Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Sister Conferences Best Papers Track

Improving Inductive Link Prediction Using Hyper-Relational Facts (Extended Abstract)*

Mehdi Ali[1,2,†], **Max Berrendorf**[3,†], **Mikhail Galkin**[4], **Veronika Thost**[5], **Tengfei Ma**[5], **Volker Tresp**[3,6], **Jens Lehmann**[1,2]
[1]Smart Data Analytics Group, University of Bonn, Germany
[2]Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Sankt Augustin and Dresden, Germany
[3]Ludwig-Maximilians-Universität München, Munich, Germany
[4]Mila, McGill University
[5]IBM Research, MIT-IBM Watson AI Lab
[6]Siemens AG, Munich, Germany
{mehdi.ali,jens.lehmann}@cs.uni-bonn.de, {berrendorf,tresp}@dbs.ifi.lmu.de,
mikhail.galkin@mila.quebec, vth@zurich.ibm.com, tengfei.ma1@ibm.com

Abstract

For many years, link prediction on knowledge graphs (KGs) has been a purely transductive task, not allowing for reasoning on unseen entities. Recently, increasing efforts are put into exploring semi- and fully inductive scenarios, enabling inference over unseen and emerging entities. Still, all these approaches only consider triple-based KGs, whereas their richer counterparts, hyper-relational KGs (e.g., Wikidata), have not yet been properly studied. In this work, we study the benefits of employing hyper-relational KGs on a wide range of semi- and fully inductive link prediction tasks powered by recent advancements in graph neural networks. Our experiments demonstrate that qualifiers over typed edges can lead to performance improvements of 6% of absolute gains (for the Hits@10 metric) compared to triple-only baselines. Our code is available at https://github.com/mali-git/hyper_relational_ilp.

1 Introduction

Knowledge graphs are known for their incompleteness [Nickel *et al.*, 2016], therefore predicting missing links is one of the most important applications of machine learning over KGs [Nickel *et al.*, 2011; Bordes *et al.*, 2013]. A flurry [Ali *et al.*, 2021b; Ji *et al.*, 2020] of approaches has been developed over the years. Most of them operate over *triple-based* KGs in the *transductive* setup, where all entities are known at training time. Such approaches can neither operate on unseen entities, which might emerge after updating the graph, nor on new (sub-)graphs comprised of completely new entities. Those scenarios are often unified under the *inductive* link prediction (LP) setup. A variety of NLP tasks building upon KGs have inductive nature, for instance, entity linking or information extraction. Hence, being able to work in inductive settings becomes crucial for KG representation learning algorithms. For instance (see Fig. 1 in [Ali *et al.*, 2021a]), the `director-genre` pattern from the seen graph allows to predict a missing `genre` link for `The Martian` in the unseen subgraph.

Several recent works [Teru *et al.*, 2020; Daza *et al.*, 2021] investigate inductive LP but usually focus on a custom inductive setting. Further, their underlying KG structure is based on triples only. On the other hand, new, more expressive KGs like Wikidata [Vrandecic and Krötzsch, 2014] exhibit a *hyper-relational* nature where each triple (a typed edge in a graph) can be further instantiated with a set of explicit relation-entity pairs, known as *qualifiers* in the Wikidata model. Recently, it was shown [Galkin *et al.*, 2020] that employing hyper-relational KGs yields significant gains in the transductive LP task compared to their triple-only counterparts. However, the effect of such KGs on inductive LP is unclear. Intuitively (Fig. 1 in [Ali *et al.*, 2021a]), the `(nominee: Matt Damon)` qualifier provides a helpful signal to predict `Best Actor` as an object of `nominated for` of `The Martian` given that `Good Will Hunting` received such an award with the same nominee.

Therefore, in our ISWC 2021 paper [Ali *et al.*, 2021a], we systematically study hyper-relational KGs in different inductive settings:

- We propose a classification of inductive LP scenarios that describes the settings formally and, to the best of our knowledge, integrates all relevant existing works. Specifically, we distinguish *fully-inductive* scenarios, where target links are to be predicted in a new subgraph of unseen entities, and *semi-inductive* ones where unseen nodes have to be connected to a known graph.
- We adapt two existing baseline models for the two inductive LP tasks probing them in the hyper-relational settings.
- Our experiments suggest that models supporting hyper-relational facts indeed improve link prediction in both inductive settings compared to strong triple-only baselines by more than 6% Hits@10.

*Published at International Semantic Web Conference 21 - Received Best Research Paper Award

†equal contribution

2 Background

We assume the reader to be familiar with the standard link prediction setting (e.g. from [Nickel *et al.*, 2011]) and introduce the specifics of the setting with qualifiers.

Let $G = (\mathcal{E}, \mathcal{R}, \mathcal{S})$ be a hyper-relational KG where $\mathcal{E}$ is a set of entities, $\mathcal{R}$ is a set of relations, and $\mathcal{S}$ a set of statements. Each statement can be formalized as a 4-tuple (h, r, t, q) of a head and tail entity[1] $h, t \in \mathcal{E}$, a relation $r \in \mathcal{R}$, and a set of qualifiers, which are relation-entity pairs $q \subseteq \mathfrak{P}(\mathcal{R} \times \mathcal{E})$ where $\mathfrak{P}$ denotes the power set. For example, see Fig. 1 in [Ali *et al.*, 2021a] contains a statement `(Good Will Hunting, nominated for, Best Actor, {(nominee, Matt Damon)})` where `(nominee, Matt Damon)` is a qualifier pair for the main triple. We define the set of all possible statements as set

$$\mathbb{S}(\mathcal{E}_H, \mathcal{R}, \mathcal{E}_T, \mathcal{E}_Q) = \mathcal{E}_H \times \mathcal{R} \times \mathcal{E}_T \times \mathfrak{P}(\mathcal{R} \times \mathcal{E}_Q)$$

with a set of relations $\mathcal{R}$, a set of head, tail and qualifier entities $\mathcal{E}_H, \mathcal{E}_T, \mathcal{E}_Q \subseteq \mathcal{E}$. $\mathcal{S}_{train}$ is the set of training statements and $\mathcal{S}_{eval}$ are evaluation statements. We assume that we have a feature vector $\mathbf{x}_e \in \mathbb{R}^d$ associated with each entity $e \in \mathcal{E}$. Such feature vectors can, for instance, be obtained from entity descriptions available in some KGs or represent topological features [Belkin and Niyogi, 2001].

In this work, we focus on the setting with one fixed set of known relations. That is, we do not require $\mathbf{x}_r \in \mathbb{R}^d$ features for relations but learn relation embeddings during training.

3 Inductive Link Prediction

Recent works have pointed out the practical relevance of different inductive LP scenarios. However, there exists a terminology gap as different authors employ different names for describing conceptually the same task or, conversely, use the same *inductive LP* term for practically different setups. We propose a unified framework that provides an overview of the area and describes the settings formally.

Let $\mathcal{E}_\bullet$ denote the set of entities occurring in the training statements $\mathcal{S}_{train}$ at any position (head, tail, or qualifier), and $\mathcal{E}_\circ \subseteq \mathcal{E} \setminus \mathcal{E}_\bullet$ denote a set of unseen entities. In the *transductive* setting, all entities in the evaluation statements are seen during training, i.e., $\mathcal{S}_{eval} \subseteq \mathbb{S}(\mathcal{E}_\bullet, \mathcal{R}, \mathcal{E}_\bullet, \mathcal{E}_\bullet)$. In contrast, in *inductive* settings, $\mathcal{S}_{eval}$, used in validation and testing, may contain unseen entities. In order to be able to learn representations for these entities at inference time, inductive approaches may consider an additional set $\mathcal{S}_{inf}$ of *inference statements* about (un)seen entities; of course $\mathcal{S}_{inf} \cap \mathcal{S}_{eval} = \emptyset$.

The *fully-inductive* setting (FI) is akin to transfer learning where link prediction is performed over a set of entities not seen before, i.e., $\mathcal{S}_{eval} \subseteq \mathbb{S}(\mathcal{E}_\circ, \mathcal{R}, \mathcal{E}_\circ, \mathcal{E}_\circ)$. This is made possible by providing an auxiliary inference graph $\mathcal{S}_{inf} \subseteq \mathbb{S}(\mathcal{E}_\circ, \mathcal{R}, \mathcal{E}_\circ, \mathcal{E}_\circ)$ containing statements about the unseen entities in $\mathcal{S}_{eval}$. For instance, in Fig. 1 in [Ali *et al.*, 2021a], the training graph is comprised of entities `Matt Damon, Good Will Hunting, Best Actor, Gus Van Sant, Milk, Drama`. The inference graph contains new entities `The Martian, Alien, Ridley Scott, Blade Runner, Sci-fi` with one missing link to be predicted. The fully-inductive setting is considered in [Teru *et al.*, 2020; Daza *et al.*, 2021].

In the *semi-inductive* setting (SI), new, unseen entities are to be connected to seen entities, i.e., $\mathcal{S}_{eval} \subseteq \mathbb{S}(\mathcal{E}_\bullet, \mathcal{R}, \mathcal{E}_\circ, \mathcal{E}_\bullet) \cup \mathbb{S}(\mathcal{E}_\circ, \mathcal{R}, \mathcal{E}_\bullet, \mathcal{E}_\bullet)$. Illustrating with Fig. 1 in [Ali *et al.*, 2021a], `The Martian` as the only unseen entity connecting to the seen graph, the semi-inductive statement connects `The Martian` to the seen `Best Actor`. Note that there are other practically relevant examples beyond KGs, such as predicting interaction links between a new drug and a graph containing existing proteins/drugs [Bagherian *et al.*, 2020; Gaudelet *et al.*, 2020]. We hypothesize that, in most scenarios, we are not given any additional information about the new entity, and thus have $\mathcal{S}_{inf} = \emptyset$; we will focus on this case in this paper. However, the variation where $\mathcal{S}_{inf}$ may contain k statements connecting the unseen entity to seen ones has been considered too [Albooyeh *et al.*, 2020; Bhowmik and de Melo, 2020; Clouâtre *et al.*, 2021] and is known as *k-shot learning* scenario.

A mix of the fully- and semi-inductive settings where evaluation statements may contain two instead of just one unseen entity is studied in [Daza *et al.*, 2021; Baek *et al.*, 2020; Wang *et al.*, 2020]. That is, unseen entities might be connected to the seen graph, i.e., $\mathcal{S}_{eval}$ may contain seen entities, and, at the same time, the unseen entities might be connected to each other; i.e, $\mathcal{S}_{inf} \neq \emptyset$.

Our framework is general enough to allow $\mathcal{S}_{eval}$ to contain new, unseen relations r having their features $\mathbf{x}_r$ at hand. Since research so far has focused on the setting where all relations are seen in training, we will do so, too.

We hypothesize that qualifiers, being explicit attributes over typed edges, provide a strong inductive bias for LP tasks. In this work, for simplicity, we require both qualifier relations and entities to be seen in the training graph, i.e., $\mathcal{E}_Q \subseteq \mathcal{E}_\bullet$ and $\mathcal{R}_Q \subseteq \mathcal{R}$, although the framework accommodates a more general case of unseen qualifiers given their respective features.

4 Inductive Link Prediction with Qualifiers

Both semi- and fully-inductive tasks assume node features to be given. Recall that relation embeddings are learned and, often, to reduce the computational complexity, their dimensionality is smaller than that of node features.

Encoders. In the semi-inductive setting, an unseen entity arrives without any graph structure pointing to existing entities, i.e., $\mathcal{S}_{inf} = \emptyset$. This fact renders message passing approaches [Gilmer *et al.*, 2017] less applicable, so we resort to a simple linear layer to project all entity features (including those of qualifiers) into the relation space: $\phi : \mathbb{R}^{d_f} \rightarrow \mathbb{R}^{d_r}$.

In the fully inductive setting, we are given a non-empty inference graph $\mathcal{S}_{inf} \neq \emptyset$, and we probe two encoders: (i) the same linear projection of features as in the semi-inductive scenario which does not consider the graph structure; (ii) GNNs which can naturally work in the inductive settings [Chami *et al.*, 2020]. However, the majority of existing GNN encoders for multi-relational KGs like CompGCN [Vashishth *et al.*, 2020] are limited to only triple KG representation. To the best of our knowledge, only the recently proposed StarE [Galkin

[1] We use *entity* and *node* interchangeably

Model	#QP	WD20K (100) V1					WD20K (100) V2				
		AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)	AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)
BLP	0	22.78	5.73	1.92	8.22	12.33	36.71	3.99	1.47	4.87	9.22
CompGCN	0	37.02	10.42	<u>5.75</u>	15.07	18.36	74.00	2.55	0.74	3.39	5.31
QBLP	0	28.91	5.52	1.51	8.08	12.60	35.38	4.94	2.58	5.46	9.66
StarE	2	41.89	9.68	3.73	**16.57**	20.99	40.60	2.43	0.45	3.86	6.17
StarE	4	35.33	10.41	4.82	15.84	21.76	37.16	5.12	1.41	7.93	12.89
StarE	6	34.86	**11.27**	**6.18**	15.93	21.29	47.35	4.99	1.92	6.71	11.06
QBLP	2	**18.91**	10.45	3.73	16.02	<u>22.65</u>	**28.03**	**6.69**	**3.49**	<u>8.47</u>	12.04
QBLP	4	<u>20.19</u>	<u>10.70</u>	3.99	<u>16.12</u>	**24.52**	<u>31.30</u>	5.87	2.37	7.85	**13.93**
QBLP	6	23.65	7.87	2.75	10.44	17.86	34.35	<u>6.53</u>	<u>2.95</u>	**9.29**	<u>13.13</u>

Table 1: Results on FI WD20K (100) V1 & V2. #QP denotes the number of qualifier pairs used in each statement (including padded pairs). Best results **in bold**, second best <u>underlined</u>.

et al., 2020] encoder supports hyper-relational KGs which we take as a basis for our inductive model. Its aggregation formula is:

$$\mathbf{x}'_v = f\left(\sum_{(u,r)\in\mathcal{N}(v)} \mathbf{W}_{\lambda(r)}\phi_r(\mathbf{x}_u, \gamma(\mathbf{x}_r, \mathbf{x}_q)_{vu})\right) \quad (1)$$

where γ is a function that infuses the vector of aggregated qualifiers $\mathbf{x}_q$ into the vector of the main relation $\mathbf{x}_r$. The output of the GNN contains updated node and relation features based on the adjacency matrix A and qualifiers Q:

$$\mathbf{X}', \mathbf{R}' = \textsc{StarE}(A, \mathbf{X}, \mathbf{R}, Q)$$

Finally, in both inductive settings, we linearize an input statement in a sequence using a padding index where necessary: $[\mathbf{x}'_h, \mathbf{x}'_r, \mathbf{x}'_{q_1^r}, \mathbf{x}'_{q_1^e}, \text{[PAD]}, \ldots]$. Note that statements can greatly vary in length depending on the amount of qualifier pairs, and padding mitigates this issue.

Decoder. Given an encoded sequence, we use the same Transformer-based decoder for all settings:

$$f(h,r,t,q) = g(\mathbf{x}'_h, \mathbf{x}'_r, \mathbf{x}'_{q_1^r}, \mathbf{x}'_{q_1^e}, \ldots)^T \mathbf{x}'_t \text{ with}$$
$$g(\mathbf{x}'_1, \ldots, \mathbf{x}_k) = \text{Agg}(\text{Transformer}([\mathbf{x}'_1, \ldots, \mathbf{x}'_k]))$$

In this work, we evaluated several aggregation strategies and found a simple mean pooling over all non-padded sequence elements to be preferable.

Here and below, we denote the linear encoder + Transformer decoder model as QBLP (that is, Qualifier-aware BLP, an extension of BLP [Daza *et al.*, 2021]), and the STARE encoder + Transformer decoder, as STARE.

In order to compare results with triple-only approaches, we trained the models, as usual, on subject and object prediction.

5 Experiments

We design our experiments to investigate whether the incorporation of qualifiers improves inductive link prediction. In particular, we investigate the fully-inductive setting (Section 5.1) and the semi-inductive setting (Section 5.2). We analyze the impact of the qualifier ratio (i.e., the number of statements with qualifiers) and the dataset's size on a model's performance.

5.1 Fully-Inductive Setting

In the full inductive setting, we analyzed the effect of qualifiers for four different datasets (i.e., WD20K (100) V1 & V2 and WD20K (66) V1 & V2, which have different ratios of qualifying statements, the numbers in brackets, and are of different sizes). As triple-only baselines, we evaluated CompGCN [Vashishth *et al.*, 2020] and BLP [Daza *et al.*, 2021]. To evaluate the effect of qualifiers on the fully-inductive LP task, we evaluated StarE [Galkin *et al.*, 2020] and QBLP. It should be noted that StarE without the use of qualifiers is equivalent to CompGCN.

The main findings (from all experiments) are that (i) for all datasets, the use of qualifiers leads to increased performance, and (ii) the ratio of statements with qualifiers and the size of the dataset has a major impact on the performance. CompGCN and StarE apply message-passing to obtain enriched entity representations while BLP and QBLP only apply a linear transformation. Consequently, CompGCN and StarE require $\mathcal{S}_{inf}$ to contain useful information in order to obtain the entity representations while BLP and QBLP are independent of $\mathcal{S}_{inf}$. Table 1 shows the results obtained for two of the four datasets.

We observe that the performance gap between BLP/QBLP (0) and QBLP (2,4,6) is considerably larger than the gap between CompGCN and StarE. This might be explained by the fact that QBLP does not take into account the graph structure provided by $\mathcal{S}_{inf}$, hence is heavily dependent on additional information, i.e., the qualifiers compensate for the missing graph information. The overall performance decrease observable between V1 and V2 could be explained by the datasets' composition, in particular, in the composition of the training and inference graphs: $\mathcal{S}_{inf}$ of V2 comprises more entities than V1, so that each test triple is ranked against more entities, i.e., the ranking becomes more difficult. At the same time, the training graph of V1 is larger than that of V2, i.e., during training more entities are seen which may improve generalization.

5.2 Semi-inductive Setting

In the SI setting, we evaluated BLP as a triple-only baseline and QBLP as a statement baseline (i.e., involving qualifiers) on the WD20K SI datasets. We did not evaluate CompGCN and StarE since message-passing-based approaches are not directly applicable in the absence of $\mathcal{S}_{inf}$. The results highlight that aggregating qualifier information improves the prediction

Model	#QP	WD20K (33) SI					WD20K (25) SI				
		AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)	AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)
BLP	0	**4.76**	13.95	7.37	17.28	24.65	6.01	12.45	5.98	17.29	23.43
QBLP	0	7.04	28.35	14.44	28.58	36.32	6.75	17.02	8.82	22.10	29.50
QBLP	2	11.51	**35.95**	**20.70**	**34.98**	**41.82**	5.99	20.36	11.77	**24.86**	**32.26**
QBLP	4	11.38	34.35	19.41	33.90	40.20	12.18	**21.05**	**12.32**	24.07	30.09
QBLP	6	4.98	25.94	15.20	30.06	38.70	**5.73**	19.50	11.14	24.73	31.60

Table 2: Results on the WD20K SI datasets. #QP denotes the number of qualifier pairs used in each statement (including padded pairs).Best results **in bold**, second best underlined.

WD20K (100) V1 FI		
Wikidata ID	relation name	ΔMR
P2868	subject has role	0.12
P463	member of	-0.04
P1552	has quality	-0.34
P2241	reason for deprecation	-26.44
P47	shares border with	-28.91
P750	distributed by	-29.12

Table 3: Top 3 worst and best qualifier relations affecting overall MR.

of semi-inductive links despite the fact that the ratio of statements with qualifiers is not very large (37% for SI WD20K (33), and 30% for SI WD20K (25)). In the case of SI WD20K (33), the baselines are outperformed even by a large margin. Overall, the results might indicate that in semi-inductive settings, performance improvements can already be obtained with a decent amount of statements with qualifiers.

5.3 Qualitative Analysis

We obtain deeper insights on the impact of qualifiers by analyzing the StarE model on the fully-inductive WD20K (66) V2 dataset. In particular, we study individual ranks for head/tail prediction of statements with and without qualifiers (see Fig. 2 in [Ali *et al.*, 2021a]) varying the model from zero to four pairs. First, we group the test statements by the number of available qualifier pairs. We observe generally smaller ranks which, in turn, correspond to better predictions when more qualifier pairs are available. In particular, just one qualifier pair is enough to significantly reduce the individual ranks. Note that we have less statements with many qualifiers.

We also study how particular qualifiers affect ranking and predictions (see Fig. 3 in [Ali *et al.*, 2021a]). For that, we measure ranks of predictions for distinct statements in the *test set* with and without masking the qualifier relation from the inference graph $\mathcal{S}_{inf}$. We then compute ΔMR and group them by used qualifier relations. Interestingly, certain qualifiers, e.g., `convicted of` or `including`, may deteriorate performance, which we attribute to the usage of rare, qualifier-only entities. Nevertheless, other qualifiers largely reduce ranks and hence very positively impact prediction accuracy.

Finally, we study the average impact of qualifiers on the whole graph, i.e., we take the whole *inference graph* and mask out all qualifier pairs containing one relation and compare the overall evaluation result on the test set, we count ranks of all test statements, not only those which have that particular qualifier) against the non-masked version of the same graph. We then sort relations by ΔMR and find top 3 most confusing and most helpful relations across two datasets (cf. Table 3). On the smaller WD20K (100) V1 where all statements have at least one qualifier pair, most relations tend to improve MR. For instance, qualifiers with the `distributed by` relations reduce MR by about 29 points. On the larger WD20K (66) V2 some qualifier relations, e.g., `statement is subject of`, tend to introduce more noise and worsen MR which we attribute to the increased sparsity of the graph given an already rare qualifier entity. That is, such rare entities might not benefit enough from message passing.

6 Conclusion

In our ISWC 2021 paper [Ali *et al.*, 2021a], we present a study of the inductive link prediction problem over hyper-relational KGs. In particular, we propose a theoretical framework to categorize various LP tasks to alleviate an existing terminology discrepancy pivoting on two settings, semi- and fully-inductive LP. Probing statement-aware models against triple-only baselines, we demonstrated that hyper-relational facts considerably improve LP performance in both inductive. Moreover, our qualitative analysis show that the achieved gains are consistent across different setups and still interpretable.

Our findings open up interesting prospects for employing inductive LP and hyper-relational KGs along several axes, e.g., large-scale KGs of billions statements, new application domains including life sciences, drug discovery, and KG-based NLP applications like question answering or entity linking.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A and Grant No. 01IS18050D (project "MLWin"). The authors of this work take full responsibilities for its content.

References

[Albooyeh *et al.*, 2020] Marjan Albooyeh, Rishab Goel, and Seyed Mehran Kazemi. Out-of-sample representation learning for knowledge graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP 2020, Online Event, 16-20 November 2020*, pages 2657–2666. Association for Computational Linguistics, 2020.

[Ali *et al.*, 2021a] Mehdi Ali, Max Berrendorf, Mikhail Galkin, Veronika Thost, Tengfei Ma, Volker Tresp, and Jens Lehmann. Improving inductive link prediction using hyper-relational facts. In *The Semantic Web – ISWC 2021*, pages 74–92, Cham, 2021. Springer International Publishing.

[Ali *et al.*, 2021b] Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Mikhail Galkin, Sahand Sharifzadeh, Asja Fischer, Volker Tresp, and Jens Lehmann. Bringing light into the dark: A large-scale evaluation of knowledge graph embedding models under a unified framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[Baek *et al.*, 2020] Jinheon Baek, Dong Bok Lee, and Sung Ju Hwang. Learning to extrapolate knowledge: Transductive few-shot out-of-graph link prediction. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[Bagherian *et al.*, 2020] Maryam Bagherian, Elyas Sabeti, Kai Wang, Maureen A. Sartor, Zaneta Nikolovska-Coleska, and Kayvan Najarian. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Briefings in Bioinformatics*, 22(1):247–269, 01 2020.

[Belkin and Niyogi, 2001] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani, editors, *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pages 585–591. MIT Press, 2001.

[Bhowmik and de Melo, 2020] Rajarshi Bhowmik and Gerard de Melo. Explainable link prediction for emerging entities in knowledge graphs. In *ISWC (1)*, volume 12506 of *Lecture Notes in Computer Science*, pages 39–55. Springer, 2020.

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795, 2013.

[Chami *et al.*, 2020] Ines Chami, Sami Abu-El-Haija, Bryan Perozzi, Christopher Ré, and Kevin Murphy. Machine learning on graphs: A model and comprehensive taxonomy. *CoRR*, abs/2005.03675, 2020.

[Clouâtre *et al.*, 2021] Louis Clouâtre, Philippe Trempe, Amal Zouaq, and Sarath Chandar. MLMLM: link prediction with mean likelihood masked language model. In *ACL/IJCNLP (Findings)*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 4321–4331. Association for Computational Linguistics, 2021.

[Daza *et al.*, 2021] Daniel Daza, Michael Cochez, and Paul Groth. Inductive entity representations from text via link prediction. In *WWW*, pages 798–808. ACM / IW3C2, 2021.

[Galkin *et al.*, 2020] Mikhail Galkin, Priyansh Trivedi, Gaurav Maheshwari, Ricardo Usbeck, and Jens Lehmann. Message passing for hyper-relational knowledge graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 7346–7359. Association for Computational Linguistics, 2020.

[Gaudelet *et al.*, 2020] Thomas Gaudelet, Ben Day, Arian R. Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy B. R. Hayter, Richard Vickers, Charles Roberts, Jian Tang, David Roblin, Tom L. Blundell, Michael M. Bronstein, and Jake P. Taylor-King. Utilising graph machine learning within drug discovery and development. *CoRR*, abs/2012.05716, 2020.

[Gilmer *et al.*, 2017] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *ICML*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 2017.

[Ji *et al.*, 2020] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition and applications. *CoRR*, abs/2002.00388, 2020.

[Nickel *et al.*, 2011] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 809–816. Omnipress, 2011.

[Nickel *et al.*, 2016] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proc. IEEE*, 104(1):11–33, 2016.

[Teru *et al.*, 2020] Komal Teru, Etienne Denis, and Will Hamilton. Inductive relation prediction by subgraph reasoning. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 9448–9457. PMLR, 2020.

[Vashishth *et al.*, 2020] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. Composition-based multi-relational graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[Vrandecic and Krötzsch, 2014] Denny Vrandecic and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85, 2014.

[Wang *et al.*, 2020] Bin Wang, Guangtao Wang, Jing Huang, Jiaxuan You, Jure Leskovec, and C.-C. Jay Kuo. Inductive learning on commonsense knowledge graph completion. *CoRR*, abs/2009.09263, 2020.