Characterizing the complexity of brain and mind networks

Gorka Zamora-López, Eleonora Russo, Pablo M. Gleiser, Changsong Zhou and Jürgen Kurths

References	**This article cites 63 articles, 12 of which can be accessed free** http://rsta.royalsocietypublishing.org/content/369/1952/3730.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections statistical physics (73 articles) complexity (47 articles) biophysics (472 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. A* go to:
http://rsta.royalsocietypublishing.org/subscriptions

Review

Characterizing the complexity of brain and mind networks

By Gorka Zamora–López[1,2], Eleonora Russo[3], Pablo M. Gleiser[4], Changsong Zhou[5,6,*] and Jürgen Kurths[2,7]

[1]*Bernstein Center for Computational Neuroscience, Berlin, Germany*
[2]*Department of Physics, Humboldt University, Berlin, Germany*
[3]*SISSA – Cognitive Neuroscience, Trieste, Italy*
[4]*Centro Atómico Bariloche, Instituto Balseiro, Bariloche, Argentina*
[5]*The Beijing–Hong Kong–Singapore Joint Centre for Nonlinear and Complex Systems, Hong Kong, China*
[6]*Department of Physics, Center for Nonlinear Studies, Hong Kong Baptist University, Hong Kong, China*
[7]*Potsdam Institute for Climate Impact Research, Potsdam, Germany*

Recent studies of brain connectivity and language with methods of complex networks have revealed common features of organization. These observations open a window to better understand the intrinsic relationship between the brain and the mind by studying how information is either physically stored or mentally represented. In this paper, we review some of the results in both brain and linguistic networks, and we illustrate how modelling approaches can serve to comprehend the relationship between the structure of the brain and its function. On the one hand, we show that brain and neural networks display dynamical behaviour with optimal complexity in terms of a balance between their capacity to simultaneously segregate and integrate information. On the other hand, we show how principles of neural organization can be implemented into models of memory storage and recognition to reproduce spontaneous transitions between memories, resembling phenomena of memory association studied in psycholinguistic experiments.

Keywords: brain networks; semantic networks; complexity; memory latching; free association

1. Introduction

In recent years, a novel field has emerged which aims at the analysis, characterization and modelling of intricate interrelations between the components that form complex systems. Networks of both natural and artificial systems are often coined with the term *complex* because their topology is neither regular nor random, they contain some degree of organization that is not trivial at

first sight. Systems of diverse origin often share similar topological properties indicating the presence of common mechanisms of formation and functional organization [1–3]. Such is the case of neural and cortical networks [4,5], as well as linguistic-related networks [6,7]. These observations open the door to a promising framework which is gaining experimental strength: that one of the ingredients to bridge the gap between the brain and the mind is to understand how knowledge is organized. This comprises how information is either physically stored and mentally represented.

Physiological evidences and cognitive theories have long foreseen high-level cognitive functions as emerging from the interplay of distributed but interconnected neural networks [8–13]. The novelty is that only during the last two decades have reliable data of anatomical connectivity been collected and analysed [4,5]. On the other hand, uncovering the structure of the mind, i.e. the organization of knowledge at the mental level, is rather challenging. Beyond technological limitations, the challenge lies on the conceptual difficulty of defining what the observables are and how to quantify them. Percepts, concepts, words, memories of past experiences, feelings, etc., are all elements of the mental world, and they are interrelated. To some extent, the manner in which those items are associated at the mental level is intrinsically related to the manner in which memories are stored, retrieved and associated at the neural level. As explicit memories depend on the personal history of individuals, they are unsuitable for experimental study and statistical characterization. On the other hand, languages are abstract mental objects with their own rules, distinctive constituents (words) and are shared by large human populations. Therefore, language and the words encoding conceptual memories form a suitable experimental window through which the organization of the mind can be explored.

In this paper, we summarize present knowledge on anatomical and linguistic networks. We describe their common properties and we illustrate how modelling of memory storage and retrieval in neural-like networks emerges as a promising approach to shed light into the elusive problem of linking brain and mind. In §2, a concise introduction to complex networks and graph theory is provided. Section 3 summarizes the physical organization of the nervous system at different scales, from neuronal assemblies to the macroscopic connectivity between different parts of the brain. By means of dynamical simulations, we show that the modular and hierarchical architecture of neural networks enhances the complexity of dynamical processes they host. In §4, the properties of linguistic networks are reviewed with a particular focus on free-association studies. A model is introduced that produces spontaneous retrieval of memories. The transition probabilities between the memories allow us to generate simulated networks of free association whose properties are comparable to the characteristics of empirically obtained free-association networks.

2. Complex networks in a nutshell

Many natural and artificial systems are composed of multiple elements whose interrelations can be represented as a graph. This abstract representation provides the system with a form (a topology) that is a subject of mathematical characterization and statistical description. A graph is composed of nodes that are

connected by links. The nature of the links differs depending on the system under study. They can be physical entities, such as the cables connecting a network of computers, or the axonal projections between neurons. Often, they represent more abstract interactions, e.g. the friendship ties between a group of humans or the semantic relationship between words. Graphs are represented by an adjacency matrix A, whose elements $A_{ij} = 1$ if there is a link connecting the node number i to node number j; otherwise, $A_{ij} = 0$. The matrix entries may take either integer or real values, encoding the strength of the link between the two nodes.

In order to uncover the topological architecture of a network, many statistical descriptors are applied to the adjacency matrix, allowing the organization of the network to be scanned through at different scales. The degree $k(i)$ of a node i is the number of nodes to which the node is connected. The distribution of the degrees $p(k)$ informs about the heterogeneity of nodes we can find in the network. In a regular lattice, all nodes have the same number of neighbours and hence they are indistinguishable. Many real networks display a scale-free degree distribution, meaning that most of the nodes have few connections and only a few nodes have many neighbours.

Other popular measures include the clustering coefficient C and the average path length l. The clustering characterizes the transitive probability that two nodes are connected to each other, provided they have a common neighbour. In social terms, it is well known that two persons are more likely to know each other, provided they have a common friend. In a network, the distance between two nodes d_{ij} is quantified by the number of links crossed to travel from one node i to another node j. If there is a link connecting the two nodes, then $d_{ij} = 1$. If there is no direct link but it can travel from i to another node k that is connected to j, then $d_{ij} = 2$, and so on. If there is no path to travel between two nodes, then they are at an infinite distance and $d_{ij} = \infty$. The average path length l is the average distance between all pairs of nodes.

Networks usually contain distinguishable groups of nodes, named modules or communities. The nodes within a module are densely connected to each other, but less likely connected to nodes in other modules. In a similar manner in which nodes group into modules, the modules can also join to form larger modules, giving rise to hierarchically nested structures.

3. Anatomy of brain networks

The nervous system of mammals forms a vast complex network with different scales of organization [14]. At the microscopic level, we find individual neurons and their local neighbours forming ordered structures such as layers and micro columns. At the other extreme, we find the large components of the brain: brainstem, thalamus, hippocampus, cerebral cortex, cerebellum, etc. These components are interconnected with each other. Uncovering those patterns of local and global interconnectivity and their functional influence on the working brain are very relevant questions that we are starting to understand. For example, the diversity in the microstructure (cytoarchitecture) of different brain parts may respond to the kind of specialized signal processing they perform [15], and the large-scale connectivity may support the intrinsic necessity of the brain to simultaneously segregate and integrate information [16,17]. While some parts

of the brain extract features of the sensory input, perception and awareness emerge from the combination of those features [18,19]. In this section, we will review recent discoveries on the neural connectivity and their potential functional implications. We will focus on data that have been studied through graph analysis.

(*a*) *Architecture of corticocortical connectivity*

By tracing the propagation of dyes injected into the brain, the axonal fibres between different regions can be revealed. Compilation of reported tract-tracing experiments led, in the early 1990s, to connectivity matrices between cortical areas in the brains of cats [20] and macaque monkeys [21,22]. The invasive and toxic nature of these experimental techniques makes them unsuitable for application in humans. Large-scale connectivity in the human brain might be obtained through non-invasive imaging techniques [23]. Although the reliability of these techniques to achieve accurate maps is difficult to evaluate, they serve as an initial draft of the human large-scale connectivity.

Recent analysis of the corticocortical networks of the cat and macaque have led to the discovery of striking characteristics of functional relevance. In comparison to most natural networks analysed in the literature of complex networks, the corticocortical networks are densely connected. In the cortex of the cat, 30 per cent of the pairs of areas are connected by direct links, and 60 per cent are separated by only two processing steps [24]. As a consequence, the information processed in any cortical area is highly accessible to other areas. Cortical networks are referred to as *small-world networks* because of this closeness and accessibility between cortical areas [25–27]. Additionally, information can flow through many alternative paths, which significantly enhances the richness and complexity of the information processing capabilities with a limited set of resources [24]. These observations support the idea that the cortex is a highly cooperative and interactive system, in contrast to the generally accepted conception that regards cortical function as a collection of functional regions, each highly specialized and rather independent of the processing of other areas.

The regions in the cortex of the macaque and cat are arranged into a few distinguishable modules [25,28]. The areas in one module are more frequently connected with each other than with areas in the other modules. Additionally, the modules follow functional subdivisions, i.e. they contain areas known to be specialized in the processing of information of the same modality: visual, auditory, somatosensory, motor and frontolimbic. This modular organization represents the partial segregation of information in the cortex and permits that sensory information of different modalities be processed in parallel, i.e. simultaneously by different parts of the cortex. On the other hand, the emergence of a coordinated perception and awareness requires that the multi-sensory information is combined (integrated). How and where this happens is still a source of debate. Further analysis of these corticocortical networks has shed new light onto the discussion. In the cortical network of the cat, a small set of areas have been found that could be responsible for such multi-sensory integration [16,24,29,30]. These cortical hubs have both afferent and efferent projections to areas in all the modalities and they are densely interconnected with each other, forming a module of their own on top of the hierarchy of the cortical network [17,31]. This observation strongly suggests that while sensory information is processed by the modally related

modules, it is the function of the cortical hubs to collect and cooperatively process the information of different modalities, giving rise to multi-sensory integration necessary for perception and awareness. An important observation is that, while areas of similar functions are usually located close to each other in the cortical surface, the hubs are dispersed, forming a delocalized module, only detectable through the analysis of the large-scale connectivity.

(b) Complexity of neural and brain network dynamics

The neural networks serve as the physical substrate in which the dynamical activity and the information processing of a working brain happens. On the other hand, dynamical activity can influence the architecture of the network by processes of learning and evolutionary adaptation. Hence, the architecture of cortical networks described above is the evolutionary consequence of the interplay between the functionality they can support and the functional necessities of the brain. Modelling of neural activity helps explore the range of dynamical behaviours that are possible within a given network. Neurones and cortical areas are simulated by mathematical neuronal models, and the complexity of the resulting network dynamics is quantified. Such models simulate the neuronal activity in the absence of external stimulation and can be considered as models of the 'resting-state' activity [16,29,30]. Performing the simulations require adequate selection of models for the neuronal dynamics and for the synaptic binding, otherwise the resulting dynamics could be insensitive to the underlying network topology [32].

In the following, we simulate the neuronal network of the worm *Caenoharbditis elegans* and the corticocortical network of the cat. The nervous system of the *C. elegans* consists of approximately 270 neurons and 3000 synapses and electrical junctions between them [33–35]. We simulate its neurons by noisy Hodgkin–Huxley neurons [36] and the cortical areas of the cat by neural-mass models [37,38]. The time series obtained for each node during the simulation are used to quantify their dynamical influence. In particular, we compute the pairwise cross-correlation R_{ij} and summarize the results into a correlation matrix, figure 1.

Given the distribution of the correlation values $P(R_{ij})$, we characterize the complexity of the dynamical influences as the Shannon entropy of this distribution,

$$S = -\frac{1}{S_m} \sum_{l=1}^{m} P_l \log P_l, \tag{3.1}$$

where $S_m = \log m$ and $1/S_m$ is a normalization factor for the number m of bins used to compute the distribution. By increasing the coupling strength between the nodes, the network dynamics move from non-synchronization to complete synchronization [39]. In this process, the distribution $P(R_{ij})$ shifts from a very narrow peak close to $\langle R_{ij} \rangle = 0$, to another narrow peak near $\langle R_{ij} \rangle = 1$. For intermediate values of the coupling, the network dynamics display a balance between partial independence and partial synchronization. In this regime, the values R_{ij} are widely distributed, reflecting that the various levels of synchronization and complexity are maximal.
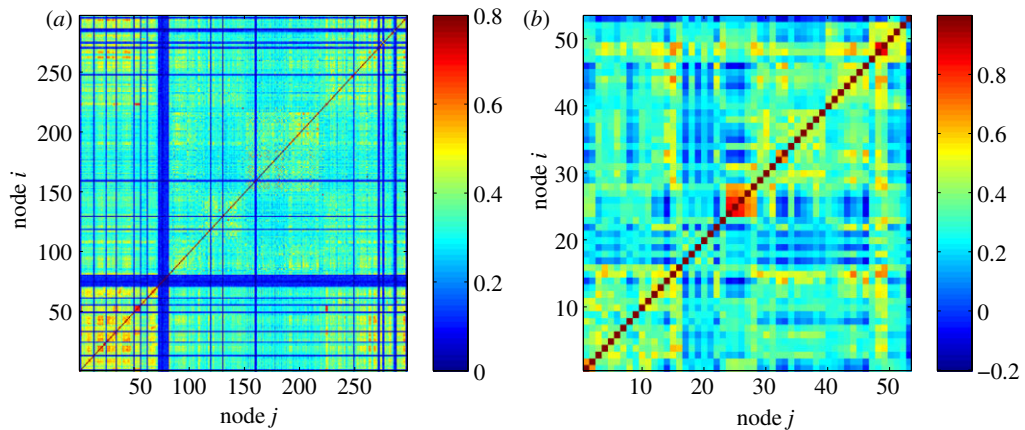
Figure 1. Correlation matrices of (*a*) the neural network of the *C. elegans* and (*b*) the corticocortical network of the cat. (Online version in colour.)

In the literature, this balance has been quantified by a measure of 'neural complexity' [40] that relies on information theoretical approaches. This measure is suitable only for very small networks because it requires all the bipartitions of the network for sizes $n = 2, 3, \ldots, N$, where N is the number of nodes, to be found. Zhao *et al.* [41] have shown that the definition of complexity in equation (3.1) is consistent with neural complexity and can be very easily calculated for networks of any size.

From the viewpoint of information processing, the regime of high complexity is of particular interest. It captures the dynamical regime at which specialized processing within the modules coexists with the capacity of the network to integrate global information. It is thus natural to speculate whether the neural network topologies have been optimized in the course of evolution to satisfy these dynamical demands. We test this speculation by studying the variation of the complexity measure S after perturbation of the network architecture. Therefore, we perform a series of modifications in the network that aim to (i) eliminate the integrative influence of the hubs and to (ii) decrease the capacity of the network to segregate dynamics. See Zhao *et al.* [41] for further details.

Type I. We rewire the network to eliminate the integrative properties of the hubs. First, we break the community of hubs by rewiring the links between 20 per cent nodes with the largest degrees while conserving their input and output degrees. After this operation, the hubs no longer form a module. Second, we rewire to randomize only the output links of the hubs while maintaining their input degrees. Third, we randomize both input and output links, and the hubs are all eliminated. In the rewiring process, the number of connections within and between the modules and their direction are preserved. A link pointing from community A to community B is rewired such that, after the rewiring, it still points from A to B. As a consequence, the topological modularity is maintained, but the community of hubs, and the hubs, are destroyed. Results are shown in figure 2*a,b*.
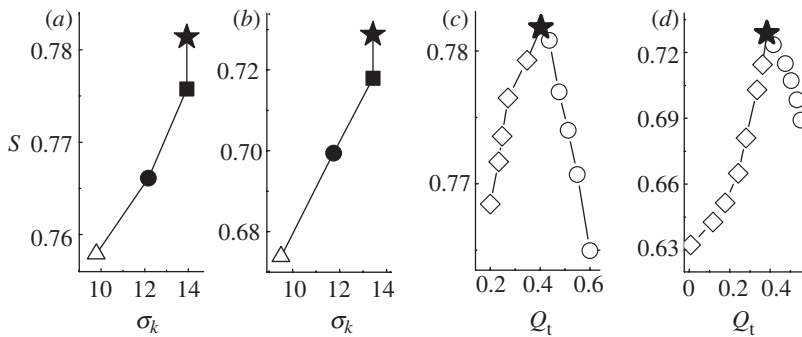
Figure 2. Complexity of the neural and corticocortical networks (filled star) compared with the rewired versions. (a,b) Type I rewiring. Complexity S is shown as a function of the standard deviation σ_k of the node degree in the network. Breaking the community of hubs, filled black square; destroying in-hubs, filled circles; destroying in- and out-hubs, open triangle. (c,d) Type II rewiring. S is shown as a function of topological modularity Q_t. Decreasing modularity, open diamonds; increasing modularity, open circle. (a,c), *C. elegans*; (b,d), cat.

Type II. We rewire the network to modify its modular organization. To reduce the modularity, links are randomly rewired. To enhance the modular architecture, the inter-community links are rewired into the communities. For example, in the cat cortical network, a link pointing from a node in the visual module to a node in the frontolimbic module is redirected such that it points to another node in the visual module. Results are shown in figure 2c,d, as the variation of complexity S depending on the modularity Q_t of the rewired networks.

In both networks, application of all the rewiring procedures proposed have the effect of reducing the dynamical complexity, figure 2. These results confirm that the current modular and hierarchical topology of the two networks are optimal in terms of our complexity.

In this section, we have summarized the topological characteristics of neural networks from a large-scale perspective. We have highlighted the importance of their modular and hierarchical architecture as serving the functional necessities of the nervous system to simultaneously process different modalities of sensory information (specialization) and to integrate that information. By perturbing the corticocortical network of the cat and the neural network of the *C. elegans*, we have shown that evolution might have shaped neural and brain connectivity to optimize dynamical complexity, supporting a balance between integration and segregation.

4. Memory networks

The results reviewed in the previous section focus on the large-scale organization of the brain and its capacity to work with sensory information from different modalities. When studying particular functions of the brain, e.g. cognitive function, one has to concentrate on smaller scales of description. The fundamental role of the prefrontal cortex for cognitive processing has been widely documented, in as much as the posterior cortex is important in the processing of sensory input.

Regardless of the general principles of organization exposed above, the presence of specialized but interconnected regions forming distributed networks of interaction could be also valid across many neural scales. Studies of brain damage, as well as functional imaging studies of healthy subjects have evidenced the existence of dissociable neural systems that are specialized in representing knowledge of different conceptual domains [42–44]: living or non-living things, man-made tools, vegetables, animals, etc. On the other hand, micro-electrode recordings have evidenced that memories are physically stored involving distributed and overlapping networks of neurons, where individual neurons participate in more than one memory [9,12,45,46].

At the mental level, memories are not isolated entities. Rather, they are always embedded into broader contextual frameworks: let them be historical, causal–effect relations, shared conceptual categories, etc. Memories may also be associated with a word, with an image, with a sound, etc. The conceptual memory of an apple is associated to a characteristic internal image of an apple, to the word 'apple' and its phonetical sound. But it can also be associated with the apple tree from which apples grow, with the grocery store we find them, with apple pies or any other personal experience where apples have been involved. This multiple contextual forms of associations, and the variability of personal experiences make it very difficult to systematically study the architecture of the mind. But there are windows through which we can look at it, for example, the language. Languages are mental constructions with their own rules and structures, and they are shared by large human populations. These characteristics, and restricting to those conceptual memories representable by words, permit us to statistically describe the interrelations between memories.

There are several approaches to construct linguistic networks, see Borge-Holthoefer & Arenas [6] for further details. (i) In text analysis, the relation between words is defined as the times they co-occur in the same sentence or paragraph and the relative position they take, e.g. whether they appear one after the other or before. Networks of co-occurrence and co-location are extracted from text corpuses and they manifest lexical semantic affinities beyond grammatical restrictions. (ii) Dictionaries provide another source for defining the relations between words, connecting them if they are synonyms, antonyms, etc. Note that these semantic relationships are the product of deliberate classifications by lexicographer expertise. (iii) Feature production norms are constructed by asking groups of participants about the characteristics of objects and concepts. For example, an apple is a fruit, it can be round, it can be red, it can taste sweet or sour, but it can neither be friendly nor rude. The relationship between words is thus defined as the number of common features they share. (iv) Finally, networks of free association are constructed by asking participants to provide the first word that comes to their mind after a cue word has been presented. Such networks are believed to capture principles of memory organization because the responses of the subjects arise from a free navigation of their mental space. Free-association networks have been collected for English [47], French [48], Spanish [49,50] and German (Mellinger & Weber, http://www.coli.uni-saarland.de/projects/nag/).

Despite the different constructions, all linguistic networks present three main features [6,7]. First, they are sparse. Second, they all display small-world properties [51,52] as encountered in other networked real and technological

systems. This property means that these networks are well communicated: it is easy to reach any given element from another one through very short paths. Patients of Alzheimer's disease display consistently larger path lengths in their linguistic networks, revealing a loss of efficiency [53]. And third, language networks are highly heterogeneous in their degree distribution with near to scale-free decays, although some of them decay exponentially [54]. In co-occurrence networks, the hub words correspond to functional words such as articles and prepositions: *a*, *and*, *for*, *of*, etc. Polysemic words also appear as highly connected.

Although the networked and statistical description of language is relevant *per se*, to answer questions related to linguistic evolution, formation and learning, it also helps to explore cognitive implications. From the standpoint of retrieval of information, the small-world property represents a maximization of efficiency: high clustering arises from the gathering of similar pieces of information, low distances allow for fast search and retrieval. It strikes a balance between the number of active connections and the number of steps required to access any node. Community analysis of the free-association networks [6] reveals that words are mentally arranged into well-defined communities. But these communities poorly correlate with semantic classification of words, they contain words of distinct semantic categories. This striking observation implies that, although memories could be stored in the cortex following certain categorical relations as reviewed above, the mental navigation processes are influenced by other relations between memories beyond the semantic categories.

Several models have tried to capture the topological characteristics of language. Most of them concentrate on the nature of language growth, learning and evolution [55–57]. Shifting towards a cognitive interest, Borge-Holthoefer & Arenas [58] studied the relationship between free associations and the semantic categories through simple models of mental navigation. In the present paper, it is our aim to glimpse beyond the purely linguistic interest and motivate lines of research towards the convergence of neuroanatomy and cognitive science, through the description of the structure of the brain and the storage of memories at the physical level, and the organization of the memory associations at the mental level. In the following, we present a model that simulates spontaneous transitions between memories. The model consists of a group of interconnected Potts units representing the physical connectivity. The system is entrained to recognize a set of memories whose retrieval is characterized by the dynamics falling into a particular attractor state. In contrast to other models of machine learning, here the states are made unstable, permitting the system to jump from one memory to another, a process that is comparable to the experimental free-association studies.

(*a*) *A model of memory retrieval and association*

The model was inspired by the work of Braitenberg [59], who noted that despite local peculiarities, the cortex is characterized by the coexistence of two distinct interacting systems: one with local and another with long-range connections. Treves characterizes this distinction by choosing the Potts unit as the elementary constituent of his network [60]. Instead of representing a single neurone, each of these units stands for a local patch of cortex. Thus, the connection of the Potts units results in a system that can be regarded as a global network of local

subnetworks. Further inspired by Braitenberg & Schüz [61], Treves proposed a model of the cortex that resembles an associative memory machine. An auto-associative network is characterized by recurrent collaterals and it has the ability to store, as attractors, a finite number of memories. In this kind of network, a memory is associated with a particular activity configuration of the whole network, and it corresponds to local minima of the system's energy landscape [62]. After entraining, when partial versions of the memories are presented, the network dynamics are spontaneously attracted towards the activity pattern associated with the configuration of the original memory.

The model is thought of as a network of local subnetworks. Being auto-associative, both the global and the local networks are loaded with global and local stored memories, respectively. In the model, the local subnetworks are represented by linear Potts units. Each unit takes values from 0 to 1 along a fixed number S of directions (active states) and in one inactive state. Each active state labels a local memory and the associated activation value represents the correlation of the subnetwork configuration with that particular local memory. The inactive state, instead, measures the inverse of the whole activity of the subnetwork. If a patch of the cortex is not involved in the representation of a particular global memory, its Potts unit will be completely aligned to the inactive state. The activity vector of a Potts unit moves in an $(S+1)$-dimensional space with the only restriction that the sum of the activation values in all the states is equal to one. In a model with N Potts units, the long-range connectivity between the units is represented by the binary adjacency matrix A_{ij} of size $N \times N$. The detailed configuration giving rise to the memories is encoded into the weighted tensor J_{ij}^{kl} where indexes $i, j = 1, \ldots, N$ are the Potts units and the indexes $k, l = 1, \ldots, S + 1$ are the states of each unit. By acting on the J matrix, it is possible to shape the energy landscape in which the system moves, determining automatically the global memories.

To simulate a dynamical process of free association by spontaneous transitions between memories, the attractors related to the memories need to become unstable. Therefore, Treves [60] introduced an adaptive process into the model, which generates instabilities. Each Potts unit activates according to the incoming field produced by its neighbouring units. An adaptive threshold in the field facilitates the spontaneous activation of silent units and obstructs the firing of active units. After the retrieval of a memory, all active units increasingly feel fatigue effects and tend to stop firing. As a consequence, the activity of the network switches to a different configuration by falling into another attractor state, and another memory is retrieved. As shown in Russo *et al.* [63], this transition is controlled by the correlation between the physical representation of the two memories in the cortex. The activation of units shared by the initial and the second memory act as a cue for the transition. Treves calls this process a *latching transition.*

Figure 3 shows an example of spontaneous latching dynamics. At time 0, an external field drives the system towards one of the stored memories, randomly chosen. After the correct retrieval of the pattern, the system adapts and makes a transition to a new attractor. This new retrieved memory adapts in its turn and acts as a cue for a new transition, and so on. In this manner, a sequence of associations between memories spontaneously emerges.
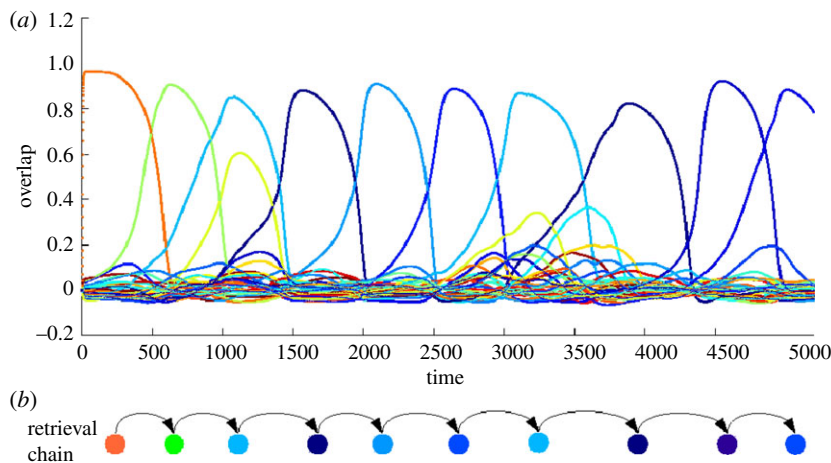
Figure 3. Latching chain: an example of a sequence of spontaneous memories retrievals after a first cued recall. (*a*) The behaviour of the overlap of the system state with the stored memories is represented in different colours and (*b*) the sequence of retrieved memories. (Online version in colour.)

(*b*) *Simulated networks of free association*

For illustrative purposes, we consider here a simple long-range connectivity, each Potts unit is connected with all other units. This pattern of connections can be easily replaced by more biologically realistic topologies. We entrain the network to recognize $p = 80$ memories, each encoded by an N-dimensional configuration vector ξ^{μ}, where $\mu = 1, \ldots, p$. The vectors are not completely random but share some degree of correlation between them in order to facilitate the spontaneous transitions. We run p^2 simulations with a duration of 30 latching transitions each. Every simulation starts at a randomly chosen memory. We count the number of times that the system jumps from one memory μ to another ν and summarize the results into the transition probability matrix, $P_{\mu\nu}$. Its entries are normalized such that the row sum $\sum_{\nu=1}^{N} P_{\mu\nu} = 1$ for all μ because the retrieval of a memory is always followed by a transition into another memory. Hence, $P_{\mu\nu} = 1$ only if the system always switches to ν after memory μ has been active. The transition probability matrix resembles the matrices of free association collected in linguistic experiments and is considered in the following as the weighted adjacency matrix of the network of associations between the p memories.

The network emerging out of the simulations, the $P_{\mu\nu}$ matrix, contains $L = 1109$ weighted links with values ranging from 0 to 1 due to the normalization. The network is not fully connected because many memory pairs were not accessible to each other by a direct transition. Graph analysis reveals that the latching process has generated a complex network of free associations. The degree of the memories is very inhomogeneous and ranges from 1 to 35 as shown in figure 4*a,b*. Many memories are associated with a few others, and some memories are associated with up to 43 per cent of the memories. These latter can be considered as hub memories. The network displays small-world characteristics, also observed in the empirical networks of free association. Its average path length is only
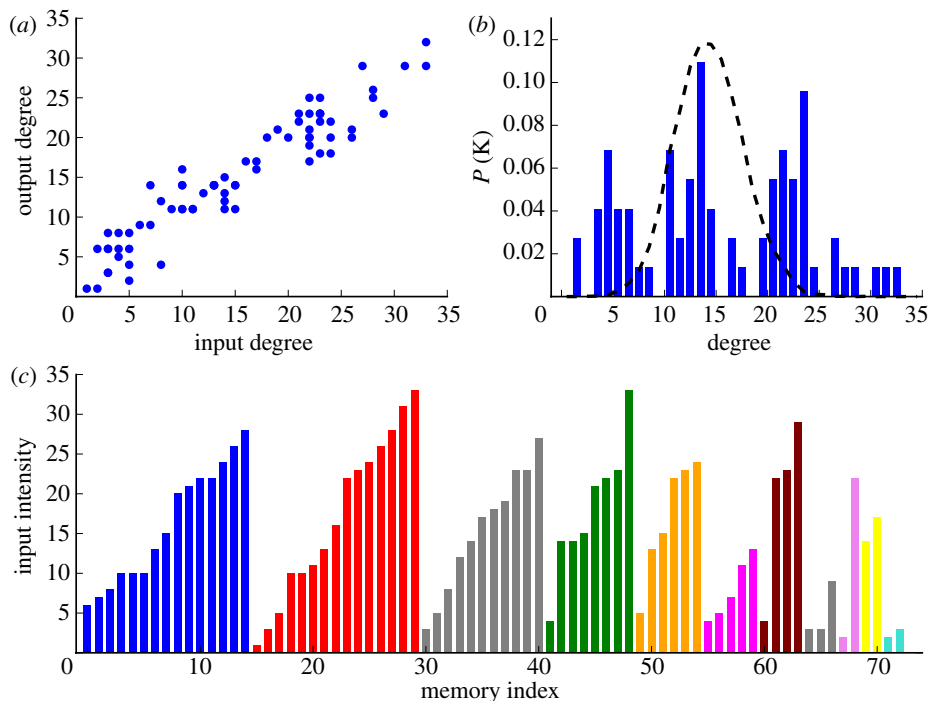
Figure 4. Network properties of the simulated free associations. (*a*) Input and output degrees of the $N = 80$ memories reflect that memories tend to reciprocally activate each other. (*b*) Degree distribution of the memories (bars) and degree distribution of equivalent random networks (dashed line). (*c*) Community analysis of the free-association network reveals 11 communities. All modules contain nodes of different centrality, revealing internal hierarchies of memories. (Online version in colour.)

$l = 2.00$, very close to the value of equivalent random networks ($l_{\mathrm{rand}} = 1.87$). It also displays a large clustering coefficient of $C = 0.41$, which is 2.3 times larger than the clustering of random networks.

We have performed a community analysis by use of the RADATOOLS software, http://deim.urv.cat/~sgomez/radatools.php. Considering the weighted links, a tabu search [64] followed by a refinement procedure by reposition detects 11 communities with a weighted modularity of $Q = 0.484$. Four of the modules contain only two or three nodes. The modular organization is followed by a hierarchical ranking of the memories. Every module contains memories of dissimilar importance, figure 4*c*, where importance is quantified here by the column sum of the elements in the P matrix. In terms of graph theory, this measure is known as the input intensity of the node $s_{\mathrm{in}}(\mu) = \sum_{\nu=1}^{N} P_{\nu\mu}$. In this particular system, it is proportional to the frequency with which a memory is retrieved.

(*c*) *Effect of memory correlation on the pattern of free association*

As described above, the connectivity between the Potts units is a simple all-to-all connectivity. Hence, the emergence of a complex network of memory

associations cannot be attributed to an underlying networked substrate. However, in order to facilitate the transitions, we have entrained the system with correlated memories. Conceptually speaking, these correlations can be regarded as an external relatedness between the memories, for example, as if the corpus used for a free-association experiment would be restricted to words of the same semantic class. At the level of the Potts network, these initial correlations between the memories result in structural correlations during the entrainment process, resembling the similarity in cortical representation of two memories.

We quantify the correlation between the memory vectors ξ^μ as the overlap of their elements and summarize the results into a matrix,

$$K_{\mu\nu} = \frac{1}{N} \sum_i^N \delta(\xi_i^\mu, \xi_i^\nu). \tag{4.1}$$

All vectors share an intermediate or high correlation with $K_{\mu\nu}$ values, between 0.55 and 0.72, figure 5a. The question is then, whether the complexity observed in the network of associations is a trivial consequence of these initial correlations, or it emerges from the dynamical process. To explore this possibility, we compare the total correlation of the memories in the K matrix (row sum) to their degree in the P adjacency matrix, figure 5b. We find that there is only a weak correlation between the two quantities, hence, the degree of association achieved by the memories is not a trivial consequence of their initial correlation. We can extract information about the dynamics from the global structure of the network but not from a pairwise comparison between couples of memories. Deviations from this trend are probably due to a history effect. Indeed, it is shown in Russo *et al.* [65] that the transition probability between two memories is also affected by the associative path followed by the system.

A community analysis of the K matrix detects three communities, although their modularity is much lower, $Q = 0.206$, than for the communities found in the network of simulated associations. To find out whether the modules encountered in the two cases are related, we compute their cross-overlap, quantified as the number of common nodes. The result, as shown in figure 5c, evidences that the communities of the K matrix serve as seeds for the emerging communities of associations, but do not fully determine them. The communities 1 and 2 of the P matrix largely conserve the composition of the communities II and III of the K matrix, but the nodes of community I end up widely dispersed across the communities of the P network.

In summary, we have found out that the model presented here generates free-association networks of artificial memories that share fundamental topological similarities to experimentally acquired free-association networks. In this illustrative case, the Potts units were all-to-all connected. However, the latching process is capable of producing a complex network of memory associations, beyond the organization imposed by the initial intrinsic correlations between the memory vectors. The degree of association achieved by two memories is not a trivial consequence of their initial correlation, but is influenced by the presence of the other memories, sharing the physical space in which they are stored. A pairwise comparison does not provide any direct understanding, but the network approaches do reveal large-scale relationships.
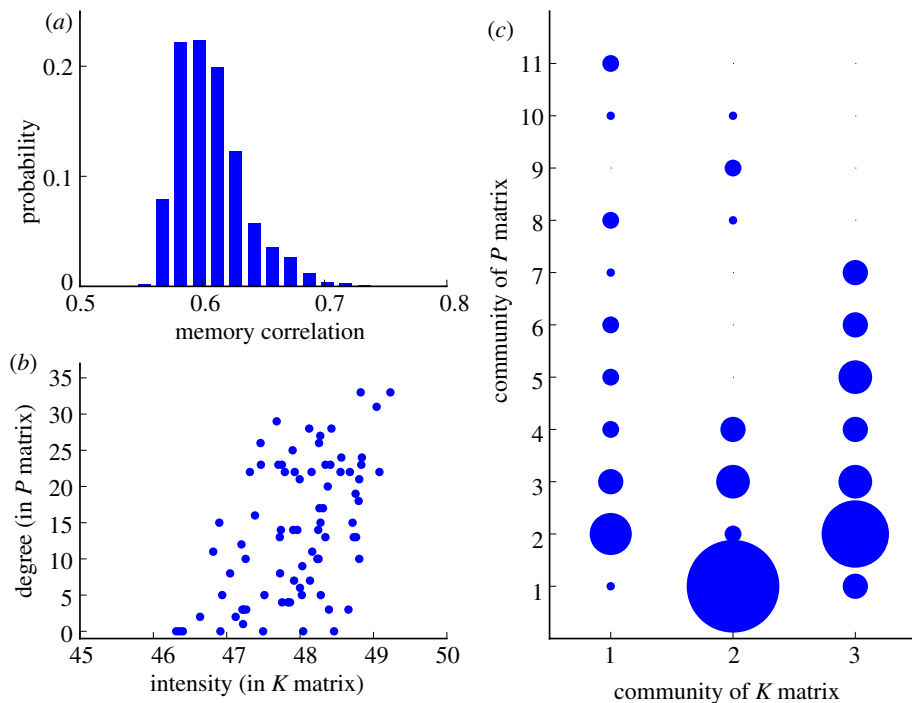
Figure 5. Relationship between the initial correlation of the memories and the emerging associations at the simulated network. (*a*) Distribution of the initial correlation between the 80 entrained memories. (*b*) Comparison of the total correlation (row sum) of the memories and their subsequent degree at the simulated network. (*c*) The overlap (number of common nodes) between the three communities out of the $K_{\mu\nu}$ matrix and the 11 communities detected in the $P_{\mu\nu}$ network. Size of the circles is proportional to the number of common nodes. (Online version in colour.)

5. Summary and discussion

This paper is motivated by the old idea that the associations between mental representations (either sensory and cognitive) are intrinsically related with the manner in which memories are physically stored and dynamically retrieved in the brain. Uncovering the organization of both the brain and mind is a difficult experimental quest, but the last two decades have seen several advances on both sides, particularly, since concepts and tools of complex networks have been applied to study their organization. Here, we have reviewed those findings and concluded that both neural and language-related networks share relevant common properties: (i) a broad degree distribution containing hubs, (ii) small-world properties, and (iii) an organization into modular and hierarchical structures.

On the other hand, we have outlined how network models serve to understand the relationship between the structure and the function of the brain. In §3, we have presented a simulation of the resting-state dynamics in the corticocortical network of cats and in the neural network of the nematode *C. elegans*. By targeted manipulation of their topology, we observe that their dynamical behaviour is largely affected. This indicates that evolutive adaptation might have shaped the

organization of these networks to favour their capacity to host complex dynamical processes, in particular their ability to segregate and integrate information of different sensory modalities.

In §4, we have revisited a model of memory storage and pattern recognition proposed by Treves [60]. This model is inspired by the combination of local and long-range projections found in corticocortical connectivity. It generates a chain of spontaneous transitions between memories, and hence, emulates processes of semantic free association. The simulated network of associations displays topological similarities to experimental free-association networks: a broad degree distribution, small-world characteristics and a modular and hierarchical organization. In our illustrative example, we have considered that the underlying physical substrate, the network of Potts units, is all-to-all connected. Future implementations should include more realistic biological topologies as those reviewed in §3. Other models of learning machines have explored the impact of the underlying network topology [66–69]. However, their goal is restricted to evaluate the performance of artificial neural networks to recognize learned patterns. The approach presented here looks forward, to using machine-learning models to simulate simple cognitive tasks and to compare them with the outcome of real-life experiments with human subjects, at least at a statistical level.

In summary, we have reviewed experimental observations on the organization of both brain and mind networks, and we have stressed their similarities. Further, we have outlined some of the recent modelling approaches that aim to bridge the gap between the structure and function of the brain. The reviewed results indicate that we are at the beginning of fascinating times in which the paths of neuroanatomy and cognitive science start to converge.

References

1 Newman, M. E. J. 2003 The structure and function of complex networks. *SIAM Rev.* **45**, 167–256. (doi:10.1137/S003614450342480)

2 Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. 2006 Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308. (doi:10.1016/j.physrep.2005.10.009)

3 Costa, L., Rodrigues, F. A., Travieso, G. & Villas-Boas, P. R. 2007 Characterization of complex networks: a survey of measurements. *Adv. Phys.* **56**, 167–242. (doi:10.1080/00018730601170527)

4 Kaiser, M. 2007 Brain architecture: a design for natural computation. *Phil. Trans. R. Soc. A* **365**, 3033–3045. (doi:10.1098/rsta.2007.0007)

5 Bullmore, E. & Sporns, O. 2009 Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198. (doi:10.1038/nrn2575)

6 Borge-Holthoefer, J. & Arenas, A. 2010 Semantic networks: structure and dynamics. *Entropy* **12**, 1264–1302. (doi:10.3390/e12051264)

7 Solé, R. V., Corominas-Murtra, B., Valverde, S. & Steels, L. 2010 Language networks: their structure, function, and evolution. *Complexity* **15**, 20–26. (doi:10.1002/cplx.20305)

8 Luria, A. R. 1966 *Higher cortical functions in man.* New York, NY: Basic Books.

9 Goldman-Rakic, P. S. 1988 Topography of cognition: parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci.* **11**, 137–156. (doi:10.1146/annurev.ne.11. 030188.001033)

10 Damasio, A. R. 1989 The brain binds entities and events by multiregional activation from convergence zones. *Neural Comp.* **1**, 123–132. (doi:10.1162/neco.1989.1.1.123)

11 van Essen, D. C., Anderson, C. H. & Felleman, D. J. 1992 Information processing in the primate visual system: an integrated systems perspective. *Science* **255**, 419–423. (doi:10.1126/ science.1734518)

12 Bressler, S. L. 1995 Large-scale cortical networks and cognition. *Brain Res. Rev.* **20**, 288–304. (doi:10.1016/0165-0173(94)00016-I)

13 Fuster, J. M. 2003 *Cortex and mind: unifying cognition.* New York, NY: Oxford University Press.

14 Sporns, O., Tononi, G. & Kötter, R. 2005 The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* **1**, e42. (doi:10.1371/journal.pcbi.0010042)

15 Hilgetag, C.-C. & Grant, S. 2010 Cytoarchitectural differences are a key determinant of laminar projection origins in the visual cortex. *Neuroimage* **51**, 1006–1017. (doi:10.1016/j.neuroimage. 2010.03.006)

16 Sporns, O., Honey, C. J. & Kötter, R. 2007 Identification and classification of hubs in brain networks. *PLoS ONE* **10**, e1049. (doi:10.1371/journal.pone.0001049)

17 Zamora-López, G., Zhou, C. S. & Kurths, J. 2010 Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.* **4**, 1.

18 Singer, W. 2001 Consciousness and the binding problem. *Ann. N.Y. Acad. Sci.* **929**, 123–146. (doi:10.1111/j.1749-6632.2001.tb05712.x)

19 Engel, A. K. & Singer, W. 2001 Temporal binding and neural correlates of sensory awareness. *Trends Cogn. Sci.* **5**, 16–25. (doi:10.1016/S1364-6613(00)01568-0)

20 Scannell, J. W. & Young, M. P. 1993 The connectional organization of neural systems in the cat cerebral cortex. *Curr. Biol.* **3**, 191–200. (doi:10.1016/0960-9822(93)90331-H)

21 Young, M. P. 1993 The organization of neural systems in the primate cerebral cortex. *Proc. R. Soc. Lond. B* **252**, 13–18. (doi:10.1098/rspb.1993.0040)

22 Stephan, K. E., Kamper, L., Bozkurt, A., Burns, G. A. P. C., Young, M. P. & Kötter, R. 2001 Advance database methodology for the collation of connectivity data on the macaque brain. *Phil. Trans. R. Soc. Lond. B* **356**, 1159–1186. (doi:10.1098/rstb.2001.0908)

23 Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J. & Sporns, O. 2008 Mapping the structural core of human cerebral cortex. *PLOS Biol.* **6**, e159. (doi:10.1371/ journal.pbio.0060159)

24 Zamora-López, G., Zhou, C. S. & Kurths, J. 2009 Graph analysis of cortical networks reveals complex anatomical communication substrate. *Chaos* **19**, 015117. (doi:10.1063/1.3089559)

25 Hilgetag, C.-C., Burns, G. A. P. C., O'Neill, M. A., Scannell, J. W. & Young, M. P. 2000 Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Phil. Trans. R. Soc. Lond. B* **355**, 91–110. (doi:10.1098/rstb.2000.0551)

26 Sporns, O. & Zwi, J. D. 2004 The small world of the cerebral cortex. *Neuroinformatics* **2**, 145–162. (doi:10.1385/NI:2:2:145)

27 Sporns, O., Chialvo, D. R., Kaiser, M. & Hilgetag, C.-C. 2004 Organization, development and function of complex brain networks. *Trends Cogn. Sci.* **8**, 418–425. (doi:10.1016/j.tics.2004. 07.008)

28 Hilgetag, C. C. & Kaiser, M. 2004 Clustered organisation of cortical connectivity. *Neuroinformation* **2**, 353–360. (doi:10.1385/NI:2:3:353)

29 Zemanová, L., Zhou, C. S. & Kurths, J. 2006 Structural and functional clusters of complex brain networks. *Physica D* **224**, 202–212. (doi:10.1016/j.physd.2006.09.008)

30 Zhou, C. S., Zemanová, L., Zamora-López, G., Hilgetag, C.-C. & Kurths, J. 2006 Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.* **97**, 238103. (doi:10.1103/PhysRevLett.97.238103)

31 Zamora-López, G., Zhou, C. S. & Kurths, J. 2011 Exploring brain function from anatomical connectivity. *Front. Neurosci.* **5**, 83. (doi:10.3389/fnins.2011.00083)

32 Schmidt, G., Zamora-López, G., Zhou, C. & Kurths, J. 2010 Simulation of large scale cortical networks by individual neuron dynamics. *Int. J. Bifurc. Chaos* **20**, 859–867. (doi:10.1142/S02181 27410026149)

33 White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. 1986 The structure of the nervous system of the nematode *Caenorhabditis elegans. Phil. Trans. R. Soc. Lond. B* **314**, 1–340. (doi:10.1098/rstb.1986.0056)

34 Durban, R. M. 1987 Studies on the developmnet and organisation of the nervous system of *Caernorhabditis Elegans.* PhD thesis, King's College, Oxford, UK

35 Varshney, L. A., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. 2011 Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066. (doi:10.1371/journal.pcbi.1001066)

36 Hodgkin, A. L. & Huxley, A. F. 1952 A quantitative description of membrane current and its application to conduction and excitation in nerve cells. *J. Physiol.* **117**, 500–544.

37 Lopes da Silva, F. H., Hoeks, A., Smits, H. & Zetterberg, L. H. 1974 Model of brain rythmic activity. *Kybernetik* **15**, 27–37. (doi:10.1007/BF00270757)

38 Wendling, F., Bellanger, J. J., Bartolomei, F. & Chauvel, P. 2000 Relevance of nonlinear lumped-parameter models in the analysis of depth-eeg epiletic signals. *Biol. Cybern.* **83**, 367–378. (doi:10.1007/s004220000160)

39 Gómez-Gardeñes, J., Zamora-López, G. Moreno, Y. & Arenas, A. 2010 From modular to centralized organization of synchronization in functional areas of the cat cerebral cortex. *PLoS ONE* **5**, e12313. (doi:10.1371/journal.pone.0012313)

40 Tononi, G., Sporns, O. & Edelman, G. M. 1994 A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl Acad. Sci. USA* **91**, 5033–5037. (doi:10.1073/pnas.91.11.5033)

41 Zhao, M., Zhou, C., Chen, Y., Hu, B. & Wang, B.-H. 2010 Complexity versus modularity and heterogeneity in oscillatory networks: combining segregation and integration in neural systems. *Phys Rev. E* **82**, 046225. (doi:10.1103/PhysRevE.82.046225)

42 Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. 1996 Neural correlates of category-specific knowledge. *Nature* **379**, 649–652. (doi:10.1038/379649a0)

43 Caramazza, A. & Mahon, B. Z. 2006 The organisation of conceptual knowledge in the brain: the future's past and some future directions. *Cogn. Neuropsychol.* **23**, 13–38. (doi:10.1080/0264329 0542000021)

44 Martin, A. 2006 The representation of object concepts in the brain. *Annu. Rev. Psychol.* **58**, 25–45. (doi:10.1146/annurev.psych.57.102904.190143)

45 Fuster, J. M. 2009 Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosc.* **21**, 2047–2072. (doi:10.1162/jocn.2009.21280)

46 Binder, J. R., Desai, R. H., Gaves, W. W. & Conant, L. L. 2009 Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796. (doi:10.1093/cercor/bhp055)

47 Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. 1998 The University of South Florida word association, rhyme, and word fragment norms. *Behav. Res. Meth.* **36**, 402–407.

48 Ferrand, L. & Alario, F. X. 1998 Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année Psychol.* **98**, 531–552.

49 Callejas, A., Correa, A., Lupiáñez, J. & Tudela, P. 2003 Normas asociativas intracategoriales para 612 palabras de seis categorías semánticas en español. *Psicológica* **24**, 185–214.

50 Fernández, A., Díez, E., Alonso, M. & Beato, M. 2004 Free-association norms for the Spanish names of the snodgrass and vanderwart pictures. *Behav. Res. Meths.* **36**, 577–583.

51 Steyvers, M. & Tenenbaum, J. B. 2005 Graph theoretic analyses of semantic networks: small worlds in semantic networks. *Cogn. Sci.* **29**, 41–78. (doi:10.1207/s15516709cog2901_3)

52 Ferrer i Cancho, R. & Solé, R. V. 2001 The small world of human language. *Proc. R. Soc. Lond. B* **268**, 2261–2265. (doi:10.1098/rspb.2001.1800)

53 Chan, A. S., Butters, N. & Salmon, D. P. 1997 The deterioration of semantic networks in patients with alzheimer's disease: a crosssectional study. *Neuropsychologia* **35**, 241–248. (doi:10.1016/S0028-3932(96)00067-X)

54 Ferrer i Cancho, R. & Solé, R. V. 2002 Zip's law in random texts. *Adv. Complex Syst.* **5**, 1–6. (doi:10.1142/S0219525902000468)

55 Dorogovtsev, S. N. & Mendes, J. F. F. 2001 Language as an evolving word web. *Proc. R. Soc. Lond. B* **268**, 2603–2606. (doi:10.1098/rspb.2001.1824)

56 Ferrer i Cancho, R. & Solé, R. V. 2003 Least effort and the origins of scaling in human language. *Proc. Natl Acad. Sci. USA* **100**, 788–791. (doi:10.1073/pnas.0335980100)

57 Steyvers, M. & Tenenbaum, J. B. 2005 The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn. Sci.* **29**, 41–78. (doi:10.1207/s15516709cog2901_3)

58 Borge-Holthoefer, J. & Arenas, A. 2010 Categorizing words through semantic memory navigation. *Eur. Phys. J. B* **74**, 265–270. (doi:10.1140/epjb/e2010-00058-9)

59 Braitenberg, V. 1978 Cortical architectonics: general and areal. In *Architectonics of the cerebral cortex* (eds M. A. B. Brazier & H. Petsche), pp. 443–465. New York, NY: Raven Press.

60 Treves, A. 2005 Frontal latching networks: a possible neural basis for infinite recursion. *Cogn. Neuropsychol.* **22**, 276–291. (doi:10.1080/02643290442000329)

61 Braitenberg, V. & Schüz, A. 1991 *Anatomy of the cortex: statistics and geometry.* Berlin, Germany: Springer.

62 Amit, M. A. 1989 *Modeling brain function: the world of attractor neural networks.* Cambridge, UK: Cambridge University Press.

63 Russo, E., Namboodiri, V. M. K., Treves, A. & Kropff, E. 2008 Free association transitions in models of cortical latching dynamics. *New J. Phys.* **10**, 015008. (doi:10.1088/1367-2630/10/1/015008)

64 Arenas, A., Fernández, A. & Gómez, S. 2008 Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* **10**, 05039. (doi:10.1088/1367-2630/10/5/053039)

65 Russo, E., Pirmoradian, S. & Treves, A. 2010 Associative latching dynamics vs. syntax. *Adv. Cogn. Neurodyn.* **2**, 125–129.

66 Kim, B. J. 2004 Performance of networks of artificial neurons: the role of clustering. *Phys. Rev. E* **69**, 045101. (doi:10.1103/PhysRevE.69.045101)

67 Oshima, H. & Odagaki, T. 2007 Storage capacity and retrieval time of small-world neural networks. *Phys. Rev. E* **76**, 036114. (doi:10.1103/PhysRevE.76.036114)

68 Costa, L. & Stauffer, D. 2003 Associative recall in non-randomly diluted neural networks. *Physica A* **330**, 37–45. (doi:10.1016/j.physa.2003.08.010)

69 Zhang, P. & Chen, Y. 2008 Topology and dynamics of attractor neural networks: the role of loopiness. *Physica A* **387**, 4411–4416. (doi:10.1016/j.physa.2008.02.073)