Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 2325–2331
Marseille, 11–16 May 2020

Multiple Knowledge GraphDB (MKGDB)

Stefano Faralli[1], Paola Velardi[2], Farid Yusifli[2]
[1] University of Rome Unitelma-Sapienza, stefano.faralli@unitelmasapienza.it
[2] University of Rome Sapienza, velardi@di.uniroma1.it,yusifli.1772848@studenti.uniroma1.it

Abstract

We present MKGDB, a large-scale graph database created as a combination of multiple taxonomy backbones extracted from 5 existing knowledge graphs, namely: ConceptNet, DBpedia, WebIsAGraph, WordNet and the Wikipedia category hierarchy. MKGDB, thanks the versatility of the Neo4j graph database manager technology, is intended to favour and help the development of open-domain natural language processing applications relying on knowledge bases, such as information extraction, hypernymy discovery, topic clustering, and others. Our resource consists of a large hypernymy graph which counts more than 37 million nodes and more than 81 million hypernymy relations.
Keywords: knowledge graphs, graph databases, hypernymy graphs

1. Introduction

Since the end of the last century, both semantic networks (Allen and Frisch, 1982) and knowledge graphs (KGs) are playing the important role of representing entities and relations with high reliability, explainability, and reusability (Bonatti et al., 2018).
KGs have had an impact on the development of many fields of research (Camacho-Collados et al., 2018) ranging from the Semantic Web (Shadbolt et al., 2006) to Natural Language Processing (NLP)[1]. In order to speed up and support the development of novel knowledge-based applications, we present MKGDB, a (very) large scale graph database created as a combination of multiple taxonomy backbones, extracted [2] from 5 existing knowledge graphs namely: ConceptNet (Speer et al., 2017), DBpedia (Lehmann et al., 2015), WebIsAGraph (Faralli et al., 2019), WordNet (Fellbaum, 1998) and the Wikipedia hierarchy of categories.
The resource combines multiple lexical knowledge graphs representing entities and hypernymy relations, both automatically harvested from corpora (i.e., WebIsAGraph) and crafted by human experts (like WordNet and others).
Thanks to the availability of different streams of knowledge and the versatility of graph database technologies, we exploit methodologies leveraging topological features from multiple knowledge graphs.
As an example, MKGDB contains a high number of cross-link edges (edges connecting nodes belonging to different knowledge graphs) which are fundamental information in algorithms such as linking and mining heterogeneous data (P and Jurek-Loughrey, 2018), entity alignment between knowledge graphs (Trisedya et al., 2019) and noisy graph pruning (Faralli et al., 2017).
Our resource is compiled on top of the Neo4j platform[3]. The Neo4j, in comparison to other state of art technologies (see Section 2. for a review on graph database systems), offers both a scalable solution enabling to manage large scale graphs and specific textual based indexing functionalities to better support natural language processing applications.
Other filed of applications and studies that might benefit from the availability of MKGDB are, for example:

- applications aimed at generating graph embeddings (Wang et al., 2017) where combining knowledge graphs can partially solve the sparsity problem;
- studies aimed at extracting or inducing faceted/multimodal domain knowledge graphs (Liu et al., 2019);
- studies devoted to the definition of novel benchmarks for the tasks of knowledge graph refinement (Paulheim, 2016), or taxonomy induction (Bordea et al., 2015) (Velardi et al., 2013);
- distributional and topological based methodologies for the enrichment of lexical resources (Biemann et al., 2018);
- empirical studies of graph algorithms applied to large scale real graphs.

The rest of this paper is organized as follows:

- Section 2. describes the state of the art on knowledge graphs and graph database technologies;
- Section 3. provides details about the MKGDB resource, and detailed statistics about the topology of the graph; and
- Section 4. summarizes the contributions of this paper.

2. Related work

Large-scale lexical Knowledge Bases: As described in (Camacho-Collados et al., 2018), the exploitation of knowledge bases in AI/NLP applications is a well established practice. In recent years, we observed many efforts on the construction of fully or partially human curated general purpose lexical knowledge bases - e.g., WordNet (Fellbaum, 1998) BabelNet (Navigli and Ponzetto, 2012), DBpedia (Lehmann et al., 2015), Yago (Mahdisoltani et al., 2015), ConceptNet (Speer et al., 2017) - and application specific knowledge bases - e.g., FrameNet (Baker et al.,

[1] Michael Galkin, Knowledge Graphs in Natural Language Processing @ ACL 2019, `https://towardsdatascience.com/knowledge-graph-bb78055a7884`
[2] at time of writing
[3] `https://neo4j.com/`

1998), SenticNet (Cambria et al., 2018). More recently, increasing attention has been paid to multi-modal knowledge graphs (Liu et al., 2019), where the knowledge is represented combining different media types.
The majority of the above mentioned resources are able to provide a human and machine readable general purpose multilingual knowledge representation, but they do not model commonsense and domain specific knowledge. To cope with the absence of domain/application specific information, other efforts are focusing on knowledge acquisition techniques to mine information from heterogeneous sources, and even from the entire Web.
Examples in this direction are: graph based approaches such as OntolearnReloaded (Velardi et al., 2013) or (Biemann et al., 2018) where the authors present a distributional semantics-based end-to-end framework for the enrichment of lexical semantic resources, or Probase (Wu et al., 2012) and WebIsADB (Seitner et al., 2016), where lexical syntactic patterns are used to mine hypernymy relations from Web-scale corpora.
The main problem of the above mining techniques is related to the acquisition of noisy or wrong information, which requires human supervision or additional algorithmic efforts to be identified and removed. In the above described context, the MKGDB resource, by combining both human curated and noisy hypernymy graphs, may facilitate the development of novel approaches dedicated to minimizing the human supervision required in current state of the art methodologies.

Knowledge Base management technologies: To deploy our resource, we analyzed several alternative graph database platforms, since other knowledge representation models/technologies, such as RDF stores (Faye et al., 2012), are mainly oriented to navigational, reasoning and interlinking purposes, while graph databases are scalable technologies enabling the development of graph algorithm-based applications.
As surveyed in (Patil et al., 2018), there are no industry standards for graph database technologies. We briefly review in this section standard-de-facto technologies able to manage graph data models. The reader can find an interesting comparative study in (Fernandes and Bernardino, 2018).
The graph database technologies we analyzed and candidated for the deployment of MKGDB are: i) Sparksee (formerly known as DEX), a lightweight fast and scalable graph database manager.[4]; ii) Neo4j (Lal, 2015), a transactional graph database manager able to handle large scale graphs; iii) Hyper GraphDB (Iordanov, 2010), an open source software with a similar architecture when compared to Neo4j and Sparksee. GraphDB is also able to handle hyper graphs data models (Levene and Poulovassilis, 1990), (Levene and Poulovassilis, 1991).[5] iv) with a particular attention to distributed computing capabilities, we mention also ArangoDB[6], Trinity (Shao et al., 2013), ThingSpan (formerly known as Infinite Graph)[7] and Titan[8].
Finally, we decided to adopt the graph database platform Neo4j for both the development and the deployment of MKGDB.
Our choice is based on four important main factors: i) the ability to efficiently handle and query extremely large graphs made of billions of nodes and relationship.[9]; ii) the Neo4j software is supported by a large and growing community of users and in comparison to others technologies, and therefore it was evaluated as the most promising in the medium and long terms; iii) Neo4j supports efficient indexing mechanisms that are well suited for large scale graphs where nodes are labelled with textual features (such as lexical knowledge graphs); iv) finally, Neo4j provides a specific query language called ”cypher” which is particularly effective (we provide examples of queries we designed to perform the statistical analysis in Section 3.3.) and easy to extend[10].

3. Resource

MKGDB is an hypernymy graph database including (at the time of writing) the backbone terminological taxonomies from five knowledge graphs: ConceptNet, DBpedia (both ontology and instances), WebIsAGraph, the Wikipedia hierarchy of categories and WordNet.

- ConceptNet (Speer et al., 2017) is the result of a project intended to provide a large semantic graph that describes general human knowledge and how it is expressed in natural language;
- DBpedia (Lehmann et al., 2015), is the result of a crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects. As reported from the project web site ”DBpedia describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology”;
- WordNet (Fellbaum, 1998) is a human expert curated lexical database that groups English words into sets of synonyms called synsets, and provides also a number of relations among these synonym sets (hypernymy and hyperonymy relations included);
- Wikipedia Categories hierarchy (Ponzetto and Strube, 2007), is a folksonomy built on the set of Wikipedia categories used for the classification of Wikipedia articles;
- WebIsAGraph (Faralli et al., 2019) is a large noisy hypernymy graph induced automatically by means of syntactic lexical pattern matches on the Common Crawl[11] Web corpus.

[4] http://www.sparsity-technologies.com/index#sparksee
[5] http://www.hypergraphdb.org/
[6] https://www.arangodb.com/
[7] https://www.objectivity.com/products/thingspan/thingspanfeatures/
[8] https://titan.thinkaurelius.com/
[9] https://neo4j.com/
[10] the Neo4j platform provides an interface for the development and inclusion of external plugins written in Java.
[11] https://commoncrawl.org/

# nodes	**37,095,451**
# edges	**81,229,350**
Avg Degree	4.1
Min Degree	1
Max Degree	166,289
self loops	1,122
cycles of length 2	3,208,394

Table 1: Structural statistics of MKGDB.

Future releases of MKGDB will include additional hypernymy graphs, for example derived from BabelNet (Navigli and Ponzetto, 2012) and Probase (Wu et al., 2012) (among the others).
Starting from an empty graph database, we created the resource by parsing the previously listed datasets.
To better describe our resource, in Section 3.1. we describe the graph data model, in Section 3.3. we provide an in-depth analysis of the resulting graph, and in Section 3.4. we provide additional information on how to integrate MKGDB in other systems' pipelines.

3.1. Graph Data Model

The resulting graph database model consists of nodes of type ":Term" and directed edges of type ":IsA". Both nodes and edges are decorated with a maximum of six properties (namely ConceptNet, DBPediaInstances, DBPediaOntology, WebIsAGraph, WikiCategories and WordNet) to indicate if a terminological node or hypernymy relationship belongs to a specific knowledge graph.

3.2. Source Datasets

In this Section, we describe the collection of datasets we combined in MKGDB. We extracted hypernymy relation of the form (t, h) where t is a term (e.g., "cat") and h a hypernym of t (e.g., "feline") from the following data sources:

- **ConceptNet:** from the release 5 of the knowledge base accessible from `http://conceptnet.io/`. Hypernymy pairs are extracted from the assertions of kind "/r/IsA".
- **DBpedia:** from the release 3.9 of the English version by parsing the dataset "instance_types_en.nt" accessible from `http://downloads.dbpedia.org/3.9/en/`, and "dbpedia_2016-10.nt" accessible from `https://wiki.dbpedia.org/downloads-2016-10`; We mine hypernymy relations from triples with predicates:
  `<http://www.w3.org/2000/01/rdf-schema#subClassOf>`
  `<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>`
  for the ontology hierarchy and for the instances' types respectively.
- **WebIsAGraph:** from the first release of the graph database accessible from `https://sites.google.com/unitelmasapienza.it/webisagraph/home` which is already in the form of graph database;
- **WikiCategories:** from the release 3.9 of the English version by parsing the dataset "skos_categories_en.nt" accessible from `http://downloads.dbpedia.org/3.9/en/` and mining hypernymy relations from the triples having the predicate: `<http://www.w3.org/2004/02/skos/core#broader>`
- **WordNet:** from the release 3.1 of the data base accessible from `https://wordnet.princeton.edu/download/current-version/`, we programmatically queried the database mining for hypernymy relations between noun synsets.

3.3. Statistics

To the best of our knowledge, MKGDB represents the first step towards the construction of the largest available hypernymy graph.[12] As shown in Table 1, our resource includes a total of 37,095,451 nodes and a total of 81,229,350 hypernymy relationships with an average node degree of 4.1. We also observed the presence of cycles, and in particular we counted about 1K self-loops and about 3M cycles of length 2. In the remaining of this Section, we provide an in-depth structural analysis focusing on: i) the distribution of shared nodes and edges across data sources (see Section 3.3.1.), and ii) the distribution of cross-link edges across different data sources (see Section 3.3.2.).

3.3.1. Shared nodes and edges

We report in Table 2 (diagonal), for each source knowledge graph, the total count of nodes included in MKGDB. Values are obtained, with simple *cypher*[13] queries of the form:

```
MATCH (n)
WHERE EXISTS(n.<KB>)
RETURN count(n);
```

where <KB> $\in$ {*ConceptNet*, *DBPediaInstances*, *DBPediaOntology*, *WebIsAGraph*, *WikiCategories*, *WordNet*}.
The remaining cells of Table 2 correspond to the total count of common nodes for each pair of source knowledge graphs. In this case, we computed the number of shared terminological nodes, with simple *cypher* queries of the form:

```
MATCH (n)
WHERE EXISTS(n.<KB1>)
AND EXISTS(n.<KB2>)
RETURN count(n);
```

where <KB_1>,<KB_1> $\in$ {*ConceptNet*, *DBPediaInstances*, *DBPediaOntology*, *WebIsAGraph*, *WikiCategories*, *WordNet*}, <KB_1> $\neq$ <KB_2>.
We observe that the majority of nodes (about 89%) are originated from the WebIsAGraph, which we recall is a large

[12] Next MKGDB releases will incrementally integrated others human curated and automatically acquired hypernymy graphs.
[13] cypher is the query language to query a neo4j graph `https://neo4j.com/developer/cypher-basics-i/`

KB_1 / KB_2	*ConceptNet*	*DBPediaInstances*	*DBPediaOntology*	*WebIsAGraph*	*WikiCategory*	*WordNet*
ConceptNet	**153,241 [0.413%]**	32,132 [0.087%]	381 [0.001%]	47,077 [0.127%]	10,851 [0.029%]	64,850 [0.175%]
DBPediaInstances	32,132 [0.087%]	**3,240,978 [8.7369%]**	406 [0.001%]	70,843 [0.191%]	46,699 [0.126%]	13,339 [0.036%]
DBPediaOntology	381 [0.001%]	406 [0.001%]	**767 [0.002%]**	422 [0.001%]	50 [0.000%]	378 [0.001%]
WebIsAGraph	47,077 [0.127%]	70,843 [0.191%]	422 [0.001%]	**33,030,457 [89.042%]**	16,400 [0.044%]	41,387 [0.112%]
WikiCategory	10,851 [0.029%]	46,699 [0.126%]	50 [0.000%]	16,400 [0.044%]	**837,327 [2.257%]**	5,920 [0.016%]
WordNet	64,850 [0.175%]	13,339 [0.036%]	378 [0.001%]	41,387 [0.112%]	5,920 [0.016%]	**104,404 [0.281%]**

Table 2: Total number of nodes by data source ($KB_1 = KB_2$ diagonal) and pairwise intersections ($KB_1 \neq KB_2$). Percentage values represent the ratio over the total number of nodes (i.e., 37,095,451)

KB_1 / KB_2	*ConceptNet*	*DBPediaInstances*	*DBPediaOntology*	*WebIsAGraph*	*WikiCategory*	*WordNet*
ConceptNet	**222,038 [0.273%]**	17,721 [0.022%]	82 [0,000%]	19,244[0.024%]	490 [0.001%]	73,130 [0.090%]
DBPediaInstances	17,721 [0.022%]	**13,304,353 [16.379%]**	18 [0.000%]	12,878 [0.000%]	5 [0.016%]	390 [0.000%]
DBPediaOntology	82 [0.000%]	18 [0.000%]	**769 [0.001%]**	199 [0.000%]	0 [0.000%]	40 [0.000%]
WebIsAGraph	19,244 [0.024%]	12,878[0.016%]	199 [0.000%]	**65,681,899 [80.859%]**	3,126 [0.004%]	18,219 [0.022%]
WikiCategory	490 [0.001%]	5 [0.000%]	0 [0.000%]	3,126 [0.004%]	**1,895,410 [2.333%]**	412 [0.000%]
WordNet	73,130 [0.090%]	390 [0.000%]	40 [0.000%]	18,219 [0.022%]	412 [0.000%]	**256,896 [0.316%]**

Table 3: Total number of edges by data source ($KB_1 = KB_2$ diagonal) and pairwise intersections ($KB_1 \neq KB_2$). Percentage values represent the ratio over the total number of edges (i.e., 81,229,350)

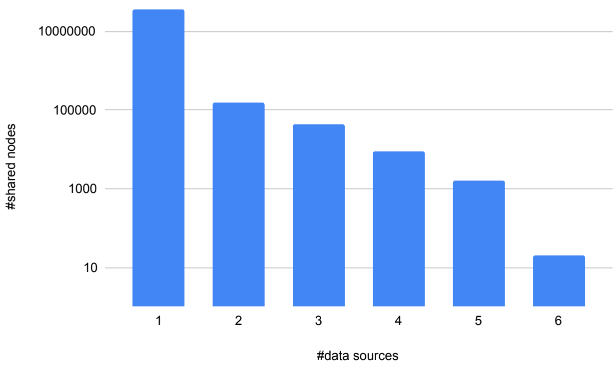

Figure 1: Distribution of the number of shared nodes over the number of data sources (e.g., 20 nodes are shared by 6 data sources).

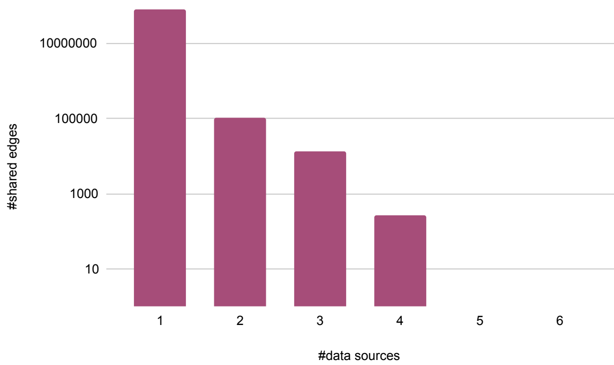

Figure 2: Distribution of the number of shared edges over the number of data sources (e.g., 268 hypernymy relations are shared by 4 data sources).

scale noisy hypernymy graph, and the second major contribution (8.7%) is due by the instance nodes of DBpedia.
In Table 3 with similar simple *cypher* queries of the form:

```
MATCH ()-[r]->()
WHERE EXISTS(r.<KB>)
RETURN count(r);

MATCH ()-[r]->()
WHERE EXISTS(r.<KB1>)
AND EXISTS(r.<KB2>)
RETURN count(r);
```

we report the number of total hypernymy relations derived from each source dataset and shared by pair of knowledge graphs. We observed that similarly to what happened for the nodes, the majority of edges comes from WebIsAGraph (81%) and DBpedia (16%).
In Figures 1 and 2, we show two distributions: the first histogram describes the number of shared nodes over the number of sharing data sources and the second describes the number of shared edges over the number of sharing data sources.
As reported in Section 3., WebIsAGraph is the largest lexical KB in MKGDB, but due to its noisy nature, it also includes many wrong hypernymy links. A subset of reliable relationships can be derived by simply analyzing the overlapping among this resource and the others, as shown in Figures 3 and 4. Figure 3 shows the distribution of nodes (lexical entities) in WebIsAGraph that are also found in at least one, two, or more other KBs, while Figure 4 shows the same information for shared hypernymy links. In general, matching error-prone and error-free KBs can be used to identify a backbone of reliable relationships in noisy resources, that can be used as a basis to assess the quality of other unverified relationships (e.g., exploiting lexical chains (Wei et al., 2015)).

3.3.2. Cross-link edges

In Table 4 we report the number of cross-link edges. Cross-link edges are defined as those edges connecting different knowledge graphs. Such kind of edges are important in all those applications where it is relevant to traverse multiple knowledge graphs.

KB_1 / KB_2	*ConceptNet*	*DBPediaInstances*	*DBPediaOntology*	*WebIsAGraph*	*WikiCategory*	*WordNet*
ConceptNet		9,610,269 [11.831%]	3,065 [0.004%]	10,333,708[12.722%]	457,911 [0.564%]	228,888[0.282%]
DBPediaInstances	561,332 [0.691]		15,825 [0.0195%]	2296617 [2.827%]	278,165 [0.342%]	474,818 [0.584%]
DBPediaOntology	460,176 [0.566%]	12,643,395 [15.565%]		1,701,140 [2.094%]	330,505 [0.407%]	318,588 [0.392%]
WebIsAGraph	145,805 [0.179%]	9,801,764 [12.067%]	217 [0.000%]		214,961 [0.265%]	100,906 [0.124%]
WikiCategory	883,762 [1.088%]	567,077 [0.698%]	60,942 [0.075%]	4,373,623 [5.384%]		778,388 [0.958%]
WordNet	852,673 [1.049%]	9,844,036 [11.352%]	2,843 [0.004%]	9,220,797 [11.352%]	721,615 [0.888%]	

Table 4: Total number of cross-link hypernymy edges of the form (t, h) where $h \in KB_1, t \in KB_2, h \notin KB_2$, $(KB_1 \neq KB_2)$. Percentages values represent ratio over the total number of edges (i.e., 81,229,350)

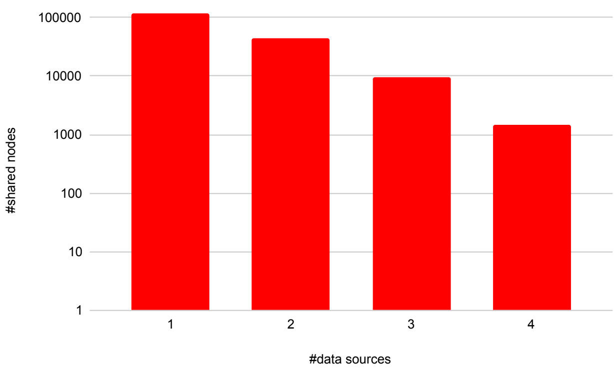

Figure 3: Distribution of the number of WebIsAGraph shared nodes over the number of other sharing data sources, e.g., 9,787 WebIsAGraph nodes are confirmed by other 3 resources (DBPediaOntology and DBPediaInstances are combined as a one single data source).

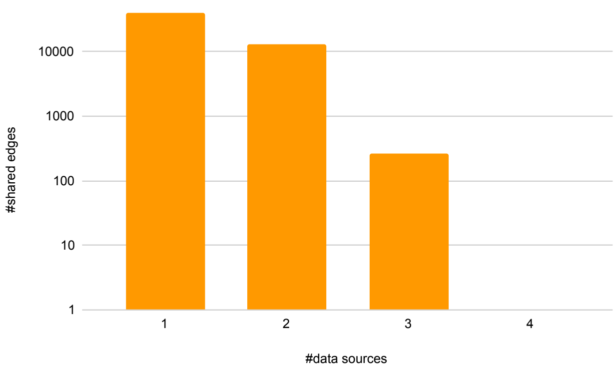

Figure 4: Distribution of the number of WebIsAGraph shared edges over the number of other sharing data sources, e.g., 1,480 WebIsAGraph edges are confirmed by other 4 resources (DBPediaOntology and DBPediaInstances are combined as a one single data source) .

Also in this case, values are calculated with simple *cypher* queries of the form:

```
MATCH (n)-[r]->(m)
WHERE EXISTS(n.<KB1>)
AND NOT EXISTS(m.<KB1>)
AND EXISTS(m.<KB2>)
RETURN count(r);
```

We observed that thanks to our resource we can count a significant number of cross-link edges connecting pairs of data sources, e.g., around the 12% of the total number of edges connect nodes from ConceptNet to instance nodes of Dbpedia.

As an example of usage of cross-links, ContrastMedium (Faralli et al., 2017) is a graph pruning approach which drives the pruning of noisy automatically acquired hypernymy graphs by transferring, hence traversing cross-link edges (see the example in Figure 5), topologically derived node features from a ground truth knowledge graph. Figure 6 shows an excerpt of the graph resulting from the simple *cypher* query:

```
MATCH (n)-[r]->(m)
WHERE  EXISTS(r.ConceptNet)
AND EXISTS(m.WordNet)
AND NOT EXISTS(m.DBPediaOntology)
AND EXISTS(n.DBPediaOntology)
AND NOT EXISTS(n.WordNet)
RETURN *;
```

Where thanks to nature of MKGDB and ”by triangulation”, one is able to discover on a ”third” knowledge graph (e.g.,ConceptNet) cross-link edges connecting nodes from a source (e.g.,DBPediaOntology) to a target (e.g., WordNet) knowledge graph.

3.4. Resource Availability

MKGDB and all the related material (data and source code) are publicly available under a CC BY 4.0 license at `https://github.com/FaridYusifli/MKGDB`.

4. Conclusions

We presented MKGDB, a resource in the form of a graph database merging multiple hypernymy graphs. Differently from other interlinked resources and knowledge representation models, MKGDB enables and facilitates the application of graph-based algorithms on very large knowledge graphs. Furthermore, MKGDB is useful for applications where it is important to identify noise-free hypernymy relationships in error-prone lexical databases, as well as cross-link edges across noisy and error-prone hypernymy graphs. Thanks to both the versatility and the scalability of graph database managers such as *Neo4j*, we plan to enrich our resource by including more data sources and more properties to nodes and edges, including features to help the generation of hybrid topological and distributional embedded representations.

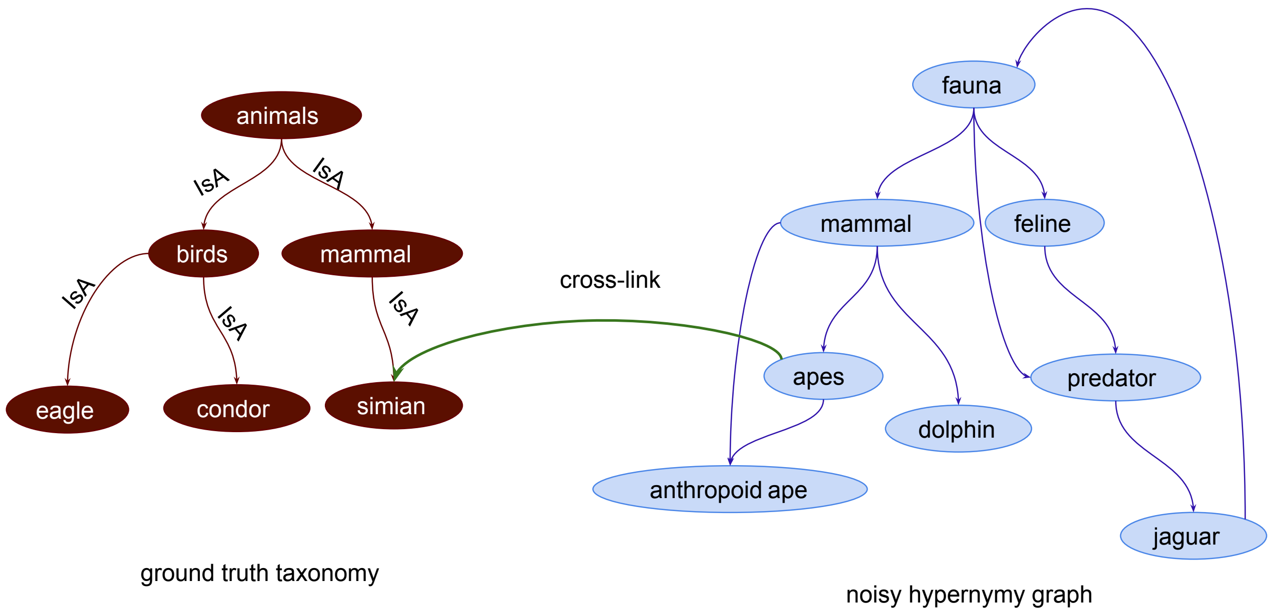

Figure 5: Example of a cross-link edge between a noisy hypernymy graph (left) and a ground truth taxonomy (right). Algorithms, such as "ContrastMedium" (Faralli et al., 2017), use to first compute topological-based features on a ground truth graph, and second, to "transpose" such features to a noisy taxonomic structure, to define a traversal strategy on top of which perform some graph processing (e.g., breaking cycles).

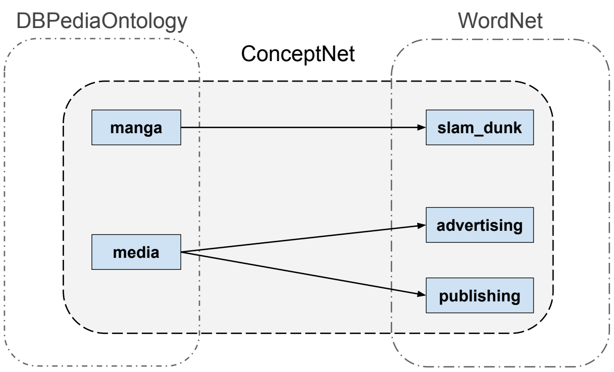

Figure 6: Excerpt of cross-link edges starting form DB-pedia ontology nodes and ending to WordNet nodes, and belonging to ConceptNet knowledge graph.

5. Acknowledgements

This work has been supported by the MIUR (Ministery of Instruction, University and Research) Project SCN_0166 SMARTOUR.

6. Bibliographical References

Allen, J. F. and Frisch, A. M. (1982). What's in a semantic network? In *20th Annual Meeting of the Association for Computational Linguistics, University of Toronto, Toronto, Ontario, Canada, June 16-18, 1982*, pages 19–27.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL '98, August 10-14, 1998, Université de Montréal, Montréal, Quebec, Canada. Proceedings of the Conference*, pages 86–90.

Biemann, C., Faralli, S., Panchenko, A., and Ponzetto, S. P. (2018). A framework for enriching lexical semantic resources with distributional semantics. *Natural Language Engineering*, 24(2):265–312.

Bonatti, P. A., Decker, S., Polleres, A., and Presutti, V. (2018). Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371). *Dagstuhl Reports*, 8(9):29–111.

Bordea, G., Buitelaar, P., Faralli, S., and Navigli, R. (2015). Semeval-2015 task 17: Taxonomy extraction evaluation (texeval). In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 902–910.

Camacho-Collados, J., Espinosa Anke, L., and Pilehvar, M. T. (2018). The interplay between lexical resources and natural language processing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 17–23, New Orleans, Louisiana, June. Association for Computational Linguistics.

Cambria, E., Poria, S., Hazarika, D., and Kwok, K. (2018). Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1795–1802.

Faralli, S., Panchenko, A., Biemann, C., and Ponzetto, S. P. (2017). The contrastmedium algorithm: Taxonomy induction from noisy knowledge graphs with just A few links. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*, pages 590–600.

Faralli, S., Finocchi, I., Ponzetto, S. P., and Velardi, P. (2019). Webisagraph: A very large hypernymy graph from a web corpus. In *Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy, November 13-15, 2019*.

Faye, D. C., Curé, O., and Blin, G. (2012). A survey of RDF storage approaches. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 15:11–35.

Fellbaum, C. (1998). A semantic network of english: The mother of all wordnets. *Computers and the Humanities*, 32(2):209–220, Mar.

Fernandes, D. and Bernardino, J. (2018). Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb. In *Proceedings of the 7th International Conference on Data Science, Technology and Applications, DATA 2018, Porto, Portugal, July 26-28, 2018*, pages 373–380.

Iordanov, B. (2010). Hypergraphdb: A generalized graph database. In Heng Tao Shen, et al., editors, *Web-Age Information Management*, pages 25–36, Berlin, Heidelberg. Springer Berlin Heidelberg.

Lal, M. (2015). *Neo4J Graph Data Modeling*. Packt Publishing.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C. (2015). DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 6(2):167–195.

Levene, M. and Poulovassilis, A. (1990). The hypernode model and its associated query language. In *Proceedings of the 5th Jerusalem Conference on Information Technology, 1990. 'Next Decade in Information Technology'*, pages 520–530, Oct.

Levene, M. and Poulovassilis, A. (1991). An object-oriented data model formalised through hypergraphs. *Data & Knowledge Engineering*, 6(3):205 – 224.

Liu, Y., Li, H., Garcia-Duran, A., Niepert, M., Onoro-Rubio, D., and Rosenblum, D. S. (2019). Mmkg: Multi-modal knowledge graphs. In Pascal Hitzler, et al., editors, *The Semantic Web*, pages 459–474, Cham. Springer International Publishing.

Mahdisoltani, F., Biega, J., and Suchanek, F. M. (2015). YAGO3: A knowledge base from multilingual wikipedias. In *CIDR 2015, Seventh Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 4-7, 2015, Online Proceedings*.

Navigli, R. and Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

P, D. and Jurek-Loughrey, A. (2018). *Linking and Mining Heterogeneous and Multi-view Data*. Unsupervised and Semi-Supervised Learning. Springer International Publishing.

Patil, N., Kiran, P., Kavya, N., and Patel, K. (2018). A survey on graph database management techniques for huge unstructured data. *International Journal of Electrical and Computer Engineering*, 81:1140–1149, 04.

Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8:489–508.

Ponzetto, S. P. and Strube, M. (2007). Deriving a large scale taxonomy from wikipedia. In *Proceedings of the 22Nd National Conference on Artificial Intelligence - Volume 2*, AAAI'07, pages 1440–1445. AAAI Press.

Seitner, J., Bizer, C., Eckert, K., Faralli, S., Meusel, R., Paulheim, H., and Ponzetto, S. P. (2016). A large database of hypernymy relations extracted from the web. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*.

Shadbolt, N., Berners-Lee, T., and Hall, W. (2006). The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101.

Shao, B., Wang, H., and Li, Y. (2013). Trinity: a distributed graph engine on a memory cloud. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2013, New York, NY, USA, June 22-27, 2013*, pages 505–516.

Speer, R., Chin, J., and Havasi, C. (2017). Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pages 4444–4451. AAAI Press.

Trisedya, B. D., Qi, J., and Zhang, R. (2019). Entity alignment between knowledge graphs using attribute embeddings. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 297–304.

Velardi, P., Faralli, S., and Navigli, R. (2013). Ontolearn reloaded: A graph-based algorithm for taxonomy induction. *Computational Linguistics*, 39(3):665–707.

Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, Dec.

Wei, T., Lu, Y., Chang, H., Zhou, Q., and Bao, X. (2015). A semantic approach for text clustering using wordnet and lexical chains. *Expert Syst. Appl.*, 42(4):2264–2275, March.

Wu, W., Li, H., Wang, H., and Zhu, K. Q. (2012). Probase: A probabilistic taxonomy for text understanding. In *ACM International Conference on Management of Data (SIGMOD)*, May.