Path-based Semantic Relatedness on Linked Data and its use to Word and Entity Disambiguation

Ioana Hulpuş, Narumol Prangnawarat, Conor Hayes

Insight Centre for Data Analytics,
National University of Ireland, Galway (NUIG)
{first.last}@insight-centre.org

Abstract. Semantic relatedness and disambiguation are fundamental problems for linking text documents to the Web of Data. There are many approaches dealing with both problems but most of them rely on word or concept distribution over Wikipedia. They are therefore not applicable to concepts that do not have a rich textual description. In this paper, we show that semantic relatedness can also be accurately computed by analysing only the graph structure of the knowledge base. In addition, we propose a joint approach to entity and word-sense disambiguation that makes use of graph-based relatedness. As opposed to the majority of state-of-the-art systems that target mainly named entities, we use our approach to disambiguate both entities and common nouns. In our experiments, we first validate our relatedness measure on multiple knowledge bases and ground truth datasets and show that it performs better than related state-of-the-art graph based measures. Afterwards, we evaluate the disambiguation algorithm and show that it also achieves superior disambiguation accuracy with respect to alternative state-of-the-art graph-based algorithms.

1 Introduction

With the advancements in Linked Data, more and more graph-based (i.e. RDF) structured knowledge bases become available. Still, most of the digital content we produce as a society is in text format. Linking unstructured text to structured data is fundamental for leveraging the benefits of the vast amounts of knowledge (in text as well as in structured format) available.

In this paper, we tackle two strongly interdependent problems, semantic relatedness and disambiguation. The aim of semantic relatedness is to weight the semantic associations between pairs of concepts. The aim of entity and word-sense disambiguation, is to link strings in the text to the corresponding concepts in external knowledge bases (KBs). These problems are fundamental for the integration between text and structured data. The most important cue for disambiguation is the semantic relatedness between the concepts mentioned in a particular context (i.e., text), therefore the two problems are highly interdependent.

With respect to these two problems, with the exception of some very recent approaches, most systems use distributional semantics techniques and traditionally require detailed textual description of concepts. Furthermore, since the relatedness is distilled from vast amount of text documents, these approaches do not have the capability of extracting the explicit relations between concepts.

These limitations can be overcome by using knowledge-based systems. While the idea of using structured knowledge for assessing semantic relatedness can be tracked back more than fifty years, research is still needed in order to understand how the relatively new, very broad KBs like DBpedia, Freebase, YAGO, can be most effectively used. In this paper, we first introduce a novel graph-based relatedness measure that uses the paths in the KB in order to score the association between pairs of concepts. Afterwards, we propose a joint disambiguation approach that can use any path-based pairwise relatedness. Our experiments for assessing the quality of our relatedness measure show higher positive correlations to human judgements than the current state of the art. Similarly, our experiments for assessing the quality of disambiguation show that graph-based joint disambiguation produces superior results as compared to very recent alternative graph-based approaches.

1.1 Related Work on Semantic Relatedness

Semantic relatedness of entities has been heavily researched over the past couple of decades. Two main directions can be identified. The first one, which we call *corpus-based*, models entities as multi-dimensional vectors that are computed based on distributional semantics techniques [4, 9]. The *de facto* standard corpus is Wikipedia. The second direction, which we call *structure-based* or *graph-based* and which makes the focus of this paper, relies on a graph structured KB. Approaches of this type have been very prolific since the publication of WordNet [30, 29]. However, most WordNet based semantic relatedness measures rely on hierarchical relations (isA, broaderOf). The problem with such measures is that they cannot exploit other semantically rich properties of concepts in more complex KBs.

Other structure-based approaches use the network of Wikipedia pages formed by their hyperlink connections [18, 24, 7]. Their drawback is that they require pages that contain hyperlinks to the targeted concepts, or that the concepts themselves have corresponding pages. Furthermore, the hyperlinks do not provide any semantics to the relation between the source and target concepts.

Recent approaches that are motivated by Linked Data make use of the different types of relations that exist in structured KBs (i.e., DBpedia). Some of them suffer from the drawback of requiring domain adaptation, and focus on manually selected types of concepts and relations [15, 20]. Other measures are very restrictive, computing semantic similarity between either neighbouring concepts, or concepts connected through a single intermediate node by the same relation type [23].

The approach that is most related to ours is the very recent work of Schumacher and Ponzetto [27]. Like us, the authors automatically weight relations in the knowledge graph and use them to compute relatedness between concepts that are not directly connected. However, their weighting scheme considers information theoretic global measures for the relationship type and object, while our measures are local, specific to the targeted pair and therefore less computationally demanding. Furthermore, our local measures have the added advantage of requiring very little update overhead when the background KB changes, while the global ones require the update of all scores. In our work, we have compared all our methods to their approach and we report our findings in the Evaluation section of this paper.

We evaluate our approach with both DBpedia and Freebase, from three perspectives: (1) named entity (NE) relatedness; (2) common noun similarity; and (3)common noun relatedness. Our extensive evaluation sheds light not only on our measures, but also on the general use of path-based relatedness measures on the used knowledge graphs.

1.2 Related Work on Entity and Word-Sense Disambiguation

An important class of related methods to disambiguation is formed by the *centrality-based* approaches. For each ambiguous word, the selected sense is the one that has the highest graph centrality with respect to the candidate senses of the other words in context. They have mostly been used on WordNet [17, 21, 2]. A very recent centrality-based approach is AGDISTIS [32]. To the best of our knowledge, it is the only previous approach that achieves entity disambiguation by using only DBpedia knowledge. After finding the candidate sets for all the ambiguous named entities, AGDISTIS extracts a subgraph of DBpedia that contains all the candidate senses of all the targeted entities as well as their n-hop neighbours and the relations between them. Then, the HITS algorithm is run over the extracted subgraph and for each targeted entity, DBpedia concept that has the highest authority score is selected. All these *centrality-based* suffer from the drawback that the selection of the senses of entities and words is "infested" by the wrong candidate senses.

Another important research direction related to ours uses graph-based relatedness measures on a semantic network that is built on-the-fly on top of the words that make the definitions that describe the candidate senses [28, 6, 11]. A similar dependence to the text that describes senses is noticed in most of the other systems [16, 5, 8] that link to DBpedia, as they apply their algorithms on Wikipedia text. In this paper, we research novel graph-based methods that do not require textual description of senses.

2 Path-based Relatedness Measures on Knowledge Graphs

2.1 Preliminaries

In this section, we define and formalise the main concepts we refer to in this paper. By *knowledge graph* we refer to any graph used to represent knowledge about concepts and relations between them. A knowledge graph can be seen as a *property graph*, a graph whose nodes and edges have properties. Also, knowledge graphs are a superset of *multigraphs* because they can contain multiple edges between the same pair of nodes. An RDF KB is in this case also a knowledge graph. Given a triple of the form $< s, p, o >$, the predicate p and object o resources become nodes in the graph connected by an edge of type p.

Definition 1. *We define a* ***knowledge graph*** *as a directed graph* $G(V, E, \mathcal{T}, \tau)$*, where* V *represents the set of all vertices,* E *represents the set of all edges (that we also call relations), connecting vertices in* V*,* $\mathcal{T}$ *is the set of edge types, and* $\tau : E \rightarrow \mathcal{T}$ *is a function that maps every edge in* E *to a type in* $\mathcal{T}$*.*

Although the edges are directed, we consider that the reverse relations also hold and can be traversed. The assumption behind this decision is that all semantic relations can be considered to have a semantically sound inverse relation. We use $E^{\mp}$ to denote the set of edges in the graph united with the set of their reversed edges, and $\mathcal{T}^{\mp}$ to denote the set of relationship types united to the set of their reversed types.

Definition 2. *A* ***path*** $\mathcal{P}$ *through the knowledge graph* $G(V, E, \mathcal{T}, \tau)$ *is a sequence of nodes and relations* $n_1 \xrightarrow{\tau_1} n_2 \xrightarrow{\tau_2} ..., \xrightarrow{\tau_{K-1}} n_K$ *such that for every two consecutive nodes in the sequence,* n_{k-1}*,* n_k*, there exists an edge* $e \in E^{\mp}$ *of type* $\tau_{k-1} \in \mathcal{T}^{\mp}$*.*

Using these definitions, we now introduce the path-based relatedness measures that we analyse in this paper. We start with a baseline measure inspired from social network analysis and afterwards we describe in detail the measure that makes the main contribution of this paper.

2.2 Baseline - Katz Relatedness

The length of the shortest path between two nodes is a common way of measuring proximity between nodes in a graph. However, it lacks the ability to discriminate between the relatedness of many node pairs, for example, a node will be considered of equal relatedness to all its 2-hop neighbours. To better differentiate, other methods make use of more and longer paths than just the shortest. Here, we adapt Katz's [13] centrality measure that is commonly used in social network analysis. This centrality measure has inspired another previously proposed semantic relatedness measure [22]. The idea is that the effectiveness of a link between two nodes is governed by a known, constant probability, α. In case of a path made up of k nodes, the probability of the path is α^k. We use this idea in a relatedness measure, where the relatedness between two nodes is the accumulated score over the top-k shortest paths between them.

$$rel^{(k)}_{Katz}(x, y) = \frac{\sum\limits_{p \in SP^{(k)}_{xy}} \alpha^{length(p)}}{k} \tag{1}$$

where $SP^{(k)}_{xy}$ is the set of the top-k shortest paths between concepts x and y.

2.3 Exclusivity-based Relatedness

The rationale behind the previous relatedness measure is that the more and shorter relation paths between two nodes, the higher their relatedness. However, it has been long known that not all direct relationships weight the same. Manual assignment of weights based on relationship type is infeasible, given the great amount of relationship types in knowledge graphs (almost 14000 in Freebase and more than 1100 in DBpedia). Therefore, we must devise automatic ways of assessing the importance of individual direct relations.

At the core of our next suggested measure, is one main rationale: a relation between two concepts is stronger if each of the concepts is related through the same type of relation to fewer other concepts. We name this property of relations *exclusivity* and we formalise it in the following.

Definition 3. *Given an edge e of type τ between two adjacent nodes x and y, directed from x to y, we define the* ***exclusivity*** *of edge e as the probability that, if we randomly select an edge e' out of the set of all edges of type τ that exit node x and all edges of type τ entering node y, that edge e' is edge e. Formally,*

$$exclusivity(x \xrightarrow{\mathcal{T}} y) = \frac{1}{|x \xrightarrow{\mathcal{T}} *| + |* \xrightarrow{\mathcal{T}} y| - 1}; \tag{2}$$

where $|x \xrightarrow{\mathcal{T}} *|$ denotes the number of relations of type $\tau \in \mathcal{T}$ that exit node x, and $|* \xrightarrow{\mathcal{T}} y|$ denotes the number of relations of type $\tau \in \mathcal{T}$ that enter node y.

1 is subtracted from the denominator because the relation $x \xrightarrow{\mathcal{T}} y$ is otherwise counted twice, once for the relations of x and once for the relations of y. As of Formula 2, the exclusivity score of a relation lies inside the $(0, 1]$ interval, with value 1 being obtained when the targeted relation is the only relation of its type for both x and y.

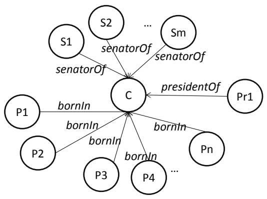

Fig. 1: Example exclusivity

Example 1. Let us look at the toy example in Figure 1, where we consider node C a country and all the other nodes, people. The exclusivity of the $bornIn$ relations is $1/n$, for the $senatorOf$ relations it is $1/m$, and for the $presidentOf$ relation exclusivity is 1. Naturally, n would be much higher than m, giving the $bornIn$ relation a smaller exclusivity than the $senatorOf$ and $presidentOf$ relations.

Since the exclusivity is computed for each individual relation, the *bornIn* relations to a small country will have a higher exclusivity than the *bornIn* relations to a bigger country. Extrapolating this measure to nodes that are not directly connected, people born in the same small country will be more related to each other than people born in a bigger country, and senators of a country will be more related to each other than random citizens born in the country.

An important property of our exclusivity property of relations is **symmetry**: $exclusivity(x \xrightarrow{\tau} y) = exclusivity(y \xrightarrow{\tau^-} x)$. Symmetry of exclusivity is crucial for consistency with our assumption that relations in the knowledge graph can be traversed in both directions.

Given a path through G, $\mathcal{P} = n_1 \xrightarrow{\tau_1} n_2 \xrightarrow{\tau_2}, ..., n_K$, with $\tau_i \in \mathcal{T}^{\mp}$ its weight can be computed by Formula 3.

$$weight(\mathcal{P}) = \frac{1}{\sum_i 1/exclusivity(n_i \xrightarrow{\tau_i} n_{i+1})}; \tag{3}$$

Then, given two nodes x and y we compute their relatedness as the sum of the path weights of the top-k paths with highest weight between them. In order to give preference to shorter paths, we introduce a constant length decay factor, $\alpha \in (0, 1]$. When $\alpha = 1$ longer paths are not penalised.

$$rel^{(k)}_{Excl}(x, y) = \sum_{\mathcal{P}_i \in P^{(k)}_{xy}} \alpha^{length(\mathcal{P}_i)} weight(\mathcal{P}_i); \tag{4}$$

This being the exclusivity based relatedness measure, we now move on to the problem of word-sense disambiguation and our proposed solution.

3 Joint Disambiguation on Knowledge Graphs

Joint disambiguation approaches treat disambiguation as a combinatorial optimisation problem. Given multiple ambiguous words, the correct senses for all words are selected simultaneously, by maximising a function of relatedness between the selected senses. Therefore, this methodology avoids the influence that wrong senses might have on the final solution. Having a context of n words, with each word w_i having m_i possible senses, a solution R contains n senses, one sense for each word. There are $\prod\limits_{i \in [1,n]} m_i$ solutions. We denote the set of all solutions as $\mathbf{R}$. A solution R^* is chosen that has the highest *coherence*. The most common way of computing the coherence of a solution R is by summing up all the pairwise relatedness scores of the senses in R, as shown in Formula (5). In this approach, the disambiguated sense s^*_w of word w is the sense of w that belongs to solution R^* as shown in Formula (6):

$$R^* = \arg\max_{R \in \mathbf{R}} \sum_{s \in R} \sum_{\substack{s' \in R;\\ s' \neq s}} rel(s, s'); \tag{5}$$

$$s^*_w = R^*[w]; \tag{6}$$

In Formula (5), the $rel(s, s')$ factor represents any pairwise relatedness measure. What sets graph-based joint disambiguation apart from other methods of joint disambiguation, is that $rel(s, s')$ is a graph-based measure. This problem is equivalent to the problem of finding the clique with the maximum sum of edge weights which is an NP-hard problem. We solve it by using the branch-and-bound algorithm wrapped in an approximate search routine. For complete details about the algorithm we refer to Hulpuş [10], page 114. However, any maximum edge weight clique finding algorithm can be used instead.

3.1 Kan-Dis: The Knowledge Graph based Disambiguation System

In order to evaluate the joint disambiguation with DBpedia, we implemented a system whose disambiguation process is illustrated in Figure 2.

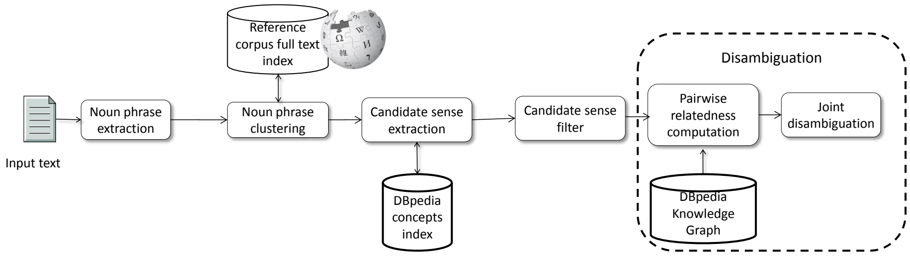

Fig. 2: Disambiguation process with Kan-Dis

We are only interested in disambiguating the nouns and noun-phrases of the document. We extract them by using the Stanford CoreNLP toolkit[1]. After the noun-phrases are extracted, the possible senses of the noun-phrases are retrieved from a Lucene Index[2] where we have indexed all DBpedia concepts based on their names.

In order to form groups of words to be simultaneously disambiguated (disambiguation context), we cluster the noun-phrases based on their co-occurrence in a reference corpus (i.e., Wikipedia). We experimented with two clustering algorithms: Louvain, which is a modularity-based community finding algorithm and hierarchical clustering with various linkage types and dendrogram cutting thresholds. The relatedness measures are computed between all pairs of candidates for the noun-phrases in each cluster. These relatedness scores are then sent to the joint disambiguation algorithm. The last two steps are part of the joint disambiguation algorithm, and they can be replaced with any other disambiguation algorithm.

[1] http://nlp.stanford.edu/software/corenlp.shtml
[2] http://lucene.apache.org/core/

4 Experiments and Results

We have described our relatedness measures and how we plan to use it for disambiguation. In the following, we detail the experiments we made to validate our hypothesis that our exclusivity based measure for relatedness correlates with human assessments. Afterwards we detail our experiments that show that path-based joint disambiguation outperforms centrality based disambiguation.

4.1 Evaluation of Relatedness Measures

We now present the experiments we carried out in order to verify the suitability of the proposed measures for assessing semantic relatedness. We follow the most established methodology for validating semantic relatedness measures, which consists of computing the correlation between human assessed scores and the proposed automatic measures. Our main hypothesis is that the exclusivity-based measure of relatedness will improve over the baseline and show high positive correlation to human assessments.

Ground-truth Datasets We experiment with five of the most commonly used datasets:

R&G [26] is one of the oldest and most used datasets that contain human assessment of word **similarity**. It contains 65 pairs of words together with the overall assessment of humans, gathered from 51 subjects. The users were requested to judge the "similarity of meaning" on a scale from 0.0 to 4.0, where a high score means high similarity.

WordSim353 [3, 1] contains 353 pairs of words assessed on a scale from 0 to 10 by 13 to 16 human users. Agirre et al [1] split the dataset through another user study in two overlapping parts

- **WS353-Sim** - containing 203 pairs that the users considered suitable for similarity computation;
- **WS353-Rel** - containing 252 pairs that the users considered suitable for relatedness computation;

R122 [31] is more recent and was created specifically for measuring **relatedness** [31] . It contains 122 pairs of words, scored within a range from 0.0 (completely unrelated) to 4.0 (very strongly related), each pair being evaluated by 14 to 22 annotators out of a total of 92 participants.

KORE [9] has also been created for measuring **relatedness**, but between NEs. It consists of 21 main entities, whose relatedness to other 20 entities each has been manually assessed, leading to 420 entity pairs.

Except for the last dataset, all others contain pairs of words rather than DBpedia concepts. Table 1 shows the exact number of pairs of words that we could directly and unambiguously link to concepts from DBpedia and Freebase.

Dataset	R&G	WS353-Sim	R-122	WS353-Rel	KORE
#pairs	38	139	93	168	419

Table 1: Number of concept pairs per ground truth dataset

Knowledge Bases In order to verify the generalisability of our measures, we evaluate them with both DBpedia and Freebase. All reported experiments were run on DBpedia 2014 version[3], and Freebase dump[4] from 18th January 2015. We remove from the graph of DBpedia the so-called stopURIs [12, 27]. Regarding Freebase, we remove all edges with an exclusivity score lower than 10^{-7} as they bring no impact on our measures due to their very small contribution, but they dramatically impact the performance of graph traversal algorithms. Table 2 shows the sizes of the resulting knowledge graphs.

KB	**#nodes**	**#relationships**	**#relationship types**
DBpedia	7,514,827	35,762,630	1,198
Freebase	41,527,432	253,813,430	13,991

Table 2: Number of elements in the knowledge bases

Regarding the DBpedia graph, we experiment with two settings:

- the full graph: DBpedia Full;
- the categories and types graph: DBpedia Categories;

In the case of the latter, we restrict the graph traversals to the relationships: *rdf:type, dcterms:subject, skos:broaderOf, skos:narrowerOf, rdfs:subClassOf*. We expect that the aforementioned properties are mostly useful in assessing similarity of concepts. We similarly expect that the full set of properties is most useful when assessing relatedness.

Compared methods We report our results on the methods described in Section 2. We have experimented with various values for the α parameter for both Katz relatedness and exclusivity based relatedness. In the following, we report the values obtained with $\alpha \in \{0.25, 0.5\}$ for Katz relatedness, and with $\alpha \in \{0.25, 0.5, 0.75, 1\}$ for exclusivity-based methods. We have also experimented with different k (1 to 20) values for the top-k paths in Katz as well as the exclusivity based methods. We report our results for top-1, top-5 and top-10 paths.

For comparison to related work, we have also implemented the *combIC* measure of Schuhmacher & Ponzetto [27]

[3] http://wiki.dbpedia.org/Downloads2014

[4] https://developers.google.com/freebase/data

Results Tables 3, 4, and 5 show the Spearman correlations obtained. In Table 3, we show the results on the datasets assessed for semantic similarity of common nouns and noun-phrases.

Dataset	**Method**	**DBpedia CAT**			**DBpedia Full**			**Freebase**		
		top-1	top-5	top-10	top-1	top-5	top-10	top-1	top-5	top-10
R & G	Katz $\alpha = 0.25$	**0.79**	0.78	0.79	0.71	0.67	0.64	0.45	0.41	0.51
	Katz $\alpha = 0.5$	**0.79**	0.78	0.79	0.71	0.67	0.63	0.45	0.41	0.20
	Katz $\alpha = 0.75$	**0.79**	0.78	0.78	0.71	0.67	0.61	0.45	0.41	0.11
	ER $\alpha = 0.25$	**0.79**	**0.81**	0.81	**0.72**	0.73	0.73	**0.67**	**0.66**	**0.67**
	ER $\alpha = 0.5$	0.78	**0.82**	**0.81**	0.66	0.73	0.73	**0.67**	**0.66**	**0.67**
	ER $\alpha = 0.75$	0.76	0.79	**0.81**	0.66	**0.75**	**0.75**	**0.67**	**0.66**	**0.67**
	ER $\alpha = 1$	0.76	0.79	**0.81**	0.66	0.72	0.74	0.61	0.60	0.61
	CombIC	0.74			0.57			0.59		
WS 353-Sim	Katz $\alpha = 0.25$	0.74	0.75	0.75	0.69	0.66	0.64	0.29	0.30	0.33
	Katz $\alpha = 0.5$	0.74	0.75	0.74	0.69	0.65	0.61	0.29	0.30	0.20
	Katz $\alpha = 0.75$	0.74	0.74	0.72	0.69	0.64	0.58	0.29	0.29	0.17
	ER $\alpha = 0.25$	**0.77**	**0.78**	**0.78**	0.68	**0.67**	**0.66**	0.58	0.57	0.57
	ER $\alpha = 0.5$	0.76	0.77	0.77	0.63	0.64	0.63	**0.59**	**0.59**	**0.59**
	ER $\alpha = 0.75$	0.71	0.73	0.73	0.59	0.61	0.61	**0.59**	**0.59**	**0.59**
	ER $\alpha = 1$	0.63	0.68	0.69	0.54	0.57	0.58	**0.59**	**0.59**	**0.59**
	CombIC	0.72	-	-	0.59	-	-	0.40	-	-

Table 3: Spearman correlations with ground truth on common nouns similarity datasets: R&G and WS353-Sim.

We notice that all measures have very high correlation with human assessment of similarity, when used on the DBpedia Categories. Our Exclusivity based relatedness performs best, reaching 0.82 correlation with the R&G dataset for $\alpha = 0.5$ and when top-5 paths are used. For comparison, $CombIC$ reaches 0.74 in the same setup. ER also performs better than $Katz$ and this is visible especially on the Freebase corpus.

Table 4 presents the results on the datasets assessed for semantic relatedness of common nouns and noun-phrases. We notice that overall, the results are much lower than for the similarity datasets (Table 3). No method correlates more than 0.57 with human assessment. Most likely, the cause of this poor assessment of noun relatedness has to do with the type of knowledge within the analysed KBs. They contain encyclopedic knowledge rather than common sense knowledge. For example, humans assess relatedness of concepts in pairs *(caffeine, headache)* and *(game, victory)* as moderately to strongly related but the KBs do not have any path shorter than 6 between them.

Nevertheless, ER performs much better than both $CombIC$ and $Katz$. ER with $\alpha = 0.25$ produces the best results on the noun relatedness datasets, which means that the smaller the influence of the longer paths, the better. We also notice that all measures perform extremely poorly when used on Freebase. This indicates that Freebase's graph structure connects in a less meaningful way the concepts referred to by common nouns, than DBpedia.

Table 5 shows the results obtained on the KORE dataset, which contains pairs of NEs assessed for semantic relatedness. On this dataset, the first thing to notice is that all measures perform very bad on DBpedia Categories, but very

Dataset	**Method**	**DBpedia CAT**			**DBpedia Full**			**Freebase**		
		top-1	top-5	top-10	top-1	top-5	top-10	top-1	top-5	top-10
R 122	Katz $\alpha = 0.25$	0.56	0.54	0.52	0.46	0.45	0.44	0.17	0.15	0.15
	Katz $\alpha = 0.5$	0.56	0.52	0.50	0.46	0.43	0.40	0.17	0.15	0.15
	Katz $\alpha = 0.75$	0.56	0.50	0.49	0.46	0.41	0.40	0.17	0.15	0.15
	ER $\alpha = 0.25$	**0.57**	**0.57**	**0.55**	**0.56**	**0.55**	**0.54**	**0.33**	**0.32**	**0.32**
	ER $\alpha = 0.5$	0.56	0.55	0.53	0.53	0.52	0.51	0.32	**0.32**	0.31
	ER $\alpha = 0.75$	0.52	0.53	0.50	0.50	0.49	0.48	**0.33**	**0.32**	0.31
	ER $\alpha = 1$	0.46	0.49	0.46	0.49	0.47	0.45	**0.33**	**0.32**	0.31
	CombIC	0.53	-	-	0.41	-	-	0.20	-	-
WS 353-Rel	Katz $\alpha = 0.25$	0.44	0.44	0.44	0.45	0.40	0.38	0.14	0.16	0.20
	Katz $\alpha = 0.5$	0.44	0.44	0.44	0.45	0.40	0.38	0.14	0.16	0.21
	Katz $\alpha = 0.75$	0.44	0.44	0.44	0.45	0.40	0.38	0.14	0.16	0.22
	ER $\alpha = 0.25$	**0.48**	**0.48**	0.47	**0.47**	**0.46**	**0.46**	**0.35**	**0.35**	**0.35**
	ER $\alpha = 0.5$	0.47	**0.48**	**0.48**	0.44	0.44	0.45	**0.35**	**0.35**	**0.35**
	ER $\alpha = 0.75$	0.47	**0.48**	0.47	0.42	0.43	0.44	**0.35**	0.34	0.34
	ER $\alpha = 1$	0.46	0.47	0.47	0.40	0.41	0.42	0.34	0.33	0.34
	CombIC	0.45	-	-	0.42	-	-	0.14	-	-

Table 4: Spearman correlations to ground truth on common nouns relatedness datasets: R-122 and WS353-Rel

well on Freebase. This indicates that Freebase's structure has a focus on named entities. On this dataset as well, *ER* outperforms both other methods, but from a smaller distance than on the other datasets.

The results clearly show that the exclusivity-based measure we introduce in this paper outperforms the Katz relatedness as well as the *CombIC* measure [27]. The measure that overall obtains the highest results is exclusivity based relatedness, with the α parameter set to 0.25 (ER 0.25). These results show that our graph-proximity measures are able to accurately capture semantic relatedness and similarity on both DBpedia and Freebase. We also notice the trend that the lower α values lead to better performance. This indicates that the lower the influence of longer paths, the better.

KORE									
	DBpedia CAT			**DBpedia Full**			**Freebase**		
	top-1	top-5	top-10	top-1	top-5	top-10	top-1	top-5	top-10
Katz $\alpha = 0.25$	0.31	0.33	0.32	0.56	0.53	0.49	0.57	0.58	0.57
Katz $\alpha = 0.5$	0.31	0.33	0.32	0.57	0.50	0.48	0.57	0.58	0.56
Katz $\alpha = 0.75$	0.31	0.32	0.32	0.56	0.48	0.46	0.57	0.58	0.55
ER $\alpha = 0.25$	**0.35**	**0.35**	**0.35**	**0.62**	0.62	0.62	**0.64**	**0.64**	**0.64**
ER $\alpha = 0.5$	**0.35**	**0.35**	**0.35**	**0.62**	**0.63**	**0.63**	0.63	0.63	0.63
ER $\alpha = 0.75$	**0.35**	**0.35**	0.34	0.60	0.61	0.61	0.60	0.61	0.61
ER $\alpha = 1$	0.34	0.34	0.34	0.60	0.61	0.61	0.60	0.60	0.61
CombIC	0.33	-	-	0.60	-	-	0.61	-	-

Table 5: Spearman correlations to ground truth on NE relatedness dataset KORE

4.2 Evaluation of Joint Disambiguation with DBpedia

We now present the experiments we carried out in order to evaluate the proposed approach to disambiguation. Our hypothesis is that joint disambiguation approaches perform better than the graph centrality based approaches.

Evaluated Methods In order to test our hypothesis, we implement Kan-Dis introduced earlier. We use it with three settings:

Joint ER implements joint disambiguation with our Relation Exclusivity based relatedness measure, with $\alpha = 0.25$ and top-5 most relevant paths.

Joint CombIC implements joint disambiguation with *CombIC* [27] relatedness measure;

HITS Authority implements the HITS centrality based disambiguation algorithm used by AGDISTIS [32].

Ground Truth Datasets We are using five commonly used datasets of texts that have been manually annotated by humans:

NYT10 dataset consists of ten excerpts from news articles published by New York Times [16]. Each text has all the meaning bearing phrases annotated with *at most one* DBpedia resource.

Aquaint50 dataset contains 50 documents from the AQUAINT corpus, that were used by Milne and Witten [19]. They have been linked and disambiguated to Wikipedia articles by their system, and the results were evaluated using Amazon Mechanical Turk[5].

IITB dataset contains 103 manually annotated documents. It has been published by Kulkarni et al [14].

RSS500 [25] dataset contains 500 manually annotated sentences mainly from news documents, automatically scrapped from RSS feeds.

Reuters-21578 [25] dataset contains 145 news randomly sampled from Reuters-21578 news articles dataset. The sampled news items were manually annotated with the linked named entities by domain experts.

Each of these datasets was produced with a particular purpose. NYT10 and IITB try to link as many meaning bearing words as possible. The other three focus on named entities only. Some datasets link every mention of a concept, while others link only the first occurrence. To deal with these differences, we use various performance measures, as follows.

Performance Measures In order to understand the performance of the relatedness measures and the joint disambiguation algorithm, we evaluate them under two tests.

The first test is an *annotation test* and consists of the traditional information retrieval evaluation measures, precision and recall. This test is highly influenced

[5] https://www.mturk.com

by the noun-phrase extraction phase. In case our text analysis component extracts different noun-phrases than the ground truth, it is penalised. Similarly, if the ground truth targets only named entities, the precision of our systems is penalised, since we target both named entities and common nouns. As we consider noun-phrase extraction a complementary problem to word-sense disambiguation, we reduce its influence by devising the so-called *disambiguation test*.

The Noun-Phrase Disambiguation Test The second test is a "disambiguation test" as it verifies to what extent a noun-phrase whose corresponding DBpedia concept is set by humans is linked and disambiguated by the system to the same DBpedia concept. It computes the *disambiguation accuracy* measure (denoted by Acc) by dividing the number of correctly linked noun phrases that are both annotated in the ground truth and extracted by the system, to the total number of noun-phrases both annotated by humans and extracted by the system. As such, as opposed to the "annotation test", the results obtained at the "disambiguation test" cancel out the impact of the noun phrase extraction. Furthermore, since many entities are not ambiguous in DBpedia, we also report a variation of this measure, in which we compute the disambiguation accuracy only on the entities and words that have more than one candidate sense (denoted by $Acc^{>1}$).

Results Table 6 shows the results achieved by the evaluated algorithms. It is easily noticeable that joint disambiguation performs generally better than the centrality based one, especially when used with $CombIC$ relatedness. On both hierarchical and Louvain clustering, joint disambiguation with $CombIC$ achieves best precision. With respect to recall, $HITSAuthority$ used by AGDISTIS tends to perform better. The most relevant performance measure for our setup are the accuracies Acc and $Acc^{>1}$. The accuracy of 0.916 achieved by joint disambiguation with $CombIC$ on the Reuters dataset means that out of the noun phrases that are annotated in the dataset and extracted by Kan_Dis, 91.6% are linked to the correct DBpedia concept. The 0.727 $Acc^{>1}$ score means that out of the noun-phrases that are annotated in the ground truth dataset, were extracted by Kan_Dis, and have more than one disambiguation candidate, 72% were disambiguated to the correct DBpedia concept. The $Acc^{>1}$ scores of joint disambiguation with $CombIC$ are with 0.1 higher than those of $HITSAuthority$ in average.

Regarding the datasets, there is a noticeable decrease of precision for the datasets that only annotate named entities (AQUAINT, RSS500, Reuters). This is because Kan_Dis links and disambiguates both common nouns and named entities. Therefore, in order to get an idea of its performance, the Acc and $Acc^{>1}$ measures are the most conclusive. We notice that on $RSS500$ the accuracies of all the methods are very poor. This is due to the fact that $RSS500$ contains single sentences, therefore there might be not sufficient context for achieving correct disambiguation.

The disambiguation context produced with hierarchical clustering leads to higher precision but lower recall than Louvain clustering. This is due to the used dendrogram cutting threshold that we use (0.8) with hierarchical clustering and that leads to smaller clusters than the Louvain clusters. Small clusters tend to be cleaner, therefore the disambiguation accuracy improves. However, the small clusters have less disambiguation cues, and lead to more words not being disambiguated, producing lower recall.

Dataset	Method	Hierarchical					Louvain				
		P	R	F	Acc	$Acc^{>1}$	P	R	F	Acc	$Acc^{>1}$
NYT10	HITS Authority	0.567	**0.584**	0.572	0.834	0.658	0.576	0.593	0.581	0.851	0.693
	Joint ER	0.607	0.573	0.585	**0.919**	**0.842**	0.586	0.590	0.585	0.885	0.79
	Joint CombIC	**0.613**	0.568	**0.586**	0.912	0.82	**0.595**	**0.601**	**0.595**	**0.892**	**0.802**
IITB	HITS Authority	0.417	**0.435**	0.420	0.761	0.591	0.417	0.436	0.421	0.760	0.587
	Joint ER	0.437	0.432	0.429	0.775	0.653	0.429	**0.437**	0.428	0.780	0.626
	Joint CombIC	**0.472**	0.413	**0.435**	**0.813**	**0.717**	**0.446**	0.428	**0.431**	**0.802**	**0.710**
AQUAINT	HITS Authority	0.247	**0.583**	0.341	0.801	0.594	0.247	0.582	0.340	0.800	0.589
	Joint ER	0.257	0.563	0.346	0.809	0.638	0.250	0.574	0.341	0.816	0.688
	Joint CombIC	**0.264**	0.564	**0.353**	**0.824**	**0.708**	**0.256**	**0.583**	**0.348**	**0.831**	**0.733**
RSS500	HITS Authority	0.171	**0.543**	0.238	0.760	**0.315**	0.171	**0.543**	0.238	0.760	0.315
	Joint ER	0.190	0.524	0.252	0.784	0.224	0.174	0.540	0.240	0.763	0.307
	Joint CombIC	**0.194**	0.524	**0.256**	**0.789**	0.241	**0.176**	0.540	**0.242**	**0.768**	**0.321**
REUTERS	HITS Authority	0.152	**0.700**	0.235	0.894	0.704	0.152	**0.701**	0.235	0.894	0.704
	Joint ER	0.156	0.653	0.237	0.906	0.652	0.153	0.692	0.236	0.892	0.638
	Joint CombIC	**0.162**	0.633	**0.241**	**0.929**	**0.727**	**0.159**	0.676	**0.241**	**0.906**	**0.729**

Table 6: Disambiguation results: P - precision; R - recall; Acc- linking accuracy, Acc$^{>1}$ - disambiguation accuracy for words with more than 1 candidate. Joint ER uses decay 0.25 and top 5 paths.

5 Conclusion and Future Work

In this paper, we have proposed a novel measure for assessing strength of relations in knowledge graphs, called relation exclusivity. We used this measure for computing semantic relatedness as well as similarity. Besides, we also proposed an entity and word-sense disambiguation pipeline *Kan_Dis* that uses the proposed relatedness measures. We analysed our approach from different perspectives, on five ground truth datasets, and three knowledge graphs. We showed that when used with full DBpedia or Freebase it achieves better results than state-of-the-art approaches.

With respect to our disambiguation approach, we focused specifically on graph-based algorithms. We implemented algorithms from the related work and used them in the same experimental setup with ours, in order to obtain a conclusive comparison. We then showed that joint path-based disambiguation achieves better performance than the graph centrality based approach.

An interesting outcome of our experiments is that while *CombIC* achieved much worse performance when evaluated against human assessment of relatedness, it achieved the best disambiguation capability. This indicates that for disambiguation, measures must have additional properties than correlation to human assessment of relatedness. One such property might be the scale of the resulted scores. We plan to investigate this in future work.

A very interesting future research path is that of extraction of relevant relation paths between given concepts. We plan to investigate and formally evaluate if the paths deemed relevant by our measure are indeed relevant to humans.

Acknowledgements

This work was supported by Science Foundation Ireland (SFI) under Grant No. 12/RC/2289 (Insight Centre for Data Analytics).

References

1. Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Paşca, M., Soroa, A.: A study on similarity and relatedness using distributional and wordnet-based approaches. pp. 19–27. NAACL '09, ACL (2009)
2. Agirre, E., Soroa, A.: Personalizing pagerank for word sense disambiguation. In: Proc. 12th Conf. of the European Chapter of the Association for Computational Linguistics (2009)
3. Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin, E.: Placing search in context: The concept revisited. ACM Trans. Inf. Syst. 20(1), 116–131 (Jan 2002), http://doi.acm.org/10.1145/503104.503110
4. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. pp. 1606–1611. IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2007)
5. Garcia, A., Szomszor, M., Alani, H., Corcho, O.: Preliminary results in tag disambiguation using dbpedia. In: Knowledge Capture (K-Cap'09) - 1st International Workshop on Collective Knowledge Capturing and Representation (2009)
6. Gentile, A.L., Zhang, Z., Xia, L., Iria, J.: Semantic relatedness approach for named entity disambiguation. In: Digital libraries, pp. 137–148. Springer (2010)
7. Grieser, K., Baldwin, T., Bohnert, F., Sonenberg, L.: Using ontological and document similarity to estimate museum exhibit relatedness. ACM Journal of Computing and Cultural Heritage" 3(3), 1–20 (2011)
8. Hakimov, S., Oto, S.A., Dogdu, E.: Named entity recognition and disambiguation using linked data and graph-based centrality scoring. In: Proceedings of the 4th International Workshop on Semantic Web Information Management. pp. 4:1–4:7. SWIM '12, ACM, New York, NY, USA (2012)
9. Hoffart, J., Seufert, S., Nguyen, D.B., Theobald, M., Weikum, G.: Kore: Keyphrase overlap relatedness for entity disambiguation. pp. 545–554. CIKM '12, ACM (2012)
10. Hulpuş, I.: Semantic Network Analysis for Topic Linking and Labelling. Ph.D. thesis, National University of Ireland, Galway (2014)
11. Hulpuş, I., Hayes, C., Karnstedt, M., Greene, D.: An Eigenvalue-Based Measure for Word-Sense Disambiguation. In: FLAIRS '12 (2012)
12. Hulpuş, I., Hayes, C., Karnstedt, M., Greene, D.: Unsupervised graph-based topic labelling using dbpedia. pp. 465–474. WSDM, ACM, New York,USA (2013)
13. Katz, L.: A new status index derived from sociometric analysis. Psychometrika 18(1), 39–43 (1953)

14. Kulkarni, S., Singh, A., Ramakrishnan, G., Chakrabarti, S.: Collective annotation of wikipedia entities in web text. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 457–466. KDD '09, ACM, New York, NY, USA (2009)
15. Leal, J.P., Rodrigues, V., Queirs, R.: Computing semantic relatedness using dbpedia. In: Simes, A., Queirs, R., da Cruz, D.C. (eds.) SLATE. OASICS, vol. 21, pp. 133–147 (2012)
16. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: DBpedia spotlight: shedding light on the web of documents. In: I-Semantics '11. pp. 1–8 (2011)
17. Mihalcea, R., Tarau, P., Figa, E.: Pagerank on semantic networks, with application to word sense disambiguation. In: Proceedings of the 20th International Conference on Computational Linguistics. COLING '04, ACL (2004)
18. Milne, D., Witten, I.H.: An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In: In Proceedings of AAAI 2008 (2008)
19. Milne, D., Witten, I.H.: Learning to link with wikipedia. In: Proceedings of the 17th ACM CIKM. pp. 509–518. CIKM '08, ACM (2008)
20. Mirizzi, R., Di Noia, T., Ragone, A., Ostuni, V.C., Di Sciascio, E.: Movie recommendation with dbpedia. CEUR Workshop Proceedings, vol. 835 (2012)
21. Navigli, R., Lapata, M.: Graph connectivity measures for unsupervised word sense disambiguation. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence. pp. 1683–1688. IJCAI'07 (2007)
22. Nunes, B.P., Dietze, S., Casanova, M., Kawase, R., Fetahu, B., Nejdl, W.: Combining a co-occurrence-based and a semantic measure for entity linking. In: ESWC 2013 - 10th Extended Semantic Web Conference (2013)
23. Passant, A.: Measuring semantic distance on linking data and using it for resources recommendations. In: Linked Data Meets Artificial Intelligence, Papers from the 2010 AAAI Spring Symposium, Stanford, California, USA (2010)
24. Ratinov, L., Roth, D., Downey, D., Anderson, M.: Local and global algorithms for disambiguation to wikipedia. pp. 1375–1384. HLT '11, Association for Computational Linguistics (2011)
25. Röder, M., Usbeck, R., Hellmann, S., Gerber, D., Both, A.: N3 - a collection of datasets for named entity recognition and disambiguation in the nlp interchange format. In: The 9th edition of LREC, 26-31 May, Reykjavik, Iceland (2014)
26. Rubenstein, H., Goodenough, J.B.: Contextual correlates of synonymy. Commun. ACM 8(10), 627–633 (Oct 1965), `http://doi.acm.org/10.1145/365628.365657`
27. Schuhmacher, M., Ponzetto, S.P.: Knowledge-based graph document modeling. In: Proceedings of the 7th ACM WSDM. pp. 543–552. WSDM '14, ACM (2014)
28. Sinha, R., Mihalcea, R.: Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. In: Proc. International Conference on Semantic Computing. pp. 363–369. IEEE Computer Society (2007)
29. St-Onge, D.: Detecting and Correcting Malapropisms with Lexical Chains. Master's thesis, University of Toronto (1995)
30. Sussna, M.: Word sense disambiguation for free-text indexing using a massive semantic network. In: Proceedings of the second CIKM. pp. 67–74. CIKM '93, ACM, New York, NY, USA (1993)
31. Szumlanski, S.R., Gomez, F., Sims, V.K.: A new set of norms for semantic relatedness measures. In: ACL (2). pp. 890–895 (2013)
32. Usbeck, R., Ngonga Ngomo, A.C., Rder, M., Gerber, D., Coelho, S., Auer, S., Both, A.: Agdistis - graph-based disambiguation of named entities using linked data. In: The Semantic Web ISWC 2014, vol. 8796, pp. 457–471. Springer (2014)