Rule Mining for Semantifying Wikilinks

Luis Galárraga, Danai Symeonidou, Jean-Claude Moissinac
Télécom ParisTech, Paris, France
{luis.galarraga, danai.symeonidou, jean-claude.moissinac}@telecom-paristech.fr

ABSTRACT

Wikipedia-centric Knowledge Bases (KBs) such as YAGO and DBpedia store the hyperlinks between articles in Wikipedia using *wikilink* relations. While wikilinks are signals of semantic connection between entities, the meaning of such connection is most of the times unknown to KBs, e.g., for 89% of wikilinks in DBpedia no other relation between the entities is known. The task of discovering the exact relations that hold between the endpoints of a wikilink is called *wikilink semantification*. In this paper, we apply rule mining techniques on the already semantified wikilinks to propose relations for the unsemantified wikilinks in a subset of DBpedia. By mining highly supported and confident logical rules from KBs, we can semantify wikilinks with very high precision.

1. INTRODUCTION

Some of the most prominent KBs such as DBpedia [1] or YAGO [19] build upon accurate information extraction on the semi-structured parts of Wikipedia articles such as infoboxes, Wikipedia categories and hyperlinks between articles, namely wikilinks. Even though wikilinks account for more than 25% of the non-literal facts in DBpedia, they are rarely exploited. Nevertheless, the fact that two entities are connected via a hyperlink accurately suggests a semantic connection between them. The goal of this paper is to discover the exact meanings of such connections.

Some wikilinks are already semantified in KBs. YAGO and DBpedia, for example, know that Barack Obama links to USA and is also a citizen and the President of that country. KBs can extract such information because it is usually available in the infoboxes; however if the information lies somewhere outside the infoboxes, KBs will not see it, leading to unsemantified wikilinks (see [9, 21] for automatic population of infoboxes from text). This is the case for 89% of wikilinks in DBpedia. For instance, the Wikipedia article of Barack Obama links to the article of the 2009 Nobel Prize, but DBpedia does not know that he won the Nobel Prize. In some other cases, the semantic connection encoded in a wikilink can be vague and opaque and even not modeled in the schema of the KB. For example, Obama's article also links to the articles for cocaine and ovarian cancer.

In this work, we show how to leverage the already semantified wikilinks to semantify the others. This is achieved by learning frequent semantic patterns from the relations in the KB and the wikilinks. If we observe that people often link to the countries where they come from, we can suggest that unsemantified wikilinks from people to countries convey a nationality relationship. This example also implies that the types of entities play an important role when semantifying wikilinks. For instance, the fact that France links to Spain suggests that the implicit relation carried by the wikilink holds between countries (or even places) and therefore discards any relation with an incompatible signature. If we assume that wikilinks between countries encode a trade partnership, we can formulate this pattern as a logical rule:

$$linksTo(x, y) \wedge is(x, Country) \wedge is(y, Country) \Rightarrow deals(x, y)$$

Given an unsemantified wikilink between two countries, this rule will predict that they must be trade partners. Such predictions could be proposed as candidate facts to populate KBs. Still, this application scenario would require the rules to have certain quality, i.e., they should be statistically significant and draw correct conclusions in most cases. This would avoid capturing noisy or irrelevant patterns and make wrong predictions.

The process of learning logical rules from structured data is known as Rule Mining. In this paper, we resort to a method called AMIE [6] to mine logical rules from KBs. We then use the rules to draw conclusions and compute a list of the most likely candidate meanings (relations) between the entities of unsemantified wikilinks. Using a straightforward inference method, we can discover meanings for 180K unsemantified wikilinks with very high precision.

In addition to the semantification of wikilinks, and to further emphasize their value, we discuss their effect in the task of rule mining. We observe that sometimes, they can increase the confidence of the obtained rules. For instance, assuming that a rule mining approach learns the rule:

$$currentMember(x, y) \Rightarrow team(x, y)$$

we observe that by requiring the existing of a wikilink between the entities:

$$linksTo(x, y) \wedge currentMember(x, y) \Rightarrow team(x, y)$$

we achieve higher confidence. This observation could be

Copyright is held by the owner/author(s).
WWW2015 Workshop: Linked Data on the Web (LDOW2015).

leveraged by data inference and link prediction approaches. It also provides additional insights about the KB.

2. RELATED WORK

Link prediction. The task of discovering semantic links between entities in KBs is often referred in the literature as *link prediction*. Due to the prominence of the Semantic Web, the problem has been extensively studied using multiple paradigms.

Statistical graphical models such as Bayesian Networks [5] and Markov Logic Networks (MLN) [17] offer a theoretically rigorous framework for data inference in KBs. Given a KB and a set of soft weighted rules expressed in first order logic, MLNs support multiple inference tasks such as probability calculation for queries and predictions, and MAP (Maximum a Posteriori) inference. The major drawback of such methods is that in the original formulation they do not scale to the size of current KBs. Nevertheless, there have been initiatives to extend the applicability of MLNs to large datasets [14].

Some approaches represent KBs as matrices or tensors [12, 13]. Under this paradigm, for instance, a KB can be represented as a three-dimensional tensor where the fact $r(x, y)$ is encoded as 1 in the cell with coordinates (r, x, y). Methods such as RESCAL [12], among others [13, 18] resort to tensor factorization and latent factor analysis on the matrix representation of the KB, in order to estimate the confidence of the missing cells, i.e., how likely the missing facts are true based on the latent features in the data. Even though the scores are often given a probabilistic interpretation, they are not probabilities in a strict sense. Unlike our approach, this line of methods does not rely on explicitly formulated rules to perform inference.

A third family of approaches [7, 20, 3] resorts to embedding models to formulate the link prediction problem. In [20], entities are represented as vectors in an embedding space, while relations are defined as transformations on those vectors, e.g., the transformation *nationality* maps the vector of Barack Obama to the vector of USA. Methods based on embedding methods are very effective at predicting values for functional relations, e.g., place of birth and still perform fairly well for one-to-many relations, e.g., children.

Unlike the previous methods, the approach proposed in [10] relies on a graph representation for KBs and applies random walks and path ranking methods to discover new facts in large KBs. In a similar fashion [11] mines frequent metapaths on data graphs, i.e., sequences of data types connected by labeled edges, and uses them to predict links between entities.

All the approaches mentioned so far tackle the link prediction problem in KBs in a general way. Our approach in contrast, has a more focused scope, since we aim at predicting semantic links for entities for which there exists a signal of semantic connection, namely a wikilink.

Wikilinks for type induction. Some approaches have leveraged the semantic value conveyed by wikilinks for the task of type inference in KBs. The work presented in [15] represents the set of wikilinks as a directed graph where each entity is replaced by its more specific type in the DBpedia type hierarchy. The method discovers frequent subgraph patterns on such graph. These are called Encyclopedic Knowledge Patterns (EKP). EKPs can be used to describe classes of entities and therefore predict the types for untyped entities, e.g., instances of soccer players will often link to instances of coaches and soccer leagues. While this method also makes use of the instance information to mine patterns, it does not aim to discover relations between entities. Thus, it does not make use of any other relations holding between the endpoints of wikilinks. In the same spirit, [16] builds upon EKPs and uses the instance information to map both entities and classes to a vector space. A similarity function on this space is used to compute the distance of an entity to the prototypical vector of classes and predict the types for untyped entities.

3. PRELIMINARIES

3.1 Rule Mining

Our proposal to semantify wikilinks relies on logical rules mined from a KB and its wikilinks. In this paper we use a logical notation to represent rules and facts in a KB, e.g., the fact that Angela Merkel is a citizen of Germany is expressed as *nationality*(Angela Merkel, Germany). An *atom* is a fact where at least one of the arguments of the relation is a variable, e.g., *nationality*(x, Germany). We say that an atom holds in a KB if there exists an assignment for the variables in the atom that results in a fact in the KB. Moreover,we say that two atoms are *connected* if they share at least one variable. The building blocks for logical rules are conjunctions of transitively connected atoms. For example, the rule that says that married couples have the same nationality can be expressed as:

$$nationality(x, y) \wedge spouse(x, z) \Rightarrow nationality(z, y)$$

This is a *Horn rule*. The left-hand side of the implication is a conjunction of connected atoms called the *body*, whereas the right-hand side is the *head*. In this paper, we focus on *closed* Horn rules, i.e., rules where each variable occurs in at least two atoms of the rule. Closed Horn rules always conclude concrete facts for assignments of the variables to values in the KB. If the KB knows *nationality*(Barack Obama, USA) and *spouse*(Barack Obama, Michelle Obama), our example rule will conclude *nationality*(Michelle Obama, USA). If the conclusion of a rule does not exist in the KB, we call it a *prediction*. Rule Mining approaches require a notion of counter-examples and precision for rules, to account for the cases where the rules err. In the next section we describe such notions as well as a method to learn closed Horn rules from potentially incomplete KBs.

3.2 AMIE

AMIE [6] is a system that learns closed Horn rules of the form:

$$B_1 \wedge \cdots \wedge B_n \Rightarrow r(x, y) \;\; \text{Abbrev.} \;\; \boldsymbol{B} \Rightarrow r(x, y)$$

AMIE assesses the quality of rules in two dimensions: statistical significance and confidence. The first dimension is measured by the *support* of the rule. This metric is defined according to the following formula:

$$supp(\boldsymbol{B} \Rightarrow r(x, y)) := \#(x, y) : \exists z_1, ..., z_m : \boldsymbol{B} \wedge r(x, y)$$

In other words, the support is the number of distinct assignments of the head variables for which the rule concludes a fact in the KB. Support is defined to be monotonic; given a rule, the addition of a new atom will never increase its support. Moreover, support is a measure of statistical evidence, thus, it does not gauge the precision of the rule, i.e., how often it draws correct or incorrect conclusions. This requires a notion of negative examples. Since KBs do not encode negative information, rule mining approaches resort to different assumptions to derive counter-evidence. Methods based on traditional association rule mining [8] resort to the Closed World Assumption (CWA). Under the CWA, any conclusion of the rule that is absent in the KB, is a counter-example. This mechanism, however, contradicts the Open World Assumption that KBs make. In constrast, AMIE uses the Partial Completeness Assumption (PCA) to deduce counter-examples. The PCA is the assumption that if a KB knows some r-values for an instance, then it knows **all** its values. If a rule predicts a second nationality for Barack Obama, knowing that he is American, the PCA will count such deduction as a counter-example. On other hand if the KB did not know any nationality for Obama, then such case would be disregarded as evidence, while the CWA would still count it as negative evidence. Notice that, the PCA is perfectly safe for functional relations, e.g., place of birth and still feasible for quasi-functions such as nationality.

The confidence of a rule under the PCA follows the formula:

$$pcaconf(\boldsymbol{B} \Rightarrow r(x,y)) := \frac{supp(\boldsymbol{B} \Rightarrow r(x,y))}{\#(x,y) : \exists z_1, \ldots, z_k, y' : \boldsymbol{B} \wedge r(x,y')}$$

The PCA confidence normalizes the support of the rule (number of positive examples) over the number of both the positive and the negative examples according to the PCA.

AMIE uses support and confidence as quality metrics for rules and the user can threshold on these metrics. In addition, AMIE implements a set of strategies to guarantee good runtime and rules of good quality. Examples of such strategies are *prune by support* and the *skyline technique*. To prune the search space efficiently, AMIE relies on the monotonicity of support, that is, once a rule has dropped below the given support threshold, the system can safely discard the rule and all its derivations with more atoms. The skyline technique, on the other hand, is an application of the Occam Razor principle: among a set of hypotheses with the same predictive power, the one with fewer assumptions (the simplest) should be preferred. If the system has already learned a rule of the form $\boldsymbol{B} \Rightarrow r(x,y)$ and then finds a more specific version of the rule, i.e., $\boldsymbol{B} \wedge r_n(x_n, y_n) \Rightarrow r(x,y)$, the more specific rule will be output only if it has higher confidence.

4. SEMANTIFYING WIKILINKS

Our approach to semantify wikilinks relies on the intuition that (a) wikilinks often convey a semantic connection between entities, (b) some of them are already semantified in KBs, (c) the types of the entities in the wikilink define the signature of its implicit relation and (d) the already semantified wikilinks can help us semantify the others. The already semantified wikilinks constitute our training set. From this training set, we mine a set of semantic patterns in the form of logical rules.

To justify our intuition, we look at the types of the endpoints of semantified wikilinks in DBpedia. We restrict our analysis to the classes *Person*, *Place* and *Organization*. Table 1 shows the most common relations holding between pairs of those entities for which there exists at least one wikilink. For example, we observe that when a person links to a place, in 56% of the cases, the person was born in that place. Similarly, when an organization links to a place, in 19% of the cases, this corresponds to its location. We also observe that in our dataset, 81% of the links for these classes are not semantified. Rule mining techniques can help us learn the patterns suggested by Table 1 and semantify more links. For example, the fact that organizations link to the places where they are located can be expressed as:

$$linksTo(x,y) \wedge is(x, Org) \wedge is(y, Loc) \Rightarrow location(x,y)$$

Such a rule would allow us to predict the relation *location* for unsemantified wikilinks between organizations and locations. This is a link prediction task and has a great value for web-extracted KBs such as YAGO or DBpedia.

We start by constructing a training set $\mathcal{K}$ from DBpedia 3.8[1] consisting of 4.2M facts and 1.7M entities, including people, places and organizations. We enhance this dataset with the type information about the entities, i.e., 8M *rdf:type* statements, and the wikilinks between those entities. Since we can only learn from already semantified wikilinks, we restrict the set of wikilinks to those where both endpoints participate in a relation in the data, i.e., $linksTo(a,b) \in \mathcal{K}$ iff $\exists\, r, r', x, y : (r(x,a) \vee r(a,x)) \wedge (r'(y,a) \vee r'(a,y))$. This procedure led us to a training set $\mathcal{K}$ with a total of 18M facts. We ran AMIE on this dataset and configured it to mine closed Horn rules of the form:

$$linksTo^{*}(x,y) \wedge \boldsymbol{B} \wedge is(x,C) \wedge is(y,C') \Rightarrow r(x,y)$$

where *linksTo* is an alias for *wikiPageWikiLink*, *linksTo** denotes either *linksTo* or $linksTo^{-1}$, "*is*" is a synonym for *rdf:type* and $\boldsymbol{B}$ is a conjunction of up to 2 atoms. We call them *semantification rules*. With support and PCA confidence thresholds 100 and 0.2 respectively, AMIE found 3546 semantification rules on the training set $\mathcal{K}$. Table 2 shows examples of those rules.

We then use the rules to draw predictions of the form $p := r(a,b)$, i.e., $r(a,b) \notin \mathcal{K}$. We restrict even further the set of predictions, by requiring the arguments to be the endpoints of unsemantified wikilinks, more precisely, $\nexists\, r' : r' \neq linksTo \;\wedge\; r'(a,b) \in \mathcal{K}$. Recall that those predictions may have a different degree of confidence depending on the confidence of the rules that are used to deduce them. Moreover, a prediction can in principle be deduced by multiple rules since AMIE explores the search space of rules in an exhaustive fashion. To take this observation into account, we define the *confidence* of a prediction p according to the following formula:

$$conf(p) := 1 - \prod_{i=1}^{|\mathcal{R}|} (1 - [\phi(R_i, p) \times pcaconf(R_i)]) \qquad (1)$$

where $\mathcal{R}$ is the set of semantification rules and $\phi(R_i, p) = 1$ if $R_i \vdash p$, i.e., if p is concluded from rule R_i; otherwise $\phi(R_i, p) = 0$. The rationale behind Formula 1 is that the more rules lead to a prediction, the higher the confidence on that prediction should be. The confidence is then defined as the probability that at least one of the rules R_i that concludes p applies. This can be calculated as 1 minus the probability that none of the rules holds. The latter probability is defined as the product of the probabilities that each rule in isolation does not hold, in other words $(1 - pcaconf(R_i))$. Formula 1 thus, makes two strong assumptions. First, it confers a probabilistic interpretation to the PCA confidence.

[1]We learn rules on DBpedia 3.8 to corroborate some of their predictions automatically in DBpedia 3.9

Domain	Range	Relation - % occurrences					
Person	Person	successor	18%	associatedBand	11%	associatedMusicalArtist	11%
Person	Place	birthPlace	56%	deathPlace	18%	nationality	8%
Person	Organization	team	53%	almaMater	8%	party	5%
Place	Place	isPartOf	29%	country	28%	location	13%
Place	Person	leaderName	42%	architect	32%	saint	12%
Place	Organization	owner	24%	tenant	16%	operatedBy	12%
Organization	Organization	sisterStation	18%	associatedBand	15%	associatedMusicalArtist	15%
Organization	Person	currentMember	22%	bandMember	20%	formerBandMember	20%
Organization	Place	location	19%	city	17%	hometown	13%

Table 1: Top-3 relations encoded in wikilinks between instances of Person, Place and Organization in DBpedia.

Rule	PCA. Conf.
$linksTo(x,y) \wedge parent(x,y) \wedge successor(y,x) \wedge is(x, Person) \wedge is(y, Person) \Rightarrow predecessor(x,y)$	1.0
$linksTo(x,y) \wedge picture(x,y) \wedge is(x, ArchitecturalStructure) \wedge is(y, PopulatedPlace) \Rightarrow location(x,y)$	0.94
$linksTo(y,x) \wedge owner(x,y) \wedge subsidiary(y,x) \wedge is(y, Co.) \wedge is(x, Co.) \Rightarrow owningCompany(x,y)$	1.0

Table 2: Some semantification rules mined by AMIE on DBpedia.

Precision@1	Precision@3
0.77 $\pm$ 0.10	0.67 $\pm$ 0.07

Table 3: Average MAP@1 and MAP@3 scores for semantification of wikilinks on DBpedia.

Second, it assumes that rules are independent events. While we do not claim these assumptions to be correct, they still provide a naive baseline to estimate the likelihood of facts without resorting to more sophisticated approaches for data inference. As we show later, such a naive estimator delivers satisfactory results in our scenario.

Given an unsemantified wikilink $l := linksTo(a, b)$, Formula 1 allows us to propose a list of candidate meanings for l. If among the set of predictions there are several facts of the form $r_i(a, b)$, then each relation r_i is a semantification candidate for l with confidence $conf(r_i(a, b))$. For each unsemantified link, we propose a list of semantification candidates sorted by confidence. Our procedure proposes relation candidates for 180K unsemantified wikilinks in the training set. Since, we can semantify only 1% of them by automatically checking our predictions in DBpedia 3.9, we evaluate the precision of our approach on a sample of 60 unsemantified wikilinks. We then evaluate the correcteness of their rankings of semantification candidates as follows: for each wikilink we count the number of correct candidates at top 1 and top 3 of the ranking, we then add up these counts and divide them by the total number of candidates at top 1 and top 3 respectively. This gives us an estimation of the precision of our approach. Table 3 shows the estimated precision values drawn from the sample as well as the size of the Wilson Score Interval [4] at confidence 95%. The results imply that, for example, the precision at top 1 for the whole set of wikilinks lies in the interval 77% $\pm$ 10% with 95% probability.

Table 4 shows some examples of wikilinks and the ranking of semantification candidates proposed by our approach. The number in parentheses corresponds to the confidence of the semantification candidate. The candidates evaluated as correct according to the our evaluation are in italics.

Rules without wikilink	857
Rules with wikilink	1509
Rules with confidence gain	1389
Weighted average gain (wag)	0.03
Rules with gain $\geq$ 0.1	139

Table 5: Statistics about rule mining with and without wikilinks.

5. WIKILINKS FOR RULE MINING

The skyline technique implemented in AMIE prevents the system from reporting low quality rules. If AMIE finds two rules $\boldsymbol{B} \Rightarrow r(x,y)$ and $\boldsymbol{B} \wedge r_n(x_n, y_n) \Rightarrow r(x,y)$ and the latter has lower confidence, the system will not output it because it is worse in all dimensions, i.e., it has also lower support. We therefore investigate the confidence gain carried by the addition of wikilink atoms in rules.

We first run AMIE on the DBpedia mapping-based triples. In a second run, we add the wikilinks to the mapping-based triples and instruct the system to mine, when possible, rules of the form $linksTo^*(x,y) \wedge \boldsymbol{B} \Rightarrow r(x,y)$, i.e., if the skyline technique does not prune the longer rule. In both cases, we set a threshold of 100 positive examples for support and no confidence threshold. We report our findings in Table 5. We observe that requiring the head variables to be connected via a wikilink increases the number of rules from 857 to 1509. This occurs because in the second run, AMIE sometimes mines versions of the rules with and without the $linksTo^*$ atom. In other words, for some rules the addition of a wikilink atom provides a confidence gain. This is the case for 1389 rules as Table 5 shows. We are interested in finding how much confidence gain is carried by those rules. Thus, we define the *gain* of a wikilink rule as a variant of the gain metric used in association rule mining [2]:

$$gain(R) := supp(R) \times (pcaconf(R) - pcaconf(R_{\neg linksTo}))$$

That is, the gain of a wikilink rule is the product of its support and the difference in confidence with respect to the rule without the $linksTo^*$ atom. Table 5 reports an average gain of 0.03. We find, however, that for 10% of rules,

WikiLink	Semantification candidates
Interstate 76 (west) → Colorado State Highway	*routeJunction* (1.0)
J. Bracken Lee → Herbert B. Maw	*predecessor* (1.0), parent(0.998), governor(0.882)
WHQX → WTZE	*sisterStation* (1.0)

Table 4: Some examples of semantification candidates for wikilinks. The correct candidates are in italics.

Rule	Δ-conf
$producer(x, y) \wedge recordLabel(x, y) \Rightarrow artist(x, y)$	0.34
$debutTeam(x, y) \Rightarrow team(x, y)$	0.28
$officialLanguage(x, y) \Rightarrow spokenIn(x, y)$	0.19

Table 6: Confidence gain for some rules when specialized with a *linksTo* atom on the head variables.

the gain can be higher than 0.1. We show some of those rules with their corresponding confidence gain in Table 6. It follows that, in the majority of cases, the wikilinks do not provide a significant confidence gain to rule mining in DBpedia. The reason lies on the fact that for 99% of the triples in the DBpedia mapping-based dataset, there is a wikilink between the arguments of the triples, that is, the addition of a wikilink atom does not provide additional information to the rule. On the other hand, for certain relations, the arguments are not sometimes not connected with a wikilink. This is the case for 100K triples. In such cases, the addition of a $linksTo^*$ atom may convey a confidence gain that can be used to improve the quality of the rules.

All our datasets and experimental results are available under `http://luisgalarraga.de/semantifying-wikilinks`.

6. CONCLUSIONS

While none of the major Wikipedia-centric KBs make further use of the wikilinks, in this work we have shown that they often encode latent relations between entities. Such relations may not be captured in KBs. We have shown that rule mining techniques and naive inference methods are a feasible alternative to accurately discover those implicit semantics. This wikilink semantification task can be seen as a particular case of the link prediction problem in KBs. With this work, we aim at turning the attention to the wikilinks, as they convey valuable information that can help improve the completeness of KBs.

7. ACKNOWLEDGMENTS

This work is supported by the Chair "Machine Learning for Big Data" of Télécom ParisTech and Labex DigiCosme (project ANR-11-LABEX-0045-DIGICOSME) operated by ANR as part of the program "In- vestissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02).

8. REFERENCES

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A nucleus for a web of open data. In *ISWC*, 2007.

[2] R. J. Bayardo. Mining the most interesting rules. pages 145–154, 1999.

[3] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, 2013.

[4] L. D. Brown, T. T. Cai, and A. DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 2001.

[5] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *IJCAI*, 1999.

[6] L. Galárraga, C. Teflioudi, K. Hose, and F. Suchanek. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. In *WWW*, 2013.

[7] A. García-Durán, A. Bordes, and N. Usunier. Effective blending of two and three-way interactions for modeling multi-relational data. In *ECML-PKDD*, 2014.

[8] B. Goethals and J. Van den Bussche. Relational Association Rules: Getting WARMER. In *Pattern Detection and Discovery*, volume 2447. Springer Berlin / Heidelberg, 2002.

[9] D. Lange, C. Böhm, and F. Naumann. Extracting structured information from wikipedia articles to populate infoboxes. In *CIKM*, 2010.

[10] N. Lao, T. Mitchell, and W. W. Cohen. Random walk inference and learning in a large scale knowledge base. In *EMNLP*, 2011.

[11] C. Meng, R. Cheng, S. Maniu, P. Senellart, and W. Zhang. Discovering meta-paths in large heterogeneous information networks. In *WWW*, 2015.

[12] M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. In *ICML*, 2011.

[13] M. Nickel, V. Tresp, and H.-P. Kriegel. Factorizing YAGO: Scalable machine learning for linked data. In *WWW*, 2012.

[14] F. Niu, C. Ré, A. Doan, and J. Shavlik. Tuffy: Scaling up statistical inference in markov logic networks using an rdbms. *VLDB Endowment.*, 2011.

[15] A. Nuzzolese, A. Gangemi, V. Presutti, and P. Ciancarini. Encyclopedic knowledge patterns from wikipedia links. In *ISWC*. 2011.

[16] A. G. Nuzzolese, A. Gangemi, V. Presutti, and P. Ciancarini. Type inference through the analysis of wikipedia links. In *LDOW*, 2012.

[17] M. Richardson and P. Domingos. Markov logic networks. *Mach. Learn.*, 62(1-2):107–136, Feb. 2006.

[18] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. In *KDD*, 2008.

[19] F. M. Suchanek, G. Kasneci, and G. Weikum. Yago: A Core of Semantic Knowledge. In *WWW*, 2007.

[20] Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, 2014.

[21] F. Wu and D. S. Weld. Autonomously semantifying wikipedia. In *CIKM*, 2007.