The Modular Community Structure of Linguistic Predication Networks

Aaron Gerow
Computation Institute
University of Chicago
Chicago, IL, USA
gerow@uchicago.edu

James Evans
Dept. of Sociology & Computation Institute
University of Chicago
Chicago, IL, USA
jevans@uchicago.edu

Abstract

This paper examines the structure of linguistic predications in English text. Identified by the copular "is-a" form, predications assert category membership (hypernymy) or equivalence (synonymy) between two words. Because predication expresses ontological structure, we hypothesize that networks of predications will form modular groups. To measure this, we introduce a semantically motivated measure of *predication strength* to weight relevant predications observed in text. Results show that predications do indeed form modular structures without any weighting ($Q \approx 0.6$) and that using predication strength increases this modularity ($Q \approx 0.9$) without discarding low-frequency items. This high level of modularity supports the network-based analysis and the use of predication strength as a way to extract dense semantic clusters. Additionally, words' centrality within communities exhibits slight correlation with hypernym depths in WordNet, underscoring the ontological organization of predication.

1 Introduction & Background

Statistical patterns in language use are evident at many levels and have proved useful in an increasingly wide range of computational and cognitive applications. Statistical regularities offer a way to quantify and model how people create, encode and use knowledge about the world. Statements specifically about "what things are" (ie. ontological statements) offer uniquely transparent evidence about peoples' knowledge of the world. Our research adopts a corpus-based approach in which networks of predications are analyzed to assess the underlying structure of ontological assertions.

Word-word predications, observed as the copular `is-a` form in English, are important because, unlike most grammatical constructions that have few semantic constraints, predications tend to imply category membership or equivalence. Take (i) and (ii) for example:

(i) Safety is always a primary concern.
(ii) This organization is an institution where [...].

(i) is a category assertion (*safety* as a type of *concern*) and (ii) is an equivalence assertion (*organization* is an *institution*). Most predications can be interpreted as category memberships like (i); explicit articulation of equivalence is actually quite rare in language (Cimiano, 2006; Cimiano and Völker, 2005). Although some categorical predications are metaphorical, many of these are interpreted using category matching or analogical mapping processes (Glucksberg et al., 1997; Bowdle and Gentner, 2005). In both semantic interpretations, predications naturally form a directed network of words. Consisting primarily of category assertions, the structure of this network should exhibit a degree of natural clustering owing natural categories of the those things it represents.

Network representations of language have been used to describe a wide range of structures in language, including word-word and word-document co-occurrences, term collocations, dependency structure and named entity relations. Networks of grammatical relations have been found to differentiate word-classes (Ferrer i Cancho et al., 2004) and semantic networks can be used to model vocabulary growth (Steyvers and Tenenbaum, 2005). Co-occurrence networks, which are perhaps the most widely studied natural language network, are the foundation of many vector-space models (Landauer and Dumais, 1997; Turney et al., 2010) and can be used to mine synonyms (Cohen et al., 2005), disambiguate word senses (Agirre et

al., 2014; Biemann, 2006) and even help mark the quality of essays (Foltz et al., 1999). Spectral methods applied to linguistic networks have been used to differentiate languages (Ferrer i Cancho et al., 2004), word-classes (Sun and Korhonen, 2009) and genres of text (Ferrer i Cancho et al., 2007). Using spectral methods, research has also found that syntactic and semantic distributional similarity networks have considerably different structure (Biemann et al., 2009). The use of lexical graphs (networks of words) in particular, pre-dates modern NLP (Rapoport et al., 1966), though the approach continues to influence a variety of NLP and information retrieval tasks like summarization and retrieval (Erkan and Radev, 2004; Véronis, 2004). Network-based methods have even used community detection, similar to the algorithm described in this paper, to extract specialist terms from sets of multi-theme documents (Grineva et al., 2009) as well as unstructured texts (Gerow, 2014).

Because predications naturally form directed chains of ontological assertions, we hypothesize that their underlying structure is systematic and modular, given its representation of naturally organized things in the world. Our method employs community detection on networks of noun-noun predications as a way to assess the *overall* structure of predication, but it could be extended to hypernym and category extraction tasks (Hearst, 1992; Caraballo, 1999). Specifically, we test for community structure in predications and explore whether this structure becomes more highly resolved when using a semantic measure of *predication strength* introduced in the following section. We also predict that central nodes (i.e. words) in individual modules will correlate to categorical super-ordinance or hypernymy. Thus, we first seek to assess the *overall* community structure of predication, testing whether or not it is more resolved using a novel measure of *predication strength*. Second, within communities of predications, we compare the words' closeness centrality to their positions in WordNet's hypernym tree (Miller, 1995).

2 Method

Unlike co-occurrence networks, where words are related simply by proximity, predication networks are built using extracted grammatical relationships. The implied relationship in a co-occurrence network provides a natural way to weight edges, but predications have no analogue to a proximity-based weighting scheme. One option would be to weight edges by the number of times given predications were observed. While this is perhaps the most obvious way to account for important predications, it risks exaggerating high-frequency items that are common for reasons other than importance (perhaps they are idioms, collocates or found in abnormally strong colligational structures). Frequency weighting would also be susceptible to noise from the many low-frequency items. To address these concerns, we introduce a semantically informed measure of predication relevance.

Wilks' (1975; 1978) theory of preference semantics proposes that subject- and object-verb relationships evince "selectional preference", which can be thought of as the disposition verbs have to select certain arguments – particular classes of subjects or objects. To operationalize selectional preference, Resnik (1997) introduced *selectional preference strength* to measure the disposition or "preference" of a verb, v:

$$S_R(v) = \sum_{c \in C} P(c|v) \log \frac{P(c|v)}{P(c)} \qquad (1)$$

where C is a set of semantic classes from which v can select and R is the grammatical relation in question. Note that $S_R(v)$ is effectively the sum K-L divergence between the probabilities of v and c for all classes. In a corpus-based setting, the probability of any word can be estimated by its relative frequency: $P(x) = \frac{f(x)}{\sum_i f(x_i)}$. Resnik goes on to define a measure of *selectional association* between a verb and a specific class, c:

$$A_R(v, c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)} \qquad (2)$$

In the typical form of selectional preference induction – the task of estimating likelihoods over all classes – Eq. 2 is used to measure a verb's preference for classes of nominal subjects or objects like vehicles, insects, birds, etc. (Resnik, 1997; Shutova et al., 2013).

To test our assumptions regarding the modular structure of predications in English, a measure like selectional association should account for predicates' diversity (or uniformity) of attachment. That is, the preference a predicate has to

operate on a wide or narrow range of words. To account for this, we add a term, $U(p)$, to account for the relative number of *unique* words a predicate p has been observed to predicate. Note that this is not the total number of predications involving p, which would produce problematically high values were p to collocate strongly with the words it predicates. Instead, $U(p)$ addresses and normalizes for the diversity with which p is applied. Additionally, instead of using a pre-set collection of semantic classes on which predication is assumed to operate, each predicate is treated as its own class. For a predication consisting of word w predicated by p, predication strength is defined as follows:

$$PS(w,p) = \frac{1}{S_R(p)} \log \frac{P(w|p)}{U(p)P(w)} \qquad (3)$$

PS thus combines three important properties of predications: the relative frequency of a given predication $P(w|p)$, the relative frequency of a word $P(w)$ and the diversity of a word's potential predications $U(p)$. Defined like this, $U(p)$ helps diminish the contribution of predicates that are widely applicable, under the assumption that being widely used, they are in-fact somewhat less significant. Using this measure to weight edges in a predication network should help diminish the contribution of exceptionally frequent predications as well as that from low-frequency predications without excluding them.

An example predication network is shown in Figure 1. In these networks, a network is constructed over a set of documents where nodes are the words in a predication, the direction following the `is-a` link. Thus, example (i) would result in a link from *safety* to *concern* with weight 1. Were another predication involving *concern* to be observed, another edge would be added from that node. Note that circularities are allowed even though this example is acyclic. To assess predication strength as a relevance function, we compare the community structure of weighted and unweighted networks. The example in Figure 1 shows a sample network (top) and the communities extracted from the unweighted and weighted versions of the same network (middle and bottom). Note the changes in community assignments from the unweighted version to the weighted. In particular, observe the new clusters in the weighted network around *money*, *factor* and *murder*. If our
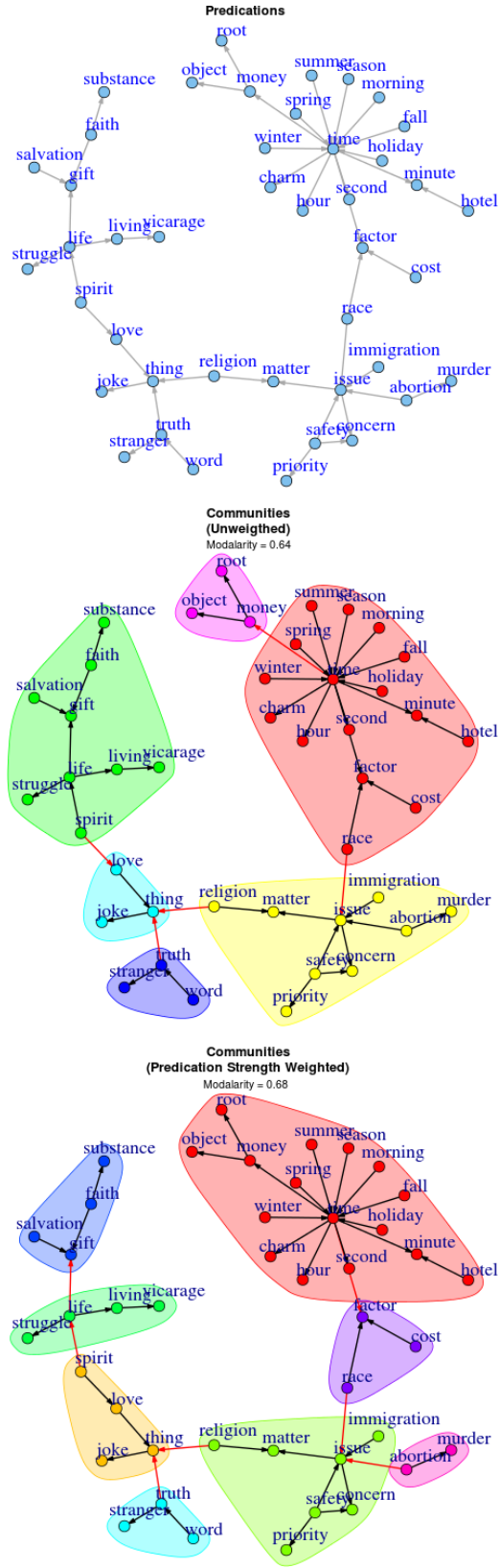

Figure 1: Predication network from the enTen-Ten corpus (pruned by frequency $\geq$ 170): the initial network (top), communities assigned by the Infomap algorithm for the unweighted network (middle) and for the network weighted by predication strength (bottom).

intuitions about the systematic nature of linguistic predication is correct, there should be at least a moderate degree of community structure in the unweighted networks, and if predication displays semantic preference similar to selectional association, this community structure should be stronger for networks weighted by predication strength.

The school *librarian* may be the **person** that controls [...]
You may find *Rachel* is the one **person** who may [...]
Neither the state nor its *government* is a **person**.
An *arbitrator* is a **person** who is appointed [...]
On the other hand, an *expert* is a **person** to fix [...]
After all, the *vendor* is the **person** best able to [...]
An *expert* need not be an individual **person**.
The innocent *party* is a natural **person**.
If the *indemnifier* is a natural **person**, [...]
consumers who are natural **persons** under the Directive.

Table 1: Sample predications involving forms of the word *person* as the target in the BNC. In each case an edge would connect the predicate (in italics) to *person* (in bold).

3 Results

To explore the structure of predication networks, we analyzed two corpora using the method described above: the British National Corpus (BNC) (Leech et al., 2001) and the enTenTen web corpus. Predications were extracted templates over a POS-tagged version of each corpus using the Sketch Engine[1] tool (Kilgarriff et al., 2004). The BNC contained about 112 million tokens and the enTenTen collection has 3.2 billion tokens. For each collection, the top 1,000 most frequent nouns provided a seed set from which to extract all `predicate` and `predicate_of` relations[2] (see examples in Table 1). For the BNC, this resulted in 40,721 predications (14,319 unique) and 260,555 (20,651 unique) for the enTenTen collection. Predication strength scores were computed for every predication using within-corpus relative frequencies. These scores were used to weight edges in one version of the predication network, whereas the edge-weights of the "unweighted" version were uniformly set to 1.0. No node-weighting was applied in either case.

___

[1] http://www.sketchengine.co.uk/
[2] `"NN.?.?" [tag="WP"|tag="PNQ"|tag="CJT"]`
`?[tag="RB.?"|tag="RB"|tag="VM"]0,5`
`[lemma="be" & tag="V.*"] "RB.?"0,2`
`[tag="DT.?"|tag="PP$"]0,1 "CD"0,2`
`[tag="JJ.?"|tag="RB.?"|word=","]0,3`
`"NN.?.?"0,2 2:"NN.?.?" [tag!="NN.?.?"]`

Two methods were used to extract communities from the predication networks: the Infomap and walktrap algorithms. By using two methods, we attain some assurance that our findings are not artifacts of the assumptions underlying either algorithm. The Infomap algorithm is an information-theoretic method that exploits the analogue between optimizing a compression dictionary and simplifying a graph by describing "flow" through nodes (Rosvall and Bergstrom, 2008). Infomap assumes edges in a network induce such flow and by deriving a minimum description of this flow, the algorithm can find multi-level communities in large networks (Rosvall and Bergstrom, 2011). The second method, walktrap, operationalizes the intuition that a large set of short random walks on a network will leave walkers on some groups of nodes more often than others (Pons and Latapy, 2005). By setting the walk distance to a small value, relative to a network's density, walkers will tend toward communities if the walker sample is sufficient. These algorithms are both known to work well with large, directed networks and neither imposed intractable computational burdens at our scale (Fortunato, 2010; Lancichinetti and Fortunato, 2009). Because both algorithms require a connected network, our analysis is restricted to the largest connected component (LCC) for all networks, though we have no reason to believe results would differ significantly for other components.

Community assignments can be assessed by measuring how self-contained or "modular" the resulting communities are. Modularity was introduced as a way to choose the level of an optimal cut for hierarchical partitioning algorithms, analogous to the level in the dendrogram that yields the best communities (Newman and Girvan, 2004). For a network with adjacency matrix **A** and community assignments c, modularity is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \mathbf{A}_{ij} - \frac{k_i k_j}{2m} \delta(c_i, c_j) \qquad (4)$$

where m is the number of edges and k_i is the degree of node i. $\delta(c_i, c_j)$ is 1 when the community assignment of node i is the same as that for node j. Modularity measures how likely it is that nodes in a community are connected to one another as opposed to nodes in other communities. Modularity is defined from -1.0 to 1.0 and graphs where $Q > 0.6$ are conventionally said to have relatively strong community structure (New-

man, 2010). Here, we use modularity instead of a measure of semantic similarity or semantic coherence because predication is seldom an assertion of equivalence or similarity. This means that although words in predication communities may be related in an ontological sense, such an assessment would not expose the level of independence between the communities.

Weighted and unweighted networks from both corpora were submitted to each community detection algorithm, the results of which were assessed using modularity. We also carried out this analysis on frequency-weighted networks, the results of which were similar to the unweighted configuration, but are not reported for sake of brevity. Figure 2 shows the modularity for each configuration with varying minimum predication frequency (the number of times a predication had to occur to be included). Varying the minimum frequency thresholds helps simulate the effect of corpus-size on the algorithm. In the BNC, unweighted networks with no minimum edge frequency show slight modularity ($Q = 0.30$), whereas in weighted networks it is quite strong ($Q = 0.89$). The enTenTen corpus exhibits a gap between the unweighted ($Q = 0.61$) and weighted networks ($Q = 0.88$) at low edge thresholds. This shows that predication strength is helpful in weighting relevant items without excluding low-frequency observations. The lower modularity scores (Q; Eq. 4) in the unweighted networks may be due to more novel, loose or figurative associations found in low-frequency predications that inappropriately connect unrelated communities. Interestingly, scores for unweighted and weighted networks converge up to a point as the minimum frequency increases (reducing the size of the network). This pruning is helpful for the unweighted networks, but has little effect on the weighted versions. In all cases, sparsity takes a toll as the LCC becomes quite small. The reason for the eventual decline as the LCC shrinks below 70 nodes is because communities are less likely to form at all in small networks.

In addition to the highly modular structure, the communities of predications themselves are likely to represent some semantic organization. Specifically, we looked for a categorical structure within the communities by comparing words to the hypernym tree in WordNet (Miller, 1995). Intuitively, one would expect words that are central in
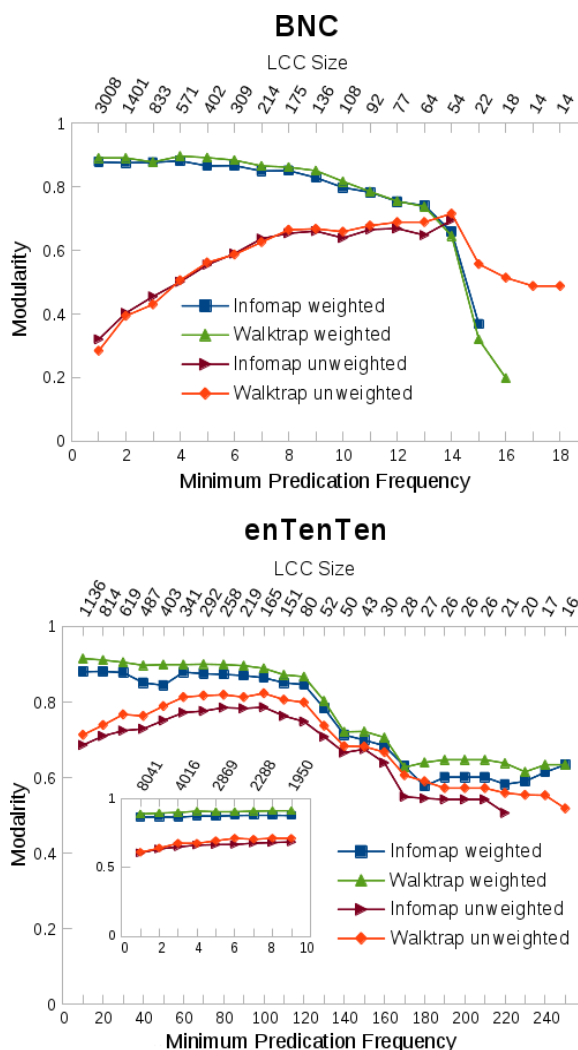

Figure 2: Modularity of predication networks in the BNC (top) and enTenTen (bottom). Note, as the minimum frequency increases (bottom axis) and the LCC contains fewer and fewer nodes (top axis), the community detection algorithms may not produce a solution with more than one community, resulting in undefined modularity.

a community to be members of higher-level categories. In figure 1, for example, *summer*, *hour* and *holiday* all point to *time*, one could infer that *time* is a shared hypernym. We use closeness centrality, a graph-theoretic measure of node's average proximity to other nodes, as a within-community measure of super-ordinance (i.e. hypernymy). Though there a number of network centrality measures, closeness centrality is a robust measure, though it tends not to scale well to larger networks because it requires computing the distance between every pair of nodes (Friedl et al., 2010).

The centrality scores in the communities were compared to WordNet using the first sense-entry for each node (which is typically the most common) and words not found in the tree were discarded. For the unweighted networks across both corpora, we found a mean Spearman correlation of $r=0.35$ ($p < 0.01$; using Fisher's transformation) for the Infomap algorithm and $r=0.38$ ($p < 0.01$) for walktrap. In the weighted versions, Infomap produced $r=0.41$ ($p < 0.01$) and walktrap produced $r=0.44$ ($p < 0.01$). This confirms that predication communities tend to specify categorical knowledge is moderately similar to WordNet. Note these correlation values are comparable between the weighted and unweighted networks, implying that relevance, as selectional association, is not an important marker of the communities' hypernymic composition.

4 Discussion

The analysis in this paper is an attempt to identify whether or not ontological knowledge expressed in text consists of meaningful clusters. With the network representation and our measure of predication strength, results indicate that predication forms strong community structures. Overall, results point to the highly modular nature of predication, previously unreported in language. This confirms our prediction that predication comprises systematic clusters of related things and the higher modularity observed in networks weighted by predication strength implies that predication exhibits a form of selectional preference. Predication's strong community structure is important because it supports the use of linguistic patterns in establishing ontological representations, which naturally form higher-level groups.

Technically, our measure of predication strength, which is built on prior assessments of selectional preference, identifies the modular semantic structure of predication even when low frequency predications are included. This may be because low-frequency predications are more likely to inscribe novel, loose or figurative associations that reach between semantic clusters to inappropriately decrease the overall modularity if not down-weighted. As a result, more systematic comparison of weighted and unweighted networks, and the relative location of predication within these structures, will reveal where semantic innovation and figurative assertions are

most likely to occur. The predication networks analyzed rely on a relatively tight definition of predication, one that, in other languages, may not be accessible by the copular form. Additionally, the two literal interpretations of linguistic predications, equivalence or category membership, may also not be common in all languages. To the extent that parsers or taggers are available, a comparative analysis would broaden the understanding of predication in general.

Given their high modularity, predication structures could be exploited further for a number of NLP tasks. The correlations between centrality and hypernym depth mean that predication networks could help construct or update categorical taxonomies. For example, these networks could help automate the construction of a hypernym taxonomy with weighted branches, potentially augmenting resources like WordNet (Ruiz-Casado et al., 2005; Miller, 1995). One could also examine the growth, combination and bifurcation of specific communities to help track ontological commitments, either over time as shifts in language structures (Gerow and Ahmad, 2012), or across genre and domain (Davies, 2010). Further, because predication encodes categorical information, its community structure may also encode higher-level relations where strong inter-community links imply relationships between classes of objects.

Our study examined the topographical structure of English predications in general, structure that consists, in large part, of hypernym relations. Though the relations in the examined networks are defined by copular `is-a` predication structure, within-community hierarchies correlated only moderately with the hypernym hierarchy in WordNet. This implies that the predications comprising our networks are either not entirely hypernymic or that WordNet is not a good baseline. Indeed, predication is a grammatical relationship that often asserts synonymy or figurative hypernymy (perhaps sometimes also metonymy) and it is not apparent from the surface structure how these semantic interpretation could be disambiguated. One reason this correlation is not higher is likely to do with the low coverage of the copular form as evidence of hypernymy (Hearst, 1992).

Further work regarding the structure of predications could build on the network framework to evaluate the communities themselves. What prop-

erties differentiate communities? Are there semantic, lexical or statistical properties that contribute to the formation of communities? Are there discernible differences between words that typify communities as opposed to those that bridge communities? Predication communities are primarily semantic in nature, implying that central nodes would typify meaningful aspects of their community. It would also be relatively easy to extend network representations to address more qualitative aspects such as coherence, word norms and word associations. Indeed, a variety of corpus-based research could employ network-based methods like those exemplified in this paper, capitalizing on graph-theory, social network analysis and statistical physics, without departing from relational structures inherent to language.

Acknowledgments

References

Eneko Agirre, Oier Lpez de Lacalle, and Aitor Soroa. 2014. Random walks for knowledge-based word sense disambiguation. *Computational Linguistics*, 40(1):57–84.

Chris Biemann, Monojit Choudhury, and Animesh Mukherjee. 2009. Syntax is from mars while semantics from venus!: insights from spectral analysis of distributional similarity networks. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 245–248.

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*, pages 73–80.

Brian F. Bowdle and Dedre Gentner. 2005. The career of metaphor. *Psychological Review*, 112(1):193.

Sharon A. Caraballo. 1999. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 120–126.

Philipp Cimiano and Johanna Völker. 2005. Text2onto. In *Natural language processing and information systems*, pages 227–238. Springer.

Philipp Cimiano. 2006. *Ontology learning from text*. Springer.

Aaron M. Cohen, William R. Hersh, Christopher Dubay, and K. Spackman. 2005. Using co-occurrence network structure to extract synonymous gene and protein names from medline abstracts. *BMC Bioinformatics*, 6(1):103.

Mark Davies. 2010. The corpus of contemporary american english as the first reliable monitor corpus of english. *Literary and linguistic computing*, 25(4):447–464.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22(1):457–479.

Ramon Ferrer i Cancho, Ricard V Solé, and Reinhard Köhler. 2004. Patterns in syntactic dependency networks. *Physical Review E*, 69(5):051915.

Ramon Ferrer i Cancho, Andrea Capocci, and Guido Caldarelli. 2007. Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation and Chaos*, 17(07):2453–2463.

Peter W. Foltz, Darrell Laham, and Thomas K. Landauer. 1999. Automated essay scoring: Applications to educational technology. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 939–944.

Santo Fortunato. 2010. Community detection in graphs. *Physics Reports*, 486(3):75–174.

Dipl-Math Bettina Friedl, Julia Heidemann, et al. 2010. A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, 2(6):371–385.

Aaron Gerow and Khurshid Ahmad. 2012. Diachronic variation in grammatical relationships. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*.

Aaron Gerow. 2014. Extracting clusters of specialist terms from unstructured text. In *Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing (forthcoming)*.

Sam Glucksberg, Matthew S. McGlone, and Deanna Manfredi. 1997. Property attribution in metaphor comprehension. *Journal of memory and language*, 36(1):50–67.

Maria Grineva, Maxim Grinev, and Dmitry Lizorkin. 2009. Extracting key terms from noisy and multi-theme documents. In *Proceedings of the 18th international conference on World wide web*, pages 661–670.

Marti A Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on computational linguistics-Volume 2*, pages 539–545.

Adam Kilgarriff, Pavel Rychl, Pavel Smr, and David Tugwell. 2004. The sketch engine. In *Proceedings of EURALEX 2004*, pages 105–116.

Andrea Lancichinetti and Santo Fortunato. 2009. Community detection algorithms: a comparative analysis. *Physical Review E*, 80(5):056117.

Thomas K. Landauer and Susan T. Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211.

Geoffrey Leech, Paul Rayson, and Andrew Wilson. 2001. *Word frequencies in written and spoken English: based on the British National Corpus*. Longman.

George A. Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Mark E. J. Newman and M. Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113.

Mark E. J. Newman. 2010. *Networks: an introduction*. Oxford University Press.

Pascal Pons and Matthieu Latapy. 2005. Computing communities in large networks using random walks. In *Computer and Information Sciences-ISCIS 2005*, pages 284–293. Springer.

Anatol Rapoport, Amnon Rapoport, William P Livant, and John Boyd. 1966. A study of lexical graphs. *Foundations of Language*, pages 338–376.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How*, pages 52–57.

Martin Rosvall and Carl T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123.

Martin Rosvall and Carl T. Bergstrom. 2011. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PloS one*, 6(4):e18209.

Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. 2005. Automatic extraction of semantic relationships for wordnet by means of pattern learning from wikipedia. In *Natural Language Processing and Information Systems*, pages 67–79. Springer.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.

Mark Steyvers and Joshua B. Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78.

Lin Sun and Anna Korhonen. 2009. Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 638–647.

Peter D. Turney, Patrick Pantel, et al. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1):141–188.

Jean Véronis. 2004. Hyperlex: lexical cartography for information retrieval. *Computer Speech & Language*, 18(3):223–252.

Yorick Wilks. 1975. A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence*, 6(1):53–74.

Yorick Wilks. 1978. Making preferences more active. *Artificial Intelligence*, 11(3):197–223.