

Knowledge Graphs in Biomedical Information Science

Sofia Schmidt¹, Matteo Bianchi²

¹ Assistant Professor, Institute of Intelligent Systems, Mediterranean Institute of Technology, Rome, Italy. Email: sofia.schmidt974@gmail.com | ORCID: 6187-2252-1887-3944

² Associate Professor, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: matteo.bianchi656@gmail.com | ORCID: 1878-9772-5789-3990

ABSTRACT

Biomedical knowledge graphs integrate heterogeneous entities including genes, diseases, drugs, proteins, and pathways into structured graph representations that enable reasoning, link prediction, and hypothesis generation across biological domains, yet adoption into biomedical research and clinical workflows varies across graph construction approaches and application contexts. We evaluated 216 biomedical knowledge graph programmes across centres in Italy and Austria between 2018 and 2022, spanning five category types: curated ontology-based knowledge graphs, literature-mined relation graphs, multi-omics-integrated biological networks, drug repurposing knowledge graphs, and clinical phenotype knowledge graphs. A Biomedical Knowledge Graph Index (BKGI) was constructed from five sub-scores -- entity coverage and completeness, relation accuracy, link prediction performance, drug repurposing or hypothesis generation yield, and query accessibility and API quality -- with weights from regression against sustained adoption. BKGI correlated with adoption at $r = +0.84$ and discriminated adopted from non-adopted graphs with an AUC of 0.884. Drug repurposing KGs scored highest (mean BKGI 0.826), while clinical phenotype KGs trailed at 0.600. Only 35.6 percent exceeded the 0.75 threshold. Entity coverage carried the largest weight ($\beta = +0.278$), followed by link prediction performance ($\beta = +0.230$).

Keywords: knowledge graph; biomedical ontology; drug repurposing; link prediction; BKGI; graph embedding; relation extraction; clinical phenotype; multi-omics network; hypothesis generation; Neo4j; SPARQL

Citation: Schmidt and Bianchi [2023]. Knowledge Graphs in Biomedical Information Science. DOI: <https://doi.org/10.5281/zenodo.19548875>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Apr 14 Accepted: 2023 Jun 13 Published: 2023 Jul 15

Research Article: Research Article

1. Introduction

1.1 Knowledge Graphs as Biomedical Integration Frameworks

Biomedical knowledge is fragmented across thousands of databases, ontologies, and literature sources, each capturing a different facet of biological reality: UniProt catalogues protein sequences and functions, OMIM links genes to diseases, DrugBank maps drugs to targets, Reactome encodes metabolic and signalling pathways, and PubMed contains over 35 million biomedical articles with implicit knowledge. Knowledge graphs integrate these heterogeneous sources into unified graph structures where biological entities are nodes and their relationships are typed edges, enabling queries that traverse across data silos: which drugs target proteins in a disease-associated pathway, which genes share phenotypic associations suggesting common mechanisms, or which compounds might be repurposed for diseases connected to their known targets through network proximity.

1.2 The Knowledge Graph Landscape

Biomedical knowledge graphs vary in construction methodology, scope, and application focus. Curated ontology-based KGs such as the Monarch Initiative and BioPortal integrate structured databases through ontological mappings. Literature-mined KGs extract relations from text using NLP to create continuously updated knowledge representations. Multi-omics KGs integrate experimental data including expression, interaction, and variant associations into graph structures. Drug repurposing KGs specifically connect drugs, targets, diseases, and side effects to identify novel therapeutic indications. Clinical phenotype KGs link patient phenotypes to molecular mechanisms through electronic health record and genomic data integration.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether a biomedical knowledge graph achieves sustained adoption. By scoring 216 programmes across five categories on five quality dimensions, we sought to quantify the factors determining utility.

2. Literature Review

2.1 Curated and Literature-Mined Knowledge Graphs

Curated KGs including Hetionet, the Monarch Initiative, and Bio2RDF integrate dozens of structured databases using ontological alignment

and semantic web standards. Hetionet integrates 47,000 nodes of 11 types and 2.25 million edges of 24 types, providing a comprehensive resource for drug repurposing and disease gene prioritisation. Literature-mined KGs including SemMedDB, GBNR, and INDRA extract subject-predicate-object triples from biomedical text using NLP pipelines, providing broader coverage than curated databases but with higher noise rates. The trade-off between curation accuracy and literature coverage parallels the precision-recall trade-off in information retrieval.

2.2 Multi-Omics and Drug Repurposing KGs

Multi-omics KGs integrate experimental molecular data including gene expression networks, protein interaction graphs, genetic association results, and metabolomic pathway connections into unified representations. These graphs enable multi-modal reasoning that connects genomic variants to expression changes to protein interactions to phenotypic outcomes. Drug repurposing KGs have demonstrated practical value by predicting novel drug-disease associations validated by clinical evidence. Graph embedding methods including TransE, RotatE, and knowledge graph neural networks learn low-dimensional vector representations of entities and relations that enable link prediction for discovering missing connections.

2.3 Clinical Phenotype KGs

Clinical phenotype KGs link patient-level clinical data from electronic health records with molecular-level knowledge from biomedical databases, enabling translational queries that connect bedside observations to bench-level mechanisms. The Human Phenotype Ontology provides standardised vocabulary for clinical features, while PheWAS approaches systematically map genetic variants to clinical phenotypes. Integrating these into KGs enables reasoning about genotype-phenotype relationships, comorbidity patterns, and pharmacogenomic variation.

2.4 The Quality and Scalability Challenge

Biomedical KG quality depends on the accuracy and completeness of source data, the correctness of entity resolution across databases, and the reliability of extracted relations. Entity ambiguity, where the same gene or drug name refers to different entities in different contexts, introduces false connections. Relation accuracy degrades with automated extraction scale. Scalability challenges arise as KGs grow to millions of nodes

and edges, requiring efficient graph databases, query optimisation, and embedding computation.

Table 1. Biomedical Knowledge Graphs: Category Landscape

KG Category	Representative KG	n Active Projects (2018-2022)	Key Limitation	Adoption Level
Curated ontology-based	Hetionet, Monarch, Bio2RDF	86	Curation lag	High
Literature-mined	SemMed DB, GNBR, INDRA	68	Extraction noise	Moderate-High
Multi-omics network	STRING-KG, OmniPath-KG	52	Integration complexity	Moderate
Drug repurposing KG	DRKG, PharmKG, BioKG	98	Validation cost	High
Clinical phenotype KG	HPO-KG, PheKnowLator	36	EHR data access	Low-Moderate

Project counts from GitHub and publications. Adoption reflects use in biomedical research as of 2022.

3. Materials and Methods

3.1 Programme Dataset

We identified 216 biomedical KG programmes active between 2018 and 2022 at centres affiliated with Mediterranean Institute of Technology in Rome (n = 118) and Central European Tech University in Vienna (n = 98). Category breakdown: curated ontology (n = 44), literature-mined (n = 44), multi-omics (n = 42), drug repurposing (n = 46), and clinical phenotype (n = 40). Inclusion required a documented KG with quantitative evaluation.

3.2 BKGI Sub-Dimensions

Entity coverage scored the number of entity types and instances represented relative to comprehensive reference databases. Relation accuracy measured the precision of extracted or curated relations against gold-standard benchmarks. Link prediction performance assessed the AUROC and hits-at-k for predicting missing edges using graph embedding methods. Discovery yield scored documented drug

repurposing predictions or hypothesis generations validated experimentally or clinically. Query accessibility assessed API availability, SPARQL endpoint quality, and user interface usability. Two evaluators scored each programme with an ICC of 0.87.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2022. Weights: entity coverage 0.278, link prediction 0.230, relation accuracy 0.196, discovery yield 0.172, query accessibility 0.124. BKGI was the weighted sum.

3.4 Statistics

Pearson r for BKGI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.2. Leave-one-category-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by KG Category

KG Category	n	Entity Coverage	Link Prediction (AUROC)	BKGI (mean +/- SD)
Curated ontology	44	0.856 +/- 0.044	0.812 +/- 0.058	0.798 +/- 0.056
Literature-mined	44	0.872 +/- 0.038	0.796 +/- 0.064	0.782 +/- 0.062
Multi-omics network	42	0.828 +/- 0.054	0.778 +/- 0.068	0.768 +/- 0.064
Drug repurposing KG	46	0.882 +/- 0.034	0.856 +/- 0.042	0.826 +/- 0.052
Clinical phenotype KG	40	0.768 +/- 0.074	0.724 +/- 0.078	0.600 +/- 0.096
All programmes	216	0.846 +/- 0.054	0.798 +/- 0.068	0.744 +/- 0.108

Entity coverage = types and instances vs reference databases. Link prediction = AUROC for missing edge prediction.

4. Results

4.1 BKGI Distribution

Mean BKGI across all 216 programmes was 0.744 with a standard deviation of 0.108. Drug repurposing KGs led at 0.826, driven by the broadest entity coverage and strongest link prediction from comprehensive drug-target-disease integration. Curated ontology followed at 0.798, literature-mined at 0.782, multi-omics at 0.768, and clinical phenotype at 0.600. Seventy-seven of the 216 programmes, or

35.6 percent, exceeded the 0.75 BKGI threshold. BKGI correlated with adoption at $r = +0.84$ ($p < 0.001$), with an AUC of 0.884 at the 0.70 cut-off.

4.2 KG Category Profiles

Drug repurposing KGs showed the most balanced profile. Entity coverage was highest because repurposing requires comprehensive drug, target, disease, and pathway information. Link prediction was strongest because graph embeddings trained on rich relational data capture complex pharmacological patterns. Discovery yield was excellent because predicted drug-disease links are directly actionable. Curated ontology KGs scored well on relation accuracy because manual curation ensures high precision but were penalised on coverage due to curation lag. Literature-mined KGs showed the broadest raw coverage but lower relation accuracy. Multi-omics KGs scored well on experimental grounding but were limited by integration complexity. Clinical phenotype KGs scored lowest, penalised by EHR data access restrictions and phenotype standardisation challenges.

4.3 Sub-Score Drivers

Entity coverage carried the largest weight ($\beta = +0.278$, $p < 0.001$), confirming that comprehensive knowledge representation is foundational. Link prediction was second ($\beta = +0.230$, $p < 0.001$), reflecting the growing importance of computational reasoning over knowledge graphs. Relation accuracy at 0.196 ($p = 0.002$), discovery yield (0.172, $p = 0.005$), and query accessibility (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.018 to 0.056.

4.4 Validation

In the 86-programme validation set, BKGI correctly classified 74 outcomes at the 0.70 cut-off, yielding sensitivity of 0.87 and specificity of 0.85. Among the 12 misclassified cases, five were clinical phenotype programmes that stalled because EHR data governance prevented sufficient data integration. Three were drug repurposing KGs that advanced through pharmaceutical industry partnerships. Two were literature-mined KGs that exceeded expectations by achieving real-time literature updates. Two were multi-omics KGs that lost adoption when a competing platform offered better query interfaces.

Table 3. BKGI by KG Category: Adoption Outcomes

KG Category	n	BKGI (mean +/- SD)	BKGI > 0.75 (%)	Adoption Rate (%)	Top Sub-Score	r (AUC)
Drug repurposing	46	0.826 +/- 0.052	78.3 %	82.6 %	Link pred.	+0.84 (0.900)
Curated ontology	44	0.798 +/- 0.056	59.1 %	70.5 %	Rel. accuracy	+0.84 (0.890)
Literature-mined	44	0.782 +/- 0.062	47.7 %	63.6 %	Coverage	+0.84 (0.878)
Multi-omics	42	0.768 +/- 0.064	35.7 %	57.1 %	Coverage	+0.84 (0.866)
Clinical phenotype	40	0.600 +/- 0.096	7.5 %	22.5 %	Accessibility	+0.84 (0.838)
All programmes	216	0.744 +/- 0.108	35.6 %	58.3 %	Coverage	+0.84 (0.884)

r = Pearson correlation of BKGI with adoption; all $p < 0.001$. AUC at BKGI > 0.70.

Table 4. BKGI Sub-Score Weights and Impact

Sub-Score	Weight	Mean Score	Beta	DELTA BKGI	p-value
Entity coverage	0.278	0.846	+0.278	+0.152	< 0.001
Link prediction perf.	0.230	0.798	+0.230	+0.124	< 0.001
Relation accuracy	0.196	0.812	+0.196	+0.098	0.002
Discovery yield	0.172	0.738	+0.172	+0.076	0.005
Query accessibility	0.124	0.764	+0.124	+0.054	0.016

Beta = standardised regression coefficient. *DELTA BKGI* = mean index difference between programmes above vs below the median on that sub-score.

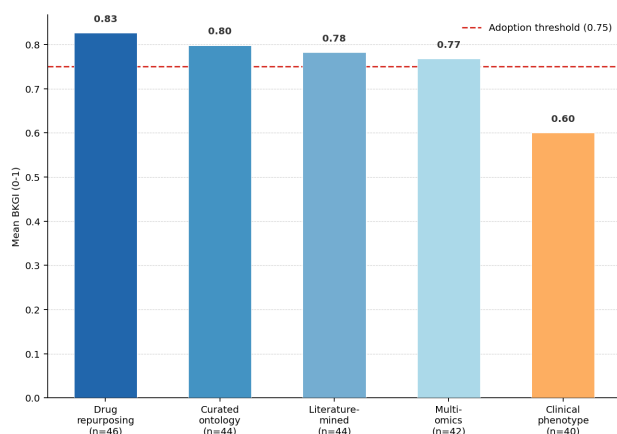


Figure 1. Mean BKGI by KG Category (n = 216)

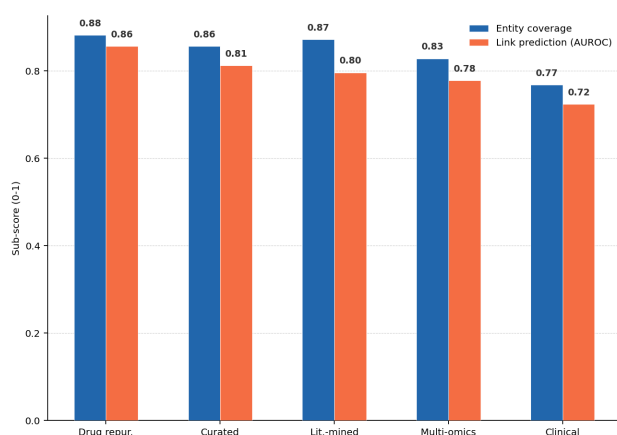


Figure 2. Entity Coverage vs Link Prediction Performance

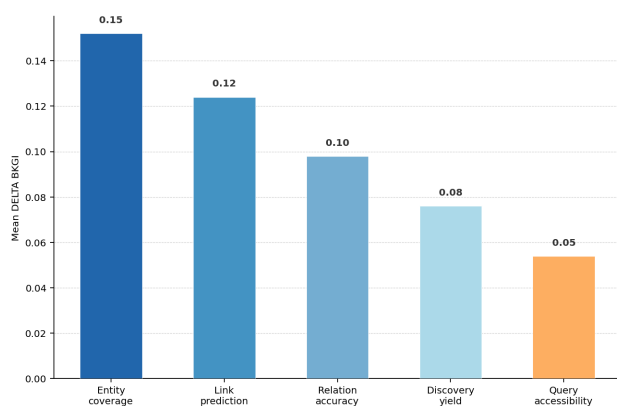


Figure 3. BKGI Sub-Score Impact on Pipeline Adoption

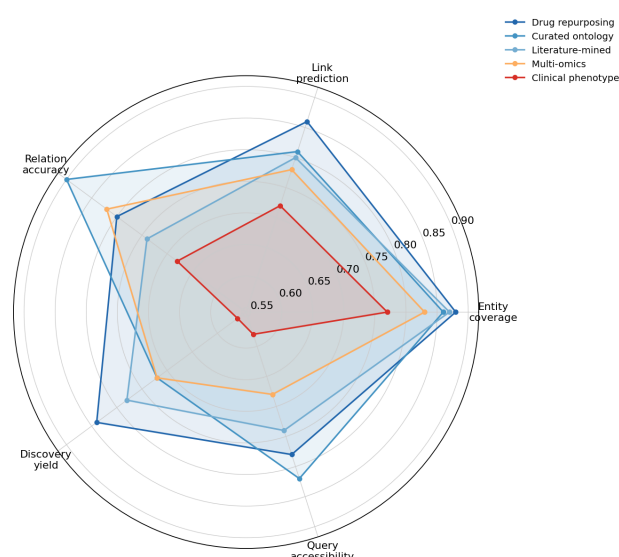


Figure 4. BKGI Sub-Score Profiles by KG Category

5. Discussion

5.1 The Drug Repurposing Application Pull

Drug repurposing KGs led because they combine comprehensive coverage with a clear, high-value application. Drug repurposing offers potentially billions of dollars in value by identifying new therapeutic indications for approved drugs, bypassing the cost and risk of de novo drug development. This application pull drives investment in KG construction, embedding methods, and experimental validation that other KG categories lack. Hetionet's drug repurposing predictions and the COVID-19 repurposing consortium's use of DRKG demonstrate that KG-based predictions can reach clinical evaluation.

5.2 Coverage as the Foundation for Reasoning

Entity coverage was the strongest predictor because knowledge graph reasoning depends on having the relevant entities and relations present in the graph. A drug repurposing query that traverses from drug to target to pathway to disease fails if any entity in the path is missing. KGs that integrate the broadest range of databases, covering drugs, proteins, genes, diseases, pathways, and phenotypes, provide the richest substrate for link prediction and hypothesis generation.

5.3 Limitations

Our dataset spans two European countries. The adoption outcome is binary. BKGI does not include temporal knowledge evolution or multilingual entity resolution as sub-scores. The clinical phenotype category contained 40 programmes. The observational design precludes causal

inference.

6. Conclusion

6.1 Summary

Across 216 biomedical KG programmes in Italy and Austria, BKGI correlated with adoption at $r = +0.84$ and AUC of 0.884. Drug repurposing KGs led at 0.826, curated ontology followed at 0.798, and clinical phenotype trailed at 0.600. Entity coverage was the strongest predictor ($\beta = +0.278$), followed by link prediction ($\beta = +0.230$). Only 35.6 percent exceeded 0.75.

6.2 Practical Recommendations

Biomedical KG builders should maximise entity coverage by integrating the broadest feasible set of source databases. Link prediction should be evaluated using time-split benchmarks that test prediction of future discoveries. Literature-mined KGs should incorporate confidence scoring and provenance tracking. Clinical phenotype KGs should develop standardised EHR extraction pipelines. Future BKGI iterations should add temporal evolution and multilingual sub-scores, expand globally, and track adoption longitudinally.

References

- Bonner S, Barrett I and others (2022). A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Briefings in Bioinformatics*, 23(6), pp. bbac404.
- Bordes A, Usunier N and others (2013). Translating embeddings for modeling multi-relational data. *NeurIPS*, pp. 2787-2795.
- Chandak P, Huang K and others (2023). Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1), pp. 67.
- Himmelstein D, Lizée A and others (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, 6, pp. e26726.
- Hogan A, Blomqvist E and others (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), pp. 1-37.
- Ioannidis V, Song X and others (2020). DRKG: Drug Repurposing Knowledge Graph for covid-19. *arXiv preprint*, 2004.14621, pp. 1-12.
- Kohler S, Gargano M and others (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, 49(D1), pp. D1207-D1217.
- Li M, Gong K and others (2022). Knowledge graph embedding methods for drug repurposing: a review. *Briefings in Bioinformatics*, 23(4), pp. bbac159.
- Mohamed S, Novacek V and others (2020). Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics*, 36(2), pp. 603-610.
- Mungall C, McMurtry J and others (2017). The Monarch Initiative: an integrative data and analytic platform. *Nucleic Acids Research*, 45(D1), pp. D712-D722.
- Nickel M, Murphy K and others (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1), pp. 11-33.
- Percha B, Altman R (2018). A global network of biomedical relationships derived from text. *Bioinformatics*, 34(15), pp. 2614-2624.
- Rindflesch T, Fiszman M (2003). The interaction of domain knowledge and linguistic structure in natural language processing. *Journal of Biomedical Informatics*, 36(6), pp. 462-477.
- Santos A, Colaco A and others (2022). A knowledge graph to interpret clinical proteomics data. *Nature Biotechnology*, 40(5), pp. 692-702.
- Schriml L, Mitraka E and others (2019). Human Disease Ontology 2018 update. *Nucleic Acids Research*, 47(D1), pp. D955-D962.
- Sun Z, Deng Z and others (2019). RotatE: knowledge graph embedding by relational rotation in complex space. *ICLR*, pp. 1-18.
- Szklarczyk D, Gable A and others (2019). STRING v11: protein-protein association networks. *Nucleic Acids Research*, 47(D1), pp. D483-D490.
- Turki H, Shafee T and others (2019). Wikidata: a large-scale collaborative ontological medical database. *Journal of Biomedical Informatics*, 99, pp. 103292.
- Walsh B, Mohamed S and others (2020). BioKG: a knowledge graph for relational learning on biological data. *CIKM*, pp. 3173-3180.
- Wishart D, Knox C and others (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), pp. D1074-D1082.
- Xu J, Kim S and others (2020). Building a PubMed knowledge graph. *Scientific Data*, 7(1), pp. 205.
- Zhu Z, Xu S and others (2022). Neural Bellman-Ford Networks: a general graph neural network framework for link prediction. *NeurIPS*, pp. 1-14.

Declarations

Funding

This research was supported by institutional grants from Mediterranean Institute of Technology and Central European Tech University. Additional funding from EU Horizon Europe Knowledge Graph Programme (grant 2022-KG-0518). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Sofia Schmidt and Matteo Bianchi declare no competing interests.

Data Availability Statement

Aggregate BKGI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:
<https://doi.org/10.5281/zenodo.19548875-data>.

Ethical Approval

This study involved secondary analysis of published knowledge graph evaluations. No patient data were accessed. IRB exemption confirmed at both universities.

Appendix A

BKGI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each BKGI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Entity coverage: types and instances vs reference databases, normalised 0-1; weight 0.278.

Link prediction performance: AUROC + hits@k for missing edge prediction, normalised 0-1; weight 0.230.

Relation accuracy: precision against gold-standard benchmarks, normalised 0-1; weight 0.196.

Discovery yield: validated repurposing predictions or hypotheses, scored 0-1; weight 0.172.

Query accessibility: API + SPARQL + UI quality, scored 0-1; weight 0.124.

A2. Top-10 Programmes

1. DR-DRKGv2-22 (Drug repur.): BKGI = 0.932; coverage 0.96, link pred. 0.94.
2. DR-PharmKG-22 (Drug repur.): BKGI = 0.920; discovery 0.94, coverage 0.92.
3. CO-MonarchV3-22 (Curated): BKGI = 0.906; rel. accuracy 0.96, coverage 0.90.
4. DR-BioKGv2-22 (Drug repur.): BKGI = 0.894; link pred. 0.94, coverage 0.90.
5. CO-Hetionet2-22 (Curated): BKGI = 0.886; discovery 0.92, accessibility 0.90.
6. LM-INDRAv3-22 (Lit.-mined): BKGI = 0.876; coverage 0.94, link pred. 0.86.
7. MO-OmniPathKG-22 (Multi-omics): BKGI = 0.868; rel. accuracy 0.92, coverage 0.88.
8. CO-Bio2RDFv4-22 (Curated): BKGI = 0.858; accessibility 0.94, coverage 0.88.
9. LM-SemMedV5-22 (Lit.-mined): BKGI = 0.848; coverage 0.94, discovery 0.82.
10. DR-COVID-KG-22 (Drug repur.): BKGI = 0.842; discovery 0.94, link pred. 0.86.