Multi-Ontology Integration with Dual-Axis Propagation for Medical Concept Representation

Mohsen Nayebi Kerdabai, Arya Hadizadeh Moghaddam, Dongie Wang, Zijun Yao*

KU THE UNIVERSITY OF KANSAS

CIKM SEOUL 2025
November 10 - 14

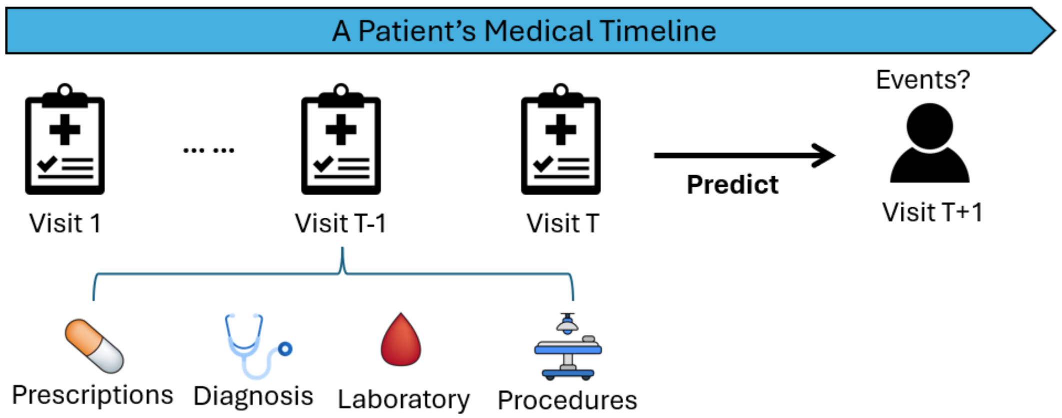
Outline

- **Background and Motivation**
- **Proposed Method**
- **Evaluation**
- **Conclusion**

Background and Motivation

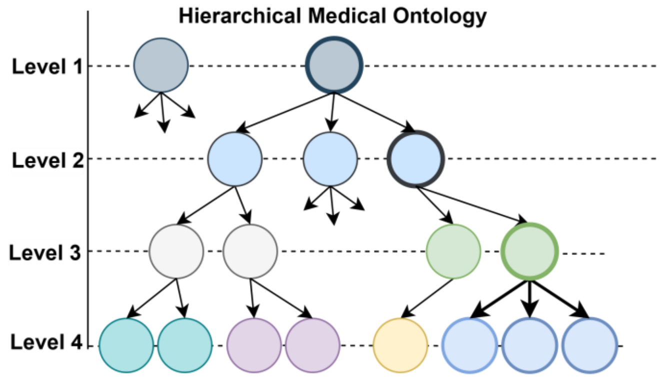
Electronic Health Records (EHRs)

- Contain vast patient history as sequences of medical visits, each containing diagnosis, procedure, and drug codes.
- Rich, context-aware code representations are critical for robust clinical predictions.



Two Source of External information

- **Medical Ontology**
  - Medical ontologies (e.g., ICD, ATC) organize diagnoses, drugs, and procedures into hierarchical codes.
  - They encode domain relationships, but are often underutilized in predictive models.
- **Large Language Models**
  - pre-trained on extensive corpora such as biomedical literature, provide a rich source of domain-specific knowledge



Background and Motivation

Challenges:

- EHR codes are sparse, diverse, and complex, especially for rare diseases.
- Ontology graphs capture domain knowledge, but most models use only single or unconnected systems, missing cross-ontology links.
- LLMs have potential as knowledge bases, but can produce factually inconsistent or misleading content in free-text.
- Result: Limited representations and poor predictions for rare or complex cases.
- Need: unified, multi-ontology representation learning

Objective

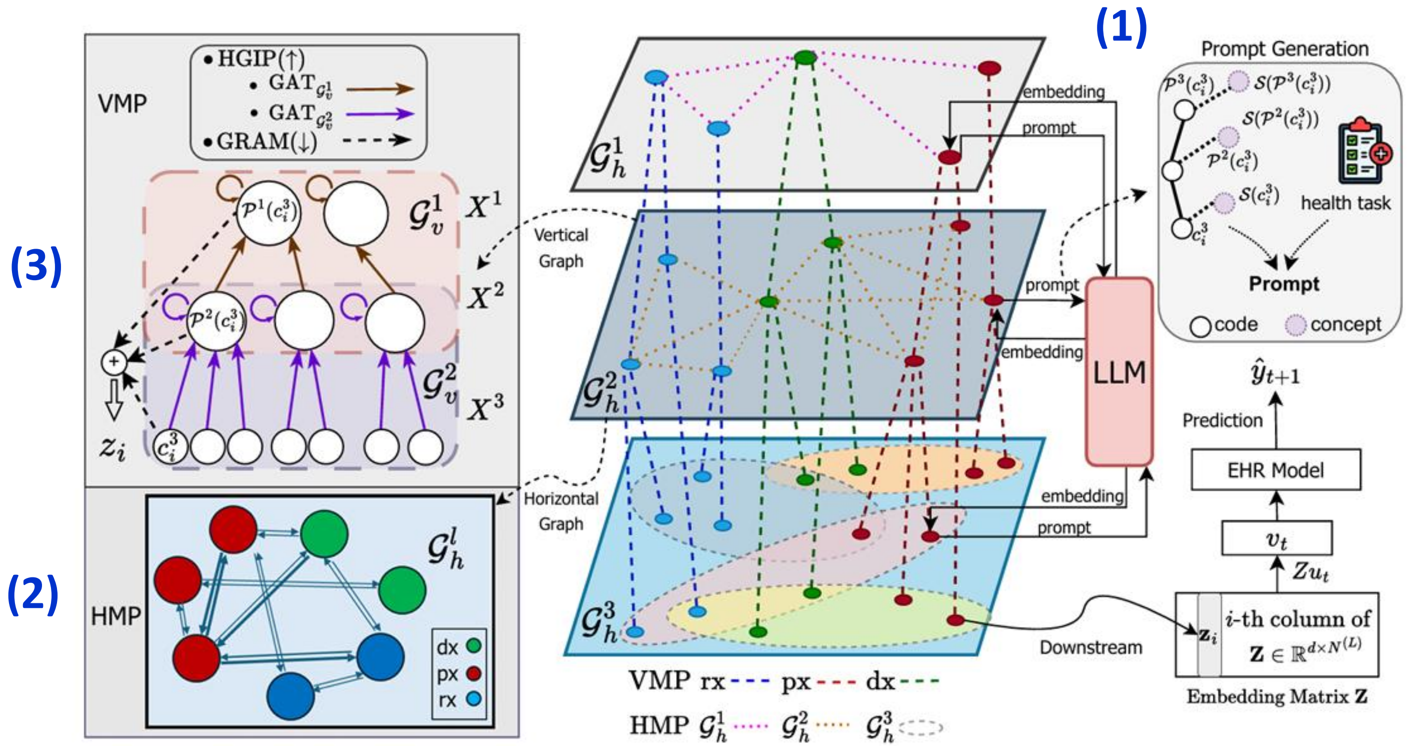
- Enhance medical concept representation learning by integrating multiple ontologies (disease, drug, procedure) in a unified learning structure.
- Leverage LLMs for expert knowledge while ensuring reliability in healthcare predictions.

Outline

- Background and Motivation
- **Proposed Method**
- Evaluation
- Conclusion

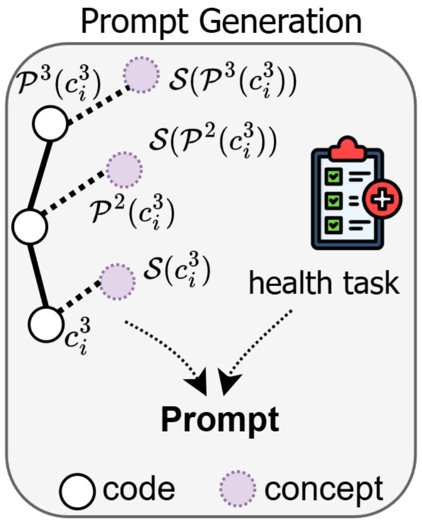
Proposed Method

- We present **LINKO: LLM-augmented INtegrative Knowledge Propagation over Ontology Graphs,** a medical concept encoder with three key steps:
  1. **Meta-KG construction** - integrates multiple ontologies, initializes embeddings using LLM
  2. **Horizontal Message Passing (HMP)** – between codes across and within ontologies
  3. **Vertical Message Passing (VMP)** – between hierarchical ontology levels within each ontology

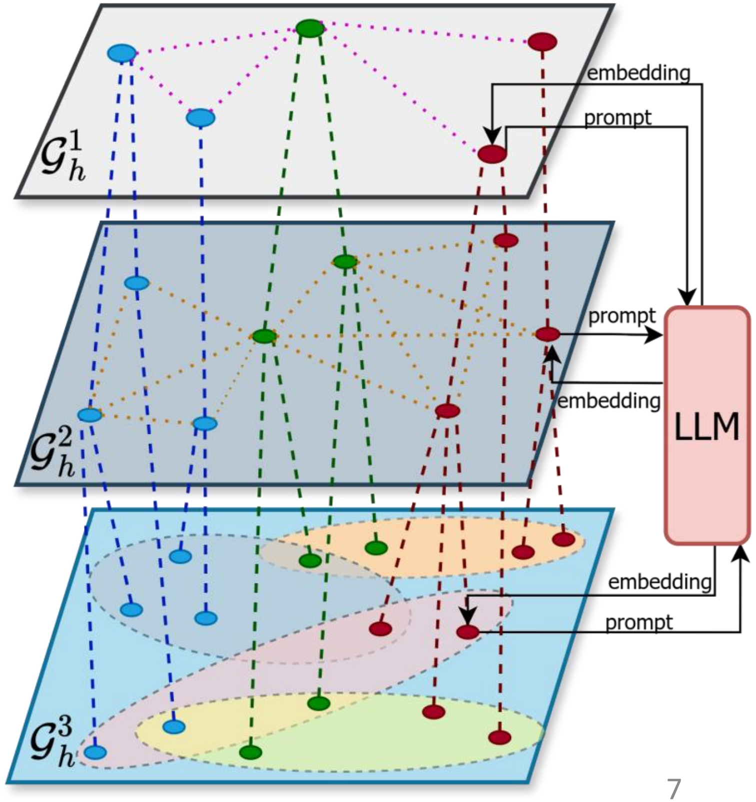


Proposed Method - Meta-KG construction

- Formulating the Meta-KG, a multifaceted knowledge graph (KG) that joins multiple ontology graphs through every level of hierarchy.
- The initialization of node embeddings in Meta-KG is enabled through a curated LLM prompting strategy and dense embedding retrieval.

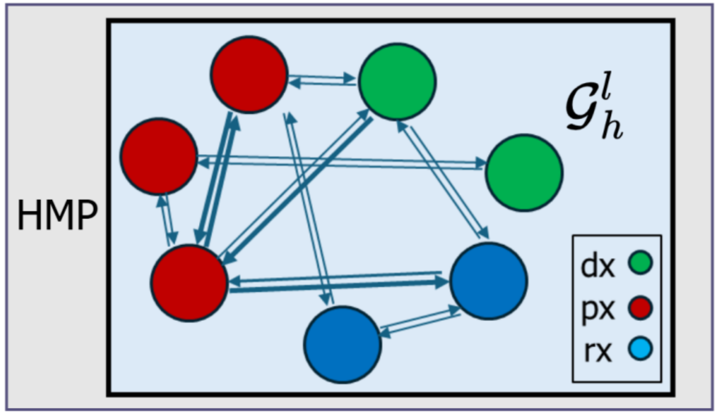


ICD-9 Diagnosis 250.7:
Prompt: << For the task of diagnosis prediction, provide a semantic representation for ICD-9 diagnosis code 250.7, which represents Diabetes with peripheral circulatory disorders. It is a specific medical concept under the broader ICD-9 Diagnosis categories of 250 (Diabetes mellitus), 249-259 (Diseases of Other Endocrine Glands), and 240-279 (Endocrine, Nutritional, and Metabolic Diseases, and Immunity Disorders).>>

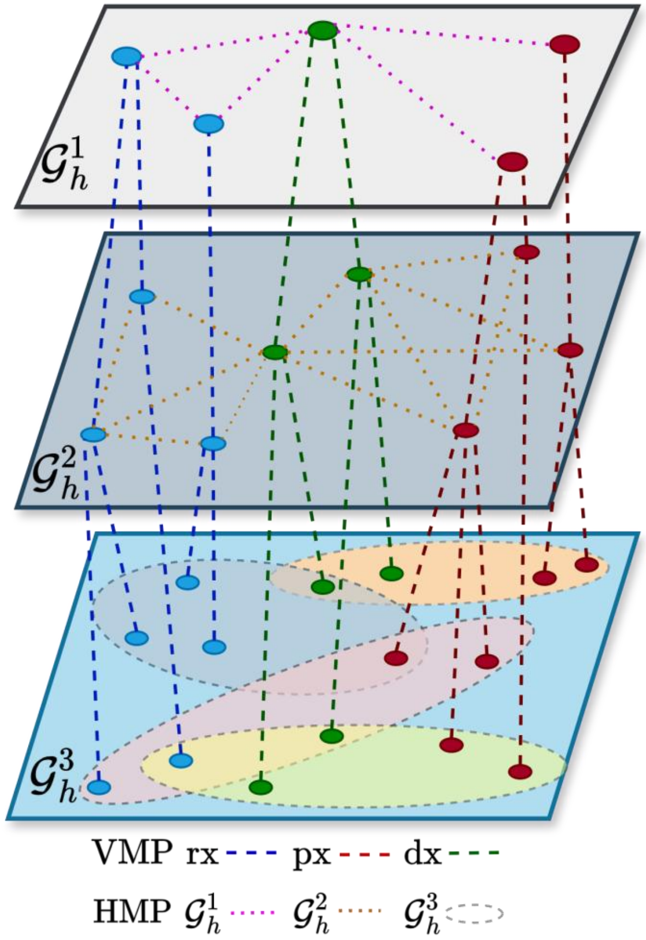


Proposed Method - Horizontal Message Passing (HMP)

- **Core Idea:**
  - Connects heterogeneous concepts, e.g., *disease* ↔ *drug* ↔ *procedure*
  - Enables **horizontal propagation** of knowledge at each level
  - Bridges ontology graphs using **EHR-driven co-occurrence edges**



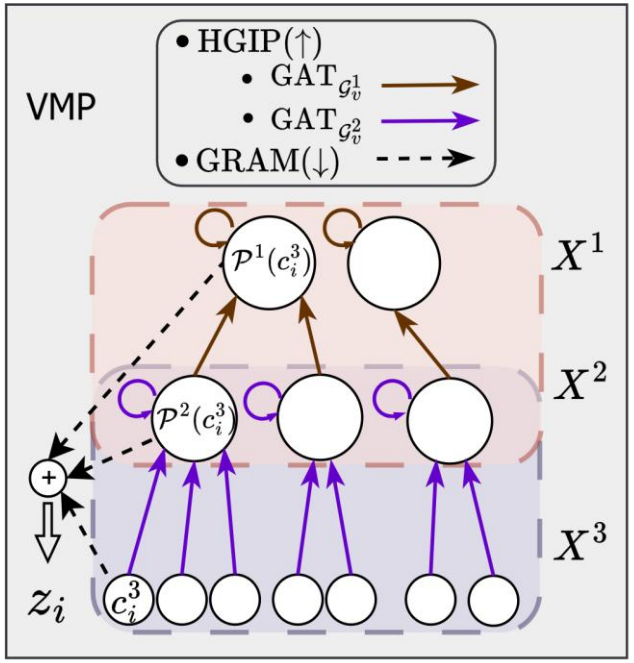
- **How It Works:**
  - Build **co-occurrence conditional probabilities** from EHR visits to define directed edges
  - Filter edges by threshold $\tau^{(l)}$ to control connectivity
  - Encode graphs using:
    - **GAT** (Graph Attention Network) → for parent-level graphs
    - **HAT** (Hypergraph Attention Network) → for leaf-level hypergraph



Proposed Method - Vertical Message Passing (VMP)

- **Core Idea:**
  - Goal: Propagate information **within each ontology** — from general to specific concepts and back.
  - Two rounds of message passing:
    - **Bottom-Up Propagation (HGIP)** — children → parents
    - **Top-Down Propagation (GRAM)** — parents → children
  - Enables multi-level information sharing across ontology hierarchy

- **How It Works:**
  - Construct sequential **vertical subgraphs** between adjacent levels
  - Apply **Graph Attention (GAT)** to aggregate information between levels
  - Compute final leaf-level embeddings via **attention-weighted combination**



Proposed Method – Plug-In into Downstream Task

Integrating LINKO into EHR Models

- LINKO can be used as a plug-in medical concept encoder for any EHR prediction framework.
- Outputs dynamic, ontology-informed embeddings tailored for clinical tasks.
- No changes needed to the core architecture of existing EHR models.

Embedding for code i

Raw input feature for medical concept i at level L

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N^{(L)}}] \leftarrow \mathrm{LINKO}(\mathbf{x}_1^{(L)}, \mathbf{x}_2^{(L)}, \ldots, \mathbf{x}_{N^{(L)}}^{(L)})$$

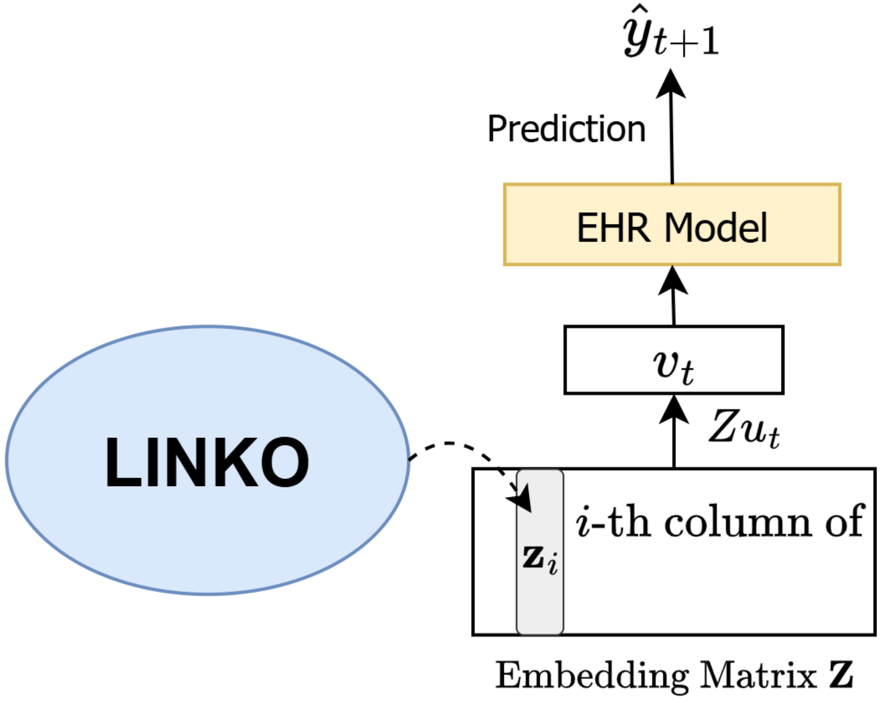
$$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t = \mathbf{Z}[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_t]$$

$$\mathbf{h}_p = \mathrm{Model}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t)$$

$$\hat{\mathbf{y}}_{t+1} = \mathrm{Sigmoid}(\mathbf{W}\mathbf{h}_p + \mathbf{b})$$

Visit representation for time t

Patient representation



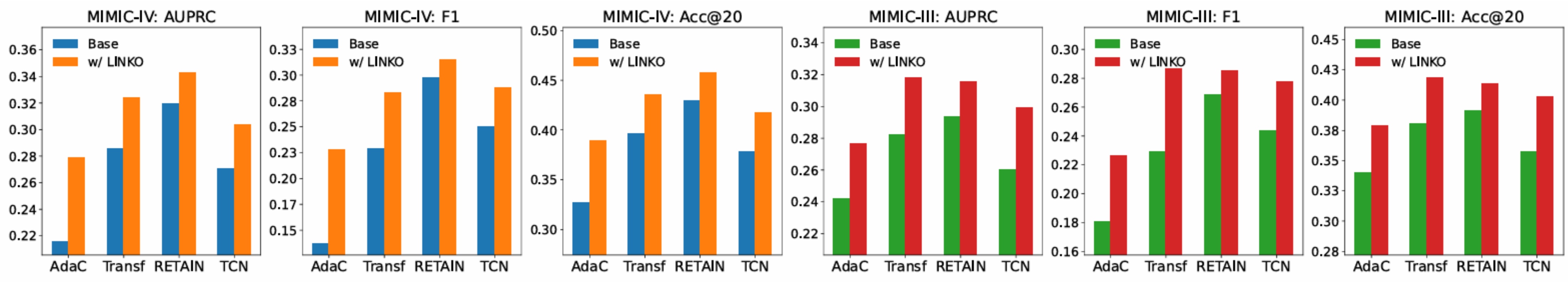
Outline

- Background and Motivation
- Proposed Method
- **Evaluation**
- Conclusion

Evaluation - Experimental Setup

- **EHRs Dataset**
  - MIMIC-III, MIMIC-IV
- **Medical Ontologies:**
  - Drug codes: Anatomical Therapeutic Chemical (ATC) classification
  - Diagnosis/Procedure : The International Classification of Diseases (ICD-9)
- **Task**
  - Sequential Diagnosis Prediction: Given a patient's visit sequence $X_t = \{V_1, V_2, \dots, V_t\}$, the objective is to predict diagnosis codes for the next visit V_{t+1}
- **Evaluation Metrics**
  - AUPRC : The area under the precision-recall curve
  - Acc@k: The number of correct diagnosis codes among the top k predictions.
  - F1-score: The harmonic mean of precision and recall
- **Scope**
  - In this paper, we have tackled 6 Research Questions (RQ)

Evaluation - RQ1: Plug-in Enhancement Evaluation



RQ1: Does our method enhance the performance of EHR models as a complementary concept encoder?

- Integrated **LINKO** as a plug-in encoder into multiple EHR models
- Consistently improved downstream prediction performance
- Demonstrated strong **model-agnostic generalization** across architectures
- Validated LINKO's effectiveness in **enhancing medical concept representations**

Evaluation - RQ2: Baseline Comparison

RQ2: How does LINKO perform compared to existing medical code encoders?

- Base model is Transformer, while subsequent rows represent Base with added medical concept encoders (e.g., GRAM = Base + GRAM)
- Compared to leading ontology-based encoders LINKO consistently achieved the **outperforms** across datasets
- Grouping Diagnosis labels into four frequency bands, label category performance showes **robustness for rare and low-frequency medical codes**

	Model	General Performance					Label Category Performance (AUPRC)			
		AUPRC	F1	Acc@15	Acc@20	Acc@30	0-25%	25-50%	50-75%	75-100%
MIMIC-IV	Base	28.54$_{\pm0.39}$	22.90$_{\pm0.10}$	37.88$_{\pm0.53}$	39.53$_{\pm0.60}$	44.10$_{\pm0.52}$	28.33$_{\pm0.44}$	58.73$_{\pm0.86}$	46.54$_{\pm0.30}$	67.55$_{\pm0.18}$
	GRAM	29.96$_{\pm0.45}$	24.17$_{\pm0.18}$	39.42$_{\pm0.25}$	41.23$_{\pm0.48}$	45.90$_{\pm0.61}$	29.73$_{\pm0.53}$	60.62$_{\pm0.27}$	47.67$_{\pm0.27}$	68.45$_{\pm0.21}$
	MMORE	30.06$_{\pm0.25}$	25.11$_{\pm1.60}$	39.56$_{\pm0.18}$	41.46$_{\pm0.32}$	46.19$_{\pm0.34}$	29.84$_{\pm0.22}$	60.60$_{\pm0.37}$	48.84$_{\pm0.34}$	68.07$_{\pm1.3}$
	KAME	29.13$_{\pm0.32}$	23.39$_{\pm0.32}$	39.18$_{\pm0.32}$	40.28$_{\pm0.32}$	45.67$_{\pm0.25}$	28.84$_{\pm0.32}$	59.62$_{\pm0.32}$	47.21$_{\pm0.42}$	68.08$_{\pm0.32}$
	G-BERT	30.13$_{\pm0.36}$	25.19$_{\pm0.33}$	39.65$_{\pm0.73}$	41.66$_{\pm0.81}$	46.29$_{\pm0.95}$	29.91$_{\pm0.59}$	60.83$_{\pm0.69}$	49.88$_{\pm0.74}$	69.03$_{\pm1.0}$
	HAP	30.02$_{\pm0.23}$	24.22$_{\pm1.30}$	39.55$_{\pm0.23}$	41.37$_{\pm0.37}$	46.22$_{\pm0.33}$	29.64$_{\pm0.34}$	60.76$_{\pm0.35}$	49.53$_{\pm0.33}$	68.94$_{\pm1.4}$
	ADORE	30.11$_{\pm0.34}$	25.21$_{\pm0.87}$	39.87$_{\pm0.24}$	41.51$_{\pm0.42}$	46.25$_{\pm0.34}$	30.15$_{\pm0.30}$	60.89$_{\pm0.45}$	49.44$_{\pm1.02}$	68.18$_{\pm0.93}$
	KAMPNet	30.21$_{\pm0.23}$	25.32$_{\pm1.10}$	40.18$_{\pm0.42}$	41.88$_{\pm0.54}$	46.41$_{\pm0.56}$	30.21$_{\pm0.34}$	61.05$_{\pm0.89}$	48.93$_{\pm0.53}$	70.30$_{\pm1.0}$
	LINKO$_{w/\ GAT}$	32.13$_{\pm0.12}$	28.56$_{\pm0.09}$	42.60$_{\pm0.14}$	43.56$_{\pm0.14}$	48.12$_{\pm0.18}$	31.83$_{\pm0.13}$	63.03$_{\pm0.26}$	50.96$_{\pm0.40}$	71.35$_{\pm0.29}$
	LINKO$_{w/\ HAT}$	32.38$_{\pm0.32}$	30.02$_{\pm0.13}$	42.05$_{\pm0.65}$	43.70$_{\pm0.66}$	48.63$_{\pm0.73}$	32.31$_{\pm0.41}$	63.25$_{\pm0.73}$	51.17$_{\pm0.23}$	71.58$_{\pm0.12}$
MIMIC-III	Base	28.23$_{\pm0.24}$	22.36$_{\pm0.33}$	36.15$_{\pm0.22}$	38.03$_{\pm0.34}$	43.19$_{\pm0.43}$	28.07$_{\pm0.28}$	54.28$_{\pm0.10}$	49.89$_{\pm0.29}$	73.34$_{\pm0.26}$
	GRAM	28.99$_{\pm0.34}$	24.10$_{\pm0.48}$	37.17$_{\pm0.36}$	39.13$_{\pm0.52}$	44.32$_{\pm0.51}$	28.79$_{\pm0.41}$	54.81$_{\pm1.3}$	50.58$_{\pm0.34}$	74.44$_{\pm0.77}$
	MMORE	29.11$_{\pm0.38}$	23.67$_{\pm0.70}$	37.15$_{\pm0.46}$	39.14$_{\pm0.53}$	44.36$_{\pm0.50}$	28.87$_{\pm0.43}$	54.92$_{\pm0.87}$	51.29$_{\pm0.25}$	74.42$_{\pm0.30}$
	KAME	28.52$_{\pm0.32}$	23.18$_{\pm0.09}$	36.76$_{\pm0.47}$	38.36$_{\pm0.45}$	43.75$_{\pm0.36}$	28.19$_{\pm0.38}$	55.13$_{\pm0.16}$	50.09$_{\pm0.32}$	73.79$_{\pm0.23}$
	G-BERT	29.22$_{\pm0.81}$	24.92$_{\pm0.83}$	37.36$_{\pm0.12}$	39.55$_{\pm0.12}$	45.15$_{\pm0.44}$	29.16$_{\pm0.10}$	54.47$_{\pm0.99}$	52.45$_{\pm0.62}$	75.40$_{\pm0.18}$
	HAP	29.28$_{\pm0.41}$	23.25$_{\pm0.88}$	37.47$_{\pm0.56}$	39.64$_{\pm0.55}$	45.05$_{\pm0.57}$	29.10$_{\pm0.45}$	55.11$_{\pm0.99}$	52.19$_{\pm0.27}$	75.50$_{\pm0.24}$
	ADORE	29.31$_{\pm0.29}$	23.61$_{\pm0.81}$	37.53$_{\pm0.31}$	39.68$_{\pm0.32}$	45.01$_{\pm0.58}$	29.18$_{\pm0.73}$	55.14$_{\pm0.42}$	52.14$_{\pm0.45}$	75.41$_{\pm0.27}$
	KAMPNet	29.38$_{\pm0.65}$	25.01$_{\pm0.72}$	37.83$_{\pm0.39}$	39.91$_{\pm0.61}$	45.26$_{\pm0.53}$	29.31$_{\pm0.45}$	55.20$_{\pm0.87}$	52.56$_{\pm0.34}$	75.91$_{\pm0.29}$
	LINKO$_{w/\ HAT}$	30.78$_{\pm0.99}$	27.67$_{\pm0.23}$	39.11$_{\pm0.92}$	41.18$_{\pm0.73}$	46.65$_{\pm0.83}$	30.72$_{\pm0.85}$	54.74$_{\pm0.19}$	57.16$_{\pm0.26}$	79.23$_{\pm0.10}$
	LINKO$_{w/\ GAT}$	31.79$_{\pm0.11}$	28.66$_{\pm0.18}$	39.95$_{\pm0.11}$	41.84$_{\pm0.14}$	46.96$_{\pm0.96}$	31.62$_{\pm1.09}$	55.68$_{\pm0.15}$	56.90$_{\pm0.59}$	80.76$_{\pm0.47}$

Evaluation - RQ3: Ablation Study

Dataset	Model	General Performance					Label Category Performance (AUPRC)			
		AUPRC	F1	Acc@15	Acc@20	Acc@30	0-25%	25-50%	50-75%	75-100%
MIMIC-IV	LINKO$_{w/\ GAT}$	$32.13_{\pm0.12}$	$28.56_{\pm0.09}$	$42.60_{\pm0.14}$	$43.56_{\pm0.14}$	$48.12_{\pm0.18}$	$31.83_{\pm0.13}$	$63.03_{\pm0.26}$	$50.96_{\pm0.40}$	$71.35_{\pm0.29}$
	LINKO$_{w/\ HAT}$	$32.38_{\pm0.32}$	$30.02_{\pm0.13}$	$42.05_{\pm0.65}$	$43.70_{\pm0.66}$	$48.63_{\pm0.73}$	$32.31_{\pm0.41}$	$63.25_{\pm0.73}$	$51.17_{\pm0.23}$	$71.58_{\pm0.12}$
	w/o HGIP	$30.65_{\pm0.28}$	$26.19_{\pm0.22}$	$40.16_{\pm0.20}$	$42.03_{\pm0.21}$	$46.81_{\pm0.40}$	$30.36_{\pm0.45}$	$61.34_{\pm0.13}$	$49.15_{\pm0.19}$	$68.63_{\pm0.55}$
	w/o LLM	$30.86_{\pm0.59}$	$26.69_{\pm0.52}$	$40.40_{\pm0.83}$	$42.06_{\pm0.75}$	$46.54_{\pm0.88}$	$30.66_{\pm0.68}$	$61.42_{\pm0.10}$	$48.35_{\pm0.18}$	$67.10_{\pm0.32}$
	w/o leaf-HMP	$30.70_{\pm0.33}$	$26.89_{\pm0.88}$	$40.27_{\pm0.43}$	$42.09_{\pm0.38}$	$46.76_{\pm0.60}$	$30.41_{\pm0.45}$	$61.30_{\pm0.99}$	$48.77_{\pm0.79}$	$69.46_{\pm0.99}$
	w/o parent-HMP	$30.32_{\pm0.69}$	$25.49_{\pm0.25}$	$39.83_{\pm0.59}$	$41.72_{\pm0.92}$	$46.47_{\pm0.90}$	$30.07_{\pm0.78}$	$61.40_{\pm0.72}$	$46.46_{\pm0.23}$	$68.06_{\pm0.29}$
	w/o HMP	$30.10_{\pm0.39}$	$25.26_{\pm0.88}$	$39.67_{\pm0.41}$	$41.50_{\pm0.34}$	$46.25_{\pm0.53}$	$29.85_{\pm0.51}$	$60.94_{\pm0.75}$	$47.39_{\pm0.28}$	$67.33_{\pm0.62}$
MIMIC-III	LINKO$_{w/\ HAT}$	$30.78_{\pm0.99}$	$27.67_{\pm0.23}$	$39.11_{\pm0.92}$	$41.18_{\pm0.73}$	$46.65_{\pm0.83}$	$30.72_{\pm0.85}$	$54.74_{\pm0.19}$	$57.16_{\pm0.26}$	$79.23_{\pm0.10}$
	LINKO$_{w/\ GAT}$	$31.79_{\pm0.11}$	$28.66_{\pm0.18}$	$39.95_{\pm0.11}$	$41.84_{\pm0.14}$	$46.96_{\pm0.96}$	$31.62_{\pm1.09}$	$55.68_{\pm0.15}$	$56.90_{\pm0.59}$	$80.76_{\pm0.47}$
	w/o HGIP	$30.25_{\pm0.37}$	$25.75_{\pm0.14}$	$38.54_{\pm0.51}$	$40.47_{\pm0.55}$	$46.00_{\pm0.66}$	$30.04_{\pm0.44}$	$55.61_{\pm0.12}$	$53.36_{\pm0.31}$	$77.00_{\pm0.22}$
	w/o LLM	$30.38_{\pm0.10}$	$25.88_{\pm0.90}$	$38.23_{\pm0.57}$	$40.38_{\pm0.66}$	$45.82_{\pm0.90}$	$30.33_{\pm0.10}$	$55.05_{\pm0.99}$	$55.03_{\pm0.64}$	$79.16_{\pm0.33}$
	w/o leaf-HMP	$30.09_{\pm0.43}$	$25.44_{\pm0.15}$	$38.25_{\pm0.54}$	$40.35_{\pm0.54}$	$45.92_{\pm0.59}$	$29.88_{\pm0.51}$	$55.57_{\pm0.57}$	$53.62_{\pm0.30}$	$76.68_{\pm0.17}$
	w/o parent-HMP	$29.76_{\pm0.43}$	$26.02_{\pm0.10}$	$37.97_{\pm0.39}$	$39.93_{\pm0.56}$	$45.43_{\pm0.52}$	$29.53_{\pm0.48}$	$55.24_{\pm0.88}$	$52.48_{\pm0.26}$	$75.78_{\pm0.17}$
	w/o HMP	$29.41_{\pm0.40}$	$25.28_{\pm0.88}$	$37.55_{\pm0.41}$	$39.61_{\pm0.52}$	$45.25_{\pm0.60}$	$29.26_{\pm0.50}$	$54.52_{\pm0.63}$	$52.58_{\pm0.20}$	$75.46_{\pm0.21}$

RQ3: What is the impact of each component of LINKO on performance?

Removing any key component of LINKO leads to **performance degradation**

- **w/o leaf-HMP:** removing ontology fusion at the leaf level weakens fine-grained concept interaction
- **w/o parent-HMP:** greater drop — parent-level fusion is crucial for coarse-grained alignment
- **w/o HMP:** removing all horizontal propagation severely hurts performance; cross-ontology links are vital
- **w/o HGIP:** loss of hierarchical information flow reduces structured representation quality
- **w/o LLM:** random initialization degrades semantic grounding and performance

Evaluation - RQ3: Ablation Study

	Concept Type	w/ Multi-level Integration					w/o Multi-level Integration				
		AUPRC	F1	Acc@15	Acc@20	Acc@30	AUPRC	F1	Acc@15	Acc@20	Acc@30
MIMIC-IV	rx,px	22.74$_{\pm0.04}$	16.92$_{\pm0.17}$	31.56$_{\pm0.20}$	33.57$_{\pm0.27}$	38.75$_{\pm0.26}$	21.35$_{\pm0.27}$	15.15$_{\pm0.26}$	30.03$_{\pm0.53}$	32.21$_{\pm0.65}$	37.23$_{\pm0.39}$
	dx,px	30.67$_{\pm0.44}$	25.87$_{\pm0.13}$	40.25$_{\pm0.62}$	42.21$_{\pm0.59}$	46.99$_{\pm0.42}$	29.28$_{\pm0.31}$	22.49$_{\pm0.15}$	38.62$_{\pm0.31}$	40.68$_{\pm0.37}$	45.56$_{\pm0.24}$
	dx,rx	30.86$_{\pm0.10}$	25.34$_{\pm0.23}$	40.44$_{\pm0.10}$	42.53$_{\pm0.12}$	47.31$_{\pm0.10}$	30.17$_{\pm0.10}$	25.08$_{\pm0.14}$	39.77$_{\pm0.57}$	41.71$_{\pm0.70}$	46.59$_{\pm0.64}$
	dx,rx,px	32.38$_{\pm0.32}$	30.02$_{\pm0.13}$	42.05$_{\pm0.65}$	43.70$_{\pm0.66}$	48.63$_{\pm0.73}$	30.10$_{\pm0.39}$	25.26$_{\pm0.23}$	39.67$_{\pm0.41}$	41.50$_{\pm0.34}$	46.25$_{\pm0.53}$
MIMIC-III	rx,px	24.24$_{\pm0.97}$	18.29$_{\pm0.97}$	31.92$_{\pm0.86}$	34.43$_{\pm0.14}$	39.87$_{\pm0.13}$	22.98$_{\pm0.22}$	16.55$_{\pm0.14}$	30.81$_{\pm0.31}$	32.95$_{\pm0.49}$	38.61$_{\pm0.33}$
	dx,px	30.19$_{\pm0.11}$	26.40$_{\pm0.26}$	38.03$_{\pm0.81}$	40.18$_{\pm0.79}$	46.03$_{\pm0.11}$	29.01$_{\pm0.67}$	24.48$_{\pm0.10}$	37.04$_{\pm0.74}$	39.16$_{\pm0.61}$	44.89$_{\pm0.68}$
	dx,rx	30.64$_{\pm0.40}$	27.06$_{\pm0.15}$	39.31$_{\pm0.58}$	41.27$_{\pm0.64}$	46.62$_{\pm0.57}$	29.56$_{\pm0.38}$	24.44$_{\pm0.65}$	37.88$_{\pm0.61}$	39.98$_{\pm0.65}$	45.46$_{\pm0.60}$
	dx,rx,px	31.79$_{\pm0.11}$	28.66$_{\pm0.18}$	39.95$_{\pm0.11}$	41.84$_{\pm0.14}$	46.96$_{\pm0.96}$	29.41$_{\pm0.40}$	25.28$_{\pm0.88}$	37.55$_{\pm0.41}$	39.61$_{\pm0.52}$	45.25$_{\pm0.60}$

RQ3: What is the impact of each component on performance? Closer look at HMP

Evaluated LINKO with different ontology combinationsof (**dx**, **rx**, **px) and** Repeating experiments with and without **Horizontal Message Passing (HMP)**

- **Without HMP:** ontologies learned independently → weaker and less consistent performance
- **With HMP:** cross-ontology integration captures both heterogeneous and homogeneous code relations
- Dropping any ontology reduces performance; **dx** has the largest impact, **px** the smallest
- **HMP enables synergy**, even sparse ontologies (like px) contribute meaningfully through integration

Evaluation - RQ4: Prompt Analysis

Prompt Information	MIMIC-IV					MIMIC-III				
	AUPRC	F1	Acc@15	Acc@20	Acc@30	AUPRC	F1	Acc@15	Acc@20	Acc@30
no LLM	$30.86_{\pm0.59}$	$26.69_{\pm0.52}$	$40.40_{\pm0.83}$	$42.06_{\pm0.75}$	$46.54_{\pm0.88}$	$30.38_{\pm0.10}$	$25.88_{\pm0.90}$	$38.23_{\pm0.57}$	$40.38_{\pm0.66}$	$45.82_{\pm0.90}$
type-code-noise	$31.28_{\pm0.10}$	$28.46_{\pm0.30}$	$40.99_{\pm0.75}$	$43.25_{\pm0.10}$	$47.39_{\pm0.11}$	$30.57_{\pm0.74}$	$27.88_{\pm0.21}$	$38.99_{\pm0.46}$	$40.96_{\pm0.57}$	$46.57_{\pm0.63}$
type-code	$31.96_{\pm0.40}$	$27.77_{\pm0.67}$	$41.57_{\pm0.35}$	$43.24_{\pm0.42}$	$47.92_{\pm0.50}$	$31.15_{\pm0.98}$	$27.89_{\pm0.10}$	$39.28_{\pm0.69}$	$41.26_{\pm0.44}$	$46.79_{\pm0.53}$
type-code-concept	$32.00_{\pm0.25}$	$29.07_{\pm0.56}$	$41.73_{\pm0.74}$	$43.52_{\pm0.83}$	$48.17_{\pm0.78}$	$31.18_{\pm1.00}$	$28.35_{\pm0.65}$	$39.39_{\pm0.91}$	$41.55_{\pm0.83}$	$46.90_{\pm0.94}$
type-code-concept-children	$32.05_{\pm1.00}$	$29.03_{\pm1.02}$	$41.57_{\pm0.11}$	$43.41_{\pm0.13}$	$47.94_{\pm0.14}$	$31.09_{\pm1.00}$	$27.67_{\pm1.02}$	$39.23_{\pm0.13}$	$41.18_{\pm0.10}$	$46.66_{\pm0.99}$
type-code-concept-parent	$32.35_{\pm0.50}$	$28.27_{\pm0.48}$	$42.01_{\pm0.73}$	$43.59_{\pm0.38}$	$48.49_{\pm0.41}$	$31.21_{\pm0.95}$	$28.48_{\pm0.29}$	$39.46_{\pm0.79}$	$41.64_{\pm0.10}$	$46.91_{\pm0.97}$
type-code-concept-parent-task	$\mathbf{32.38}_{\pm0.32}$	$\mathbf{30.02}_{\pm0.13}$	$\mathbf{42.05}_{\pm0.65}$	$\mathbf{43.70}_{\pm0.66}$	$\mathbf{48.63}_{\pm0.73}$	$\mathbf{31.79}_{\pm0.11}$	$\mathbf{28.66}_{\pm0.18}$	$\mathbf{39.95}_{\pm0.11}$	$\mathbf{41.84}_{\pm0.14}$	$\mathbf{46.96}_{\pm0.96}$
w/ LLM-emb-freezed	$22.13_{\pm0.77}$	$13.35_{\pm0.89}$	$30.79_{\pm0.90}$	$32.91_{\pm0.92}$	$37.70_{\pm0.97}$	$25.75_{\pm0.37}$	$19.96_{\pm0.36}$	$33.82_{\pm0.39}$	$36.28_{\pm0.38}$	$42.00_{\pm0.38}$

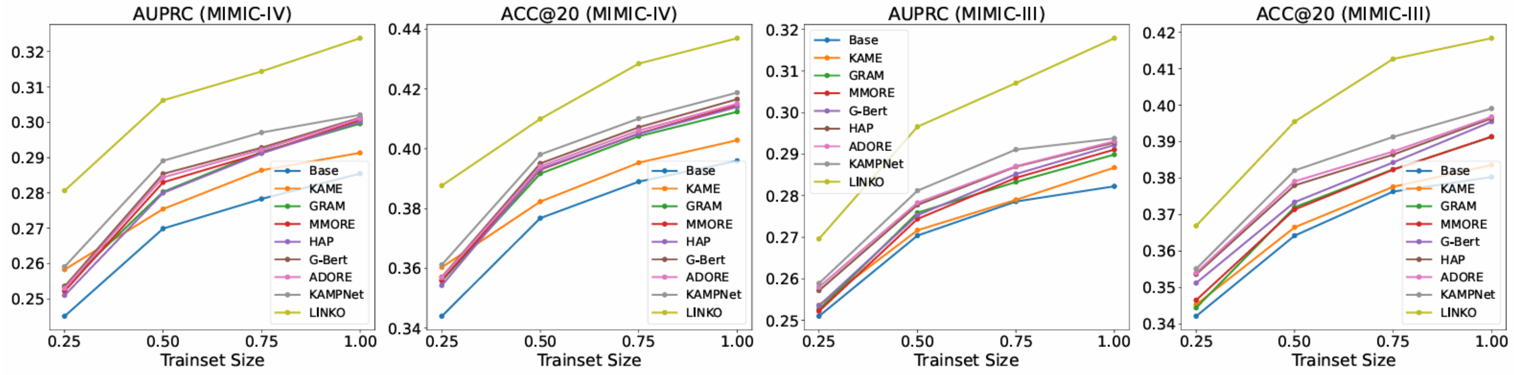
RQ4: How does the prompting strategy affect performance?

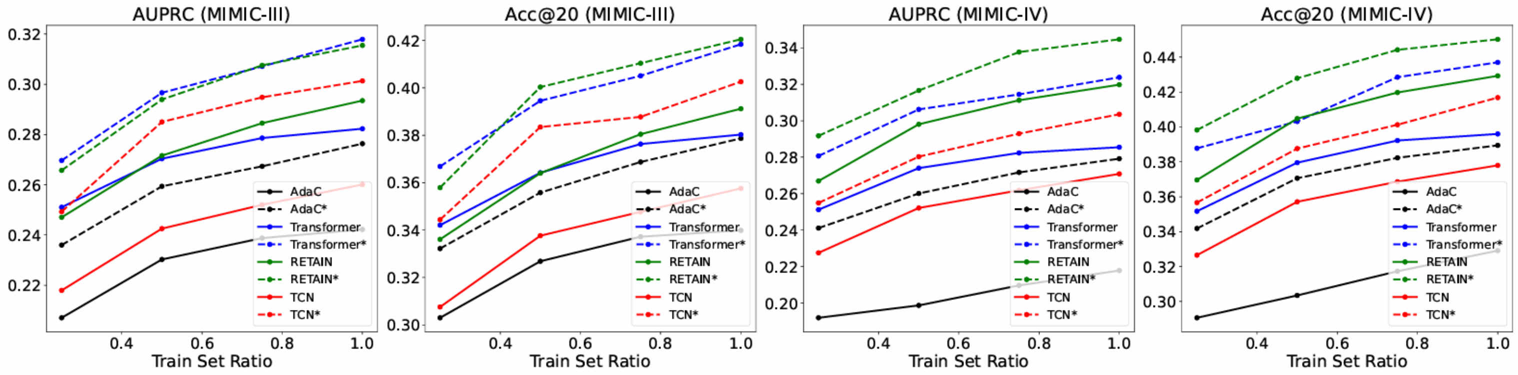
- **Best performance:** prompt includes code ID, name, description, parent info, and task context
- Adding **child-level details** does not help with performance
- Injecting **irrelevant text/noise** degrades embedding quality
- **Freezing LLM embeddings** weakens representation learning => fine-tuning is essential
- **LLM initialization accelerates convergence** (~4× faster than random initialization)

Evaluation - RQ5: Data Insufficiency Analysis

RQ5: How does our method mitigate data insufficiency limitations and rare disease predictions?

- Evaluating LINKO under varying training dataset sizes to simulate limited-data scenarios, it consistently outperforms baselines in data-scarce settings.
- In the Plug-in Enhancement setting, LINKO significantly improves EHR model performance even with reduced training data.



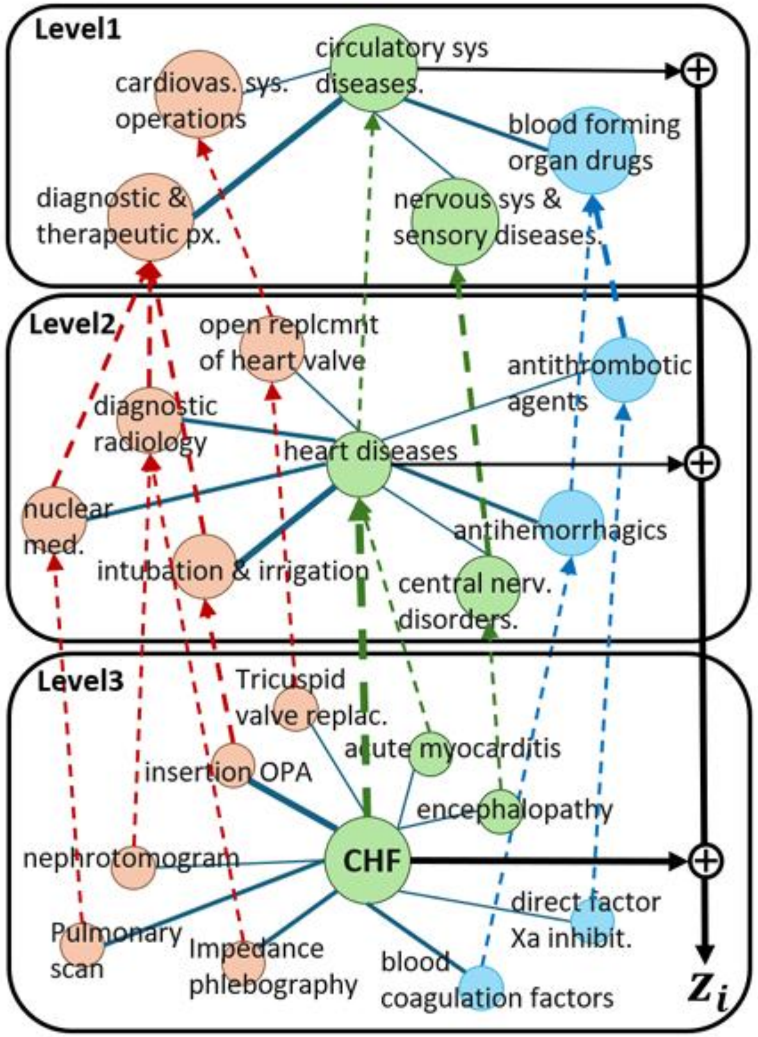


Evaluation – RQ6: Case Study Analysis

RQ6: How does LINKO learn representations for individual codes?

Case Study – Encoding CHF (ICD-9 428.0: Congestive Heart Failure) with LINKO

- Dual-axis propagation: vertical (hierarchical ancestors) & horizontal (co-occurring codes in all ontologies)
- LLM prompts initialize code embeddings with semantic context
- Graph attention learns edge importance; shown by line thickness
- Color-coded edges: procedures (red), diagnoses (green), drugs (blue)
- Final embedding combines own and ancestor features



Outline

- Background and Motivation
- Proposed Method
- Evaluation
- **Conclusion**

Conclusion

We present **LINKO: LLM-augmented INtegrative Knowledge Propagation over Ontology Graphs,** featuring following contributions:

- **Unified Multi-Ontology Framework:** Jointly represents diagnoses, drugs, and procedures in a single meta-graph.
- **Dual-axis Knowledge Propagation:** Combines vertical (hierarchical) and horizontal (cross-ontology) message passing.
- **Graph-Augmented LLM Initialization:** Uses LLM-based, graph-aware embeddings for strong initialization, then refines in-graph.
- **Plug-in, Robust Gains:** Outperforms baselines, boosts robustness for rare diseases, and works as a plug-in encoder.

Chekput our paper!

Thank you

For questions, Please contact with the authors:

mohsen.nayebi@ku.edu
zyao@ku.edu