

# Literature-Based Discovery and Imperfect Knowledge In Biomedicine

First International Workshop on Representing and Reasoning with Imperfect knowledge – KGC 2022

**Ali Daowd** ([ali.daowd@dal.ca](mailto:ali.daowd@dal.ca))

NICHE Research Group

Dalhousie University



# Undiscovered Public Knowledge

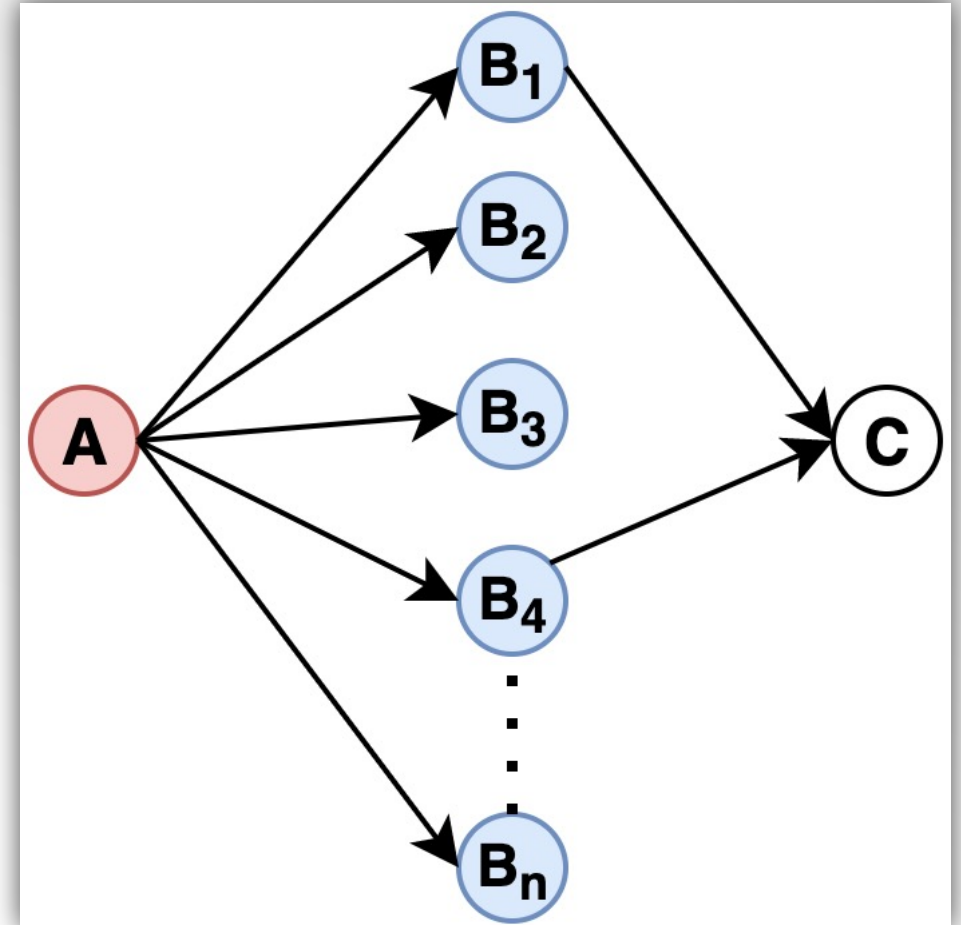
- Exponential growth in volume of scientific literature:
  - PubMed > 800,000 new articles/year in last 10 years <sup>(1)</sup>
  - Nature Portfolio > 100 new articles/year in 156 specialized journals <sup>(2)</sup>
- Fragmentation of evidence-based knowledge
- Wealth of undiscovered complementary and dynamic knowledge in published literature

(1) Landhuis, E. (2016). Scientific literature: Information overload. *Nature*, 535(7612), 457-458.

(2) Springer Nature, Nature portfolio, accessed 30 April 2022, <<https://www.nature.com/siteindex>>

# Literature-Based Discovery (LBD)

- Aims to *discover* and *synthesize* knowledge by connecting and reasoning over fragmented literature
- Pioneered by Don Swanson's *ABC model* in 1985 <sup>(1)</sup>
  - **{A, B}** and **{B, C}** are *complementary* but *disjoint*
  - Implicit relationship between A and C
- Knowledge triples (s, p, o) represented as a knowledge graph



**ABC discovery model**

(1) Swanson, D. R. (1986). Undiscovered public knowledge. The Library Quarterly, 56(2), 103-118.

# Literature-Based Discovery Applications

1. Drug repurposing <sup>(1, 2)</sup>
2. Identifying adverse drug events <sup>(3, 4)</sup>
3. Explaining underlying molecular mechanisms of diseases <sup>(5,6)</sup>
4. Hypothesis generation

- (1) Yang, H. T., Ju, J. H., Wong, Y. T., Shmulevich, I., & Chiang, J. H. (2017). Literature-based discovery of new candidates for drug repurposing.
- (2) Daowd, A., Abidi, S., & Abidi, S. S. R. (2022). A Knowledge Graph Completion Method Applied to Literature-Based Discovery for Predicting Missing Links Targeting Cancer Drug Repurposing. *Artificial Intelligence in Medicine (forthcoming)*.
- (3) Hristovski, D., Kastrin, A., Dinevski, D., Burgun, A., Žiberna, L., & Rindflesch, T. C. (2016). Using literature-based discovery to explain adverse drug effects. *Journal of medical systems*, 40(8), 1-5.
- (4) Rastegar-Mojarad, M., Elayavilli, R. K., Wang, L., Prasad, R., & Liu, H. (2016, October). Prioritizing adverse drug reaction and drug repositioning candidates generated by literature-based discovery. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 289-296).
- (5) Henry, S., Wijesinghe, D. S., Myers, A., & McInnes, B. T. (2021). Using Literature Based Discovery to Gain Insights Into the Metabolomic Processes of Cardiac Arrest. *Frontiers in Research Metrics and Analytics*, 6, 32.
- (6) Daowd, A., Barrett, M., Abidi, S., & Abidi, S. S. R. (2021, August). A Framework To Build A Causal Knowledge Graph for Chronic Diseases and Cancers By Discovering Semantic Associations from Biomedical Literature. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)* (pp. 13-22).

# Literature-Based Discovery Challenges

- Accuracy of knowledge extraction
  - Co-occurrence based models: high recall, low precision
  - Semantic based models: low recall, high precision
- Missing facts in the literature
  - E.g., tacit knowledge

# How to Handle Incomplete and Imperfect Literature-Based Knowledge?

## I. Incomplete knowledge

- Knowledge graph completion (KGC)

## II. Imperfect knowledge

- Language models to classify meaningful semantic triples

# How to Handle Incomplete and Imperfect Literature-Based Knowledge?

- Knowledge Graph Completion (KGC) for literature-based KGs:
  - Infer missing knowledge based on existing triples
  - Knowledge graph embeddings – *TransE*, *ComplEx*, *DistMult*, *HolE*, *ConvE*
  - Entity prediction (*head, relation, ?*)
  - Relation prediction (*head, ?, tail*)

# How to Handle Incomplete and Imperfect Literature-Based Knowledge?

- Knowledge Graph Completion for literature-based KGs:
  - Completing large KGs is unrealistic
    - Computationally expensive, dense ‘completed’ KG
  - Need to focus KGC on a subset of knowledge:
    - MeSH descriptors provide concept-level implicit summary of biomedical literature

Daowd, A., Abidi, S., & Abidi, S. S. R. (2022). A Knowledge Graph Completion Method Applied to Literature-Based Discovery for Predicting Missing Links Targeting Cancer Drug Repurposing. *Artificial Intelligence in Medicine* (forthcoming)

> [Cell Signal](#). 2020 Feb;66:109488. doi: 10.1016/j.cellsig.2019.109488. Epub 2019 Nov 27.

## CXCL12–CXCL4 heterodimerization prevents CXCL12–driven breast cancer cell migration

Khanh T P Nguyen<sup>1</sup>, Lawrence J Druhan<sup>2</sup>, Belinda R Avalos<sup>2</sup>, Li Zhai<sup>3</sup>, Lubica Rauova<sup>4</sup>, Irina V Nesmelova<sup>5</sup>, Didier Dréau<sup>6</sup>

Affiliations + expand

PMID: 31785332 DOI: [10.1016/j.cellsig.2019.109488](#)

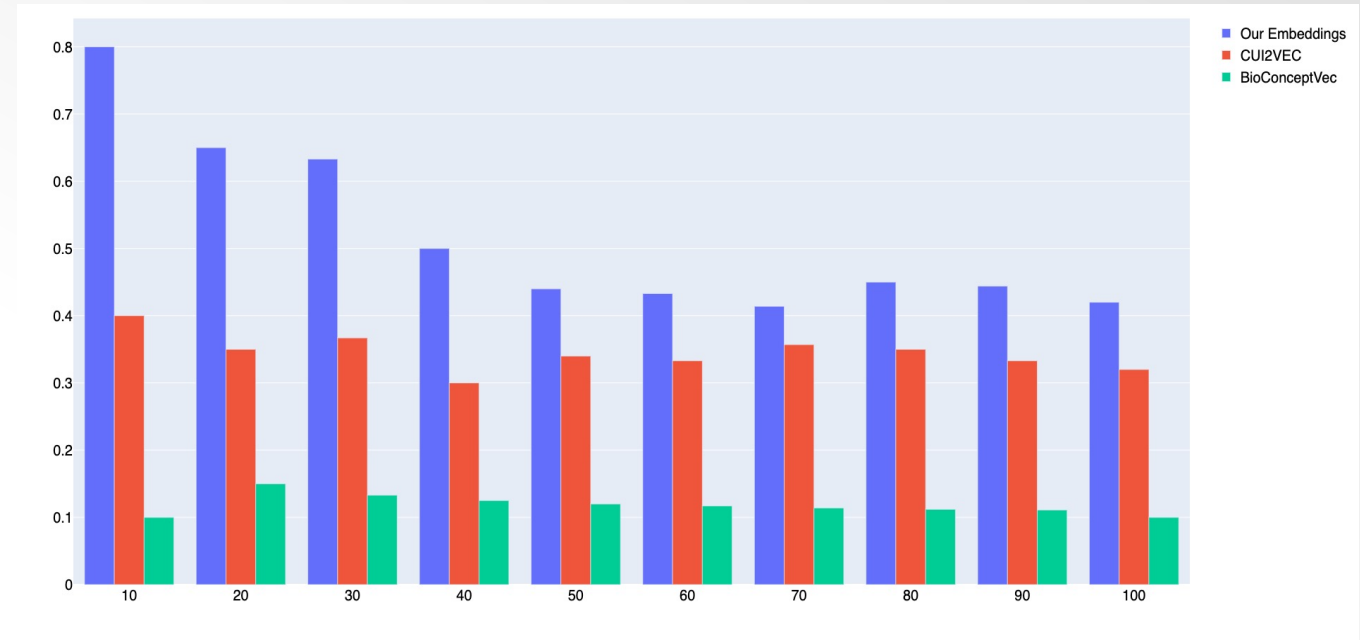
### MeSH terms

- > [Adenocarcinoma / metabolism\\*](#)
- > [Adenocarcinoma / pathology](#)
- > [Cell Line, Tumor](#)
- > [Cell Movement](#)
- > [Chemokine CXCL12 / metabolism\\*](#)
- > [Female](#)
- > [Humans](#)
- > [Platelet Factor 4 / metabolism\\*](#)
- > [Protein Multimerization](#)
- > [Triple Negative Breast Neoplasms / metabolism\\*](#)
- > [Triple Negative Breast Neoplasms / pathology](#)



# How to Handle Incomplete and Imperfect Literature-Based Knowledge?

- Language models to identify meaningful semantic triples
  - Word2Vec, BioBERT
  - Vector-oriented reasoning



Daowd, A., Barrett, M., Abidi, S., & Abidi, S. S. R. (2021, August). A Framework To Build A Causal Knowledge Graph for Chronic Diseases and Cancers By Discovering Semantic Associations from Biomedical Literature. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)* (pp. 13-22).

# Remaining Challenges

- Incompleteness is still a problem in almost all biomedical KGs – including literature-based ones
  - Complexity of biomedical language
  - Contradictory knowledge
  - Biased research and invalid findings – “streetlight effect”