

Review

Semantic Networks: Structure and Dynamics

Javier Borge-Holthoefer * and Alex Arenas

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Avda. Països Catalans 26 43007, Tarragona, Catalonia, Spain; E-Mail: alexandre.arenas@urv.cat

* Author to whom correspondence should be addressed; E-Mail: javier.borge@urv.cat;
Tel.: +34-977-55-9687; Fax: +34-977-55-9710.

Received: 21 February 2010 / Accepted: 1 May 2010 / Published: 14 May 2010

Abstract: During the last ten years several studies have appeared regarding language complexity. Research on this issue began soon after the burst of a new movement of interest and research in the study of *complex networks*, *i.e.*, networks whose structure is irregular, complex and dynamically evolving in time. In the first years, network approach to language mostly focused on a very abstract and general overview of language complexity, and few of them studied how this complexity is actually embodied in humans or how it affects cognition. However research has slowly shifted from the language-oriented towards a more cognitive-oriented point of view. This review first offers a brief summary on the methodological and formal foundations of complex networks, then it attempts a general vision of research activity on language from a complex networks perspective, and specially highlights those efforts with cognitive-inspired aim.

Keywords: complex network; language; universality; cognition

Language is surely a strikingly complex object to study. It is built upon many thousands of items (WordNet database [1], for instance, currently gathers over 200,000 words in English), which in turn are often ambiguous. Also, they hold manifold relationships with each other, sometimes for the sake of structure (syntax), sometimes for the sake of their meaning (semantics). Words, and the relationships they hold, evolve across time, new words appear, some of them die, some simply change. As difficult as it may be to study or model these facts, complexity is even greater when language is placed in a cognitive context: linguistic production and comprehension processes occur proficiently in the mind while many other processes are concurrently interacting. Consider, among them, actions from the

sensorimotor system (vocalization), the perceptual system (listening, reading) or memory (retrieval, recall and recognition). Finally, a last step to complexity is to consider linguistic performance as a result of neural activity. Language, thus, is a complex object efficiently managed in a complex mental context, which in turn is embodied in the most complex known system, the brain.

Linguistics and psycholinguistics devote much efforts to disentangle the details of the aforementioned facts. However, some fundamental questions can not be addressed from this fine-grained perspective: what is the general structure of language? Is such structure common to every language? Can we describe the general trends of the mechanisms that provide for linguistic efficient performance? Is it possible to describe the principles of language growth (from a child to an adult)? Such questions demand a complementary point of view from that of linguistics and psycholinguistics, one that abstracts and simplifies as much as possible the intricate nature of language. This general view makes the minimum assumptions, in the end language is reduced to a set of entities which are related with each other. Following this line, cognitive processes are characterized as phenomena occurring on top of that structure. These processes are conceived as naïve mechanisms.

The basics of this viewpoint fit naturally in complex systems approach. Empirical evidence from experiments with subjects and other lexical resources (*thesauri* [2], *corpus* [3], *etc.*) suggest that language can be suitably represented as a network. This article reviews some of the achievements gained along these lines in the last decade, which take a complex network perspective to tackle linguistic phenomena.

Although the concept of small-world was already well known by sociologists [4,5], it was in 1998 when Watts and Strogatz introduced the model of “small world” network [6], which eventually became the seed for the modern theory of complex networks. Soon it turned out that the nature of many interaction patterns observed both in natural and artificial scenarios (for instance, the World-Wide-Web, metabolic networks or scientific collaboration networks) was even more complex than the small world model. In the next decade we have witnessed the evolution of the field of complex networks, and language has not been left out of this process: these advances have made it possible to address the previous questions from a statistical physics point of view, characterizing the structure of language, comparing such characterizations for different languages (even for different domains), setting up growth models for them, simulating dynamics on the structures, *etc.*

Research on language include syntax, prosody, semantics, neuroscience, *etc.* Some of them deal with physical observables but are not suitably approached from a statistical physics point of view yet (as far as the authors know). That is the case of prosody, which tries to extract useful linguistic information from the loudness, pitch or frequency of language sounds. Others, like syntax, have been subject of study from a network perspective, for example by dealing with syntactic trees understood as graphs. Although this latter line has received much attention [7–12] (or rather, *because* of it), it probably deserves special attention in a separate work. Thus the natural framework of this overview is semantics at the lexical (word) level and some adjacent phenomena (lexicon formation and change). This means that works devoted to linguistic superstructures (phrases and sentences) are not considered in this manuscript; neither are sub-lexical units (lemmas, phonemes, *etc.*), although there also exists some work on them in the complex systems bibliography [13,14].

The review is organized as follows. It starts with an overview of some notions of graph theory as used subsequently, see Section 1, this section gives the necessary formal background to fully understand the review. Besides the mathematical descriptors of complex networks, the most influential models, from Erdős–Rényi random graphs [15] to the mentioned “small world” [6] and “scale-free network” [16] are also reviewed here. Section 2 introduces the question of data acquisition and network construction, pointing some sources that have been used to build up language networks and what interpretation they should receive.

In the next three sections 3–5 we focus on the central areas of the review: (i) characterization of language: the organization of language is characterized in terms of general network structural principles (Section 3); (ii) cognitive growth and development: we attempt to reveal how structural features reflect general processes of language acquisition (Section 4); and (iii) cognitive processes: a few models that relate human performance in semantic processing tasks with processes operating on complex networks are presented (Section 5).

Finally, the last section rounds off the review by pointing at open questions (with special attention to neuroscience) and future research directions, as well as offering some conclusions.

1. Introduction to complex networks

There exist many excellent reviews and books in the literature about the structure and dynamics of complex networks [17–26]. Here we overview only those minimal requirements of the theory that will be mentioned along the current work.

1.1. Terminology in complex networks

A network is a graph with N nodes and L links. If the network is *directed* links are then named *arcs*, and account for the directionality of the connections. Otherwise the network is *undirected*, and we refer to links or edges indistinctly. Besides direction, the links can also be valued. A *weighted network* associates a label (weight) to every edge in the network. Two vertices i and j are *adjacent*, or neighbors, if they have an edge connecting them. Notice that, in a directed network, i being adjacent to j does not entail j being adjacent to i . Networks with multiple links (multigraphs) are not considered.

A *path* in a network is a sequence of vertices i_1, i_2, \dots, i_n such that from each of its vertices there is an edge to the next vertex in the sequence. The first vertex is called the *start* vertex and the last vertex is called the *end* vertex. The *length* of the path or *distance* between i_1 and i_n is then the number of edges of the path, which is $n - 1$ in unweighted networks. For weighted networks, the length is the addition of each weight in every edge. When i_1 and i_n are identical, their distance is 0. When i_1 and i_n are unreachable from each other, their distance is defined to be infinity (∞).

A *connected network* is an undirected network such that there exists a path between all pairs of vertices. If the network is directed, and there exists a path from each vertex to every other vertex, then it is a *strongly connected network*. A network is considered to be a *complete network* if all vertices are connected to one another by one edge. We denote the complete network on n vertices K_n . A *clique* in a network is a set of pairwise adjacent vertices. Since any subnetwork induced by a clique is a complete

subnetwork, the two terms and their notations are usually used interchangeably. A k -clique is a clique of order k . A *maximal clique* is a clique that is not a subset of any other clique.

1.2. Complex network descriptors

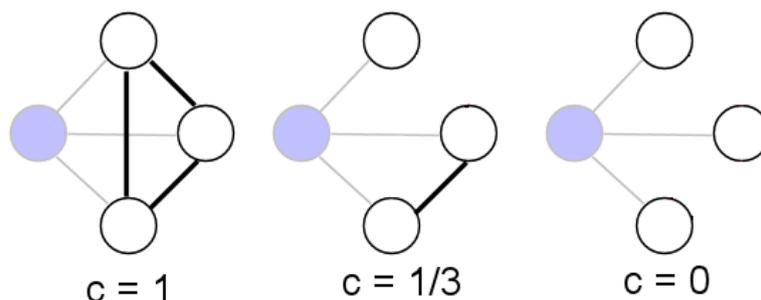
Degree and Degree Distribution The simplest and the most intensively studied one vertex characteristic is degree. Degree, k , of a vertex is the total number of its connections. If we are dealing with a directed graph, in-degree, k_i , is the number of incoming arcs of a vertex. Out-degree, k_o is the number of its outgoing arcs. Degree is actually the number of nearest neighbors of a vertex. Total distributions of vertex degrees of an entire network, $p(k)$, $p_i(k_i)$ (the in-degree distribution), and $p_o(k_o)$ (the out-degree distribution) are its basic statistical characteristics. We define $p(k)$ to be the fraction of vertices in the network that have degree k . Equivalently, $p(k)$ is the probability that a vertex chosen uniformly at random has degree k . Most of the work in network theory deals with cumulative degree distributions, $P(k)$. A plot of $P(k)$ for any given network is built through a cumulative histogram of the degrees of vertices, and this is the type of plot used throughout this article (and often referred to just as “degree distribution”). Although the degree of a vertex is a local quantity, we shall see that a cumulative degree distribution often determines some important global characteristics of networks. Yet another important parameter measured from local data and affecting the global characterization of the network is average degree $\langle k \rangle$. This quantity is measured by the equation:

$$\langle k \rangle = \frac{\sum_{i=1}^N k_i}{N} \quad (1)$$

Strength Distribution In weighted networks the concept of degree of a node i (k_i) is not as important as the notion of strength of that node, $\omega_i = \sum_{j \in \Gamma_i} \omega_{ij}$, *i.e.*, the sum over the nodes j in the of i , of weights from node i towards each of the nodes j in its neighborhood Γ_i . In this type of network it is possible to measure the average strength $\langle k \rangle$ with a slight modification of eq.1. On the other hand, it is also possible to plot the cumulative strength distribution $P(s)$, but it is important to make a good choice in the number of bins of the histogram (this depends on the particular distribution of weights for each network).

Shortest Path and Diameter For each pair of vertices i and j connected by at least one path, one can introduce the shortest path length, the so-called *intervertex distance* d_{ij} , the corresponding number of edges in the shortest path. Then one can define the distribution of the shortest-path lengths between pairs of vertices of a network and the average shortest-path length L of a network. The average here is over all pairs of vertices between which a path exists and over all realizations of a network. It determines the effective “linear size” of a network, the average separation of pairs of vertices. In a fully connected network, $d = 1$. Recall that shortest paths can also be measured in weighted networks, then the path’s cost equals the sum of the weights. One can also introduce the maximal intervertex distance over all the pairs of vertices between which a path exists. This descriptor determines the maximal extent of a network; the maximal shortest path is also referred to as the *diameter* (D) of the network.

Figure 1. An illustration of the concept of clustering C , calculated on the gray node. In the left figure, every neighbor of the mentioned node is connected to each other; therefore, clustering coefficient is 1. In the middle picture, only two of the gray node neighbors' are connected, yielding a clustering coefficient of $1/3$; finally, in the last illustration none of the gray node's neighbors are linked to each other, which yields a clustering coefficient of 0. From Wikipedia Commons.



Clustering Coefficient The presence of connections between the nearest neighbors of a vertex i is described by its clustering coefficient. Suppose that a node (or vertex) i in the network has k_i edges and they connect this node to k_i other nodes. These nodes are all neighbors of node i . Clearly, at most

$$\binom{k}{2} = \frac{k(k-1)}{2} \tag{2}$$

edges can exist among them, and this occurs when every neighbor of node i connected to every other neighbor of node i (number of loops of length 3 attached to vertex i). The clustering coefficient C_i of node i is then defined as the ratio between the number E_i of edges that actually exist among these k_i nodes and the total possible number:

$$C_i = \frac{2E_i}{k(k-1)} \tag{3}$$

Equivalently, the clustering coefficient of a node i can be defined as the proportion of 3-cliques in which i participates. The clustering coefficient C of the whole network is the average of C_i over all i , see Figure 1. Clearly, $C \leq 1$; and $C = 1$ if and only if the network is globally coupled, which means that every node in the network connects to every other node. By definition, *trees* are graphs without loops, *i.e.*, $C = 0$.

The clustering coefficient of the network reflects the transitivity of the mean closest neighborhood of a network vertex, that is, the extent to which the nearest neighbors of a vertex are the nearest neighbors of each other [6]. The notion of clustering was much earlier introduced in sociology [27].

Centrality Measures Centrality measures are some of the most fundamental and frequently used measures of network structure. Centrality measures address the question, “Which is the most important or central node in this network?”, that is, the question whether nodes should all be considered equal in significance or not (whether exists some kind of hierarchy or not in the system). The existence of such hierarchy would then imply that certain vertices in the network are more *central* than others. There are many answers to this question, depending on what we mean by important. In this Section we

briefly explore two centrality indexes (*betweenness* and *eigenvector centrality*) that are widely used in the network literature. Note however that betweenness or eigenvector centrality are not the only method to classify nodes' importance. Within graph theory and network analysis, there are various measures of the centrality of a vertex within a graph that determine the relative importance of a vertex within the graph. For instance, besides betweenness, there are two other main centrality measures that are widely used in network analysis: *degree centrality* and *closeness*. The first, and simplest, is degree centrality, which assumes that the larger is the degree of a node, the more central it is. The closeness centrality of a vertex measures how easily other vertices can be reached from it (or the other way: how easily it can be reached from the other vertices). It is defined as the number of vertices minus one divided by the sum of the lengths of all geodesics from/to the given vertex.

a. Betweenness One of the first significant attempts to solve the question of node centrality is Freeman's proposal (originally posed from a social point of view): *betweenness* as a centrality measure [28]. As Freeman points out, a node in a network is central to the extent that it falls on the shortest path between pairs of other nodes. In his own words, "suppose that in order for node i to contact node j , node k must be used as an intermediate station. Node k in such a context has a certain "responsibility" to nodes i and j . If we count all the minimum paths that pass through node k , then we have a measure of the "stress" which node k must undergo during the activity of the network. A vector giving this number for each node of the network would give us a good idea of stress conditions throughout the system" [28]. Computationally, betweenness is measured according to the next equation:

$$C_B(i) = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \quad (4)$$

with σ_{jk} as the number of shortest paths from j to k , and $\sigma_{jk}(i)$ the number of shortest paths from j to k that pass through vertex i . Note that shortest paths can be measured in a weighted and/or directed network, thus it is possible to calculate this descriptor for any network [29]. Commonly, betweenness is normalized by dividing through by the number of pairs of vertices not including v , which is $(n-1)(n-2)$. By means of normalization it is possible to compare the betweenness of nodes from different networks.

b. Eigenvector centrality A more sophisticated version of the degree centrality is the so-called eigenvector centrality [30]. Where degree centrality gives a simple count of the number of connections a vertex has, eigenvector centrality acknowledges that not all connections are equal. In general, connections to people who are themselves influential will lend a person more influence than connections to less influential people. If we denote the centrality of vertex i by x_i , then we can allow for this effect by making x_i proportional to the average of the centralities of i 's network neighbors:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^N A_{ij} x_j \quad (5)$$

where λ is a constant. Defining the vector of centralities $x = (x_1, x_2, \dots)$, we can rewrite this equation in matrix form as

$$\lambda x = Ax \quad (6)$$

and hence we see that x is an eigenvector of the adjacency matrix with eigenvalue λ . Assuming that we wish the centralities to be non-negative, it can be shown (using the Perron-Frobenius theorem) that λ must be the largest eigenvalue of the adjacency matrix and x the corresponding eigenvector. The eigenvector centrality defined in this way accords each vertex a centrality that depends both on the number and the quality of its connections: having a large number of connections still counts for something, but a vertex with a smaller number of high-quality contacts may outrank one with a larger number of mediocre contacts. In other words, eigenvector centrality assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Eigenvector centrality turns out to be a revealing measure in many situations. For example, a variant of eigenvector centrality is employed by the well-known Web search engine Google to rank Web pages, and works well in that context. Specifically, from an abstract point of view, the World Wide Web forms a directed graph, in which nodes are Web pages and the edges between them are hyperlinks [31]. The goal of an Internet search engine is to retrieve an ordered list of pages that are relevant to a particular query. Typically, this is done by identifying all pages that contain the words that appear in the query, then ordering those pages using a measure of their importance based on their link structure. Although the details of the algorithms used by commercial search engines are proprietary, the basic principles behind the PageRank algorithm (part of Google search engine) are public knowledge [32], and such algorithm relies on the concept of eigenvector centrality. Despite the usefulness of centrality measures, hierarchy detection and node's role determination is not a closed issue. For this reason, other classifying techniques will be explored in subsequent Sections.

Degree-Degree correlation: assortativity It is often interesting to check for correlations between the degrees of different vertices, which have been found to play an important role in many structural and dynamical network properties. The most natural approach is to consider the correlations between two vertices connected by an edge. A way to determine the degree correlation is by considering the Pearson correlation coefficient of the degrees at both ends of the edges [33,34]

$$r = \frac{\frac{1}{N} \sum_{j>i} k_i k_j a_{ij} - [\frac{1}{N} \sum_{j>i} \frac{1}{2}(k_i + k_k) a_{ij}]^2}{\frac{1}{N} \sum_{j>i} \frac{1}{2}(k_i^2 + k_j^2) a_{ij} - [\frac{1}{N} \sum_{j>i} \frac{1}{2}(k_i + k_k) a_{ij}]^2} \quad (7)$$

where N is the total number of edges. If $r > 0$ the network is assortative; if $r < 0$, the network is disassortative; for $r = 0$ there are no correlation between vertex degrees.

Degree correlations can be used to characterize networks and to validate the ability of network models to represent real network topologies. Newman computed the Pearson correlation coefficient for some real and model networks and discovered that, although the models reproduce specific topological features such as the power law degree distribution or the small-world property, most of them (e.g., the Erdős–Rényi and Barabási–Albert models) fail to reproduce the assortative mixing ($r = 0$ for the mentioned models) [33,34]. Further, it was found that the assortativity depends on the type of network. While social networks tend to be assortative, biological and technological networks are often disassortative. The latter property is undesirable for practical purposes, because assortative networks are known to be resilient to simple target attack, at the least.

There exist alternative definitions of degree-degree relations. Whereas correlation functions measure linear relations, information-based approaches measure the general dependence between two variables [35]. Specially interesting is *mutual information* provided by the expression

$$I(q) = H(q) - H_c(q|q') = \sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') \log \frac{q_c(k, k')}{q(k)q(k')} \tag{8}$$

See the work by Solé and Valverde [35] for details.

1.3. Network models

Regular Graphs Although regular graphs do not fall under the definition of complex networks (they are actually quite far from being complex, thus their name), they play an important role in the understanding of the concept of “small world”, see below. For this reason we offer a brief comment on them.

In graph theory, a regular graph is a graph where each vertex has the same number of neighbors, *i.e.*, every vertex has the same degree. A regular graph with vertices of degree k is called a k -regular graph or regular graph of degree k [36].

Random Graphs Before the burst of attention on complex networks in the decade of 1990s, a particularly rich source of ideas has been the study of random graphs, graphs in which the edges are distributed randomly. Networks with a complex topology and unknown organizing principles often appear random; thus random-graph theory is regularly used in the study of complex networks. The theory of random graphs was introduced by Paul Erdős and Alfréd Rényi [15,37,38] after Erdős discovered that probabilistic methods were often useful in tackling problems in graph theory. A detailed review of the field is available in the classic book of Bollobás [39]. Here we briefly describe the most important results of random graph theory, focusing on the aspects that are of direct relevance to complex networks.

a. The Erdős–Rényi Model In their classic first article on random graphs, Erdős and Rényi define a random graph as N labeled nodes connected by n edges, which are chosen randomly from the $N(N - 1)/2$ possible edges [15].

In a random graph with connection probability p the degree k_i of a node i follows a binomial distribution with parameters $N - 1$ and p :

$$P(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{9}$$

This probability represents the number of ways in which k edges can be drawn from a certain node. To find the degree distribution of the graph, we need to study the number of nodes with degree k , N_k . Our main goal is to determine the probability that N_k takes on a given value, $P(N_k = r)$. According to equation 9, the expectation value of the number of nodes with degree k is

$$E(N_k) = NP(k_i = k) = \lambda_k \tag{10}$$

with

$$\lambda_k = N \binom{k}{N-1} p^k (1-p)^{N-1-k} \quad (11)$$

The distribution of the N_k values, $P(N_k = r)$, approaches a Poisson distribution,

$$P(X_k = r) = e^{-\lambda_k} \quad (12)$$

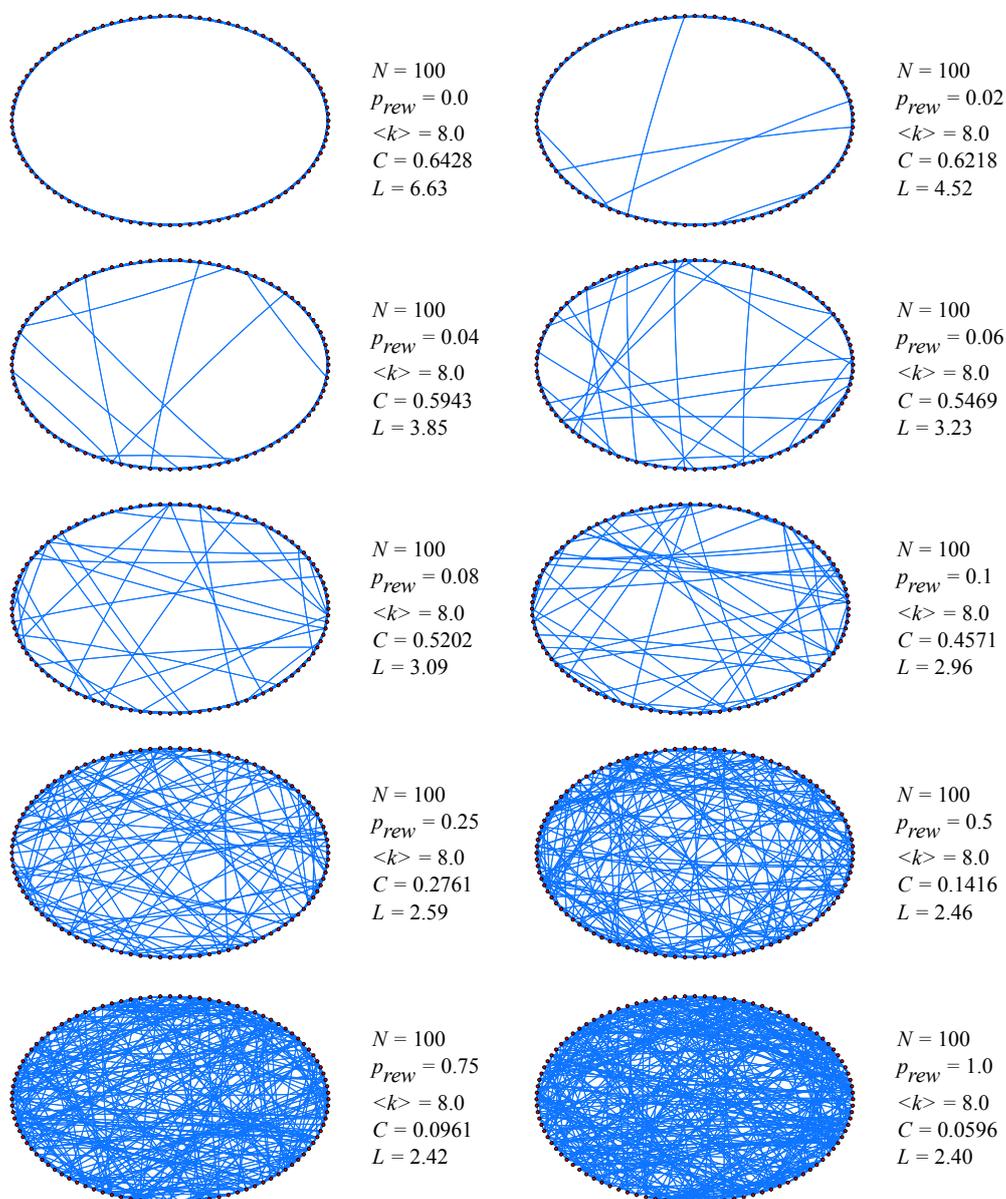
Thus the number of nodes with degree k follows a Poisson distribution with mean value λ_k . Although random graph theory is elegant and simple, and Erdős and other authors in the social sciences, like Rapoport [40–43], believed it corresponded fundamental truth, reality interpreted as a network by current science is not aleatory. The established links between the nodes of various domains of reality follow fundamental natural laws. Despite some edges might be randomly set up, and they might play a non-negligible role, randomness is not the main feature in real networks. Therefore, the development of new models to capture real-life systems' features other than randomness has motivated novel developments. Specially, two of these new models occupy a prominent place in contemporary thinking about complex networks. Here we define and briefly discuss them.

b. Watts–Strogatz small-world network In simple terms, the small-world concept describes the fact that despite their often large size, in most networks there is a relatively short path between any two nodes. The distance between two nodes is defined as the number of edges along the shortest path connecting them. The most popular manifestation of small worlds is the “six degrees of separation” concept, uncovered by the social psychologist Stanley Milgram [4,5], who concluded that there was a path of acquaintances with a typical length of about six between most pairs of people in the United States. This feature (short path lengths) is also present in random graphs. However, in a random graph, since the edges are distributed randomly, the clustering coefficient is considerably small. Instead, in most, if not all, real networks the clustering coefficient is typically much larger than it is in a comparable random network (*i.e.*, same number of nodes and edges as the real network). Beyond Milgram’s experiment, it was not until 1998 that Watts and Strogatz’ work [6] stimulated the study of such phenomena. Their main discovery was the distinctive combination of high clustering with short characteristic path length, which is typical in real-world networks (either social, biological or technological) that cannot be captured by traditional approximations such as those based on regular lattices or random graphs. From a computational point of view, Watts and Strogatz proposed a one-parameter model that interpolates between an ordered finite dimensional lattice and a random graph. The algorithm behind the model is the following [6]:

- *Start with order*: Start with a ring lattice with N nodes in which every node is connected to its first k neighbors ($k/2$ on either side). In order to have a sparse but connected network at all times, consider $N \gg k \gg \ln(N) \gg 1$.
- *Randomize*: Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded. This process introduces $pNk/2$ long-range edges which connect nodes that otherwise would be part of different neighborhoods. By varying p one can closely monitor the transition between order ($p=0$) and randomness ($p=1$).

The simple but interesting result when applying the algorithm was the following. Even for a small probability of rewiring, when the local properties of the network are still nearly the same as for the original regular lattice and the average clustering coefficient does not differ essentially from its initial value, the average shortest-path length is already of the order of the one for classical random graphs (see Figure 2).

Figure 2. From regularity to randomness: note the changes in average path length and clustering coefficient as a function of the rewiring probability $L(p)$, $C(p)$ for the family of randomly rewired graphs. For low rewiring probabilities the clustering is still close to its initial value, whereas the average path length has already decreased significantly. For high probabilities, the clustering has dropped to an order of 10^{-2} . This figure illustrates the fact that small-world is not a network, but a family of networks.



As discussed in [44], the origin of the rapid drop in the average path length L is the appearance of shortcuts between nodes. Every shortcut, created at random, is likely to connect widely separated parts of the graph, and thus has a significant impact on the characteristic path length of the entire graph. Even a relatively low fraction of shortcuts is sufficient to drastically decrease the average path length, yet locally the network remains highly ordered. In addition to a short average path length, small-world networks have a relatively high clustering coefficient. The Watts–Strogatz model (SW) displays this duality for a wide range of the rewiring probabilities p . In a regular lattice the clustering coefficient does not depend on the size of the lattice but only on its topology. As the edges of the network are randomized, the clustering coefficient remains close to $C(0)$ up to relatively large values of p .

Scale-Free Networks Certainly, the SW model initiated a revival of network modeling in the past few years. However, there are some real-world phenomena that small-world networks can't capture, the most relevant one being evolution. In 1999, Barabási and Albert presented some data and formal work that has led to the construction of various scale-free models that, by focusing on the network dynamics, aim to offer a universal theory of network evolution [16].

Several empirical results demonstrate that many large networks are scale free, that is, their degree distribution follows a power law for large k . The important question is then: what is the mechanism responsible for the emergence of scale-free networks? Answering this question requires a shift from modeling network topology to modeling the network assembly and evolution. While the goal of the former models is to construct a graph with correct topological features, the modeling of scale-free networks will put the emphasis on capturing the network dynamics.

In the first place, the network models discussed up to now (random and small-world) assume that graphs start with a fixed number N of vertices that are then randomly connected or rewired, without modifying N . In contrast, most real-world networks describe open systems that *grow* by the continuous addition of new nodes. Starting from a small nucleus of nodes, the number of nodes increases throughout the lifetime of the network by the subsequent addition of new nodes. For example, the World Wide Web grows exponentially in time by the addition of new web pages.

Second, network models discussed so far assume that the probability that two nodes are connected (or their connection is rewired) is independent of the nodes degree, *i.e.*, new edges are placed randomly. Most real networks, however, exhibit *preferential attachment*, such that the likelihood of connecting to a node depends on the nodes degree. For example, a web page will more likely include hyperlinks to popular documents with already high degrees, because such highly connected documents are easy to find and thus well known.

a. The Barabási–Albert model These two ingredients, growth and preferential attachment, inspired the introduction of the Barabási–Albert model (BA), which led for the first time to a network with a power-law degree distribution. The algorithm of the BA model is the following:

1. *Growth*: Starting with a small number (m_0) of nodes, at every time step, we add a new node with $m(\leq m_0)$ edges that link the new node to m different nodes already present in the system.

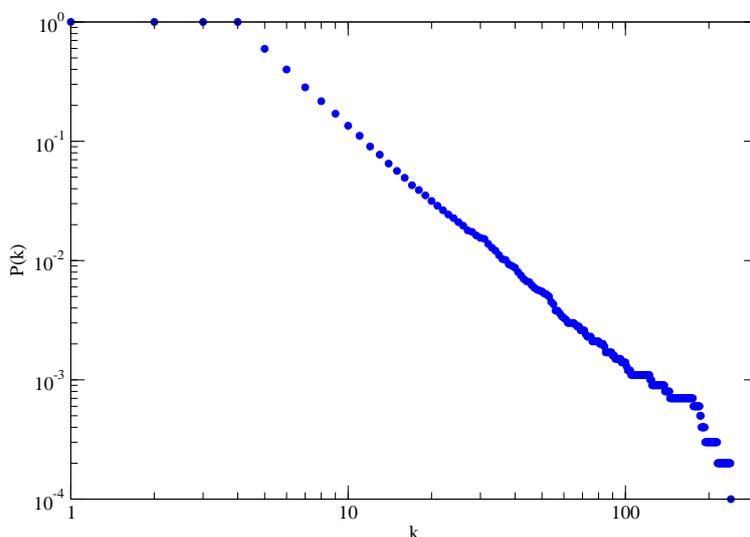
2. *Preferential attachment*: When choosing the nodes to which the new node connects, we assume that the probability Π that a new node will be connected to node i depends on the degree k_i of node i , such that

$$\Pi = \frac{k_i}{\sum_j k_j} \tag{13}$$

It is specially in step (1) of the algorithm that the scale-free model captures the dynamics of a system. The power-law scaling in the BA model indicates that growth and preferential attachment play important roles in network development. However, some question arise when considering step (2): admitting that new nodes' attachment might be preferential, is there only one equation (specifically, the one mentioned here) that grasps such preference across different networks (social, technological, *etc.*)? Can preferential attachment be expressed otherwise?

In the limit $t \rightarrow \infty$ (network with infinite size), the BA model produces a degree distribution $P(k) \approx k^{-\gamma}$, with an exponent $\gamma = 3$, see Figure 3.

Figure 3. Cumulative degree distribution for a SF with $N = 10000$, constructed according to the BA model. For each node entering the network, 3 new edges are placed. The horizontal axis is vertex degree k and the vertical axis is the cumulative probability distribution of degrees, *i.e.*, the fraction of vertices that have degree greater than or equal to k .



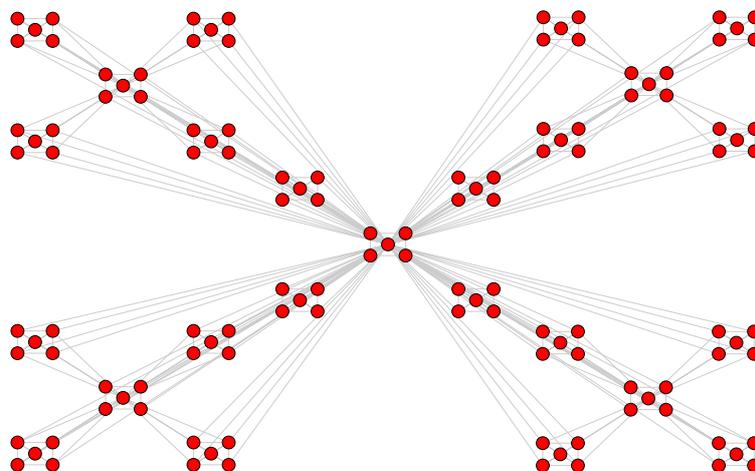
The average distance in the BA model is smaller than in a ER-random graph with same N , and increases logarithmically with N . Analytical results predict a double logarithmic correction to the logarithmic dependence $L \sim \frac{\log N}{\log(\log N)}$. The clustering coefficient vanishes with the system size as $C \sim N^{-0.75}$. This is a slower decay than that observed for random graphs, $C \sim \langle k \rangle N^{-1}$, but it is still different from the behavior in small-world models, where C is independent of N .

b. Other SF models The BA model has attracted an exceptional amount of attention in the literature. In addition to analytic and numerical studies of the model itself, many authors have proposed

modifications and generalizations to make the model a more realistic representation of real networks. Various generalizations, such as models with nonlinear preferential attachment, with dynamic edge rewiring, fitness models and hierarchically and deterministically growing models, can be found in the literature. Such models yield a more flexible value of the exponent γ which is restricted to $\gamma = 3$ in the original BA construction. Furthermore, modifications to reinforce the clustering property, which the BA model lacks, have also been considered.

Among these alternative models we can find the *Dorogovtsev–Mendes–Samukhin* (DMS) model, which considers a linear preferential attachment; or the *Ravasz–Barabási* (RB) model, which aims at reproducing the hierarchical organization observed in some real systems (this makes it useful as an appropriate benchmark for multi-resolution community detection algorithms, see next Section and Figure 4).

Figure 4. The RB model yields a hierarchical network, that combines the scale-free property with a high degree of clustering. The starting point is a small cluster of five densely linked nodes; next, four replicas of this hypothetical module are generated. The four external nodes of the replicated clusters are connected to the central node of the old cluster, obtaining a large 25-node module. This replication and connection can be repeated recursively, thus obtaining networks of size 25, 125, *etc.*



The *Klemm–Eguiluz* (KE) model seeks to reproduce the high clustering coefficient usually found in real networks, which the BA model fails to reproduce [45]. To do so, it describes the growth dynamics of a network in which each node of the network can be in two different states: active or inactive. The model starts with a complete graph of m active nodes. At each time step, a new node j with m outgoing links is added. Each of the m active nodes receives one incoming link from j . The new node j is then activated, while one of the m active nodes is deactivated. The probability \prod_i^{deact} that node i is deactivated is given by

$$\prod_i^{deact} = \frac{1}{k_i + a} \left(\sum_{l \in N_{act}} \frac{1}{k_l + a} \right)^{-1} \quad (14)$$

where k_i is the in-degree of node i , a is a positive constant and the summation runs over the set N_{act} of the currently active nodes. The procedure is iteratively repeated until the desired network size is reached. The model produces a scale-free network with $\gamma = 2 + a/m$ and with a clustering coefficient $C = 5/6$ when $a = m$. Since the characteristic path length is proportional to the network size ($L \sim N$) in the KE model, additional rewiring of edges is needed to recover the small-world property. Reference [19] thoroughly discusses these and other models.

1.4. The mesoscale level

Research on networks cannot be solely the identification of actual systems that mirror certain properties from formal models. Therefore, the network approach has necessarily come up with other tools that enrich the understanding of the structural properties of graphs. The study of networks (or the methods applied to them) can be classified in three levels:

- The study at the *micro* level attempts to understand the behavior of single nodes. Such level includes degree, clustering coefficient or betweenness and other parameters.
- *Meso* level points at group or community structure. At this level, it is interesting to focus on the interaction between nodes at short distances, or classification of nodes, as we shall see.
- Finally, *macro* level clarifies the general structure of a network. At this level, relevant parameters are average degree $\langle k \rangle$, degree distribution $P(k)$, average path length L , average clustering coefficient C , etc.

The first and third levels of topological description range from the microscopic to the macroscopic description in terms of statistical properties of the whole network. Between these two extremes we find the mesoscopic level of analysis of complex networks. In this level we describe an inhomogeneous connecting structure composed by subsets of nodes which are more densely linked, when compared to the rest of the network.

This mesoscopic scale of organization is commonly referred as *community structure*. It has been observed in many different contexts, including metabolic networks, banking networks or the worldwide flight transportation network [46]. Moreover, it has been proved that nodes belonging to a tight-knit community are more than likely to have other properties in common. For instance, in the world wide web community analysis has uncovered thematic clusters.

Whatever technique applied, the belonging of a node to one or another community cannot depend upon the “meaning” of the node, *i.e.*, it can't rely on the fact that a node represents an agent (sociology), a computer (the internet), a protein (metabolic network) or a word (semantic network). Thus communities must be determined solely by the topological properties of the network: nodes must be more connected within its community than with the rest of the network. Whatever strategy applied, it must be *blind* to content, and only *aware* of structure.

The problem of detection is particularly tricky and has been the subject of discussion in various disciplines. In real complex networks there is no way to find out, *a priori*, how many communities can be discovered, but in general there are more than two, making the process more costly. Furthermore, communities may also be hierarchical, that is communities may be further divided into sub-communities

and so on [47–49]. Summarizing, it is not clear at what point a community detection algorithm must stop its classification, because no prediction can be made about the right level of analysis.

A simple approach to quantify a given configuration into communities that has become widely accepted was proposed in [50]. It rests on the intuitive idea that random networks do not exhibit community structure. Let us imagine that we have an arbitrary network, and an arbitrary partition of that network into N_c communities. It is then possible to define a $N_c \times N_c$ size matrix e where the elements e_{ij} represent the fraction of total links starting at a node in partition i and ending at a node in partition j . Then, the sum of any row (or column) of e , $a_i = \sum_j e_{ij}$ corresponds to the fraction of links connected to i . If the network does not exhibit community structure, or if the partitions are allocated without any regard to the underlying structure, the expected value of the fraction of links within partitions can be estimated. It is simply the probability that a link begins at a node in i , a_i , multiplied by the fraction of links that end at a node in i , a_i . So the expected number of intra-community links is just $a_i a_i$. On the other hand we know that the real fraction of links exclusively within a community is e_{ii} . So, we can compare the two directly and sum over all the communities in the graph.

$$Q = \sum_j (e_{ii} - a_i)^2 \quad (15)$$

This is a measure known as *modularity*. Equation 15 has been extended to a directed and weighted framework, and even to one that admits negative weights [51]. Designing algorithms which optimize this value yields good community structure compared to a null (random) model. The problem is that the partition space of any graph (even relatively small ones) is huge (the search for the optimal modularity value seems to be a *NP*-hard problem due to the fact that the space of possible partitions grows faster than any power of the system size), and one needs a guide to navigate through this space and find maximum values. Some of the most successful heuristics are outlined in [52,53]. The first one relies on a genetic algorithm method (Extremal Optimization), while the second takes a greedy optimization (hill climbing) approach. Also, there exist methods to decrease the search space and partially relieve the cost of the optimization [54]. In [55] a comparison of different methods is developed, see also [56].

Modularity-based methods have been extended to analyze the community structure at different resolution levels, thus uncovering the possible hierarchical organization of the mesoscale [49,57,58].

With the methodological background developed in the previous Section 1, it is now possible to turn to language. The following Sections are devoted to acknowledge the main achievements of a complex network approach to language and the cognitive processes associated to it.

2. Building language networks

An often expressed concern around complex networks is their arbitrary character. When modeling actual, real-world systems using network methodology, the researcher needs to take some decisions: what kind of object must be understood as a vertex, in the first place; and more critical, what must be understood as a link between vertices. In our case, it is not straightforward to define the notion of word interaction in a unique way. For instance, one can connect the nearest neighbors in sentences. Also, one could take into account linguistic standard relations, like synonymy, hyper- or hyponymy, *etc.* Finally, one can assemble networks out of laboratory data, *i.e.*, data coming from experiments with subjects in

psycholinguistics. We detail these three lines in the subsequent paragraphs, closely following the ideas in [59].

2.1. Text analysis: co-occurrence graphs

Intuitively, the simplest strategy to collect relations among entities is to construct a network whose topology reflects the co-occurrence of words. Such intuition is rooted in *collocation analysis*, a well established field of corpus linguistics [60–62]. It follows a tradition according to which collocations manifest lexical semantic affinities beyond grammatical restrictions [63].

Typically, text co-occurrence networks are obtained with the minimum assumptions and cost, *i.e.*, a fixed adjacency window of width d is predefined, such that two words w_1 and w_2 are connected by an edge (link) if $d_{w_1-w_2} \leq d$. Thus, a *two-word* adjacency network automatically connects a word with any two words before and after it. Often articles and other connecting words are excluded. Their topology quantified by several measurements can provide information on some properties of the text, such as style and authorship [64].

Some limitations must be taken into account under this constructive method: if d is long, the risk of capturing spurious co-occurrences increases. If d is too short, certain strong co-occurrences can be systematically not taken into account [65].

The textual sources for these type of networks can be varied. In some cases a single source is chosen (for example, a book from a particular author). In other cases, collections of newspapers or magazines are used (as in the ACE corpus [66]). This subtle difference is important, in the first case the resulting structure reflects (at least partially) the lexical organization of an individual; whereas the latter provides an access to the semantic collective system of a language, that is, to the overall organization of its lexical subsystem [67]. This distinction already points in two research poles, the more cognitive- and the more language-oriented, which shall appear later.

2.2. Dictionaries and Thesauri

As in the case of multi-source text analysis, again a collective view on language is predominant in the case of dictionaries. *Lexical reference systems* or *terminological ontologies* (e.g., WordNet, [1]), *thesauri* (e.g., Roget's thesaurus, [2]) and related systems build on expert knowledge of lexicographers in order to define *sense relations* (e.g., synonymy, antonymy, hyponymy) between words or conceptual relations between concepts (therefore, they are meaning-based). Following [59], in the case of *thesaurus* graphs based on the expertise of lexicographers and corpus linguists, the characteristics of the network can be interpreted as indicators of *thesaurus* quality or consistency. For instance, a graph representing hyponymy relations within a *thesaurus* should induce a hierarchical structure, whereas polysemy should provide for the small world nature of the semantic system of the language under consideration. Such is the case of Wordnet in the study by Sigman and Cecchi [67].

2.3. Semantic features

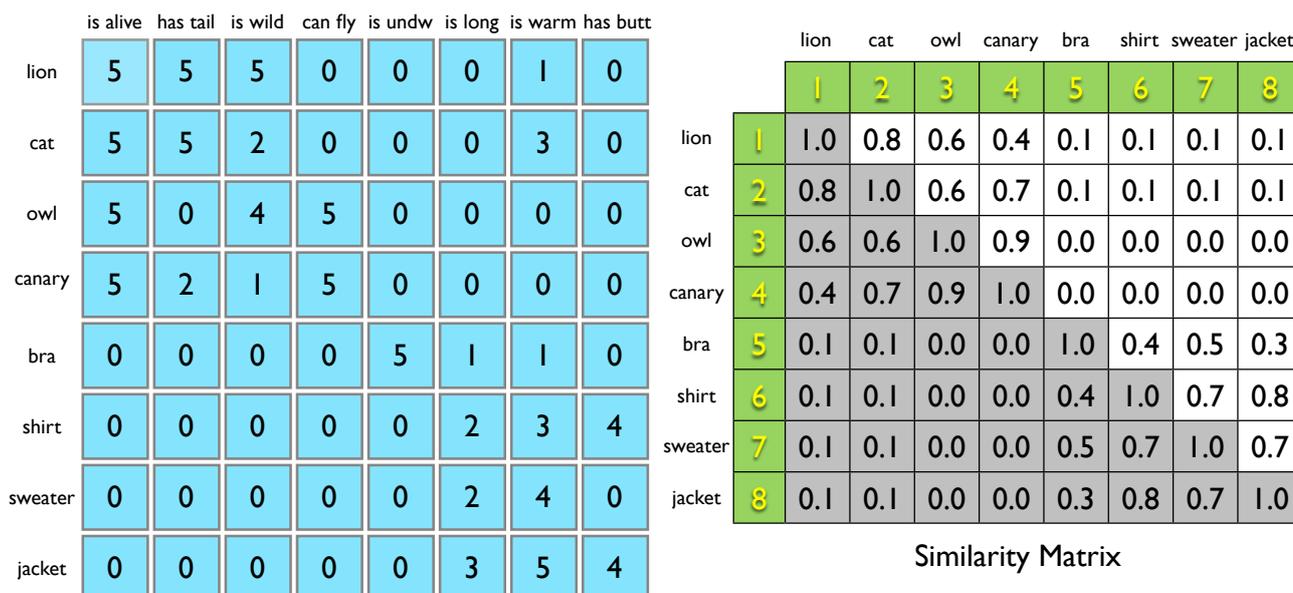
In many of the most influential theories of word meaning and of concepts and categorization, semantic features have been used as their representational currency. Numerous vector models of memory are based

on feature representations. For this reason, the major purpose of collecting semantic feature production norms is to construct empirically derived conceptual representation and computation.

One of the most relevant example of such data collection is that of McRae *et al.* [68] Feature Production Norms, which were produced 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 represents a feature. In particular, participants are presented with a set of concept names and are asked to produce features they think are important for each concept. Each feature stands as a vector component, with a value that represents its production frequency across participants. These norms include 541 living and nonliving thing concepts, for which semantic closeness or *similarity* is computed as the cosine (overlap) between pairs of vectors of characteristics. The cosine is obtained as the dot product between two concept vectors, divided by the product of their lengths:

$$\cos \theta = \frac{v_1w_1 + v_2w_2 + \dots + v_nw_n}{\| v \| \| w \|} \tag{16}$$

Figure 5. A network structure out of semantic features data. Left: each subject assigns semantic features to given nouns, and features build up a semantic vector. In the example, features are *is alive*, *has tail*, *is wild*, *can fly*, *is underwear*, *is long*, *is warm* and *has buttons*. The number in each cell reflects the number of participants who assigned that feature to the corresponding item. Right: cosine overlapping between each pair of vectors from the left matrix. This new similarity matrix can be suitably interpreted as a semantic network. Note that values in both matrices do not represent actual results, and have been put merely for illustrative purposes.



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).

In terms of network modeling, 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 (or above a predefined threshold τ). The meaning of an edge in this network is thus the features similarity between two words. The network is undirected (symmetric relationships) and weighted by the value of the projections. See Figure 5 for illustration.

Although these measures are not obtained from an individual, but rather averaged out of many participants in an experiment, this type of data is in the line of cognitive research, in which network modeling is a tool to understand actual mechanisms of human language usage. The same can be said regarding associative networks, in the next subsection.

2.4. Associative networks

Association graphs are networks in which vertices denote words, whereas links represent association relations as observed in cognitive-linguistic experiments. Such graphs are considered the most relevant from a psychological point of view. According to the hypothesis that association is one of the principles of memory organization, the question that has to be addressed is which network topologies support an efficient organization in terms of time and space complexity.

The best known Free Association data set in English are University of South Florida Free Association Norms (USF-FA from now on; [69]). 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. Weights represent the frequency of association in the sample. These same data exist in Spanish [70,71], German [72] or French [73].

Generally speaking, Free-Association Norms represent a more complex scenario than Feature Production Norms when considering the semantics of edges. Free-Association Norms are heterogeneous by construction, they 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.

From this data set, two networks can be created. A directed network, where two word nodes i and j are joined by an arc (from i to j) if the cue i evoked j as an associative response for at least two of the participants in the database. In an undirected version, word nodes are joined by an edge if the words were associatively related regardless of associative direction. Although the directed network is clearly a more natural representation of word associations, most of the literature on small-world and scale-free networks has focused on undirected networks.

The following sections attempt to review some works centered on network modeling of language. We will move gradually from the language-oriented pole, which is concerned with general structural patterns

and dynamics of language; towards the cognitive-oriented one, which is confronted with a greater degree of detail and complexity.

3. Language networks: topology, function, evolution

Soon after the seminal works by Watts and Strogatz, and Barabási and Albert in the late '90s, network scientists focused upon language as an interesting subject. Unsurprisingly, the insights were general in this initial stage, and became deeper from then on.

Table 1. Results for the conceptual network defined by the Thesaurus dictionary, and a comparison with a corresponding random network with the same parameters. N is the total number of nodes, $\langle k \rangle$ is the average number of links per node, C is the clustering coefficient, and L is the average shortest path. After [74].

	N	$\langle k \rangle$	C	L
Moby Thesaurus	30244	59.9	0.53	3.16
Randomized MT	30244	59.9	0.002	2.5

One of the first approaches is that of Motter *et al.* in [74], where the network structure of language is studied. The author presents results for the English language, which are expected to hold for any other language. A conceptual network is built from the entries in the Moby Thesaurus, and considers two words connected if they express similar concepts. Motter *et al.*'s resulting network includes over 30,000 nodes (words). Table 1 summarizes the results of the analysis. The random counterpart of the Moby Thesaurus [75] highlights the small-world features of the original, *i.e.*, high clustering coefficient and low average path length.

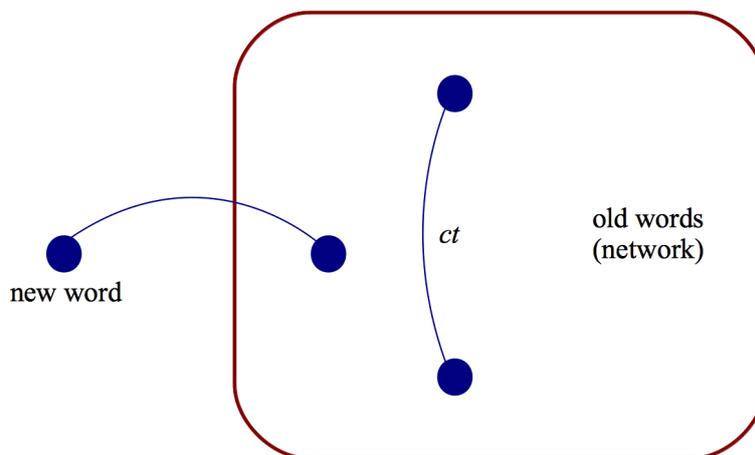
Similarly, Sigman and Cecchi thoroughly characterize the WordNet database [67], with similar results. Their analysis of the degree distribution results in power-law distributions, the fingerprint of self-organizing, evolving systems.

Dorogovstev and Mendes explore the mentioned possibility in [76], namely that language (or, more precisely, lexicon) is a self-organized growing system. Specifically, they discuss whether empirical degree distributions might be the result of some type of preferential attachment dynamics. The authors propose a stochastic theory of evolution of human language based on the treatment of language as an evolving network of interacting words. It is well known that language evolves, then the question is what kind of growth (in the sense of increase of lexical repertoire) leads to a self-organized structure? Although in the general framework of Barabási and Albert's preferential attachment, their proposal adds a second growth mechanism inspired in observations from real collaboration networks. This variation includes, at each time step, the appearance of new edges between already-existing (old) words, besides the birth of new words that link to some old ones (see Figure 6).

The model can be described in a precise analytical form. It is possible to detail the evolution of the degree of a certain word "born" at time s and observed at time t :

$$\frac{\partial k(s, t)}{\partial t} = (1 + 2ct) \frac{k(s, t)}{\int_0^t du k(u, t)} \quad (17)$$

Figure 6. Dorogovstev and Mendes’ scheme of the language network growth [76]: a new word is connected to some old one i with the probability proportional to its degree k_i (Barabási and Albert’s preferential attachment); in addition, at each increment of time, ct new edges emerge between old words, where c is a constant coefficient that characterizes a particular network.



The development of Equation 17 leads to a description of the evolution of the degree distribution $P(k, t)$, which matches the empirical findings in [65], *i.e.*, a two-regime power-law with different exponents, see comments below.

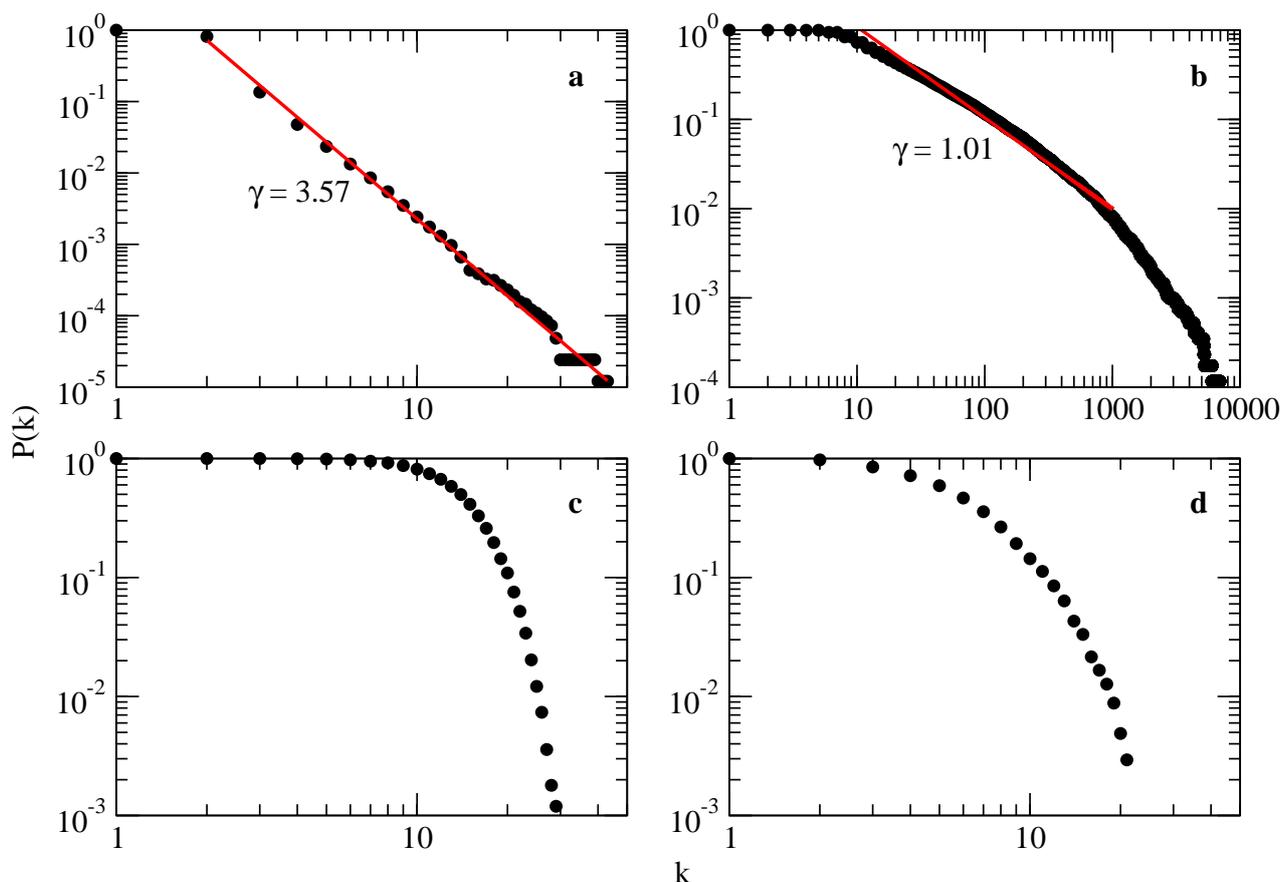
Table 2. Some parameters obtained from four different data-sets: the University of South Florida word association (USF-FA, [69]), Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures (SFA-SV, [71]), association norms in Spanish (SFA, [70]) and association norms for the German names of the Snodgrass and Vanderwart pictures (GFA-SV, [72]). As the ones reported on Table 1, they all conform sparse structures with very low L (if compared to the size of the network). However, only USF-FA and SFA clearly fit in the small-world definition. Low C in the data sets based on the drawings from Snodgrass and Vanderwart [77] can be explained by the specific experimental setup with this material. N is the total number of nodes, $\langle k \rangle$ is the average number of links per node, C is the clustering coefficient, L is the average shortest path, and D is the diameter. The latter descriptors (L and D) have been measured from the undirected, unweighted networks of the data sets.

	N	$\langle k \rangle$	C	L	D
USF-FA	5018	22	0.1928	3.04	5
SFA-SV	7759	3.05	0.0867	3.71	5
SFA	2901	4.9	0.1487	4.50	8
GFA-SV	3632	2.05	0.034	4.57	8

Different language networks display as well similar small-world characteristics, see Table 2. Also, their degree distribution corresponds in some cases to scale-free networks, see Figure 7 (remarkably, the

high interest in scale-free networks might give the impression that all complex networks in nature have power-law degree distributions. As is shown in the mentioned figure, this is far from being the case).

Figure 7. Plots of the cumulative degree distribution in four networks. All of them have been converted to unweighted and undirected. (a) WordNet, hypernymy relationships; (b) Co-occurrence networks for variable window size, from the ACE corpus; (c) English Free Association Norms (USF-FA); (d) Roget’s thesaurus. Note that the plots are drawn in log-log scale. Only (a) and (b) display a power-law decay, whereas (c) and (d) do not follow a scale-free distribution. All of them, nonetheless, fit in the small-world definition.



Most interestingly, these early results led to the claim that they have deep cognitive implications. From the standpoint of retrieval of information, the small-world property of the network represents a maximization of efficiency: high clustering gathers similar pieces of information, low distances makes fast search and retrieval possible. The expression “mental navigation” arises: irrespective of the specifics of the neuronal implementation, it can be thought that the small-world property is a desirable one in a navigation network (it strikes a balance between the number of active connections and the number of steps required to access any node); and, taking mental navigation for granted, one could also expect that the hubs of the network should display a statistical bias for priming in association and related tasks [67]. Navigation, in this context, corresponds to *retrieval* in semantic memory, understood as intentional recovery of a word. “Mental exploration” would instead correspond to search processes (such as when

trying to produce words that begin for a certain letter): there is no topological information to achieve this purpose in the network. In both processes shortcuts and hubs must significantly affect proficiency.

These intuitions probably point at the right direction, but there is a need to focus the attention on some specific phenomena. Then, since linguistic phenomena does not occur outside the boundaries of cognition, research necessarily turned towards the cognitive pole.

The work by Ferrer i Cancho and Solé represents significant steps in this direction. For instance, a difference is settled between single- and multi-author linguistic sources. In [65], a network $N \approx 5 \times 10^5$ words is built out of the British National Corpus (BNC). The degree distribution of such network evidences a two-regime power law, one of them with an average exponent close to the Barabási-Albert model ($\gamma_{BA} = -3$). From this twofold behavior the authors claim that the lexicon is divided into a set of core words (kernel, $\gamma = -2.7$) and a set of peripheral words ($\gamma = -1.5$). The kernel lexicon contains words that are common to the whole community of speakers, while in the periphery a certain word is unknown for one speaker and familiar for another. Results suggest that language has grown under the dynamics of preferential attachment, the core of the network (with $\gamma \approx \gamma_{BA}$) containing at least functional words, *i.e.*, those with low or null semantic content. This approach takes into account not only the features of complex physical systems (self-organization, *etc.*), but how can they be explained in terms of collective behavior.

This “physical system–cognitive phenomena” mapping is again visible in [78,79]. The question here is to give account of Zipf’s least effort principle [80] using network methodology and information theory [81]. Again, the center of the discussion is a cognitive phenomenon (communication) in which a speaker and a listener are involved. As it is well known, word frequencies in human language obey a universal regularity, the so-called Zipf’s law. If $P(f)$ is the proportion of words whose frequency is f in a text, we obtain $P(f) \propto f^{-\beta}$, with $\beta \in [1.6, 2.4]$. Given this interval, the author’s claim is that the exponent of Zipf’s law depends on a balance between maximizing information transfer and saving the cost of signal use. This trade-off is in close relation to the one reported in [82] according to the expression

$$\Omega = \lambda I(S, R) - (1 - \lambda)H(S) \quad (18)$$

where Ω is the energy function that a communication system must minimize, $I(S, R)$ denotes the Shannon information transfer between the set of signals S and the set of stimuli R ; and $H(S)$ is the entropy associated to signals, *i.e.*, the cost of signal use present in any communication [78]. In this context, $\lambda \in [0, 1]$ is a parameter regulating the balance between the goals of communication (maximize transfer of information) and its cost. Of course, $\lambda = 1$ results in a completely effective communication, whereas $\lambda = 0$ leads to a costless (though null) communication.

Given this framework, energy Ω can be minimized for different values of λ . Results show a sudden jump from close to null information transfer (low values of λ) to a maximum information transfer at a critical value $\lambda^* \approx 0.5$. For values $\lambda > \lambda^*$, $I(S, R)$ does not increase. These results are in turn interpreted in the context of networks in [79], by showing that favoring information transfer without regard of the cost (low values of λ) corresponds to a dense, richly interconnected network (information availability); above a threshold, the situation is reversed and the network of signals and stimulus (language) is broken or disconnected (certain parts of language remain unreachable). The change from one to another scenario occurs, again, in the form of a phase transition at a certain critical value.

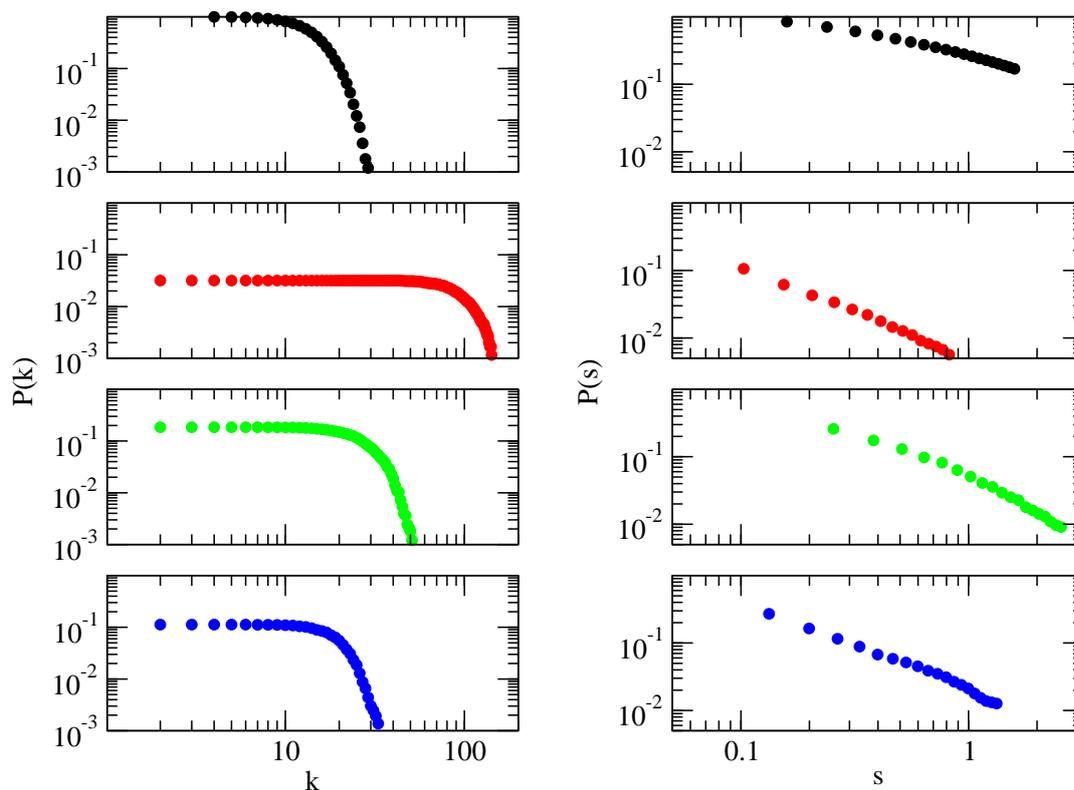
Table 3. Language-oriented complex networks works.

Graph	Source Network	Vertex	Edge	Orient.	N	$\langle k \rangle$	L	C	Reference
thesaurus graph	Moby's thesaurus	word	sense relation	undir.	30,244	59.9	3.16	0.53	[74]
collocation graph	BNC corpus	word	collocation	undir.	460,902	70.13	2.67	0.44	[65]
co-occurrence graph	BNC corpus	word	co-occurrence	undir.	478,773	74.2	2.63	0.69	[65]
thesaurus graph	Roget's thesaurus	word	sense relation	undir.	29,381	S. (3.3)	5.60	0.87	[83]
concept graph	WordNet	word	sense relation	undir.	122,005	3.3	10.56	0.03	[83]
association graph	free assoc. data	word	association	undir.	5,018	22.0	3.04	0.19	[83]
association graph	free assoc. data	word	association	dir.	5,018	12.7	4.27	0.19	[83]

Up to now, we have been able to assess the existence of certain universal statistical trends (see Table 3 and references therein), and we have placed language networks in the framework of information and communication theory, which approaches them to its natural place, *i.e.*, embedded in human cognition.

Thus, we now fully turn to the cognitive-oriented research. As Solé *et al.* [9] point out, some (possibly interacting) factors must be considered for a more comprehensive view on linguistic phenomena, for instance: a common brain architecture and vocalization system, or the need for optimization in communication and learnability. These new considerations have turned the attention of research towards a cognitive-oriented work, where the network is not the object of analysis anymore (or not exclusively, at least); rather it is the object *on top of which* language cognitive mechanisms operate. Furthermore, more attention is put both on the type of data and its original meaning: while a coarse-grain general study on structural principles usually treats with undirected, unweighted networks, the cognitive approach tries to preserve as much as possible the original structures. By doing so, the natural heterogeneity and bias in cognitive phenomena are preserved. For instance, Figure 8 illustrates how misleading it can be to oversee the details in data. Summarizing, the study of cognitive processes demands a lower level of detail, where it matters whether a word facilitates another one, but not the other way around; or whether two words are semantically similar up to 0.9, whereas another pair reaches only 0.1. Both situations are treated as symmetric unweighted relationships in most complex network overviews of language.

Figure 8. Directions and weights matter. Left: log-log plots of the cumulative degree distributions for psycholinguistic data in four languages (from top to bottom: USF, SFA-SV, SFA and GFA). Directions are symmetrized and weights are not taken into account. Right: log-log plots of the cumulative in-strength distribution for the same data without manipulation. Note that there exist striking differences between degree and strength distributions of psycholinguistic data. These differences are also evident in other descriptors, which suggests that comprehension about cognitive-linguistic processes demand attention to such details.

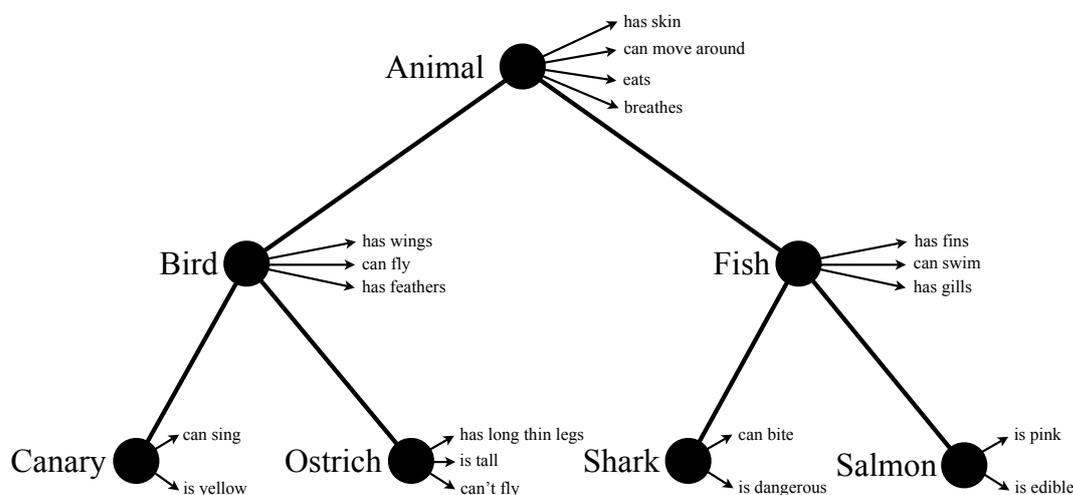


4. The cognitive pole I: Language and conceptual development

The work by Steyvers and Tenenbaum in 2005 [83] represents, up to date, the most comprehensive effort to join cognitive science with complex systems. As a confluence of these disciplines the authors vindicate the group of theories in psychology of memory which, under the label of *semantic networks*, were developed forty years ago [84,85]. These classic semantic networks often represent defined relationships between entities, and the topological structure is typically defined by the designer. A classical example of this type of semantic network is Collins and Quillian's groundbreaking work in 1969 [84]. These authors suggested that concepts are represented as nodes in a tree-structured hierarchy, with connections determined by class-inclusion relations. Additional nodes for characteristic attributes or predicates are linked to the most general level of the hierarchy to which they apply, see Figure 9. Collins and Quillian proposed algorithms for efficiently searching these inheritance hierarchies to retrieve or verify facts such as *Robins have wings*, and they showed that reaction times of human subjects often

seemed to match the qualitative predictions of their model. Word retrieval and recognition processes involve, in this proposal, tracing out the structure in parallel (simulated in the computer by a breadth-first search algorithm) along the links from the node of each concept specified by the input words. Such tracing process is known as “spreading activation”. The spread of activation constantly expands, first to all the nodes linked to the first node, then to all the nodes linked to each of these nodes, and so on. At each node reached in this process, an activation tag is left that specifies the starting node and the immediate predecessor. When a tag from another starting node is encountered, an intersection between the two nodes has been found. By following the tags back to both starting nodes, the path that led to the intersection can be reconstructed. Interestingly, the relation between structure and performance is addressed in terms of the *cognitive economy principle*. Such principle, in its weak version, imposes certain constraints on the amount of information stored per node, thus affecting the structure (and its growth) in behalf of better future performance, see [84,85] for further development.

Figure 9. The Collins and Quillians tree data structure provides a particularly economical system for representing knowledge about categories. The cognitive economy principle prevents the structure from having redundant information, thus features which belong to one level do not appear in any other. Despite some positive experimental results with humans, the structure is far too rigid to accommodate actual semantic knowledge.



A tree-structured hierarchy provides a particularly economical system for representing default knowledge about categories, but it places too strong constraints on the possible ways of organizing knowledge. Moreover, it has severe limitations as a general model of semantic structure. Inheritance hierarchies are clearly appropriate only for certain taxonomically organized concepts, such as classes of animals or other natural kinds.

The second classical proposal is that of Collins and Loftus [85] which, although accepting many of Collins and Quillians premises, assumes a quite different data structure: a graph (notice that a graph is a general case of a tree; or, to put it the other way around, a tree is a particular case of a graph). Collins and Loftus model does not differentiate between concepts and their attributes. Therefore, nodes in the graph can either be nouns (such as “apple”), adjectives (such as “red”), or even compounded expressions (such

as “fire engine”). Edges connecting them express a semantic relationship between them (not necessarily a category or similarity relationship), and it is assigned a number (a weight). Therefore, Collins and Loftus proposal yields an undirected, weighted graph which formally resembles very much the type of network that has been reviewed along this work.

Note that conceptually there is not much distance between Collins and Loftus graph proposal and complex networks. However, perhaps because of the limited prediction power of these proposals, perhaps because other points of view evidenced higher success at that time, the following decades did not witness a prolongation of these seminal works. As a consequence, there is relatively small agreement about general principles governing the large-scale structure of semantic memory, or how that structure interacts with processes of memory search or knowledge acquisition.

A complex network approach to language emerges naturally from this tradition, thus the work of Steyvers and Tenenbaum can be thought of as an update, both from the point of view of methodology and data availability. Although this work has a wide scope, part of it reports similar results as those reviewed in the previous Section, for instance a structural characterization of WordNet, Roget’s Thesaurus and USF-FA. Our interest is focused now on the genuine cognitive approach to language learning or growth in an individual (*lexical development*).

The first part of the question can be stated: is it possible to find a variation on Barabási and Albert’s preferential attachment which guarantees the emergence of a small-world, scale-free network? This question was already tackled by Dorogovstev and Mendes, as we have seen above. The novelty lies on the fact that the goal is to explain the statistics of semantic networks *as the products of a general family of psychologically plausible developmental processes*. In particular, (i) it is assumed that semantic structures grow primarily through a process of differentiation: the meaning of a new word or concept typically consists of some kind of variation on the meaning of an existing word or concept; (ii) it is assumed that the probability of differentiating a particular node at each time step is proportional to its current complexity (how many connections it has); and finally, (iii) nodes are allowed to vary in a “utility” variable, which modulates the probability that they will be the targets of new connections.

These constraints are translated to an algorithm which departs from a clique (fully connected network) of M initial nodes. Then, a node i is chosen to be differentiated at time t with probability $P_i(t)$ to be proportional to the complexity of the corresponding word or concept, as measured by its number of connections:

$$P_i(t) = \frac{k_i(t)}{\sum_{l=1}^{n(t)} k_l(t)} \quad (19)$$

where $k_i(t)$ is the degree (number of connections) of node i at time t . Secondly, given that node i has been selected for differentiation, we take the probability $P_{ij}(t)$ of connecting to a particular node j in the neighborhood of node i to be proportional to the utility of the corresponding word or concept:

$$P_{ij}(t) = \frac{u_j}{\sum_{l \in \Gamma_i} u_l} \quad (20)$$

where Γ_i stands for the neighborhood of node i . One possibility is to equate a word's utility with its frequency; for a simpler model, one may also take all utilities to be equal, then connection probabilities are simply distributed uniformly over the neighborhood of node i :

$$P_{ij}(t) = \frac{1}{k_i(t)} \quad (21)$$

With these equations (Equations 19–20, or Equations 19–21) each new node is connected to M old nodes. Nodes are added to the network until the desired size N is reached. With these constructive algorithm a synthetic network is obtained, and its statistical features can be compared to the empirical counterparts. Steyvers and Tenenbaum report a significant agreement on degree distribution $P(k)$ match, as well as on some quantities, which are reproduced in Table 4.

Table 4. Results of model simulations (undirected version). γ is the exponent of the power-law that describes $P(k)$. Standard deviations of 50 simulations given between parentheses.

	N	$\langle k \rangle$	C	L	D	γ
USF-FA	5018	23.5	0.1928	3.04	5	3.01
Synthetic USF	5018	22	0.174	3.00(.012)	5(.000)	2.95(.054)

5. The cognitive pole II: Cognitive-linguistic processes on language networks

In the following subsections, we report three examples of the application of complex systems techniques to gain insight on genuine cognitive phenomena. All the network concepts that appear subsequently have been developed in Section 1.

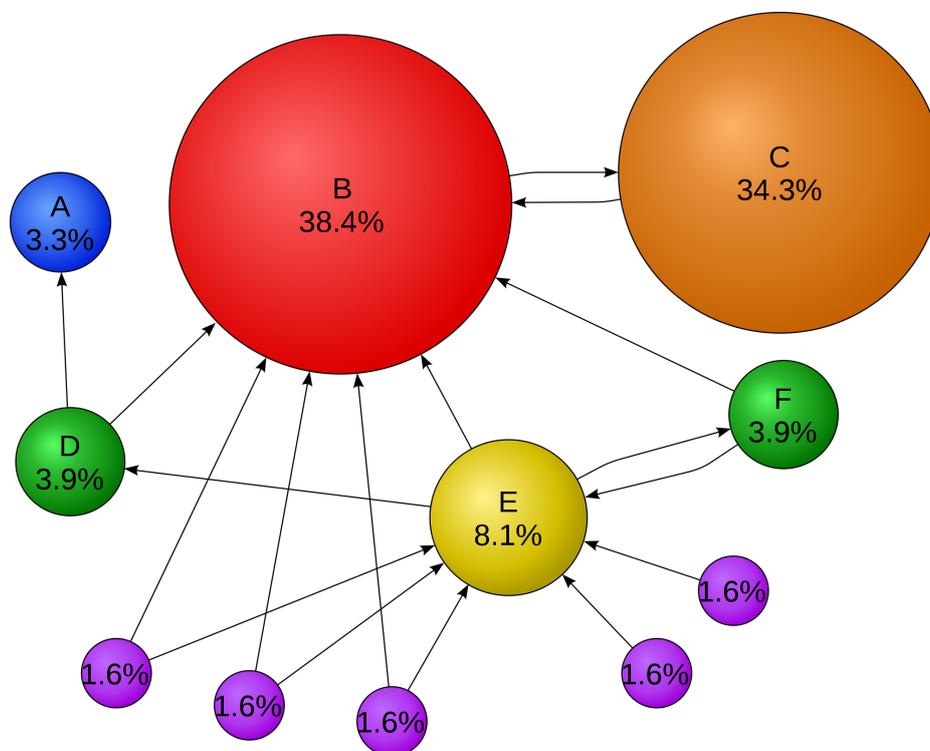
5.1. Google and the mind

The world wide web (WWW) presents at least two resemblances to associative models of language. First, it is organized as a directed network (nodes are web pages and the links between those nodes are hyperlinks, in the case of the WWW); second, its structure is dominated by the contents of its nodes. These factors add up to the fact that both human semantic memory and Internet face a shared computational problem, namely the necessity to retrieve stored pieces of information in an efficient way.

Given this, Griffiths and co-authors point out a very interesting parallelism between the PageRank algorithm [32] (see Figure 10) and human performance on certain cognitive processes [86].

To explore the correspondence between PageRank and human memory, the authors used a task that closely parallels the formal structure of Internet search. In this task, people were shown a letter of the alphabet (the query) and asked them to say the first word beginning with that letter that came to mind. The aim was to mimic the problem solved by Internet search engines, which retrieve all pages containing the set of search terms, and thus to obtain a direct estimate of the prominence of different words in human memory. In memory research, such a task is used to measure *fluency* (the ease with which people retrieve different facts). With this experimental setup, accordance between the word's rank given by the algorithm and by empirical data is measured.

Figure 10. An illustration of the output of the PageRank algorithm. A link from an important web page is a better indicator of importance than a link from an unimportant web page. Under such a view, an important web page is one that receives many links from other important web pages. From Wikipedia Commons.



Results evidence that verbal fluency can be predicted, at least partially, attending the prominence (*i.e.*, centrality) of words in memory. Furthermore, PageRank performs better predictions than those obtained attending word usage frequency.

In the context of this review, note that the work of Griffiths and co-authors involves experimental design and direct, detailed comparison between the theoretical hypothesis and empirical results. From this point of view, the mixture of cognitive research and complex network methodology represents a real advance in the comprehension of knowledge organization in humans. Also, this novel orientation places research on language networks in the general framework of traffic and navigation on complex networks. The hypothesis suggests that search and retrieval are affected by the way information flows, this issue has received much attention during the past years, see for instance [87,88].

5.2. Clustering and switching dynamics

Previous Section deals with a dynamic cognitive search process where subjects' production is independent of meaning, the task depends on the *form* of words, rather than their content. An alternative scenario might be that where subjects are demanded to produce words according to a certain category (for instance, "name any animal you can think of"). This approach has been studied in [89], under the theoretical framework of Troyer's model for optimal fluency [90], in which search and retrieval cognitive

processes exist on a double time-scale, a short one regarding local exploration (clustering), and a long one accounting for switch-transitions times.

The authors' proposal shares some aspects with the previous one. However, the issue here is not prominence or availability of words (centrality), but rather the fact that words are organized in communities or modules. Such modules are not only topological clusters, but also thematic groups or topics. From this point of view, the *switching and clustering* mechanism, understood as a double-level navigation process, can be used to predict human performance in such task as it is reported in [91]. The switcher-random-walker model (SRW) is then a cognitive inspired strategy that combines random-walking with switching for random exploration of networks. It is found that the number of steps needed to travel between a pair of nodes decreases when following this strategy, and thus the overall exploration abilities of a SRW within networks improves respect to mere random walkers.

Interestingly, a highly modular organization plus a two-level exploration scheme allows the system to organize information or to evolve without compromising exploration and retrieval efficiency. In this sense, semantic memory might be organizing information in a strongly modular or locally clustered way without compromising retrieval performance of concepts.

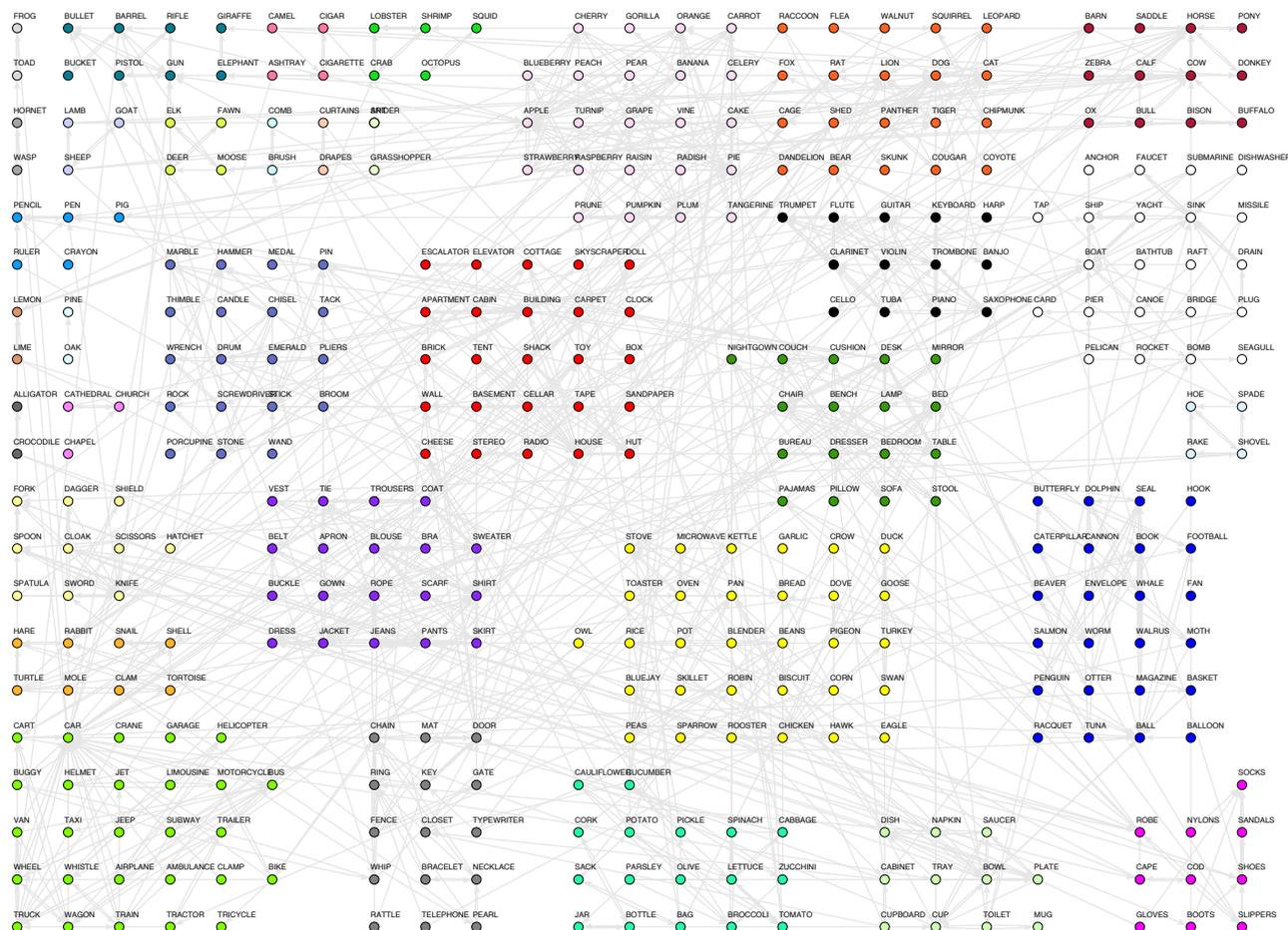
Community detection on empirical databases reveals the highly modular structure of word association. Analysis of USF-FA's mesoscale yields a modularity value $Q = 0.6162$, about 150 standard deviations above its randomized counterpart; similar results have been obtained with SFA ($Q = 0.7930$). See an example of detected modular structure for a subset of USF-FA data in Figure 11. The partition has been obtained for this review using a combination of algorithms (Extremal Optimization [52], Fast Algorithm [53] and Tabu Search [49]) available at [56]. These values seem a good starting point from which empirical work can be taken ahead.

5.3. Encoding semantic similarity

As it has been stated, free association data reflects many possible ways by which two words can be related (semantic similarity, causal or functional relationship, *etc.*); whereas feature production norms strictly inform about semantic similarity. The work reviewed here, after [92,93], explores whether it is possible to disentangle similarity relationships from general word association network (USF-FA) by the navigation of the semantic network. The authors construct upon these hypothesis and propose an algorithm that allows the disentanglement of a type of relationship embedded on the structure of the more general association network.

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

Figure 11. Community analysis for a subset of USF-FA with $N = 376$ nodes. The modularity value for this analysis is $Q = 0.8630$. The partition has been obtained for this review using a combination of algorithms (Extremal Optimization [52], Fast Algorithm [53] and Tabu Search [49]) available at [56]



Let us define the transition probability of the USF-FA network. The elements of USF-FA (a_{ij}) correspond to frequency of first association reported by the participants of the experiments. The data have to be normalized before having a transition probability matrix. We define the transition probability matrix P as:

$$P_{ij} = \frac{a_{ij}}{\sum_j a_{ij}} \tag{22}$$

Note that this matrix is asymmetric, as well as the original matrix USF-FA. This asymmetry property is maintained 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 :

$$T = \lim_{S \rightarrow \infty} \sum_{i=1}^S P^i \quad (23)$$

Finally, the matrix that will represent the feature similarity network (synthetic), where similarity is calculated as the cosine of the vectors in the new space, is given by the scalar product of the matrix and its transpose, $FS = TT^\dagger$.

The results obtained show macro-statistical coincidences (functional form of the distributions and descriptors) between the real (McRae's feature production norms [68]) and the synthetic obtained network, see Table 5. Moreover, the model yields also significant success at the microscopic level, *i.e.*, is able to reproduce to a large extent FP empirical relationships, see Table 6 as an example. These results support the general hypothesis about implicit entangled information in USF-FA, and also reveals a possible mechanism of navigation to recover feature information in semantic networks. Results are also compared with those obtained using the well known Latent Semantic Analysis (LSA; [94,95]) and Word Association Space (WAS; [96]), which are well-known in the psycholinguistic literature but not related to network methodology.

Table 5. 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

Assuming that Free Association semantic networks are good exposures of human semantic knowledge, the authors 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.

Table 6. Some illustrative examples of LSA, WAS and RIM's predictive capacity, when comparing closest neighbors to McRae's FP.

TUBA			
FP	LSA	WAS	RIM
trombone	clarinet	bath tub	trombone
trumpet	violin	faucet	saxophone
drum	flute	sink	trumpet
cello	guitar	bucket	flute
clarinet	trombone	bridge	clarinet
saxophone	fork	submarine	cello
flute	trumpet	drain	violin
harp	cake	raft	harp
banjo	drum	tap	banjo
piano	piano	dishwasher	stereo
ERROR	4.83	11	2.5
ROOSTER			
chicken	cat	chicken	chicken
goose	gate	crow	turkey
pigeon	donkey	skillet	crow
sparrow	barn	rice	robin
penguin	turnip	spinach	sparrow
pelican	owl	bowl	bluejay
bluejay	pig	beans	pigeon
dove	fence	robin	pelican
hawk	lion	tomato	goose
turkey	strawberry	sparrow	hawk
ERROR	11	8.5	2.87

6. Conclusions and perspectives

In this article we have reviewed some important work from the last decade on language as a networked system. Work in this area has been strongly motivated by the uprise of a prolific branch of statistical mechanics, complex networks. Its foundations have been outlined in Section 1, focusing on a number of *macro* and *micro* statistical properties of networks that have received particular attention, and on some tools to scrutinize the *meso* level.

Section 2 bridges methodological issues and the review of works specifically devoted to language. This Section elucidates the variety of sources and points of view from which language can be modeled as a network.

In Section 3 we have concentrated on the so-called language-oriented works. Inspired by empirical studies of real-world networks ranging from the Internet to citation networks, researchers have

approached language so as to propose models of networks that seek to explain either how networks come to have the observed structure, or what the expected effects of that structure will be. Such advances have brought to light two important facts: (i) that language resembles in many aspects other complex systems; and (ii) that different languages are also similar to each other regarding statistical descriptors. These results allow us to talk about the existence of certain universal trends that underlie linguistic structures. Within this Section, we have also seen some efforts to link language topology and linguistic activity in humans.

In the last part of this review (Sections 4 and 5) we have discussed work on the behavior of processes that take place on networks. This implies a shift from an interest in structures *per se* towards an interest in the mechanisms that operate on them. It also implies a greater transdisciplinary effort, aiming at a convergence with knowledge from cognitive science. We have paid attention to some topics of a cognitive-oriented complex network research, namely lexical development (network growth with a cognitive accent) and mental navigation (dynamical processes on language networks).

The progress in this field is so rapid, that we have failed to discuss and even cite a number of relevant results.

We believe that these results are only the tip of the iceberg. In looking forward to future developments in this area it is clear that there is much to be done. From a methodological point of view, the techniques for analyzing networks are at present no more than a collection of miscellaneous and largely unrelated tools. A systematic program for characterizing network structure is still missing.

On the linguistic side we are just in the first attempts at answering a few questions; this means that almost everything is yet to begin. Some topics that might be important in the future are: are there common mechanisms in the emergence of SF language network structures in artificial communities of agents [97–99] and language acquisition in children? How can be mental navigation so efficient on a network which displays many different types of links between words? Is it possible to construct a typology of languages where the genealogical relations are reflected in network features? How do semantic categories evolve? Can semantic memory's malfunctions (blocking, persistence, bias, *etc.*) be explained in terms of topological changes? How are language networks modified through aging and brain damage? If we can gain some understanding for these questions, it will give us new insight into complex and previously poorly understood phenomena.

Finally, in the long run questions will necessarily turn towards neuroscience: is it possible to find a mapping between neural and language networks? Complex neural topologies have already been spotted [100–103], which suggest that complex network methods might be adequate in this area. Furthermore, data indicate that there are several pathways connecting the language-relevant brain areas [104], which suggest a networked structure. Finally, there exists strong evidence about specific localization of semantic categories. Brain imaging studies have shown that different spatial patterns of neural activation are associated with thinking about different semantic categories of pictures and words (for example, tools, buildings, and animals) [105–109]. These works suggest that the lexico-semantic system's organization we observe at an abstract level, *i.e.*, semantically coherent modular structure (see Figure 11), may have a close correlate at the physical level. Although more fine grained resolution of fiber tracts and crossings is necessary and unavailable nowadays, we can envisage some future research issues: what is the direction of the information flow in the fiber tracts connecting language areas? Is

there a distinctive area where linguistic information is integrated? Is the modular structure detected in language networks mirrored at the neural level? These key questions open up a whole new and intriguing research scenario.

References

1. Miller, G.A.; Beckwith, R.; Fellbaum, C.; Gross, D.; Miller, K.J. Introduction to wordnet: an on-line lexical database. *Int. J. Lexico.* **1990**, *3*, 235–244.
2. Roget, P. *Roget's Thesaurus of English Words and Phrases*; TY Crowell co.: New York, NY, USA, 1911.
3. *WordNet: An electronic lexical database*; Fellbaum, C., Ed.; MIT press: Cambridge, MA, USA, 1998.
4. Milgram, S. The small world problem. *Psychol. Today* **1967**, *2*, 61–67.
5. Travers, J.; Milgram, S. An experimental study of the small world problem. *Sociometry* **1969**, *32*, 425–443.
6. Watts, D.; Strogatz, S. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440.
7. Ferrer i Cancho, R.; Solé, R.; Köhler, R. Patterns in syntactic dependency networks. *Phys. Rev. E* **2004**, *69*, 051915.
8. Ferrer i Cancho, R. Euclidean distance between syntactically linked words. *Phys. Rev. E* **2004**, *70*, 56135.
9. Solé, R.V.; Murtra, B.; Valverde, S.; Steels, L. Language Networks: their structure, function and evolution. *Trends Cogn. Sci.* **2006**, cplx.20305.
10. Ferrer i Cancho, R. Why do syntactic links not cross? *Europhys. Lett.* **2006**, *76*, 1228–1234.
11. Corominas-Murtra, B.; Valverde, S.; Solé, R.V. The ontogeny of scale-free syntax networks through language acquisition. *Adv. Compl. Syst.* **2009**, *12*, 371–392.
12. Corominas-Murtra, B.; Valverde, S.; Solé, R.V. Evolution of Communication and Language in Embodied Agents. *Evol. Commun. Lang. Embodied Agent.* **2010**, pp. 83–101.
13. Kello, C.; Beltz, B. Scale-free networks in phonological and orthographic wordform lexicons. In *Approaches to Phonological Complexity*; Chitoran, I., Coup, C., Marsico, E., Pellegrino, F., Eds.; Mouton De Gruyter: Berlin, Germany, 2009; p. 171.
14. Vitevitch, M. The clustering coefficient of phonological neighborhoods influences spoken word recognition. *J. Acoust. Soc. Am.* **2006**, *120*, 3252.
15. Erdős, P.; Rényi, A. On Random Graphs. *Publ. Math. (Debrecen)* **1959**, *6*, 290–297.
16. Barabási, A.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509.
17. Albert, R.; Barabási, A. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97.
18. Newman, M.E.J. The structure and function of complex networks. *SIAM Rev* **2003**, *45*, 167–256.
19. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D. Complex networks: structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308.
20. Chung, F.; Lu, L. *Complex Graphs and Networks*; American Mathematical Society: Providence, RI, USA, 2006.

21. Newman, M.; Barabási, A.; Watts, D. *The Structure and Dynamics of Networks*; Princeton University Press: Princeton, NJ, USA, 2006.
22. Caldarelli, G. *Scale-free Networks: Complex Webs in Nature and Technology*; Oxford University Press: New York, NY, USA, 2007.
23. Barrat, A.; Barthélemy, M.; Vespignani, A. *Dynamical Processes on Complex Networks*; Cambridge University Press: New York, NY, USA, 2008.
24. Dorogovtsev, S.; Goltsev, A.; Mendes, J. Critical phenomena in complex networks. *Rev. Mod. Phys.* **2008**, *80*, 1275–1335.
25. Cohen, R.; Havlin, S. *Complex Networks: Structure, Robustness and Function*; Cambridge University Press: New York, NY, USA, 2010.
26. Rozenfeld, H. *Structure and Properties of Complex Networks: Models, Dynamics, Applications*; VDM Verlag: Saarbrücken, Germany, 2010.
27. Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press: Cambridge, UK, 1994 .
28. Freeman, L. A set of measures of centrality based upon betweenness. *Sociometry* **1977**, *40*, 35–41.
29. Brandes, U. A Faster Algorithm for Betweenness Centrality. *J. Math. Soc.* **2001**, *25*, 163–177.
30. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **1972**, *2*, 113–120.
31. Adamic, L. The small world web. *Res. Adv. Tech. Digit. Libr.* **1999**, pp. 852–852.
32. Page, L.; Brin, S.; Motwani, R.; Winograd, T. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, Stanford, CA, USA, 1998.
33. Newman, M. Assortative mixing in networks. *Phys. Rev. Lett.* **2002**, *89*, 208701.
34. Newman, M. Mixing patterns in networks. *Phys. Rev. E* **2003**, *67*, 26126.
35. Solé, R.V.; Valverde, S. Chapter: Information Theory of Complex Networks: On Evolution and Architectural Constraints, In *Networks: Structure, Dynamics and Function; Lecture Notes in Physics*; Springer-Verlag: Berlin, Germany, 2004; pp.169–190.
36. Read, R.; Wilson, R. *An atlas of graphs*; Oxford University Press: Oxford, UK, 1998.
37. Erdős, P.; Rényi, A. On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **1960**, *5*, 17.
38. Erdős, P.; Rényi, A. On the Evolution of Random Graphs. *Bull. Inst. Int. Stat.* **1961**, *38*, 343.
39. Bollobás, B. *Random Graphs*, 2nd ed.; Academic Press: New York, NY, USA, 2001.
40. Rapoport, A. Spread of Information through a Population with Socio- Structural Bias I. Assumption of Transitivity. *Bull. Math. Biophys.* **1953**, *15*, 523–533.
41. Rapoport, A. Spread of Information through a Population with Socio- Structural Bias II. Various Models with Partial Transitivity. *Bull. Math. Biophys.* **1953**, *15*, 533–546.
42. Rapoport, A. Spread of Information through a Population with Socio- Structural Bias III. Suggested Experimental Procedures. *Bull. Math. Biophys.* **1954**, *16*, 75–81.
43. Rapoport, A.; Horvath, W. A Study of a Large Sociogram. *Behav. Sci.* **1961**, *6*, 79–91.
44. Watts, D.J. *Small Worlds: The Dynamics of Networks between Order and Randomness*; Princeton University: Princeton, NJ, USA, 1999.

45. Klemm, K.; Eguiluz, V. Highly clustered scale-free networks. *Phys. Rev. E* **2002**, *65*, 36123.
46. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7821–7826.
47. Guimerà, R.; Danon, L.; Díaz-Guilera, A.; Giralt, F.; Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **2003**, *68*, 065103.
48. Gleiser, P.; Danon, L. Community structure in jazz. *Adv. Compl. Syst.* **2003**, *6*, 565–573.
49. Arenas, A.; Fernandez, A.; Gomez, S. Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* **2008**, *10*, 053039.
50. Newman, M.E.J.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113.
51. Gómez, S.; Jensen, P.; Arenas, A. Analysis of community structure in networks of correlated data. *Phys. Rev. E* **2009**, *80*, 016114.
52. Duch, J.; Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E* **2005**, *72*, 027104.
53. Newman, M.E.J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **2004**, *69*, 321–330.
54. Arenas, A.; Duch, J.; Fernández, A.; Gómez, S. Size reduction of complex networks preserving modularity. *New J. Phys.* **2007**, *9*, 176.
55. Danon, L.; Duch, J.; Arenas, A.; Díaz-Guilera, A. Comparing community structure identification. *J. Stat. Mech.* **2005**, P09008.
56. Sergio Gómez homepage. Available online: <http://deim.urv.cat/~sgomez/radatools.php> (accessed on 15 April 2010).
57. Arenas, A.; Díaz-Guilera, A.; Pérez-Vicente, C. Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.* **2006**, *96*, 114102.
58. Arenas, A.; Fernández, A.; Gómez, S. A complex network approach to the determination of functional groups in the neural system of *C. Elegans*. *Lect. Notes Comput. Sci.* **2008**, *5151*, 9–18.
59. Mehler, A. Chapter: Large Text Networks as an Object of Corpus Linguistic Studies, *Corpus Linguistics. An International Handbook*; de Gruyter: Berlin, Germany/New York, NY, USA, 2007.
60. Sinclair, J. *Corpus, Concordance, Collocation*; Oxford University Press: Oxford, UK, 1991.
61. Stubbs, M. *Words and Phrases: Corpus Studies of Lexical Semantics*; Blackwell Pub.: Cambridge, MA, USA, 2001.
62. Stubbs, M. *Text and Corpus Analysis*; Blackwell Pub.: Cambridge, MA, USA, 1996.
63. Halliday, M. Lexis as a linguistic level. *In memory of JR Firth* **1966**, pp. 148–162.
64. Antequiera, L.; Nunes, M.; Oliveira, O.; Costa, L. Strong correlations between text quality and complex networks features. *Phys. A* **2006**, *373*, 811–820.
65. Ferrer i Cancho, R.; Solé, R. The small-world of human language. *Proc. R. Soc. Lond., B, Biol. Sci.* **2001**, *268*, 2261–2265.
66. Collins, P.; Peters, P. The Australian corpus project. In *Corpus linguistics, hard and soft*, Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora; Editions Rodopi: Atlanta, GA, USA, 1988; p. 103.

67. Sigman, M.; Cecchi, G. Global organization of the WordNet lexicon. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1742–1747.
68. McRae, K.; Cree, G.; Seidenberg, M.; McNorgan, C. Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Meth.* **2005**, *37*, 547–559.
69. Nelson, D.L.; McEvoy, C.L.; Schreiber, T.A. The University of South Florida word association, rhyme, and word fragment norms. *Behav. Res. Meth.* **1998**, *36*, 402–407.
70. Callejas, A.; Correa, A.; Lupiáñez, J.; Tudela, P. Normas asociativas intracategoriales para 612 palabras de seis categorías semánticas en español. *Psicológica* **2003**, *24*, 185–214.
71. Fernández, A.; Díez, E.; Alonso, M.; Beato, M. Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behav. Res. Meth.* **2004**, *36*, 577–583.
72. Melinger, A.; Weber, A. Database of Noun Associations for German. Available online: <http://www.coli.uni-saarland.de/projects/nag/> (accessed on 10 April 2010).
73. Ferrand, L.; Alario, F.X. Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année Psychol.* **1998**, *98*, 531–552.
74. Motter, A.E.; de Moura, A.P.S.; Lai, Y.C.; Dasgupta, P. Topology of the conceptual network of language. *Phys. Rev. E* **2002**, *65*, e065102.
75. <ftp://ibiblio.org/pub/docs/books/gutenberg/etext02/mthes10.zip> (accessed on 27 December 2009).
76. Dorogovtsev, S.; Mendes, J. Language as an evolving word web. *Proc. R. Soc. Lond., B, Biol. Sci.* **2001**, *268*, 2603–2606.
77. Snodgrass, J.; Vanderwart, M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn.* **1980**, *6*, 174–215.
78. Ferrer i Cancho, R. Zipf's law from a communicative phase transition. *Eur. Phys. J. B* **2005**, *47*, 449–457.
79. Ferrer i Cancho, R. When language breaks into pieces. A conflict between communication through isolated signals and language. *BioSystems* **2006**, *84*, 242–253.
80. Zipf, H. *Human behaviour and the Principle of Least Effort: An Introduction to Human Ecology*; Addison Wesley: Cambridge, MA, USA, 1949.
81. Shannon, C.; Weaver, W. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623.
82. Ferrer i Cancho, R.; Solé, R. Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 788.
83. Steyvers, M.; Tenenbaum, J.B. The largescale structure of semantic networks: statistical analyses and a model of semantic growth. *Cog. Sci.* **2005**, *29*, 41–78.
84. Collins, A.M.; Quillian, M.R. Retrieval time from semantic memory. *J. Verbal Learn. Verbal Beh.* **1969**, *8*, 240–247.
85. Collins, A.; Loftus, E. A spreading activation theory of semantic memory. *Psychol. Rev.* **1975**, *82*, 407–428.
86. Griffiths, T.; Steyvers, M.; Firl, A. Google and the mind: Predicting fluency with PageRank. *Psychol. Sci.* **2007**, *18*, 1069–1076.

87. Arenas, A.; Cabrales, A.; Danon, L.; Díaz-Guilera, A.; Guimerà, R.; Vega-Redondo, F. Optimal information transmission in organizations: search and congestion. *Rev. Econ. Des.* **2010**, *14*, 75–93.
88. Danon, L.; Arenas, A.; Díaz-Guilera, A. Impact of community structure on information transfer. *Phys. Rev. E* **2008**, *77*, 36103.
89. Goñi, J.; Martincorena, I.; Corominas-Murtra, B.; Arrondo, G.; Ardanza-Trevijano, S.; Villoslada, P. Switcher-random-walks: a cognitive inspired mechanism for network exploration. *Arxiv* **2009**, 0903.4132.
90. Troyer, A.; Moscovitch, M.; Winocur, G. Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology* **1997**, *11*, 138–146.
91. Bousfield, W.; Sedgewick, C. An analysis of sequences of restricted associative responses. *J. Gen. Psychol.* **1944**, *30*, 149–165.
92. Borge-Holthoef, J.; Arenas, A. Navigating word association norms to extract semantic information. In Proceedings of the 31st Annual Conference of the Cognitive Science Society, Amsterdam, The Netherlands, 2009.
93. Borge-Holthoef, J.; Arenas, A. Categorizing words through semantic memory navigation. *Eur. Phys. J. B* **2010**, *74*, 265–270.
94. Landauer, T.; Dumais, S. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **1997**, *104*, 211–240.
95. Landauer, T.; McNamara, D.; Dennis, S.; Kintsch, W. *Handbook of Latent Semantic Analysis*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2007.
96. Steyvers, M.; Shiffrin, R.M.; Nelson, D.L. Word association spaces for predicting semantic similarity effects in episodic memory. In *Experimental cognitive psychology and its applications*; Healy, A.F., Ed.; American Psychological Association: Washington, DC, USA, 2004; pp. 237–249.
97. Steels, L. Language as a complex adaptive system. *Lect. Notes Comput. Sci.* **2000**, *1917*, 17–28.
98. Steels, L. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Conn. Sci.* **2005**, *17*, 213–230.
99. Baronchelli, A.; Felici, M.; Loreto, V.; Caglioti, E.; Steels, L. Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech.* **2006**, *6*, P06014.
100. Sporns, O.; Kötter, R. Motifs in brain networks. *PLoS Biol.* **2004**, *2*, e369.
101. Sporns, O.; Chialvo, D.; Kaiser, M.; Hilgetag, C. Organization, development and function of complex brain networks. *Trends Cogn. Sci.* **2004**, *8*, 418–425.
102. Zhou, C.; Zemanová, L.; Zamora-López, G.; Hilgetag, C.; Kurths, J. Hierarchical Organization Unveiled by Functional Connectivity in Complex Brain Networks. *Phys. Rev. Lett.* **2006**, *97*, 238103.
103. Zhou, C.; Zemanová, L.; Zamora-López, G., G.; Hilgetag, C.; Kurths, J. Structure–function relationship in complex brain networks expressed by hierarchical synchronization. *New J. Phys.* **2007**, *9*, 178.
104. Friederici, A. Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* **2009**, *13*, 175–181.

105. Hillis, A.; Caramazza, A. Category-specific naming and comprehension impairment: A double dissociation. *Brain* **1991**, *114*, 2081–2094.
106. Martin, A.; Wiggs, C.; Ungerleider, L.; Haxby, J. Neural correlates of category-specific knowledge. *Nature* **1996**, *379*, 649–652.
107. Polyn, S.; Natu, V.; Cohen, J.; Norman, K. Category-specific cortical activity precedes retrieval during memory search. *Science* **2005**, *310*, 1963.
108. Mitchell, T.; Shinkareva, S.; Carlson, A.; Chang, K.; Malave, V.; Mason, R.; Just, M. Predicting human brain activity associated with the meanings of nouns. *Science* **2008**, *320*, 1191.
109. Binder, J.; Desai, R.; Graves, W.; Conant, L. Where is the semantic system? *Cereb Cortex* **2009**, *120*, 2767–2796.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license <http://creativecommons.org/licenses/by/3.0/>.