Navigating word association norms to extract semantic information

Javier Borge-Holthoefer (javier.borge@urv.cat)
Department of Computer Science and Mathematics, Avda. Països Catalans,
Tarragona, 43007 Catalonia, Spain

Alex Arenas (alexandre.arenas@urv.cat)
Department of Computer Science and Mathematics, Avda. Països Catalans,
Tarragona, 43007 Catalonia, Spain

Abstract

We present a simple model that allows the extraction of semantic similarity relations from free association information. In our study, we use two acclaimed databases of linguistic relationships between pairs of words, feature-based and association-based. We apply a complex networks methodology to disentangle feature based relationships on top of a free association network. As a consequence, we broaden complex networks' applications in the field of psycholinguistics, from a merely descriptive to a predictive level. Results are systematically compared to those of two powerful well-known computational models (LSA and WAS).

Keywords: complex networks; random walks; Latent Semantic Analysis; Word Association Space; free association norms; feature production norms

Introduction

The problem of semantic representation has been one of the focus of attention of the cognitive psychology community over the past decades. Although computational approaches to such problem were proposed as soon as the late '60s (Collins & Quillian, 1969; Collins & Loftus, 1975), only in the past decade technology has made it feasible to deal with huge amounts of empirical data, where models could be tested with reliability. It is in this scenario where the most powerful models have emerged (Lund & Burgess, 1996; Landauer & Dumais, 1997; Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2004; Steyvers, Shiffrin, & Nelson, 2004; Griffiths, Steyvers, & Tenenbaum, 2007). Along with these, recent studies have used also the perspective offered by the theory of complex networks to gain insight on language-related problems (Sigman & Cecchi, 2002; Steyvers & Tenenbaum, 2005). The main idea behind these works is to map empirical data onto a graph (usually called complex network) that summarizes the observed relations between words in a given experiment. Once the structure is set up, it is possible to statistically characterize it (with a wide range of existing descriptors) and elucidate properties that can help to better understand the large-scale structure of semantic relations in the specific set.

However, while the network approach has been merely descriptive up to now, computational models like LSA (Landauer & Dumais, 1997), HAL (Lund & Burgess, 1996), WAS (Steyvers et al., 2004) or the more recent Topic Model (Griffiths & Steyvers, 2004) have an intrinsic predictive capability. In particular, some of these models are used to reveal interaction between episodic and semantic memory, considering empirical data that reflects the impact of environmental (i.e. nonlinguistic) experience upon linguistic phenomena (Steyvers et al., 2004; Andrews, Vigliocco, & Vinson, 2005; Silberman, Bentin, & Miikkulainen, 2007).

The description of semantic knowledge as a complex network of interactions between words, does not suffice to get a clear picture of the specific relations between complex networks representing different semantic empirical data sets. One of the main reasons for this is that while the notion of node is quite uncontroversial (in our case a word), the concept of edge is not so because it must be committed to a definition of relationship. In what semantics is concerned, we can consider that a word is related to another one if they belong to the same class (category-related, such as *car* and *wagon*); or if they tend to co-occur in many contexts (*car* and *road*); or if they have a cause-effect relationship (*fire* and *smoke*), and so on. For some of these types of relationship there exist empirical data that quantify how strong two words are related. (Notice that two words may have several of these relationships).

It is clear that different semantic networks will arise depending on the type of association used to link words by the subjects of a cognitive experiment. Moreover, given the intricate complexity of human mind, the more free the association scenario, the more rich the types of relationship will appear. These different association scenarios can reflect semantic or episodic memory contents, depending on the experiment. One of the main challenges is to understand the interaction between both memory representations. In Steyvers et al. (2004) the authors propose the prediction of semantic similarity effects in episodic memory using empirical data. The procedure applied is a modification of the general LSA scheme, using singular value decomposition and multidimensional scaling over a specific data set (Nelson, McEvoy, & Schreiber, 1998). The results show the emergence of feature association groups in a multidimensional space known as Word Association Space (WAS).

We will consider the same problem from a complex network perspective adding a different interpretation of the disentanglement process with plausible cognitive implications. In our work, this prediction is reformulated in the following terms: whether is possible to disentangle similarity relationships from general association words networks by the navigation of the semantic network. We address this question assuming that: (1) Each available data set is a partial exposure to semantic knowledge; (2) Some data sets are more general than others, they grasp the heterogeneity of the se-

mantic knowledge more precisely; and (3) as a consequence of (2), some information from a less general data set might be partially implicit in a more general one. We will construct upon these hypothesis and propose an algorithm that allows the disentanglement of a type of relationship embedded on the structure of a more general association network. In particular, we will focus on two well-known data sets in English: the free-association database constructed by Nelson et al. (1998), and the semantic feature production norms by McRae, Cree, Seidenberg, and McNorgan (2005). Interestingly, the algorithm takes advantage of both a probabilistic and a semantic space approach.

Feature Production Norms description

Feature Production Norms (FP from now on) were produced by McRae et al. by asking subjects to conceptually recognize features when confronted with a certain word. This feature collection is used to build up a vector of characteristics for each word, where each dimension (vector component) represents a feature, with a value that represents its production frequency across participants. These norms include 541 concepts, for which semantic closeness or *similarity* is computed as the cosine (overlap) between pairs of vectors of characteristics. The cosine is obtained as follows:

$$\cos\theta = \frac{v_1 w_1 + v_2 w_2 + \ldots + v_n w_n}{\| v \| \| w \|} \quad (1)$$

that is, the dot product between two concept vectors *v* and *w*, divided by the product of their lengths.

As a consequence, words like *banjo* and *accordion* are very similar (i.e. they have a projection close to 1) because their vector representations show a high overlap, essentially provoked by their shared features as musical instruments, while the vectors for *banjo* and *spider* are very different, showing an overlap close to 0 (almost orthogonal vectors).

We will represent this information under the form of a network, where each node represents a word, and an edge (or link) is set up between a pair of nodes whenever their vectors projection is different from 0. The meaning of an edge in this network is thus the features similarity between two words. The network is undirected (since similarity is symmetric) and weighted by the value of the projections.

Free-Association Norms description

Nelson et al. produced these norms by asking over 6000 participants to write down the first word (*target*) that came to their mind when confronted with a *cue* (word presented to the subject). The experiment was performed using more than 5000 cues. Among other information, a frequency of coincidence between subjects for each pair of words is obtained. As an example, words *mice* and *cheese* are neighbors in this database, because a large fraction of the subjects related this target to this cue. Note, however, that the association of these two words is not directly represented by similar features but other relationships (in this case mice eat cheese). The network empirically obtained is directed and weighted, where

weights represent the frequency of association in the sample. Although free association data is often transformed into symmetric information (as in Steyvers et al., 2004), FA has been treated here in its original form.

Generally speaking, Free-Association Norms (FA from now on) represent a more complex scenario than FP when considering the semantics of edges. FA is heterogeneous by construction, it may grasp any relation between words e.g. a causal-temporal relation (*fire* and *smoke*), an instrumental relation (*broom* and *floor*) or a conceptual relation (*bus* and *train*), among others. This heterogeneity will be on the basis of our approach because we assume that some similarity information is implicit in FA.

The Random Inheritance Model (RIM)

Our specific goal is to propose a computational model to extract a FP-like network from the track of a dynamical process upon FA. *Dynamics* is in this framework related to the navigation of the network (diffusion process), whereas temporal dynamics (growth and change) of the network is not considered in this work (see Steyvers & Tenenbaum, 2005 for a complex network approach to this problem).

The idea is to simulate a naïve cognitive navigation on top of a general association semantic network to relate words with a certain similarity, in particular we want to recover feature similarities. We schematize this process as uncorrelated *random walks* from node to node that propagate an inheritance mechanism among words, converging to a feature vectors network. Our intuition about the expected success of our approach relies on two facts: the modular structure of the FA network surely retains significant meta-similitude relationships, and random walks are a the most simple dynamical processes capable of revealing the local neighborhoods of nodes when they persistently get trapped into modules. The inheritance mechanism is a simple reinforcement of similarities within these groups. We call this algorithm the Random Inheritance Model (RIM).

The results obtained show macro-statistical coincidences (functional form of the distributions and descriptors) between the real and the synthetic FP network, moreover, the model yields also significant success at the microscopic level, i.e. is able to reproduce to a large extent FP empirical relationships. These results support the general hypothesis about implicit entangled information in FA, and also reveals a possible mechanism of navigation to recover feature information in semantic networks. Finally, we compare these results with those obtained using LSA and WAS on the same data sets.

FA and FP norms can be represented as semantic networks of words, which in turn can be analyzed in terms of descriptors (see last section). Both empirical networks are topologically different, that is, the statistical local and global properties differ significantly from each other. The main differences are concerned to the sparsity of FA, in contrast to the strong density of FP. Since our goal is to compare a synthetic network obtained from FA, to FP up to a microscopic level, we need both networks to have the same nodes (words). To this

end, we have extracted from the databases those words that appear in both, which has left two subnetworks of 376 nodes each (polysemous words, such as *bat*, were left aside). The statistical characteristics of the extracted subnetworks do not differ very much from their complete versions but they do between them, see Table 1. From now on, the subnetworks of 376 words common in FA and FP will be used for comparison purposes.

Table 1
Main statistical descriptors of the networks FA and FP data, and their respective common words' subnetworks (FA-s and FP-s). N is the number of nodes; ⟨s⟩ is the average strength; L is the average shortest path length; D is the diameter of the network; C is clustering coefficient and r is the assortativity coefficient (see last section).

Descriptor	FA (whole)	FP (whole)	FA-s	FP-s
N	5018	541	376	376
$\langle s \rangle$	0.77	20.20	0.26	13.43
L	3.04	1.68	4.41	1.68
D	5	5	9	3
C	0.1862	0.6344	0.1926	0.6253
r	0.097	0.2609	0.3258	0.2951

Finally, it is worthwhile to mention the fact that the databases, although they both belong to the psycholinguistic field, they were created in different places and years (affecting the use of language); a different number of subjects were used to build up the norms (affecting the robustness of data), etc. Even the intention (i.e. the type of problem they seek to tackle) of the collections is different. It is important to realize about all these facts in order to understand the amount of uncertainty any model faces when trying to reproduce a particular empirical dataset.

Keeping in mind all these general considerations we can move on to specify how our model works. In what follows, we first specify the logic behind our proposal and, after, we describe the mathematical framework that unifies the different steps. The main logic stages in RIM are:

1. Initialization:
First, every word in the FA network is tagged with an initial vector of characteristics. To avoid initial bias, we choose the vectors to be orthogonal in the canonical basis. That means that every word has associated a vector of N-dimensions, being N the number of words in the network, with a component at 1 and the rest at zero.

2. Navigation and Inheritance:
Then a random walk of S steps[1] starting at a node i is performed. At every step of the walk, we propose an inheritance mechanism that changes v_i (the initial vector of the word i) depending on the visited nodes. Let $s = s_1, s_2, ..., s_n$ the set of visited nodes. Then the new vector for node i is computed as:

$$v_i := v_i + \sum_{j=1}^{n} v_{s_j} \qquad (2)$$

The process is performed in parallel, i.e. the update of the feature vectors is done after completion of the inheritance for every word. At the end, we have a synthetic vector of features for every word in the network.

3. Averaging:
Once the feature vectors have been computed, we build up a synthetic feature similarity network. The network is the result of projecting all pairs of vectors and prescribing a weighted link between two words according to this projection. The whole process is iterated (by simulating several runs) up to convergence of the average of the synthetic feature similarity networks generated at each run. The average, after convergence, is the synthetic feature similarity network we compare with FP.

Let us define the transition probability of the FA network. The elements of FA (a_{ij}) correspond to frequency of first association reported by the participants of the experiments. However, note that the 5018 words that appear on the data set are not all the words that appeared at the experiment, but only those that where at the same time cues in the experiment. That means that the data have to be normalized before having a transition probability matrix. We define the transition probability matrix P as:

$$P_{ij} = \frac{a_{ij}}{\sum_j a_{ij}} \qquad (3)$$

Note that this matrix is asymmetric, as well as the original matrix FA. We maintain this asymmetry property in our approach to preserve the meaning of the empirical data. Once the matrix P is constructed, the random walkers of different lengths are simply represented by powers of P. For example, if we perform random walks of length 2, after averaging over many realizations we will converge to the transition matrix P^2, every element $(P^2)_{ij}$ represents the probability of reaching j, from i, in 2 steps, and the same applies to other length values. The inheritance process proposed, corresponds, in this scenario, to a change of basis, from the canonical basis of the N-dimensional space, to the new basis in the space of transitions T:

$$T = \lim_{S \to \infty} \sum_{i=1}^{S} P^i \qquad (4)$$

The convergence of Eq.(4) is guaranteed by the Perron-Frobenius theorem. In practice, the summation in Eq.(4) converges very fast, limiting the dependence on indirect associative strengths (Nelson & Zhang, 2000). We tested the behavior up to S=10, although with S=4 we already achieve convergence in T up to 10^{-4} in terms of the Hamming distance. All the results for RIM will be expressed for $S = 4$ from now on. Finally, the matrix that will represent in our model the feature similarity network (synthetic FP), where similarity is calculated as the cosine of the vectors in the new space, is

---

[1] A random walk is a time-reversible finite Markov chain, see (Lovàsz, 1996) for a survey on the topic.

given by the scalar product of the matrix and its transpose, $FS = TT^\dagger$.

Results of RIM

In this section we present the performance of the RIM in disentangling a Feature Production Norm from the empirical FA. To compare the results with the empirical FP we define a set of measures that can be classified in macroscopic and microscopic similarities. To evaluate macroscopic similarities we will use basic descriptors of complex networks (see Appendix), the strength distribution $P(s)$, along with average (global) quantities already computed in Table 1. To evaluate the microscopic similarities we will compute the success rate on the local structure of the neighborhood of words in both real and synthetic networks. We will also compare the results of our model with those obtained using the well known Latent Semantic Analysis (LSA; Landauer & Dumais, 1997; Landauer, McNamara, Dennis, & Kintsch, 2007) and Word Association Space (WAS; Steyvers et al., 2004). Although LSA's applicability goes beyond the scope of this work, it stands as an appropriate benchmark model to compare the performance of our proposal. In particular, we have used LSA vector representation based on the corpus TASA for the subset of common words in FA and FP, with a space dimensionality of $d = 300$. This LSA TASA-based representation is suitable for comparison because it has been assessed as a simulation of human vocabulary test synonym judgments (Landauer, Foltz, & Laham, 1998). WAS model is specially pertinent for the current comparison because: on one hand, the model is formally similar to LSA; on the other, it makes use of mediated strength between non-direct associates as in RIM, and has been reported on the same data set we use in order to extract semantic information. Accordingly, we have performed the procedure described in the cited article upon the whole network, and after extracted the mentioned 376 subset of words. We only compare to the best results of WAS for this data set (which correspond to Singular Value Decomposition under $S^{(2)}$, $N = 5018$, and $d = 400$, see Steyvers et al., 2004 for details).
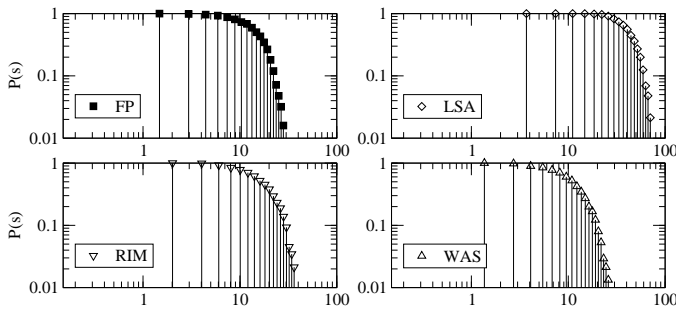

Figure 1. Log-log plot of the cumulative strength distribution of the networks: Feature Production norms FP (empirical target), and the synthetic networks obtained using Latent Semantic Analysis LSA, Word Association Space WAS and Random Inheritance Model RIM.

Table 2

Statistical parameters for Free Association norms FA (substrate of the dynamic process), Feature Production norms FP (empirical target), and the synthetic networks obtained using Latent Semantic Analysis LSA, Word Association Space WAS and Random Inheritance Model RIM.

Descriptor	FA	FP	LSA	WAS	RIM
N	376	376	376	376	376
$\langle s \rangle$	0.26	13.43	39.60	10.29	15.62
L	4.41	1.68	0.02	2.00	1.77
D	9	3	2	4	3
C	0.192	0.625	0.961	0.492	0.584
r	0.325	0.295	0.125	0.303	0.305

Macroscopic similarities

First we plot the cumulative strength distribution of the empirical network FP, and the respective synthetic networks provided by LSA or by applying WAS and RIM to FA, see Figure 1. The statistical agreement between FP and RIM and WAS is remarkable. The general observation is that all distributions present an exponential decay instead of a power-law decay. This specific form of the distributions is characteristic of random homogeneous networks. In Table 2 we present the main descriptors of the four previous networks, plus the substrate of the dynamic process FA, for comparison purposes. Again, the agreement between the empirical FP and RIM is marked, RIM reproduces with significant accuracy the average strength, the average path length, diameter, clustering and assortativity, of the FP target network. WAS also succeeds largely on the determination of macroscopic properties of the network, while LSA can not be so similar.

Microscopic similarities

The necessity for a detailed comparison between synthetic and empirical sets is double: first, the statistical characterization presented before is informative and important, but not definitive to state the validity of the model to disentangle actual information in the original FA network. At most, it is capable of a correct description of the empirical target's structure. And second, the difference between our particular network, and general examples used in complex networks theory is that nodes are tagged and then not interchangeable. The specific neighborhood of every word matters, because it reveals semantic relations, and then the degree or the clustering become less relevant than the specific list of neighbors synthetically obtained. Therefore a model dealing with tagged elements can only be predictive if it succeeds at this level of comparison.

To this end, we pursue the evaluation of RIM in predicting the specific words of each node's neighborhood.

The first question concerning the models at stake (WAS and RIM, since they are here built upon FA) is whether they are useful at all: that is, whether they are capable of uncovering semantic relations which were not in FA already (i.e.
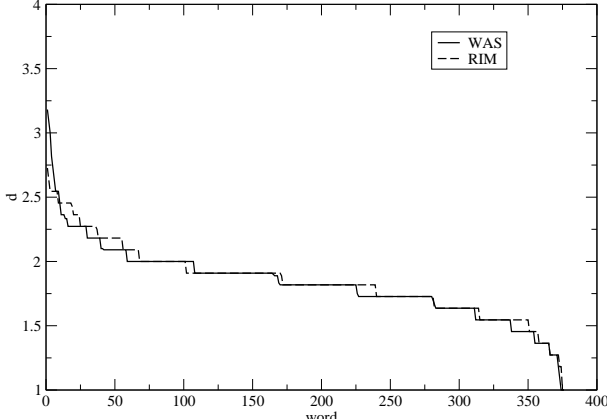
Figure 2. Inheritance distances for $l = 15$ for WAS and RIM, evidencing the fact that both models are capable of capturing semantic relationships beyond a word's immediate neighborhood.

Figure 3. For each synthetic network (LSA, WAS and RIM) we have measured the mean error (for $l = 1$ to $l = 15$) against FP, according to Eq. 5.

as first neighbors). Figure 2 points at this question by evidencing that both WAS and RIM are actually collecting semantic relationships further than $d = 1$. The bulk of words is inheriting features from approximately $d = 2$ on average, which means that the models are in fact capturing semantic similarities from a significant distance (it is important to keep in mind that FA has a diameter of 5, see Table 1).

A second problem we face on this microscopic evaluation of the model is that of proposing pertinent measures. We proceed as follows, given a specific word i, we start sorting its neighbors according to their linking strength. We apply this for each word in our data sets forming lists. The reference list, is the list of each word in FP, and the lists we want to compare with, are those obtained for each word in the synthetic data sets, RIM, WAS and LSA. We restrict our analysis up to the first 15 ordered neighbors, assuming that these are the most significant ones. We have designed an expression that assigns an error score between a list and its reference, depending on the number of mismatches between both lists, and also on the number of misplacements in them. A mismatch (M) corresponds to a word that exist in the reference list and not in the synthetic list and vice versa, these are considered the main errors in our approach. Since the model seeks powerful predictive capacity, misplacements are also taken into account. A misplacement (O) is an error in the order of appearance of both words in each list. The error score E is then defined as:

$$E = E_M + \frac{E_O}{l - E_M} \quad (5)$$

where E_M stands for the number of mismatches, E_O the number of displacements and l the length of the list.

This quantity recalls well-known *edit distances* such as Levenshtein Distance (Levenshtein, 1966) or its generalization, Damerau-Levenshtein Distance (Damerau, 1964), where the similarity between two strings depends on the amount of insertions/deletions (mismatches in our case) and transpositions (movements or misplacements) that one has to perform on a string in order to completely match the other
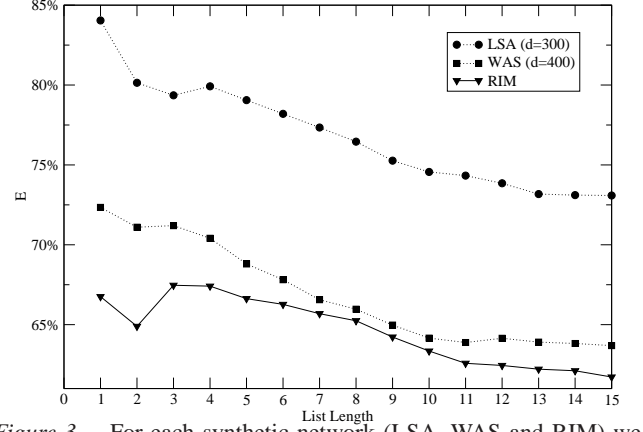
one. Notice that measure E is strongly increased when a mismatch appears, whereas misplacements are less punished. In particular, $E = 0$ when the vectors under comparison are identical. In the other extreme, if a vector has only one match with the other one, and the matching element is not placed correctly, then $E = l$, where l is the length of the involved vectors. Beyond this, there exists only a worse situation, i.e. complete mismatch between vectors. In this case $E = \infty$. Since we intend to compute an average error score, we can not allow an ∞ value, and therefore we prescribe $E = l + 1$, expressing the fact that such score is worse than any case where any match occurs.

The error defined in formula 5 is plotted in Figure 3, on average the error of RIM is about 10% lower than the error of LSA, and 4% lower than that of WAS.

Conclusions

In this work, we have proposed an algorithm to extract feature similarity information from an empirical words' Free Association network. Building upon the idea that free association entangles, in particular, semantic traits of association based on similar characteristics between concepts, we have proposed a simple algorithm to disentangle this information. The results reproduce to a large extent the findings in an empirical Feature Production norms network. The simple strategy of a random navigation process of the actual FA topology and a reinforcement inheritance mechanism suffice to produce relationships comparable to those experimentally obtained.

The comparison with the powerful LSA and WAS models is indicative of the level of macroscopic and microscopic success of our proposal, notwithstanding the fact that both these models provide useful semantic spaces, from a theoretical and an empirical point of view. Furthermore, beyond the level of success of any of these models, we propose that RIM is an approach that enriches other existing models, in the sense that it introduces a dynamical perspective to the formation of semantic spaces. The random navigation mechanism introduced, far from been an optimal strategy in the search space,

ought to start exploring dynamic approaches to the problem of semantic cognition. Finally, the network representation and the dynamics on it provide an intuitive and useful representation, which can be broadened with more realistic navigation and inheritance strategies.

Assuming that Free Association semantic networks are good exposures of human semantic knowledge, we speculate that some cognitive tasks can rely on a specific navigation of this network, in particular a simple navigation mechanism based on randomness, structure of the network and reinforcement could be enough to reproduce non trivial relationships of feature similarity between concepts represented as words. Moreover, explicit metadata associated to semantic structural patterns seem to play an important role on information recovery, that could be extended to other cognitive tasks. Given the already detected importance of modular structure in the study of semantic representation (see Topic Model, Griffiths et al., 2007) we think that disambiguation is perhaps the next affordable challenge along this line of research.

About Complex Network Theory

Complex networks refers to networks (graphs) whose topological characteristics are *non-trivial* in contrast with simple networks where regularities and symmetries dominate their structure. Complex network are found in the representation of interacting elements, in many fields of science ranging from biology to social sciences (Newman, 2003). Related to cognitive science, the application of complex networks has been specially successful to better understand some aspects of brain dynamics in neuroscience (Sporns, Chialvo, Kaiser, & Hilgetag, 2004) and linguistics (Sigman & Cecchi, 2002; Motter, Moura, Lai, & Dasgupta, 2002; Bales & Johnson, 2005; Mehler, 2007), and there is evidence that a step further to psycholinguistics is also on its way (Steyvers & Tenenbaum, 2005; Solé, Corominas, Valverde, & Steels, 2008).

Complex networks main assets comprise a wide range of measures that help on the quantification of its structural characteristics, either at a *micro* (node), *meso* (group) and *macro* (network) level. For extensive network theory reviews and foundational works, see Newman, 2003; Watts & Strogatz, 1998; Barabási & Albert, 1999; Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006.

References

Andrews, M., Vigliocco, G., & Vinson, D. (2005). The role of attributional and distributional information in semantic representation in bara. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the twenty seventh annual conference of the cognitive science society*.

Bales, M., & Johnson, S. (2005). Graph theoretic modeling of large-scale semantic networks. *Journal of Biomedical Informatics*.

Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Phys. Rep.*, *424*, 175-308.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic memory. *Psychological Review*, *82*, 407-428.

Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240-247.

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *comm. ACM*.

Griffiths, T., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(1), 5228-5235.

Griffiths, T., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211-244.

Landauer, T., & Dumais, S. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Landauer, T., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

Landauer, T., McNamara, D., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklad.*, *10*(8), 707-710.

Lovàsz, L. (1996). Random walks on graphs: A survey. *Bolyai Soc. Math. Stud.*, *2*, 353-397.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203-208.

McRae, K., Cree, G., Seidenberg, M., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547-559.

Mehler, A. (2007). Corpus linguistics. an international handbook. In A. Ludeling & M. Kyto (Eds.), (chap. Large Text Networks as an Object of Corpus Linguistic Studies). Berlin/New York: de Gruyter.

Motter, A. E., Moura, A. P. S. de, Lai, Y., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, *65*.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The university of south florida word association, rhyme, and word fragment norms.* Available from http://www.usf.edu/FreeAssociation/

Nelson, D. L., & Zhang, N. (2000). The ties that bind what is known to the recognition of what is new. *Psychonomic Bulletin and Review*, *7*, 604-617.

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

Sigman, M., & Cecchi, G. (2002). Global organization of the wordnet lexicon. *Proceedings of the National Academy of Sciences*, *99*(3), 1742-1747.

Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: An empirically-based computational model. *Cognitive Science: A Multidisciplinary Journal*, *31*(4), 645-671.

Solé, R., Corominas, B., Valverde, S., & Steels, L. (2008). Language networks: their structure, function and evolution. *Trends in Cognitive Science*.

Sporns, O., Chialvo, D., Kaiser, M., & Hilgetag, C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Science*, *8*, 418-425.

Steyvers, M., Shiffrin, R., & Nelson, D. (2004). Experimental cognitive psychology and its applications. In A. Healy (Ed.), (p. 237-249). Washington, D.C: American Psychological Association.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science: A Multidisciplinary Journal*, *29*(1), 41-78.

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440.