Joint Workshop of the **5th Extraction and Evaluation of Knowledge Entities from Scientific** Documents (EEKE2024) and the 4th AI + Informetrics (AII2024)

# Biomedical association inference on pandemic knowledge graphs: A comparative study

Mengjia Wu[1], Chao Yu[2], Jian Xu[2], Ying Ding[3] and Yi Zhang[1]

1 Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

2 School of Information Management, Sun Yat-sen University, China

3 School of Information, University of Texas, Austin, USA

Biomedical knowledge graph
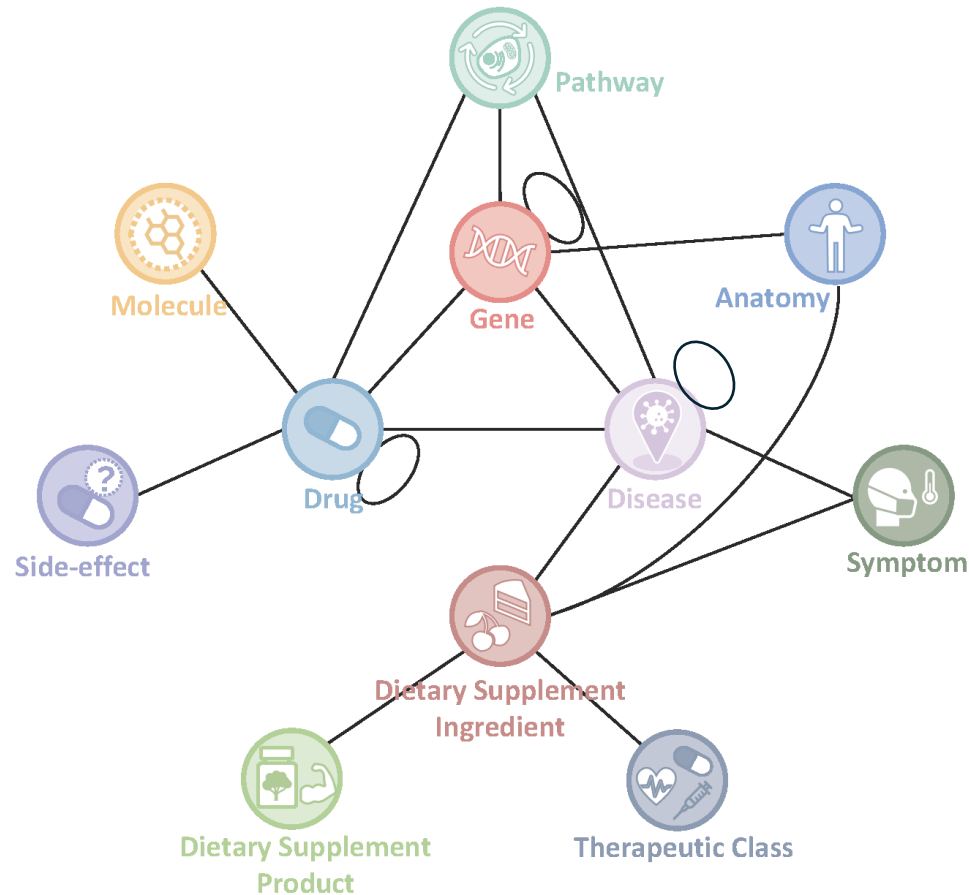
iBKH dataset

Relationships from 17 publicly available biomedical databases.

- Drug-Disease: Treat, palliate, effect, associate, inhibit, alleviate, …

- Drug-gene: target, transporter, enzyme, carrier, downregulates, …

- Disease-disease: is_a, resemble

- Disease-gene: Associate, downregulate, upregulate, …

- Gene-gene: covary, interact, regulate, associate, …

- Drug-drug: interaction, resemble, …

1 Su, C., Hou, Y., Zhou, M., Rajendran, S., Maasch, J. R., Abedi, Z., … & Wang, F. (2023). Biomedical discovery through the integrative biomedical knowledge hub (iBKH). Iscience, 26(4).

PubMed search strategy[1]

SARS: (((("Severe acute respiratory syndrome" OR "SARS") AND (coronavirus*)) OR ("SARS virus" OR "SARS disease" OR "Severe acute respiratory syndrome disease" OR "Severe acute respiratory syndrome virus" OR "SARS-Cov"))) NOT ((("covid" OR "nCov" OR "Covid-19" OR "covid19" OR "SARS-Cov-2" OR "Severe acute respiratory syndrome-2" OR "MERS" OR "middle east respiratory syndrome"))) Filters: from 2001 - 3000/12/12

MERS: (((("Middle east respiratory syndrome" OR "MERS") AND (coronavirus)) OR ("MERS-Cov" OR "MERS virus" OR "MERS disease" OR "Middle east respiratory syndrome virus" OR "Middle east respiratory syndrome disease"))) NOT ((("nCov" OR "Covid-19" OR "covid19" OR "SARS-Cov" OR "SARS-Cov-2" OR "SARS" OR "Severe acute respiratory syndrome")))) Filters: from 2011 - 3000/12/12

Covid-19: ("covid-19" OR "covid19" OR "coronavirus disease 2019" OR "2019-nCov" OR "Novel Coronavirus" OR "Novel Corona virus" OR "SARS-Cov-2") Filters: from 2011 - 3000/12/12
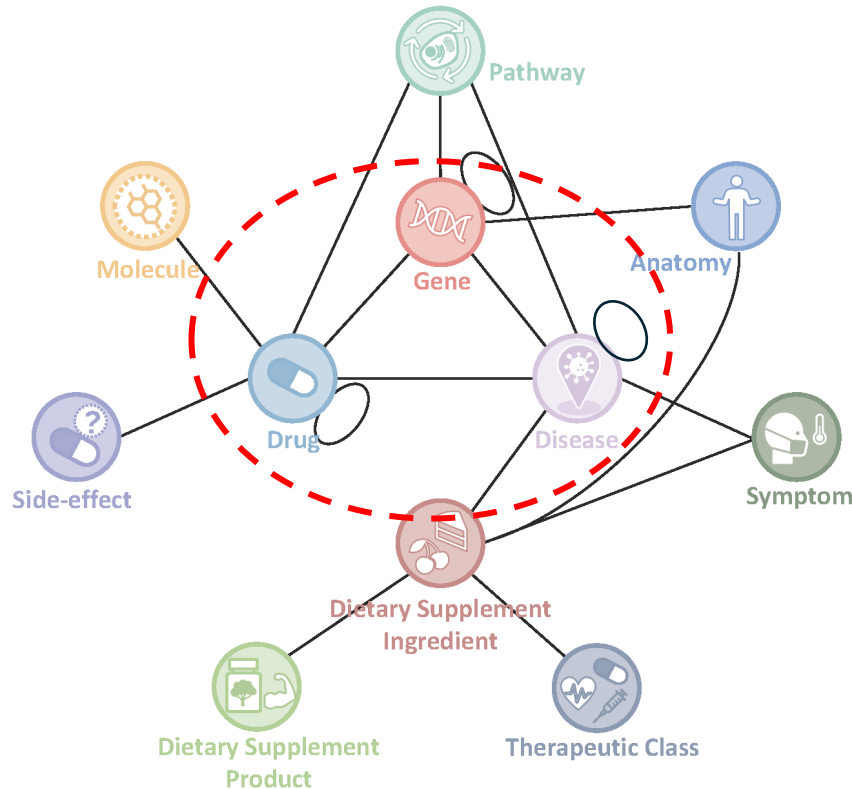
[1] Haghani, M., & Bliemer, M. C. (2020). Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCoV literature. Scientometrics, 125, 2695-2726.

Pandemic paper search

Table 1

The basic information of pandemic knowledge graphs — iBKH dataset

	SARS	MERS	COVID-19	Pandemic graph
#Paper	9,991	1,494	281,569	293,054
Drug	439	46	1,429	1,507
Disease	522	94	1,841	1,814
Gene	1,939	345	5,435	5,821
Drug-drug	145	13	1,626	1,678
Drug-disease	951	59	9,085	9,381
Drug-Gene	710	49	3,709	4,135
Disease-disease	148	26	928	939
Disease-gene	6,256	575	54,503	57,236
Gene-gene	2,199	347	7,461	8,338

Methodology
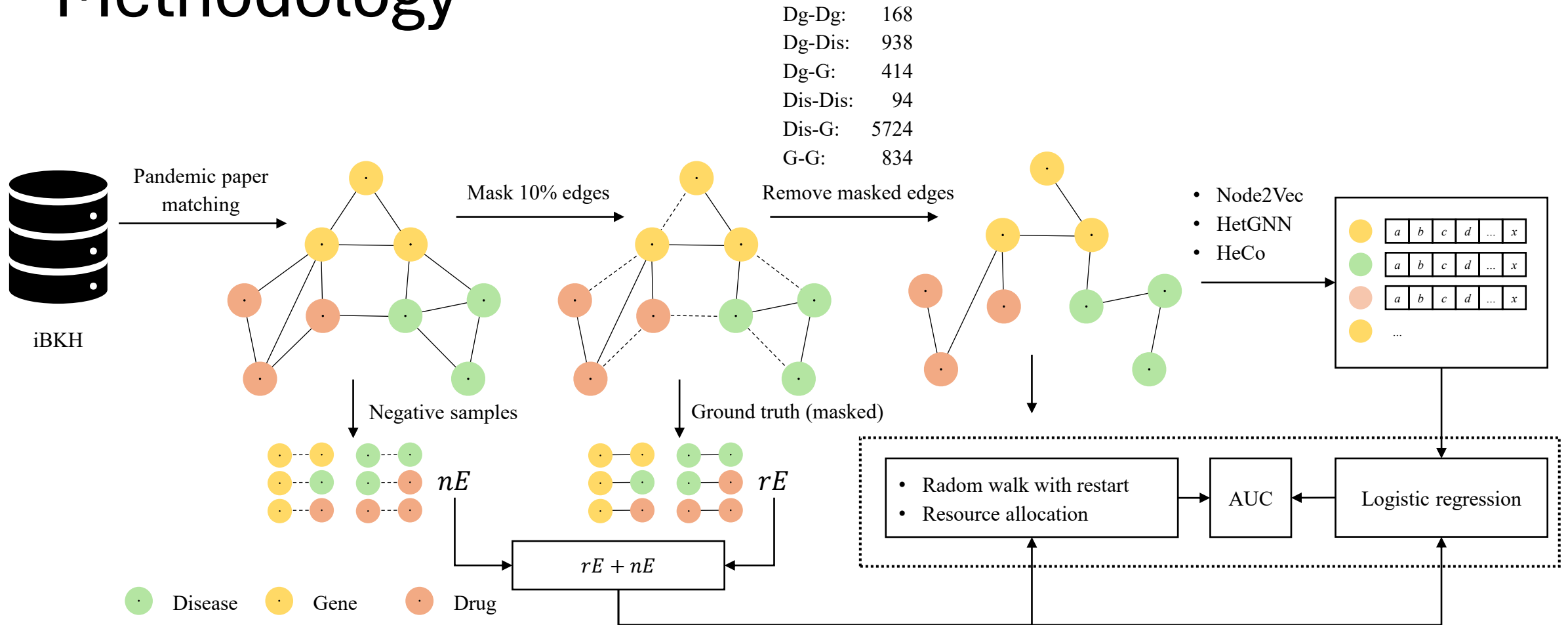
Results

Table 2

Performance comparison of selected algorithms

Method	RWR	RA	Node2Vec	HeCo	**HetGNN**
E_{dg}^{dg}	0.5827	0.5830	0.7257	-	**0.9566**
E_{dg}^{dis}	0.7081	0.7651	0.8079	-	**0.8315**
E_{dg}^{g}	0.8298	0.8741	0.9250	0.9120	**0.9584**
E_{dis}^{dis}	0.7585	0.7893	0.7086	-	**0.8495**
E_{dis}^{g}	0.5327	0.5410	0.7802	0.7990	**0.8001**
E_{g}^{g}	0.7561	0.8110	0.8327	0.8530	**0.9050**

Insights:

- Gene-drug and drug-drug associations prediction are the most promising tasks.
- Heterogeneity in graphs contributes significantly to the association prediction task.

Future directions

- Add edge attribute classification task

- Examine whether LLM outputs can enhance prediction accuracy

- Infer inner-pandemic & cross-pandemic associations

Thanks!

mengjia.wu@uts.edu.au