Analysis of the Wikipedia Category Graph for NLP Applications

Torsten Zesch and Iryna Gurevych
Ubiquitous Knowledge Processing Group
Telecooperation Division
Darmstadt University of Technology, Hochschulstraße 10
D-64289 Darmstadt, Germany
{zesch,gurevych} (at) tk.informatik.tu-darmstadt.de

Abstract

In this paper, we discuss two graphs in Wikipedia (i) the article graph, and (ii) the category graph. We perform a graph-theoretic analysis of the category graph, and show that it is a scale-free, small world graph like other well-known lexical semantic networks. We substantiate our findings by transferring semantic relatedness algorithms defined on WordNet to the Wikipedia category graph. To assess the usefulness of the category graph as an NLP resource, we analyze its coverage and the performance of the transferred semantic relatedness algorithms.

1 Introduction

Wikipedia[1] is a free multi-lingual online encyclopedia that is constructed in a collaborative effort of voluntary contributors and still grows exponentially. During this process, Wikipedia has probably become the largest collection of freely available knowledge. A part of this knowledge is encoded in the network structure of Wikipedia pages. In particular, Wikipedia articles form a network of semantically related terms, while the categories are organized in a taxonomy-like structure called **Wikipedia Category Graph** (**WCG**).

In this paper, we perform a detailed analysis of the WCG by computing a set of graph-theoretic parameters, and comparing them with the parameters reported for well-known graphs and classical lexical semantic networks. We show that the WCG, which is constructed collaboratively, shares many properties with other lexical semantic networks, such as

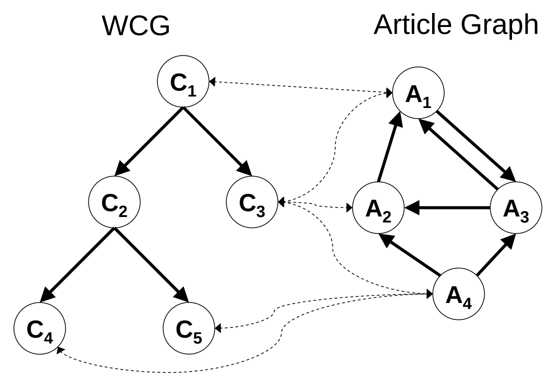

Figure 1: Relations between article graph and WCG.

WordNet (Fellbaum, 1998) or Roget's Thesaurus[2] that are constructed by expert authors. This implies that the WCG can be used as a resource in NLP applications, where other semantic networks have been traditionally employed.

To further evaluate this issue, we adapt algorithms for computing semantic relatedness on classical semantic networks like WordNet to the WCG. We evaluate their performance on the task of computing semantic relatedness using three German datasets, and show that WCG based algorithms perform very well.

Article graph Wikipedia articles are heavily linked, as links can be easily inserted while editing an article. If we treat each article as a node, and each link between articles as an edge running from one node to another, then Wikipedia articles form a directed graph (see right side of Figure 1). The article graph has been targeted by numerous studies, and is not addressed in this paper. Buriol et al. (2006) analyze the development of the article graph over time, and find that some regions are fairly stable, while others are advancing quickly. Zlatic et al.

[1] http://www.wikipedia.org

[2] http://thesaurus.reference.com

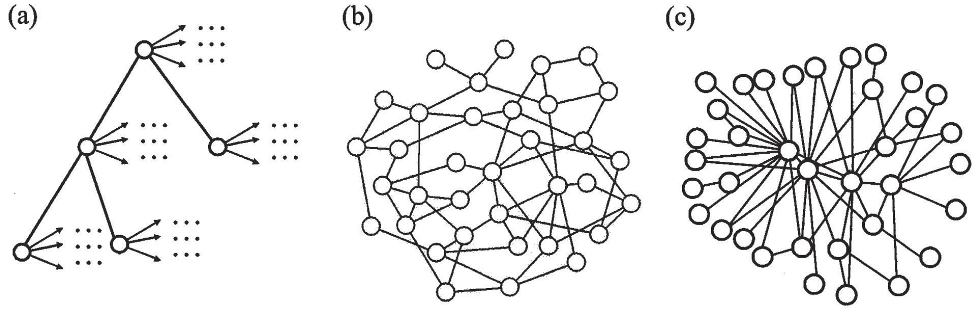

Figure 2: Structures of semantic networks after Steyvers and Tenenbaum (2005). a) a taxonomy, b) an arbitrary graph, c) scale-free, small-world graph.

(2006) give a comprehensive overview of the graph parameters for the largest languages in Wikipedia. Capocci et al. (2006) study the growth of the article graph and show that it is based on preferential attachment (Barabasi and Albert, 1999). Voss (2005) shows that the article graph is scale-free and grows exponentially.

Category graph Categories in Wikipedia are organized in a taxonomy-like structure (see left side of Figure 1 and Figure 2-a). Each category can have an arbitrary number of subcategories, where a subcategory is typically established because of a hyponymy or meronymy relation. For example, a category *vehicle* has subcategories like *aircraft* or *watercraft*. Thus, the WCG is very similar to semantic wordnets like WordNet or GermaNet (Kunze, 2004). As Wikipedia does not strictly enforce a taxonomic category structure, cycles and disconnected categories are possible, but rare. In the snapshot of the German Wikipedia[3] from May 15, 2006, the largest connected component in the WCG contains 99,8% of all category nodes, as well as 7 cycles.

In Wikipedia, each article can link to an arbitrary number of categories, where each category is a kind of semantic tag for that article. A category backlinks to all articles in this category. Thus, article graph and WCG are heavily interlinked (see Figure 1), and most studies (Capocci et al., 2006; Zlatic et al., 2006) have not treated them separately. However, the WCG should be treated separately, as it differs from the article graph. Article links are established because of any kind of relation between articles, while links between categories are typically established because of hyponymy or meronymy relations.

Holloway et al. (2005) create and visualize a category map based on co-occurrence of categories. Voss (2006) pointed out that the WCG is a kind of thesaurus that combines collaborative tagging and hierarchical indexing. Zesch et al. (2007a) identified the WCG as a valueable source of lexical semantic knowledge, but did not analytically analyze its properties. However, even if the WCG seems to be very similar to other semantic wordnets, a graph-theoretic analysis of the WCG is necessary to substantiate this claim. It is carried out in the next section.

2 Graph-theoretic Analysis of the WCG

A graph-theoretic analysis of the WCG is required to estimate, whether graph based semantic relatedness measures developed for semantic wordnets can be transferred to the WCG. This is substantiated in a case study on computing semantic relatedness in section 4.

For our analysis, we treat the directed WCG as an undirected **graph** $G := (V, E)$,[4] as the relations connecting categories are reversible. V is a set of vertices or nodes. E is a set of unordered pairs of distinct vertices, called edges. Each page is treated as a **node** n, and each link between pages is modeled as an **edge** e running between two nodes.

Following Steyvers and Tenenbaum (2005), we characterize the graph structure of a lexical semantic resource in terms of a set of graph parameters: The

[3]Wikipedia can be downloaded from `http://download.wikimedia.org/`

[4]Newman (2003) gives a comprehensive overview about the theoretical aspects of graphs.

PARAMETER	$Actor$	$Power$	$C.elegans$	AN	$Roget$	$WordNet$	$Wiki_{Art}$	**WCG**
$\|V\|$	225,226	4,941	282	5,018	9,381	122,005	190,099	27,865
D	-	-	-	5	10	27	-	17
$\overline{k}$	61.0	2.67	14.0	22.0	49.6	4.0	-	3.54
γ	-	-	-	3.01	3.19	3.11	2.45	2.12
$\overline{L}$	3.65	18.7	2.65	3.04	5.60	10.56	3.34	7.18
$\overline{L}_{random}$	2.99	12.4	2.25	3.03	5.43	10.61	$\sim$ 3.30	$\sim$ 8.10
C	0.79	0.08	0.28	0.186	0.87	0.027	$\sim$ 0.04	0.012
C_{random}	0.0003	0.005	0.05	0.004	0.613	0.0001	$\sim$ 0.006	0.0008

Table 1: Parameter values for different graphs.

Values for *Actor* (collaboration graph of actors in feature films), *Power* (the electrical power grid of the western United States) and *C.elegans* (the neural network of the nematode worm C. elegans) are from Watts and Strogatz (1998). Values for AN (a network of word associations by Nelson et al. (1998)), Roget's thesaurus and WordNet are from Steyvers and Tenenbaum (2005). Values for $Wiki_{art}$ (German Wikipedia article graph) are from Zlatic et al. (2006). We took the values for the page set labelled M on their website containing 190,099 pages for German, as it comes closest to a graph of only articles. Values marked with '-' in the table were not reported in the studies. The values for the WCG are computed in this study.

degree k of a node is the number of edges that are connected with this node. Averaging over all nodes gives the **average degree** $\overline{k}$. The degree distribution $P(k)$ is the probability that a random node will have degree k. In some graphs (like the WWW), the degree distribution follows a power law $P(k) \approx k^{-\gamma}$ (Barabasi and Albert, 1999). We use the **power law exponent** γ as a graph parameter.

A **path** $p_{i,j}$ is a sequence of edges that connects a node n_i with a node n_j. The **path length** $l(p_{i,j})$ is the number of edges along that path. There can be more than one path between two nodes. The **shortest path length** L is the minimum of all these paths, i.e. $L_{i,j} = \min l(p_{i,j})$. Averaging over all nodes gives the **average shortest path length** $\overline{L}$. The **diameter** D is the maximum of the shortest path lengths between all pairs of nodes in the graph.

The **cluster coefficient** of a certain node n_i can be computed as

$$C_i = \frac{T_i}{\frac{k_i(k_i-1)}{2}} = \frac{2T_i}{k_i(k_i-1)}$$

where T_i refers to the number of edges between the neighbors of node n_i and $k_i(k_i-1)/2$ is the maximum number of edges that can exist between the k_i neighbors of node n_i.[5] The cluster coefficient C for the whole graph is the average of all C_i. In a fully connected graph, the cluster coefficient is 1.

[5] In a social network, the cluster coefficient measures how many of my friends (neighboring nodes) are friends themselves.

For our analysis, we use a snapshot of the German Wikipedia from May 15, 2006. We consider only the largest connected component of the WCG that contains 99,8% of the nodes. Table 1 shows our results on the WCG as well as the corresponding values for other well-known graphs and lexical semantic networks. We compare our empirically obtained values with the values expected for a random graph. Following Zlatic et al. (2006), the cluster coefficient C for a random graph is

$$C_{random} = \frac{(\overline{k}^2 - \overline{k})^2}{|V|\overline{k}}$$

The average path length for a random network can be approximated as $\overline{L}_{random} \approx \log|V| \;/\; \log\overline{k}$ (Watts and Strogatz, 1998).

From the analysis, we conclude that all graphs in Table 1 are small world graphs (see Figure 2-c). Small world graphs (Watts and Strogatz, 1998) contain local clusters that are connected by some long range links leading to low values of $\overline{L}$ and D. Thus, small world graphs are characterized by (i) small values of $\overline{L}$ (typically $\overline{L} \gtrsim \overline{L}_{random}$), together with (ii) large values of C ($C \gg C_{random}$).

Additionally, all semantic networks are scale-free graphs, as their degree distribution follows a power law. Structural commonalities between the graphs in Table 1 are assumed to result from the growing process based on preferential attachment (Capocci et al., 2006).

Our analysis shows that $WordNet$ and the WCG are (i) scale-free, small world graphs, and (ii) have a very similar parameter set. Thus, we conclude that algorithms designed to work on the graph structure of WordNet can be transferred to the WCG.

In the next section, we introduce the task of computing semantic relatedness on graphs and adapt existing algorithms to the WCG. In section 4, we evaluate the transferred algorithms with respect to correlation with human judgments on SR, and coverage.

3 Graph Based Semantic Relatedness Measures

Semantic similarity (**SS**) is typically defined via the lexical relations of synonymy ($automobile$ – car) and hypernymy ($vehicle$ – car), while **semantic relatedness** (**SR**) is defined to cover any kind of lexical or functional association that may exist between two words (Budanitsky and Hirst, 2006). Dissimilar words can be semantically related, e.g. via functional relationships ($night$ – $dark$) or when they are antonyms ($high$ – low). Many NLP applications require knowledge about semantic relatedness rather than just similarity (Budanitsky and Hirst, 2006).

We introduce a number of competing approaches for computing semantic relatedness between words using a graph structure, and then discuss the changes that are necessary to adapt semantic relatedness algorithms to work on the WCG.

3.1 Wordnet Based Measures

A multitude of semantic relatedness measures working on semantic networks has been proposed.

Rada et al. (1989) use the path length (**PL**) between two nodes (measured in edges) to compute semantic relatedness.

$$dist_{PL} = l(n_1, n_2)$$

Leacock and Chodorow (1998, ***LC***) normalize the path-length with the depth of the graph,

$$sim_{LC}(n_1, n_2) = -\log \frac{l(n_1, n_2)}{2 \times depth}$$

where $depth$ is the length of the longest path in the graph.

Wu and Palmer (1994, ***WP***) introduce a measure that uses the notion of a lowest common subsumer of two nodes $lcs(n_1, n_2)$. In a directed graph, a lcs is the parent of both child nodes with the largest depth in the graph.

$$sim_{WP} = \frac{2\, depth(lcs)}{l(n_1, lcs) + l(n_2, lcs) + 2\, depth(lcs)}$$

Resnik (1995, ***Res***), defines semantic similarity between two nodes as the information content (IC) value of their lcs. He used the relative corpus frequency to estimate the information content value.

Jiang and Conrath (1997, ***JC***) additionally use the IC of the nodes.

$$dist_{JC}(n_1, n_2) = IC(n_1) + IC(n_2) - 2IC(lcs)$$

Note that JC returns a distance value instead of a similarity value.

Lin (1998, ***Lin***) defined semantic similarity using a formula derived from information theory.

$$sim_{Lin}(n_1, n_2) = 2 \times \frac{IC(lcs)}{IC(n_1) + IC(n_2)}$$

Because polysemous words may have more than one corresponding node in a semantic wordnet, the resulting semantic relatedness between two words w_1 and w_2 can be calculated as

$$SR = \begin{cases} \min\limits_{n_1 \in s(w_1), n_2 \in s(w_2)} dist(n_1, n_2) & \text{path} \\ \max\limits_{n_1 \in s(w_1), n_2 \in s(w_2)} sim(n_1, n_2) & \text{IC} \end{cases}$$

where $s(w_i)$ is the set of nodes that represent senses of word w_i. That means, the relatedness of two words is equal to that of the most related pair of nodes.

3.2 Adapting SR Measures to Wikipedia

Unlike other wordnets, nodes in the WCG do not represent synsets or single terms, but a generalized concept or category. Therefore, we cannot use the WCG directly to compute SR. Additionally, the WCG would not provide sufficient coverage, as it is relatively small. Thus, transferring SR measures to the WCG requires some modifications. The task of estimating SR between terms is casted to the task of SR between Wikipedia articles devoted to these terms. SR between articles is measured via the categories assigned to these articles.

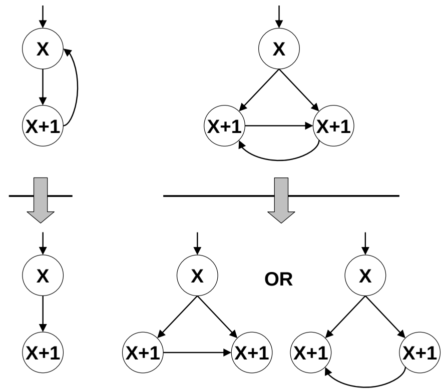

Figure 3: Breaking cycles in the WCG.

We define C_1 and C_2 as the set of categories assigned to article a_i and a_j, respectively. We then determine the SR value for each category pair (c_k, c_l) with $c_k \in C_1$ and $c_l \in C_2$. We choose the best value among all pairs (c_k, c_l), i.e. the minimum for path based and the maximum for information content based measures.

$$SR_{best} = \begin{cases} \min\limits_{c_k \in C_1, c_l \in C_2} (sr(c_k, c_l)) & \text{path based} \\ \max\limits_{c_k \in C_1, c_l \in C_2} (sr(c_k, c_l)) & \text{IIC based} \end{cases}$$

See (Zesch et al., 2007b) for a more detailed description of the adaptation process.

We substitute Resnik's information content with the **intrinsic information content** (**IIC**) by Seco et al. (2004) that is computed only from structural information of the underlying graph. It yields better results and is corpus independent. The IIC of a node n_i is computed as a function of its hyponyms,

$$IIC(n) = 1 - \frac{\log(hypo(n_i) + 1}{\log(|C|)}$$

where $hypo(n_i)$ is the number of hyponyms of node n_i and $|C|$ is the number of nodes in the taxonomy.

Efficiently counting the hyponyms of a node requires to break cycles that may occur in a WCG. We perform a colored depth-first-search to detect cycles, and break them as visualized in Figure 3. A link pointing back to a node closer to the top of the graph is deleted, as it violates the rule that links in the WCG typically express hyponymy or meronymy relations. If the cycle occurs between nodes on the same level, we cannot decide based on that rule and simply delete one of the links running on the same level. This strategy never disconnects any nodes from a connected component.

4 Semantic Relatedness Experiments

A commonly accepted method for evaluating SR measures is to compare their results with a gold standard dataset based on human judgments on word pairs.[6]

4.1 Datasets

To create gold standard datasets for evaluation, human annotators are asked to judge the relatedness of presented word pairs. The average annotation scores are correlated with the SR values generated by a particular measure.

Several datasets for evaluation of semantic relatedness or semantic similarity have been created so far (see Table 2). Rubenstein and Goodenough (1965) created a dataset with 65 English noun pairs (**RG65** for short). A subset of RG65 has been used for experiments by Miller and Charles (1991, **MC30**) and Resnik (1995, **Res30**).

Finkelstein et al. (2002) created a larger dataset for English containing 353 pairs (**Fin353**), that has been criticized by Jarmasz and Szpakowicz (2003) for being culturally biased. More problematic is that Fin353 consists of two subsets, which have been annotated by a different number of annotators. We performed further analysis of their dataset and found that the inter-annotator agreement[7] differs considerably. These results suggest that further evaluation based on this data should actually regard it as two independent datasets.

As Wikipedia is a multi-lingual resource, we are not bound to English datasets. Several German datasets are available that are larger than the existing English datasets and do not share the problems of the Finkelstein datasets (see Table 2). Gurevych (2005) conducted experiments with a German translation of an English dataset (Rubenstein and Goodenough, 1965), but argued that the dataset is too small and only contains noun-noun pairs connected

[6]Note that we do not use multiple-choice synonym question datasets (Jarmasz and Szpakowicz, 2003), as this is a different task, which is not addressed in this paper.

[7]We computed the correlation for all annotators pairwise and summarized the values using a Fisher Z-value transformation.

Dataset	Year	Language	# Pairs	POS	Type	Scores	# Subjects	Correlation r	
								Inter	Intra
RG65	1965	English	65	N	SS	continuous 0–4	51	-	.850
MC30	1991	English	30	N	SS	continuous 0–4	38	-	-
Res30	1995	English	30	N	SS	continuous 0–4	10	.903	-
Fin353	2002	English	353	N, V, A	SR	continuous 0–10	13/16	-	-
			153				13	.731	-
			200				16	.549	-
Gur65	2005	German	65	N	SS	discrete {0,1,2,3,4}	24	.810	-
Gur350	2006	German	350	N, V, A	SR	discrete {0,1,2,3,4}	8	.690	-
ZG222	2006	German	222	N, V, A	SR	discrete {0,1,2,3,4}	21	.490	.647

Table 2: Comparison of German datasets used for evaluating semantic relatedness.

by either synonymy or hyponymy. Thus, she created a larger German dataset containing 350 word pairs (**Gur350**). It contains nouns, verbs and adjectives that are connected by classical and non-classical relations (Morris and Hirst, 2004). However, word pairs for this dataset are biased towards strong classical relations, as they were manually selected. Thus, Zesch and Gurevych (2006) used a semi-automatic process to create word pairs from domain-specific corpora. The resulting **ZG222** dataset contains 222 word pairs that are connected by all kinds of lexical semantic relations. Hence, it is particularly suited for analyzing the capability of a measure to estimate SR.

4.2 Results and Discussion

Figure 4 gives an overview of our experimental results of evaluating SR measures based on the WCG on three German datasets. We use Pearson's product moment correlation r to compare the results with human judgments. From each dataset, we only use word pairs where Wikipedia articles corresponding to these words are available (see section 4.3 for a detailed discussion of word pair coverage). For comparison, we give the best results obtained by GermaNet based measures (abbreviated as **GN**).[8]

Our results show that the graph-based SR measures have been successfully transferred to the WCG. Results on the Gur65 dataset (containing only word pairs connected by strong classical relations) are lower than values computed using GermaNet. This is to be expected, as the WCG is created collaboratively without strictly enforcing a certain type

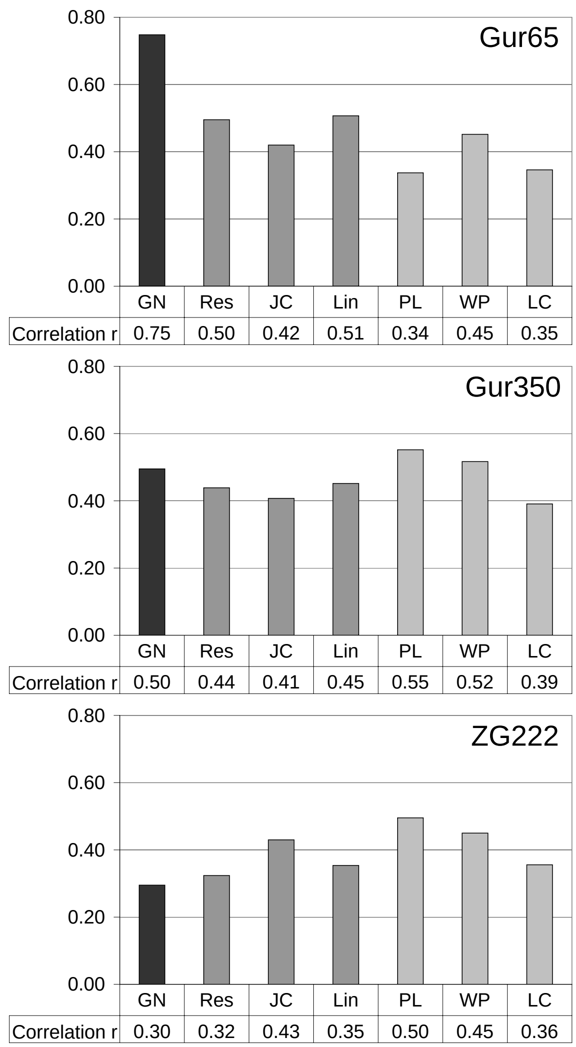

Figure 4: Correlations on different datasets.

[8]Additionally, Table 2 gives the inter annotator agreement for each subset. It constitutes an upper bound of a measure's performance on a certain dataset.

of semantic relation between categories, while GermaNet is carefully modelled to represent the strong classical relations captured by Gur65. Results on the two other datasets, which contain a majority of word pairs connected by non-classical semantic relations, show that the WCG is better suited than GermaNet to estimate SR.

Performance of WCG based measures depends on the dataset and the kind of knowledge used. IIC based measures (Res, JC and Lin) outperform path based measures (PL, LC and WP) on the Gur65 dataset, while path based measures are clearly better on SR datasets (Gur350 and ZG222). The impressive performance of the simple PL measure on the SR datasets cannot be explained with the structural properties of the WCG, as they are very similar to those of other semantic networks. Semantically related terms are very likely to be categorized under the same category, resulting in short path lengths leading to high SR. The generalization process that comes along with classification seems to capture the phenomenon of SR quite well. As each article can have many categories, different kinds of semantic relations between terms can be established, but the type of relation remains unknown.

4.3 Coverage of Word Pairs

If the WCG is to be used as a lexical semantic resource in large scale NLP applications, it should provide broad coverage. As was described in section 3.2, computing SR using the WCG relies on categories assigned to articles. Thus, we consider a word to be covered by the WCG, if there is a categorized article with matching title.

Table 3 gives an overview of the number of word pairs covered in GermaNet or the WCG. Only few words from Gur65 were not found in one of the resources. This proportion is much higher for Gur350 and ZG222, as these datasets contain many domain specific terms that are badly covered in GermaNet, and many word pairs containing verbs and adjectives that are badly covered in the WCG.[9] A number of word pairs (mostly containing combinations of verbs or adjectives) were found neither in GermaNet nor Wikipedia (see GN $\cup$ WCG). If we consider only noun-noun pairs (NN), the coverage of Wikipedia exceeds that of GermaNet. The high proportion of word pairs that are either only found in GermaNet or in the WCG indicates that they are partially complementary with respect to covered vocabulary.

[9] Resulting from an editorial decision, Wikipedia only contains articles devoted to terms of encyclopedic interest - mainly nouns. Adjectives and verbs redirect to their corresponding nouns, if they are covered at all.

5 Conclusion

In this paper, we performed a graph-theoretic analysis of the Wikipedia Category Graph and showed that it is a scale-free, small-world graph, like other semantic networks such as WordNet or Roget's thesaurus. From this result, we concluded that the WCG can be used for NLP tasks, where other semantic networks have been traditionally employed. As Wikipedia is a multi-lingual resource, this enables the transfer of NLP algorithms to languages that do not have well-developed semantic wordnets.

To substantiate this claim, we described how measures of semantic relatedness operating on semantic wordnets, like WordNet or GermaNet, can be adapted to work on the WCG. We showed that the WCG is well suited to estimate SR between words. This is due to the categorization process that connects terms which would not be closely related in a taxonomic wordnet structure. Consequently, GermaNet outperforms the WCG on the task of estimating semantic similarity. Furthermore, the WCG cannot be used for tasks that require knowledge about the exact type of semantic relation.

We performed an analysis of the coverage of Wikipedia. It covers nouns very well, but is less suited to compute semantic relatedness across parts-of-speech. In this case, conventional semantic wordnets are likely to provide a better knowledge source. In Zesch et al. (2007b), we show that knowledge from wordnets and from Wikipedia is complementary, and can be combined to improve the performance on the SR task. As the simple PL measure performs remarkably well on the SR datasets, in our future work, we will also consider computing SR using the path length on the Wikipedia article graph rather than on the WCG.

Acknowledgments

This work was supported by the German Research Foundation under the grant "Semantic Information

DATASET	# PAIRS	GN	WCG	GN ∪ WCG	GN \ WCG	WCG \ GN	GN ∩ WCG
Gur65	65	57	61	65	4	8	53
Gur350	350	208	161	248	87	40	121
Gur350 NN	173	109	115	129	14	20	95
ZG222	222	86	86	118	32	30	56
ZG222 NN	119	57	61	73	12	16	45

Table 3: Number of covered word pairs based on GermaNet (GN) and the WCG on different datasets.

Retrieval from Texts in the Example Domain Electronic Career Guidance" (SIR), GU 798/1-2.

References

A. Barabasi and R. Albert. 1999. Emergence of scaling in random networks. *Science*, 286:509–512.

A. Budanitsky and G. Hirst. 2006. Evaluating WordNet-based Measures of Semantic Distance. *Computational Linguistics*, 32(1).

L. Buriol, C. Castillo, D. Donato, S. Leonardi, and S. Millozzi. 2006. Temporal Analysis of the Wikigraph. In *Proc. of Web Intelligence*, Hong Kong.

A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. 2006. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, 74:036116.

C. Fellbaum. 1998. *WordNet An Electronic Lexical Database*. MIT Press, Cambridge, MA.

L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, and G. Wolfman. 2002. Placing Search in Context: The Concept Revisited. *ACM TOIS*, 20(1):116–131.

I. Gurevych. 2005. Using the Structure of a Conceptual Network in Computing Semantic Relatedness. In *Proc. of IJCNLP*, pages 767–778.

T. Holloway, M. Bozicevic, and K. Börner. 2005. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. *ArXiv Computer Science e-prints*, cs/0512085.

M. Jarmasz and S. Szpakowicz. 2003. Roget's thesaurus and semantic similarity. In *Proc. of RANLP*, pages 111–120.

J. J. Jiang and D. W. Conrath. 1997. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In *Proc. of the 10th International Conference on Research in Computational Linguistics*.

C. Kunze, 2004. *Computerlinguistik und Sprachtechnologie*, chapter Lexikalisch-semantische Wortnetze, pages 423–431. Spektrum Akademischer Verlag.

C. Leacock and M. Chodorow, 1998. *WordNet: An Electronic Lexical Database*, chapter Combining Local Context and WordNet Similarity for Word Sense Identification, pages 265–283. Cambridge: MIT Press.

D. Lin. 1998. An Information-Theoretic Definition of Similarity. In *Proc. of ICML*.

G. A. Miller and W. G. Charles. 1991. Contextual Correlates of Semantic Similarity. *Language and Cognitive Processes*, 6(1):1–28.

J. Morris and G. Hirst. 2004. Non-Classical Lexical Semantic Relations. In *Proc. of the Workshop on Computational Lexical Semantics, NAACL-HLT*.

D. L. Nelson, C. L. McEvoy, and T. A. Schreiber. 1998. The University of South Florida word association, rhyme, and word fragment norms. Technical report, U. of South Florida.

M. E. J. Newman. 2003. The structure and function of complex networks. *SIAM Review*, 45:167–256.

R. Rada, H. Mili, E. Bicknell, and M. Blettner. 1989. Development and Application of a Metric on Semantic Nets. *IEEE Trans. on Systems, Man, and Cybernetics,*, 19(1):17–30.

P. Resnik. 1995. Using Information Content to Evaluate Semantic Similarity. In *Proc. of IJCAI*, pages 448–453.

H. Rubenstein and J. B. Goodenough. 1965. Contextual Correlates of Synonymy. *Communications of the ACM*, 8(10):627–633.

N. Seco, T. Veale, and J. Hayes. 2004. An Intrinsic Information Content Metric for Semantic Similarity in WordNet. In *Proc. of ECAI*.

M. Steyvers and J. B. Tenenbaum. 2005. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29:41–78.

J. Voss. 2005. Measuring Wikipedia. In *Proc. of the 10th International Conference of the International Society for Scientometrics and Informetrics*, Stockholm, Sweden.

J. Voss. 2006. Collaborative thesaurus tagging the Wikipedia way. *ArXiv Computer Science e-prints*, cs/0604036.

D. J. Watts and S. H. Strogatz. 1998. Collective Dynamics of Small-World Networks. *Nature*, 393:440–442.

Z. Wu and M. Palmer. 1994. Verb Semantics and Lexical Selection. In *Proc. of ACL*, pages 133–138.

T. Zesch and I. Gurevych. 2006. Automatically Creating Datasets for Measures of Semantic Relatedness. In *Proc. of the Workshop on Linguistic Distances, ACL*, pages 16–24.

T. Zesch, I. Gurevych, and M. Mühlhäuser. 2007a. Analyzing and Accessing Wikipedia as a Lexical Semantic Resource. In *Proc. of Biannual Conference of the Society for Computational Linguistics and Language Technology*, pages 213–221.

T. Zesch, I. Gurevych, and M. Mühlhäuser. 2007b. Comparing Wikipedia and German Wordnet by Evaluating Semantic Relatedness on Multiple Datasets. In *Proc. of NAACL-HLT*, page (to appear).

V. Zlatic, M. Bozicevic, H. Stefancic, and M. Domazet. 2006. Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 74:016115.