GraphCare: Enhancing Healthcare Predictions with Personalized Knowledge Graphs

Pengcheng Jiang[1], Cao Xiao[2], Adam Cross[3], Jimeng Sun[1]
[1] University of Illinois Urbana-Champaign
[2] GE HealthCare [3] OSF HealthCare

Paper

Code

ICLR

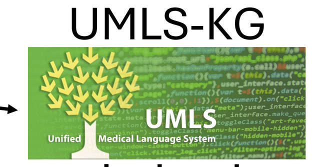
Introduction

1. Personalized KG is needed for personalized treatment.

2. Existing works mainly focus on
   1) simple hierarchical relations
   2) the inner graphical structure of EHR
   while there are many external biomedical KGs

LLM as KB

UMLS-KG



Q1: How to construct personalized KGs utilizing external knowledge bases?

A1: We construct medical concept-specific KGs

Q2: How to improve time-series clinical (EHR) predictions with those KGs?

A2: We treat personalized knowledge graphs as patient representations

Performance of GraphCare on MIMIC-III/IV across 4 tasks

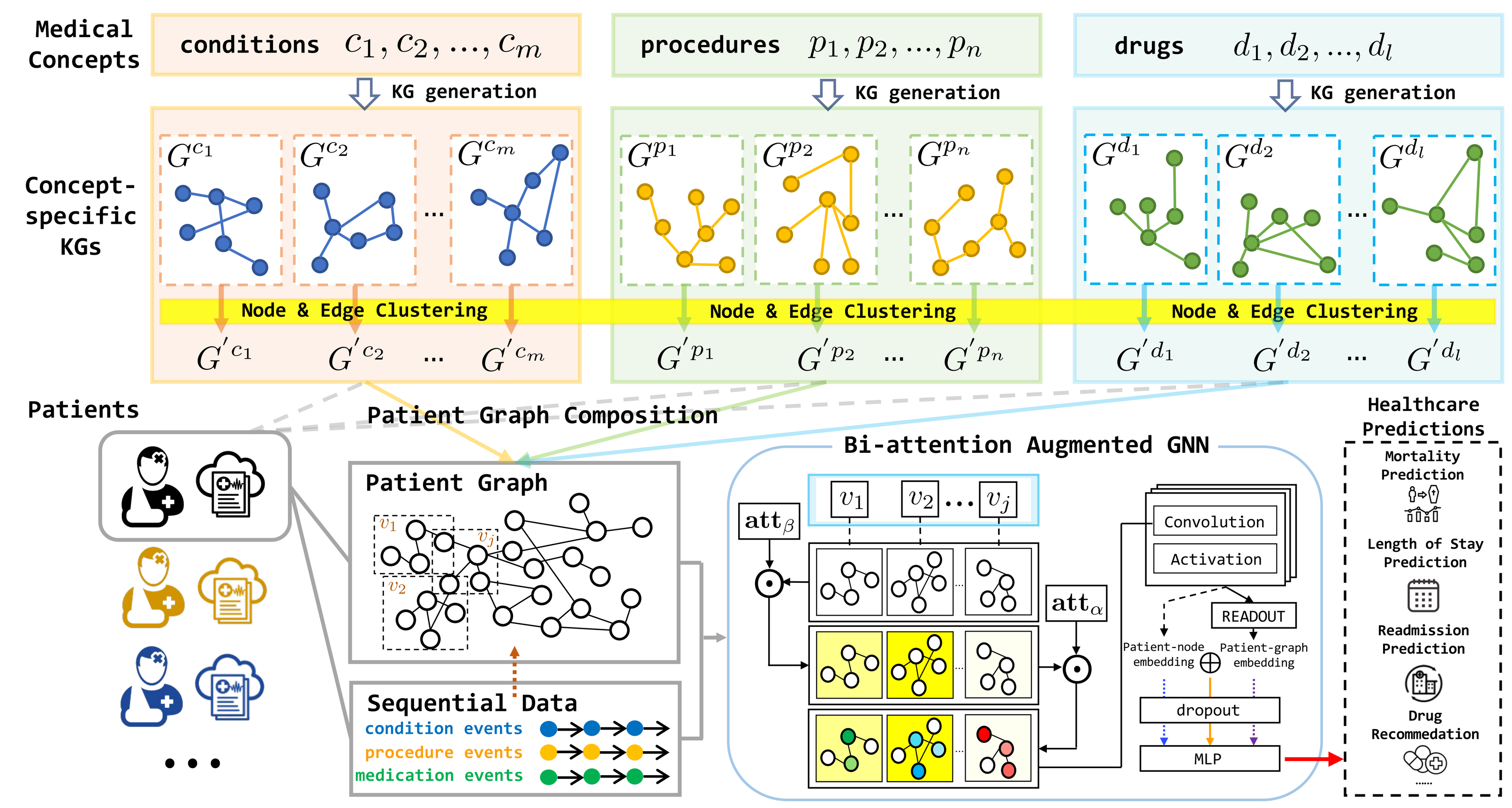
Model		Task 1: Mortality Prediction MIMIC-III AUPRC	MIMIC-III AUROC	MIMIC-IV AUPRC	MIMIC-IV AUROC	Task 2: Readmission Prediction MIMIC-III AUPRC	MIMIC-III AUROC	MIMIC-IV AUPRC	MIMIC-IV AUROC
GRU		$11.8_{(0.5)}$	$61.3_{(0.9)}$	$4.2_{(0.1)}$	$69.0_{(0.8)}$	$68.2_{(0.4)}$	$65.4_{(0.8)}$	$66.1_{(0.1)}$	$66.2_{(0.1)}$
Transformer		$10.1_{(0.9)}$	$57.2_{(1.3)}$	$3.4_{(0.4)}$	$65.1_{(1.2)}$	$67.3_{(0.7)}$	$63.9_{(1.1)}$	$65.7_{(0.3)}$	$65.3_{(0.4)}$
RETAIN		$9.6_{(0.6)}$	$59.4_{(1.5)}$	$3.8_{(0.4)}$	$64.8_{(1.6)}$	$65.1_{(1.0)}$	$64.1_{(0.7)}$	$66.2_{(0.3)}$	$66.3_{(0.2)}$
GRAM		$11.4_{(0.7)}$	$60.4_{(0.9)}$	$4.4_{(0.3)}$	$66.7_{(0.7)}$	$67.2_{(0.8)}$	$64.3_{(0.4)}$	$66.1_{(0.2)}$	$66.3_{(0.3)}$
Deepr		$13.2_{(1.1)}$	$60.8_{(0.4)}$	$4.2_{(0.2)}$	$68.9_{(0.9)}$	$68.8_{(0.9)}$	$66.5_{(0.4)}$	$65.6_{(0.1)}$	$65.4_{(0.2)}$
AdaCare		$11.1_{(0.4)}$	$58.4_{(1.4)}$	$4.6_{(0.3)}$	$69.3_{(0.7)}$	$68.6_{(0.6)}$	$65.7_{(0.3)}$	$65.9_{(0.0)}$	$66.1_{(0.0)}$
GRASP		$9.9_{(1.1)}$	$59.2_{(1.4)}$	$4.7_{(0.1)}$	$68.4_{(1.0)}$	$69.2_{(0.4)}$	$66.3_{(0.6)}$	$66.3_{(0.3)}$	$66.1_{(0.2)}$
StageNet		$12.4_{(0.3)}$	$61.5_{(0.7)}$	$4.2_{(0.3)}$	$69.6_{(0.8)}$	$69.3_{(0.6)}$	$66.7_{(0.4)}$	$66.1_{(0.1)}$	$66.2_{(0.1)}$
GraphCare	w/ GAT	$14.3_{(0.8)}$	$67.8_{(1.1)}$	$5.1_{(0.1)}$	$71.8_{(1.0)}$	$71.5_{(0.7)}$	$68.1_{(0.6)}$	$67.4_{(0.4)}$	$67.3_{(0.4)}$
	w/ GINE	$14.4_{(0.4)}$	$67.3_{(1.3)}$	$5.7_{(0.1)}$	$72.0_{(1.1)}$	$71.3_{(0.8)}$	$68.0_{(0.4)}$	$68.3_{(0.3)}$	$67.5_{(0.4)}$
	w/ EGT	$15.5_{(0.5)}$	$69.1_{(1.0)}$	$6.2_{(0.2)}$	$71.3_{(0.7)}$	$72.2_{(0.5)}$	$68.8_{(0.4)}$	$68.9_{(0.2)}$	$67.6_{(0.3)}$
	w/ GPS	$15.3_{(0.9)}$	$68.8_{(0.8)}$	$\mathbf{6.7}_{(0.2)}$	$72.7_{(0.9)}$	$71.9_{(0.6)}$	$68.5_{(0.6)}$	$69.1_{(0.4)}$	$67.9_{(0.4)}$
	w/ BAT	$\mathbf{16.7}_{(0.5)}$	$\mathbf{70.3}_{(0.5)}$	$\mathbf{6.7}_{(0.3)}$	$\mathbf{73.1}_{(0.5)}$	$\mathbf{73.4}_{(0.4)}$	$\mathbf{69.7}_{(0.5)}$	$\mathbf{69.6}_{(0.3)}$	$\mathbf{68.5}_{(0.4)}$

Model		Task 3: Length of Stay Prediction MIMIC-III AUROC	MIMIC-III Kappa	MIMIC-III Accuracy	MIMIC-III F1-score	MIMIC-IV AUROC	MIMIC-IV Kappa	MIMIC-IV Accuracy	MIMIC-IV F1-score
GRU		$78.3_{(0.1)}$	$26.2_{(0.2)}$	$40.3_{(0.3)}$	$34.9_{(0.5)}$	$78.7_{(0.1)}$	$26.0_{(0.1)}$	$35.2_{(0.1)}$	$31.6_{(0.2)}$
Transformer		$78.3_{(0.2)}$	$25.4_{(0.4)}$	$40.1_{(0.3)}$	$34.8_{(0.2)}$	$78.3_{(0.3)}$	$25.3_{(0.4)}$	$34.4_{(0.2)}$	$31.4_{(0.3)}$
RETAIN		$78.2_{(0.1)}$	$26.1_{(0.4)}$	$40.6_{(0.3)}$	$34.9_{(0.4)}$	$78.9_{(0.3)}$	$26.3_{(0.2)}$	$35.7_{(0.2)}$	$32.0_{(0.2)}$
GRAM		$78.2_{(0.1)}$	$26.3_{(0.3)}$	$40.4_{(0.4)}$	$34.5_{(0.2)}$	$78.8_{(0.2)}$	$26.1_{(0.4)}$	$35.4_{(0.2)}$	$31.9_{(0.3)}$
Deepr		$77.9_{(0.1)}$	$25.3_{(0.4)}$	$40.1_{(0.6)}$	$35.0_{(0.4)}$	$79.5_{(0.3)}$	$26.4_{(0.2)}$	$35.8_{(0.3)}$	$32.3_{(0.1)}$
StageNet		$78.3_{(0.2)}$	$24.8_{(0.2)}$	$39.9_{(0.2)}$	$34.4_{(0.4)}$	$79.2_{(0.3)}$	$26.0_{(0.2)}$	$35.0_{(0.2)}$	$31.3_{(0.3)}$
GraphCare	w/ GAT	$79.4_{(0.3)}$	$28.2_{(0.2)}$	$41.9_{(0.2)}$	$36.1_{(0.4)}$	$80.3_{(0.3)}$	$28.4_{(0.4)}$	$36.2_{(0.1)}$	$33.3_{(0.3)}$
	w/ GINE	$79.2_{(0.2)}$	$28.3_{(0.3)}$	$41.5_{(0.3)}$	$36.0_{(0.4)}$	$79.9_{(0.2)}$	$27.5_{(0.3)}$	$36.3_{(0.3)}$	$32.8_{(0.2)}$
	w/ EGT	$80.3_{(0.3)}$	$28.8_{(0.2)}$	$42.8_{(0.4)}$	$36.3_{(0.5)}$	$80.5_{(0.2)}$	$28.7_{(0.3)}$	$36.7_{(0.2)}$	$33.5_{(0.1)}$
	w/ GPS	$80.9_{(0.3)}$	$28.8_{(0.4)}$	$43.0_{(0.3)}$	$36.8_{(0.4)}$	$80.7_{(0.3)}$	$28.8_{(0.4)}$	$36.7_{(0.3)}$	$33.9_{(0.3)}$
	w/ BAT	$\mathbf{81.4}_{(0.3)}$	$\mathbf{29.5}_{(0.4)}$	$\mathbf{43.2}_{(0.4)}$	$\mathbf{37.5}_{(0.2)}$	$\mathbf{81.7}_{(0.2)}$	$\mathbf{29.8}_{(0.3)}$	$\mathbf{37.3}_{(0.3)}$	$\mathbf{34.2}_{(0.3)}$

Model		Task 4: Drug Recommendation MIMIC-III AUPRC	MIMIC-III AUROC	MIMIC-III F1-score	MIMIC-III Jaccard	MIMIC-IV AUPRC	MIMIC-IV AUROC	MIMIC-IV F1-score	MIMIC-IV Jaccard
GRU		$77.0_{(0.1)}$	$94.4_{(0.0)}$	$62.3_{(0.3)}$	$47.8_{(0.3)}$	$74.1_{(0.1)}$	$94.2_{(0.1)}$	$60.2_{(0.2)}$	$44.0_{(0.4)}$
Transformer		$76.1_{(0.1)}$	$94.2_{(0.0)}$	$62.1_{(0.4)}$	$47.1_{(0.4)}$	$71.3_{(0.1)}$	$93.4_{(0.1)}$	$55.9_{(0.2)}$	$40.4_{(0.1)}$
RETAIN		$77.1_{(0.1)}$	$94.4_{(0.0)}$	$63.7_{(0.2)}$	$48.8_{(0.2)}$	$74.2_{(0.1)}$	$94.3_{(0.0)}$	$60.3_{(0.1)}$	$45.0_{(0.1)}$
GRAM		$76.7_{(0.1)}$	$94.2_{(0.1)}$	$62.9_{(0.3)}$	$47.9_{(0.3)}$	$74.3_{(0.2)}$	$94.2_{(0.1)}$	$60.1_{(0.2)}$	$45.3_{(0.3)}$
Deepr		$74.3_{(0.1)}$	$93.7_{(0.0)}$	$60.3_{(0.4)}$	$44.7_{(0.3)}$	$73.7_{(0.1)}$	$94.2_{(0.1)}$	$59.1_{(0.4)}$	$43.8_{(0.4)}$
StageNet		$74.4_{(0.1)}$	$93.0_{(0.1)}$	$61.4_{(0.3)}$	$45.8_{(0.4)}$	$74.4_{(0.1)}$	$94.2_{(0.0)}$	$60.2_{(0.3)}$	$45.4_{(0.4)}$
SafeDrug		$68.1_{(0.3)}$	$91.0_{(0.1)}$	$46.7_{(0.4)}$	$31.7_{(0.3)}$	$66.4_{(0.5)}$	$91.8_{(0.2)}$	$56.2_{(0.4)}$	$44.3_{(0.3)}$
MICRON		$77.4_{(0.0)}$	$94.6_{(0.1)}$	$63.2_{(0.4)}$	$48.3_{(0.4)}$	$74.4_{(0.1)}$	$94.3_{(0.1)}$	$59.3_{(0.3)}$	$44.1_{(0.3)}$
GAMENet		$76.4_{(0.0)}$	$94.2_{(0.1)}$	$62.1_{(0.4)}$	$47.2_{(0.4)}$	$74.2_{(0.1)}$	$94.2_{(0.1)}$	$60.4_{(0.4)}$	$45.3_{(0.3)}$
MoleRec		$69.8_{(0.1)}$	$92.0_{(0.1)}$	$58.1_{(0.1)}$	$43.1_{(0.3)}$	$68.6_{(0.1)}$	$92.1_{(0.1)}$	$56.3_{(0.4)}$	$40.6_{(0.3)}$
GraphCare	w/ GAT	$78.5_{(0.2)}$	$94.8_{(0.1)}$	$64.4_{(0.3)}$	$49.2_{(0.4)}$	$74.7_{(0.5)}$	$94.4_{(0.3)}$	$60.4_{(0.3)}$	$45.7_{(0.4)}$
	w/ GINE	$78.2_{(0.1)}$	$94.7_{(0.1)}$	$63.6_{(0.4)}$	$47.9_{(0.3)}$	$74.8_{(0.3)}$	$94.6_{(0.1)}$	$60.6_{(0.4)}$	$46.1_{(0.4)}$
	w/ EGT	$79.6_{(0.2)}$	$95.3_{(0.0)}$	$66.4_{(0.2)}$	$49.6_{(0.4)}$	$75.4_{(0.4)}$	$95.0_{(0.1)}$	$61.6_{(0.3)}$	$47.3_{(0.3)}$
	w/ GPS	$79.9_{(0.3)}$	$\mathbf{95.5}_{(0.1)}$	$66.2_{(0.3)}$	$\mathbf{49.8}_{(0.4)}$	$75.9_{(0.5)}$	$94.9_{(0.1)}$	$62.1_{(0.3)}$	$46.8_{(0.4)}$
	w/ BAT	$\mathbf{80.2}_{(0.2)}$	$\mathbf{95.5}_{(0.1)}$	$\mathbf{66.8}_{(0.2)}$	$49.7_{(0.3)}$	$\mathbf{77.1}_{(0.1)}$	$\mathbf{95.4}_{(0.2)}$	$\mathbf{63.9}_{(0.3)}$	$\mathbf{48.1}_{(0.3)}$

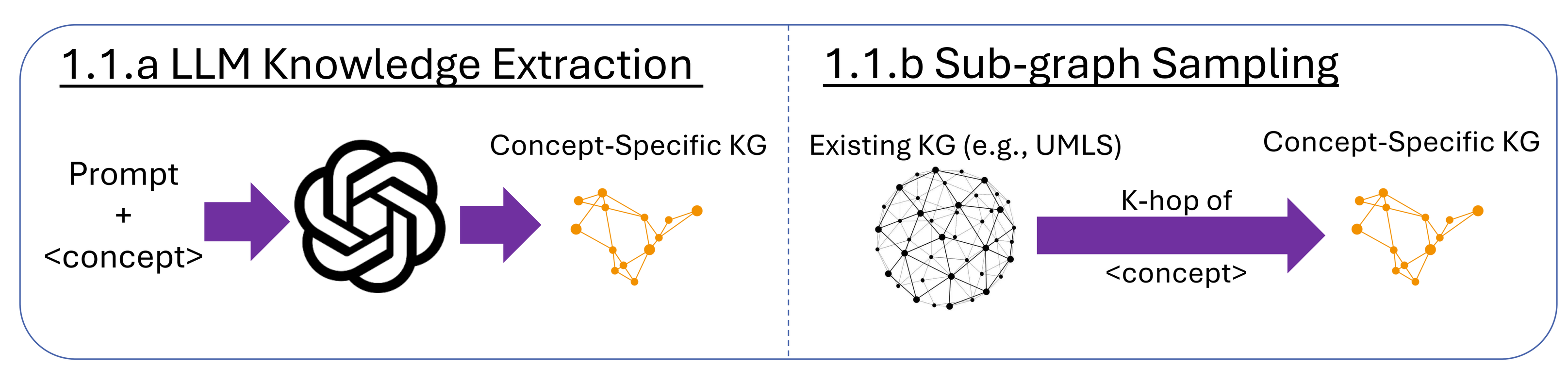
1. GraphCare consistently outperforms other methods on all tasks/datasets.
2. BAT outperforms other GNNs and graph transformers.

Method Overview



GraphCare Framework – Constructing Personalized Knowledge Graph for Various Healthcare Predictions.

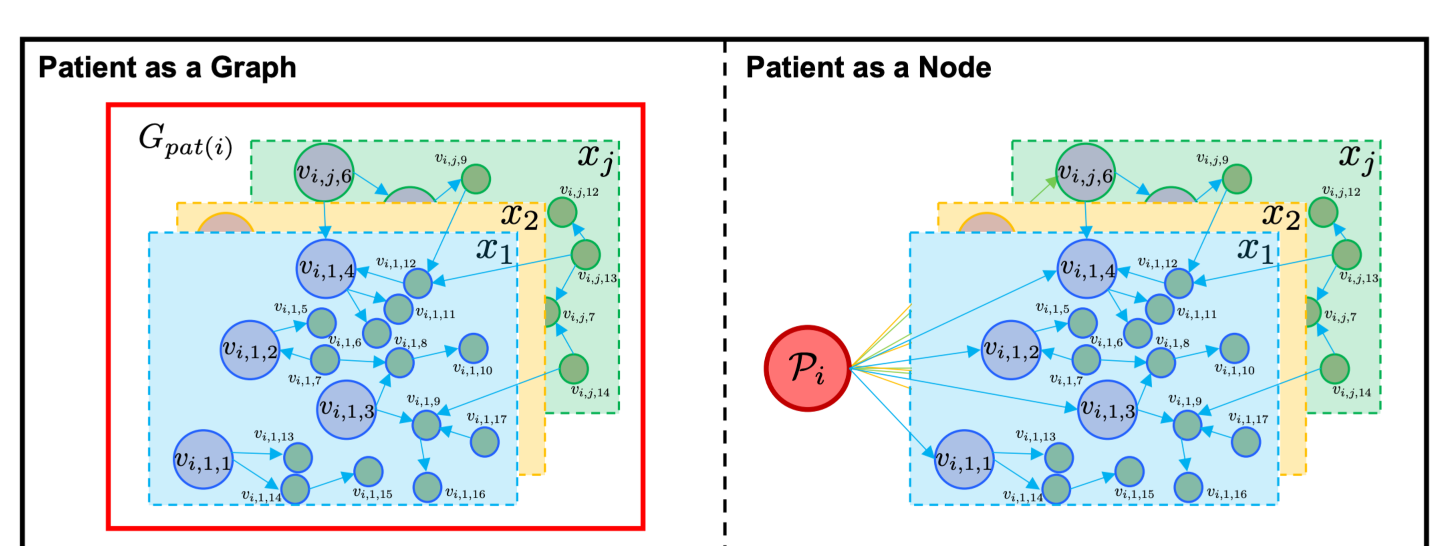
Step 1:
1.1 Generate concept-specific KGs for every medical concept using LLM prompts and by subsampling from existing KGs.



1.2 Perform node/edge clustering on across KGs.

Step 2:
For each patient, merge relevant concept-specific KGs to form a personalized KG.

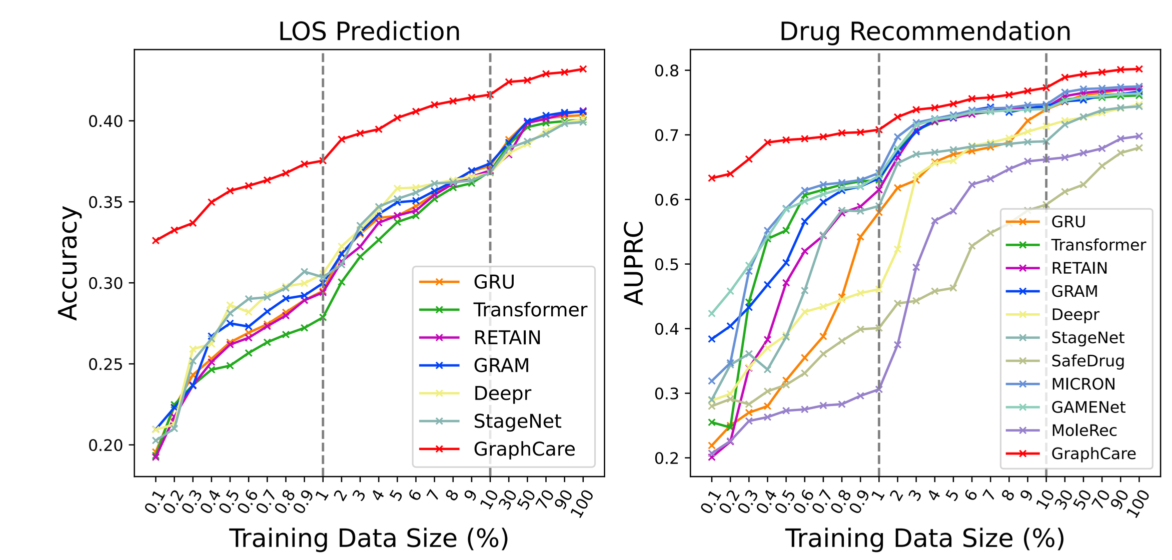
Step 3:
Employ Bi-attention Augmented (BAT) Graph Neural Network (GNN) to predict based on the personalized KGs.



$$\mathbf{h}_i^{G_{\text{pat}}} = \text{MEAN}(\sum_{j=1}^{J}\sum_{k=1}^{K_j}\mathbf{h}_{i,j,k}^{(L)}), \quad \mathbf{h}_i^{\mathcal{P}} = \text{MEAN}(\sum_{j=1}^{J}\sum_{k=1}^{K_j}\mathbb{1}_{i,j,k}^{\Delta}\mathbf{h}_{i,j,k}^{(L)}),$$

$$\mathbf{z}_i^{\text{graph}} = \text{MLP}(\mathbf{h}_i^{G_{\text{pat}}}), \quad \mathbf{z}_i^{\text{node}} = \text{MLP}(\mathbf{h}_i^{\mathcal{P}}), \quad \mathbf{z}_i^{\text{joint}} = \text{MLP}(\mathbf{h}_i^{G_{\text{pat}}} \oplus \mathbf{h}_i^{\mathcal{P}})$$

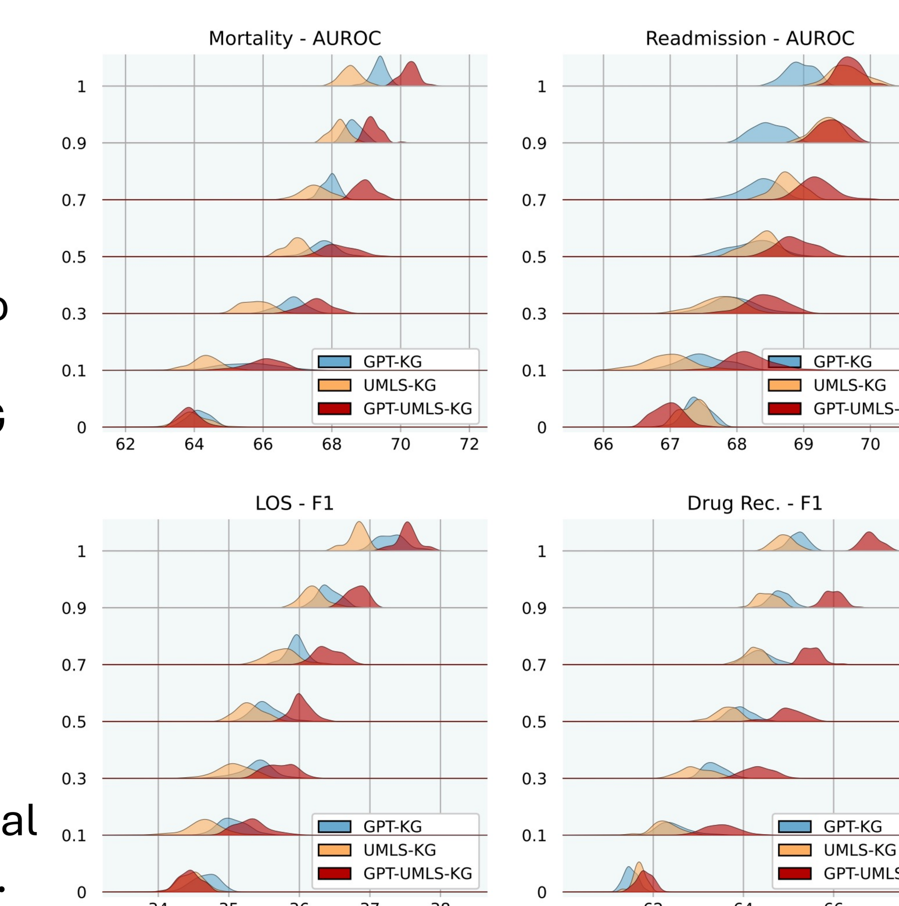
Performance vs. EHR Training Data



1. GraphCare exhibits a considerable edge over other models when confronted with scarce EHR training data.
2. Other graph-augmented methods (e.g. GRAM, GAMENet) also show a certain level of resilience against scarce data.

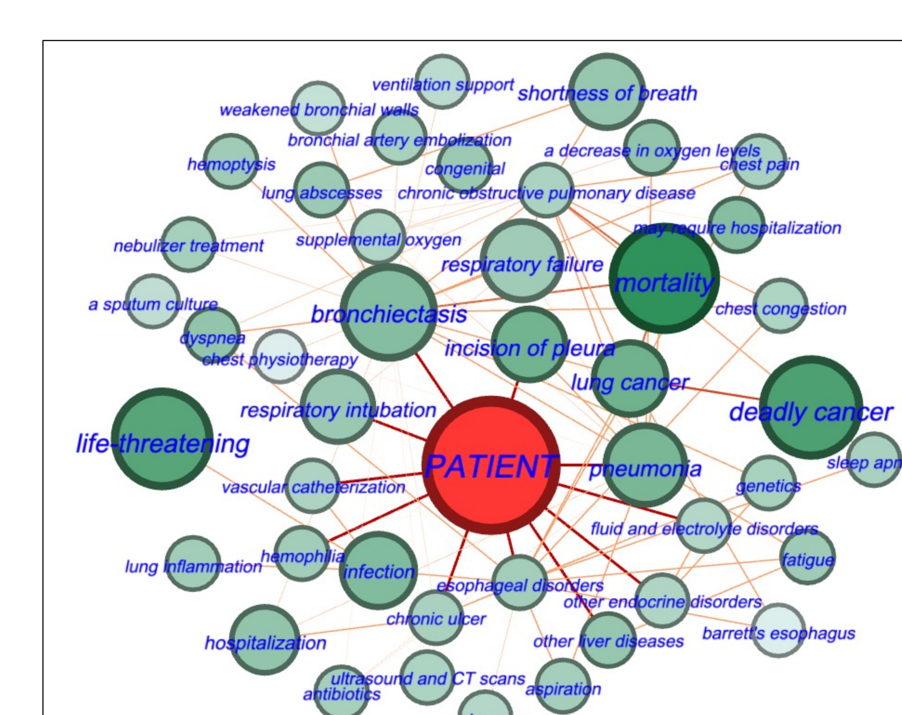
Performance vs. Knowledge Graph Size

1. Combined GPT-UMLS-KG outperforms both GPT-KG and UMLS-KG consistently.
2. GPT-KG contributes more to mortality and LOS predictions, while UMLS-KG edges out in readmission prediction.
3. Lower KG ratios are associated with larger standard deviations, due to the reduced likelihood of vital knowledge being contained.



Visualization of Personalized KG

Removing the indirect node "lung cancer" connecting crucial nodes "mortality" and "deadly cancer" results in a failure of mortality prediction.



Emphasizes the value of comprehensive health data and considering all potential health factors

Personalized KG is important!