Review

Feature Review

Networks in Cognitive Science

Andrea Baronchelli[1], Ramon Ferrer-i-Cancho[2], Romualdo Pastor-Satorras[3], Nick Chater[4], and Morten H. Christiansen[5,6]

[1] Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA 02115, USA
[2] Complexity and Quantitative Linguistics Lab, TALP Research Center, Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Campus Nord, Edifici Omega, E-08034 Barcelona, Spain
[3] Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, E-08034 Barcelona, Spain
[4] Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK
[5] Department of Psychology, Cornell University, Uris Hall Ithaca, NY 14853, USA
[6] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Networks of interconnected nodes have long played a key role in Cognitive Science, from artificial neural networks to spreading activation models of semantic memory. Recently, however, a new Network Science has been developed, providing insights into the emergence of global, system-scale properties in contexts as diverse as the Internet, metabolic reactions, and collaborations among scientists. Today, the inclusion of network theory into Cognitive Sciences, and the expansion of complex-systems science, promises to significantly change the way in which the organization and dynamics of cognitive and behavioral processes are understood. In this paper, we review recent contributions of network theory at different levels and domains within the Cognitive Sciences.

Introduction

Humans have more than 10^{10} neurons and between 10^{14} and 10^{15} synapses in their nervous system [1]. Together, neurons and synapses form neural networks, organized into structural and functional subnetworks at many scales [2]. However, understanding the collective behavior of neural networks starting from the knowledge of their constituents is infeasible. This is a common feature of all complex systems, summarized in the famous motto 'more is different' [3]. The study of complexity has yielded important insights into the behavior of complex systems over the past decades, but most of the toy models that proliferated under its umbrella have failed to find practical applications [4]. However, in the past decade or so a revolution has taken place. An unprecedented amount of data, available thanks to technological advances, including the Internet and the Web, has transformed the field. The data-driven modeling of complex systems has led to what is now known as Network Science [5].

Network Science has managed to provide a unifying framework to put different systems under the same conceptual lens [5], with important practical consequences [6]. The resulting formal approach has uncovered widespread properties of complex networks and led to new experiments [4,7,8]. The potential impact on Cognitive Science is considerable. The newly available concepts and tools have already provided insights into the collective behavior of neurons [9], but they have also inspired new empirical work designed, for example, to identify large-scale functional networks [10,11]. Moreover, very different systems such as semantic networks [12], language networks [13], and social networks [14,15] can now be investigated quantitatively using the unified framework of Network Science.

These developments suggest that concepts and tools from Network Science will become increasingly relevant to the study of cognition. Here, we review recent results showing how a network approach can provide insights into Cognitive Science, introduce Network Science to the interested cognitive scientist without prior experience of the subject, and give pointers to further readings. After a gentle overview of complex networks, we survey existing work in three subsections concerning the neural, cognitive, and social levels of analysis. A final section considers dynamical processes taking place on networks, which is likely to be an important topic for Cognitive Science in the future.

Introduction to Network Science

The study of networks (or graphs) is a classical topic in mathematics whose history began in the 17th century [16]. In formal terms, networks are objects comprising a set of points, called vertices or nodes, joined in pairs by lines, termed edges (see Figure 1 for basic network definitions). They provide a simple and powerful representation of complex systems comprising interacting units, with nodes representing the units and edges denoting pairwise interactions between units. Mathematical graph theory [17], based mainly on the rigorous demonstration of the topological properties of particular graphs or in general extremal properties, has been dramatically expanded by the recent availability of large digital databases, which have allowed exploration of the properties of very large real networks. This work, mainly conducted within the statistical physics community, has led to the discovery that many natural and artificial systems can be usefully described in terms of networks [8]. The new Network Science

Corresponding author: Baronchelli, A. (a.baronchelli.work@gmail.com).

1364-6613/$ – see front matter
 http://dx.doi.org/10.1016/j.tics.2013.04.010

Glossary

Assortativity: the preference for nodes to attach to other vertices that are similar in some way. Assortative mixing by degree, for example, describes the case in which nodes of similar degree tend to form connections preferentially among themselves. For instance, in most social networks, high-degree nodes are connected to high-degree nodes and poorly connected vertices tend to link each other. By contrast, disassortativity describes the opposite tendency (e.g., high-degree nodes tend to be connected to low-degree nodes in technological networks).
Betweenness of a node: the number of shortest paths between pairs of vertices passing through a given vertex.
Centrality: the centrality of a node measures its relative importance inside the network; for example. in terms of degree, betweenness, or distance. In the latter case, the so-called closeness centrality is defined as the inverse of the sum of the shortest path lengths from the considered vertex to all other vertices in the network.
Clique: a subset of a network where every pair of nodes is connected.
Clustering coefficient: c_i of vertex i is defined as the ratio between e_i, the actual number of edges between its nearest neighbors, and the maximum possible number $k_i(k_i - 1)/2$; that is,

$$c_i = \frac{2e_i}{k_i(k_i - 1)}. \quad \text{[I]}$$

The clustering coefficient quantifies the transitivity of a network, measuring the probability that two vertices with a common neighbor are also neighbors of each other. The average clustering coefficient $\langle c \rangle$ is the average value of c_i over all vertices in a network; that is,

$$\langle c \rangle = \frac{1}{N}\sum_i c_i \quad \text{[II]}$$

In real networks, $\langle c \rangle$ usually takes values of order unity, in stark contrast with the clustering inversely proportional to network size that is expected for a random network (Box 2).
Community: although the precise definition of community remains an open question, a minimal and generally accepted description is that the subset of the nodes in a community is more tightly connected to one another than to the rest of the network.
Connected component: the maximal subset of vertices in a network such that there is a path joining any pair of vertices in it.
Connectome: the detailed 'wiring diagram' of the neurons and synapses in the brain.
Co-occurrence network: a network with nodes representing the elements present in a given context (e.g., words in a text) and edges representing the 'co-occurrence' in the same context according to some criterion; for example, in the case of words in a text, a simple criterion is that they appear in the text one next to the other.
Core of a network: a powerful subset of the network because of the high frequency of occurrence of its nodes [54], their importance for the existence of the remainder of nodes [53], or the fact that it is both densely connected and central (in terms of graph distance) [55].
Degree: the degree of a vertex, k_i, is defined as the number of other vertices to which vertex i is connected (or the 'number of neighbors' [65]).
Degree distribution: the probability $P(k)$ that a randomly chosen vertex has degree k for every possible k. For large networks, the degree distribution represents a convenient statistical characterization of a network's topology.
Diameter: the longest of the shortest paths between any pair of vertices in a network.
Directed network: a network in which each link has an associated direction of flow.
Hubs: the vertices in a network with the largest degree (number of connections).
Network or graph: a collection of points, called vertices (or nodes), joined by lines, referred as edges (or links). Vertices represent the elementary components of a system, whereas edges stand for the interactions or connections between pairs of components.
PageRank: a network analysis algorithm that assigns a numerical weight to each edge of a directed network, aimed at measuring its relative importance. The algorithm is used by the Google search engine to rank Word Wide Web search results.
Percolation threshold: percolation theory describes the behavior of connected clusters in a graph. A network is said to percolate when its largest connected component contains a finite fraction of the nodes that form the whole network. Percolation depends in general on some topological quantity (e.g., the average degree in the Erdös–Rényi random graph). The percolation threshold is the value of this quantity above which the network percolates.
Rich-club phenomenon: property observed in many real networks in which the hubs have a strong tendency to be connected to each other rather than with vertices of small degree.
Scale-free networks: networks with a broad, heavy-tailed degree distribution that can often be approximated by a power-law, $P(k) \sim k^{-\gamma}$, where γ is a characteristic exponent usually between 2 and 3. This heavy-tailed power-law form underlies many of the surprising features shown by real complex networks.
Shortest path length: the shortest path length, or distance, ℓ_{ij}, between vertices i and j is the length (in number of edges) of the shortest path joining i and j. The shortest path length thus represents a measure of the distance between pairs of vertices. The average shortest path length $\langle \ell \rangle$ is the average of the shortest path length over all pairs of vertices in the network; that is,

$$\langle \ell \rangle = \frac{2}{N(N-1)}\sum_{i<j} \ell_{ij}, \quad \text{[III]}$$

where N is the total number of vertices in the network.
Small-world property: a property shown by many real complex networks that exhibit a small value of the average shortest path length $\langle \ell \rangle$, increasing with network size logarithmically or slower. This property is in stark contrast to the larger diameter of regular lattices, which grows algebraically with lattice size.
Strength of a node: the sum of the weights of the edges incident on a vertex.
Transitivity of a network : the propensity of two nodes in a network to be connected by an edge if they share a common neighbor.
Tree: a network that has as many edges as vertices minus one and is connected; that is, a walk from one node can reach any other node in the network.
Weighted network: a network whose links are characterized by different capacities, or weights, defining the strength of the interaction between the nodes they connect.
Word-association network: a network where vertices are words and a link connects a cue word with the word that is produced as response.

has been successfully applied in fields ranging from computer science and biology to social sciences and finance, describing systems as diverse as the World Wide Web, patterns of social interaction and collaboration, ecosystems, and metabolic processes (see [7] for a review of empirical results).

Interest in real complex networks has been boosted by three empirical observations. The first is the so-called small-world effect, first observed experimentally by the social psychologist Stanley Milgram [18], which implies that there is a surprisingly small shortest path length, measured in traversed connections in direct paths, between any two vertices in most natural networks. In Milgram's experiment, a set of randomly chosen people in Omaha, Nebraska, were asked to navigate their network of social acquaintances to find a designated target, a person living in Boston, Massachusetts. The navigation should be performed by sending a letter to someone the recipients knew on a first-name basis who they thought should be closer to the target and asking them to do the same until the target was reached. The average number of people that the letters passed through before reaching the target led to the popular aphorism 'six degrees of separation'. Although the number six is not universal, the average distance between pairs of vertices in real networks is typically very small in relation to network size.

The second observation concerns the high transitivity of many real networks. The concept of transitivity is borrowed from usage in the social sciences [19] and refers to the fact that, for example, two friends of any given individual are themselves also likely to be friends. Transitivity can be quantitatively measured by means of the clustering coefficient [20], which takes large values in almost all real networks.

Third, the connectivity structure of many real systems is strongly heterogeneous, with a skewed distribution in the number of edges attached to each vertex (the so-called degree distribution) (Figures 2 and 3). This kind of network has been dubbed scale free [21]. The scale-free hallmark underlies many of the most surprising properties of complex

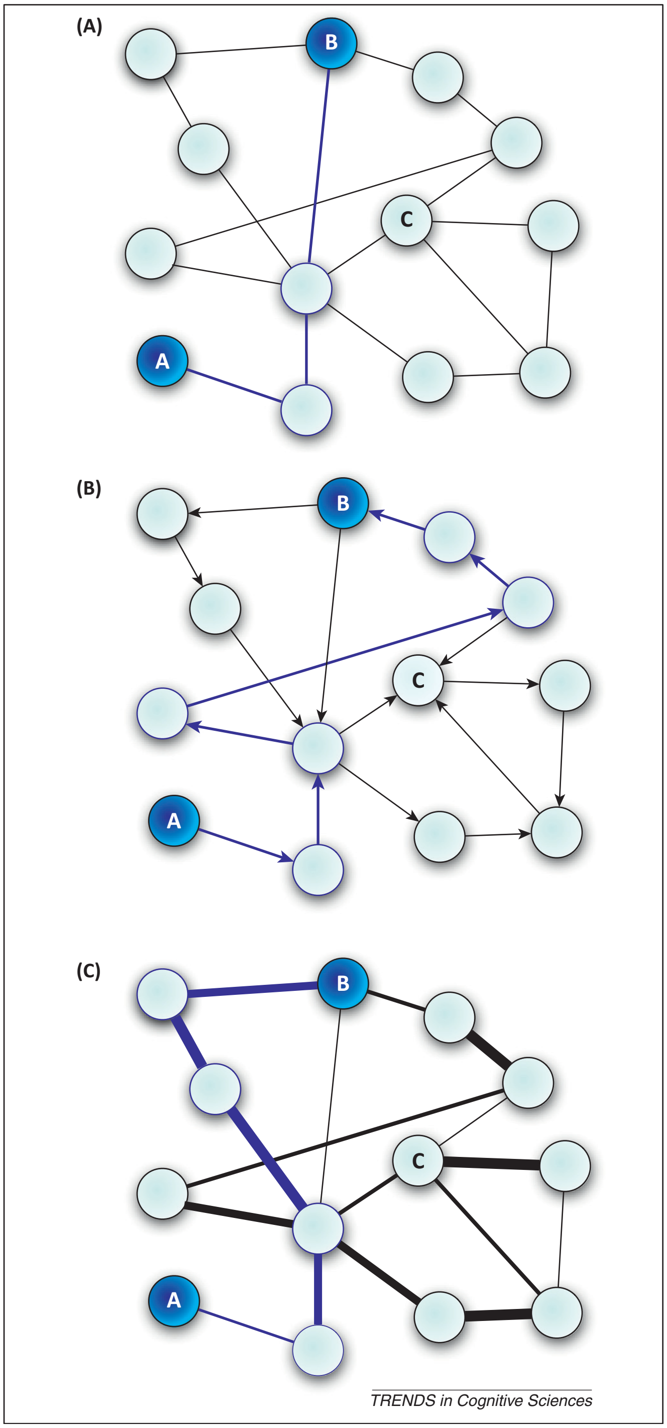

Figure 1. Basic network properties. (**A**) Circles represent vertices; unbroken lines connecting pairs of vertices correspond to edges. The degree k of a vertex is given by the number of its neighbors; that is, the number of other vertices to which it is connected by edges. For example, node C has degree $k = 4$. The distance (shortest path length) ℓ between two nodes is given by the minimum number of edges that connect them in a continuous path. For example, nodes A and B are at distance ℓ=3. (**B**) In a directed network, vertices are unidirectional, indicating that the flow of information can proceed in only one direction between adjacent nodes. The distance between nodes A and B is now ℓ=6, node C has in-degree $k_{in} = 3$ and out-degree $k_{out} = 1$, meaning that it can receive information from three nodes and pass it to just one neighbor. (**C**) In a weighted network, links have different capacities, or weights, indicating the amount of information they can carry. Many definitions of distance can be adopted. The path between nodes A and B highlighted in blue is obtained by following the maximum weight link at each step. Beyond its degree, a node is characterized also by its strength; that is, the sum of the weights of the links that connect it to the rest of the network.

networks, such as their extreme resilience to random deletion of vertices coupled with extreme sensitivity to the targeted deletion of the most connected vertices [22], and strongly impacts processes such as the propagation of diseases [23].

Applications of network theory in Cognitive Science

The brain and neural networks

The network framework provides a natural way to describe neural organization [24]. Indeed, cognition emerges from the activity of neural networks that carry information from one cell assembly or brain region to another (Box 1). The advent of Network Science suggests modifying the traditional 'computer metaphor' for the brain [25] to an 'Internet metaphor', where the neocortex takes on the task of 'packet switching' [26]. More broadly, network theory allows the shift from a reductionist to a 'complex-system' view of brain organization [2,9,10,27]. In this framework, optimal brain functioning requires a balance between local processing and global integration [28,29]. In particular, clustering

Box 1. Computing with networks

The central tenet of Cognitive Science is that thought is computation and hence that the enormously rich network of neurons that composes the human brain is a computational device. Thus, a central intellectual challenge for Cognitive Science is to understand how networks of simple neuron-like units can conduct the spectacularly rich range of computations that underlie human thought, language, and behavior. Connectionism, or parallel distributed processing (see [150] for the historical pedigree), uses networks comprising simplified neural processing units where adjustments of the connections between units allow the models to learn from experience. This approach has been applied to many aspects of cognition from cognitive development [151] to language [152], including connectionist implementations [153] of symbolic semantic networks [114]. In parallel, an active tradition has aimed to provide computational models of actual neural circuitry; such models are more biologically realistic, but typically focus less on abstract cognitive tasks and more on elementary processes of learning, early visual processing, and motor control [154].

Since the 1980s, there has been increasing interest in a related, but distinct, research program using networks to represent, make inferences over, and learn complex probability distributions [155]. In such probabilistic graphical models, nodes correspond to elementary states of affairs and links encode probabilistic relationships, or even causal connections [156], between states of affairs. These models have proved to be powerful tools for artificial intelligence and machine learning, as well as the basis for many models in Bayesian Cognitive Science (e.g., [157]). Crucially, inference and learning in such models typically requires no 'supervision' – nodes modify their level of activity in response to activity on incoming links, while the strength of a link is modified in response to signals at the nodes that it connects.

In both connectionist networks and probabilistic graphical models, the network itself autonomously carries out inference and learning. However, the possible relationship between biological neural networks and these classes of psychological network model is less well understood. One suggestion is that neuromodulation, such as long-term potentiation (activity-dependent synaptic strengthening), corresponds to strengthening a 'connection' in a computational network; more concretely, the detection of 'prediction error' (crucial in many network-learning models) relates to activity of the dopamine system [158]. Moreover, populations of neurons, and network operations over these, may implement probabilistic calculations (e.g., [159]). Nonetheless, understanding how networks can compute remains a central challenge for the cognitive and brain sciences.

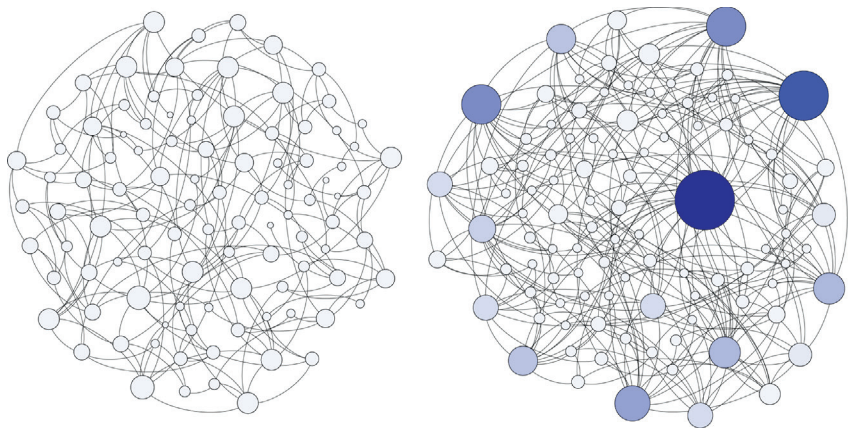

Figure 2. Graphical representation of homogeneous and scale-free networks. In homogeneous networks (left), nodes have similar topological properties, which are well captured by their average values. In heterogeneous networks (right), very different nodes coexist, including some so-called hubs (i.e., extremely well-connected nodes). In both cases, the degree of each node is visually stressed by color and size. The left panel depicts an Erdös–Rényi random graph, the right a Barabasi–Albert graph, both containing $N = 100$ nodes and the same average degree $\langle k \rangle = 2.5$.

facilitates local processing, whereas a short path length (a low degree of separation) across the neural network is required for global integration of information among brain regions. Indeed, these two factors may shape neural network structure and performance [30,31].

The map of brain connectivity, the so-called connectome (see Glossary) and its network properties are crucial for understanding the link between brain and mind [29]. The connectome is characterized by short path lengths (a small-word topology), high clustering, and assortativity, the tendency of hubs to be connected to hubs, forming a so-called 'rich club' and an overlapping community structure [32–35]. The latter observation challenges earlier reductionist views of the brain as a highly modular structure (e.g., [36]).

Alterations of fundamental network properties are often associated with pathologies [28,37–39]. For instance, smaller clustering, larger path length, and greater modularity are found in autistic spectrum disorder [38]. Similarly, the multimodal cortical network has a shorter path length and a trend to increased assortativity in people with schizophrenia [37]. It is unclear whether Alzheimer's disease has a unique signature at the brain network level, but in different studies path lengths and clustering have been found to be altered, both above and below controls [28].

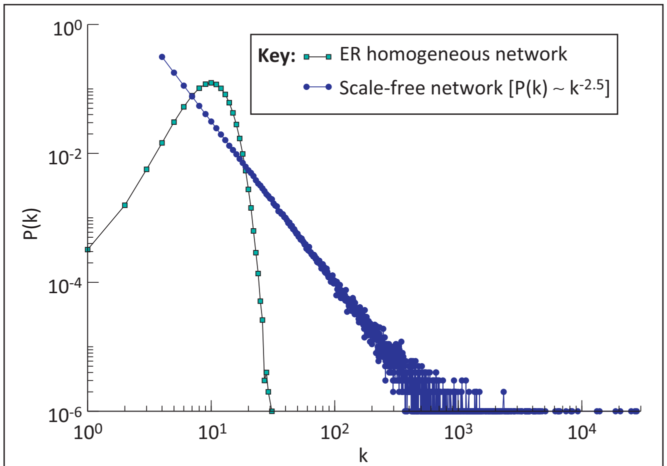

Figure 3. Degree distribution. The degree distribution of a network, $P(k)$, tells us the probability that a randomly chosen node will have degree k. Here, the degree distribution of an Erdös–Rényi (ER) graph is plotted next to one of a scale-free network (with $P(k) \sim k^{-2.5}$). It is clear that, whereas in the ER graph the probability of observing a node with degree $k > 30$ is practically zero, in the scale-free graph there is a finite, if small, probability of observing hubs connected to thousands or even tens of thousands of nodes. Both graphs have the size $N = 10^6$ nodes and the same average degree $\langle k \rangle = 10.5$.

Intriguingly, Network Science may provide the tools to describe different kinds of brain networks in a coherent fashion and to compare their properties even across different scales. Particularly remarkable is the identification of large-scale brain networks, defined according to structural connectivity or functional interdependence [10,27]. The network approach has also been a driving force in the analysis of functional networks in neuroimaging data [2]. For example, functional MRI (fMRI) techniques, an indirect measure of local neuronal activity [40], have shown dynamic reconfiguration of the modular organization of a large-scale functional network during learning [41]. Moreover, various pathologies have been related to alterations of the properties of large-scale networks [10]. Different neurodegenerative diseases have been connected with the degradation of different large-scale functional networks [42] and age-related changes in face perception have been linked to the degeneration of long-range axonal fibers [43].

Cognitive processes

At the level of cognition (i.e., the information-processing operations in the brain), a wide range of networks has been considered [44–49] (see http://www.lsi.upc.edu/~rferrericancho/linguistic_and_cognitive_networks.html). One of the most studied examples is networks of free word associations, which are in general weighted and directed, with weights reflecting the frequency of a given association [12,50]. Short path lengths, high clustering, and assortativity have been reported across datasets [44,51]. High clustering and short path lengths have been attributed to a network dynamics combining 'duplication' and 'rewiring' (Box 2).

A key theoretical question is whether the properties of networks at the level of information processing are inherited from the brain network substrate or instead arise from independent converging processes [10]. Cognitive impairment was found to be associated with a decrease of path lengths and an increase of clustering in word-fluency networks in Alzheimer's patients [46], whereas the opposite trend (increased path lengths and decreased clustering) was found in associative networks of late talkers [52]. Understanding the relationship, if any, between these alterations at the cognitive and neural levels is a challenge for future research.

Network Science has also shown how to single out the most important elements of a complex system. The simplest approach focuses on the concept of 'degree'; 'hubs' are highly connected nodes whose removal causes greater impact than low-degree nodes [22]. The internal organization of cognitive networks has been analyzed also at a larger scale, identifying the network's 'core' [53–55] and dividing ensembles of nodes into 'communities' that map into semantic [56,57] or syntactic [45] categories. It has been hypothesized that the lexicon may contain a basic vocabulary from which the meaning of the remaining words can be covered via circumlocution [58,59]. This hypothesis has been supported by the analysis of language networks [13] of word co-occurrence in many languages [60,61] and Web search queries [62], where the degree distribution shows a power law with two regimes, one containing essential vocabulary and the other containing specialized terms. The two regimes may emerge naturally from a type of preferential-attachment dynamics [54] (Box 2). Similarly, a network analysis of cross-referencing between dictionary entries has shown that dictionaries have a so-called grounding kernel, a subset of a dictionary comprising about 10% of words (typically with a concrete meaning and acquired early) from which other words can be defined [53].

As far as semantics is concerned [12], in word-association networks, names of musical instruments or color terms form strongly interconnected subsets of words; that is, communities of nodes [56,57]. Similarly, parts of speech (e.g., verbs and nouns) cluster together in a syntactic-dependency network [45]. This organization may help explain why brain damage can affect particular semantic fields [63] or specific parts of speech [64].

Network theory offers many new perspectives for understanding cognitive complexity. The ease with which a word is recognized depends on its degree or clustering coefficient [65–67]. Network theory has also helped to quantify the cognitive complexity of navigating labyrinths, whose structure, including the distance between relevant points, can be coded as a weighted network, distinguishing purely aesthetic labyrinths from those that were designed to have a complex solution [68]. The time needed to find the way out of a labyrinth is strongly correlated with that needed by a random walker (Box 3) to reach the exit (absorption time), which is in turn strongly correlated with the various network metrics including vertex strength and betweenness [68]. An interesting possible research direction is to investigate whether similar analysis applies to search problems in more abstract cognitive contexts, such as problem solving or reasoning.

The study of sequential processing has also been impacted by Network Science. For example, the length of a dependency between two elements of a sequence provides a measure of the cognitive cost of that relationship [69]. Thus, the mean of such lengths may measure the cognitive cost of processing a sequence such as a sentence [70,71]. The minimum linear arrangement problem is to determine the ordering of elements of the sequence that minimizes such a sum of lengths, given a network defining the dependencies between elements (Box 4, Figure 4) [70,72,73]. The rather low frequency of dependency crossings in natural language (Figure 4C,D) and related properties could be a side effect of dependency length minimization [73–75], suggesting that crossings and dependency lengths cannot be treated as independent properties, as is customary in Cognitive Sciences [70,76]. These findings suggest that a universal grammar is not needed to explain the origins of some important properties of syntactic-dependency structures; the limited capacity of the human brain may severely constrain the space of possible grammars. The network approach additionally allows for a reappraisal of existing empirical evidence. For example, the second moment of the degree distribution, $\langle k^2 \rangle$, is positively correlated with the

Box 2. Network models

The Erdös–Rényi random graph model (see Figures 2 and 3 in main text) has been the paradigm of network generation for a long time. It considers N isolated nodes connected at random, in which every link is established with an independent connection probability p [131]. The result is a graph with a binomial degree distribution that is centered at the average degree and has little clustering. The availability of large-scale network data made clear that different models were needed to explain the newly observed properties, in particular a large clustering coefficient and a power-law degree distribution [8]. The Watts–Strogatz model is one attempt to reconcile the high clustering characteristic of ordered lattices and small shortest paths lengths observed in complex networks [20]. In this model, in an initially ordered lattice, some edges are randomly rewired. For a small rewiring probability, clustering is preserved, whereas the introduction of a few shortcuts greatly reduces the network diameter. The Barabási–Albert model (see Figure 2 in main text) represents a first explanation of the power-law degree distributions found in many complex networks (see Figure 3 in main text) [21]. It is based on the principle of growth and preferential attachment. At each time step, a new node enters the network and connects to old nodes proportionally to their degree; therefore, 'richer nodes' (nodes with higher degree) 'get richer'. This rule leads to a degree distribution scaling as $P(k) \sim k^{-3}$. Exponents other than three can be found by, for example, allowing for edge rewiring [22]. Other growth models displaying power-law degree distributions have been considered, involving mechanisms such as duplicating a node and its connections, with some edge rewiring [78,132], or random growth by adding triangles to randomly chosen edges [160]. Non-growing alternatives to the origin of a scale-free topology have applied optimization mechanisms, seeking an explanation in terms of trade-offs, optimizing the conflicting objectives pursued in the set up of the network. Such models, elaborating on the highly optimized tolerance framework [161], find examples in the class of heuristically optimized trade-off (HOT) network models [162]. Other approaches, such as the class of models with 'hidden variables' [163], represent a generalization of the classical random graph in which the connection probability depends on some non-topological (hidden) variable attached to each node. The proper combination of connection probability and hidden variable distribution can lead to a scale-free topology without reference either to growth or to preferential attachment [164].

Box 3. Dynamical processes on networks

Processes taking place on networks are widespread across a large number of domains, from epidemics spreading through the airplane transportation network to gossip spreading through networks of acquaintances [101]. In all cases, the topological properties of the underlying networks play a crucial role in the behavior of the process and extremely simple models can provide vital insights into large classes of apparently distant phenomena. This is why the study of processes occurring on networks has recently gained much attention in Cognitive Science. In the main text, we describe how the structure of the social network affects the spreading of a linguistic innovation [15], while random walk processes have been used in different contexts, from word-association experiments [121] to language modeling [115].

The random walk is an ideal example to understand the insights that studying an apparently trivial process can provide. At each time step, a particle (the walker) hops from the node it occupies to a randomly selected neighboring node. The properties of such simple dynamics are enlightening in many respects. For example, the so-called occupation probability ρ_i of the walker (i.e., the asymptotic probability of finding it on node i) is simply proportional to the degree k_i of that node (i.e., $\rho_i \sim k_i$ in a connected network [165]. This node degree is also crucial in many more complex situations [8]. Other important properties particularly relevant to the issues of searching and spreading in networks are mean first-passage time (MFPT) and coverage [166].

- The MFPT τ_i of a node i is the average time taken by the random walker to arrive for the first time at vertex i, starting from a random source. This corresponds to the average number of messages that have to be exchanged among the nodes to identify the location of vertex i. Interestingly, in typical cases, this time is proportional to the inverse of the occupation probability.
- The coverage $C(t)$ is defined as the number of different vertices that have been visited by the walker at time t, averaged for different random walks starting from different sources. The coverage can thus be interpreted as the searching efficiency of the network, measuring the number of different individuals that can be reached from an arbitrary origin in a given number of time steps.

Box 4. The minimum linear arrangement

The minimum linear arrangement problem comprises finding a sequential ordering of the vertices of a network that minimizes the sum of edge lengths [72]. If $\pi(v)$ is the position of vertex v and $u{\sim}v$ indicates that that vertices u and v are connected, the length of the edge $u{\sim}v$ is the absolute value of the difference of their positions; that is $|\pi(v) - \pi(u)|$. The sum of edge lengths is

$$D = \sum_{u \sim v} |\pi(v) - \pi(u)|. \quad \text{[IV]}$$

In a tree of n vertices, the mean distance between edges is $\langle d \rangle = D/(2(n-1))$. Imagine that a tree has only three vertices that are labeled with the numbers 1, 2, and 3. There are only $3! = 6$ possible linear arrangements of the vertices (see Figure 4A in main text), but the minimum $\langle d \rangle$ (or equivalently the minimum D) is achieved by only two orderings, (2, 1, 3) and its reverse (3, 1, 2), with $\langle d \rangle = 1$ (see Figure 4A in main text). We say that these two orderings are minimum linear arrangements. $\langle d \rangle = 1.5$ for the remainder of the orderings.

In a star tree, where all vertices have degree one except one (i.e., the hub) (see Figure 4B in main text), D is determined by the position of the hub in the sequence. For that tree, the optimal placement of the hub is at the center of the sequence [77].

The ordering of the words in the sentence given in Figure 4C in main text, which yields $\langle d \rangle = 11/8 \approx 1.375$, is also a minimum linear arrangement; that is, none of the $9! = 362\,880$ permutations of the words of the sentence is able to achieve a smaller $\langle d \rangle$ given the syntactic dependency tree of the sentence. Finding the minimum linear arrangement problem of a network is very difficult computational problem [72], but if the network is a tree (e.g., see Figure 4C in main text), computationally efficient solutions exist [167,168].

$\langle d \rangle$ would grow linearly ($\langle d \rangle = (n+1)/3$) with the number of vertices if the vertices were ordered at random [71]. By contrast, $\langle d \rangle$ grows sublinearly as a function of the number of vertices in real syntactic dependency trees [71].

$\langle k^2 \rangle$, the degree second moment. determines the minimum value that $\langle d \rangle$ could achieve [77].

$$\langle d \rangle \geq \frac{n\langle k^2 \rangle}{8(n-1)} + \frac{1}{2}. \quad \text{[V]}$$

The worst case is a star tree (see Figure 4B in main text) with the maximum $\langle k^2 \rangle$ [77]. Therefore, the tendency to have 'hubs' (i.e., a high degree variance in degrees of different vertices) and a low $\langle d \rangle$ are incompatible.

minimum sum of dependency lengths (Box 5) and therefore sufficiently long sentences cannot have hubs [77]. Although the minimum linear arrangement problem has so far been investigated mostly in language, it applies whenever a dependency structure over elements of a sequence is defined by a network. A promising avenue for future research is to extend network analysis to sequences of non-linguistic behavior, such as music, dance, and action sequencing.

Various studies address the origin of the properties of cognitive networks (Box 2). For example, the double power-law degree distribution observed in word co-occurrence networks with two different exponents has been attributed to a dynamics combining the growth and preferential-attachment rules, where a pair of disconnected nodes becomes connected with a probability proportional to the product of their degrees [54]. The model is only a starting point, because it fails to reproduce other important properties of real networks; for example, the distribution of eigenvalues of the corresponding adjacency matrix [60]. A different model, not based on preferential attachment and mirroring a previous model of protein-interaction networks [78], introduced the concepts of growth via node duplication and link rewiring to Cognitive Science, to provide a unified explanation of the power-law distribution, the short path length, and the high clustering of semantic networks [44]. However, a simple network-growth dynamics is not necessarily the best mechanism. In a network of Wikipedia pages, the distribution of connected component sizes at the percolation threshold was found to be inconsistent with a randomly growing network [79]. In phonological similarity networks, five key properties – the largest connected component including about 50% of all vertices, small path lengths, high clustering, exponential degree distribution, and assortativity [80] – may arise from a network of predefined vertices and connections defined simply by overlap between properties of the node, rather than a growth model [81]. Overall, the debate over the different origins of cognitive networks highlights the importance of defining suitable model selection methods (see Section IV).

The network approach also suggests potentially revolutionary insights into the fast or even abrupt emergence of new cognitive functions during development, as well as the degradation of those functions with aging or neurodegenerative illness. Such abrupt changes can arise from smooth change, if the system crosses a percolation threshold; that is, a crucial point where the network becomes suddenly connected (e.g., during development) or disconnected (during aging or illness). The existence of such a point has been demonstrated in a semantic network extracted by

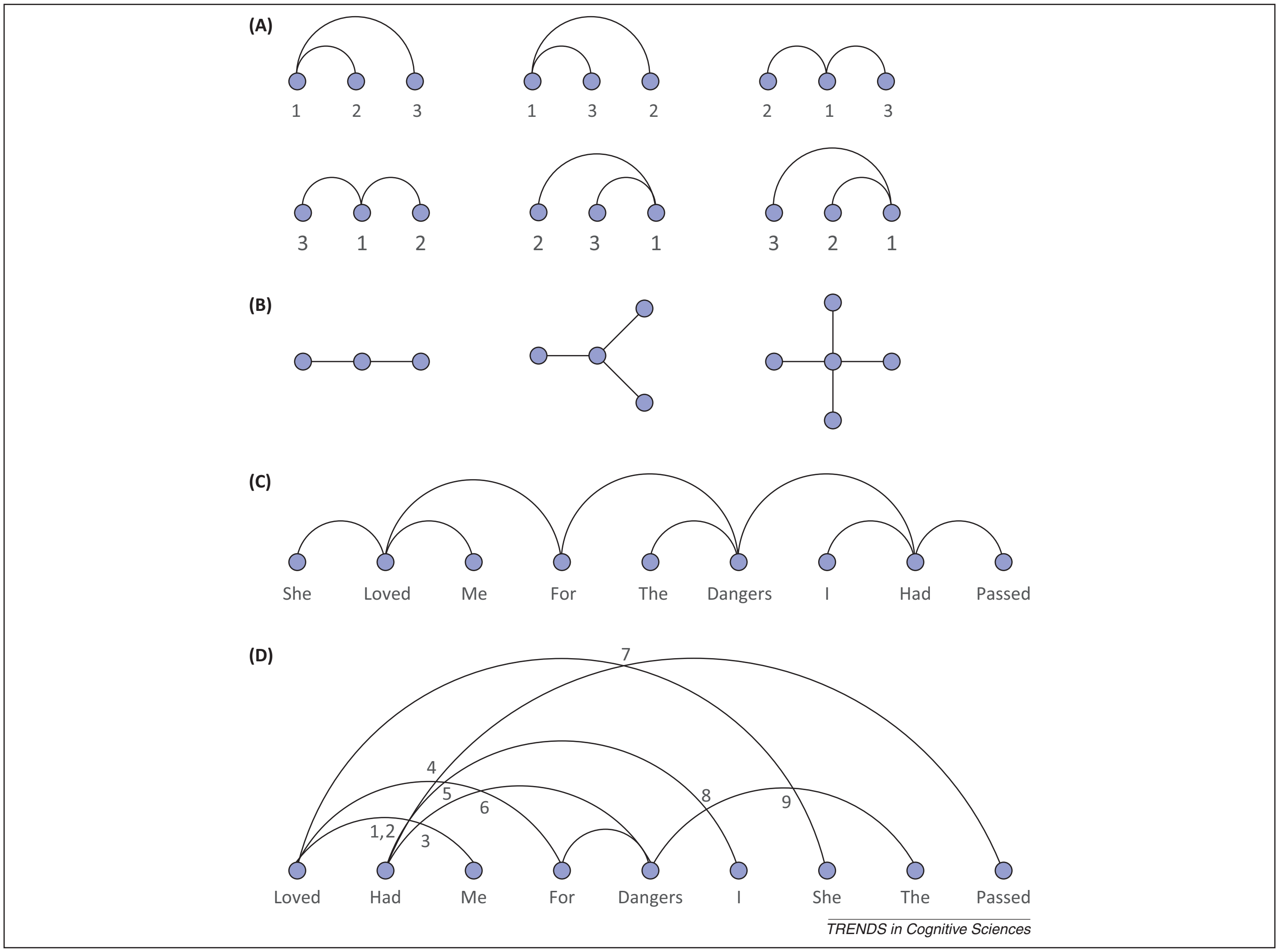

Figure 4. (**A**) The six possible linear arrangements of the three vertices of a tree. (**B**) Star trees of three, four, and five vertices. (**C**) The syntactic dependency tree of an English sentence (borrowed from [169]). Vertices are words and edges indicate syntactic dependencies between words. (**D**) A random linear arrangement of the sentence in (C) with nine edge crossings indicated by the numbers 1 to 9 (adapted from [73]). Two edges cross if they do not share vertices and one of the vertices making one of the edges is placed between the pair of vertices making the other edge. For instance, the seventh crossing is formed by the edge between 'loved' and 'she' and the edge between 'had' and 'passed'.

Wikipedia evolving by the addition of new pages [79]. Furthermore, the concept of percolation has inspired a recent explanation of hyperpriming and related phenomena exhibited by Alzheimer's disease patients in a theoretical model that qualitatively captures aspects of the experimental data [82].

Social networks and cognition

Network Science has been fruitfully applied to the investigation of networks of interactions between people, highlighting the interplay between individual cognition and social structure. For example, collaboration networks, both in scientific publications [83] and in Wikipedia [84], where a link is established between two authors if they have collaborated on at least one paper or page, provide insights into the large-scale patterns of cooperation among individuals and show a pronounced small-world property and high clustering [83]. Similarly, a 'rich-get-richer' phenomenon drives the dynamics of citation networks, between both papers and authors [21]. Scientific authors tend to cite already highly cited papers, leaving importance or quality in second place [85]. Moreover, pioneer authors benefit from a 'first-mover advantage' according to which the first paper in a particular topic often collects more citations than the best one [86]. The same approach has also allowed identification of the mechanisms that govern the emergence of (unfounded) authority among scientists, and their consequences [87].

One recent focus of research has been the large-scale validation of the so-called Dunbar number. Dunbar compared typical group size and neocortical volume in a wide range of primate species [88], concluding that biological and cognitive constraints would limit the immediate social network of humans to 100–200 individuals [89]. Analyzing a network of Twitter conversations involving 1.7 million individuals, it has been possible to confirm that users can maintain a limited number of stable relationships and that this number agrees well with Dunbar's predictions [90].

Social networks also play a fundamental role in collective problem-solving tasks [91]. For example, the speed of discovery of and convergence on an optimal solution is strongly affected by the underlying topology of the group

Box 5. Frontiers in Network Science

In the main text, we reviewed key contributions of network theory to Cognitive Science, highlighting that, along with the traditional study of the properties of fixed networks (Section 'Applications of network theory in Cognitive Science'), a recent wave also considers dynamical process on networks (Section 'Simple dynamics on networks'). Here, we sketch a brief overview of some topics at the frontiers of Network Science [170] that may have a substantial impact on Cognitive Science and many other disciplines in the near future.

A first challenge concerns the problem of timescale separation. Traditionally, two limits have been considered in the study of dynamical processes on networks: either the network is considered to be effectively static, meaning that it evolves on a timescale much slower than the one of the process under consideration, or, by contrast, it is described as rapidly varying with a pace that allows the process to perceive only the statistical properties of the graph (e.g., the degree distribution only) [8]. The issue is now to develop tools to describe what happens in the intermediate situations; that is, when the timescale of the dynamical process is comparable to the rate of network evolution [171]. Real-world examples of this can be found in social and cognitive processes occurring on face-to-face interaction networks [97] or on online messenger sites such as Twitter [90].

The second challenge is deeply connected to the first and goes one step further. What happens when the dynamical process coevolves with the underlying network, so that both dynamics interact with each other through feedback mechanisms? Recent research has shown that, when this is the case, interesting self-organization phenomena may arise, such as the possible fragmentation of social networks when links can be rewired depending on the dynamical state (i.e., the opinion) of the nodes (i.e., the individuals) they connect [172,173].

Finally, apart from the challenges of describing, modeling and understanding complex networks, a further question is how they can be controlled [174]. Control theory offers important mathematical tools to address this question, but network heterogeneity introduces non-trivial issues that have just started to be taken into account. Identifying driver nodes that can guide the system's entire dynamics over time, for example, might help the engineering of an observed system to a perform desired function or prevent malfunctioning. Interestingly, such nodes tend not to be the hubs of the network [174].

in a way that depends on the problem at hand [14,92]. More spatially based cliques seem to be advantageous for problems that benefit from broad exploration of the problem space, whereas long-distance connections enhance the results in problems that require less exploration [14], despite recent experiments suggesting that long-distance connections might always be advantageous [92]. Similarly, the amount of accessible information impacts problem solving in different ways on different social network structures, more information having opposite effects on different topologies [93].

Human behavior in social interactions has been revealed through the empirical analysis of telephone calls [94,95] and face-to-face interaction networks [96,97]. This research has clarified the relationship between the number and the duration of individual interactions or, put in network terms, between the degree and the strength of the nodes. Surprisingly, this relation differs in telephone versus face-to-face interactions; the more calls an individual makes, the less time per call he or she will allot [98], but for face-to-face interactions, popular individuals are 'super-connectors', with not only more but also longer contacts [97]. Other insights into the effect of social networks have been obtained through controlled experiments on the spread of a health behavior through artificially structured online communities [99]. Behavior spreads faster across clustered-lattice networks than across corresponding random networks. The impacts of network structure in understanding how societies solve problems and passing information may have strong parallels with how the 'society of mind' [100] within a single individual is implemented in information-processing mechanisms and neural structure.

Simple dynamics on networks

So far we have considered the structure of networks and the dynamical principles of growth or deletion (re)shaping these structures. Recently, however, new approaches have adopted a different perspective [101]: the neural, cognitive or social process is modeled as a dynamic process taking place on a network. Researchers can then ask how the network structure affects the dynamics.

An illustrative example concerns interactions among neural or cortical neurons, which often yield network-level synchrony [102–104]. Various studies reveal that abnormal synchrony in the cortex is observed in different pathologies, ranging from Parkinson's disease (excessive synchrony) [105] to autism (weak synchrony) [106,107]. Neural avalanches constitute another important process occurring on brain networks [108]. The size distribution of these bursts of activity approximate a power law, often a signature of complex systems [109]. The Kinouchi–Copelli (KC) model suggested that the neuronal dynamic range is optimized by a specific network topology tuned to signal propagation among interacting excitable neurons that leads to neural synchronization as a side-effect [110]. Remarkably, the predictions of this model have been confirmed empirically in cultures of cortex neurons where excitatory and inhibitory interactions were tuned pharmacologically [111]. Similar phenomena have been identified in connection to maximal synchronizability [103], information transmission [108,112], and information capacity [112] in cortical networks.

In the same way, it is interesting to speculate that some aspects of memory, thought, and language may be usefully modeled as navigation (i.e., the process of finding the way to a target node efficiently [48,113,114]) or exploration (i.e., navigation without a target) on network representations of knowledge by means of various strategies, such as simple random walks [57,115] or refined versions combining local exploration and 'switching' [116]. Statistical regularities such as Zipf's law can arise even from a random walk through a network where vertices are words [115]. Semantic categories and semantic similiarity between words can then emerge from properties of random walks on a word-association network [57]. Improved navigation strategies (random walks with memory) help to build efficient maps of the semantic space [117]. Furthermore, people apparently use nodes with high closeness centrality to navigate from one node to another in an experiment on navigating an artificial network [48,118]. These nodes are reminiscent of the landmarks used to navigate in the physical environment [119].

Network analysis casts light on the so-called 'function' words [120] (e.g., in, the, over, and, of). These are hubs of the semantic network and they are indeed 'authorities' according to PageRank, a sophisticated technique used by Google to determine the importance of a vertex (e.g., a word) from its degree and the importance of its neighbors [47]. Such hubs provide efficient methods for the exploration of semantic networks [116]. Moreover, the ease with which a word is recognized depends on its degree [65] and its clustering coefficient [66,67]. PageRank is a better predictor of the fluency with which a word is generated by experimental participants than the frequency or the degree of a word [118].

Another example is found in the collective dynamics of social annotation [121], occurring on websites (such as *Bibsonomy*) that allow users to tag resources; that is, to associate keywords to, for example, photos or links. First, a co-occurrence graph is obtained by establishing a link between two tags if they appear together in at least one post. The study of the network's evolution generates interesting observations, such as yet another power law, Heaps' law, which relates the number of word types ('the observed vocabulary size') and word tokens in a language corpus [122]. In addition, the mental space of the user is represented in terms of a synthetic semantic network and a single synthetic post is then generated by a finite random walk (Box 3) exploring this graph. Many synthetic random walk-generated posts are then created and an artificial co-occurrence network is built. Different synthetic mental spaces are then tested. The artificial co-occurrence network reproduces many of the features of the real graph, if the synthetic semantic graph has the small-world property and finite connectivity [121].

In the study of language dynamics and evolution, social networks describing the interactions between individuals have been central [15,123]. The role of the topology of such networks has been studied extensively for the Naming Game [124,125], a simple model of the emergence of shared linguistic conventions in a population of individuals. When the social network is fully connected, the individuals reach a consensus rapidly, but the possibility of interacting with anybody else requires a large individual memory to take into account the conventions used by different people [125]. When the population is arranged on a lattice, however, individuals are forced to interact repeatedly with their neighbors [126], so that while local agreement emerges rapidly with the agents using a very little memory, global convergence is reached slowly through the competition of the different locally agreeing groups (local clusters). Small-world networks, by contrast, are optimal in the sense that finite connectivity allows the individuals to use a finite amount of memory, as in lattices, whereas the small-world property prevents the formation of local clusters [15,127]. Similar analyses have been performed for the case of competition not between specific linguistic conventions, but between entire languages [128–130]. Overall, these studies highlight the importance of the properties of social networks for the emergence and maintenance of complex cognition, language, and culture. The study of dynamics on networks is also likely to clarify the relevance of properties of network structure, such as path lengths and clustering, for cognitive processes and their pathologies. A take-home message from this research is that network theory challenges radically the view that the unique requirement for complex cognition and its multiple manifestations is the human brain. Instead, the key for the emergence and maintenance of such skills might be the properties of the network defining how the individuals interact.

Methodological issues for future research

Despite of the enormous potential of network theory for the cognitive and brain sciences, important methodological challenges remain regarding network construction, analysis, and modeling.

Challenges for network construction

A basic challenge for the analysis of co-occurrence networks is determining whether two vertices have co-occurred above chance – that is, inferring whether an edge should be drawn between them (an issue that arises in the parallel literature on probabilistic graphical models [131,132]). In networks of co-occurrence, typically no statistical filter is used [133] or the filter is not well defined [61]. For this reason, proper statistical filters (e.g., [134]) or more precise ways of linking vertices have been considered; for example, syntactic dependency instead of word occurrence [135,136]. In general, however, defining an appropriate null hypothesis for the existence of an edge is crucial. In cases of networks induced from co-occurrences of elements in a sequence [133,137], this would distinguish between significant above-random properties, identified by the ensemble of permutations of the original sequence (e.g., the permutations of a text), and non-significant findings. Even the latter are important, however, as they may suggest that some features of the network could be a side effect purely of the frequency with which the elements occur. The issue is a general one; in brain network research, it arises when determining whether the activities of two brain regions is really correlated [138], whereas in collaboration networks, connecting two scientists because they have 'co-occurred' in the coauthor list of an article does not imply that they have actually collaborated [83]. This variety of applications highlights the value of Network Science in offering a unified framework to the various areas of Cognitive Science.

Challenges for network analysis

The most commonly used null hypothesis for the statistical properties of a network is the Erdös–Rényi (or binomial) network (Box 2, Figures 2 and 3). A better null hypothesis is a network that preserves the original degree sequence but in which edges are randomized [8], which in general clarifies the role of the degree distribution and how it could be responsible for the properties of the observed network. For instance, a power-law distribution of degrees may lead to an apparently large clustering coefficient in networks of not too large a size [8]. Other properties, however, can depend on further details apart from the degree distribution. For example, apparently harmless manipulations such as banning loops (edges from a node to itself) and multiple edges (more than two edges joining a pair of nodes) can lead to degree correlations and disassortative behavior in power-law degree distributions [139].

Table 1. Cognitive Science through the eyes of network theory: translation of Cognitive Science terms into network theory concepts

Cognitive Science and neighboring fields	Network theory
Semantic field	Community in a network (e.g., word-association network) [56,57]
Island	Connected component [80]
Brain module	Community in a brain network [35]
Semantic memory	Semantic network [57]
Mental exploration (mental navigation without a target)	Random walk in a cognitive network [57,115]
Tagging activity by users	Random walk in a mental semantic network [121]
Landmark (in a way-finding problem)	Node with high closeness centrality [48]
Pathological brain or pathological cognition	Anomalous network metrics; for example, clustering and path lengths [28,46,52]
Unfounded scientific authority, first-mover advantage	Rich-get-richer phenomenon on a citation network [86,87]

Another challenging problem concerns the degree distribution, which is often assumed to be a power law (Box 2). First, where a power law is certain (Figure 3), direct regression methods to determine the degree exponent are potentially biased and an estimation by maximum likelihood is more convenient [140]. However, various distributions, not only the power law, are able to mimic an approximate straight line in double logarithmic scale [141,142] and a power-law degree distribution has been found not to be sufficiently supported in biological networks, contrary to previous beliefs [143,144]. In general, the analysis of the degree distribution would require the use of standard model-selection techniques from an ensemble of candidate distributions [145]. Equivalent evaluations of a power law in cognitive networks are not available as far as we know.

Challenges for dynamical models

A big challenge for understanding the dynamical processes underlying brain and cognitive networks is determining which underlying network model is most appropriate. Because many network models can account for a power-law distribution (Box 2), other network features must be introduced in the evaluation of the most likely model. However, perhaps the most valuable information is how the network has evolved to reach a certain configuration. Different dynamical rules may lead to the same end product and it is possible to use sophisticated techniques to assess the importance of different evolutionary mechanisms [146–148]. These methods could help clarify the debate on the dynamical principles guiding the evolution of semantic networks; for example, preferential attachment and its variants in normal and late talkers [52,149]. Incorporating the statistical methods mentioned above is vital to harness the power of Network Science to reveal the dynamical principles by which the brain is structured and by which brain functions emerge, develop, and decay.

Concluding remarks and outlook

Our survey of the vast literature on network theory for brain and Cognitive Sciences has necessarily been selective, but it allows us to draw several encouraging conclusions. Network Science offers concepts for a new understanding of traditional terms in Cognitive Science (Table 1) and illuminates a wide range of phenomena, such as the organization of pathological brains or cognition (e.g., [38]), the development of vocabulary in children (e.g., [47,149]), and language competition (e.g., [130]) under the same theoretical umbrella. Many new questions arise concerning how far network properties at the neural level translate into network properties at higher levels and vice versa (Box 6). Network theory also may help bridge the gap between the brain and the mind, shedding new light on how knowledge is stored and exploited as well as reducing the gulf that separates the study of individual and collective behavior. Moreover, understanding the origin of the observed properties of networks through the tools of Network Science may help unify research on the development of cognition during childhood with the study of processing in the adult state and its decay during aging or illness. Network Science is a young discipline (Box 5), but it promises to be a valuable integrative framework for understanding and relating the analysis of mind and behavior at a wide range of scales, from brain processes to patterns of social and cultural interaction. Overall, network theory can help Cognitive Science become more internally coherent and more interconnected with the many other fields where network theory has proved fruitful.

Box 6. Outstanding questions

- Is network theory a framework that can unify the representation of structure across levels and domains in Cognitive Science and neighboring disciplines (e.g., from neural organization to knowledge representation)?
- To what extent do the underlying brain networks determine the properties of cognitive networks and vice versa? Which well-known properties of brain networks are also found at higher levels in cognitive networks and vice versa?
- What are the optimal values of path lengths and/or clustering for proper brain functioning, cognitive processing, or social dynamics? Do these optimal values depend on the cognitive domain? Do very low or high values indicate pathology? If so, do such indicators apply across different explanatory levels; for example, do the aberrant statistical properties of brain networks observed in Alzheimer's disease, schizophrenia, or autism also arise at the cognitive level?
- Are the properties of the network structure in social interactions a key factor for the emergence of complex individual abilities such as language (e.g., syntax)? Conversely, to what extent are the properties of these social interactions determined by individual cognitive abilities (e.g., Dunbar's number)?

Acknowledgments

The authors are grateful to B. Elvevåg for making them aware of relevant work and helpful discussions. R.F-i-C. was supported by the grant Iniciació i Reincorporació a la Recerca from the Universitat Politècnica de Catalunya and the grants BASMATI (TIN2011-27479-C04-03) and OpenMT-2 (TIN2009-14675-C03) from the Spanish Ministry of Science and Innovation. R.P-S. acknowledges financial support from the Spanish MICINN, under project FIS2010-21781-C02-01, and additional support through ICREA Academia, funded by the Generalitat de Catalunya. N.C. is supported by ERC Advanced Grant 295917-RATIONALITY.

References

1 Herculano-Houzel, S. (2009) The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* 3, 31
2 Bullmore, E. and Sporns, O. (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198
3 Anderson, P.W. (1972) More is different. *Science* 177, 393–396
4 Pastor-Satorras, R. and Vespignani, A. (2004) *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press
5 Barabási, A.L. (2011) The network takeover. *Nat. Phys.* 8, 14
6 Barabási, A.L. *et al.* (2011) Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68
7 Caldarelli, G. (2007) *Scale-free networks: Complex Webs in Nature and Technology*. Oxford University Press
8 Newman, M.E.J. (2010) *Networks: An Introduction*. Oxford University Press
9 Bassett, D.S. and Gazzaniga, M.S. (2011) Understanding complexity in the human brain. *Trends Cogn. Sci.* 15, 200–209
10 Bressler, S.L. and Menon, V. (2010) Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290
11 Mehler, A. (2007) Large text networks as an object of corpus linguistic studies. In *Corpus Linguistics. An International Handbook of the Science of Language and Society* (Lüdeling, A. and Kytö, M., eds), pp. 328–382, de Gruyter
12 Borge-Holthoefer, J. and Arenas, A. (2010) Semantic networks: structure and dynamics. *Entropy* 12, 1264–1302
13 Solé, R.V. *et al.* (2010) Language networks: their structure, function, and evolution. *Complexity* 15, 20–26
14 Mason, W.A. *et al.* (2008) Propagation of innovations in networked groups. *J. Exp. Psychol. Gen.* 137, 422
15 Dall'Asta, L. *et al.* (2006) Nonequilibrium dynamics of language games on complex networks. *Phys. Rev. E* 74, 036105
16 Euler, L. (1732) Solution problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 8, 128–140 (in Latin)
17 Bollobás, B. (1998) *Modern Graph Theory*. Springer-Verlag
18 Milgram, S. (1967) Small-world problem. *Psychol. Today* 1, 61–67
19 Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press
20 Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature* 393, 440–442
21 Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509–512
22 Albert, R. *et al.* (2000) Error and attack tolerance of complex networks. *Nature* 406, 378–382
23 Pastor-Satorras, R. and Vespignani, A. (2001) Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200–3203
24 Sporns, O. *et al.* (2004) Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425
25 Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Company
26 Graham, D. and Rockmore, D. (2011) The packet switching brain. *J. Cogn. Neurosci.* 23, 267–276
27 Deco, G. *et al.* (2008) The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* 4, e1000092
28 Zhao, X. *et al.* (2012) Disrupted small-world brain networks in moderate Alzheimer's disease: a resting-state fMRI study. *PLoS ONE* 7, e33540
29 Zamora-Lopez, G. *et al.* (2011) Characterizing the complexity of brain and mind networks. *Phil. Trans. R. Soc. A* 369, 3730–3747
30 Kim, B.J. (2004) Performance of networks of artificial neurons: the role of clustering. *Phys. Rev. E* 69, 045101
31 Oshima, H. and Odagaki, T. (2007) Storage capacity and retrieval time of small-world neural networks. *Phys. Rev. E* 76, 036114
32 van den Heuvel, M.P. and Sporns, O. (2011) Rich-club organization of the human connectome. *J. Neurosci.* 31, 15775–15786
33 Sporns, O. (2011) The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125
34 van den Heuvel, M.P. *et al.* (2012) High-cost, high-capacity backbone for global brain communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11372–11377
35 Wu, K. *et al.* (2011) The overlapping community structure of structural brain network in young healthy individuals. *PLoS ONE* 6, e19608
36 Fodor, J.A. (1983) *The Modularity of Mind*. MIT Press
37 Bassett, D.S. *et al.* (2008) Hierarchical organization of human cortical networks in health and schizophrenia. *J. Neurosci.* 28, 9239–9248
38 Barttfeld, P. *et al.* (2011) A big-world network in ASD: dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. *Neuropsychologia* 49, 254–263
39 Achard, S. *et al.* (2013) Hubs of brain functional networks are radically reorganized in comatose patients. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20608–20613
40 Lee, J.H. *et al.* (2010) Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature* 465, 788–792
41 Bassett, D.S. *et al.* (2011) Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7641–7646
42 Seeley, W.W. *et al.* (2009) Neurodegenerative diseases target large-scale human brain networks. *Neuron* 62, 42
43 Thomas, C. *et al.* (2008) Reduction in white matter connectivity, revealed by diffusion tensor imaging, may account for age-related changes in face perception. *J. Cogn. Neurosci.* 20, 268–284
44 Steyvers, M. and Tenenbaum, J.B. (2005) The large scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn. Sci.* 29, 41–78
45 Ferrer i Cancho, R. *et al.* (2007) Spectral methods cluster words of the same class in a syntactic dependency network. *Int. J. Bifurcation Chaos* 17, 2453–2463
46 Lerner, A.J. *et al.* (2009) Network graph analysis of category fluency testing. *Cogn. Behav. Neurol.* 22, 45–52
47 Ke, J.Y. and Yao, Y. (2008) Analysing language development from a network approach. *J. Quant. Linguist.* 15, 70–99
48 Sudarshan Iyengar, S. *et al.* (2012) Understanding human navigation using network analysis. *Top. Cogn. Sci.* 4, 121–134
49 Costa, L.d.F. *et al.* (2011) Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Adv. Phys.* 60, 329–412
50 Nelson, D.L. *et al.* (2004) The University of South Florida free association, rhyme, and word fragment norms. *Behav. Res. Methods* 36, 402–407
51 Gravino, P. *et al.* (2012) Complex structures and semantics in free word association. *Adv. Complex Syst.* 15, 1250054
52 Beckage, N. *et al.* (2011) Small worlds and semantic network growth in typical and late talkers. *PLoS ONE* 6, e19348
53 Picard, O. *et al.* (2009) Hierarchies in dictionary definition space. In *23rd Annual Conference on Neural Information Processing Systems: Workshop on Analyzing Networks and Learning With Graphs, Vancouver BC, Canada, 11–12 December 2009*, University of Southampton
54 Dorogovtsev, S.N. and Mendes, J.F.F. (2001) Language as an evolving word web. *Proc. R. Soc. B* 268, 2603–2606
55 Holme, P. (2005) Core-periphery organization of complex networks. *Phys. Rev. E* 72, 046111
56 Fortunato, S. (2010) Community detection in graphs. *Phys. Rep.* 486, 75–174
57 Borge-Holthoefer, J. and Arenas, A. (2010) Categorizing words through semantic memory navigation. *Eur. Phys. J. B* 74, 265–270
58 Ferrer i Cancho, R. and Solé, R.V. (2001) Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revisited. *J. Quant. Linguist.* 8, 165–173

59 Petersen, A.M. *et al.* (2012) Languages cool as they expand: allometric scaling and the decreasing need for new words. *Sci. Rep.* http://dx.doi.org/10.1038/srep00943
60 Choudhury, M. *et al.* (2010) Global topology of word co-occurrence networks: beyond the two-regime power-law. In *Coling* (2010) Huang, C-R. and Jurafsky, D.,eds In pp. 162–170, Tsinghua University Press
61 Ferrer i Cancho, R. and Sole, R.V. (2001) The small world of human language. *Proc. R. Soc. B* 268, 2261–2265
62 Roy, R.S. *et al.* (2011) Complex network analysis reveals kernel-periphery structure in web search queries. In *2nd International ACM SIGIR (Association for Computing Machinery Special Interest Group on Information Retrieval) Workshop on Query Representation and Understanding (QRU 2011)*, pp. 5–8, Association for Computing Machinery
63 Hagoort, P. (1998) The shadows of lexical meaning in patients with semantic impairments. In *Handbook of Neurolinguistics* (Stemmer, B. and Whitaker, H.A., eds), pp. 235–248, Academic Press
64 Gainotti, G. (1998) Category specific disorders of nouns and verbs. A very old and very new problem. In *Handbook of Neurolinguistics* (Stemmer, B. and Whitaker, H.A., eds), pp. 3–12, Academic Press
65 Luce, P.A. and Pisoni, D.B. (1998) Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36
66 Chan, K.Y. and Vitevitch, M.S. (2009) The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1934–1949
67 Vitevitch, M.S. *et al.* (2011) Simulating retrieval from a highly clustered network: implications for spoken word recognition. *Front. Psychol.* 2, 369
68 Amancio, D.R. *et al.* (2011) On the concepts of complex networks to quantify the difficulty in finding the way out of labyrinths. *Phys. A* 390, 4673–4683
69 Morrill, G. (2000) Incremental processing and acceptability. *Comput. Linguist.* 26, 319–338
70 Park, Y.A. and Levy, R. (2009) Minimal-length linearizations for mildly context-sensitive dependency trees. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 335–343, Association for Computational Linguistics
71 Ferrer i Cancho, R. (2004) Euclidean distance between syntactically linked words. *Phys. Rev. E* 70, 056135
72 Díaz, J. *et al.* (2002) A survey of graph layout problems. *ACM Comput. Surv.* 34, 313–356
73 Ferrer i Cancho, R. (2006) Why do syntactic links not cross? *Europhys. Lett.* 76, 1228
74 Ferrer-i-Cancho, R. (2008) Some word order biases from limited brain resources. A mathematical approach. *Adv. Complex Syst.* 11, 393–414
75 Liu, H. (2008) Dependency distance as a metric of language comprehension difficulty. *J. Cogn. Sci.* 9, 159–191
76 Temperley, D. (2008) Dependency-length minimization in natural and artificial languages. *J. Quant. Linguist.* 15, 256–282
77 Ferrer-i-Cancho, R. (2013) Hubiness, length, crossings and their relationships in dependency trees. *Glottometrics* 25, 1–21
78 Pastor-Satorras, R. *et al.* (2003) Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* 222, 199–210
79 Masucci, A.P. *et al.* (2011) Wikipedia information flow analysis reveals the scale-free architecture of the Semantic Space. *PLoS ONE* 6, e17333
80 Vitevitch, M.S. (2008) What can graph theory tell us about word learning and lexical retrieval? *J. Speech Lang. Hear. Res.* 51, 408–422
81 Gruenenfelder, T.M. and Pisoni, D.B. (2009) The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons. *J. Speech Lang. Hear. Res.* 52, 596–609
82 Borge-Holthoefer, J. *et al.* (2011) Modeling abnormal priming in Alzheimer's patients with a free association network. *PLoS ONE* 6, e22651
83 Newman, M.E.J. (2001) The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U.S.A.* 98, 404–409
84 Brandes, U. *et al.* (2009) Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th International Conference on World Wide Web*, pp. 731–740, Association for Computing Machinery
85 Price, D.S. (1976) A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.* 27, 292–306
86 Newman, M. (2009) The first-mover advantage in scientific publication. *Europhys. Lett.* 86, 68001
87 Greenberg, S.A. (2009) How citation distortions create unfounded authority: analysis of a citation network. *BMJ* http://dx.doi.org/10.1136/bmj.b2680
88 Dunbar, R.I.M. (1992) Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* 22, 469–493
89 Dunbar, R.I.M. (1998) The social brain hypothesis. *Brain* 9, 10
90 Gonçalves, B. *et al.* (2011) Modeling users' activity on Twitter networks: validation of Dunbar's number. *PLoS ONE* 6, e22656
91 Bavelas, A. (1950) Communication patterns in task-oriented groups. *J. Acoust. Soc. Am.* 22, 725–730
92 Mason, W. and Watts, D.J. (2012) Collaborative learning in networks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 764–769
93 Kearns, M. *et al.* (2006) An experimental study of the coloring problem on human subject networks. *Science* 313, 824–827
94 Onnela, J.P. *et al.* (2007) Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7332–7336
95 Karsai, M. *et al.* (2011) Small but slow world: how network topology and burstiness slow down spreading. *Phys. Rev. E* 83, 025102
96 Hui, P. *et al.* (2005) Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking*, pp. 244–251, Association for Computing Machinery
97 Cattuto, C. *et al.* (2010) Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE* 5, e11596
98 Onnela, J.P. *et al.* (2007) Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys.* 9, 179
99 Centola, D. (2010) The spread of behavior in an online social network experiment. *Science* 329, 1194–1197
100 Minsky, M. (1988) *Society of Mind*. Simon and Schuster
101 Barrat, A. *et al.* (2008) *Dynamical Processes on Complex Networks*. Cambridge University Press
102 Kelso, J.A.S. (1995) *Dynamic Patterns: The Self-organization of Brain and Behavior*. MIT Press
103 Yang, H. *et al.* (2012) Maximal variability of phase synchrony in cortical networks with neuronal avalanches. *J. Neurosci.* 32, 1061–1072
104 Arenas, A. *et al.* (2008) Synchronization in complex networks. *Phys. Rep.* 469, 93–153
105 Boraud, T. *et al.* (2005) Oscillations in the basal ganglia: the good, the bad, and the unexpected. *Adv. Behav. Biol.* 56, 1–24
106 Wilson, T.W. *et al.* (2007) Children and adolescents with autism exhibit reduced MEG steady-state gamma responses. *Biol. Psychiatry* 62, 192–197
107 Yizhar, O. *et al.* (2011) Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* 477, 171–178
108 Beggs, J.M. and Plenz, D. (2003) Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177
109 Stanley, H.E. (1987) *Introduction to Phase Transitions and Critical Phenomena*. p. 336, Oxford University Press
110 Kinouchi, O. and Copelli, M. (2006) Optimal dynamical range of excitable networks at criticality. *Nat. Phys.* 2, 348–351
111 Shew, W.L. *et al.* (2009) Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. *J. Neurosci.* 29, 15595–15600
112 Shew, W.L. *et al.* (2011) Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *J. Neurosci.* 31, 55–63
113 Boguna, M. *et al.* (2008) Navigability of complex networks. *Nat. Phys.* 5, 74–80
114 Collins, A.M. and Quillian, M.R. (1969) Retrieval time from semantic memory. *J. Verb. Learn. Verb. Behav.* 8, 240–247
115 Allegrini, P. *et al.* (2004) Intermittency and scale-free networks: a dynamical model for human language complexity. *Chaos Soliton. Fract.* 20, 95–105
116 Goñi, J. *et al.* (2010) Switcher-random-walks: a cognitive-inspired mechanism for network exploration. *Int. J. Bifurcation Chaos* 20, 913–922
117 Capitán, J.A. *et al.* (2012) Local-based semantic navigation on a networked representation of information. *PLoS ONE* 7, e43694

118 Griffiths, T.L. *et al.* (2007) Google and the mind. *Psychol. Sci.* 18, 1069–1076
119 Golledge, G.T.M.R.G. (1976) *Environmental Knowing: Theories, Research and Methods*. Dowden, Hutchinson and Ross
120 Chung, C. and Pennebaker, J. (2007) The psychological functions of function words. In *Social Communication* (Fiedler, K., ed.), pp. 343–359, Psychology Press
121 Cattuto, C. *et al.* (2009) Collective dynamics of social annotation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10511–10515
122 Heaps, H.S. (1978) *Information Retrieval: Computational and Theoretical Aspects*. Academic Press
123 Gong, T. *et al.* (2004) A computational framework to simulate the coevolution of language and social structure. In *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Artificial Life*, pp. 158–164, MIT Press
124 Steels, L. (1995) A self-organizing spatial vocabulary. *Artif. Life* 2, 319–332
125 Baronchelli, A. *et al.* (2006) Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech. Theor. Exp.* 2006, P06014
126 Baronchelli, A. *et al.* (2006) Topology-induced coarsening in language games. *Phys. Rev. E* 73, 015102
127 Dall'Asta, L. *et al.* (2006) Agreement dynamics on small-world networks. *Europhys. Lett.* 73, 969
128 Castelló, X. *et al.* (2006) Ordering dynamics with two non-excluding options: bilingualism in language competition. *New J. Phys.* 8, 308
129 Castelló, X. *et al.* (2009) Consensus and ordering in language dynamics. *Eur. Phys. J. B* 71, 557–564
130 Vázquez, F. *et al.* (2010) Agent based models of language competition: macroscopic descriptions and order–disorder transitions. *J. Stat. Mech. Theor. Exp.* 2010, P04007
131 Erdős, P. and Rényi, A. (1959) On random graphs. *Publ. Math-Debrecen* 6, 290–297
132 Kumar, R. *et al.* (2000) Stochastic models for the web graph. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, pp. 57–65, Institute of Electrical and Electronics Engineers
133 Milo, R. *et al.* (2004) Superfamilies of evolved and designed networks. *Science* 303, 1538–1542
134 Goñi, J. *et al.* (2011) The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. *Cogn. Process* 12, 183–196
135 Liu, H. and Li, W. (2010) Language clusters based on linguistic complex networks. *Chin. Sci. Bull.* 55, 3458–3465
136 Ferrer i Cancho, R. *et al.* (2004) Patterns in syntactic dependency networks. *Phys. Rev. E* 69, 051915
137 Masucci, A.P. and Rodgers, G.J. (2009) Differences between normal and shuffled texts: structural properties of weighted networks. *Adv. Complex Syst.* 12, 113–129
138 Sporns, O. (2011) *Networks of the Brain*. MIT Press
139 Park, J. and Newman, M.E.J. (2003) Origin of degree correlations in the Internet and other networks. *Phys. Rev. E* 68, 026112
140 Goldstein, M. *et al.* (2004) Problems with fitting to the power-law distribution. *Eur. Phys. J. B* 41, 255–258
141 Clauset, A. *et al.* (2009) Power-law distributions in empirical data. *SIAM Rev.* 51, 661–663
142 Li, W. *et al.* (2010) Fitting ranked linguistic data with two-parameter functions. *Entropy* 12, 1743–1764
143 Tanaka, R. *et al.* (2005) Some protein interaction data do not exhibit power law statistics. *FEBS Lett.* 579, 5140–5144
144 Khanin, R. and Wit, E. (2006) How scale-free are biological networks. *J. Comput. Biol.* 13, 810–818
145 Kelly, W.P. *et al.* (2011) The degree distribution of networks: statistical model selection. In *Bacterial Molecular Networks*, pp. 245–262, Humana Press
146 Ratmann, O. *et al.* (2007) Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*. *PLoS Comput. Biol.* 3, e230
147 Takahashi, D.Y. *et al.* (2012) Discriminating different classes of biological networks by analyzing the graphs spectra distribution. *PLoS ONE* 7, e49949
148 Wiuf, C. *et al.* (2006) A likelihood approach to analysis of network data. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7566–7570
149 Hills, T.T. *et al.* (2009) Longitudinal analysis of early semantic networks preferential attachment or preferential acquisition? *Psychol. Sci.* 20, 729–739
150 Thomas, M.S. and McClelland, J.L. (2008) Connectionist models of cognition. In *The Cambridge Handbook of Computational Psychology* (Sun, R., ed.), pp. 23–58, Cambridge University Press
151 Elman, J.L. (2005) Connectionist models of cognitive development: where next? *Trends Cogn. Sci.* 9, 111–117
152 Onnis, L. *et al.* (2009) Connectionist models of language processing. In *New Encyclopedia of Neuroscience* (Squire, L-R., ed.), pp. 83–90, Elsevier
153 Rogers, T.T. and McClelland, J.L. (2004) *Semantic Cognition: A Parallel Distributed Processing Approach*. MIT Press
154 Dayan, P. *et al.* (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. *J. Cogn. Neurosci.* 15, 154–155
155 Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausble Inference*. Morgan Kaufmann
156 Pearl, J. (2000) *Causality: Models, Reasoning and Inference*. Cambridge University Press
157 Chater, N. *et al.* (2006) Probabilistic models of cognition: conceptual foundations. *Trends Cogn. Sci.* 10, 287–291
158 Dayan, P. (2012) Twenty-five lessons from computational neuromodulation. *Neuron* 76, 240–256
159 Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438
160 Dorogovtsev, S.N. *et al.* (2001) Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E* 63, 062101
161 Carlson, J.M. and Doyle, J. (1999) Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E* 60, 1412
162 Fabrikant, A. *et al.* (2002) Heuristically optimized trade-offs: a new paradigm for power laws in the Internet. In *Automata, Languages and Programming* (Widmayer, P. *et al.*, eds), pp. 110–122, Springer
163 Boguná, M. and Pastor-Satorras, R. (2003) Class of correlated random networks with hidden variables. *Phys. Rev. E* 68, 036112
164 Caldarelli, G. *et al.* (2002) Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* 89, 258702
165 Noh, J.D. and Rieger, H. (2004) Random walks on complex networks. *Phys. Rev. Lett.* 92, 118701
166 Baronchelli, A. and Pastor-Satorras, R. (2010) Mean-field diffusive dynamics on weighted networks. *Phys. Rev. E* 82, 011111
167 Chung, F.R.K. (1984) On optimal linear arrangements of trees. *Comput. Math. Appl.* 10, 43–60
168 Hochberg, R.A. and Stallmann, M.F. (2003) Optimal one-page tree embeddings in linear time. *Inform. Process. Lett.* 87, 59–66
169 Mel'čuk, I. (1988) *Dependency Syntax: Theory and Practice*. SUNY Press
170 Vespignani, A. (2011) Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* 8, 32–39
171 Holme, P. and Saramäki, J. (2012) Temporal networks. *Phys. Rep.* 519, 97–125
172 Nardini, C. *et al.* (2008) Who's talking first? consensus or lack thereof in coevolving opinion formation models. *Phys. Rev. Lett.* 100, 158701
173 Bohme, G.A. and Gross, T. (2012) Fragmentation transitions in multistate voter models. *Phys. Rev. E* 85, 066117
174 Liu, Y.Y. *et al.* (2011) Controllability of complex networks. *Nature* 473, 167–173