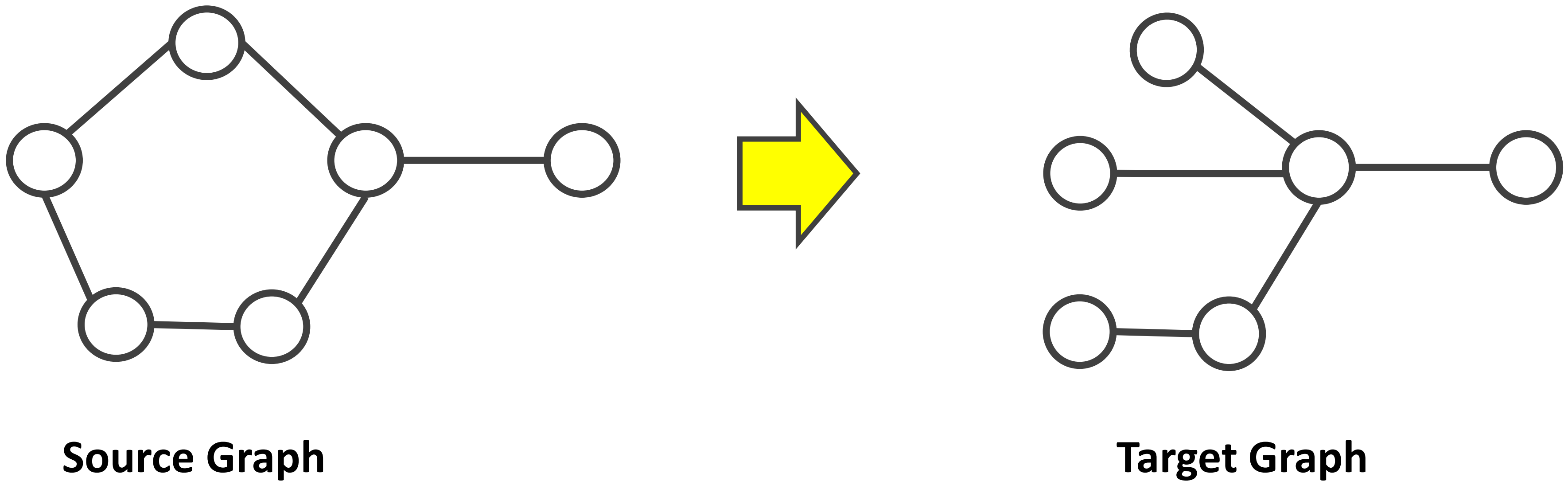
Slow Learning and Fast Inference: Efficient Graph Similarity Computation via Knowledge Distillation

Can Qin[1], Handong Zhao[2], Lichen Wang[2], Huan Wang[1], Yulun Zhang[1], Yun Fu[1]
[1]Northestern University [2]Adobe Research

Problem Introduction

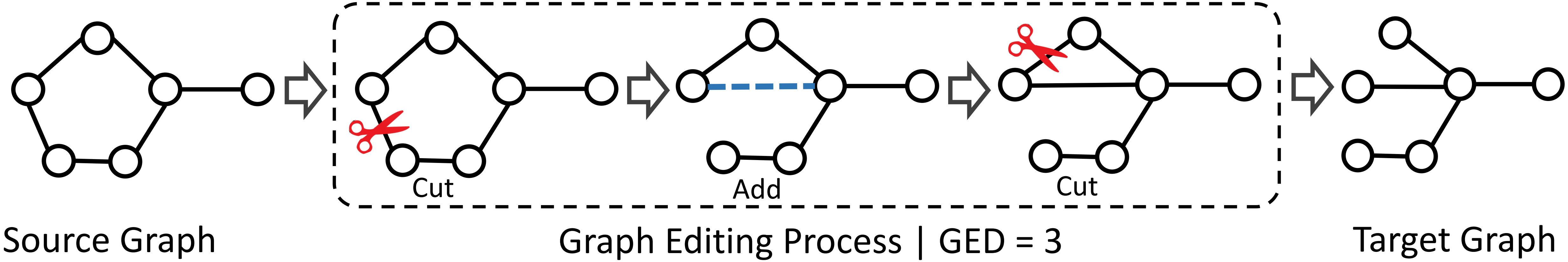
Topic: Graph Similarity Computation



How to Measure Their Similarity?

Problem Introduction

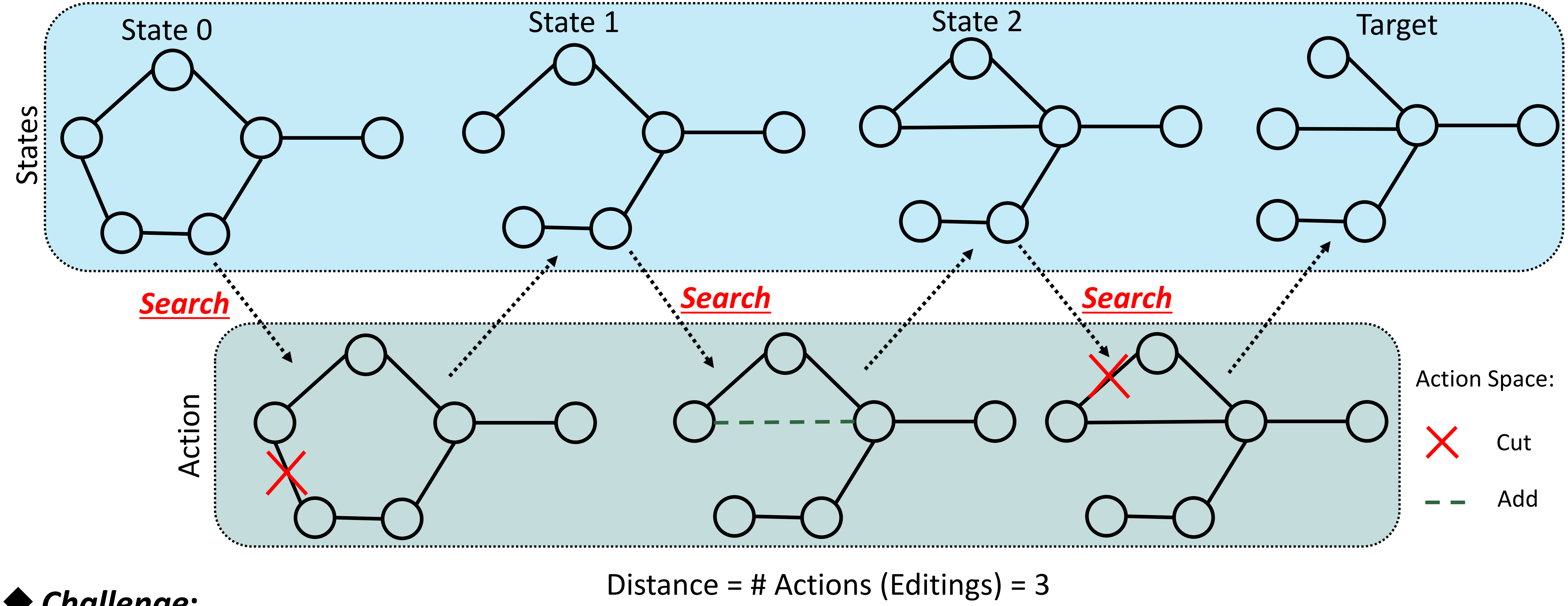
Topic: Graph Similarity Computation - GED



Concept of Graph Editing Distance (GED)

Challenges of Exact GED Solvers

Topic: NP-Hard

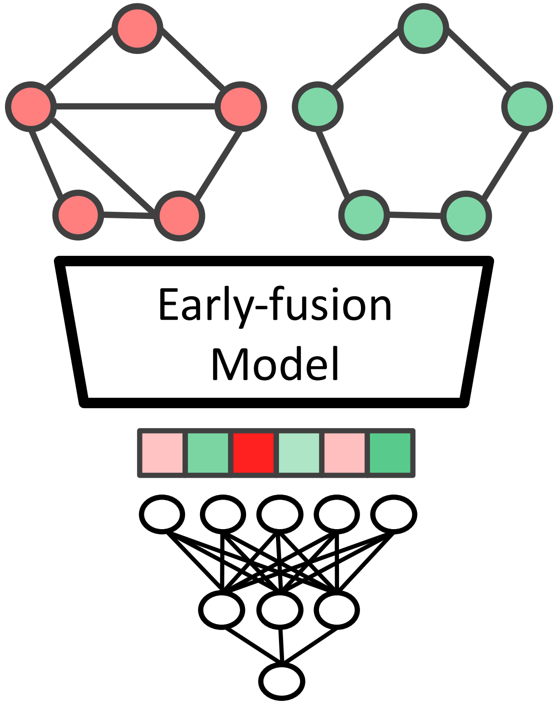


◆ ***Challenge***:

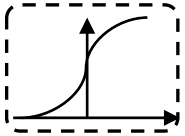
■ **Exact computation of GED is an NP-Hard problem, which is unable to scale up due to the complexity.**

Soft/Approximate GED Solution

Topic: Co-attention Model



$$nGED(G_i, G_j) = \frac{GED(G_i, G_j)}{(|G_i| + |G_j|)/2}$$

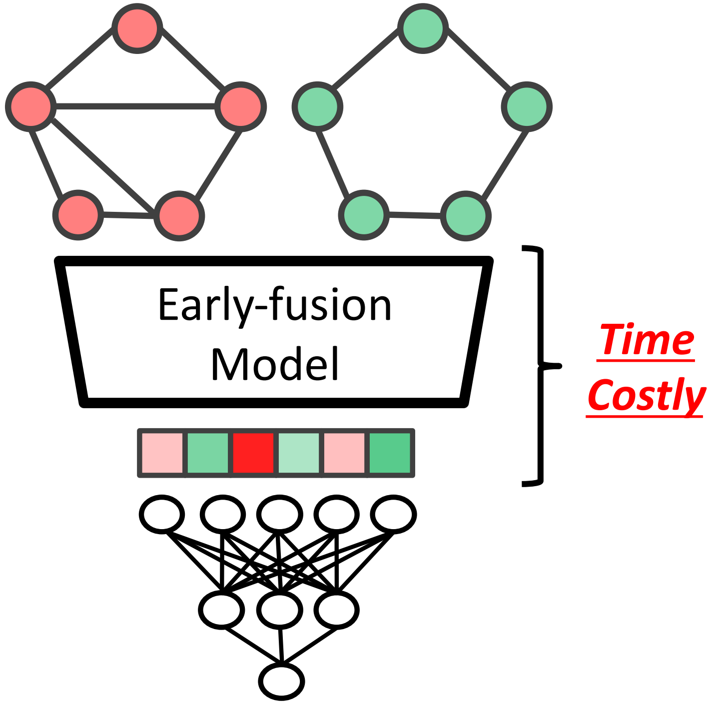


Deep Regression

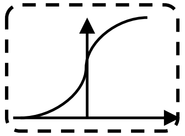
Bai, Yunsheng, et al. "Simgnn: A neural network approach to fast graph similarity computation." WSDM. 2019.
Li, Yujia, et al. "Graph matching networks for learning the similarity of graph structured objects." ICML. 2019.

Limitations of Co-attention Models

Topic: Low Efficiency of Co-attention Models



$$nGED(G_i, G_j) = \frac{GED(G_i, G_j)}{(|G_i| + |G_j|)/2}$$

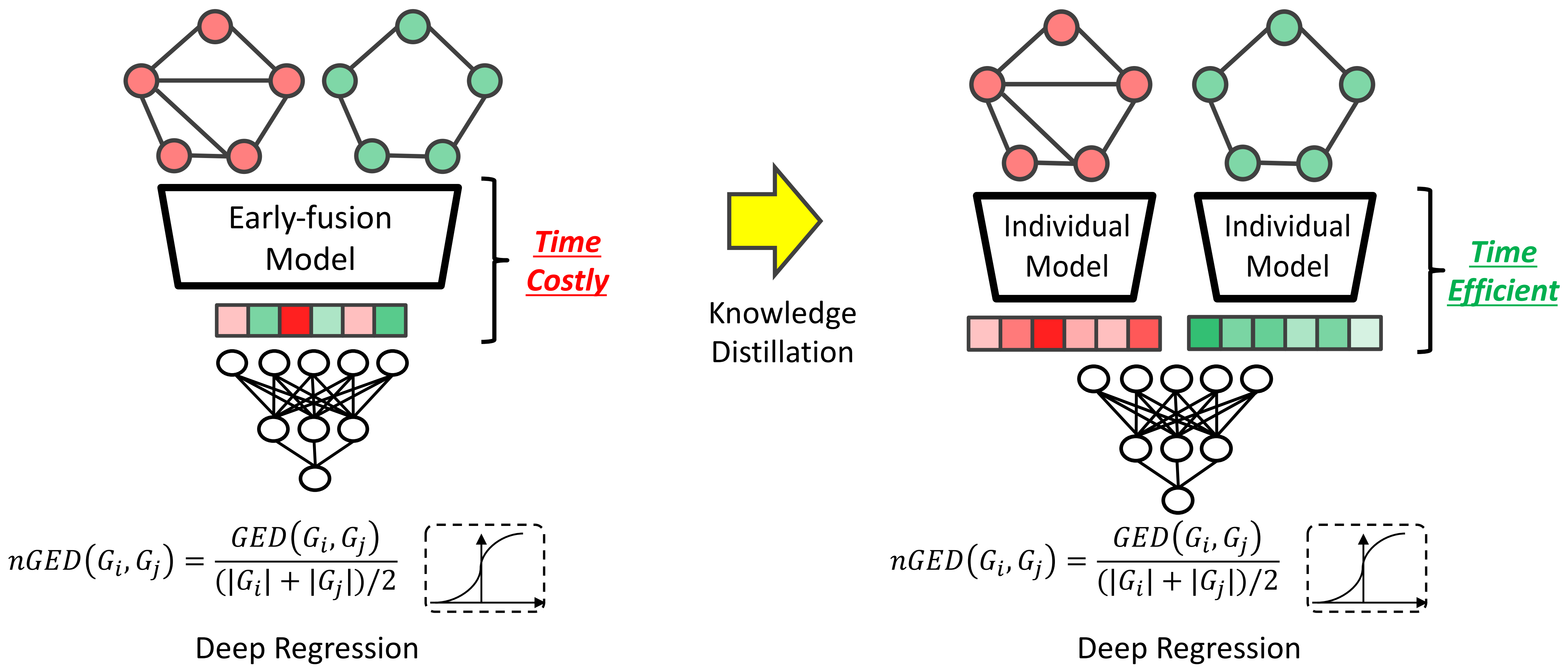


Deep Regression

Bai, Yunsheng, et al. "Simgnn: A neural network approach to fast graph similarity computation." WSDM. 2019.
Li, Yujia, et al. "Graph matching networks for learning the similarity of graph structured objects." ICML. 2019.

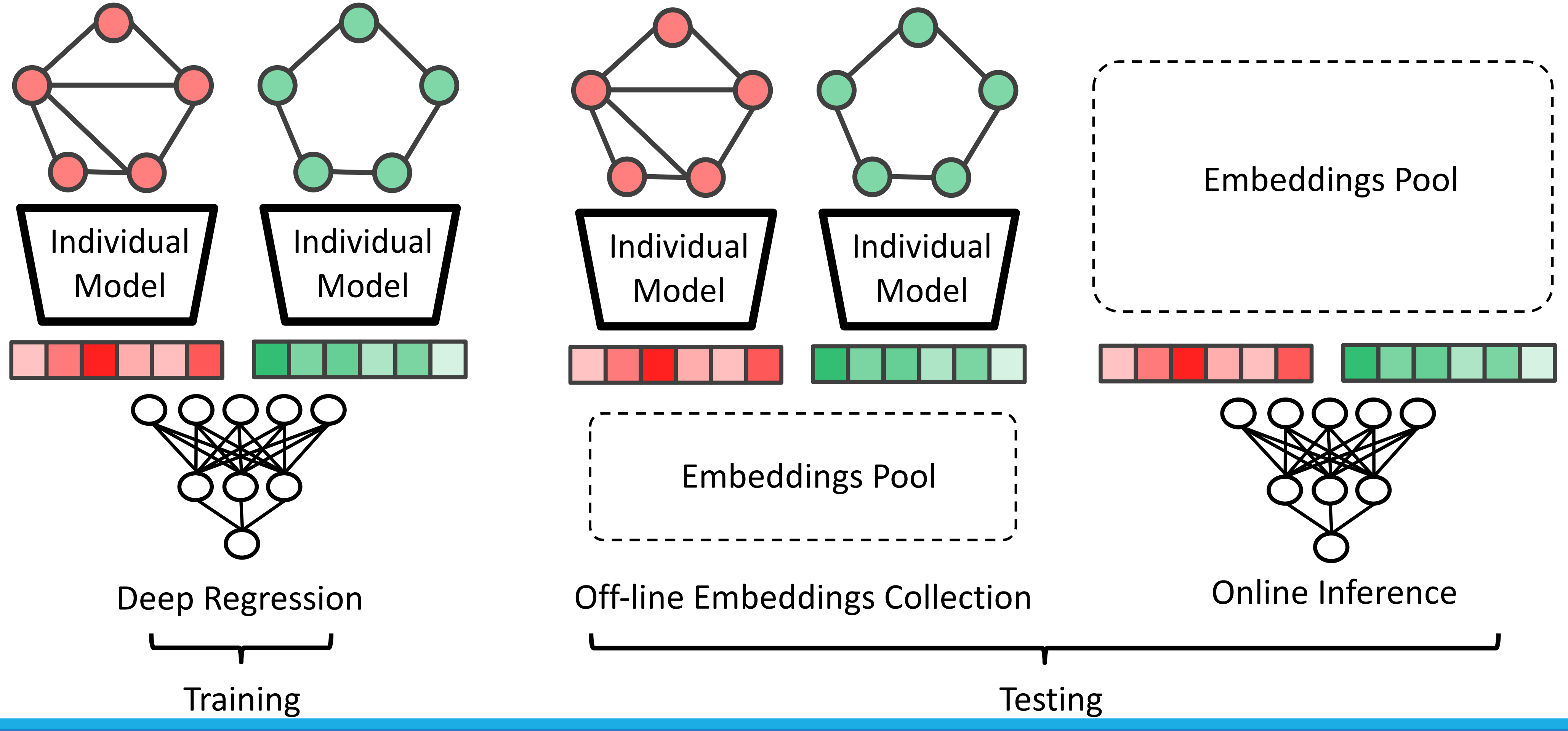
Our Motivation

Topic: Distill Co-attention Model to Siamese Models



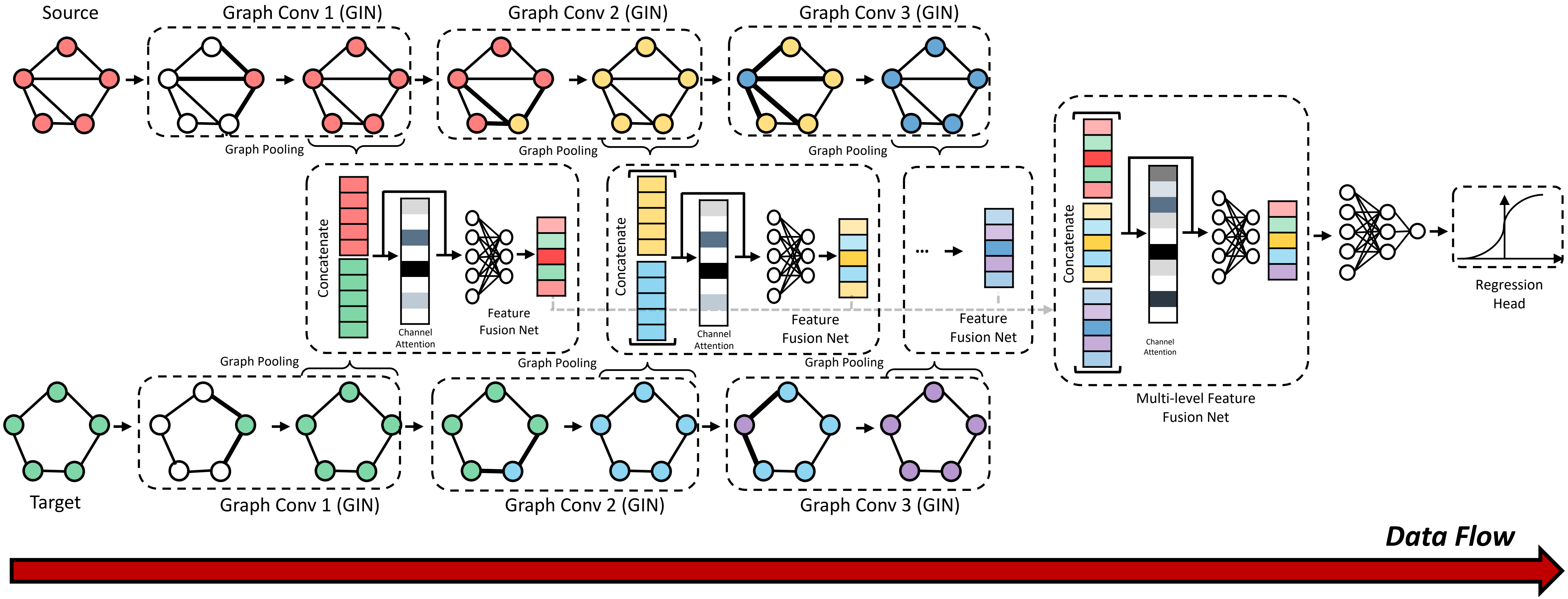
Our Motivation

Topic: Offline Embeddings Collection and Online Inference



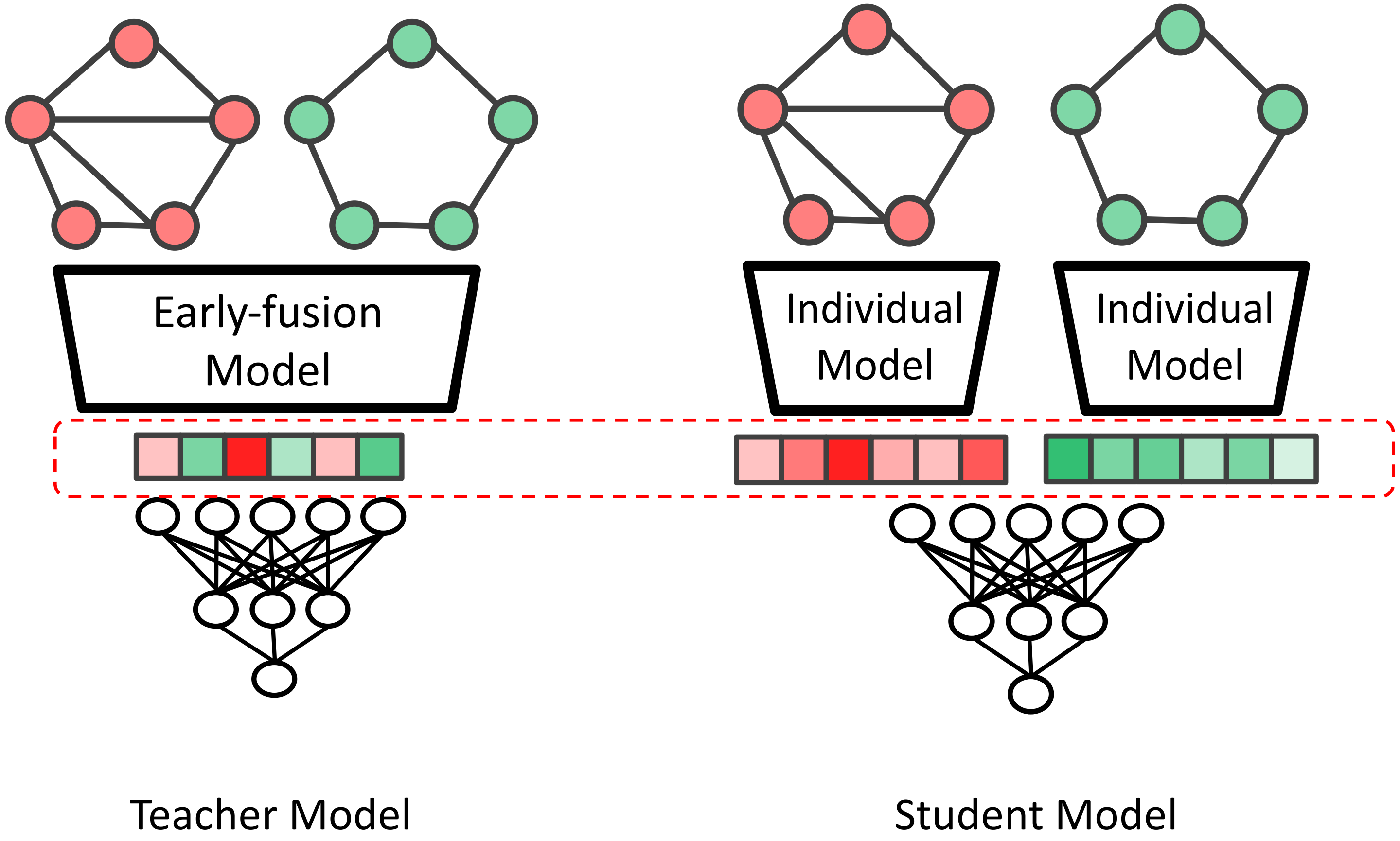
Proposed Approach: Teacher Network

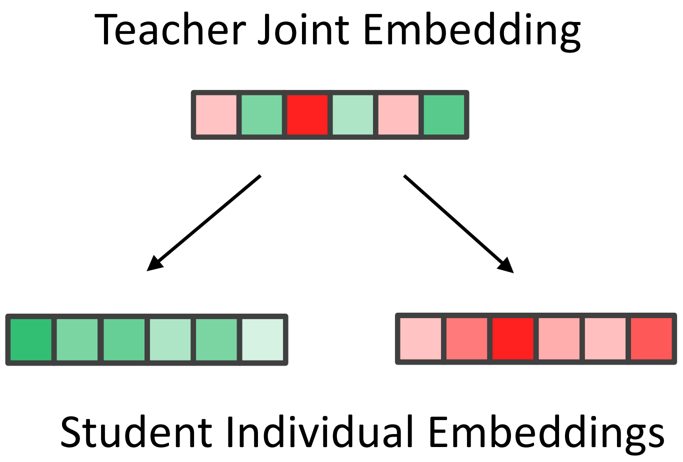
Topic: Proposed Early-fusion/Co-attention Network



Challenge: 1-to-2 Knowledge Distillation

Topic: How to distill individual embeddings from joint embedding?



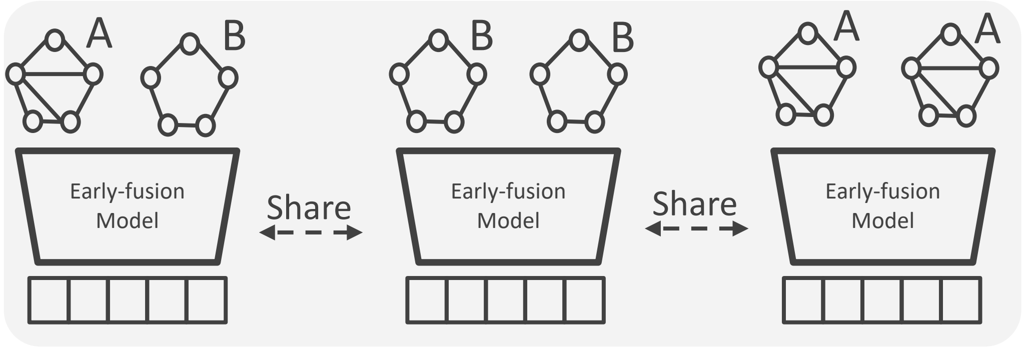


How to distill?

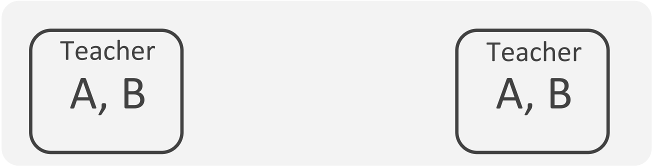
Proposed Approach: Knowledge Distillation

Topic: 1-to-2 Knowledge Distillation

Teacher Network



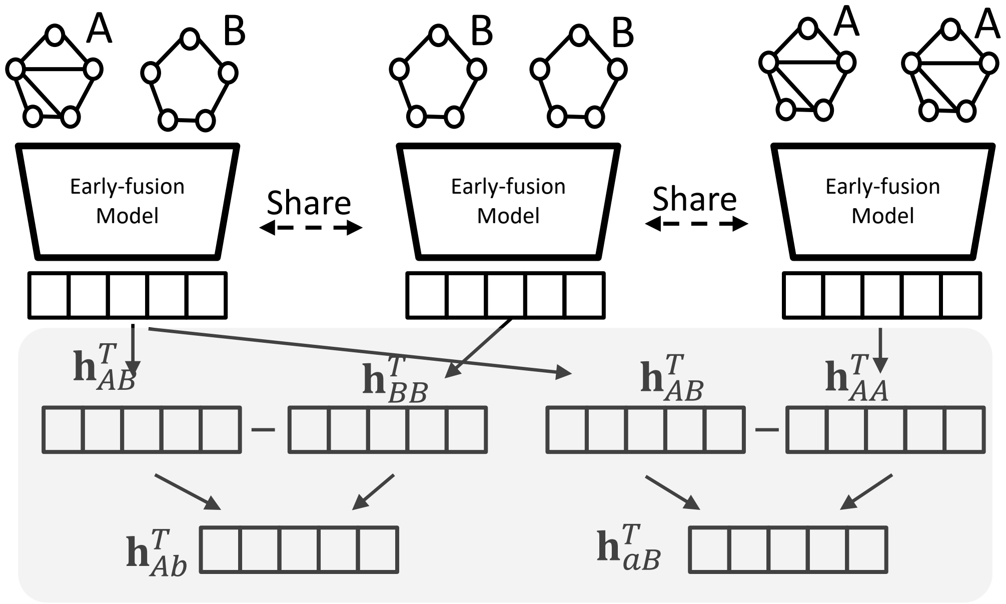
Outline of Embedding Decomposition



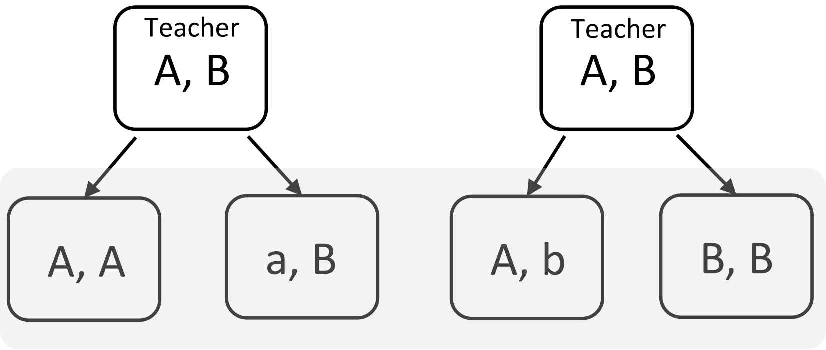
Proposed Approach: Knowledge Distillation

Topic: Offline Embeddings Collection and Online Inference

Teacher Network



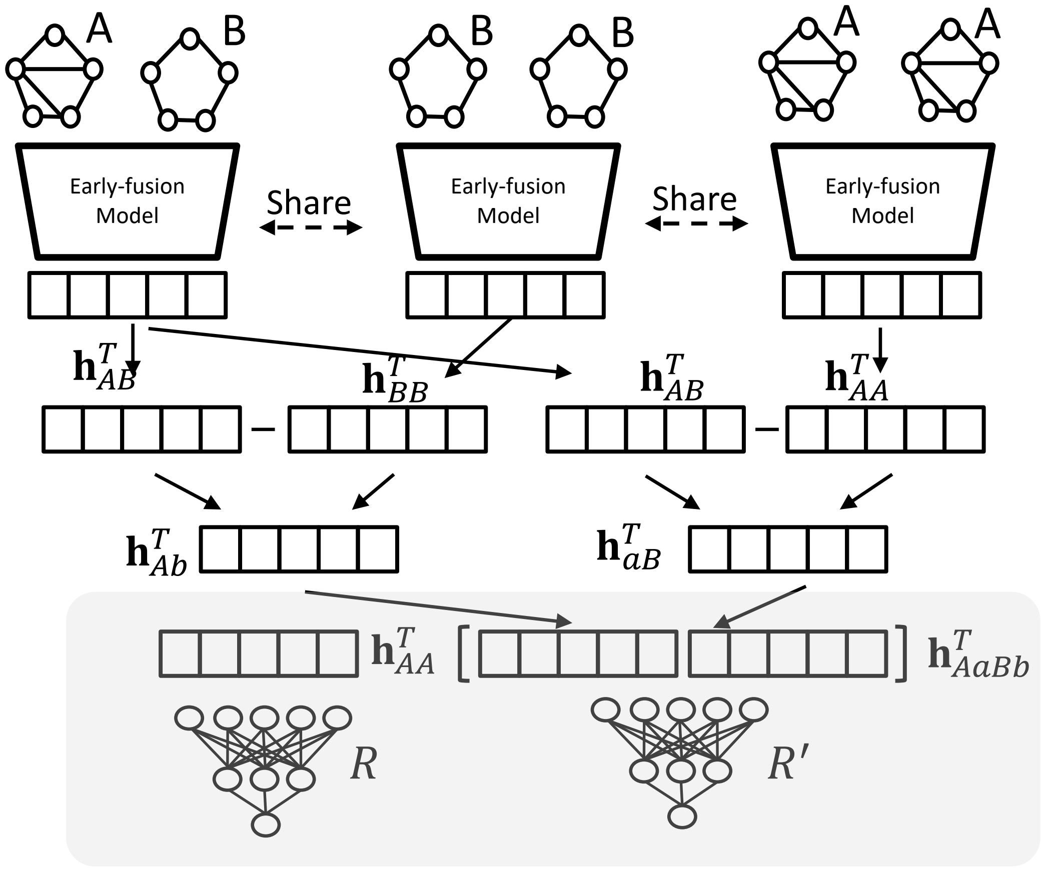
Outline of Embedding Decomposition



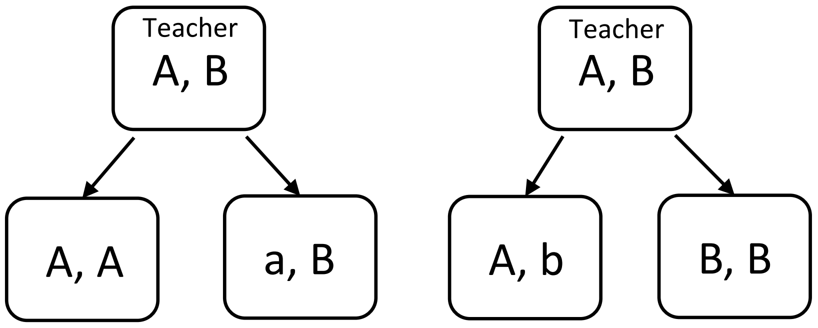
Proposed Approach: Knowledge Distillation

Topic: Offline Embeddings Collection and Online Inference

Teacher Network

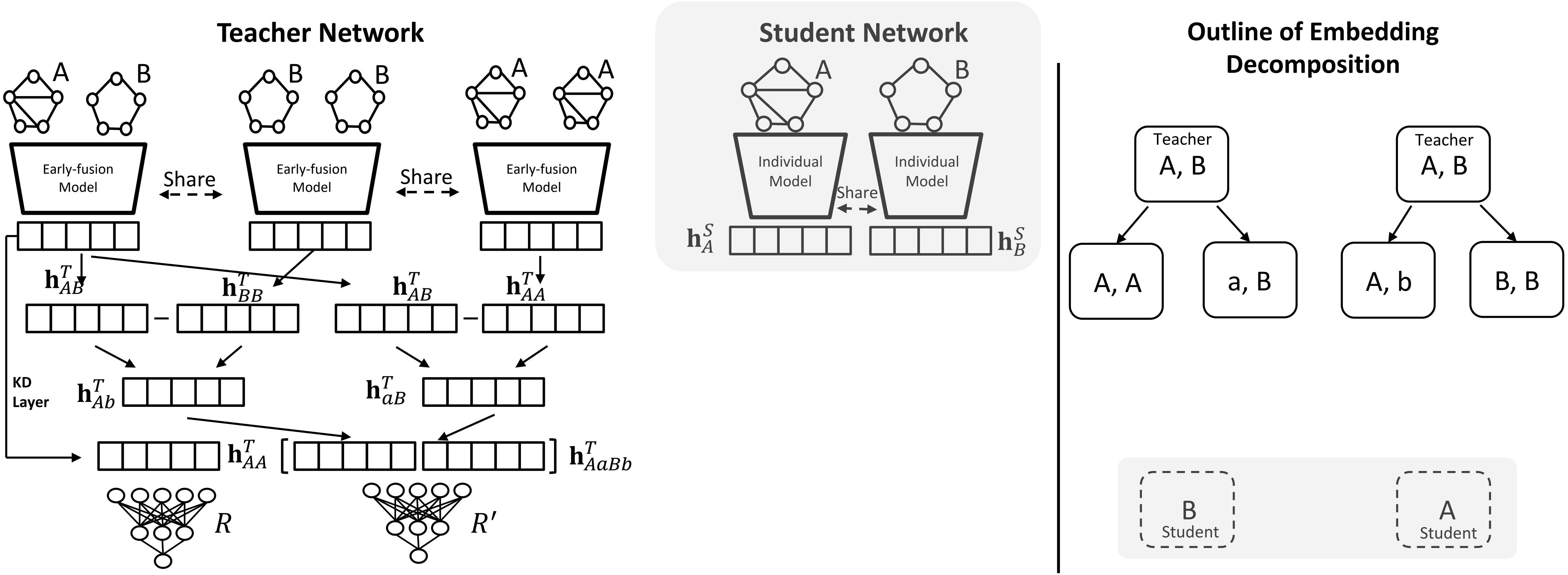


Outline of Embedding Decomposition



Proposed Approach: Knowledge Distillation

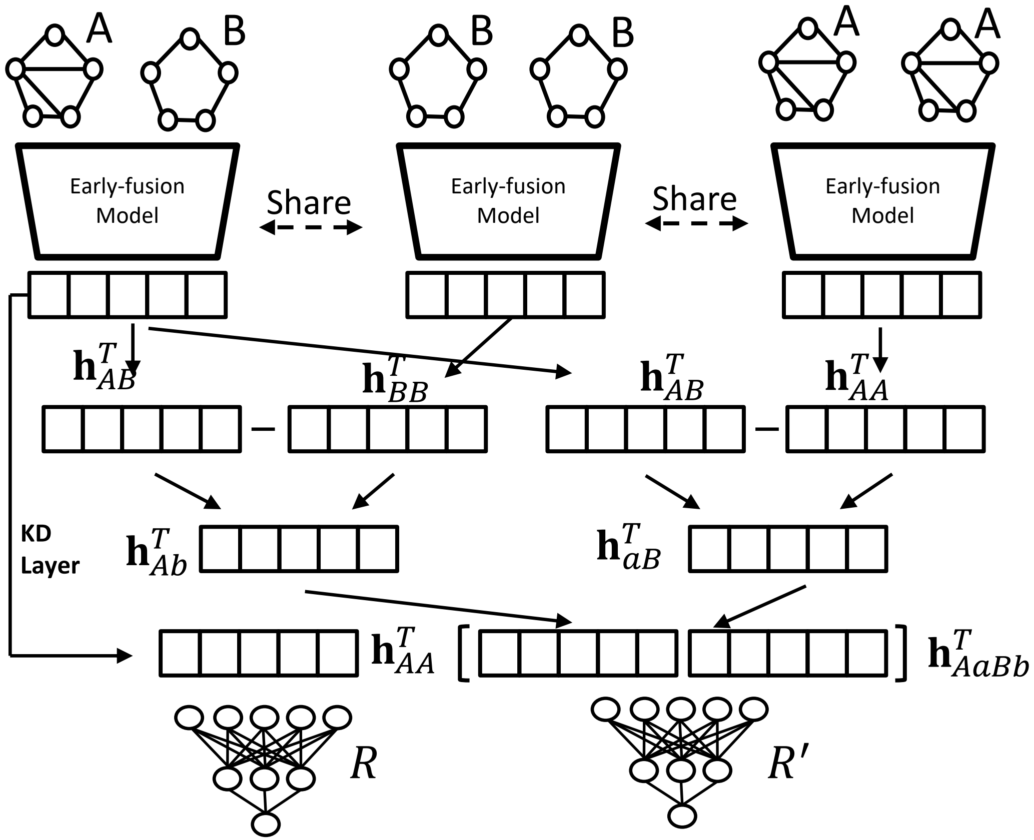
Topic: Offline Embeddings Collection and Online Inference



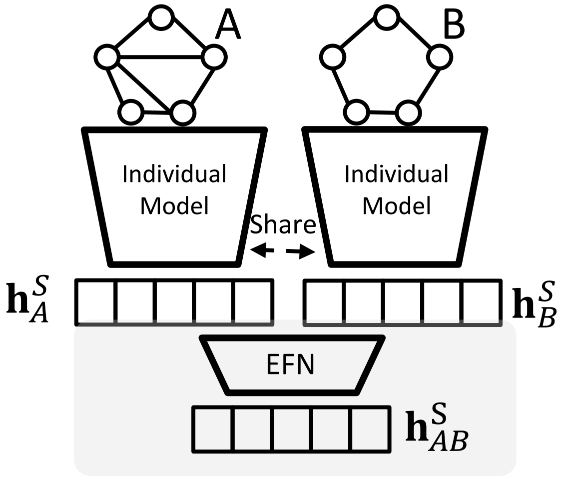
Proposed Approach: Knowledge Distillation

Topic: Offline Embeddings Collection and Online Inference

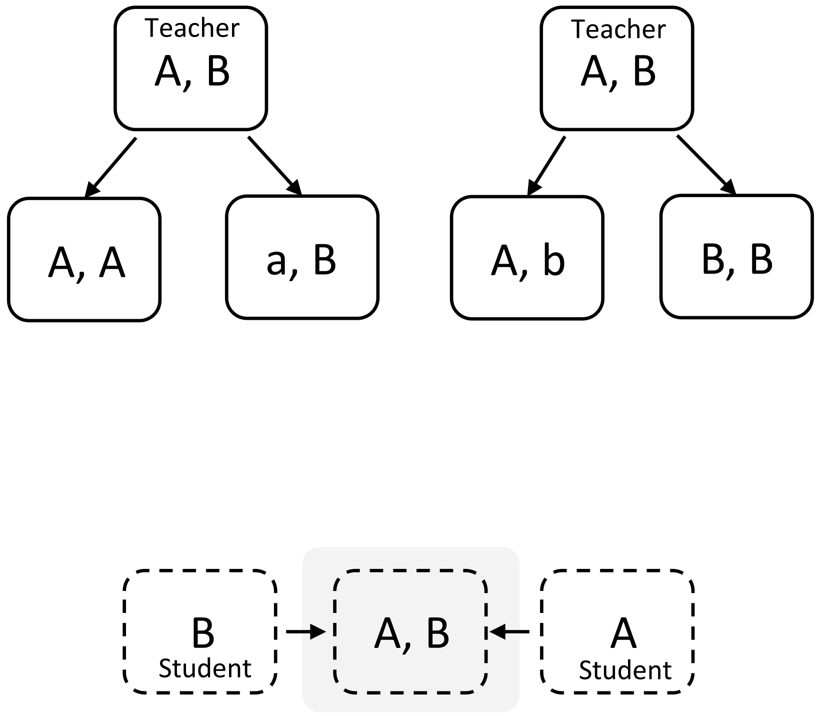
Teacher Network



Student Network

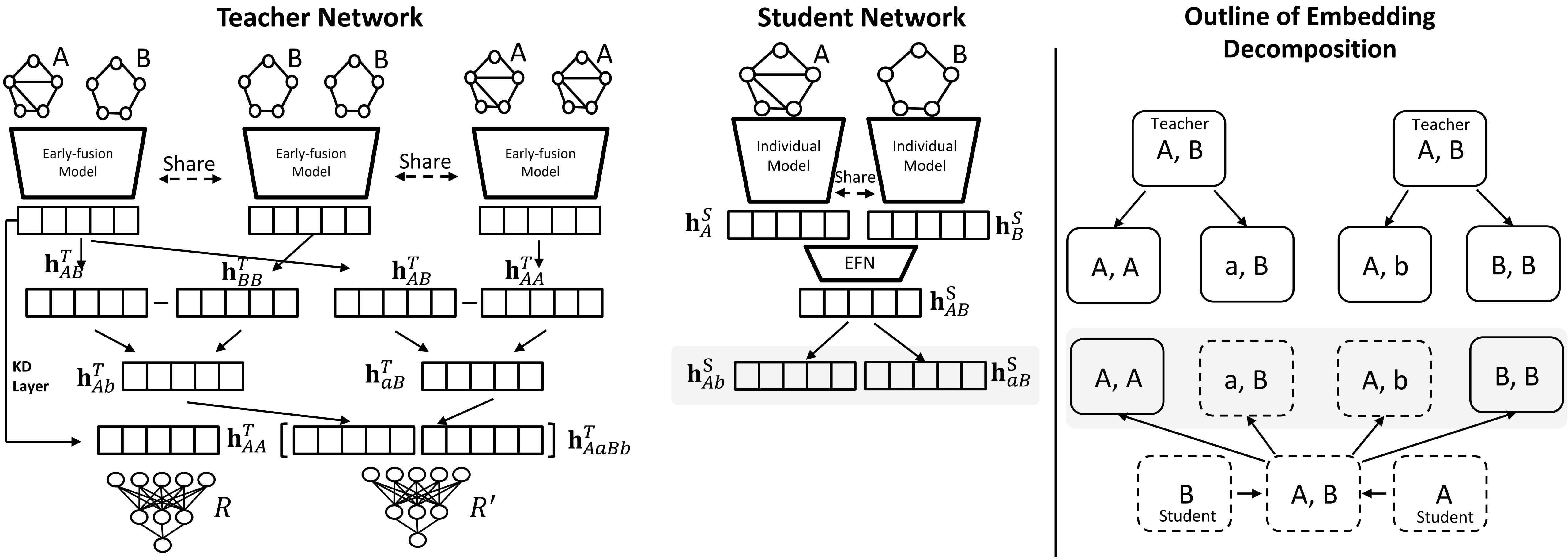


Outline of Embedding Decomposition



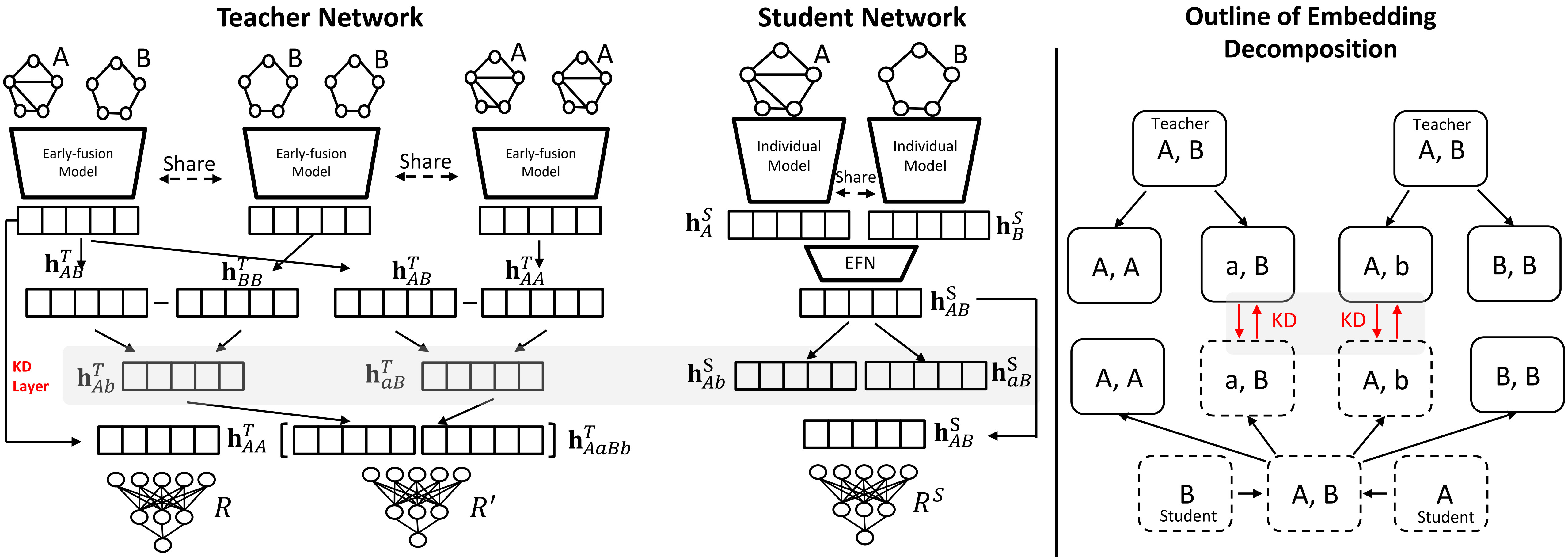
Proposed Approach: Knowledge Distillation

Topic: Offline Embeddings Collection and Online Inference



Proposed Approach: Knowledge Distillation

Topic: Offline Embeddings Collection and Online Inference



Experiments

Synergetic Media Learning Lab

Topic: Setup

- **Benchmarks**:
  - AIDS
  - LINUX
  - IMDB
  - ALKANE
- **Baselines:**
  - Beam, Hungarian, VJ
  - SimGNN, Extended-SimGNN
  - GMN
  - GENN-A*
- **Matrices**
  - Mean Squared Error (mse)
  - Spearman's Rank Correlation Coefficient
  - Kendall's Rank Correlation Coefficient
  - Precision at k (p@k), e.g., p@10, p@20
- **Framework**
  - PyG

Experiments

Synergetic Media Learning Lab

Topic: Quantitative Results

Table 1: Quantitative GED results of baselines and our method over AIDS, LINUX, IMDB and ALKANE.

Methods	AIDS					LINUX				
	mse ↓	ρ ↑	τ ↑	p@10 ↑	p@20 ↑	mse ↓	ρ ↑	τ ↑	p@10 ↑	p@20 ↑
Beam	12.09	0.609	0.463	0.481	0.493	9.268	0.827	0.714	0.973	0.924
Hungarian	25.30	0.510	0.378	0.360	0.392	29.81	0.638	0.517	0.913	0.836
VJ	29.16	0.517	0.383	0.310	0.345	63.86	0.581	0.450	0.287	0.251
GENN-A*	**0.635**	**0.959**	-	**0.871**	-	0.324	**0.991**	-	0.962	-
SimGNN	1.189	0.843	0.690	0.421	0.514	1.509	0.939	0.830	0.942	0.933
E-SimGNN	2.096	0.869	0.699	0.534	0.641	0.469	0.982	0.892	0.971	0.968
GMN	1.886	0.751	-	0.401	-	1.027	0.933	-	0.833	-
GraphSim	0.787	0.874	-	0.534	-	**0.058**	0.981	-	0.992	-
Teacher	1.601	0.901	**0.739**	0.658	**0.729**	0.163	0.988	**0.908**	**0.994**	**0.998**
Student	1.546	0.898	0.736	0.649	0.724	0.293	0.984	0.898	0.978	0.983

Methods	IMDB					ALKANE				
	mse	ρ	τ	p@10	p@20	mse	ρ	τ	p@10	p@20
SimGNN	1.264	0.878	0.770	0.759	0.777	2.446	0.859	0.686	0.87	0.782
E-SimGNN	1.148	0.864	0.75	0.806	0.807	1.622	0.886	0.722	0.982	0.955
GMN	4.422	0.725	-	0.604	-	-	-	-	-	-
GraphSim	0.743	0.926	-	0.828	-	-	-	-	-	-
Teacher	**0.553**	**0.938**	**0.829**	**0.872**	**0.878**	**0.533**	**0.930**	**0.787**	**0.998**	**0.991**
Student	0.581	0.935	0.826	0.857	0.869	1.198	0.899	0.741	0.993	0.978

Experiments

Synergetic Media Learning Lab

Topic: Ablation Study

Table 2: Ablation study results over the AIDS and IMDB datasets. **KD** represents the knowledge distillation.

Methods	KD	AIDS					IMDB				
		mse	ρ	τ	p@10	p@20	mse	ρ	τ	p@10	p@20
w/o Attn	✗	1.762	0.899	0.737	0.651	0.724	0.752	0.933	0.823	0.856	0.868
w/o GIN	✗	2.158	0.863	0.691	0.535	0.637	0.594	0.926	0.803	0.862	0.866
Single Level	✗	1.824	0.875	0.706	0.576	0.658	0.690	0.930	0.815	0.850	0.865
Student	✗	1.770	0.882	0.717	0.601	0.683	0.763	0.928	0.813	0.829	0.851
Teacher	✗	**1.601**	**0.901**	**0.739**	**0.658**	**0.729**	**0.553**	**0.938**	**0.829**	**0.872**	**0.878**
Joint Feat	✓	2.258	0.874	0.703	0.588	0.679	1.032	0.872	0.761	0.814	0.829
1st Order	✓	1.604	0.894	0.731	0.614	0.715	**0.548**	0.934	0.824	0.856	0.865
2nd Order	✓	1.647	0.893	0.731	0.631	0.715	0.692	0.929	0.814	0.847	0.866
w/o $\mathcal{L}'_{reg}$	✓	1.711	0.890	0.726	0.612	0.710	0.694	0.926	0.811	0.842	0.860
Student	✓	**1.546**	**0.898**	**0.736**	**0.649**	**0.724**	0.581	**0.935**	**0.826**	**0.857**	**0.869**

Experiments

Topic: Time Cost and Case Study

Table 3: Inference time to solve GED computation on AIDS. Student-R means the student model with raw input graphs. Student-F denotes that the embeddings are stored offline, which can be online loaded for inference.

Model	GENN-A*	SimGNN	E-SimGNN	E-SimGNN-F	Teacher	Student-R	Student-F
Time	290.1h	11.139s	9.672s	3.464s	11.139s	10.149s	**0.148s**

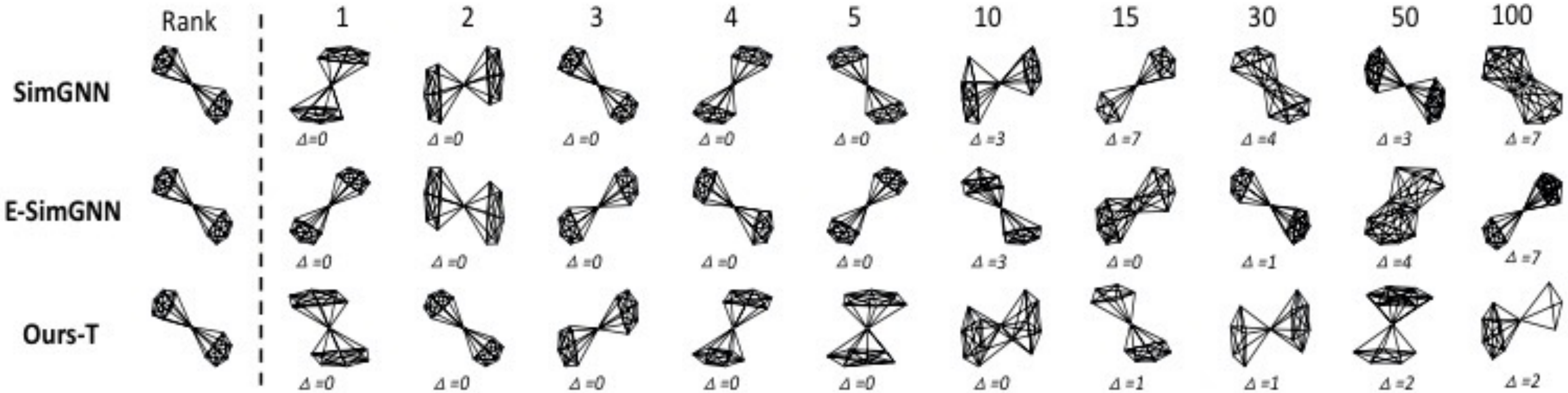


Figure 7: Ranking results of SimGNN, E-SimGNN and our teacher model on IMDB. Δ represents the absolute difference between the ground truth GED and the GED of predicted result.

Thank you!

Please contact: **qin.ca@northeastern.edu** for questions.

SMILE Lab

Electrical & Computer Engineering

Northeastern University