Knowledge Graphs & Drug Repurposing

Preparatory Work for the Master Thesis 2024-25

Siddharth Sahay
Promoter: Prof. Tom Lenaerts
Advisers: Dr. Nassim Versbraegen, Inas Bosch

UNIVERSITÉ LIBRE DE BRUXELLES

Outline

1. Drug Repurposing
2. Knowledge Graphs & Knowledge Graph Embeddings
3. State-of-the-art
4. Biomedical KGs
5. Evaluation metrics
6. Main challenges
7. Thesis roadmap

Drug Repurposing and Discovery

- ~7000 rare diseases; <6% have approved therapy
- $2.5B and 10+ years per drug
- Repurposing can cut costs and save time, drastically
- Drug-disease search space is huge
- KGs + KGEs organise and explore this space

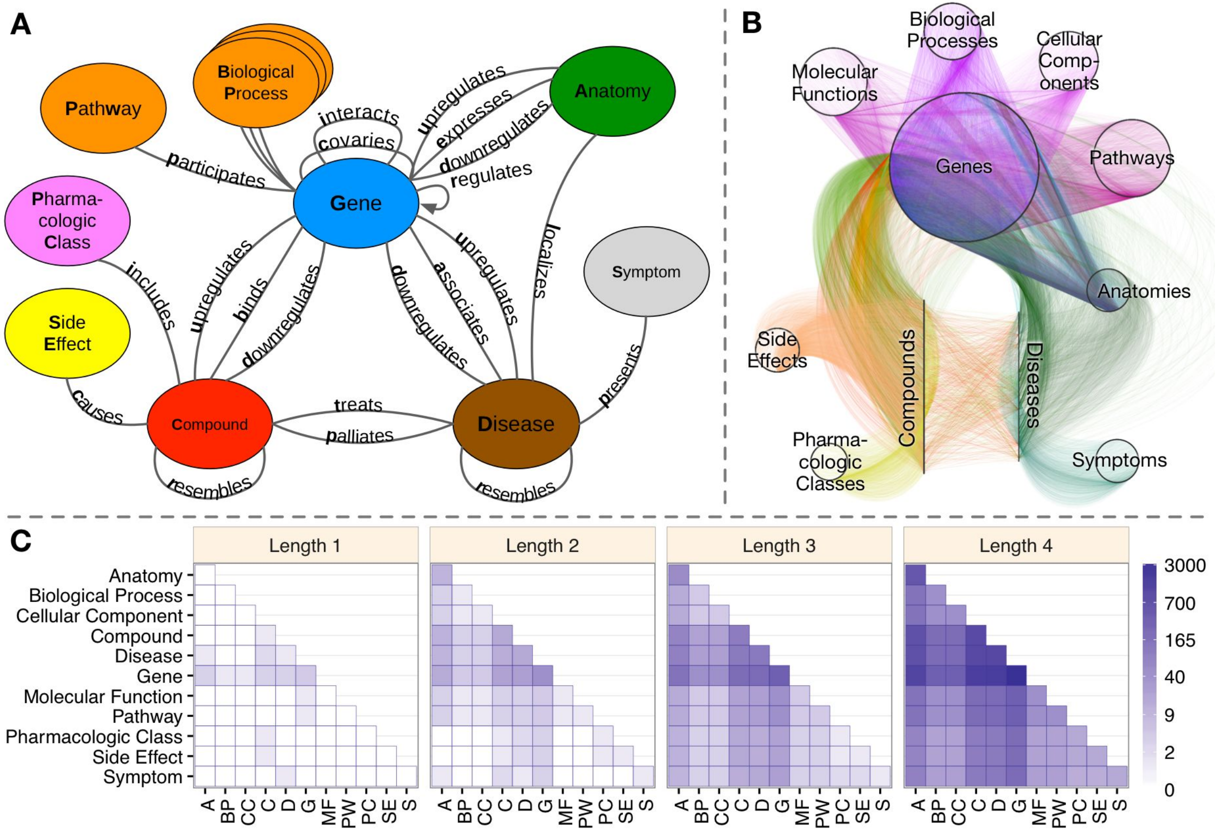
Knowledge Graphs

- Consists of triples (head, relation, tail)
- Example: (Luke Skywalker, SonOf, Darth Vader)

Biomedical KGs

- Nodes: drugs, diseases, genes, pathways, edges
- Edges: relationships between these nodes
- Great for human intuition and biomedical knowledge representation

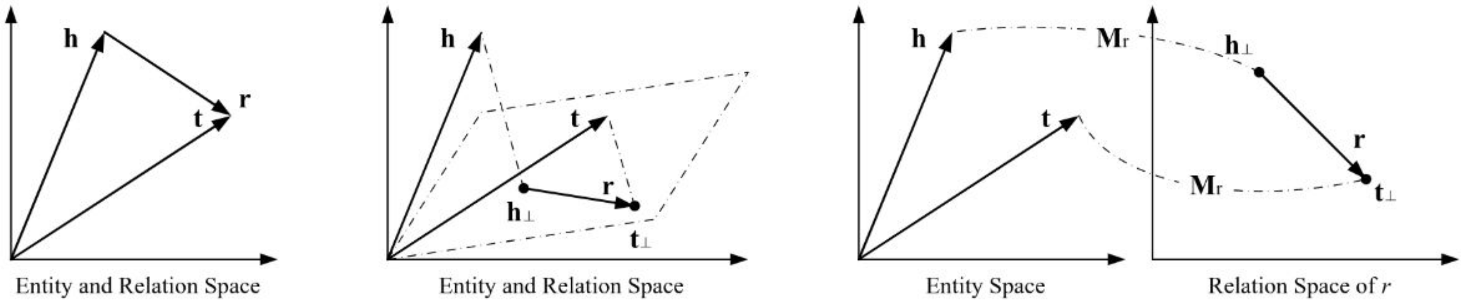
Hetionet KG



A
Biological Process
Pathway
Anatomy
Gene
Pharma-cologic Class
Side Effect
Symptom
Compound
Disease
interacts
covaries
participates
upregulates
expresses
downregulates
regulates
includes
upregulates
binds
downregulates
downregulates
associates
upregulates
localizes
presents
causes
treats
palliates
resembles
resembles
B
Biological Processes
Cellular Comp-onents
Molecular Functions
Genes
Pathways
Anatomies
Side Effects
Compounds
Diseases
Pharma-cologic Classes
Symptoms
C
Length 1
Length 2
Length 3
Length 4
Anatomy
Biological Process
Cellular Component
Compound
Disease
Gene
Molecular Function
Pathway
Pharmacologic Class
Side Effect
Symptom
A BP CC C D G MF PW PC SE S
3000
700
165
40
9
2
0

Knowledge Graph Embeddings

- Project relations into high-dimensional vector space
- Easier for ML models to use for link prediction
- Various methods: scoring function-based, path-based and semantic matching models



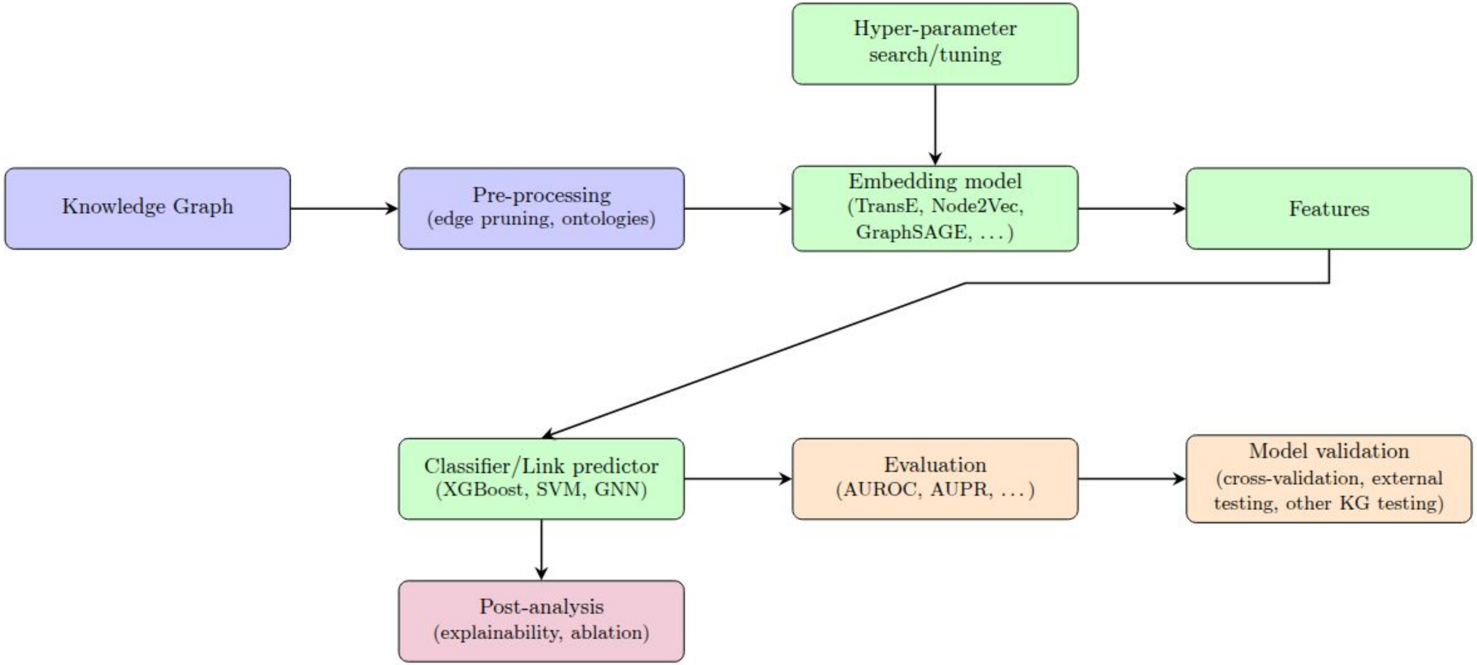
(a) TransE. (b) TransH. (c) TransR.

UNIVERSITÉ LIBRE DE BRUXELLES

State of the art

- Traditional ML methods: DT2Vec+
- Random-walk based: DREAMwalk, AnyBURL
- Deep Learning (GNN) based: GDRNet, DRAGNN, EKGDR, DTD-GNN
- LLM based: DrugChat, MoCoSA, LMKE
- Other: RPath, PoLo, GNBR

Generalised pipeline for Drug Repurposing



UNIVERSITÉ LIBRE DE BRUXELLES

eXplainable AI & Interpretability

- XAI makes ML models more transparent and understandable
- Many methods:
  - Path-based reasoning
  - Subgraph extraction
  - Logical pattern recognition
  - Attention interpretation with GATs
  - Counterfactual reasoning

Key biomedical KGs

KG	Last Updated	Nodes	Edges	Types (N/E)	Sources	Focus
Hetionet	2017	47k	2.3M	11 / 24	29 DBs	Repurposing, gene-disease
PharMeBINet	2024	2.9M	15.9M	66 / 208	Hetionet + 19	Repurposing, gene-disease
GNBR	2018	130k+	2M+	3 / 32	PubMed abstracts	Literature-based
Bioteque	2022	450k+	30M+	12 / 67	150+ DBs	Precomputed embeddings
CKG	2024	~16M	220M+	19 / 57	35 sources	Clinical
DRKG	2020	97k	5.9M	13 / 107	6 DBs + COVID pubs	Repurposing
BOCK	2023	159k	2.7M	10 / 17	Curated + networks	Oligogenic
OREGANO	2023	89k	825k	11 / 19	7 DBs	Natural compounds, repurposing

Evaluation metrics

- AUROC, AUPR
- Hits@K
- Mean Rank and Mean Reciprocal Rank

Limitations

- Bias towards PPI
- Data incompleteness
- Scalability
- Beyond second-order neighbourhoods
- Interpretability

Thesis roadmap Q1-Q4

1. Baseline benchmarking
    a. Systematically compare pipelines
    b. KGEs: TransE, DistMult, random-walk based
    c. Classifiers: XGBoost, SVMs, GNNs
2. Optimisations and Oligogenic extension
    a. Hyperparameter search
    b. Integration with BOCK
3. Designing a novel method
    a. Fill all gaps in baseline
    b. Experiment further with GNNs
4. Testing and writing

References

Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). *Knowledge graph embedding: A survey of approaches and applications.* IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2017.2754499

Mohamed, S. K., Nounu, A., & Nováček, V. (2020). *Biological applications of knowledge-graph embedding models.* Briefings in Bioinformatics. https://doi.org/10.1093/bib/bbaa012

Bordes, A., Usunier, N., García-Durán, A., Weston, J., & Yakhnenko, O. (2013). *Translating embeddings for modeling multi-relational data (TransE).* Advances in Neural Information Processing Systems.

Ali, M., Berrendorf, M., Hoyt, C. T., et al. (2021). *Bringing light into the dark: A large-scale evaluation of KGE models under a unified framework.* CoRR preprint arXiv:2006.13365.

Himmelstein, D. S., Lizee, A., Hessler, C., et al. (2017). *Systematic integration of biomedical knowledge prioritizes drugs for repurposing.* eLife. https://doi.org/10.7554/eLife.26726

Bang, D., Lim, S., Lee, S., & Kim, S. (2023). *Biomedical KG learning for drug repurposing by extending guilt-by-association to multiple layers.* Nature Communications. https://doi.org/10.1038/s41467-023-39301-y

Tayebi, J., & BabaAli, B. (2024). *EKGDR: An end-to-end KG-based method for computational drug repurposing.* Journal of Chemical Information and Modeling. https://doi.org/10.1021/acs.jcim.3c01925

Johnson, R., Li, M. M., Noori, A., Queen, O., & Zitnik, M. (2024). *Graph artificial intelligence in medicine.* Annual Review of Biomedical Data Science. https://doi.org/10.1146/annurev-biodatasci-110723-024625

Perdomo-Quinteiro, P., & Belmonte-Hernández, A. (2024). *Knowledge graphs for drug repurposing: a review of databases and methods.* Briefings in Bioinformatics. https://doi.org/10.1093/bib/bbae461

Jiménez, A., Merino, M. J., Parras, J., & Zazo, S. (2024). *Explainable drug repurposing via path-based KG completion.* Scientific Reports. https://doi.org/10.1038/s41598-024-67163-x

ULB UNIVERSITÉ LIBRE DE BRUXELLES