Metaresearch Recommendations using Knowledge Graph Embeddings

Veronika Henk,[1] Sahar Vahdati,[1] Mojataba Nayyeri,[1] Mehdi Ali,[1] Hamed Shariat Yazdi,[1] Jens Lehmann[1,2]
[1]Smart Data Analytics Group, University of Bonn
[2]Enterprise Information Systems Department, Fraunhofer IAIS
familyname@cs.uni-bonn.de

Abstract

Discovering relevant research collaborations is crucial for performing extraordinary research and promoting the careers of scholars. Therefore, building recommender systems capable of suggesting relevant collaboration opportunities is of huge interest. Most of the existing approaches for collaboration and co-author recommendation focus on semantic similarities using bibliographic metadata such as publication counts, and citation network analysis. These approaches neglect relevant and important metadata information such as author affiliation and conferences attended, affecting the quality of the recommendations. To overcome these drawbacks, we formulate the task of scholarly recommendation as a link prediction task based on knowledge graph embeddings. A knowledge graph containing scholarly metadata is created and enriched with textual descriptions. We tested the quality of the recommendations based on the TransE, TransH, TransR and DistMult models that consider only triples in the knowledge graph and DKRL which in addition incorporates natural language descriptions of entities during training.

Introduction

Research is becoming increasingly digital, interdisciplinary, and data-driven and affects different environments in addition to academia, such as industry, and government. Research output representation, publication, mining, analysis, and visualization are taken to a new level, driven by the increased use of Web standards and digital scholarly communication initiatives. The number of scientific publications produced by new players and the increasing digital availability of scholarly artifacts, and associated metadata are other drivers of the substantial growth in scholarly communication. Assisting researchers with a deeper analysis of scholarly metadata and providing recommendations can lead to new opportunities in research. Especially, discovery and recommendation about potential collaborations between researchers can lead to new ways of conducting research.

Most of the techniques use semantic similarities and graph clustering approaches. Thus, the predicted items for recommendations are those which are similar to the items clearly specified. However, this limits the recommendations to user profiles only which leaves a lot of other available information unused. Approaches for knowledge extraction from huge networks by uncovering patterns and predicting emergent properties of the network can facilitate link prediction activities. Link prediction using *knowledge graph embbedings* (KGEs) received strong interest in the last years. The idea behind KGEs is to represent entities and relations of a knowledge graph (KG) into a low dimensional vector space. These approaches can be roughly divided into *translational distance models* and *semantic matching models*, whereas the former predicts the plausibility of a link between entities by means of a distance-based scoring function, and the latter based on a similarity-based scoring function (Wang et al. 2017). An established distance-based model is *TransE* (Bordes et al. 2013) that interprets a relation/link as the translation from the head to the tail entity. The *TransH* model (Wang et al. 2014) which is an extension of TransE aiming at handling certain relation types, such as reflexive relations, better is also considered in the evaluation set of this research. Another model called *TransR* (Lin et al. 2015) focuses on various relations and different aspects of entities. One prominent semantic matching model is *DistMult* (Yang et al. 2014) that encodes each relation as a diagonal matrix and considers pairwise interactions of the latent features of the entity and relation representations to compute the plausibility of facts.

Recently, KGE models were proposed that incorporate additional sources of information such as text and logical rules instead of considering only the triples of the KG. Description-Embodied Knowledge Representation Learning (DKRL) (Xie et al. 2016) is a model that includes textual descriptions of entities while learning KGEs (Wang et al. 2017). In this work, we employ already existing embedding models in order to provide metaresearch recommendation. To the best of our knowledge, our work is the first attempt in creating such a scholarly knowledge graph and using embedding models in the scholarly communication domain. Knowledge graphs are suitable for capturing complex structures beyond simple author profiles. Therefore, we believe that KGEs can serve well in the domain of scholarly communication in which artifacts and metadata are heterogeneous and often spread over different sources. For example, it becomes difficult to keep track of relevant scientific results, to stay up-to-date with new and relevant scientific events and running projects, as well as to find potential future collabo-

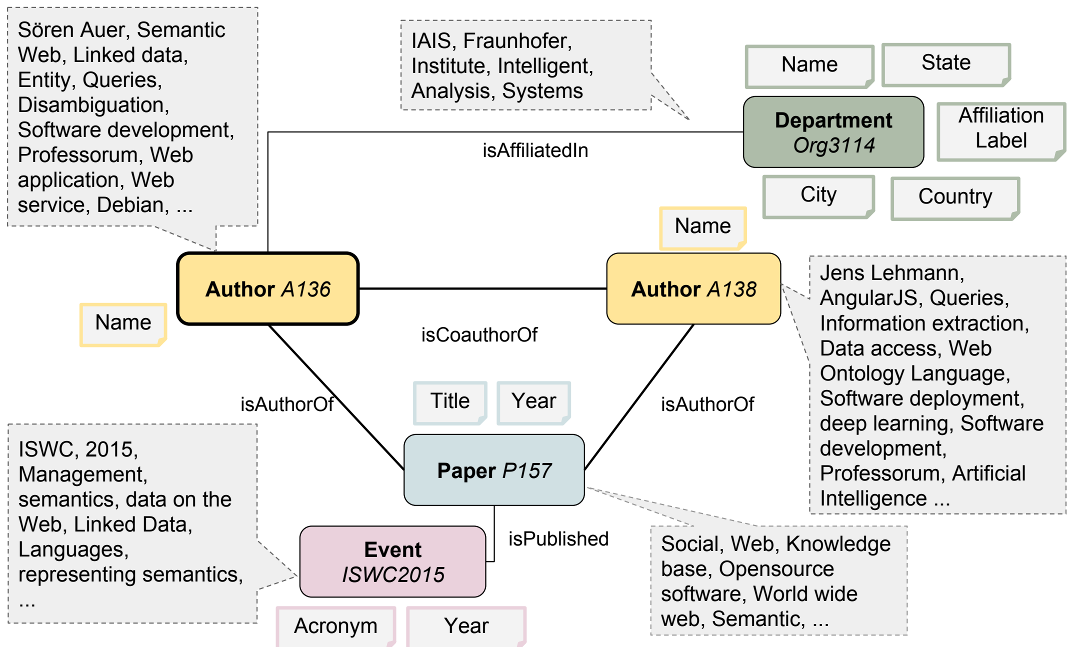

Figure 1: **Scholarly entities and relations in SG4MR.** SG4MR contains both structured metadata and unstructured text description of the entities created from keywords.

rators. Thus, assisting researchers with analytics over scholarly metadata can lead to new opportunities in research and to new ways of conducting research. However, due to the information overload and variety of artifacts being published daily in scholarly communication, it is a challenging task to generate recommendations of the relevant information for scholars. The aim of this paper is to generate a new knowledge graph for the scholarly community and use it for co-authorship recommendation. In the creation of the KG, we consider both structural descriptions as well as natural language descriptions and, therefore, unite recommender systems and natural language processing techniques. Specifically, DKRL as a recent work in KGE which can include text in its formulation to improve the result of KGE is applied to the generated KG for co-authorship recommendation.

Knowledge Graph Creation

In order to prototype the idea of using KGEs for co-author recommendation (along with similar tasks such as venue recommendation), a conceptualization of the scholarly communication domain has been done to create a knowledge graph for metaresearch recommendations (SG4MR). The data acquisition for *SG4MR* started with a systematic study over the existing information spaces containing scholarly metadata. Four different resources have been selected as the main sources of metadata acquisition: DBLP[1], Semantic Scholar[2], Springer Nature SciGraph Explorer[3] and Global Research Identifier Database [4]. Figure 1 shows a portion of the created scholarly artifacts describing the entities and relations between them. *SG4MR* includes structured metadata about a list of selected core entities and the relations between them. To facilitate using KGE models, textual descriptions of the core entities have been added to the knowledge graph. The following sections provide a detailed description of SG4MR and its components.

Core Entities and Relations

Based on the target recommendations, the core entities of the knowledge graph have been initially identified with the structured metadata as:

- **Authors:** of the scientific papers are indicated with their *names*. A post processing step has been applied over the collected data for deduplication and disambiguation of the author names. A *co-author* relationship between two persons exists if they have at least one joint paper.
- **Papers:** are the scientific results which are represented in textual descriptions authored by researchers. The structured metadata of papers, which we incorporated are the *title* of the paper, the *year of publication* and the authors (encoded by the *isAuthor* relation in SG4MR).
- **Departments:** are the organizations to which the authors of the scientific papers are affiliated. The representative properties of the department entities are *name*, *city*, *state* and *country*. *isAffiliatedIn* is the relation between an author entity and the department which is indicated inside in any specific scientific paper of the author.

[1] `https://dblp.uni-trier.de`
[2] `https://www.semanticscholar.org`
[3] `https://scigraph.springernature.com`
[4] `https://www.grid.ac`

- **Conferences:** are the hosting events of the scientific papers and have *acronyms* and *year* metadata. The venue of a paper in which it *isPublished* is described with a *full name* and *acronym*.

In order to enrich the graph,extra metadata such as year, acronym, state have been added. With a focus on Computer Science and the Semantic Web community, the corresponding metadata associated with the core entities was initially extracted from DBLP (the largest bibliographic database of scholarly papers for this domain). The metadata about affiliated departments is indicated inside the papers and corresponds to the time of publishing. To gain additional information associated with authors, affiliation information was extracted from the Springer Nature SciGraph Explorer (SG). We retrieved affiliations for *3,718* of the *4,495* authors in the knowledge graph. Some of the affiliation information contained in the SciGraph dataset has been identified with references to GRID data. The corresponding information about the affiliation of the researchers acquired from these two sources refers to research organizations, departments, and institutions. For affiliations referencing an entity in the GRID dataset, *name* and *location* information were acquired. Figure 2 indicates the number of instances per core entities.

Natural Language Descriptions

Additional metadata corresponding to publications was acquired from Semantic Scholar (S2) and filtered to extract only data related to 15 different semantic web related conferences. Using the filtered data, keywords for the papers in the knowledge graph were retrieved.

The natural language descriptions used in our approach provide additional information for various entity types. The descriptions for paper entities consist of the papers title, the year it was published and the keywords extracted from S2. The keywords for the author entities are composed of the authors name and the set of all keywords for all papers published by this author. For affiliation entities, we retrieved either the label from S2 or the entities name and location (GRID). As the label stored in S2 contains name and address in most cases, only one of these sources was used for each affiliation to avoid the acquisition of redundant information. For each conference event, keywords are composed of the events acronym and the year it took place. This information was retrieved from DBLP. Furthermore, the topics of interest for each event were acquired from the corresponding website in a manual step. Finally, in a pre-processing step stop-words were removed from the sets of keywords. The knowledge graph created in this research and corresponding evaluation results have been made publicly available [5].

Knowledge Graph Embedding Models

In this article, we are primarily concerned with link prediction using KGE models. Link prediction is defined as the task of deciding whether a fact (represented by a triple) is true or false given a KG. More formally, let $\mathcal{E} = \{e_1, \cdots, e_{N_e}\}$ be the set of entities, $\mathcal{R} = \{r_1, \cdots, r_{N_r}\}$ be the set of relations connecting two entities, $\mathcal{D} = \{d_1, \cdots, d_{N_d}\}$ be the set of relations connecting an entity and a literal, i.e., the data relations, and $\mathcal{L}$ be the set of all literal values. A knowledge graph $\mathcal{G}$ is a subset of $(\mathcal{E} \times \mathcal{R} \times \mathcal{E}) \cup (\mathcal{E} \times \mathcal{D} \times \mathcal{L})$ representing the facts that are assumed to hold. Link prediction can be formulated by a function $\psi : \mathcal{E} \times \mathcal{R} \times \mathcal{E} \rightarrow R$ mapping each possible triple $(e_i, e_j, r_k) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ to a score value, where a higher value implies the triple is more likely to be true.

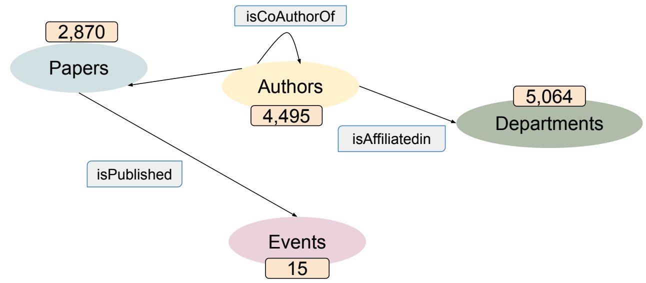

Figure 2: **Number of instances corresponding to the core entities.** A simple view of the knowledge graph is shown with entities and relations and the number of the instances of the core entities.

Most of the earlier embedding models only use structural information contained in a KG, i.e. ignore literal values such as natural language descriptions. However, recently some of works are proposed that can take advantage of textual descriptions of entities for computing embeddings. In this subsection, first, the TransE, TransH and DistMult embedding models and their formulations are reviewed. Second, DKLR as a recent work that uses both structural information and textual description for embedding is reviewed in detail.

The **TransE** transitional embedding model is one of the earlier works in KGE. Assume that (h, r, t) is a triple in a KG where (h, t) refer to the head and tail entities and r is relation between them. $(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t})$ denotes the embedding vectors of head entity, relation and tail entity. TransE uses the relation vector $(\boldsymbol{r})$ to transform the head entity vector $(\boldsymbol{h})$ to the tail entity vector $(\boldsymbol{t})$. Mathematically, the following equation should hold for each triple (h, r, t):

$$\boldsymbol{h} + \boldsymbol{r} \approx \boldsymbol{t} \tag{1}$$

The score function in embedding methods receives a triple and gives the degree of correctness of it. In order to follow Equation 1, TransE uses the following score function:

$$f_r(h, t) = -\|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\|. \tag{2}$$

If for a triple (h, r, t) the score of TransE i.e., $f_r(h, t)$ is closer to zero, i.e. below a threshold, the triple is considered as true. For the recommendation application, the score function is used for ranking the top model suggestion.

The TransE model initializes the entity and relation vectors randomly by a probability distribution function and op-

[5] `https://github.com/SmartDataAnalytics/OpenResearch`

timizes the following margin ranking loss function to obtain embedding vectors:

$$L = \sum (f_r(h^{'}, t^{'}) - f_r(h, t) + \lambda)_+, \tag{3}$$

where $f_r(h^{'}, t^{'})$ is the score of corrupted triple (the triple which is generated by corruption of head or tail entity of positive triple).

TransH stands for Translating on Hyperplanes. It is a modified version of *TransE* and deals with embedding in large scale knowledge graphs by translating on hyperplanes. TransE has limitations in modeling 1-N, N-1 and N-M relations, which is addressed in *TransH* (Wang et al. 2014). To overcome the limitations of TransE, TransH represents additionally each relation as a hyper-plane and performs the translation from the head to tail entity in the the relation specific hyper-plane (Wang et al. 2014):

$$\boldsymbol{h}_\perp = \boldsymbol{h} - \boldsymbol{w_r}^\top \boldsymbol{h} \boldsymbol{w_r} \tag{4}$$

$$\boldsymbol{t}_\perp = \boldsymbol{t} - \boldsymbol{w_r}^\top \boldsymbol{t} \boldsymbol{w_r}, \tag{5}$$

where $\boldsymbol{w_r}$ is the normal vector of the relation specific hyper-plane. The projected entities are then used to compute the score for the triple *(h,r,t)*:

$$f_r(h, t) = -\|\boldsymbol{h}_\perp + \boldsymbol{d_r} - \boldsymbol{t}_\perp\|_2^2, \tag{6}$$

where $\boldsymbol{d_r}$ is the relation specific translation vector in the relation specific hyper-plane.

DistMult Some of the embedding models focus on capturing the relational semantics and the composition of relations as characterized by matrix multiplication (Yang et al. 2014). DistMult considers learning representations of entities and relations within the knowledge graphs. Each relation is represented as a diagonal matrix $diag(\boldsymbol{r})$, and scores of triples are computed by considering pairwise interactions of the latent features of the entity and relation representations: (Wang et al. 2017):

$$f_r(h, t) = \boldsymbol{h}^\top diag(\boldsymbol{r}) \boldsymbol{t} \tag{7}$$

DKRL: In contrast to most of existing embedding models in which each entity in a KG has one low dimensional vector representation obtained based on structural knowledge included in KG, DKRL takes advantages textual descriptions of entities in KG together with structural information. Therefore, each entity in a KG has two representations: a structural representation $(\boldsymbol{h_S}, \boldsymbol{t_S})$ and a natural language description based representation $(\boldsymbol{h_D}, \boldsymbol{t_D})$.

DKRL defines a joint energy function for structural and textual information as follows:

$$E = -(E_D + E_S), \tag{8}$$

where E is the total energy function, E_S is the structured based energy function and E_D isthe description based energy function defined as follows:

$$\begin{cases} E_S = \|\boldsymbol{h_S} + \boldsymbol{r} - \boldsymbol{t_S}\|, \\ E_D = E_{DD} + E_{SD} + E_{DS}, \\ E_{DD} = \|\boldsymbol{h_D} + \boldsymbol{r} - \boldsymbol{t_D}\|, \\ E_{SD} = \|\boldsymbol{h_S} + \boldsymbol{r} - \boldsymbol{t_D}\|, \\ E_{DS} = \|\boldsymbol{h_D} + \boldsymbol{r} - \boldsymbol{t_S}\|. \end{cases} \tag{9}$$

To obtain description based representation of entities $(\boldsymbol{h_D}, \boldsymbol{t_D})$, DKRL uses a convolutional neural network which receives a vector representation of a set of words of entities which are obtained by concatenation of all vectors of words. The word vectors are obtained by *Word2vec* method.

The margin ranking loss is used for optimization of parameters of the model (embedding of entities and relations, network parameters).

Relation Name	Train Dataset	Validation Dataset	Test Dataset
Collaboration	12,711	651	1,953
Publication	8,670	438	1,264
Affiliation	12,143	588	1,770
Venue	6,428	323	1,013
Sum	**39,952**	**2,000**	**6,000**

Table 1: **Dataset Statistics.** The number of triples that are used in different datasets are shown per each relationship.

Experimental Setup

In total, the created KG comprises 45,952 triples as shown in Table 1. The triples were split into a training-set containing 39,952 triples and a test-set with 6,000 triples. Finally, the validation-set was generated by randomly selecting and extracting 2,000 triples from the training-set. Different embedding models have been trained in order to provide general recommendation for all entity types and relations in the knowledge graph. However, an example of such recommendations is shown for co-authorship recommendations using those models which were described in the previous section.

We separate our data into three parts; training, testing and validation. For each test triple, we corrupt head and then tail and replace by all other authors (entities which are authors). Then we report the mean rank and Hit@10 of different models. Finally, predictions were retrieved using only the co-author relation as the main goal of this paper is co-authorship recommendation. To train and evaluate the TransE, TransH, TransR and DistMult models we used the toolkit PyKEEN (Ali et al. 2018). All the dataset types are created for general recommendations.

Experimental Results

Different types of recommendations for scholarly community (co-author recommendation for future collaboration, event recommendation for future attending, etc.) can be done by generating a scholarly knowledge graph, enriched by textual descriptions for entities, and using knowledge graph

Setting	Mean Rank Raw	Mean Rank Filtered	Hits @ 10 Raw	Hits @ 10 Filtered
DKRL	-	1893	-	25% [6]
DistMult	2157.2	2046.72	11.1 %	13.25 %
TransE	693.03	**647.42**	44.53 %	**50.72** %
TransH	974.88	985.16	18 %	21.7 %
TransR	1258.36	1200.35	23.52 %	28.09 %

Table 2: **Experimental Results.** The results of the experimental setup.

embedding models that can take advantages of textual descriptions of entities. Recommendation can be done by the entity ranking obtained from score function of embedding models.

The remain of this section is as follows: First the results of structural based embedding models such as TransE, TransH, TransR, DistMult and textual structural based model such as DKRL are reported in the Table 2. Mean Rank and Hit@10 are two main evaluation metrics which are reported in this paper. The specification of these models are reported in the previous section. The second subsection reports some very interesting example results on co-authorship recommendation obtained from the TransE model.

Results of the Embedding Models

In this part, the results of TransE, TransH, TransR and Distmult as structural based embedding and DKRL as textual and structural based embeddings on scholarly knowledge graph are reported. The generated data is divided into training and testing parts. Mean Rank and Hit@10 are are used for evaluation. To calculate Mean Rank, as mentioned previously, we corrupt head and then tail entities and find the rank of the correct triple. The average of ranking over testing samples are reported as Mean Rank. Similar to mean rank calculation, we corrupt head and tail entities in the testing sample. If each positive test triple is ranked less that 11, then the counter increases by one. The average of the final results are reported as final Hit@10. The Table 2 reports the results of this part.

One unexpected result from the table Table 2 is that the structural based embeddings such as TransE outperform DKRL which is the textual and structural based embedding. The reason may be related to the formulation of DKRL in which the loss function has many local unwanted solutions. One possible unwanted local solution of DKRL is as follows:

$$\begin{cases} \boldsymbol{h_D} + \boldsymbol{r} \approx \boldsymbol{h_S}, \\ \boldsymbol{h_D} + \boldsymbol{r} \approx \boldsymbol{h_D}, \\ \boldsymbol{h_S} + \boldsymbol{r} \approx \boldsymbol{h_S}, \\ \boldsymbol{h_S} + \boldsymbol{r} \approx \boldsymbol{h_D}, \end{cases} \tag{10}$$

In this case, the structured and description based representations for entities would be close; consequently the model cannot capture the textual information well.

Author	#Recom.	Rank of Recom.
A136	7	1, 6, 12, 14, 21, 25, 35
A88	3	2, 10, 30
A816	5	3, 9, 13, 20, 28
A1437	8	4, 5, 7, 15, 16, 18, 23, 24
A138	4	8, 11, 22, 26
A128	2	17, 27
A295	1	19
A940	6	29, 31, 32, 33, 34, 36

Table 3: **Co-authorship Recommendations.** The rank links of discovered potential co-authorship for 8 (9 - 1 by removing the symmetric predictions and considering the highest ranked recommendation) selected researchers.

An Example of Metaresearch Recommendation

We used PyKEEN to create and rank triples of the form (author1, coAuthor, author2) representing co-author recommendations. The results represented here are focused on TransE because of its high performance on this data. In our experiments we considered 9 researchers associated with the Linked Data and Information Retrieval communities (Vahdati et al. 2018) as a foundation for the recommendations. After computing the recommendations, we applied a post-processing step to filter out reflexive triples as well as the symmetric triples. The list of recommendations triples is ranked from highest ones. This resulted in a list of all possible predictions for 8 of the authors because for one of the authors (A976), all the other symmetric predictions have been ranked higher than the predictions to A976. After filtering existing co-authorships from the 81 possible co-authorships, the remaining ranked list contains 36 new recommendations.

Table 3 summarizes the results of the recommendations for each of the candidate authors. It shows the number of recommendations for each researcher, rank of the recommendations (1 being highest rank), and the score of highest rank and lowest rank of the recommendation. The list in the table is ordered descending using the predicted scores. Overall, 8 is the highest number of recommendations for Author with id A1437. The second highest number of predictions are for A136 where their rank is in the range of 1 to 35. The third rank is placed for A940 with 6 recommendations and the fourth ranked predication is the the discovered relation for Author A816. A138 gets 4 recommendations with the best prediction in rank 8. For authors A128 and A295, there are 2 and 1 recommendations discovered. Looking at the author profiles and the data in the KG, there has never been any collaboration between them but the potential is very high considering the research topics of interest for these two researchers. Overall, the results validate the objective of using embedding models for metaresearch recommendations on the example of co-authorship.

Related Work

The usage of the knowledge graphs for recommendation systems attracted an increasing interest in Semantic Web applications. (Passant 2010) proposes a music recommendation system which is built upon the computation of the semantic distance of the entities within a knowledge graph. The DBpedia [7] knowledge base is used in (Cheekula et al. 2015) to identify entities as recommendations. In another work (Vahdati et al. 2018), we introduced KORONA a platform to uncover scholarly networks inside a knowledge graph using semantic similarities. With only focused on a simple knowledge graph without consideration of the additional information such as the textual description of the entities. In recent years, knowledge graph embedding models are increasingly used in building the recommendation systems (Cheng et al. 2016). (Ortega et al. 2018) provides a collaborative filtering dataset containing scientific documentation. It uses matrix factorization methods to implement recommendations for scholars. Sachan and Ichise (Sachan and Ichise 2010) propose a syntactic approach considering dense subgraphs of a co-author network created from the DBLP. They discover relations between authors and propose pairs of researchers belonging to the same community. Several link discovery tools have been analyzed on specific knowledge graphs such as biomedical domain (Kastrin, Rindflesch, and Hristovski 2014). In a recent work (Kadlec, Bajgar, and Kleindienst 2017), a comparison of the already existing embedding models is investigated. In conclusion of this work, the simple models e.g., DistMult are selected to perform better with regard to accuracy by increasing batch size. Comparatively relevant to our experiments.

Conclusion and Future Work

In this paper, we adapted KGE models for the first time in the area of metaresearch recommendations with a particular focus on co-author recommendation. We built up a knowledge graph and presented early results for four different KGE models. The reported experimental results show that valuable recommendations can be provided in the domain of scholarly communication using embedding models. We plan to extend the created knowledge graph with more textual and visual descriptions and use other embedding models to improve the results. The aim is to provide an online service for researchers and facilitate scholarly communication in metaresearch explorations. In future work, we will also perform a larger scale user evaluation allowing us a more precise evaluation.

Acknowledgement

We acknowledge support of the European Union H2020 framework ITN projects WDAqua (grant no. 642795) and Cleopatra (grant no. 812997).

References

[Ali et al. 2018] Ali, M.; Hoyt, C. T.; Domingo-Fernandez, D.; Lehmann, J.; and Jabeen, H. 2018. Biokeen: A library for learning and evaluating biological knowledge graph embeddings. *bioRxiv* 475202.

[Bordes et al. 2013] Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, 2787–2795.

[Cheekula et al. 2015] Cheekula, S. K.; Kapanipathi, P.; Doran, D.; Jain, P.; and Sheth, A. P. 2015. Entity recommendations using hierarchical knowledge bases.

[Cheng et al. 2016] Cheng, H.-T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, 7–10. ACM.

[Kadlec, Bajgar, and Kleindienst 2017] Kadlec, R.; Bajgar, O.; and Kleindienst, J. 2017. Knowledge base completion: Baselines strike back. *arXiv preprint arXiv:1705.10744*.

[Kastrin, Rindflesch, and Hristovski 2014] Kastrin, A.; Rindflesch, T. C.; and Hristovski, D. 2014. Link prediction on the semantic MEDLINE network - an approach to literature-based discovery. In *The Discovery Science Conference*.

[Lin et al. 2015] Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, volume 15, 2181–2187.

[Ortega et al. 2018] Ortega, F.; Bobadilla, J.; Gutiérrez, A.; Hurtado, R.; and Li, X. 2018. Artificial intelligence scientific documentation dataset for recommender systems. *IEEE Access* 6:48543–48555.

[Passant 2010] Passant, A. 2010. dbrecmusic recommendations using dbpedia. In *International Semantic Web Conference*, 209–224. Springer.

[Sachan and Ichise 2010] Sachan, M., and Ichise, R. 2010. Using semantic information to improve link prediction results in network datasets. *IJET* 2(4).

[Vahdati et al. 2018] Vahdati, S.; Palma, G.; Nath, R. J.; Lange, C.; Auer, S.; and Vidal, M.-E. 2018. Unveiling scholarly communities over knowledge graphs. In *International Conference on Theory and Practice of Digital Libraries*, 103–115. Springer.

[Wang et al. 2014] Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, volume 14, 1112–1119.

[Wang et al. 2017] 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.

[Xie et al. 2016] Xie, R.; Liu, Z.; Jia, J.; Luan, H.; and Sun, M. 2016. Representation learning of knowledge graphs with entity descriptions. In *AAAI*, 2659–2665.

[Yang et al. 2014] Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

[7] https://wiki.dbpedia.org/