

# VU Research Portal

## Exploiting Subgraphs and Attributes for Representation Learning on Knowledge Graphs

Daza Cruz, Daniel Fernando

2024

**DOI (link to publisher)**

[10.5463/thesis.823](https://doi.org/10.5463/thesis.823)

**document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

**citation for published version (APA)**

Daza Cruz, D. F. (2024). *Exploiting Subgraphs and Attributes for Representation Learning on Knowledge Graphs*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam]. <https://doi.org/10.5463/thesis.823>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                                    | <b>1</b>  |
| 1.1      | Knowledge Graphs . . . . .                             | 2         |
| 1.2      | Learning representations on knowledge graphs . . . . . | 4         |
| 1.3      | Research questions . . . . .                           | 5         |
| 1.4      | Research contributions . . . . .                       | 7         |
| <b>2</b> | <b>Learning from Complex Queries</b>                   | <b>10</b> |
| 2.1      | Introduction . . . . .                                 | 10        |
| 2.2      | Problem Definition . . . . .                           | 12        |
| 2.3      | Related Work . . . . .                                 | 13        |
| 2.4      | Message Passing Query Embedding . . . . .              | 14        |
|          | 2.4.1 Model Definition . . . . .                       | 15        |
|          | 2.4.2 Properties of MPQE . . . . .                     | 16        |
| 2.5      | Experiments . . . . .                                  | 17        |
|          | 2.5.1 Results . . . . .                                | 18        |
| 2.6      | Conclusion . . . . .                                   | 20        |
| <b>3</b> | <b>Learning via Complex Query Decomposition</b>        | <b>22</b> |
| 3.1      | Introduction . . . . .                                 | 22        |
| 3.2      | Problem Definition . . . . .                           | 24        |
| 3.3      | Complex Query Answering via optimization . . . . .     | 26        |
|          | 3.3.1 CQD via Continuous Optimization . . . . .        | 27        |
|          | 3.3.2 CQD via Combinatorial Optimization . . . . .     | 28        |
|          | 3.3.3 Adaptive CQD . . . . .                           | 29        |

|          |  |           |
|----------|--|-----------|
| 3.4      | Related Work . . . . .   | 30        |
| 3.5      | Experiments . . . . .  | 31        |
| 3.5.1    | Experiments with CQD . . . . .                                   | 31        |
| 3.5.2    | Experiments with CQD <sup>A</sup> . . . . .                      | 36        |
| 3.5.3    | Results . . . . .  | 38        |
| 3.6      | Conclusion . . . . .   | 41        |
| <b>4</b> | <b>Learning from Textual Attributes</b>                          | <b>43</b> |
| 4.1      | Introduction . . . . .   | 43        |
| 4.2      | Related work . . . . .   | 45        |
| 4.3      | Inductive entity representations . . . . .                       | 47        |
| 4.3.1    | BERT for entity descriptions . . . . .                           | 48        |
| 4.3.2    | Computational complexity . . . . .                               | 49        |
| 4.3.3    | Theoretical motivation . . . . .                                 | 50        |
| 4.4      | Evaluating entity representations . . . . .                      | 50        |
| 4.4.1    | Link prediction . . . . .  | 51        |
| 4.4.2    | Entity classification . . . . .                                  | 57        |
| 4.4.3    | Information retrieval . . . . .                                  | 58        |
| 4.5      | Conclusion . . . . .   | 61        |
| <b>5</b> | <b>Learning from Multimodal Biomedical Data</b>                  | <b>63</b> |
| 5.1      | Introduction . . . . .   | 63        |
| 5.1.1    | Link Prediction with Knowledge Graph Embeddings                  | 65        |
| 5.1.2    | Incorporating attributes in knowledge graph embeddings . . . . . | 67        |
| 5.1.3    | Knowledge Graph Embedding Models in Biology. . . . .             | 68        |
| 5.1.4    | Research objectives . . . . .                                    | 69        |
| 5.2      | Methods . . . . .  | 69        |
| 5.2.1    | The BioKG knowledge graph . . . . .                              | 69        |
| 5.2.2    | Biomedical benchmarks . . . . .                                  | 71        |
| 5.2.3    | Data partitioning . . . . .                                      | 72        |
| 5.2.4    | BioBLP . . . . .   | 73        |
| 5.2.5    | Evaluation . . . . .   | 79        |
| 5.3      | Results . . . . .  | 83        |
| 5.3.1    | Hyperparameter study . . . . .                                   | 83        |
| 5.3.2    | Link prediction performance . . . . .                            | 83        |
| 5.3.3    | Benchmarks . . . . .   | 86        |
| 5.3.4    | Performance analysis . . . . .                                   | 88        |
| 5.4      | Discussion . . . . .   | 92        |
| 5.5      | Conclusion . . . . .   | 93        |

|  |            |
|--|------------|
| <b>6 Conclusion</b>                                      | <b>95</b>  |
| <b>A Appendix</b>  | <b>100</b> |
| A.1 Complex Query Decomposition . . . . .                | 100        |
| A.1.1 Influence of the Embedding Size on the Results . . | 100        |
| A.1.2 Timing Experiments . . . . .                       | 100        |
| A.1.3 DistMult Experiments . . . . .                     | 103        |
| A.2 Adaptive Complex Query Decomposition . . . . .       | 103        |
| A.2.1 Impact of adaptation . . . . .                     | 103        |
| A.3 Learning from Textual Attributes . . . . .           | 105        |
| A.3.1 Technical details . . . . .                        | 105        |
| A.3.2 Computing resources . . . . .                      | 105        |
| A.3.3 Effect of description length . . . . .             | 106        |
| A.3.4 Transductive link prediction . . . . .             | 106        |
| A.4 Learning from Multimodal Biomedical Data . . . . .   | 107        |
| A.4.1 Additional benchmark results . . . . .             | 107        |
| <b>References</b>  | <b>109</b> |
| <b>List of Figures</b>                                   | <b>133</b> |
| <b>List of Tables</b>                                    | <b>137</b> |
| <b>SIKS Dissertatiereeks</b>                             | <b>141</b> |