Watset: Automatic Induction of Synsets from a Graph of Synonyms

Dmitry Ustalov[†*], Alexander Panchenko[‡], and Chris Biemann[‡]

[†]Institute of Natural Sciences and Mathematics, Ural Federal University, Russia
[*]Krasovskii Institute of Mathematics and Mechanics, Russia
[‡]Language Technology Group, Department of Informatics, Universität Hamburg, Germany
`dmitry.ustalov@urfu.ru`
`{panchenko,biemann}@informatik.uni-hamburg.de`

Abstract

This paper presents a new graph-based approach that induces synsets using synonymy dictionaries and word embeddings. First, we build a weighted graph of synonyms extracted from commonly available resources, such as Wiktionary. Second, we apply word sense induction to deal with ambiguous words. Finally, we cluster the disambiguated version of the ambiguous input graph into synsets. Our meta-clustering approach lets us use an efficient hard clustering algorithm to perform a fuzzy clustering of the graph. Despite its simplicity, our approach shows excellent results, outperforming five competitive state-of-the-art methods in terms of F-score on three gold standard datasets for English and Russian derived from large-scale manually constructed lexical resources.

1 Introduction

A *synset* is a set of mutual synonyms, which can be represented as a clique graph where nodes are words and edges are synonymy relations. Synsets represent word senses and are building blocks of WordNet (Miller, 1995) and similar resources such as thesauri and lexical ontologies. These resources are crucial for many natural language processing applications that require common sense reasoning, such as information retrieval (Gong et al., 2005) and question answering (Kwok et al., 2001; Zhou et al., 2013). However, for most languages, no manually-constructed resource is available that is comparable to the English WordNet in terms of coverage and quality. For instance, Kiselev et al. (2015) present a comparative analysis of lexical resources available for the Russian

language concluding that there is no resource compared to WordNet in terms of coverage and quality for Russian. This lack of linguistic resources for many languages urges the development of new methods for automatic construction of WordNet-like resources. The automatic methods foster construction and use of the new lexical resources.

Wikipedia[1], Wiktionary[2], OmegaWiki[3] and other collaboratively-created resources contain a large amount of lexical semantic information—yet designed to be human-readable and not formally structured. While semantic relations can be automatically extracted using tools such as DKPro JWKTL[4] and Wikokit[5], words in these relations are not disambiguated. For instance, the synonymy pairs (*bank*, *streambank*) and (*bank*, *banking company*) will be connected via the word "bank", while they refer to the different senses. This problem stems from the fact that articles in Wiktionary and similar resources list undisambiguated synonyms. They are easy to disambiguate for humans while reading a dictionary article, but can be a source of errors for language processing systems.

The contribution of this paper is a novel approach that resolves ambiguities in the input graph to perform fuzzy clustering. The method takes as an input synonymy relations between potentially ambiguous terms available in human-readable dictionaries and transforms them into a machine readable representation in the form of disambiguated synsets. Our method, called WATSET, is based on a new local-global meta-algorithm for fuzzy graph clustering. The underlying principle is to discover the word senses based on a *local* graph cluster-

---

[1]`http://www.wikipedia.org`
[2]`http://www.wiktionary.org`
[3]`http://www.omegawiki.org`
[4]`https://dkpro.github.io/dkpro-jwktl`
[5]`https://github.com/componavt/wikokit`

ing, and then to induce synsets using *global* sense clustering. We show that our method outperforms other methods for synset induction. The induced resource eliminates the need in manual synset construction and can be used to build WordNet-like semantic networks for under-resourced languages. An implementation of our method along with induced lexical resources is available online.[6]

2 Related Work

Methods based on resource linking surveyed by Gurevych et al. (2016) gather various existing lexical resources and perform their linking to obtain a machine-readable repository of lexical semantic knowledge. For instance, BabelNet (Navigli and Ponzetto, 2012) relies in its core on a linking of WordNet and Wikipedia. UBY (Gurevych et al., 2012) is a general-purpose specification for the representation of lexical-semantic resources and links between them. The main advantage of our approach compared to the lexical resources is that no manual synset encoding is required.

Methods based on word sense induction try to induce sense representations without the need for any initial lexical resource by extracting semantic relations from text. In particular, word sense induction (WSI) based on word ego networks clusters graphs of semantically related words (Lin, 1998; Pantel and Lin, 2002; Dorow and Widdows, 2003; Véronis, 2004; Hope and Keller, 2013; Pelevina et al., 2016; Panchenko et al., 2017a), where each cluster corresponds to a word sense. An ego network consists of a single node (ego) together with the nodes they are connected to (alters) and all the edges among those alters (Everett and Borgatti, 2005). In the case of WSI, such a network is a local neighborhood of one word. Nodes of the ego network are the words which are semantically similar to the target word.

Such approaches are able to discover homonymous senses of words, e.g., "bank" as slope versus "bank" as organisation (Di Marco and Navigli, 2012). However, as the graphs are usually composed of semantically related words obtained using distributional methods (Baroni and Lenci, 2010; Biemann and Riedl, 2013), the resulting clusters by no means can be considered synsets. Namely, (1) they contain words related not only via synonymy relation, but via a mixture of relations such as synonymy, hypernymy,

co-hyponymy, antonymy, etc. (Heylen et al., 2008; Panchenko, 2011); (2) clusters are not unique, i.e., one word can occur in clusters of different ego networks referring to the same sense, while in WordNet a word sense occurs only in a single synset.

In our synset induction method, we use word ego network clustering similarly as in word sense induction approaches, but apply them to a graph of semantically clean synonyms.

Methods based on clustering of synonyms, such as our approach, induce the resource from an ambiguous graph of synonyms where edges a extracted from manually-created resources. According to the best of our knowledge, most experiments either employed graph-based word sense induction applied to text-derived graphs or relied on a linking-based method that already assumes availability of a WordNet-like resource. A notable exception is the ECO approach by Gonçalo Oliveira and Gomes (2014), which was applied to induce a WordNet of the Portuguese language called Onto.PT.[7] We compare to this approach and to five other state-of-the-art graph clustering algorithms as the baselines.

ECO (Gonçalo Oliveira and Gomes, 2014) is a *fuzzy* clustering algorithm that was used to induce synsets for a Portuguese WordNet from several available synonymy dictionaries. The algorithm starts by adding random noise to edge weights. Then, the approach applies Markov Clustering (see below) of this graph several times to estimate the probability of each word pair being in the same synset. Finally, candidate pairs over a certain threshold are added to output synsets.

MaxMax (Hope and Keller, 2013) is a *fuzzy* clustering algorithm particularly designed for the word sense induction task. In a nutshell, pairs of nodes are grouped if they have a maximal mutual affinity. The algorithm starts by converting the undirected input graph into a directed graph by keeping the maximal affinity nodes of each node. Next, all nodes are marked as root nodes. Finally, for each root node, the following procedure is repeated: all transitive children of this root form a cluster and the root are marked as non-root nodes; a root node together with all its transitive children form a fuzzy cluster.

Markov Clustering (MCL) (van Dongen, 2000) is a *hard* clustering algorithm for graphs based on simulation of stochastic flow in graphs.

---

[6]https://github.com/dustalov/watset
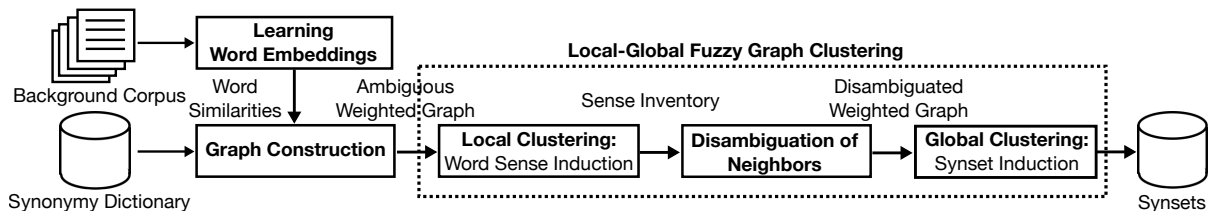
[7]http://ontopt.dei.uc.pt

Figure 1: Outline of the WATSET method for synset induction.

MCL simulates random walks within a graph by alternation of two operators called expansion and inflation, which recompute the class labels. Notably, it has been successfully used for the word sense induction task (Dorow and Widdows, 2003).

Chinese Whispers (CW) (Biemann, 2006) is a *hard* clustering algorithm for weighted graphs that can be considered as a special case of MCL with a simplified class update step. At each iteration, the labels of all the nodes are updated according to the majority labels among the neighboring nodes. The algorithm has a meta-parameter that controls graph weights that can be set to three values: (1) *top* sums over the neighborhood's classes; (2) *nolog* downgrades the influence of a neighboring node by its degree or by (3) *log* of its degree.

Clique Percolation Method (CPM) (Palla et al., 2005) is a *fuzzy* clustering algorithm for unweighted graphs that builds up clusters from k-cliques corresponding to fully connected subgraphs of k nodes. While this method is only commonly used in social network analysis, we decided to add it to the comparison as synsets are essentially cliques of synonyms, which makes it natural to apply an algorithm based on clique detection.

3 The WATSET Method

The goal of our method is to induce a set of unambiguous synsets by grouping individual ambiguous synonyms. An outline of the proposed approach is depicted in Figure 1. The method takes a dictionary of ambiguous synonymy relations and a text corpus as an input and outputs synsets. Note that the method can be used without a background corpus, yet as our experiments will show, corpus-based information improves the results when utilizing it for weighting the word graph's edges.

A synonymy dictionary can be perceived as a graph, where the nodes correspond to lexical entries (words) and the edges connect pairs of the nodes when the synonymy relation between them holds. The cliques in such a graph naturally form

densely connected sets of synonyms corresponding to concepts (Gfeller et al., 2005). Given the fact that solving the clique problem exactly in a graph is NP-complete (Bomze et al., 1999) and that these graphs typically contain tens of thousands of nodes, it is reasonable to use efficient hard graph clustering algorithms, like MCL and CW, for finding a global segmentation of the graph. However, the hard clustering property of these algorithm does not handle polysemy: while one word could have several senses, it will be assigned to only one cluster. To deal with this limitation, a word sense induction procedure is used to induce senses for all words, one at the time, to produce a disambiguated version of the graph where a word is now represented with one or many word senses. The concept of a disambiguated graph is described in (Biemann, 2012). Finally, the disambiguated word sense graph is clustered globally to induce synsets, which are hard clusters of word senses.

More specifically, the method consists of five steps presented in Figure 1: (1) learning word embeddings; (2) constructing the ambiguous weighted graph of synonyms G; (3) inducing the word senses; (4) constructing the disambiguated weighted graph G' by disambiguating of neighbors with respect to the induced word senses; (5) global clustering of the graph G'.

3.1 Learning Word Embeddings

Since different graph clustering algorithms are sensitive to edge weighting, we consider distributional semantic similarity based on word embeddings as a possible edge weighting approach for our synonymy graph. As we show further, this approach improves over unweighted versions and yields the best overall results.

3.2 Construction of a Synonymy Graph

We construct the synonymy graph $G = (V, E)$ as follows. The set of nodes V includes every lexeme appearing in the input synonymy dictionaries. The set of undirected edges E is composed of all edges

Stage 1: Ambigous Graph before the Local Clustering

Stage 2: Sense Inventory with Ambigous Neighbors

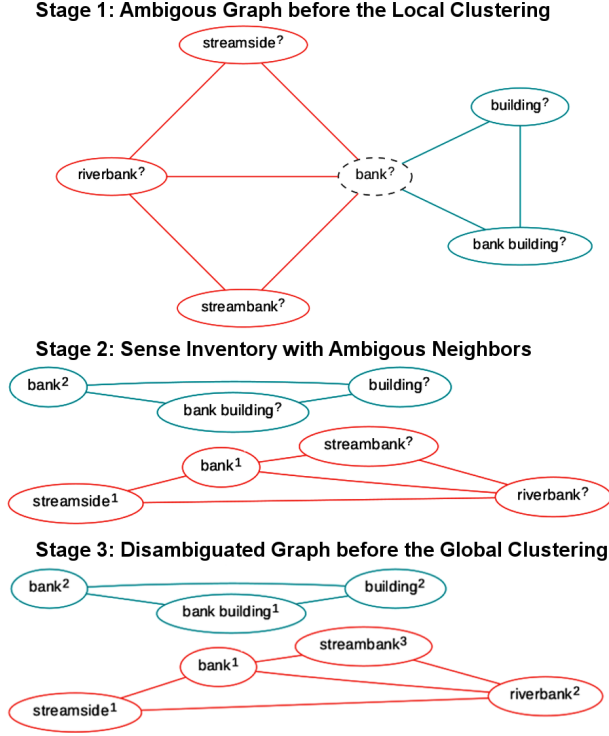**Stage 3: Disambiguated Graph before the Global Clustering**

Figure 2: Disambiguation of an ambiguous input graph using local clustering (WSI) to facilitate global clustering of words into synsets.

$(u, v) \in V \times V$ retrieved from one of the input synonymy dictionaries. We consider three edge weight representations:

- **ones** that assigns every edge the constant weight of 1;

- **count** that weights the edge (u, v) as the number of times the synonymy pair appeared in the input dictionaries;

- **sim** that assigns every edge (u, v) a weight equal to the cosine similarity of skip-gram word vectors (Mikolov et al., 2013).

As the graph G is likely to have polysemous words, the goal is to separate individual word senses using graph-based word sense induction.

3.3 Local Clustering: Word Sense Induction

In order to facilitate global fuzzy clustering of the graph, we perform disambiguation of its ambiguous nodes as illustrated in Figure 2. First, we use a graph-based word sense induction method that is similar to the curvature-based approach of Dorow and Widdows (2003). In particular, removal of the nodes participating in many triangles tends to

separate the original graph into several connected components. Thus, given a word u, we extract a network of its nearest neighbors from the synonymy graph G. Then, we remove the original word u from this network and run a hard graph clustering algorithm that assigns one node to one and only one cluster. In our experiments, we test Chinese Whispers and Markov Clustering. The expected result of this is that each cluster represents a different sense of the word u, e.g.:

$bank^1$ $\{streambank, riverbank, \dots\}$
$bank^2$ $\{bank\ company, \dots\}$
$bank^3$ $\{bank\ building, building, \dots\}$
$bank^4$ $\{coin\ bank, penny\ bank, \dots\}$

We denote, e.g., $bank^1$, $bank^2$ and other items as word senses referred to as senses(bank). We denote as $\mathrm{ctx}(s)$ a cluster corresponding to the word sense s. Note that the context words have no sense labels. They are recovered by the disambiguation approach described next.

3.4 Disambiguation of Neighbors

Next, we disambiguate the neighbors of each induced sense. The previous step results in splitting word nodes into (one or more) sense nodes. However, nearest neighbors of each sense node are still ambiguous, e.g., $(bank^3, building^?)$. To recover these sense labels of the neighboring words, we employ the following sense disambiguation approach proposed by Faralli et al. (2016). For each word u in the context $\mathrm{ctx}(s)$ of the sense s, we find the most similar sense of that word $\hat{u}$ to the context. We use the cosine similarity measure between the context of the sense s and the context of each candidate sense u' of the word u:

$$\hat{u} = \arg\max_{u' \in \text{senses}(u)} \cos(\mathrm{ctx}(s), \mathrm{ctx}(u')).$$

A context $\mathrm{ctx}(\cdot)$ is represented by a sparse vector in a vector space of all ambiguous words of all contexts. The result is a disambiguated context $\widehat{\mathrm{ctx}}(s)$ in a space of *disambiguated* words derived from its *ambiguous* version $\mathrm{ctx}(s)$:

$$\widehat{\mathrm{ctx}}(s) = \{\hat{u} : u \in \mathrm{ctx}(s)\}.$$

3.5 Global Clustering: Synset Induction

Finally, we construct the word sense graph $G' = (V', E')$ using the disambiguated senses instead of the original words and establishing the edges between these disambiguated senses:

$$V' = \bigcup_{u \in V} \text{senses}(u), \quad E' = \bigcup_{s \in V'} \{s\} \times \widehat{\mathrm{ctx}}(s).$$

Running a hard clustering algorithm on G' produces the desired set of synsets as our final result. Figure 2 illustrates the process of disambiguation of an input ambiguous graph on the example of the word "bank". As one may observe, disambiguation of the nearest neighbors is a necessity to be able to construct a global version of the sense-aware graph. Note that current approaches to WSI, e.g., (Véronis, 2004; Biemann, 2006; Hope and Keller, 2013), do not perform this step, but perform only local clustering of the graph since they do not aim at a global representation of synsets.

3.6 Local-Global Fuzzy Graph Clustering

While we use our approach to synset induction in this work, the core of our method is the "local-global" fuzzy graph clustering algorithm, which can be applied to arbitrary graphs (see Figure 1). This method, summarized in Algorithm 1, takes an undirected graph $G = (V, E)$ as the input and outputs a set of fuzzy clusters of its nodes V. This is a meta-algorithm as it operates on top of two hard clustering algorithms denoted as $\text{Cluster}_{\text{local}}$ and $\text{Cluster}_{\text{global}}$, such as CW or MCL. At the first phase of the algorithm, for each node its senses are induced via ego network clustering (lines 1–7). Next, the disambiguation of each ego network is performed (lines 8–15). Finally, the fuzzy clusters are obtained by applying the hard clustering algorithm to the disambiguated graph (line 16). As a post-processing step, the sense labels can be removed to make the cluster elements subsets of V.

4 Evaluation

We conduct our experiments on resources from two different languages. We evaluate our approach on two datasets for English to demonstrate its performance on a resource-rich language. Additionally, we evaluate it on two Russian datasets since Russian is a good example of an under-resourced language with a clear need for synset induction.

4.1 Gold Standard Datasets

For each language, we used two differently constructed lexical semantic resources listed in Table 1 to obtain gold standard synsets.

English. We use **WordNet**[8], a popular English lexical database constructed by expert lexicographers. WordNet contains general vocabulary and

---

Algorithm 1 WATSET fuzzy graph clustering

Input: a set of nodes V and a set of edges E.
Output: a set of fuzzy clusters of V.

1: **for all** $u \in V$ **do**
2:     $C \leftarrow \text{Cluster}_{\text{local}}(\text{Ego}(u))$ // $C = \{C_1, ...\}$
3:     **for** $i \leftarrow 1 \ldots |C|$ **do**
4:         $\text{ctx}(u^i) \leftarrow C_i$
5:         $\text{senses}(u) \leftarrow \text{senses}(u) \cup \{u^i\}$
6:     **end for**
7: **end for**
8: $V' \leftarrow \bigcup_{u \in V} \text{senses}(u)$
9: **for all** $s \in V'$ **do**
10:     **for all** $u \in \text{ctx}(s)$ **do**
11:         $\hat{u} \leftarrow \underset{u' \in \text{senses}(u)}{\arg \max} \cos(\text{ctx}(s), \text{ctx}(u'))$
12:     **end for**
13:     $\widehat{\text{ctx}}(s) \leftarrow \{\hat{u} : u \in \text{ctx}(s)\}$
14: **end for**
15: $E' \leftarrow \bigcup_{s \in V'} \{s\} \times \widehat{\text{ctx}}(s)$
16: **return** $\text{Cluster}_{\text{global}}(V', E')$

---

appears to be *de facto* gold standard in similar tasks (Hope and Keller, 2013). We used WordNet 3.1 to derive the synonymy pairs from synsets. Additionally, we use **BabelNet**[9], a large-scale multilingual semantic network constructed automatically using WordNet, Wikipedia and other resources. We retrieved all the synonymy pairs from the BabelNet 3.7 synsets marked as English.

Russian. As a lexical ontology for Russian, we use **RuWordNet**[10] (Loukachevitch et al., 2016), containing both general vocabulary and domain-specific synsets related to sport, finance, economics, etc. Up to a half of the words in this resource are multi-word expressions (Kiselev et al., 2015), which is due to the coverage of domain-specific vocabulary. RuWordNet is a WordNet-like version of the RuThes thesaurus that is constructed in the traditional way, namely by a small group of expert lexicographers (Loukachevitch, 2011). In addition, we use **Yet Another Russ-Net**[11] (**YARN**) by Braslavski et al. (2016) as another gold standard for Russian. The resource is constructed using crowdsourcing and mostly covers general vocabulary. Particularly, non-expert users are allowed to edit synsets in a collaborative way loosely supervised by a team of project curators. Due to the ongoing development of the re-

---

source, we selected as the gold standard only those synsets that were edited at least eight times in order to filter out noisy incomplete synsets.

Resource		# words	# synsets	# synonyms
WordNet	En	148 730	117 659	152 254
BabelNet	En	11 710 137	6 667 855	28 822 400
RuWordNet	Ru	110 242	49 492	278 381
YARN	Ru	9 141	2 210	48 291

Table 1: Statistics of the gold standard datasets.

4.2 Evaluation Metrics

To evaluate the quality of the induced synsets, we transformed them into binary synonymy relations and computed precision, recall, and F-score on the basis of the overlap of these binary relations with the binary relations from the gold standard datasets. Given a synset containing n words, we generate a set of $\frac{n(n-1)}{2}$ pairs of synonyms. The F-score calculated this way is known as *Paired F-score* (Manandhar et al., 2010; Hope and Keller, 2013). The advantage of this measure compared to other cluster evaluation measures, such as *Fuzzy B-Cubed* (Jurgens and Klapaftis, 2013), is its straightforward interpretability.

4.3 Word Embeddings

English. We use the standard 300-dimensional word embeddings trained on the 100 billion tokens Google News corpus (Mikolov et al., 2013).[12]

Russian. We use the 500-dimensional word embeddings trained using the skip-gram model with negative sampling (Mikolov et al., 2013) using a context window size of 10 with the minimal word frequency of 5 on a 12.9 billion tokens corpus of books. These embeddings were shown to produce state-of-the-art results in the RUSSE shared task[13] and are part of the Russian Distributional Thesaurus (RDT) (Panchenko et al., 2017b).[14]

4.4 Input Dictionary of Synonyms

For each language, we constructed a synonymy graph using openly available language resources. The statistics of the graphs used as the input in the further experiments are shown in Table 2.

English. Synonyms were extracted from the English Wiktionary[15], which is the largest Wiktionary at the present moment in terms of the lexical coverage, using the DKPro JWKTL tool by Zesch et al. (2008). English words have been extracted from the dump.

Russian. Synonyms from three sources were combined to improve lexical coverage of the input dictionary and to enforce confidence in jointly observed synonyms: (1) synonyms listed in the Russian Wiktionary extracted using the Wikokit tool by Krizhanovsky and Smirnov (2013); (2) the dictionary of Abramov (1999); and (3) the Universal Dictionary of Concepts (Dikonov, 2013). While the two latter resources are specific to Russian, Wiktionary is available for most languages. Note that the same input synonymy dictionaries were used by authors of YARN to construct synsets using crowdsourcing. The results on the YARN dataset show how close an automatic synset induction method can approximate manually created synsets provided the same starting material.[16]

Language	# words	# synonyms
English	243 840	212 163
Russian	83 092	211 986

Table 2: Statistics of the input datasets.

5 Results

We compare WATSET with five state-of-the-art graph clustering methods presented in Section 2: Chinese Whispers (CW), Markov Clustering (MCL), MaxMax, ECO clustering, and the clique percolation method (CPM). The first two algorithms perform hard clustering, while the last three are fuzzy clustering methods just like our method. While the hard clustering algorithms are able to discover clusters which correspond to synsets composed of unambigous words, they can produce wrong results in the presence of lexical ambiguity (one node belongs to several synsets). In our experiments, we rely on our own implementation of MaxMax and ECO as reference implementations are not available. For CW[17], MCL[18]

---

[12]https://code.google.com/p/word2vec
[13]http://www.dialog-21.ru/en/evaluation/2015/semantic_similarity
[14]http://russe.nlpub.ru/downloads
[15]We used the Wiktionary dumps of February 1, 2017.
[16]We used the YARN dumps of February 7, 2017.
[17]https://www.github.com/uhh-lt/chinese-whispers
[18]http://java-ml.sourceforge.net

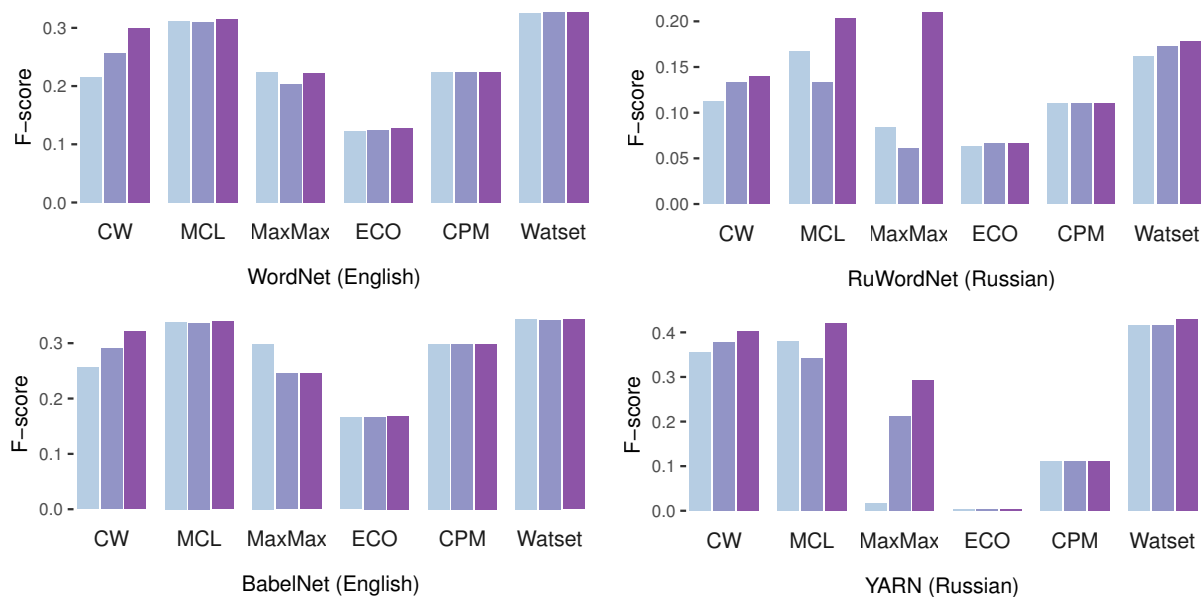Figure 3: Impact of the different graph weighting schemas on the performance of synset induction: ones, count, sim. Each bar corresponds to the top performance of a method in Tables 3 and 4.

and CPM[19], available implementations have been used. During the evaluation, we delete clusters equal or larger than the threshold of 150 words as they hardly can represent any meaningful synset. The notation WATSET[MCL, CW_top] means using MCL for local clustering and Chinese Whispers in the *top* mode for global clustering.

5.1 Impact of Graph Weighting Schema

Figure 3 presents an overview of the evaluation results on both datasets. The first step, common for all of the tested synset induction methods, is graph construction. Thus, we started with an analysis of three ways to weight edges of the graph introduced in Section 3.2: binary scores (*ones*), frequencies (*count*), and semantic similarity scores (*sim*) based on word vector similarity. Results across various configurations and methods indicate that using the weights based on the similarity scores provided by word embeddings is the best strategy for all methods except MaxMax on the English datasets. However, its performance using the *ones* weighting does not exceed the other methods using the *sim* weighting. Therefore, we report all further results on the basis of the *sim* weights. The edge weighting scheme impacts Russian more for most algorithms. The CW algorithm however remains sensitive to the weighting also for the English dataset due to its randomized nature.

5.2 Comparative Analysis

Table 3 and 4 present evaluation results for both languages. For each method, we show the best configurations in terms of F-score. One may note that the granularity of the resulting synsets, especially for Russian, is very different, ranging from 4 000 synsets for the $\text{CPM}_{k=3}$ method to 67 645 induced by the ECO method. Both tables report the number of words, synsets and synonyms after pruning huge clusters larger than 150 words. Without this pruning, the MaxMax and CPM methods tend to discover giant components obtaining almost zero precision as we generate all possible pairs of nodes in such clusters. The other methods did not show such behavior.

WATSET robustly outperforms all other methods according to F-score on both English datasets (Table 3) and on the YARN dataset for Russian (Table 4). Also, it outperforms all other methods according to recall on both Russian datasets. The disambiguation of the input graph performed by the WATSET method splits nodes belonging to several local communities to several nodes, significantly facilitating the clustering task otherwise complicated by the presence of the hubs that wrongly link semantically unrelated nodes.

Interestingly, in all the cases, the toughest competitor was a hard clustering algorithm—MCL (van Dongen, 2000). We observed that the "plain" MCL successfully groups monosemous words, but

Method	# words	# synsets	# synonyms	WordNet			BabelNet		
				P	**R**	**F1**	**P**	**R**	**F1**
WATSET[MCL, MCL]	243 840	112 267	345 883	0.345	**0.308**	**0.325**	0.400	0.301	**0.343**
MCL	243 840	84 679	387 315	0.342	0.291	0.314	0.390	0.300	**0.339**
WATSET[MCL, CW$_{log}$]	243 840	105 631	431 085	0.314	**0.325**	0.319	0.359	**0.312**	**0.334**
CW$_{top}$	243 840	77 879	539 753	0.285	**0.317**	0.300	0.326	**0.317**	0.321
WATSET[CW$_{log}$, MCL]	243 840	164 689	227 906	**0.394**	0.280	**0.327**	**0.439**	0.245	0.314
WATSET[CW$_{log}$, CW$_{log}$]	243 840	164 667	228 523	**0.392**	0.280	**0.327**	**0.439**	0.245	0.314
CPM$_{k=2}$	186 896	67 109	317 293	**0.561**	0.141	0.225	**0.492**	0.214	0.299
MaxMax	219 892	73 929	797 743	0.176	**0.300**	0.222	0.202	**0.313**	0.245
ECO	243 840	171 773	84 372	**0.784**	0.069	0.128	**0.699**	0.096	0.169

Table 3: Comparison of the synset induction methods on datasets for English. All methods rely on the similarity edge weighting (*sim*); best configurations of each method in terms of F-scores are shown for each dataset. Results are sorted by F-score on BabelNet, top three values of each metric are boldfaced.

Method	# words	# synsets	# synonyms	RuWordNet			YARN		
				P	**R**	**F1**	**P**	**R**	**F1**
WATSET[CW$_{nolog}$, MCL]	83 092	55 369	332 727	0.120	**0.349**	**0.178**	0.402	**0.463**	**0.430**
WATSET[MCL, MCL]	83 092	36 217	403 068	0.111	0.341	0.168	0.405	0.455	**0.428**
WATSET[CW$_{top}$, CW$_{log}$]	83 092	55 319	341 043	0.116	**0.351**	0.174	0.386	**0.474**	**0.425**
MCL	83 092	21 973	353 848	0.155	0.291	**0.203**	0.550	0.340	0.420
WATSET[MCL, CW$_{top}$]	83 092	34 702	473 135	0.097	**0.361**	0.153	0.351	**0.496**	0.411
CW$_{nolog}$	83 092	19 124	672 076	0.087	0.342	0.139	0.364	0.451	0.403
MaxMax	83 092	27 011	461 748	**0.176**	0.261	**0.210**	0.582	0.195	0.292
CPM$_{k=3}$	15 555	4 000	45 231	**0.234**	0.072	0.111	**0.626**	0.060	0.110
ECO	83 092	67 645	18 362	**0.724**	0.034	0.066	**0.904**	0.002	0.004

Table 4: Results on Russian sorted by F-score on YARN, top three values of each metric are boldfaced.

isolates the neighborhood of polysemous words, which results in the recall drop in comparison to WATSET. CW operates faster due to a simplified update step. On the same graph, CW tends to produce larger clusters than MCL. This leads to a higher recall of "plain" CW as compared to the "plain" MCL, at the cost of lower precision.

Using MCL instead of CW for sense induction in WATSET expectedly produces more fine-grained senses. However, at the global clustering step, these senses erroneously tend to form coarse-grained synsets connecting unrelated senses of the ambiguous words. This explains the generally higher recall of WATSET[MCL, ·]. Despite the randomized nature of CW, variance across runs do not affect the overall ranking: The rank of different versions of CW (*log, nolog, top*) can change, while the rank of the best CW configuration compared to other methods remains the same.

The MaxMax algorithm shows mixed results. On the one hand, it outputs large clusters uniting more than hundred nodes. This inevitably leads to a high recall, as it is clearly seen in the results for Russian because such synsets still pass

under our cluster size threshold of 150 words. Its synsets on English datasets are even larger and get pruned, which results in low recall. On the other hand, smaller synsets having at most 10–15 words were identified correctly. MaxMax appears to be extremely sensible to edge weighting, which also complicates its practical use.

The CPM algorithm showed unsatisfactory results, emitting giant components encompassing thousands of words. Such clusters were automatically pruned, but the remaining clusters are relatively correctly built synsets, which is confirmed by the high values of precision. When increasing the minimal number of elements in the clique k, recall improves, but at the cost of a dramatic precision drop. We suppose that the network structure assumptions exploited by CPM do not accurately model the structure of our synonymy graphs.

Finally, the ECO method yielded the worst results because the most cluster candidates failed to pass through the constant threshold used for estimating whether a pair of words should be included in the same cluster. Most synsets produced by this method were trivial, i.e., containing only a single

1586

Resource		P	R	F1
BabelNet on WordNet	En	0.729	0.998	0.843
WordNet on BabelNet	En	0.998	0.699	0.822
YARN on RuWordNet	Ru	0.164	0.162	0.163
BabelNet on RuWordNet	Ru	0.348	0.409	0.376
RuWordNet on YARN	Ru	0.670	0.121	0.205
BabelNet on YARN	Ru	0.515	0.109	0.180

Table 5: Performance of lexical resources cross-evaluated against each other.

word. The remaining synsets for both languages have at most three words that have been connected by a chance due to the edge noising procedure used in this method resulting in low recall.

6 Discussion

On the absolute scores. The results obtained on all gold standards (Figure 3) show similar trends in terms of relative ranking of the methods. Yet absolute scores of YARN and RuWordNet are substantially different due to the inherent difference of these datasets. RuWordNet is more domain-specific in terms of vocabulary, so our input set of generic synonymy dictionaries has a limited coverage on this dataset. On the other hand, recall calculated on YARN is substantially higher as this resource was manually built on the basis of synonymy dictionaries used in our experiments.

The reason for low absolute numbers in evaluations is due to an inherent vocabulary mismatch between the input dictionaries of synonyms and the gold datasets. To validate this hypothesis, we performed a cross-resource evaluation presented in Table 5. The low performance of the cross-evaluation of the two resources supports the hypothesis: no single resource for Russian can obtain high recall scores on another one. Surprisingly, even BabelNet, which integrates most of available lexical resources, still does not reach a recall substantially larger than 0.5.[20] Note that the results of this cross-dataset evaluation are not directly comparable to results in Table 4 since in our experiments we use much smaller input dictionaries than those used by BabelNet.

On sparseness of the input dictionary. Table 6 presents some examples of the obtained synsets of various sizes for the top WATSET configuration on both languages. As one might observe, the qual-

---

[20]We used BabelNet 3.7 extracting all 3 497 327 synsets that were marked as Russian.

ity of the results is highly plausible. However, one limitation of all approaches considered in this paper is the dependence on the completeness of the input dictionary of synonyms. In some parts of the input synonymy graph, important bridges between words can be missing, leading to smaller-than-desired synsets. A promising extension of the present methodology is using distributional models to enhance connectivity of the graph by cautiously adding extra relations.

Size	Synset
2	{*decimal point, dot*}
3	{*gullet, throat, food pipe*}
4	{*microwave meal, ready meal, TV dinner, frozen dinner*}
5	{*objective case, accusative case, oblique case, object case, accusative*}
6	{*radio theater, dramatized audiobook, audio theater, radio play, radio drama, audio play*}

Table 6: Sample synsets induced by the WATSET[MCL, MCL] method for English.

7 Conclusion

We presented a new robust approach to fuzzy graph clustering that relies on hard graph clustering. Using ego network clustering, the nodes belonging to several local communities are split into several nodes each belonging to one community. The transformed "disambiguated" graph is then clustered using an efficient hard graph clustering algorithm, obtaining a fuzzy clustering as the result. The disambiguated graph facilitates clustering as it contains fewer hubs connecting unrelated nodes from different communities. We apply this meta clustering algorithm to the task of synset induction on two languages, obtaining the best results on three datasets and competitive results on one dataset in terms of F-score as compared to five state-of-the-art graph clustering methods.

References

Nikolay Abramov. 1999. *The dictionary of Russian synonyms and semantically related expressions [Slovar' russkikh sinonimov i skhodnykh po smyslu vyrazhenii].* Russian Dictionaries [Russkie slovari], Moscow, Russia, 7th edition. In Russian.

Marco Baroni and Alessandro Lenci. 2010. Distributional Memory: A General Framework for Corpus-based Semantics. *Computational Linguistics* 36(4):673–721. https://doi.org/10.1162/coli_a_00016.

Chris Biemann. 2006. Chinese Whispers: An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing.* Association for Computational Linguistics, New York City, NY, USA, TextGraphs-1, pages 73–80. http://dl.acm.org/citation.cfm?id=1654774.

Chris Biemann. 2012. *Structure Discovery in Natural Language.* Theory and Applications of Natural Language Processing. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-25923-4.

Chris Biemann and Martin Riedl. 2013. Text: now in 2D! A framework for lexical expansion with contextual similarity. *Journal of Language Modelling* 1(1):55–95. https://doi.org/10.15398/jlm.v1i1.60.

Immanuel M. Bomze, Marco Budinich, Panos M. Pardalos, and Marcello Pelillo. 1999. The maximum clique problem. In *Handbook of Combinatorial Optimization*, Springer US, pages 1–74. https://doi.org/10.1007/978-1-4757-3023-4_1.

Pavel Braslavski, Dmitry Ustalov, Mukhin Mukhin, and Yuri Kiselev. 2016. YARN: Spinning-in-Progress. In *Proceedings of the 8th Global WordNet Conference.* Global WordNet Association, Bucharest, Romania, GWC 2016, pages 58–65. http://gwc2016.racai.ro/procedings.pdf.

Antonio Di Marco and Roberto Navigli. 2012. Clustering and Diversifying Web Search Results with Graph-Based Word Sense Induction. *Computational Linguistics* 39(3):709–754. https://doi.org/10.1162/COLI_a_00148.

Vyachelav G. Dikonov. 2013. Development of lexical basis for the Universal Dictionary of UNL Concepts. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue".* RGGU, Moscow, volume 12 (19), pages 212–221. http://www.dialog-21.ru/media/1238/dikonovv.pdf.

Beate Dorow and Dominic Widdows. 2003. Discovering Corpus-Specific Word Senses. In *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics - Volume 2.* Association for Computational Linguistics, Budapest, Hungary, EACL '03, pages 79–82. https://doi.org/10.3115/1067737.1067753.

Martin Everett and Stephen P. Borgatti. 2005. Ego network betweenness. *Social Networks* 27(1):31–38. https://doi.org/10.1016/j.socnet.2004.11.007.

Stefano Faralli, Alexander Panchenko, Chris Biemann, and Simone P. Ponzetto. 2016. Linked Disambiguated Distributional Semantic Networks. In *The Semantic Web – ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II*. Springer International Publishing, Cham, pages 56–64. https://doi.org/10.1007/978-3-319-46547-0_7.

David Gfeller, Jean-Cédric Chappelier, and Paulo De Los Rios. 2005. Synonym Dictionary Improvement through Markov Clustering and Clustering Stability. In *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis.* pages 106–113. https://conferences.telecom-bretagne.eu/asmda2005/IMG/pdf/proceedings/106.pdf.

Hugo Gonçalo Oliveira and Paolo Gomes. 2014. ECO and Onto.PT: a flexible approach for creating a Portuguese wordnet automatically. *Language Resources and Evaluation* 48(2):373–393. https://doi.org/10.1007/s10579-013-9249-9.

Zhiguo Gong, Chan Wa Cheang, and U. Leong Hou. 2005. Web Query Expansion by WordNet. In *Proceedings of the 16th International Conference on Database and Expert Systems Applications - DEXA '05*, Springer Berlin Heidelberg, Copenhagen, Denmark, pages 166–175. https://doi.org/10.1007/11546924_17.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics.* Association for Computational Linguistics, Avignon, France, EACL '12, pages 580–590. http://www.aclweb.org/anthology/E12-1059.

Iryna Gurevych, Judith Eckle-Kohler, and Michael Matuschek. 2016. *Linked Lexical Knowledge Bases: Foundations and Applications.* Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Kris Heylen, Yves Peirsman, Dirk Geeraerts, and Dirk Speelman. 2008. Modelling Word Similarity: an Evaluation of Automatic Synonymy Extraction Algorithms. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation.* European Language Resources Association, Marrakech, Morocco, LREC 2008, pages 3243–3249. http://www.lrec-conf.org/proceedings/lrec2008/pdf/818_paper.pdf.

David Hope and Bill Keller. 2013. MaxMax: A Graph-Based Soft Clustering Algorithm Applied to Word Sense Induction. In *Computational Linguistics*

and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part I, Springer Berlin Heidelberg, Berlin, Heidelberg, pages 368–381. https://doi.org/10.1007/978-3-642-37247-6_30.

David Jurgens and Ioannis Klapaftis. 2013. SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses. In Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, Atlanta, GA, USA, pages 290–299. http://www.aclweb.org/anthology/S13-2049.

Yuri Kiselev, Sergey V. Porshnev, and Mikhail Mukhin. 2015. Current Status of Russian Electronic Thesauri: Quality, Completeness and Availability [Sovremennoe sostoyanie elektronnykh tezaurusov russkogo yazyka: kachestvo, polnota i dostupnost']. Programmnaya Ingeneria 6:34–40. In Russian. http://novtex.ru/prin/full/06_2015.pdf.

Andrew A. Krizhanovsky and Alexander V. Smirnov. 2013. An approach to automated construction of a general-purpose lexical ontology based on Wiktionary. Journal of Computer and Systems Sciences International 52(2):215–225. https://doi.org/10.1134/S1064230713020068.

Cody Kwok, Oren Etzioni, and Daniel S. Weld. 2001. Scaling Question Answering to the Web. ACM Transactions on Information Systems 19(3):242–262. https://doi.org/10.1145/502115.502117.

Dekang Lin. 1998. An Information-Theoretic Definition of Similarity. In Proceedings of the Fifteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., Madison, WI, USA, ICML '98, pages 296–304. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.1832&rep=rep1&type=pdf.

Natalia Loukachevitch. 2011. Thesauri in information retrieval tasks [Tezaurusy v zadachakh informatsionnogo poiska]. Moscow University Press [Izd-vo MGU], Moscow, Russia. In Russian.

Natalia V. Loukachevitch, German Lashevich, Anastasia A. Gerasimova, Vladimir V. Ivanov, and Boris V. Dobrov. 2016. Creating Russian WordNet by Conversion. In Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue". RSUH, Moscow, Russia, pages 405–415. http://www.dialog-21.ru/media/3409/loukachevitchnvetal.pdf.

Suresh Manandhar, Ioannis Klapaftis, Dmitriy Dligach, and Sameer Pradhan. 2010. SemEval-2010 Task 14: Word Sense Induction & Disambiguation. In Proceedings of the 5th International Workshop on Semantic Evaluation. Association for Computational Linguistics, Uppsala, Sweden, pages 63–68. http://www.aclweb.org/anthology/S10-1011.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Advances in Neural Information Processing Systems 26, Curran Associates, Inc., Harrahs and Harveys, NV, USA, pages 3111–3119. https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.

George A. Miller. 1995. WordNet: A Lexical Database for English. Communications of the ACM 38(11):39–41. https://doi.org/10.1145/219717.219748.

Roberto Navigli and Simone P. Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence 193:217–250. https://doi.org/10.1016/j.artint.2012.07.001.

Gergely Palla, Imre Derenyi, Illes Farkas, and Tamas Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435:814–818. https://doi.org/10.1038/nature03607.

Alexander Panchenko. 2011. Comparison of the Baseline Knowledge-, Corpus-, and Web-based Similarity Measures for Semantic Relations Extraction. In Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics. Association for Computational Linguistics, Edinburgh, UK, pages 11–21. http://www.aclweb.org/anthology/W11-2502.

Alexander Panchenko, Eugen Ruppert, Stefano Faralli, Simone P. Ponzetto, and Chris Biemann. 2017a. Unsupervised Does Not Mean Uninterpretable: The Case for Word Sense Induction and Disambiguation. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Association for Computational Linguistics, Valencia, Spain, pages 86–98. http://www.aclweb.org/anthology/E17-1009.

Alexander Panchenko, Dmitry Ustalov, Nikolay Arefyev, Denis Paperno, Natalia Konstantinova, Natalia Loukachevitch, and Chris Biemann. 2017b. Human and Machine Judgements for Russian Semantic Relatedness. In Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers. Springer International Publishing, Yekaterinburg, Russia, pages 221–235. https://doi.org/10.1007/978-3-319-52920-2_21.

Patrick Pantel and Dekang Lin. 2002. Discovering Word Senses from Text. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Edmonton, Alberta, Canada, KDD '02, pages 613–619. https://doi.org/10.1145/775047.775138.

Maria Pelevina, Nikolay Arefiev, Chris Biemann, and Alexander Panchenko. 2016. Making Sense of Word Embeddings. In *Proceedings of the 1st Workshop on Representation Learning for NLP*. Association for Computational Linguistics, Berlin, Germany, pages 174–183. http://anthology.aclweb.org/W16-1620.

Stijn van Dongen. 2000. *Graph Clustering by Flow Simulation*. Ph.D. thesis, University of Utrecht.

Jean Véronis. 2004. HyperLex: lexical cartography for information retrieval. *Computer Speech & Language* 18(3):223–252. https://doi.org/10.1016/j.csl.2004.05.002.

Torsten Zesch, Christof Müller, and Iryna Gurevych. 2008. Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*. European Language Resources Association, Marrakech, Morocco, pages 1646–1652. http://www.lrec-conf.org/proceedings/lrec2008/pdf/420_paper.pdf.

Guangyou Zhou, Yang Liu, Fang Liu, Daojian Zeng, and Jun Zhao. 2013. Improving Question Retrieval in Community Question Answering Using World Knowledge. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. AAAI Press, Beijing, China, IJCAI '13, pages 2239–2245. https://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/download/6581/7029.