Deep Knowledge Graph Representation Learning for Completion, Alignment, and Question Answering

Soumen Chakrabarti
IIT Bombay
India

CCS CONCEPTS

• **Information systems** → *Searching with auxiliary databases.*

KEYWORDS

Knowledge graph embeddings; Question answering

ACM Reference Format:
Soumen Chakrabarti. 2022. Deep Knowledge Graph Representation Learning for Completion, Alignment, and Question Answering. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3477495.3532679

1 MOTIVATION

A knowledge graph (KG) has nodes and edges representing entities and relations. Nodes have canonical entity IDs and edges have canonical relation IDs. E.g., in Wikidata, Barack Obama and Honolulu have canonical IDs Q76 and Q18094, and the relation "place of birth" has canonical ID P19. A fact triple involving these as subject, relation and object is written as (Q76, P19, Q18094), or, more colloquially, (Barack Obama, place of birth, Honolulu). Curating and structuring knowledge has been a human pursuit since time immemorial, but large, collaboratively maintained KGs such as WordNet [23], Wikipedia, DBPedia, YAGO [37], Freebase [3], and Wikidata bloomed with the advent of the Internet. Wikidata has over 97 M entities, almost 10 k relations, and close to 1.5 B facts.

KGs are of central interest to search, information retrieval (IR) and question answering (QA), as evidenced by Google's purchase of Freebase in 2008, followed by Bing's development of the Satori KG, and Amazon's product KG. The use of KGs in pre-neural search and QA did see some representation [14, 30] in search and IR conferences. However, with the widespread adoption of neural networks and deep learning, the last several years have witnessed a huge surge in KG-related papers in AI, machine learning (ML) and NLP conferences, but comparatively less action in IR conferences. This stands in contrast to dense passage retrieval (DPR), where the IR community embraced deep learning early [20, 25]. My motivation behind proposing this tutorial is to bridge the gap described above, give IR researchers a thorough overview of the best practices of neural KG representation and inference from AI and NLP communities, and then explore how KG representation research of the IR community can be better driven by the needs of search and QA.

SIGIR '22, July 11–15, 2022, Madrid, Spain

ACM ISBN 978-1-4503-8732-3/22/07…$15.00
https://doi.org/10.1145/3477495.3532679

2 OBJECTIVES

By attending the proposed tutorial, participants will learn about:

- The most widely-used public KGs of today.
- Important properties of KG relations, types and entities.
- Best-practice deep representations of KG elements.
- The most successful techniques for KG completion and inference.
- Early work on representing types and time in KGs.
- The latest approaches to align and connect multiple KGs.
- Use and benefits of deep KG representations in QA applications.

3 RELEVANCE TO THE IR COMMUNITY

By 2005, search engines were racing to take the user experience beyond "ten blue links". KGs were an important part of the plan. Initially, their role was limited to cards and callouts in verticals. Shortly, the role of KGs in search was expanded to entity linking in queries [6] and corpus [16, 24], answer type prediction [15, 18], and joint query interpretation and response ranking [30, 31]. These methods were developed largely in the overlapping WebConf, SIGIR and CIKM communities, and were pre-neural and interpretable.

Around 2013, word2vec [21] and GloVE [28] word embeddings were introduced, leading to the neural/deep NLP revolution. NLP conferences exploded with search and QA papers. KG research, initially distributed among Semantic Web and database communities, moved into the AI and ML communities. The IR community remained consumers, but not key innovators, of KG and QA technology. It continued to lead efforts toward harnessing deep learning for search and retrieval, leading to a series of breakthroughs [20, 25].

It is important for the IR community to resume greater focus on KG representation, guided by the needs of search and QA.

- QA can benefit from text/search-guided KG embeddings, particularly for types and relations. Many queries fail because answer spans do not satisfy type and relation constraints that are not easily expressed within canonical KGs.
- The IR community lies somewhere between the database community with precise schemas and algebra, and the data-driven faith in deep QA in the NLP/AI communities. We can push for interpretable complex query decomposition and solution assembly, features largely missing from opaque neural QA to date.
- Cross-lingual IR has been of interest for a while, but, thanks to the invention of byte-pair and word-piece driven transformer networks [9], multilingual work has (understandably) grown much faster in the NLP community. Alignment between KGs across languages poses new research problems and can be useful in cross-lingual search and QA.

4 FORMAT AND DETAILED SCHEDULE

I will interleave slide-based presentation, scribbling on a jamboard, screen-sharing code snippets and short demos, and Q&A sessions (at least every 25 minutes). We may set up Slack or Piazza for attendees to interact with me and with each other. The schedule is for 165m (minutes) of presentation and 15m of Q&A.

4.1 Preliminaries (15m)

4.1.1 Familiarization with modern KGs. We will introduce KGs as graphs where each edge (s, r, o) represents a relation r that holds between subject s and object o. We will inspect samples from the most popular KGs: how names, descriptions, attributes are coded; how relations are associated with names; specific properties of relations — symmetry ($(s, r, o) \implies (o, r, s)$), antisymmetry ($(s, r, o) \implies \neg(o, r, s)$), asymmetry (neither implication holds), transitivity ($(e_1, r, e_2) \wedge (e_2, r, e_3) \implies (e_1, r, e_3)$, reflexivity ($(e, r, e)$ holds); representing >2-ary relations using CVT nodes [41]; etc.

4.1.2 The need for KG completion. We will sample statistics from DBpedia and Wikidata which show the extent to which top KGs are incomplete, with examples of suitable evidence via Web search which has not yet been curated into the KGs.

4.1.3 The need for entity and relation alignment. We will show samples from KGs in multiple languages that establish the synergy between KGs and the potential for augmenting a KG in one language with information gleaned from another.

4.1.4 Utility of KG in QA. We will sample queries where KGs can be harnessed for improved QA accuracy, and examples of KG-corpus synergy in answer selection and scoring [14, 30, 31].

4.2 Fact scoring and sampling (15m)

All KG representation models associate each entity e (variously, subject s and object o) with dense, continuous representations that we will denote by $\boldsymbol{e}$ (variously, $\boldsymbol{s}$ and $\boldsymbol{o}$). Similarly, relation r will be represented by $\boldsymbol{r}$. Notwithstanding the vector-like notation, the representations may be matrices representing rotations or affine tranforms, and the elements of the representations may be real or complex, in a Euclidean or non-Euclidean space, etc.

4.2.1 Scoring function. At the heart of a KG representation is a scoring function $f(s, r, o)$ that returns a real number, a possibly unnormalized belief in the validity of the fact (s, r, o). (Even invalid tuples are conventionally called 'facts'.) In principle, if f is implemented as a neural network with sufficient capacity (width and depth) — i.e., if we use a "universal approximator" — and if sufficient training data is available, we should be able to train the entity and relation embeddings and infer new facts. In practice, much trial-and-error is needed to design the structure of f.

4.2.2 Negative sampling. If a fact is valid, i.e., found in the KG, we write this as $(s, r, o) \in \text{KG}$; if it is not, we write this as $(s, r, o) \notin \text{KG}$ or $(s, r, o) \in \overline{\text{KG}}$. The parameters in entity and relation embeddings, or the score f, are often trained by demanding that $f(s, r, o) > f(s', r', o')$ when $(s, r, o) \in \text{KG}$ and $(s', r', o') \in \overline{\text{KG}}$. But the space of invalid facts $\overline{\text{KG}}$ is so astronomically large that various corners have to be (carefully) cut to keep training times practical and yet train model parameters to satisfactory predictive quality. We will discuss tricks of the trade and their performance implications.

4.3 Translation and rotation models (15m)

Word2vec introduced the surprising observation that vector translations can associate with meaningful attributes and relations, e.g., $\overrightarrow{king} - \overrightarrow{man} \approx \overrightarrow{queen} - \overrightarrow{woman}$. This directly 'translates' to the TransE [4] model $\boldsymbol{s} + \boldsymbol{r} \approx \boldsymbol{o}$. In other words, f is inversely proportional to $\|\boldsymbol{s} + \boldsymbol{r} - \boldsymbol{o}\|$. However, this clearly cannot handle many-to-1, 1-to-many, or many-to-many relations (e.g., if a person attends two different schools, their embeddings must be crushed together). To mitigate this limitation and support many-to-many relations, various projective extensions have been proposed. Unfortunately, most additive models cannot distinguish between different symmetric relations: if (s, r, o) and (o, r, s) are both valid, then we must ideally have $\boldsymbol{s} + \boldsymbol{r} = \boldsymbol{o}$ and $\boldsymbol{o} + \boldsymbol{r} = \boldsymbol{s}$, which means $\boldsymbol{s} + \boldsymbol{r} = \boldsymbol{s} - \boldsymbol{r}$, or $\boldsymbol{r} \equiv \vec{0}$ for all symmetric relations r. RotatE [39] replaces translation with rotation: it places $\boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o} \in \mathbb{C}^D$ in the complex space, enforces $\|\boldsymbol{r}\|_2 = 1$ and defines f as inversely proportional to $\|\boldsymbol{s} \odot \boldsymbol{r} - \boldsymbol{o}\|_2^2$, where $\odot$ represents elementwise product. RotatE can learn symmetry vs. antisymmetry, inversion and composition.

4.4 Factorization/multiplicative models (15m)

Given a KG with E entities and R relations, for simpler tensor notation, let $s, o \in [E]$ and $r \in [R]$.[1] The original KG can be characterized as a 3-axis tensor $\mathbf{X}$ with axes lengths $E \times E \times R$. $\mathbf{X}[s, o, r] = -1$ if $(s, r, o) \in \overline{\text{KG}}$, it is +1 if $(s, r, o) \in \text{KG}$, and 0 if (s, r, o) is not known to be valid/invalid. Various multiplicative models are a result of approximating $\mathbf{X}$ with a low-rank factorization into an entity embedding matrix $\boldsymbol{A} \in \mathbb{R}^{E \times D}$ and a relation embedding tensor $\mathbf{B} \in \mathbb{R}^{D \times D \times R}$, in the form of the tensor product

$$\mathbf{X}[s, o, r] \approx \boldsymbol{A}[s, :]^\top \mathbf{B}[r, :, :] \boldsymbol{A}[o, :] \tag{1}$$

(where $\mathbf{X}[s, o, r] \neq 0$). This is reminiscent of low-rank matrix factorization, as in Latent Semantic Indexing [8, LSI/SVD] used in the IR community. When $\mathbf{B}[r, :, :] \in \mathbb{R}^{D \times D}$ is restricted to a diagonal matrix, we get a representation $\boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o} \in \mathbb{R}^D$ and score function

$$f(s, r, o) = \langle \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o} \rangle = \textstyle\sum_{d \in [D]} \boldsymbol{s}[d]\, \boldsymbol{r}[d]\, \boldsymbol{o}[d], \tag{2}$$

the DistMult model [42]. This formula clearly cannot represent asymmetric relations, which was fixed by lifting to the complex domain, i.e., $\boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o} \in \mathbb{C}^D$, and modifying the scoring function to

$$f(s, r, o) = \operatorname{Re}\left(\textstyle\sum_{d \in [D]} \boldsymbol{s}[d]\, \boldsymbol{r}[d]\, \boldsymbol{o}[d]^\star \right), \tag{3}$$

where $c^\star = \operatorname{Re}(c) - \mathbf{i} \operatorname{Im}(c)$ is the complex conjugate of $c = \operatorname{Re}(c) + \mathbf{i} \operatorname{Im}(c)$ and $\mathbf{i} = \sqrt{-1}$. We will describe in detail the evolution of these multiplicative models DistMult [42] and ComplEx [43], and their connections to holographic embeddings [27]. All of these can support many-to-many relations. We will show how ComplEx can learn symmetric, anti-symmetric and asymmetric relations. We will discuss subtle accuracy and performance effects of negative sampling policies and computational tricks.

4.5 Embeddings hierarchies (20m)

Even though multiplicative/factor models perform better than additive models at KG completion tasks, they do not explicitly deal with transitive relations. Directed acyclic transitive relations are very important in KGs, e.g., for representing is-instance-of, is-subtype-of, and is-part-of relations. For simplicity, let us consider a single

[1] $[N]$ means $\{1, 2, \ldots, N\}$ or $\{0, 1, \ldots, N-1\}$ depending on context.

such partial order written as $e \prec e'$, e.g., *Einstein* $\prec$ *physicist* $\prec$ *scientist* $\prec$ *human* $\prec$ *living being*.

4.5.1 Order and box embeddings. A conceptually simple way to cope with partial orders is to move to an algebra for f that is not based on normed metric spaces. E.g., if e, e' are embedded to $\boldsymbol{e}, \boldsymbol{e'} \in \mathbb{R}^D$, and if $e \prec e'$, we might insist that $\boldsymbol{e} \le \boldsymbol{e'}$, where '$\le$' is interpreted elementwise. Such "order embeddings" [44] have been used for image analysis and textual entailment. 'Boxes' or multidimensional axis-aligned rectangular *regions of embedding space* [36] are more expressive than the open cones implicit in order embeddings. Ideally, if $e \prec e'$, the box representing e should be contained inside that representing e'. Much depends on how loss functions with gradience guidance are designed around violations of elementwise dominance or box containment [19, 36].

4.5.2 Generalized box embeddings. Even after box embeddings became established, it was not immediately clear how to unify these embeddings, intended for very specific transitive relations into a general embedding scheme of a KG with many other non-transitive relations. Although not always the top performer in terms of KG completion, BoxE [1] gives an elegant formalism to address this issue. It also supports general K-ary relations $r(e_1, \ldots, e_K)$ (K may vary across relations). Each entity e is associated with *two* embedding vectors $\boldsymbol{c}_e, \boldsymbol{b}_e \in \mathbb{R}^D$. The entity vector used to compute a fact score depends upon all other entities participating in it:

$$\boldsymbol{e}_i^{r(e_1,\ldots,e_K)} = \boldsymbol{c}_{e_i} - \boldsymbol{b}_{e_i} + \textstyle\sum_{k\in[K]} \boldsymbol{b}_{e_k} \tag{4}$$

In words, entity e_k 'bumps' by $\boldsymbol{b}_{e_k}$ the 'center' vector $\boldsymbol{c}_{e_i}$ of all entities e_i with which it participates in relations. Meanwhile, a K-ary relation r is represented by K boxes $\boldsymbol{r}^1, \ldots, \boldsymbol{r}^K$. If $r(e_1, \ldots, e_K)$ is valid, we want $\boldsymbol{e}_i^{r(e_1,\ldots,e_K)} \in \boldsymbol{r}^i$ for each $i \in [K]$. BoxE has many nice properties that we will discuss, along with the details of loss design and optimization.

4.5.3 Hyperbolic spaces. Embedding a tree in Euclidean space requires exponential space near the leaves, but only polynomial space in the (unit radius) Poincaré ball [26], where the distance between $\boldsymbol{x}, \boldsymbol{y}$ is $\operatorname{arcosh}\big(1 + 2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/(1-\|\boldsymbol{x}\|_2^2)(1-\|\boldsymbol{y}\|_2^2)\big)$, rather than $\|\boldsymbol{x} - \boldsymbol{y}\|_2$. As we shall show, this can be useful for embedding type hierarchies. Hyperbolic cones [10] are another alternative.

4.6 Embeddings for temporal KGs (20m)

Many KG triples are qualified with a (set of) time(s) when the fact is valid, e.g., (Obama, president-of, USA; **[2009, 2016]**) or (JFK, died-of, assassination, **1963**). In general, we can write $(s, r, o; T)$ where T is a set of times. Two common cases are that T is a sequence of 'instants'[2] (useful for periodic events) and T is a single interval (presidency, marriage, etc.). The score function $f(s, r, o, T)$ now has a fourth argument, which we also want to embed. Such temporal KGs are naturally useful for temporal QA [34].

Some methods adopt the position that an entity has an intrinsic fixed representation, from which a time-dependent representation can be obtained for each time instant. I.e., $f(s, r, o, t)$ can be written in the form $f(\boldsymbol{s}_t, r, \boldsymbol{o}_t)$. Diachronic embeddings [11] are an example. Other approaches transform all arguments to time-modulated representations, i.e., $f(\boldsymbol{s}_t, \boldsymbol{r}_t, \boldsymbol{o}_t)$. HyTE [7] allocates a unit vector $\boldsymbol{h}_t$ for each time step t, as a normal to a hyperplane. Each entity e (variously, s, o) gets a time-independent embedding $\boldsymbol{e}$ (variously $\boldsymbol{s}, \boldsymbol{o}$), which is projected to $\boldsymbol{e}{\downarrow_{\boldsymbol{h}_t}}$ on the hyperplane, to find its representation for time t. The same treatment obtains $\boldsymbol{r}{\downarrow_{\boldsymbol{h}_t}}$. Now TransE is used in the subspace of the hyperplane: $\boldsymbol{s}{\downarrow_{\boldsymbol{h}_t}} + \boldsymbol{r}{\downarrow_{\boldsymbol{h}_t}} \approx \boldsymbol{o}{\downarrow_{\boldsymbol{h}_t}}$. TNTComplex [17] takes advantage of ComplEx as the base system, rather than TransE. It defines $f(s, r, o, t) = \operatorname{Re}(\langle \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o}^\star, \boldsymbol{t}\rangle)$, where

$$\langle \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o}^\star, \boldsymbol{t}\rangle = \langle \boldsymbol{s} \odot \boldsymbol{t}, \boldsymbol{r}, \boldsymbol{o}^\star\rangle = \langle \boldsymbol{s}, \boldsymbol{r} \odot \boldsymbol{t}, \boldsymbol{o}^\star\rangle = \langle \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o}^\star \odot \boldsymbol{t}\rangle \tag{5}$$

I.e., time is allowed to modulate *any one of* subject, relation, object representations; the result is the same. This was found restrictive, and in TimePlex [12], relation r is represented in three parts $\boldsymbol{r}^{\mathrm{SO}}, \boldsymbol{r}^{\mathrm{ST}}, \boldsymbol{r}^{\mathrm{OT}}$, and the score function is defined as

$$\begin{aligned}&\operatorname{Re}(\langle \boldsymbol{s}, \boldsymbol{o}, \boldsymbol{t}^\star\rangle) + \blacklozenge \operatorname{Re}(\langle \boldsymbol{s}, \boldsymbol{r}^{\mathrm{ST}}, \boldsymbol{t}^\star\rangle) + \\ &\qquad \spadesuit \operatorname{Re}(\langle \boldsymbol{o}, \boldsymbol{r}^{\mathrm{OT}}, \boldsymbol{t}^\star\rangle) + \clubsuit \operatorname{Re}(\langle \boldsymbol{s}, \boldsymbol{r}^{\mathrm{SO}}, \boldsymbol{o}^\star\rangle)\end{aligned} \tag{6}$$

where $\blacklozenge, \spadesuit, \clubsuit$ are hyperparameters. TimePlex also learns recurrence and time gaps between relations to give state-of-the-art predictions.

4.7 Translingual KG alignment (35m)

KGs in different languages l, l' may evolve at their own pace, with occasional cross-KG 'gold' entity alignments (EA) $e_l \equiv e_{l'}$ or relation alignment (RA) $r_l \equiv r_{l'}$ via human inputs. The goal is to infer additional EAs and RAs. In each language, an entity or relation is known by different names, e.g., "Estados Unidos de América" vs. "United States of America" in Spanish and English KGs.

4.7.1 Regularizing known-equivalent embeddings. A tempting first solution is to run two copies of a competitive KG representation and completion system on the two KGs specific to the languages. If $e_l \equiv e_{l'}$ is known as a training EA pair, we add a regularization term $\|\boldsymbol{e}_l - \boldsymbol{e}_{l'}\|^2$ to the loss objective, which helps align the spaces of the two KGs [5]. During testing, entity/relation pairs with similar embeddings are ranked high for alignment.

4.7.2 Graph neural networks. To compare entity nodes in a neighborhood-cognizant manner, one can first collapse gold EA node pairs and run enhanced forms of graph convolutional network (GCN, GRN, GNN) iterations: $\boldsymbol{u}^{(k)} \longleftarrow \bigoplus_{(u,r,v)\in\mathrm{KG}} \sigma(\boldsymbol{M}_r \boldsymbol{v}^{(k-1)})$, where $\boldsymbol{M}_r$ is a trained transfer matrix which are also tied for known RAs $r_l \equiv r_{l'}$. RDGCN and RNM [46, 48] build upon GNNs and further enhance them for competitive EA performance.

4.7.3 Deep set similarity. Instead of aggregating KG neighborhoods into fixed-length node embeddings (as do GNNs), BERT-INT [40] compares entities e and e' by directly comparing their immediate neighbors $v \in N(e), v' \in N(e')$ pairwise. Neighbors v, v' are represented by embeddings $\boldsymbol{v}, \boldsymbol{v'}$ based on local text. This results in two sets of embeddings $S(e) = \{\boldsymbol{v} : v{\in}N(e)\}$ and $S(e') = \{\boldsymbol{v'} : v'{\in}N(e')\}$. The Cartesian space of $|N(e)| \times |N(e')|$ similarities is pooled into a similarity score between e and e'. AlignKGC [35] applies a similar idea to RA, by defining $\mathrm{SO}(r) = \{\langle \boldsymbol{s}, \boldsymbol{o}\rangle : (s, r, o){\in}\mathrm{KG}\}$. By additionally multitasking RA with EA and KGC, AlignKGC achieves state-of-the-art performance for multilingual KGC and RA, and close to BERT-INT performance for EA.

4.8 Application to search and QA (30m)

We will describe and analyze the use and benefits of KG representations in search and QA: QA on KG alone [33, EmbedKGQA]; QA on a composite KG and corpus graph [32, AQQUCN] [38, PullNet] [29, UNIQORN]; question-guided reasoning paths [22, KVMNet] [2]; graphs derived from tables and text [45, 47]; and temporal QA: [34, CronKGQA] and [13, EXAQT].

[2] 'Instant' interpreted at some practical level of granularity.

5 SUPPORTING MATERIAL

The tutorial will provide the following material: (1) lecture video recording, (2) annotated slides and jamboards, (3) assignments, exams and projects from related courses taught, (4) extended bibliography, and (5) compendium of public software and data sets.

REFERENCES

[1] R. Abboud, I. I. Ceylan, T. Lukasiewicz, and T. Salvatori. BoxE: A box embedding model for knowledge base completion. *NeurIPS*, abs/2007.06267, 2020. URL https://arxiv.org/pdf/2007.06267.pdf.

[2] A. Asai, K. Hashimoto, H. Hajishirzi, R. Socher, and C. Xiong. Learning to retrieve reasoning paths over wikipedia graph for question answering. *ICLR*, abs/1911.10470, 2020.

[3] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD Conference*, pages 1247–1250, 2008. URL http://ids.snu.ac.kr/w/images/9/98/sc17.pdf.

[4] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013. URL http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf.

[5] M. Chen, Y. Tian, M. Yang, and C. Zaniolo. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *IJCAI*, 2017. URL https://arxiv.org/pdf/1611.03954.pdf.

[6] M. Cornolti, P. Ferragina, M. Ciaramita, S. Rued, and H. Schuetze. The SMAPH system for query entity recognition and disambiguation. In *ERD Challenge Workshop*, 2014. URL https://research.google.com/pubs/archive/42720.pdf.

[7] S. S. Dasgupta, S. N. Ray, and P. P. Talukdar. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In *EMNLP Conference*, 2018. URL https://www.aclweb.org/anthology/D18-1225.pdf.

[8] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391–407, 1990. URL https://www.cs.bham.ac.uk/~pxt/IDA/lsa_ind.pdf.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL Conference*, 2019. URL https://www.aclweb.org/anthology/N19-1423.pdf.

[10] O.-E. Ganea, G. Bécigneul, and T. Hofmann. Hyperbolic entailment cones for learning hierarchical embeddings. *arXiv preprint arXiv:1804.01882*, 2018. URL https://arxiv.org/pdf/1804.01882.pdf.

[11] R. Goel, S. M. Kazemi, M. A. Brubaker, and P. Poupart. Diachronic embedding for temporal knowledge graph completion. In *AAAI Conference*, 2020. URL https://arxiv.org/pdf/1907.03143.pdf.

[12] P. Jain, S. Rathi, Mausam, and S. Chakrabarti. Temporal knowledge base completion: New algorithms and evaluation protocols. *EMNLP Conference*, abs/2005.05035, 2020. URL https://arxiv.org/abs/2005.05035.

[13] Z. Jia, S. Pramanik, R. S. Roy, and G. Weikum. Complex temporal question answering on knowledge graphs. *CIKM*, 2021. URL https://dl.acm.org/doi/pdf/10.1145/3459637.3482416.

[14] M. Joshi, U. Sawant, and S. Chakrabarti. Knowledge graph and corpus driven segmentation and answer inference for telegraphic entity-seeking queries. In *EMNLP Conference*, pages 1104–1114, 2014. URL http://www.emnlp2014.org/papers/pdf/EMNLP2014117.pdf. Data http://bit.ly/1OCKbVW.

[15] V. Krishnan, S. Das, and S. Chakrabarti. Enhanced answer type inference from questions using sequential models. In *EMNLP/HLT*, pages 315–322, 2005. URL http://www.aclweb.org/anthology/H05-1040.

[16] S. Kulkarni, A. Singh, G. Ramakrishnan, and S. Chakrabarti. Collective annotation of Wikipedia entities in Web text. In *SIGKDD Conference*, pages 457–466, 2009. URL http://www.cse.iitb.ac.in/~soumen/doc/CSAW/.

[17] T. Lacroix, G. Obozinski, and N. Usunier. Tensor decompositions for temporal knowledge base completion. In *ICLR*, 2020. URL https://openreview.net/forum?id=rke2P1BFwS.

[18] X. Li, D. Roth, and K. Small. The role of semantic information in learning question classifiers. In *International Joint Conference on Natural Language Processing (IJCNLP)*, 2004. URL http://l2r.cs.uiuc.edu/~danr/Papers/LiRoSm04.pdf.

[19] X. L. Li, L. Vilnis, D. Zhang, M. Boratko, and A. McCallum. Smoothing the geometry of probabilistic box embeddings. In *ICLR*, 2019. URL https://openreview.net/forum?id=H1xSNiRcF7.

[20] J. J. Lin, R. Nogueira, and A. Yates. Pretrained transformers for text ranking: Bert and beyond. *WSDM Conference*, 2021. URL https://dl.acm.org/doi/pdf/10.1145/3437963.3441667.

[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119, 2013. URL https://goo.gl/x3DTzS.

[22] A. H. Miller, A. Fisch, J. Dodge, A.-H. Karimi, A. Bordes, and J. Weston. Key-value memory networks for directly reading documents. In *EMNLP Conference*, 2016. URL https://www.aclweb.org/anthology/D16-1147.pdf.

[23] G. Miller, R. Beckwith, C. FellBaum, D. Gross, K. Miller, and R. Tengi. Five papers on WordNet. Princeton University, Aug. 1993. URL https://wordnet.princeton.edu/.

[24] D. Milne and I. H. Witten. Learning to link with Wikipedia. In *CIKM*, pages 509–518, 2008. URL http://www.cs.waikato.ac.nz/~dnk2/publications/CIKM08-LearningToLinkWithWikipedia.pdf.

[25] B. Mitra and N. Craswell. An introduction to neural information retrieval. *Foundations and Trends in Information Retrieval*, 13:1–126, 2018. URL https://www.nowpublishers.com/article/Details/INR-061.

[26] M. Nickel and D. Kiela. Poincaré embeddings for learning hierarchical representations. In *NIPS*, pages 6341–6350, 2017. URL https://arxiv.org/pdf/1705.08039.pdf.

[27] M. Nickel, L. Rosasco, T. A. Poggio, et al. Holographic embeddings of knowledge graphs. In *AAAI Conference*, pages 1955–1961, 2016. URL https://arxiv.org/abs/1510.04935.

[28] J. Pennington, R. Socher, and C. D. Manning. GloVe: Global vectors for word representation. In *EMNLP Conference*, volume 14, pages 1532–1543, 2014. URL http://www.emnlp2014.org/papers/pdf/EMNLP2014162.pdf.

[29] S. Pramanik, J. O. Alabi, R. S. Roy, and G. Weikum. UNIQORN: Unified question answering over rdf knowledge graphs and natural language text. *ArXiv*, abs/2108.08614, 2021. URL https://arxiv.org/pdf/2108.08614.pdf.

[30] D. Savenkov and E. Agichtein. When a knowledge base is not enough: Question answering over knowledge bases with external text data. In *SIGIR Conference*, pages 235–244, 2016. URL https://dl.acm.org/citation.cfm?id=2911536.

[31] U. Sawant and S. Chakrabarti. Learning joint query interpretation and response ranking. In *WWW Conference*, Brazil, 2013. URL http://arxiv.org/abs/1212.6193.

[32] U. Sawant, S. Garg, S. Chakrabarti, and G. Ramakrishnan. Neural architecture for question answering using a knowledge graph and Web corpus. *Information Retrieval Journal*, 22:324–349, 2018. URL http://arxiv.org/pdf/1706.00973.

[33] A. Saxena, A. Tripathi, and P. P. Talukdar. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In *ACL Conference*, 2020. URL https://www.aclweb.org/anthology/2020.acl-main.412.pdf.

[34] A. Saxena, S. Chakrabarti, and P. Talukdar. Question answering over temporal knowledge graphs. *ACL Conference*, 2021. URL https://arxiv.org/pdf/2106.01515.

[35] H. Singh, S. Chakrabarti, P. Jain, S. R. Choudhury, and Mausam. Multilingual knowledge graph completion with joint relation and entity alignment. In *AKBC Conference*, volume abs/2104.08804, 2021. URL https://openreview.net/forum?id=FTzU__68rp8.

[36] S. Subramanian and S. Chakrabarti. New embedded representations and evaluation protocols for inferring transitive relations. In *SIGIR Conference*, 2018. URL https://gitlab.com/soumen.chakrabarti/rectangle.

[37] F. M. Suchanek, G. Kasneci, and G. Weikum. YAGO: A core of semantic knowledge unifying WordNet and Wikipedia. In *WWW Conference*, pages 697–706. ACM Press, 2007. URL http://www2007.org/papers/paper391.pdf.

[38] H. Sun, T. Bedrax-Weiss, and W. W. Cohen. PullNet: Open domain question answering with iterative retrieval on knowledge bases and text. *EMNLP Conference*, abs/1904.09537, 2019. URL https://www.aclweb.org/anthology/D19-1242.pdf.

[39] Z. Sun, Z. Deng, J.-Y. Nie, and J. Tang. RotatE: Knowledge graph embedding by relational rotation in complex space. *ICLR*, abs/1902.10197, 2019. URL https://arxiv.org/pdf/1902.10197.pdf.

[40] X. Tang, J. Zhang, B. Chen, and C. Li. BERT-INT: A BERT-based interaction model for knowledge graph alignment. In *IJCAI*, 2020. URL https://www.ijcai.org/proceedings/2020/0439.pdf.

[41] T. P. Tanon, D. Vrandečić, S. Schaffert, T. Steiner, and L. Pintscher. From Freebase to Wikidata: The great migration. *WWW Conference*, 2016. URL http://dl.acm.org/ft_gateway.cfm?id=2874809&type=pdf.

[42] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon. Representing text for joint embedding of text and knowledge bases. In *EMNLP Conference*, pages 1499–1509, 2015. URL https://www.aclweb.org/anthology/D/D15/D15-1174.pdf.

[43] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *ICML*, pages 2071–2080, 2016. URL http://arxiv.org/abs/1606.06357.

[44] I. Vendrov, R. Kiros, S. Fidler, and R. Urtasun. Order-embeddings of images and language. *arXiv preprint arXiv:1511.06361*, 2015. URL https://arxiv.org/pdf/1511.06361.

[45] F. Wang, K. Sun, M. Chen, J. Pujara, and P. Szekely. Retrieving complex tables with multi-granular graph representation learning. In *SIGIR Conference*, pages 1472–1482, 2021. URL https://dl.acm.org/doi/pdf/10.1145/3404835.3462909.

[46] Y. Wu, X. Liu, Y. Feng, Z. Wang, R. Yan, and D. Zhao. Relation-aware entity alignment for heterogeneous knowledge graphs. *IJCAI*, 2019. URL https://arxiv.org/pdf/1908.08210.pdf.

[47] V. Zayats, K. Toutanova, and M. Ostendorf. Representations for question answering from documents with tables and text. In *EACL Conference*, 2021. URL https://www.aclweb.org/anthology/2021.eacl-main.253.pdf.

[48] Y. Zhu, H. Liu, Z. Wu, and Y. Du. Relation-aware neighborhood matching model for entity alignment. In *AAAI Conference*, 2021. URL https://arxiv.org/pdf/2012.08128.pdf.