Embedding Edge-attributed Relational Hierarchies

Muhao Chen*
University of California
Los Angeles, CA
muhaochen@ucla.edu

Chris Quirk
Microsoft Research
Redmond, WA
chrisq@microsoft.com

ABSTRACT

Relational embedding methods encode objects and their relations as low-dimensional vectors. While achieving competitive performance on a variety of relational inference tasks, these methods fall short of preserving the hierarchies that are often formed in existing graph data, and ignore the rich edge attributes that describe the relation facts. In this paper, we propose a novel embedding method that simultaneously preserve the hierarchical property and the edge information in the edge-attributed relational hierarchies. The proposed method preserves the hierarchical relations by leveraging the non-linearity of hyperbolic vector translations, for which the edge attributes are exploited to capture the importance of each relation fact. Our experiment is conducted on the well-known Enron organizational chart, where the *supervision* relations between employees of the Enron company are accompanied with email-based attributes. We show that our method produces relational embeddings of higher quality than state-of-the-art methods, and outperforms a variety of strong baselines in reconstructing the organizational chart.

ACM Reference Format:
Muhao Chen and Chris Quirk. 2019. Embedding Edge-attributed Relational Hierarchies. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331278

1 INTRODUCTION

Hierarchical relational data represents an important type of relational data that models multiple-level or tree-structured relations among real-world objects. Different forms of such data, including ontological taxonomies [10], geographic thesaurus [32], and organizational charts [16], are at the core of many application scenarios for knowledge management and information retrieval [1].

In this paper, we pay attention to the embedding learning for hierarchical data [8, 15, 19, 20]. Corresponding approaches typically encode objects from the hierarchies in low-dimensional embedding spaces, which are similar to those produced by multi-relational embedding (MRE) approaches [3, 21]. While unlike regular MRE methods that capture simple relations as vector operations between object embeddings, the hierarchical relation embeddings (HRE) require non-trivial learning techniques to help preserve the hierarchical property of the data. Such techniques include neighborhood constraints on finer objects that are associated with the same coarser one [8, 15, 19], and manifold learning that fits tree structured data in the non-linear embedding spaces [20, 28]. These HRE methods offer competitive performance on relational inference tasks.

While existing MRE and HRE approaches typically consider relations as labeled or unlabeled directed edges between objects, in real-world cases, such relations are often complex and containing much richer information than just a simple label. Consider that the employees in an engineering company not only just constitute the supervision relations, but can also share side information from email logs and project collaboration records. Such auxiliary information naturally provides edge attributes that imply the relations of employees by reflecting their communication behaviors. Ideally, the HRE model should capture both the hierarchical relations, and the edge attributes that induce the plausibility of the relation facts.

We develop a novel HRE method that leverages both the structure information and edge attributes to learn better representations of the edge-attributed relational hierarchies. Our method combines two model components to learn on two facets of the data. A *hyperbolic relational embedding model* extends the relational embedding techniques to a hyperbolic space that well suits the embedding of hierarchical structures. On top of that, an *edge attribute model* seeks to enhance the relational learning by aggregating edge attributes to infer the confidence of each relation fact. We apply our model to represent the Enron organizational hierarchies [16], where a comprehensive set of edge-attributes are extracted from the emails between Enron employees. Experimental results demonstrate that our method learns better representations of the hierarchical data when compared with a variety of MRE and HRE models, as evaluated on the challenging task of hierarchical relation prediction.

2 RELATED WORK

Multi-relational embeddings. Extensive efforts have been put to MREs. Given each relation fact (s, r, t) that models the relation r of the source and target objects s and t, MRE methods seek to capture the relation fact with a plausibility function $f_r(s, t)$. A recent survey [29] categorizes the majority of MRE methods into translational methods and similarity-based methods. Translational methods follow a common assumption $\mathbf{s}_r + \mathbf{r} \approx \mathbf{t}_r$, where $\mathbf{s}_r$ and $\mathbf{t}_r$ are either the original vectors of s and t [3], or the transformed vectors via a transformation specific to r [8, 18, 30]. Translational methods perform well on characterizing unidirectional relations, and are more robust against the sparsity of the structure [25]. Similarity-based models are more suitable for characterizing dense structures. Representative methods adopt Hadamard multiplication [31], circular correlation [21], and bilinear tensor factorization [22]. MRE methods have been used to predict missing relation facts in knowledge

*This work is done in Microsoft Research, Redmond.

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331278

graphs [7, 21], commonsense ontologies [8], co-purchase records [14] and drug interaction data [27].

More recent works propose HRE methods for the characterization of hierarchical structures, where regular MRE methods fall short. To better represent the hierarchies, some methods introduce hierarchical neighborhood constraints in node-proximity-based [15] or translational models [8, 19], which aims at embedding sibling objects (nodes) of a hierarchy closely in the Euclidean space. Other works extend MRE methods into hyperbolic spaces, and seek to use the non-linear hyperbolic distances to characterize the nodes in hierarchies [20, 28]. While these techniques effectively preserve the hierarchical properties, they do not incorporate the edge attribute information, which is demonstrated to be significant in improving representation learning of the relational hierarchies.

Embedding edge-attributed graphs. Fewer works have incorporated edge attributes into graph embeddings [5, 6, 9, 13, 26]. Goyal et al. [13] propose the use of edge attribute proximities to enhance the learning on node proximities in unlabeled graphs, and Shah et al. [26] capture the distributions of attribute values to perform edge-level anomaly detection. These methods are proposed for unlabeled graph structures where unidirectional relations are not considered for objects. Chen et al. [5, 6] incorporate various edge labels into graph embeddings, and perform label propagation to address semi-supervised relation inferences. Similarly, a probabilistic score is captured and propagated across uncertain knowledge graph embeddings following pre-defined probabilistic soft logic rules [9]. While these works incorporate side information of relations into MRE, they are only able to capture either a simple label or a score for each edge, and do not leverage comprehensive edge attributes that affect the proximity and confidence of relation facts. Most importantly, these works do not characterize hierarchical structures.

3 METHOD

We first provide the definition of edge-attributed hierarchies. We use O and R to denote the sets of objects and relations respectively. The data we seek to model is an edge-labeled graph G, where each edge $T = (s, r, t) \in G$ is a relation fact that marks a relation r between a source object s and a target object t, such that $r \in R$ and $s, t \in O$. Boldfaced $\mathbf{s}$, $\mathbf{r}$, $\mathbf{t}$ denote the embedding vectors of corresponding units. G forms hierarchies, such that each r is either a one-to-many or a many-to-one relation. A function $A : O \times O \rightarrow \mathbb{R}^k$ assigns attributes to object pairs, where k is the number of edge attributes.

Our method learns two model components on two facets of the edge-attributed hierarchies respectively. (i) A relational embedding model seeks to capture the relation facts from the hierarchical structure in the hyperbolic embedding space, and (ii) an edge attribute model utilizes edge attributes to predict the confidence of relation facts, which is used to refine the learning process of the relational embedding model. We hereby describe these model components and the learning objective in detail.

3.1 Relational Embedding Model

3.1.1 Hyperbolic Embeddings. A hyperbolic space is a non-Euclidean space of negative curvature [17], where the distance between two vectors grows rapidly with regard to their relative distances to the origin. This property has been key to the embedding of hierarchical structures, and have been leveraged to capture the binary proximity of nodes on unlabeled networks with hierarchical substructures [20, 28]. We extend HRE techniques into the hyperbolic space, so as to support the relational inferences of objects that form hierarchies.

To support the parameterization and optimization of hyperbolic embeddings in the same way of Euclidean ones, we adopt the Poincaré ball model [20]. The Poincaré ball model aims at embedding a hyperbolic space inside a Euclidean unit hyper-ball. Specifically, it is defined as a Riemannian manifold $\mathcal{P} = (\mathcal{B}^n, g_x)$, for which $\mathcal{B}^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1\}$ is an open space enclosed by an n-dimensional hyper-sphere, and $\|\cdot\|$ denotes the Euclidean norm. $g_x = \left(\frac{2}{1-\|\mathbf{x}\|^2}\right)^2 g_e$ is a Riemannian metric tensor for the Poincaré ball model, where g_e denotes the Euclidean metric tensor.

The distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ on $\mathcal{P}$ is measured by:

$$d_p(\mathbf{u}, \mathbf{v}) = \cosh^{-1}\left(1 + 2\frac{\|\mathbf{u}-\mathbf{v}\|^2}{\left(1-\|\mathbf{u}\|^2\right)\left(1-\|\mathbf{v}\|^2\right)}\right)$$

The above equation shows the locality property of the hyperbolic distance. Given any two vectors $\mathbf{u}$ and $\mathbf{v}$ with a fixed Euclidean distance $d' = ||\mathbf{u}-\mathbf{v}||$, their hyperbolic distance grows rapidly along with their Euclidean norm. This demonstrates a desired property for preserving relational hierarchies. That is to say, the root of a hierarchy can be placed near the origin, where distances to nearby nodes are relatively small. Meanwhile, the leaves can be placed close to the external boundary of $\mathcal{P}$, where the much larger distance growth allows the distribution of many fine-grained objects.

3.1.2 Relational Embedding Techniques. To embed a triple (s, r, t) in the structure, a cost function $f_r(s, t)$ is used to measure its plausibility. A lower cost indicates a more plausible triple. We adopt two representative techniques in the defined hyperbolic space, i.e. translations (TransE [3]) and circular correlation (HolE [21]). The cost functions are given as follows, where $\star : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ denotes circular correlation defined as $[\mathbf{a} \star \mathbf{b}]_k = \sum_{i=0}^{d} a_i b_{(k+i) \mod d}$.

$$f_r^{\text{Trans}}(\mathbf{s}, \mathbf{t}) = ||\mathbf{s} + \mathbf{r} - \mathbf{t}||_2$$
$$f_r^{\text{HolE}}(\mathbf{s}, \mathbf{t}) = -(\mathbf{s} \star \mathbf{t}) \cdot \mathbf{r}$$

Besides these two techniques, we have discovered that the Hadamard-product-based technique [31] does not perform well on $\mathcal{B}$. Other techniques that introduce additional parameters, including relation-specific projections [8, 18], bilinear mapping [22] and neural approaches [11], are left as future work due to that they need non-trivial adaptation to gradient computation in the hyperbolic space.

3.2 Characterizing Edge Attributes

The edge attributes represent an alternative view of a relation between two objects. Naturally, we can assess the confidence of a relation fact based on the associated edge attributes. Such confidence estimations are supported with a relation-specific function $\psi_r : \mathbb{R}^k \rightarrow [0, 1]$. Our edge attribute model is defined as a set of regressors, each represents a different ψ_r. The training process of a regressor iterates on object pairs (s, t) in a subset of $O \times O$. Given attributes $A(s, t)$ and a relation r, ψ_r is fitted with 1 if $(s, r, t) \in G$, otherwise 0. The confidence scores given by ψ_r is then incorporated to refine the learning process of relation facts as described below.

3.3 Learning Objectives

The learning objective of our model is to minimize the following marginal ranking loss,

$$J = \sum_{(s,r,t)\in G \wedge (s',r,t')\notin G} \Big[\gamma + \Big(\psi_r(A(s,t)) + \beta\Big) \cdot f_r(s,t) - \Big(\psi_r(A(s',t')) + \beta\Big) \cdot f_r(s',t')\Big]_+$$

where γ is a positive margin, f_r is a triple cost functions defined in Section 3.1.2, and (s', r, t') is a negative sampled triple generated by randomly corrupting either s or t following previous methods [21, 31]. Here, ψ_r serves as a weight over the triple cost function f_r. A higher confidence given by ψ_r magnifies the cost f_r and lead to higher gradient for the corresponding training case of triple. Hence, the model is trained towards the goal, where a triple with higher edge-attributed-based confidence is more likely to be measured higher plausibility (i.e. lower f_r). β is a positive bias which controls the contribution of ψ_r to f_r. Note that the parameters of ψ_r are fixed during the optimization of J.

Gradient conversion. We optimize the parameters using RSGD [20]. Each epoch t of training updates the parameters by $\theta_{t+1} \leftarrow \theta_t - \eta\nabla_{\mathcal{P}} J(\theta_t)$, where η is the learning rate. $\nabla_{\mathcal{P}} J(\theta_t)$ is the Riemannian gradient on the Poincaré ball $\mathcal{P}$, which is easily converted from the Euclidean gradient ∇_E by

$$\nabla_{\mathcal{P}} = \frac{(1 - ||\theta_t||^2)^2}{4} \nabla_E$$

The optimization is subject to the following norm clipping constraint to retract parameters on $\mathcal{B}$, where ϵ is a small positive constant that is set as 10^{-6}:

$$\text{clip}(\theta_{t+1}) = \begin{cases} \frac{\theta_{t+1}}{||\theta_{t+1}||} - \epsilon & \text{if } ||\theta_{t+1}|| \geq 1 \\ \theta_{t+1} & \text{otherwise} \end{cases}$$

4 EXPERIMENT

In this section, we present the detailed experimental settings. We focus on predicting the hierarchical relation facts in the Enron corpus [16], and compare our model against various baselines.

4.1 Dataset

Our experiment is based on the organization chart of core employees preprocessed by Prabhakaran et al. [24], which contains 4,796 relation facts of 3,187 objects (2,585 employees and 602 department units). There are 1,503 relation facts regarding supervision of employees (an employee *supervises* another), while the rest constitutes relation facts regarding management (an employ *manages* a unit) and affiliation (a unit *contains* an employee). The supervising relations thereof, naturally form large hierarchical structures, of which the prediction is key to the reconstruction of an incomplete organizational chart [2, 4, 12]. We randomly split-off 30% of supervision relation facts, where 5% and 25% are for validation and test respectively. The rest of the structure is used for training. The goal of the hierarchy reconstruction is to predict the correct immediate supervisor in each split-off relation fact, considering that every employee has one immediate supervisor in the organizational chart. Note that, the employees in the split-off relation facts appear at least once in the training set, hence there is no zero-shot test case.

Attribute	Description
`#Sentences`	The average number of sentences in emails sent from o_1 to o_2,
`P1to1`	The proportion of emails that are sent to o_2 by o_1 without including other recipients.
`PReply`	The proportion of emails from o_1 that are replied by o_2.
`#Recipients`	The average number of recipients of emails that are sent from o_1 to o_2, including cc'ed.
`#To`	The average number of recipients of emails that are sent from o_1 to o_2, excluding cc'ed.
`PInit`	The proportion of email sessions between o_1 and o_2 that are initiated by an email from o_1.
`PFw`	The proportion of forwarded emails among all those that are sent from o_1 to o_2.
`PRe`	The proportion of replied emails among all those that are sent from o_1 to o_2.
`#Lex`	The average count of lexicons in emails from o_1 to o_2.
`PHasGreetings`	The proportion of emails sent from o_1 to o_2 that are with greeting titles (such as "Dear" or "Hi").
`EoP`	The summed weight per email message from o_1 to o_2 based on the express-of-power words extracted by Gilbert [12].

Table 1: Descriptions of email-based edge attributes.

Besides relation facts, the corpus also contains 127,083 unique email messages among employees. We extract and aggregate 11 email-based attributes defined by Prabhakaran et al. [24] and Gilbert [12] for each pair of employees in the hierarchies. Given two employees o_1 and o_2, these attributes are described in Table 1.

These attributes are expected to reflect the supervision relations of employees. Note that such attributes have been used to characterize the dominance relations between employees[1] [23, 24]. However, we find those attributes alone to be not effective in predicting the supervision relation facts that describe the *immediate* supervisors.

4.2 Baselines and Model Configurations

We compare two variants of our models which employ the two relation embedding techniques in Section 3.1.2 respectively. Besides, we compare with the following four groups of baselines.

- `MRE methods`: These include Euclidean MRE methods TransE [3], DistMult [31], HolE [21], On2Vec [8] and RESCAL [22].
- `MRE methods with edge attributes`: These include two of the above MRE methods incorporated with the edge attribute model using our learning objective in Section 3.3.
- `Hyperbolic embedding methods`: These are formed by extending the above MRE methods in the Poincaré ball, without the consideration of edge attributes.
- `Degree centrality`: A baseline method has been used to predict the partial-order dominance of employees based on the degree centrality of nodes [2]. To predict the immediate supervisor of an employee, it retrieves the nodes in the ascending order of degree centrality, starting from the one that has the degree that is immediately larger than the node of that employee.

We initialize the parameters of our model and those of the hyperbolic embedding baselines from the uniform distribution $\mathcal{U}(-0.001, 0.001)$, and enforce the norm constraint defined in Section 3.3. The Euclidean embedding baselines employ their default initialization and regularization process. We search the hyperparameters based on the validation set, for which dimensionality k is searched among $\{15, 25, 50, 75\}$, margin γ among $\{0.25, 0.5, 1.0\}$, and bias β among

[1] o_1 dominates o_2, if o_1 has a higher position than o_2 in the report chain, while o_1 is not necessarily immediate (nearest) supervisor of o_2.

Model	$Hits@1$	$Hits@5$	MRR
Deg centrality [2]	0.23	1.90	0.019
TransE [3]	23.08	50.96	0.360
DistMult [31]	8.88	17.75	0.134
HolE [21]	13.23	17.14	0.136
On2Vec [8]	15.69	23.71	0.134
RESCAL [22]	2.88	5.77	0.183
TransE+Attr	39.42	**65.38**	0.491
HolE+Attr	17.31	23.08	0.200
Hyper-TransE	53.85	58.65	0.538
Hyper-HolE	14.42	29.81	0.213
Hyper-TransE+Attr	**54.75**	59.51	**0.548**
Hyper-HolE+Attr	16.84	31.28	0.220

Table 2: Results of hierarchy reconstruction.

$\{0.25, 0.5, 1.0\}$. Eventually, we set $k = 25$ for translational and HolE-based models, $k = 50$ for DistMult, and $k = 15$ for RESCAL. $\beta = 1.0$ and $\gamma = 1.0$ are adopted for Euclidean baselines, $\beta = 0.5$ and $\gamma = 0.25$ for hyperbolic ones. The learning rate η is fixed as 0.01. We set the ψ_r for the supervision relation as a linear SVM. For training ψ_r, we preserve all records of email attributes for relation facts in the train set as positive cases, and randomly sample the same amount of negative cases for employee pairs that do not form the supervision relation fact in the train set.

4.3 Results

We aggregate three metrics on the test cases, accuracy ($Hits@1$), the proportion of correct answers ranked no larger than 5 ($Hits@5$), and mean reciprocal ranks (MRR). The results by four groups of baselines and two variants of our model are reported in Table 2.

The degree-centrality-based baseline, although promisingly predicts the partial-order dominance of employees [2], is however not suitable for predicting the immediate supervision relations. Among the Euclidean MRE techniques, TransE offers the best performance. This is attributed to that the vanilla translational technique is more robust against the sparsity of the structure in comparison to other similarity-based techniques [25]. Considering that relational hierarchies are often sparse [1], this suggests the translation to be a suitable MRE technique for HRE. By incorporating the edge attribute model, the edge-attribute-based confidence scores effectively strengthen the characterization of relation facts. Correspondingly, we observe a drastic improvement of 16.34% in $Hits@1$ and 0.131 in MRR by TransE+Attr over TransE. Meanwhile, the hyperbolic TransE (Hyper-TransE) outperforms the Euclidean one by 30.77% in $Hits@1$ and 0.178 in MRR. Eventually, our best model variant Hyper-TransE+Attr obtains the best performance, which outperforms TransE and Hyper-TransE by 31.67% and 0.9% in $Hits@1$ respectively, and by 0.188 and 0.010 in MRR respectively.

Hence, by combining both hyperbolic embeddings and the edge attribute model, our method is competent in modeling and reconstructing the edge-attributed hierarchies.

5 CONCLUSION

We propose an embedding method to capture edge-attributed relational hierarchies. Our method leverages two model components to improve the learning on corresponding data. The hyperbolic relational embedding seeks to capture the hierarchical structures that are not preserved by traditional MRE methods. On top of that, the edge-attribute model significantly improves the prediction of relation facts by incorporating edge-attribute-based confidence into the learning process. Our method has outperformed various baselines on reconstructing the Enron organizational hierarchies.

REFERENCES

[1] Aaron B Adcock, Blair D Sullivan, and Michael W Mahoney. 2013. Tree-like structure in large social and information networks. In *ICDM*.
[2] Apoorv Agarwal, Adinoyi Omuya, Aaron Harnly, and Owen Rambow. 2012. A comprehensive gold standard for the enron organizational hierarchy. In *ACL*.
[3] Antoine Bordes, Nicolas Usunier, et al. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*.
[4] Philip Bramsen, Martha Escobar-Molano, Ami Patel, and Rafael Alonso. 2011. Extracting social power relationships from natural language. In *ACL*.
[5] Haochen Chen, Xiaofei Sun, Yingtao Tian, et al. 2018. Enhanced Network Embeddings via Exploiting Edge Labels. In *CIKM*.
[6] Haochen Chen, Yingtao Tian, Bryan Perozzi, Muhao Chen, Xiaofei Sun, and Steven Skiena. 2019. Social Relation Inference via Label Propagation. In *ECIR*.
[7] Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Carlo Zaniolo. 2018. Co-training Embeddings of Knowledge Graphs and Entity Descriptions for Cross-lingual Entity Alignment. In *IJCAI*.
[8] Muhao Chen, Yingtao Tian, Xuelu Chen, Zijun Xue, and Carlo Zaniolo. 2018. On2Vec: Embedding-based Relation Prediction for Ontology Population. In *SDM*.
[9] Xuelu Chen, Muhao Chen, Weijia Shi, et al. 2019. Embedding Uncertain Knowledge Graphs. In *AAAI*.
[10] Gene Ontology Consortium. 2018. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research* 47, D1 (2018), D330–D338.
[11] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. *AAAI* (2018).
[12] Eric Gilbert. 2012. Phrases that signal workplace hierarchy. In *CSCW*.
[13] Palash Goyal, Homa Hosseinmardi, Emilio Ferrara, and Aram Galstyan. 2018. Capturing Edge Attributes via Network Embedding. *arXiv:1805.03280* (2018).
[14] Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-based recommendation. In *RecSys*.
[15] Zhiting Hu, Poyao Huang, Yuntian Deng, Yingkai Gao, and Eric Xing. 2015. Entity hierarchy embedding. In *ACL-IJCNLP*.
[16] Bryan Klimt and Yiming Yang. 2004. The enron corpus: A new dataset for email classification research. In *ECML*.
[17] Dmitri Krioukov, Fragkiskos Papadopoulos, et al. 2010. Hyperbolic geometry of complex networks. *Physical Review E* 82, 3 (2010).
[18] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion.. In *AAAI*.
[19] Xin Lv, Lei Hou, Juanzi Li, and Zhiyuan Liu. 2018. Differentiating Concepts and Instances for Knowledge Graph Embedding. In *EMNLP*.
[20] Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *NIPS*.
[21] Maximilian Nickel, Lorenzo Rosasco, et al. 2016. Holographic Embeddings of Knowledge Graphs. In *AAAI*.
[22] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A Three-Way Model for Collective Learning on Multi-Relational Data.. In *ICML*.
[23] Vinodkumar Prabhakaran, Premkumar Ganeshkumar, and Owen Rambow. 2018. Author Commitment and Social Power: Automatic Belief Tagging to Infer the Social Context of Interactions. In *NAACL*.
[24] Vinodkumar Prabhakaran, Emily E Reid, et al. 2014. Gender and power: How gender and gender environment affect manifestations of power. In *EMNLP*.
[25] Jay Pujara, Eriq Augustine, and Lise Getoor. 2017. Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short. In *EMNLP*.
[26] Neil Shah, Alex Beutel, Bryan Hooi, Leman Akoglu, Stephan Gunnemann, Disha Makhija, Mohit Kumar, and Christos Faloutsos. 2016. Edgecentric: Anomaly detection in edge-attributed networks. In *ICDM*.
[27] Gabriel Stanovsky, Daniel Gruhl, and Pablo Mendes. 2017. Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models. In *EACL*.
[28] Tran Dang Quang Vinh, Yi Tay, Shuai Zhang, Gao Cong, and Xiao-Li Li. 2018. Hyperbolic Recommender Systems. *arXiv:1809.01703* (2018).
[29] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *TKDE* (2017).
[30] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.. In *AAAI*.
[31] Bishan Yang, Wen-tau Yih, Xiaodong He, et al. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.
[32] Youjie Zhou and Jiebo Luo. 2012. Geo-location inference on news articles via multimodal pLSA. In *ACM MM*.