



Published in final edited form as:

Trends Cogn Sci. 2016 August ; 20(8): 629–640. doi:10.1016/j.tics.2016.06.003.

Local patterns to global architectures: Influences of network topology on human learning

Elisabeth A. Karuza^{1,*}, Sharon L. Thompson-Schill^{1,2}, and Danielle S. Bassett^{3,4}

¹Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104

²Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104

³Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104

⁴Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104

Abstract

A core question in cognitive science is how humans acquire and represent knowledge about their environments. To this end, quantitative theories of learning processes have been formalized in an attempt to explain and predict changes in brain and behavior. Here we connect statistical learning approaches in cognitive science, which are rooted in learners' sensitivity to local distributional regularities, and network science approaches to characterizing global patterns and their emergent properties. We focus on innovative work that describes how learning is influenced by the topological properties underlying sensory input. The confluence of these theoretical approaches and this recent empirical evidence motivate the importance of scaling up quantitative approaches to learning at both behavioral and neural levels.

Keywords

Complex systems; Network science; Statistical learning

Relating two approaches

From the earliest stages of development, the human brain is tasked with the monumental feat of building and efficiently accessing an enormously complex constellation of knowledge. Even the most mundane interactions with our environment require a rich understanding of its component parts as well as the scales at which they relate to form a larger system. Thus, knowledge can be represented at multiple levels, ranging from local associations between elements to complex networks built from those local associations. Until recently, a dominant approach to human learning has focused on micro-level patterns, often the pairwise

*Correspondence: Elisabeth Karuza (ekaruza@sas.upenn.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

relationships between the constituents of sensory input. In the present review, we turn our attention to exciting advances in the application of network science to the study of broader architectural patterns to which human learners are sensitive.

One source of compelling support for locally-driven learning derives from demonstrations that infants can extract words from continuous speech based on the conditional probabilities between syllables [1]. Ongoing work continues to elucidate the power of statistical relationships exploited by both infants and adults, making “statistical learning” one of the most robust and deeply explored phenomena in the field of cognitive science [2–5]. An underlying rationale has been that local associations, like the conditional probabilities that facilitate word segmentation, assist in directing the learner to component parts of a cognitive system. Knowledge of these component parts not only opens up other informative cues to structure (for a review, see [6]), but also spurs the development of sophisticated representations of dependencies between higher-order units (e.g., [7]). While evidence has thus supported a key role for local computations in complex learning environments, intriguing counter-evidence suggests that statistical bootstrapping mechanisms may be overwhelmed by real-world cognitive systems (e.g., natural language [8]; but see [9,10]). As the issues of scalability in statistical learning are as yet unresolved, we stress here the value of also considering the global network structure that emerges from pairwise relationships between constituent elements in the environment.

Under a complex systems approach, the network structure of a system is studied by determining its component elements (**nodes**) as well as the relational links between them (**edges**). Once this scaffolding is constructed, it is possible to probe large-scale topological and dynamical properties over and above those present in the pairwise relations between elements. In fact, one defining characteristic of complex systems is that the explanatory power of their global architecture exceeds that of their local architecture [11]. Network science is increasingly applied to answer questions about the structure of immensely complex information: how might we represent or navigate spatial maps [12,13], object features [14], semantic concepts [15–17], and grammatical relationships [18]? It has also been effectively harnessed by cognitive neuroscientists to examine how structural and functional connections in the brain give rise to various cognitive capacities [19–25]. Despite these many advances, the integration of network science and cognitive science has tended to focus either on (1) the description of networks derived from the sensory world; or (2) the mechanisms by which the human brain engages with the sensory world, with little cross-talk between these two areas. Here, we focus on a related, but distinct line of questioning that may begin to bridge these branches of cognitive science. Namely, how can topological properties of sensory input drive the process of human learning?

In the subsequent sections, we offer examples of complex networks present in our everyday environment, focusing particularly on descriptive analyses of language networks. Next, we detail a growing body of experimental work that links topological properties of networks to knowledge acquisition. We then discuss the intersection between distributional approaches to learning, which offer insight into the acquisition of local statistical patterns, and network-based approaches to learning, which offer complementary insight into the acquisition of higher-order patterns. Finally, we describe cutting-edge neuroimaging work that construes

the brain itself as a dynamical complex system, highlighting the importance of bridging internal network models of brain function with higher-order patterns in external networks

Complex networks are pervasive

Complex systems approaches rest on the premise, not tied to any particular domain, that the world can be decomposed into parts, and that those parts interact with one another in meaningful ways. Therefore, diverse facets of human knowledge can and have been studied under the lens of network science. Cognitive systems are generally thought to adhere to a complex network structure, a type of graph structure that is neither truly random nor truly regular [26]. Random graphs are collections of nodes that are linked by edges selected at random from a uniform distribution of all possible connections. Regular graphs are collections of nodes that share connections to the same number of neighbors, thus having equivalent **degree**. Falling between these two extremes (Figure 1), complex networks display their own set of unique properties including, but not limited to: **community structure** (nodes may pattern in densely connected groupings), skewed degree distribution (a few nodes may be densely connected, forming “hubs”), and distinctive mixing patterns (nodes may be more likely to share a link with other nodes that have either similar or dissimilar properties). As we will explore in detail in the following section, human learners are adept at exploiting topological properties such as these as they extract structure from sensory input (see *Glossary*).

In principle, network analysis of cognitive systems requires only that a given dataset be parsed into discrete elements (nodes) and that some relationships between those elements be specified (edges). In practice, this process presents a number of challenges, not least among them is determining the appropriate level of granularity of the elements or what exactly constitutes a relationship (for a thoughtful assessment of these topics, see [27]). Nevertheless, network-based approaches remain extremely powerful and have been successfully applied to a number of pressing questions. In the visual domain, complex network modeling techniques have been used to understand processes essential to scene perception such as texture and shape discrimination [28–30]. Moreover, the burgeoning field of social network analysis has offered unprecedented insight into how humans transmit information and interact with one another [31,32]. Perhaps more than any other branch of cognitive science, quantitative linguistics has adopted network-based approaches as a cornerstone methodology [33–36]. Across levels of the language hierarchy, graph theoretical methods have been applied to the study of phonological [37,38], semantic [39–41], and syntactic dependency systems [42,43]. In a phonological network, for example, the nodes of the graph correspond to the phonetic transcription of a word (e.g., as drawn from a dictionary), and edges are placed between words if they differ by no more than one phoneme (Figure 2). In this way, network approaches to cognitive systems tend to be built on existing distributional measures (in this case, phonological neighborhood density [44]). However, as discussed below, the higher-order architectural properties of these cognitive networks likely have their own set of consequences for learnability.

Network topology influences learning and memory

Historically, complex network analyses of cognitive structures have had a descriptive focus. Naturally, the first step in understanding how networks structures develop [45–47] is to characterize existing topological properties (e.g., based on text or production-based corpora). Only recently has network topology been linked to empirical evidence of human learning. This relationship is typically examined in one of two ways: (1) by exposing adult learners to a set of tightly controlled stimuli and asking how they retain knowledge based on its underlying architecture; and (2) by extracting the network properties of adult speech and charting its influence on the trajectory of word learning in young children. Most of this empirical work has taken place in the language domain, likely because the topologies of language systems are richly characterized and also because mastery of a complex language system represents an unequivocal learning challenge [35].

Central to the production and comprehension of language lies the construct of a mental lexicon. To interact successfully with our language environment, it is essential to develop, adapt, and efficiently access the conceptual, grammatical, and sound-based properties of words. Accumulating evidence suggests a key role for phonological (sound-based) network properties as they mediate various aspects of this process. Interestingly, the **clustering coefficient** of a word (construed in this case as a node in a phonological network) has been shown to predict how robustly that word is acquired [48]. The clustering coefficient indicates whether the neighboring nodes of a target word are also connected to each other. In Figure 2, we offer examples of words with high and low clustering coefficient. Words in the phonological neighborhood of *beer* (i.e., those that overlap but for a single phoneme) also tend to overlap with each other (e.g., *hear-year-near-dear*), resulting in a high clustering coefficient. In contrast, words in the neighborhood of *silk* (e.g., *milk-sulk-silly-sink*) do not share a strong phonological relationship with each other, resulting in a low clustering coefficient. One week after training on pseudoword-novel object pairings, adult participants were better able to match a pseudoword to its corresponding object if that pseudoword had a high clustering coefficient. Complementary results indicate that after training on the phonological neighbors of a target word (but not the word itself), learners tended to erroneously endorse target words with low clustering coefficient. In contrast, results from a different long-term memory task (in which target words were not withheld) indicated that low clustering coefficient conferred a recognition advantage; correct endorsement was higher when compared to words with high clustering coefficient [49]. These findings are complemented by word recognition studies that demonstrate a processing [50,51] and production [52] advantage for words with low clustering coefficient, suggesting that words with these properties are more rapidly accessed from the mental lexicon. Taken together, these results suggest that dense interconnectedness between nodes in a network can confer an advantage for learning, but a disadvantage for retrieval. The effect of network topology on a given cognitive process may therefore differ depending on the precise nature of the task at hand.

To be clear, the clustering coefficient of a target word represents only one well-studied example of topological influences on learning and memory. When analyzing phonological properties of caregiver speech, it was found that children under 30 months were most likely

to produce words with a high degree and low **coreness** (a measure of how deeply a node is embedded in a network) [53]. Larger scale architectural properties may also play a role in accessing the mental lexicon. One example of such a macro-level property is **assortative mixing** by degree. This network property encapsulates the phenomenon that words with many phonological neighbors (high degree) are linked to words that also have many phonological neighbors, while words with few phonological neighbors (low degree) are linked to words that share this property. While assortative mixing has not been examined in a learning context, its effects are observable in psycholinguistic studies of lexical access. When the sound quality of a word with high degree is degraded, listeners are most likely to misperceive that item as a different word with equally high degree [54]. Other higher-level patterns in the phonological network have been implicated in memory processes. For example, words drawn from so-called “giant components” exhibiting **small-world properties** such as a high clustering coefficient and **short characteristic path length** are recalled less reliably than words drawn from “lexical islands” that share no connection to the giant component [55].

Clearly, the network structure of the lexicon has been most commonly framed in terms of phonological patterns. Nonetheless, its semantic organization (i.e., how words are related to each other via their meaning or co-occurrence in a corpus) has also been shown to influence word learning. For example, the distinctness of an object’s features, such as shape and surface properties, has been shown to predict age of acquisition of the word corresponding to that object. Phrased another way, objects that are topologically disconnected from other objects in a semantic network are labeled at the youngest age [14]. Graph analytic techniques have also been successfully applied to model child output [56,57]. In one recent study, a semantic network was constructed such that a node in the network represented a word known by children of 15-36 months, and each edge represented co-occurrence statistics as derived from a database of caregiver speech [57]. Word knowledge of typically-developing children exhibited characteristic small world properties, but the networks of late-talking children exhibited this quality to a lesser degree. Thus, language acquisition is influenced not only by the network architecture of input, but also by differences in how that input is transformed to an individual’s own distinct network topology. More broadly, these developmental studies highlight the potential power of harnessing network science to better understand individual differences in core cognitive capacities (see also [58]). In the following section, we turn to the essential question of how the learner begins to build complex network representations.

Local statistics underpin network architecture

The impact of local statistics

As introduced previously, statistical learning persists as an influential and well-supported theory of how learners extract structure from our external world. While we mainly focus on the effect of pairwise conditional probabilities, statistical learning fits into a broader distributional learning literature. In fact, longstanding interest in how we compute local contingencies can be traced to even earlier study of associative learning mechanisms in animals (e.g., [59]). As used to describe complex knowledge acquisition in humans, the term

distributional information encompasses the context in which elements in the environment appear together and with what regularity (e.g., frequency or probability). Sensitivity to distributional information is thus evident in charting the influence of patterns of co-occurrence frequency (e.g., how often do groupings of elements appear together? [60], see also a related literature on “chunking” in finite-state grammar learning [61]), or the conditional probabilities between elements (e.g., what is the probability that a given element will follow another?). Manipulation of distributional statistics has also used to examine the acquisition of phonetic categories [62,63], phrase structure [64,65], and syntactic dependencies [66]. Outside of such learning contexts, the influence of distributional statistics is also observed in the processing of already familiar input. During natural language comprehension, for example, knowledge about grammatical dependencies result in a processing benefit for expected (probable) structures [67,68]. Likewise, processing costs associated with improbable structures can be overcome by sharply shifting their frequency in the immediate sensory environment [69]. Thus, the computation of statistical regularities is a continuous process that permeates both the acquisition of novel representations and the processing of familiar structures (for further discussion of the intersection between these capacities, see [65,70,71]).

While distributional sensitivity is often demonstrated via linguistic stimuli, we note that, outside the language domain, pairwise statistical patterns have been shown to drive the parsing of tonal groupings [72], visual events [73,74], and spatial scenes [75]. Evidence therefore points to statistical learning as a powerful domain-general mechanism, or one that can perform equivalent computations regardless of both input modality (e.g., auditory or visual; but see [76,77]). Despite its wide applicability, an important, growing area of research probes constraints on our distributional learning capacities (i.e., how might this learning mechanism break down in the context of multiple patterns [78,79]? At what levels of abstraction does it operate [80]? And how does learning interact with attentional processes [81,82]?).

Local statistics give rise to complex networks

While evidence suggests that associations can be computed hierarchically (e.g., to order words into phrases and phrases into sentences [66]; more generally, [83]), many learning studies involve the manipulation of non-hierarchical, pairwise statistical information such as co-occurrence frequencies or conditional probabilities. From a complex network perspective, this amounts to constructing **dyads** in the network. However, a notable recent study suggests that learners also capitalize on regularities that emerge from global network architecture [84]. Participants in this study viewed a continuous stream of images that was generated by a random walk through a graph with community structure (i.e., three distinct groupings of interconnected nodes). Each node of the graph corresponded to a unique image, and the random walk through the graph ensured that transition (conditional) probabilities between images in the sequence were uniform (i.e., not a cue to an event boundary, as in canonical segmentation studies, e.g., [1]). Despite the absence of pairwise probabilistic information, learners were able to segment groupings of images based solely on the community structure of the underlying graph, accurately detecting when the stream shifted from one cluster to the next. While this finding represents a tremendous step forward in our understanding of

macro-level topological influences on the learning process, we stress that this account is by no means incompatible with existing statistical learning accounts. Just as early segmentation tasks demonstrated that conditional probabilities were sufficient to drive word extraction (for further discussion of this point see [4]), the present task demonstrates that community structure is also sufficient to drive event segmentation. In real-world situations, learners likely exploit a combination of the local and global-level cues available to them.

Statistical learning and complex network approaches are therefore compatible, but the marriage of these two disciplines has the potential to offer additional insight into the acquisition of large-scale structural knowledge. As they are currently implemented, statistical learning approaches provide the building blocks of graph structure, while complex systems approaches can reveal higher order relational patterns over and above those captured by co-occurrence statistics. We suggest here that the types of local relationships typically manipulated in statistical learning paradigms (e.g., syllable, word, or phrase co-occurrence) are not only correlated with certain micro-level topological properties of complex networks (i.e., the degree of a target node), but themselves underpin macro-level properties such as community structure (Figure 3). For example, in quantitative linguistics, co-occurrence networks are generated from temporal adjacency between words in a given context have been used to probe broader topological properties of semantic systems (e.g., [85]). Moreover, because the edges of a network are interpreted as any relationship between two nodes, it is possible to build probabilistic information into the weights of an edge when modelling the graph structure of a system. While the densely clustered graph structure discussed previously [84] involved edges of equal weights (uniform transitional probabilities), a random walk on a weighted graph would result in the traversal of certain edges more than others, potentially leading to novel segmentation patterns. Thus, probability of co-occurrence can be used to determine the regularity with which graph edges are traversed (e.g., as a temporal sequence unfolds). Indeed, an important line of future work might explore how learning is influenced by topological properties of a complex network in addition to the specific sequence in which its edges are traversed. Preliminary evidence suggests that the impact of general topological properties outweighs algorithms dictating the order in which its constituent elements are revealed or retrieved [86].

Already statistical learning researchers are tackling new questions with concepts closely tied to network science: how does the sparseness or density of input influence generalization [87]? What latent structures support acquisition [88]? A more formal integration of network science methods with current statistical learning approaches will allow current experimental methods to be “scaled up” to unprecedented levels (see criticisms by [8]). While mounting evidence indicates that statistical learning mechanisms act on larger lexicons (and that these learning effects can persist over years [89]), increased communication between statistical learning and complex network science will likely offer additional insight into the influence of macro-level topological properties on this process. Network topology might also have far-reaching impact on cognitive capacities that support or influence learning (e.g., attention, working memory, general intelligence, see [76]). For a more detailed treatment of how network-based approaches might be applied to the study of learning, refer to the Outstanding Questions Box.

The human brain is a dynamical complex system

As with all efforts to understand a cognitive process, probing how that process is implemented in the human brain must be considered in addition to patterns of human behavior. Particularly as related to linguistic functions such as processing and production, this integrative approach between brain and behavior has been applied with success [90]. With the advent of functional Magnetic Resonance Imaging (fMRI), similar strides have been made in increasing our understanding of the neural regions recruited during statistical learning (e.g., in segmentation tasks: [91–95]). However, this univariate approach is relatively coarse, revealing only areas that, on average, show increased neural activity above a significance threshold or related to a behavioral measure. To address this potential concern, we now survey one additional avenue of promising research: the study of the human brain itself as a dynamical complex system [19,21,23].

In functional brain networks, relationships arise from correlation or coherence in the neural response of pairs of neural areas (Figure 4). These relationships are taken to index how regions communicate with one another dynamically during a given task. Most relevant to the current review, advancements have been made in the application of complex system approaches to uncovering functional neural mechanisms of the learning process [96,97]. For example, graph theoretical techniques have been applied to track changes in modular (community) organization as participants practiced a simple motor sequence [98]. In this case, it was observed that modularity shifted dynamically over time, and also that it was possible to predict future behavior based on characteristics of network configuration in a given scanning session. Specifically, flexibility, defined by the number of changes in the connectivity of nodes to higher-level communities, was a significant predictor of reaction time speed-up. Especially as indexed in frontal areas, flexibility has also been linked to performance on various memory tasks, such as those involving short-term contextual recollection [99] and working memory components [100]. Related work has demonstrated that the separability of a richly connected “core” of primary sensorimotor regions and a sparsely connected, flexible “periphery” of association regions increase throughout learning, a measure also found to predict individual differences in performance on a motor learning task [101]. Ongoing work examines other dynamical properties that may drive learning (e.g., network centrality [102] and autonomy of sensorimotor associations [103]).

As reviewed throughout this manuscript, a key line of research in cognitive science centers on the learnability of network structures that emerge from external input (e.g., via the topological features in language). We propose here that an equally important line of work examines how internal complex system dynamics give rise to learning, and indeed how acquired knowledge might be reflected in observable topological patterns in the human brain. While most complex network approaches to brain connectivity begin with brain parcellation according to cytoarchitectonic or other anatomical division (Figure 4), some approaches have examined network structure via voxel-to-voxel associations (i.e., where individual voxels correspond to a network node [104–107]). In contrast to whole-brain connectivity approaches, which focus on the coordination of large-scale cognitive systems, voxel-based analysis techniques introduce the possibility of examining network-based representations as they develop. We submit that how learners translate network architectures

in sensory input to network-based neural representations remains an essential, but open question. More generally, we propose that the field of cognitive science stands to benefit enormously from increased incorporation of network-based approaches, specifically because the adoption of these methods offers a hitherto absent framework upon which to unify behavioral, computational, and neuroscientific studies of learning.

Concluding remarks

Learners attain a complex and highly structured representation of the world. Currently, many quantitative approaches to learning hinge on sensitivity to local statistics such as co-occurrence frequencies and transitional probabilities between adjacent elements. While local statistics are clearly one salient source of structural information, evidence reviewed here suggests that learners also perceive global organizational patterns. In fact, exciting new results suggest that learners can acquire knowledge of these large-scale patterns even when local statistics are relatively uninformative. Thus, as we deepen our understanding of the natural world as a complex system, we can harness descriptions of its large-scale organizational properties to then scale-up our approach to the study of learning. Moreover, just as we seek to understand how humans extract both local statistics and global architectural information, it is also useful to examine how micro- and macro-level brain dynamics support learning. In particular, an exciting avenue of future research will center on understanding how differences in network architecture of naturalistic stimuli lead to differences in learnability via unique neurophysiological mechanisms.

Acknowledgments

This work was supported by an NIH grant to STS (DC-009209-12), an NSF Career Award to DSB (1554488), and NSF workshop award BCS-1430087 “Quantitative Theories of Learning Memory, and Prediction.”

GLOSSARY

Assortative mixing

A measure of whether nodes with similar properties (e.g., high degree), are more likely to share an edge

Clustering coefficient

The extent to which adjacent neighbors of a given node are also connected to one another. This measure may be calculated for an individual node, or on average across a network

Community structure

A graph property wherein nodes are densely connected in clusters that in turn share only weak connections with one another. Communities are commonly also referred to as modules

Coreness

A measure of how deeply a given node is embedded in a network. A node has high coreness if it is retained in the network after recursively pruning nodes with low degree

Degree

The number of edges incident to a given node. A node has high degree if it is densely connected to many other nodes and low degree if it is only sparsely connected. Complex networks may have skewed degree distributions such that certain nodes are far more richly connected than others, forming hubs

Dyad

A pair of nodes sharing an edge

Edges

Links between the vertices in a network. If an edge is directed, then the order in which nodes are connected is meaningful (e.g., temporal order is important for a syntactic network, but not for a phonological network)

Nodes

Vertices, or connection points, which comprise a network

Shortest characteristic path length

A measure of network efficiency, it is, on average, the least possible distance between every pair of nodes when traversing along the edges of a network

Small-world network

A family of networks defined by short characteristic path length and a high degree of clustering

References

1. Saffran JR, et al. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–8. [PubMed: 8943209]
2. Romberg AR, Saffran JR. Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2010; 1:906–914. [PubMed: 21666883]
3. Aslin RN, Newport EL. Statistical learning: From acquiring specific items to forming general rules. *Curr Dir Psychol Sci*. 2012; 21:170–176. [PubMed: 24000273]
4. Aslin RN, Newport EL. Distributional Language Learning: Mechanisms and Models of Category Formation. *Lang Learn*. 2014; 64:86–105. [PubMed: 26855443]
5. Schapiro, A.; Turk-Browne, NB. Statistical Learning. In: Toga, AW., editor. *Brain Mapping: An Encyclopedic Reference*. Elsevier; 2015. p. 501-6.
6. Saffran, JR., et al. The Infant's Auditory World: Hearing, Speech, and the Beginnings of Language. In: Damon, W.; Lerner, RM., editors. *Handbook of Child Psychology*. Vol. 6. John Wiley & Sons, Inc; 2007. p. 58-108.
7. Saffran JR. The Use of Predictive Dependencies in Language Learning. *J Mem Lang*. 2001; 44:493–515.
8. Johnson EK, Tyler MD. Testing the limits of statistical learning for word segmentation. *Dev Sci*. 2010; 13:339–45. [PubMed: 20136930]
9. Pelucchi B, et al. Statistical learning in a natural language by 8-month-old infants. *Child Dev*. 2009; 80:674–85. [PubMed: 19489896]
10. Estes KG, Lew-Williams C. Listening through voices: Infant statistical word segmentation across multiple speakers. *Dev Psychol*. 2015; 51:1517–28. [PubMed: 26389607]
11. Nicolis, G.; Nicolis, C. *Foundations of Complex Systems*. World Scientific; 2012.
12. Amancio DR, et al. On the concepts of complex networks to quantify the difficulty in finding the way out of labyrinths. *Phys A Stat Mech its Appl*. 2011; 390:4673–4683.

13. Chrastil ER, Warren WH. Active and passive spatial learning in human navigation: acquisition of graph knowledge. *J Exp Psychol Learn Mem Cogn*. 2015; 41:1162–78. [PubMed: 25419818]
14. Engelthaler T, Hills TT. Feature Biases in Early Word Learning: Network Distinctiveness Predicts Age of Acquisition. *Cogn Sci*. 2016; doi: 10.1111/cogs.12350
15. Steyvers M, Tenenbaum JB. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn Sci*. 2005; 29:41–78. [PubMed: 21702767]
16. Borge-Holthoefer J, Arenas A. Semantic networks: Structure and dynamics. *Entropy*. 2010; 12:1264–1302.
17. Goñi J, et al. Switcher-random-walks: a cognitive-inspired mechanism for network exploration. *Int J Bifurc Chaos*. 2010; 20:913–22.
18. Ferrer i Cancho R, et al. Patterns in syntactic dependency networks. *Phys Rev E*. 2004; 69:051915.
19. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*. 2009; 10:186–198. [PubMed: 19190637]
20. Siegelmann HT. Complex Systems Science and Brain Dynamics: A Frontiers in Computational Neuroscience Special Topic. *Front Comput Neurosci*. 2010; 4:7. [PubMed: 20877423]
21. Sporns O. Structure and function of complex brain networks. *Dialogues Clin Neurosci*. 2013; 15:247–62. [PubMed: 24174898]
22. Papo D, et al. Complex network theory and the brain. *Philos Trans R Soc Lond B Biol Sci*. 2014; 369:47–97.
23. Goldenberg D, Galván A. The use of functional and effective connectivity techniques to understand the developing brain. *Dev Cogn Neurosci*. 2015; 12:155–164. [PubMed: 25770766]
24. Medaglia JD, et al. Cognitive network neuroscience. *J Cogn Neurosci*. 2015; 27:1471–91. [PubMed: 25803596]
25. Sporns O, Betzel RF. Modular Brain Networks. *Annu Rev Psychol*. 2016; 67:613–640. [PubMed: 26393868]
26. Strogatz SH. Exploring complex networks. *Nature*. 2001; 410:268–276. [PubMed: 11258382]
27. Butts CT, et al. Revisiting the foundations of network analysis. *Science*. 2009; 325:414–6. [PubMed: 19628855]
28. Backes AR, et al. A complex network-based approach for boundary shape analysis. *Pattern Recognit*. 2009; 42:54–67.
29. Backes AR, Bruno OM. Shape classification using complex network and Multi-scale Fractal Dimension. *Pattern Recognit*. 2010; 31:44–51.
30. Backes AR, et al. Texture analysis and classification: A complex network-based approach. *Inf Sci (Ny)*. 2013; 219:168–180.
31. Wang Z, et al. Impact of social punishment on cooperative behavior in complex networks. *Sci Rep*. 2013; 3:3055. [PubMed: 24162105]
32. Zhang, F., et al. Community detection based on links and node features in social networks. In: He, X., et al., editors. *MultiMedia Modeling: 21st International Conference, Part I*. Springer International Publishing; 2015. p. 418-29.
33. Cong J, Liu H. Approaching human language with complex networks. *Phys Life Rev*. 2014; 11:598–618. [PubMed: 24794524]
34. Vitevitch MS, Castro N. Using network science in the language sciences and clinic. *Int J Speech Lang Pathol*. 2015; 17:13–25. [PubMed: 25539473]
35. Beckage, NM.; Colunga, E. Language Networks as Models of Cognition: Understanding Cognition through Language. In: Mehler, A., et al., editors. *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Springer; 2015. p. 3-28.
36. Nastase V, et al. A survey of graphs in natural language processing. *Nat Lang Eng*. 2015; 21:665–698.
37. Arbesman S, et al. The structure of phonological networks across multiple languages. *Int J Bifurc Chaos*. 2010; 20:679–685.
38. Siew CSQ. Community structure in the phonological network. *Front Psychol*. 2013; 4:553. [PubMed: 23986735]

39. Gravino, p, et al. Complex structures and semantics in free word association. *Adv Complex Syst.* 2012; 15:1250054.
40. Morais AS, et al. Mapping the Structure of Semantic Memory. *Cogn Sci.* 2013; 37:125–145. [PubMed: 23163731]
41. Utsumi A. A Complex Network Approach to Distributional Semantic Models. *PLoS One.* 2015; 10:e0136277. [PubMed: 26295940]
42. Corominas-Murtra B, et al. The ontogeny of scale-free syntax networks: phase transitions in early language acquisition. *Adv Complex Syst.* 2009; 12:371–392.
43. Cech R, Macutek J. Word form and lemma syntactic dependency networks in Czech: a comparative study. *Glottometrics.* 2009; 19:85–98.
44. Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 1998; 19:1–36. [PubMed: 9504270]
45. Ke J, Yao Y. Analysing Language Development from a Network Approach. *J Quant Linguist.* 2008; 15:70–99.
46. Nematzadeh, A., et al. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014. A Cognitive Model of Semantic Network Learning; p. 244-254.
47. Hills TT, et al. Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychol Sci.* 2009; 20:729–739. [PubMed: 19470123]
48. Goldstein R, Vitevitch MS. The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition. *Front Psychol.* 2014; 5:1307. [PubMed: 25477837]
49. Vitevitch MS, et al. Complex network structure influences processing in long-term and short-term memory. *J Mem Lang.* 2012; 67:30–44. [PubMed: 22745522]
50. Chan KY, Vitevitch MS. The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *J Exp Psychol Hum Percept Perform.* 2009; 35:1934–49. [PubMed: 19968444]
51. Yates M. How the clustering of phonological neighbors affects visual word recognition. *J Exp Psychol Learn Mem Cogn.* 2013; 39:1649–56. [PubMed: 23565783]
52. Chan KY, Vitevitch MS. Network structure influences speech production. *Cogn Sci.* 2010; 34:685–697. [PubMed: 21564230]
53. Carlson MT, et al. How children explore the phonological network in child-directed speech: A survival analysis of children’s first word productions. *J Mem Lang.* 2014; 75:159–180. [PubMed: 25089073]
54. Vitevitch MS, et al. Insights into failed lexical retrieval from network science. *Cogn Psychol.* 2014; 68:1–32. [PubMed: 24269488]
55. Siew CSQ, Vitevitch MS. Spoken Word Recognition and Serial Recall of Words From Components in the Phonological Network. *J Exp Psychol Learn Mem Cogn.* 2015; 42:394–410. [PubMed: 26301962]
56. Kenett YN, et al. Semantic organization in children with cochlear implants: computational analysis of verbal fluency. *Front Psychol.* 2013; 4:543. [PubMed: 24032018]
57. Beckage N, et al. Small worlds and semantic network growth in typical and late talkers. *PLoS One.* 2011; 6:e19348. [PubMed: 21589924]
58. Kenett YN, et al. Investigating the structure of semantic networks in low and high creative persons. *Front Hum Neurosci.* 2014; 8:407. [PubMed: 24959129]
59. Pavlov IP. Conditioned Reflexes. 1927:17.
60. Arnon I, Snider N. More than words: Frequency effects for multi-word phrases. *J Mem Lang.* 2010; 62:67–82.
61. Perruchet P, Pacton S. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci.* 2006; 10:233–8. [PubMed: 16616590]
62. Maye J, et al. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition.* 2002; 82:B101–11. [PubMed: 11747867]

63. Liu R, Holt LL. Dimension-based statistical learning of vowels. *J Exp Psychol Hum Percept Perform.* 2015; 41:1783–98. [PubMed: 26280268]
64. Gómez RL. Variability and detection of invariant structure. *Psychol Sci a J Am Psychol Soc / APS.* 2002; 13:431–436.
65. Misyak JB, et al. Sequential expectations: the role of prediction-based learning in language. *Top Cogn Sci.* 2010; 2:138–53. [PubMed: 25163627]
66. Thompson SP, Newport EL. Statistical Learning of Syntax: The Role of Transitional Probability. *Lang Learn Dev.* 2007; 3:1–42.
67. Levy R. Expectation-based syntactic comprehension. *Cognition.* 2008; 106:1126–1177. [PubMed: 17662975]
68. MacDonald MC. How language production shapes language form and comprehension. *Front Psychol.* 2013; 4:226. [PubMed: 23637689]
69. Fine AB, et al. Rapid Expectation Adaptation during Syntactic Comprehension. *PLoS One.* 2013; 8:e77661. [PubMed: 24204909]
70. Conway CM, et al. Implicit Statistical Learning in Language Processing: Word Predictability is the Key. *Cognition.* 2011; 114:356–371. [PubMed: 19922909]
71. Karuza, EA., et al. On-line measures of prediction in a self-paced statistical learning task. In: Bello, P., et al., editors. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* 2013. p. 725-30.
72. Creel SC, et al. Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J Exp Psychol Learn Mem Cogn.* 2004; 30:1119–1130. [PubMed: 15355140]
73. Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn.* 2002; 28:458–67. [PubMed: 12018498]
74. Kirkham NZ, et al. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition.* 2002; 83:B35–42. [PubMed: 11869728]
75. Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci.* 2001; 12:499–504. [PubMed: 11760138]
76. Siegelman N, Frost R. Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *J Mem Lang.* 2015; 81:105–120. [PubMed: 25821343]
77. Frost R, et al. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci.* 2015; 19:117–125. [PubMed: 25631249]
78. Gebhart AL, et al. Changing structures in midstream: Learning along the statistical garden path. *Cogn Sci.* 2009; 33:1087–1116. [PubMed: 20574548]
79. Franco A, Destrebecqz A. Chunking or not chunking? How do we find words in artificial language learning? *Adv Cogn Psychol.* 2012; 8:144–54. [PubMed: 22723813]
80. Emberson LL, Rubinsteyn DY. Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition.* 2016; 153:63–78. [PubMed: 27139779]
81. Turk-Browne NB, et al. The automaticity of visual statistical learning. *J Exp Psychol Gen.* 2005; 134:552–64. [PubMed: 16316291]
82. Musz E, et al. Visual statistical learning is not reliably modulated by selective attention to isolated events. *Atten Percept Psychophys.* 2015; 77:78–96. [PubMed: 25172196]
83. Friston K. Hierarchical Models in the Brain. *PLoS Comput Biol.* 2008; 4:e1000211. [PubMed: 18989391]
84. Schapiro AC, et al. Neural representations of events arise from temporal community structure. *Nat Neurosci.* 2013; 16:486–92. [PubMed: 23416451]
85. Howard MW, et al. Constructing semantic representations from a gradually-changing representation of temporal context. *Top Cogn Sci.* 2011; 3:48–73. [PubMed: 21552364]
86. Abbott J, et al. Random Walks on Semantic Networks Can Resemble Optimal Foraging. *Psych Review.* 2015; 122:558–569.
87. Reeder PA, et al. From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cogn Psychol.* 2013; 66:30–54. [PubMed: 23089290]

88. Ellis NC, et al. Usage-Based Language: Investigating the Latent Structures That Underpin Acquisition. *Lang Learn.* 2013; 63:25–51.
89. Frank MC, et al. Learning and Long-Term Retention of Large-Scale Artificial Languages. *PLoS One.* 2013; 8:e52500. [PubMed: 23300975]
90. Friederici AD, Singer W. Grounding language processing on basic neurophysiological principles. *Trends Cogn Sci.* 2015; 19:329–338. [PubMed: 25890885]
91. McNealy K, et al. Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci.* 2006; 26:7629–39. [PubMed: 16855090]
92. Turk-Browne NB, et al. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci.* 2009; 21:1934–45. [PubMed: 18823241]
93. Cunillera T, et al. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage.* 2009; 48:541–53. [PubMed: 19580874]
94. Turk-Browne NB, et al. Implicit perceptual anticipation triggered by statistical learning. *J Neurosci.* 2010; 30:11177–87. [PubMed: 20720125]
95. Karuza EA, et al. The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang.* 2013; 127:46–54. [PubMed: 23312790]
96. Hampshire A, et al. Network mechanisms of intentional learning. *Neuroimage.* 2016; 127:123–34. [PubMed: 26658925]
97. Parkin BL, et al. Dynamic network mechanisms of relational integration. *J Neurosci.* 2015; 35:7660–73. [PubMed: 25995457]
98. Bassett DS, et al. Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci U S A.* 2011; 108:7641–6. [PubMed: 21502525]
99. Fornito A, et al. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proc Natl Acad Sci.* 2012; 109:12788–12793. [PubMed: 22807481]
100. Braun U, et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc Natl Acad Sci U S A.* 2015; 112:11678–83. [PubMed: 26324898]
101. Bassett DS, et al. Task-Based Core-Periphery Organization of Human Brain Dynamics. *PLoS Comput Biol.* 2013; 9:e1003171. [PubMed: 24086116]
102. Mantzaris AV, et al. Dynamic network centrality summarizes learning in the human brain. *J Complex Networks.* 2013; 1:83–92.
103. Bassett DS, et al. Learning-induced autonomy of sensorimotor systems. *Nat Neurosci.* 2015; 18:744–51. [PubMed: 25849989]
104. de Reus MA, van den Heuvel MP. The parcellation-based connectome: Limitations and extensions. *Neuroimage.* 2013; 80:397–404. [PubMed: 23558097]
105. Schwarz AJ, et al. Voxel Scale Complex Networks of Functional Connectivity in the Rat Brain: Neurochemical State Dependence of Global and Local Topological Properties. *Comput Math Methods Med.* 2012; 2012:1–15.
106. Hayasaka S, Laurienti PJ. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. *Neuroimage.* 2010; 50:499–508. [PubMed: 20026219]
107. Yu Q, et al. Altered Small-World Brain Networks in Temporal Lobe in Patients with Schizophrenia Performing an Auditory Oddball Task. *Front Syst Neurosci.* 2011; 5:7. [PubMed: 21369355]

TRENDS BOX

- Descriptive analytical approaches indicate that diverse facets of the environment adhere to a complex network structure.
- Recent advances offer insight into how learners might acquire and access network representations. Specifically, higher-order topological properties of networks have been shown to facilitate learning.
- Emerging neuroimaging techniques construe the brain itself as complex system, revealing how network dynamics support learning.
- We suggest that network science approaches are compatible with statistical learning approaches to knowledge acquisition. That is, local statistical regularities extracted from sensory input form the building blocks of complex network structures. Broader architectural properties of network structures might then explain learning effects beyond sensitivity to local statistical information.

OUTSTANDING QUESTIONS

- To what extent is network-based learning constrained? Are the richly studied topological properties that influence learning in the language domain *also* drivers of learning with non-linguistic stimuli?
- A central finding from the statistical learning literature is that learners are sensitive to non-adjacent, in addition to adjacent, dependencies. How might the learner exploit adjacent and long-distance relationships between dyads to build representations of complex networks?
- A key point here is that learners are sensitive to local distributional patterns as well as higher-order topological properties, and that these two accounts are fundamentally compatible. How might learning unfold when these sources of information are explicitly at odds? For example, in a segmentation context, how might a learner determine event boundaries when elements are densely clustered, but share only weak connections (i.e., co-occur infrequently)?
- How can empirical data, machine learning techniques, and complex systems approaches to neural representations be more tightly inter-linked? To what extent are the topological properties of sensory input reflected in neural patterns? As complex systems approaches to brain activity evolve, can computational models of cognition be given a firmer biological basis?

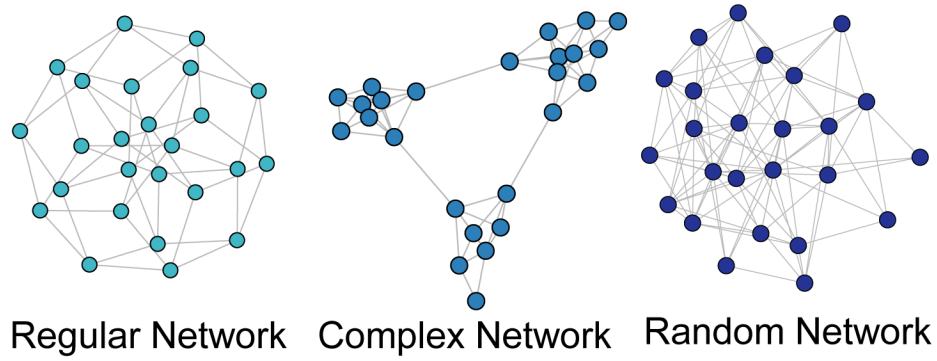


Figure 1.

Visualization of networks types. Regular networks, also known as lattices, are collections of nodes with equivalent degree (*left panel*). Random networks are collections of nodes that are linked by edges selected at random from a uniform distribution of all possible connections. Here we show a random network generated from an Erdős-Rényi model with an edge probability of 0.3 (*right panel*). In the center panel, we display a complex network with community structure, much like a network that could be derived from a learner's language environment.

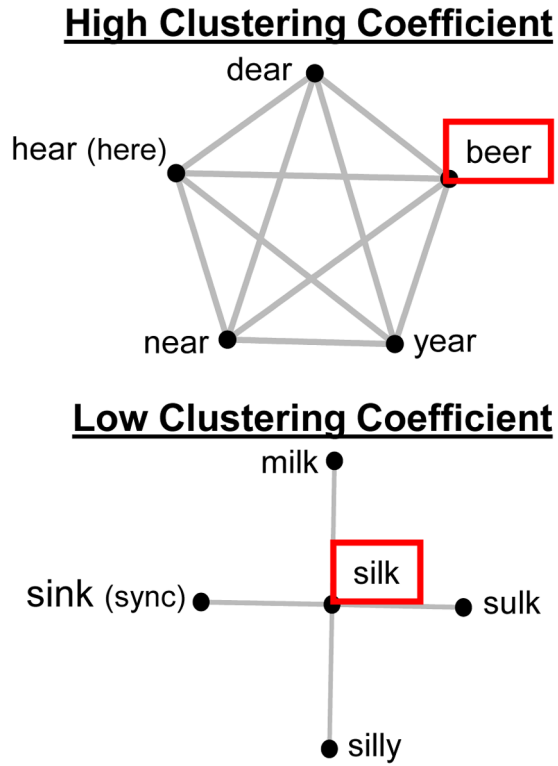
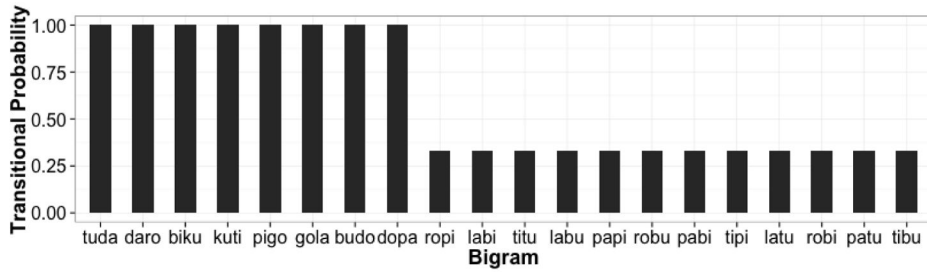


Figure 2. Sample (not exhaustive) phonological networks of two English words, *beer* and *silk*, that differ in their clustering coefficient. Note how the phonological neighbors of *beer* tend also to be phonological neighbors of each other, resulting in a high clustering coefficient. In contrast, the phonological neighbors of *silk* are not phonological neighbors, resulting in a low clustering coefficient.

Input

...TU-DA-RO-BI-KU-TI-PI-GO-LA-
 BU-DO-PA-BI-KU-TI-TU-DA-RO-
 PI-GO-LA-TU-DA-RO-BU-DO-PA-
 PI-GO-LA-BI-KU-TI-BU-DO-PA...

Pairwise statistical information



Community structure

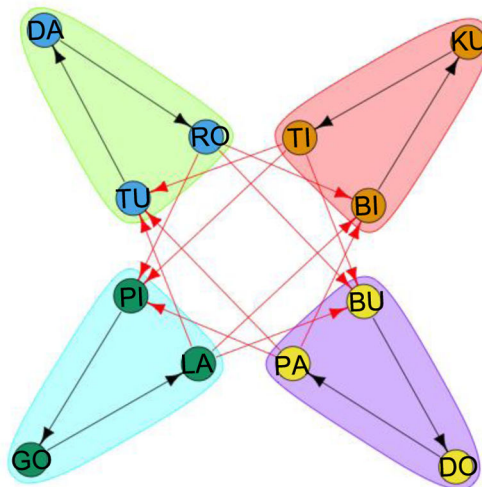


Figure 3. Co-occurrence (bigram) statistics underpin network topology. When four pseudowords (*tudaro*, *bikuti*, *pigola*, *budopa*) are concatenated together to form a continuous stream of syllables, evidence from Saffran et al. (1996) indicates that these words can be segmented via the dip in transitional probabilities at word boundaries. Here, we show that the co-occurrence between syllables can also be used to construct a weighted graph (black lines indicate a high bigram frequency and red lines indicate a low bigram frequency). A community detection algorithm consisting of a series of short random walks through this graph will then reveal robust cluster structure corresponding to each word in the stream (shown in green, pink, purple, and blue).

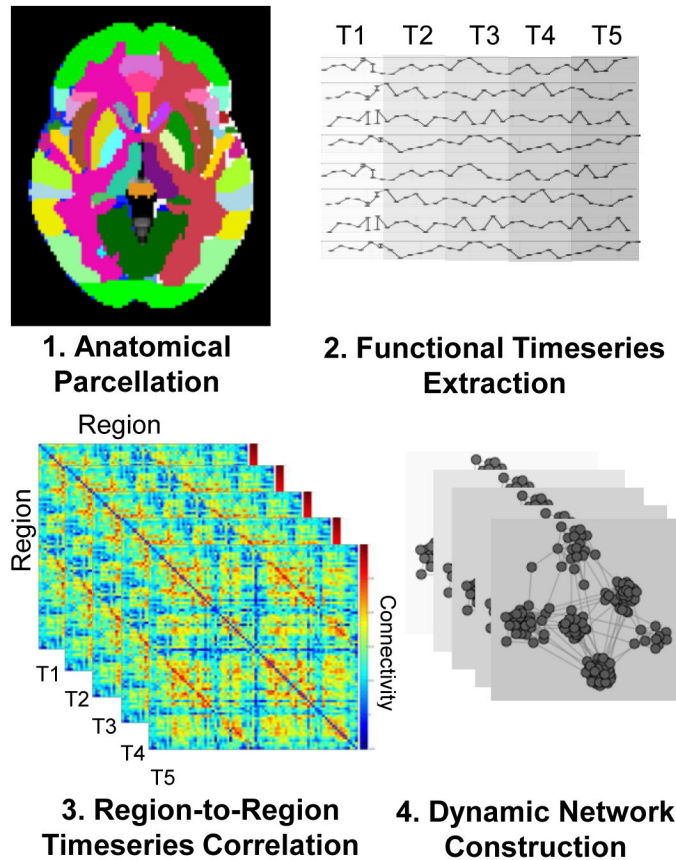


Figure 4.

Evolution of complex network structure in the brain. To investigate topological properties of task-based functional connectivity, the brain is first parcellated into anatomically defined nodes, in this example using the Harvard-Oxford structural atlas (1). Next, the moment-to-moment activity within each of these regions is extracted at different points during a scanning session (2). To construct the edges between brain regions (nodes of the network), the correlation or coherence between any two time series is computed, forming a pairwise adjacency matrix at each time point (3). These matrices can then be used to probe the temporal dynamics of a functional brain network over the course of a given task (4).