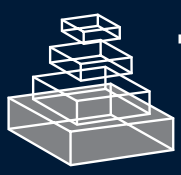
frontiers

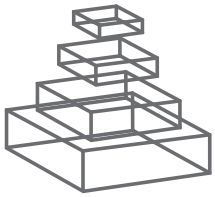
RESEARCH TOPICS

HIERARCHY AND DYNAMICS IN NEURAL NETWORKS

Hosted by
Rolf Kötter, Marcus Kaiser and
Claus Hilgetag



frontiers in
NEUROINFORMATICS



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-010-2
DOI 10.3389/978-2-88919-010-2

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

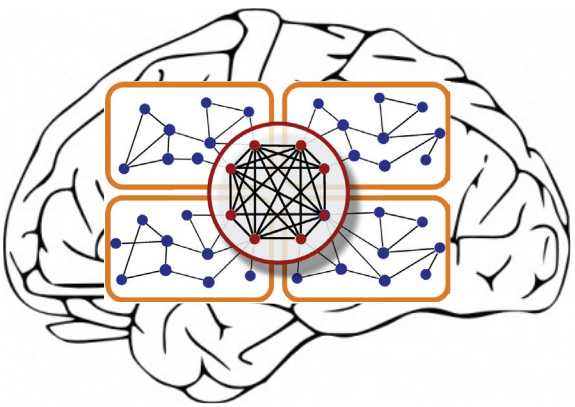
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

HIERARCHY AND DYNAMICS IN NEURAL NETWORKS

Hosted By
Rolf Kötter, Radboud University Nijmegen, Netherlands
Marcus Kaiser, Seoul National University, Korea (South)
Claus C. Hilgetag



Hierarchy is a central feature in the organisation of complex biological systems and particularly the structure and function of neural networks. While other aspects of brain connectivity such as regionalisation, modularity or motif composition have been discussed elsewhere, no detailed analysis has been presented so far on the role of hierarchy and its connection to brain dynamics. Recent discussions among many of our colleagues have shown an increasing interest in hierarchy (of spatial, temporal and dynamic features), and this is an emerging key question in neuroscience as well as generally in the field of network science, due to its links with concepts of control, efficiency and development across scales (e.g. Hilgetag et al. Science, 1996; Ravasz et al. Science, 2002; Bassett et al. PNAS, 2006; Mueller-Linow et al. PLoS Comp. Biol., in press). The proposed Research Topic will address recent findings from a theoretical as well as experimental perspective including contributions under the following four headings: 1) Topology: Detecting and characterizing network hierarchy; 2) Experiments: Neural dynamics across hierarchical scales; 3) Dynamics: Activity spread, oscillations, and synchronization in hierarchical networks; 4) Dynamics: Stable functioning and information processing in hierarchical networks.

Table of Contents

frontiers in
NEUROINFORMATICS

EDITORIAL
published: 23 August 2010
doi: 10.3389/fninf.2010.00112

Hierarchy and dynamics of neural networks

Marcus Kaiser[1,2,3]*, Claus C. Hilgetag[4,5] and Rolf Kötter[6]

[1] *School of Computing Science, Newcastle University, Newcastle upon Tyne, UK*
[2] *Institute of Neuroscience, Newcastle University, Newcastle upon Tyne , UK*
[3] *Department of Brain and Cognitive Sciences, Seoul National University, Seoul, South Korea*
[4] *School of Engineering and Science, Jacobs University Bremen, Bremen, Germany*
[5] *Department of Health Sciences, Boston University, Boston, USA*
[6] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands*
**Correspondence: m.kaiser@ncl.ac.uk*

Hierarchy is a central feature in the organization of complex biological systems and particularly the structure and function of neural networks. While other aspects of brain connectivity such as regional specialization, modularity, or motif composition have already been discussed extensively (Sporns et al., 2004; Bullmore and Sporns, 2009), no comprehensive analysis has been presented so far on the role of hierarchy and its connection to brain dynamics. Nonetheless, recent discussions among many of our colleagues have shown an increasing interest in the subject of hierarchy. This topic is an emerging key question in neuroscience, as well as generally in the field of network science, due to its links with concepts of control, efficiency, and development across scales (Ravasz and Barabasi, 2003; Barthélemy et al., 2004; Breakspear and Stam, 2005; Zhou et al., 2006; Dehmer et al., 2008).

"Hierarchy" may be understood in several different ways, and can apply to topological, spatial, temporal as well as functional properties of neural networks. The papers of this Special Topic on hierarchy and dynamics reflect this conceptual diversity.

One interpretation of hierarchy is that of a processing *sequence*. For example, the popular organizational scheme of the primate visual system (Felleman and van Essen, 1991) implies a sequential ordering of visual cortical areas from the visual sensory periphery to "higher-level" areas involved in abstract aspects of vision. This hierarchical concept is formalized by Krumnack, et al. ("Criteria for optimizing cortical hierarchies with continuous ranges") and expanded from a recent paper by Reid et al. (2009). The authors re-analyze the anatomical constraints for the hierarchical sorting of visual areas, using linear optimization and mixed integer programming, and demonstrate that there are multiple optimal solutions for visual hierarchies, as well as several alternative definitions of optimality. For instance, optimal hierarchies can be based on minimizing the number of violated constraints, or minimizing the maximal size of a constraint violation (cf. Hilgetag et al., 2000), broadening the perspective for the interpretation of the anatomical data.

Another widely used definition of hierarchy is that of a repeated *encapsulation* of smaller elements in larger ones (Kaiser et al., 2007a; Robinson et al., 2009), an organization which may also be characterized as recursive or fractal (Sporns, 2006). In that sense, neural networks show a self-similar hierarchical organization across a wide range of metric or non-metric scales. These scales may be spatial, ranging from the lobes of the brain to cortical mini-columns; temporal, stretching from plasticity and learning processes taking days and longer to neuronal firing at the millisecond scale; or topological, containing small functional elements such as "canonical circuits" (Douglas and Martin, 2004) in larger modules such as the "visual cortex."

The Special Topic contains several examples of such encapsulated hierarchies.

For instance, Meunier, et al. ("Hierarchical modularity in human brain functional networks") present techniques for the rapid detection of a hierarchy of encapsulated modules in resting-state fMRI data. They analyzed networks composed of 1,800 regional nodes, extracted from neuroimaging data for 18 human subjects, and found a good degree of similarity between the network hierarchies for different brains. Moreover, out of five modules at the highest level, the occipital modules demonstrated less sub-modular organization than modules comprising regions of multimodal association cortex. Connector nodes and hubs, with a key role in inter-modular connectivity, were also concentrated in cortical association areas. The study demonstrates the feasibility of extracting large-scale hierarchical networks from experimental imaging data, and prepares the ground for characterizing brain function by advanced network analyses.

Modules in hierarchical networks may be overlapping, rather than be cleanly delineated. Moreover, the individual nodes may differ by the topological "reach" that they have across the network. While most nodes have relatively few connections, some regions (such as amygdala and hippocampus in the cat, or the lateral intraparietal area (LIP) and area 7 in the macaque brain) are connected to many nodes of the network and thus form hubs (Kaiser et al., 2007b). Such hubs can be further distinguished into provincial (intra-modular) hubs or connector (inter-modular) hubs (Sporns et al., 2007). Zamora-López, Zhou, and Kurths ("Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks") expand such previous approaches, and identify a new element in the cat cortical connectivity network, a hub module, which consists of network nodes that possess many connections with the rest of the network as well as each other. This set of nodes forms a topologically central module of the cortex that appears to be essential for integrating multisensory information (**Figure 1**).

Several articles in the Special Topic explore the *dynamic implications* of hierarchical modular network architectures. Kaiser and Hilgetag ("Optimal hierarchical modular topologies for producing limited sustained activation of neural networks") investigate the influence of the number of hierarchical levels (scales), as well as sub-modules at each level (granularity), on the spreading of activity in hierarchical modular networks of different sizes, using a minimal dynamic node model. In particular, they characterize the conditions leading to the biologically relevant case of limited sustained activity in which activity persists between the extremes of dying out or activating the whole network (Kaiser et al., 2007a). For different network sizes, limited sustained activity is best supported when the

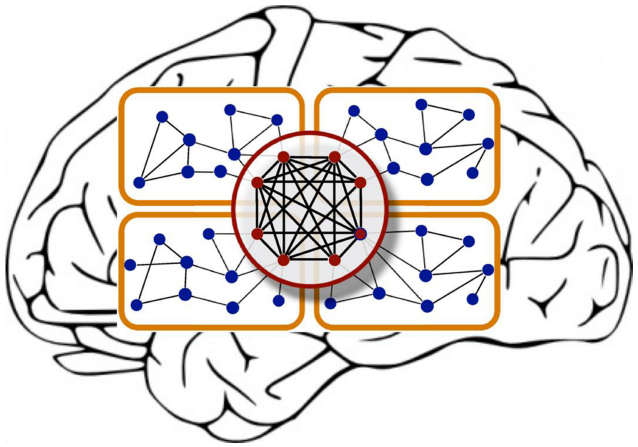


FIGURE 1 | The cartoon illustrates two important features of hierarchical brain networks: modules and hubs. Adapted from Zamora-López et al. (2010).

average number of connections per node remains similar, and the number of hierarchical levels or modules per level increases. This observation indicates that dynamic constraints may contribute to the evolution of network complexity in brain architecture.

In their work, Jarvis et al. ("Extending stability through hierarchical clusters in Echo State Networks") also find that increased intricacy of network structure aids network dynamics. They demonstrate that the stability of Echo State Networks (Jaeger and Haas, 2004) is potentially enhanced, as indicated by the range of spectral radius values, when the networks are structured in a multi-modular, hierarchical way. The more clearly the ESNs are structured, the larger the range of spectral radius values, while increasing inter-cluster connectivity decreases the maximal spectral radius. The finding suggests that insights into the organization of biological networks also have the potential to improve the functioning of networks for technical applications.

An important feature of neural systems is the occurrence of sudden changes in their dynamics (a drastic example of such phase transitions is the shift from normal brain activity to epileptic seizures). In a previous paper, Müller-Linow et al. (2008) observed that two different dynamic behaviors may emerge from hierarchical networks: waves propagating from central nodes and module-based synchronization. In the present issue, Hütt and Lesne ("Interplay between topology and dynamics in excitation patterns on hierarchical graphs") analyze more formally how excitable systems can switch from one of such states to the other. In addition to a mean-field model simulation, a formalism is introduced in which excitation waves are described as avalanches.

Shifting the perspective from network dynamics to *network function*, the flow of information in hierarchical networks is used in the brain for the processing of external or internal signals. Rodrigues and Costa ("Signal propagation in cortical networks: a digital signal processing approach") show that signal flow among network nodes can be characterized by a finite impulse response (FIR) filter. With such an approach, filters underlying the cat and macaque cortical organization are found to be low-pass, allowing signal processing to be summarized through respective cut-off frequencies. Furthermore, filtering intensity varies between network modules, and regions involved in object recognition tend to present the highest cut-off frequencies for both the cat and macaque networks.

Hierarchies can also be seen in *temporal aspects* of brain activity. Natural stimuli, such as speech, possess features at different temporal scales. Therefore, models of speech recognition should be able to represent slowly changing neuronal states that encode trajectories of faster signals. Kiebel et al. ("Perception and hierarchical dynamics") present a mathematical approach that assumes that sensory input is generated by a hierarchy of attractors in a dynamic system. Future applications of this approach might emerge from modeling perception as non-autonomous recognition dynamics enslaved by autonomous hierarchical dynamics in the sensorium.

The papers presented here offer an exciting glimpse into future directions of the field of hierarchical neural networks. But they also demonstrate that we still need a better understanding of the different kinds of network hierarchies, paralleled by the development of suitable analysis techniques. Most importantly, an improved understanding is required of how the different aspects of topological, spatial, temporal, and functional hierarchy in the brain are related to each other.

REFERENCES

Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. (2004). Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* 92, 178701.

Breakspear, M., and Stam, C. J. (2005). Dynamics of a neural system with a multiscale architecture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1051–1074.

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

Dehmer, M., Borgert, S., and Emmert-Streib, F. (2008). Entropy bounds for hierarchical molecular networks. *PLoS One* 3, e3079. doi:10.1371/journal.pone.0003079.

Douglas, R. J., and Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Ann. Rev. Neurosci.* 27, 419–451.

Felleman, D. J., and van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Hilgetag, C. C., O'Neill, M. A., and Young, M. P. (2000). Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Philos. Trans. R. Soc. Lond. Ser. B* 355, 71–89.

Hütt, M.-T., and Lesne, A. (2009). Interplay between topology and dynamics in excitation patterns on hierarchical graphs. *Front. Neuroinform.* 3:28. doi: 10.3389/neuro.11.028.2009.

Jaeger, H., and Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* 304, 78–80.

Jarvis, S., Rotter, S., and Egert, U. (2010). Extending stability through hierarchical clusters in Echo State Networks. *Front. Neuroinform.* 4:11. doi: 10.3389/fninf.2010.00011.

Kaiser, M., Görner, M., and Hilgetag, C. C. (2007a). Functional criticality in clustered networks without inhibition. *New J. Phys.* 9, 110.

Kaiser, M., Martin, R., Andras, P., and Young, M. P. (2007b). Simulation of robustness against lesions of cortical networks. *Eur. J. Neurosci.* 25, 3185–3192.

Kaiser, M., and Hilgetag, C. C. (2010). Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. *Front. Neuroinform.* 4:8. doi: 10.3389/fninf.2010.00008.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2009). Perception and hierarchical dynamics. *Front. Neuroinform.* 3:20. doi: 10.3389/neuro.11.020.2009.

Krumnack, A., Reid, A. T., Wanke, E., Bezgin, G., and Kötter, R. (2010).

Criteria for optimizing cortical hierarchies with continuous ranges. *Front. Neuroinform*. 4:7. doi: 10.3389/fninf.2010.00007.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., and Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Front. Neuroinform*. 3:37. doi: 10.3389/neuro.11.037.2009.

Müller-Linow, M., Hilgetag, C. C., and Hutt, M. T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS Comput. Biol.* 4, e1000190. doi:10.1371/journal.pcbi.1000190.

Ravasz, E., and Barabasi, A. L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E 67*, 026112.

Reid, A. T., Krumnack, A., Wanke, E., and Kotter, R. (2009). Optimization of cortical hierarchies with continuous scales and ranges. *Neuroimage* 47, 611–617.

Robinson, P. A., Henderson, J. A., Matar, E., Riley, P., and Gray, R. T. (2009). Dynamical reconnection and stability constraints on cortical network architecture. *Phys. Rev. Lett.* 103, 108104.

Rodrigues, F. A., and da Fontoura Costa, L. (2009). Signal propagation in cortical networks: a digital signal processing approach. *Front. Neuroinform*. 3:24. doi: 10.3389/neuro.11.024.2009.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems* 85, 55–64.

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425.

Sporns, O., Honey, C. J., and Kötter, R. (2007). Identification and classification of hubs in brain networks. *PLoS One* 2, e1049. doi:10.1371/journal.pone.0001049.

Zamora-López, G., Zhou, C., and Kurths, J. (2010). Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform*. 4:1. doi: 10.3389/neuro.11.001.2010.

Zhou, C., Zemanova, L., Zamora, G., Hilgetag, C. C., and Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.* 97, 238103.

Received: 04 August 2010; accepted: 05 August 2010; published online: 23 August 2010.

*Citation: Kaiser M, Hilgetag CC and Kötter R (2010) Hierarchy and dynamics of neural networks. Front. Neuroinform. **4**:112. doi: 10.3389/fninf.2010.00112*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 31 March 2010
doi: 10.3389/fninf.2010.00007

Criteria for optimizing cortical hierarchies with continuous ranges

Antje Krumnack[1]*, Andrew T. Reid[2,3], Egon Wanke[4], Gleb Bezgin[2,3] and Rolf Kötter[2]*

[1] *Department of Psychology, University of Gießen, Giessen, Germany*
[2] *Donders Institute for Brain, Cognition and Behaviour, Centre for Neuroscience, Section Neurophysiology and Neuroinformatics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands*
[3] *Center for Anatomy and Brain Research, University Clinics Düsseldorf, Heinrich Heine University, Dusseldorf, Germany*
[4] *Department of Computer Science, Heinrich Heine University, Dusseldorf, Germany*

Edited by:
Claus C. Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Basilis Zikopoulos, Boston University, USA
Michael Capalbo, University of Maastricht, Netherlands

****Correspondence:***
Antje Krumnack, Allgemeine Psychologie und Kognitionsforschung, Fachbereich 06: Psychologie und Sportwissenschaft, Justus-Liebig-Universität Gießen, Otto-Behaghel-Strasse 10F, 35394 Giessen, Germany. e-mail: antje.krumnack@psychol.uni-giessen.de; Rolf Kötter, Donders Institute for Brain, Cognition and Behaviour, Centre for Neuroscience, Section Neurophysiology and Neuroinformatics, Radboud University Nijmegen Medical Centre, POB 9101, 6500 HB Nijmegen, Netherlands.
e-mail: rk@donders.ru.nl

In a recent paper (Reid et al., 2009) we introduced a method to calculate optimal hierarchies in the visual network that utilizes continuous, rather than discrete, hierarchical levels, and permits a range of acceptable values rather than attempting to fit fixed hierarchical distances. There, to obtain a hierarchy, the sum of deviations from the constraints that define the hierarchy was minimized using linear optimization. In the short time since publication of that paper we noticed that many colleagues misinterpreted the meaning of the term "optimal hierarchy". In particular, a majority of them were under the impression that there was perhaps only one optimal hierarchy, but a substantial difficulty in finding that one. However, there is not only more than one optimal hierarchy but also more than one option for defining optimality. Continuing the line of this work we look at additional options for optimizing the visual hierarchy: minimizing the number of violated constraints and minimizing the maximal size of a constraint violation using linear optimization and mixed integer programming. The implementation of both optimization criteria is explained in detail. In addition, using constraint sets based on the data from Felleman and Van Essen (1991), optimal hierarchies for the visual network are calculated for both optimization methods.

Keywords: connectivity, hierarchy, visual system, macaque, optimality, linear programming, mixed integer programming

INTRODUCTION

In 1991, Felleman and Van Essen formalized the idea of a visual cortical hierarchy using a large number of tract tracing results obtained from macaque monkeys. Their general premise was that the laminar source and termination patterns of corticocortical projections contained information about their hierarchical directionality, which allowed projections to be labelled as ascending, descending, or lateral. Using this information as a constraint, the authors presented a cortical hierarchy in which most of these direction relationships were satisfied (see Figure 4, Felleman and Van Essen, 1991).

Since the publication of this article, the question emerged: what is the *optimal* visual cortical hierarchy? Using the same set of criteria and notion of optimality, Hilgetag et al. (1996) demonstrated that there are at least 100,000 hierarchies which are even more optimal than that introduced by Felleman and Van Essen (i.e., having less constraint violations). In Reid et al. (2009), we introduced a new approach to calculating hierarchies, which combined the laminar data from Felleman and Van Essen (1991) with new concepts from Vezoli et al. (2004) that permitted the additional representation of the hierarchical *distance* of a projection. Our approach utilized a continuous, rather than discrete scale for describing hierarchical level, and introduced the measurement of *deviation* from a constraint as a cost function for optimization, rather than a count of discrete violations. Although this method does not produce a unique optimal hierarchy – an indeterminacy problem similar to that reported by Hilgetag et al. (1996) – it does always produce an optimal hierarchy. The method is also easily implemented, such that an optimal hierarchy can be calculated for any arbitrary set of cortical areas with tract tracing information, and easily updated if new data are produced.

A further question which emerges from this process is that of suitable optimization criteria. Whether a hierarchy is considered optimal depends on the notion of optimality that is employed, and there are many options for defining optimality. In Reid et al. (2009), we utilized the sum of deviations as an objective function to be minimized, but it is uncertain whether this is the best choice of criterion, and to what extent the addition of further objectives might improve the resulting hierarchy. This also introduces the related issue of how sensitive the optimization is to this choice of criteria. In the present article we explore these possibilities, investigating in particular the effects of adding (1) the number of violations, and (2) the maximal deviation, to the objective function.

MATERIALS AND METHODS

GRAPH REPRESENTATION AND THE HIERARCHY FUNCTION

Consider a network representing a hierarchy given as a directed graph $G = (V, E)$, where V is a set of vertices and E a set of directed edges. Each edge in the graph is assigned a weight that signifies the possible range of distances of its endpoints within the hierarchy (see **Figure 1** for an example).

If the edge runs from vertex u to vertex v then an interval $[x, y]$, $x, y \in \mathbb{R}$ is assigned to that edge and the hierarchical distance between u and v should lie within this interval (see **Figure 2**). A positive distance value implies the edge is going up and v is above u in the hierarchy. A negative value means that the edge is descending with regard to the hierarchy and v is below u in the hierarchy. The value 0 signifies that the edge does not cross levels in the hierarchy and u and v are on the same level.

To find the hierarchy implied by these edge values we are looking for a function h: $V \to \mathbb{R}$ so that for every edge (u, v) with assigned interval $[x, y]$, $x, y \in \mathbb{R}$ the following conditions hold:

$$h(u) + x \le h(v) \text{ and } h(u) + y \ge h(v). \tag{1}$$

This function h is called the hierarchy function.

However for the example in **Figure 1** it is not possible to find such hierarchy functions because the distance information is not consistent. The alternative is to find a hierarchy function that "best" fits the data. To do so it is necessary to allow some deviations from the given data. To measure this deviation we introduce a variable $\Delta_{(u,v)}$ for every edge (u, v). A variable $\Delta_{(u,v)}$ measures for the two conditions defined by the edge (u, v) how much the hierarchy violates these conditions. This alters the condition for the hierarchy function resulting from an edge (u, v) with a range $[x, y]$ in the following way (compare **Figure 2** and Eq. 1):

$$h(u) + x - \Delta_{(u,v)} \le h(v) \text{ and } h(u) + y + \Delta_{(u,v)} \ge h(v). \tag{2}$$

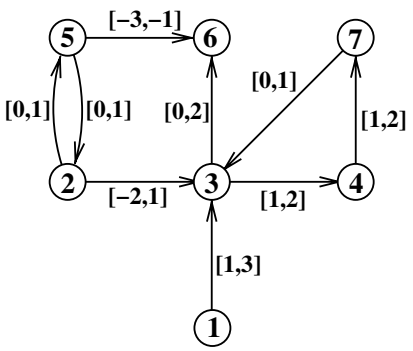


FIGURE 1 | A directed graph with vertices $v_1, v_2, v_3, v_4, v_5, v_6, v_7$ and edge ranges.

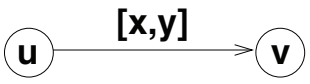


FIGURE 2 | Vertices *u* and *v* with the hierarchical distance [*x*,*y*].

With these inequalities we allow a deviation of $\Delta_{(u,v)}$ from the constraints assigned to an edge (u, v) by the interval $[x, y]$. The $\Delta_{(u,v)}$ is specific for every edge (u, v) which means one distinct variable $\Delta_{(u,v)}$ for every edge is needed. Ideally, the value of these $\Delta_{(u,v)}$ should be kept as small as possible, preferable 0. Since all $\Delta_{(u,v)}$ measure a deviation, their values are always non-negative. Also note that at most one of the two conditions for an edge (u,v) can require an $\Delta_{(u,v)}$ larger than 0. The objective is now to find a hierarchy that best fits the data, i.e., with as little overall deviation as possible. To accomplish this the sum of all deviations $\Delta_{(u,v)}$ should be minimal.

To calculate such an hierarchy we use a well known method called linear programming. A detailed introduction on linear programming can be found (for example) in the book of Papadimitiou and Steiglitz (1998). In brief the aim of linear programming is the optimization of a linear objective function, subject to linear equality and inequality constraints. Those linear problems have the following form, which is also called linear program:

Maximize/minimize the expression

$$c_1x_1 + c_2x_2 + \ldots + c_nx_n \tag{3}$$

subject to constraints

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &\le b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &\le b_m \end{aligned}$$

Here $x_1, x_2, \ldots, x_n$ are variables, for which values need to be found. All other elements are constants.

There are different ways to solve these kinds of problems. The oldest and most widely used is the simplex algorithm which was developed in 1947 by Dantzig (see for example Dantzig, 1963). It has an exponential worst case run time but most instances can be solved much faster. Today the inner point method invented by Karmarkar (1984) is often used as well, which has a polynomial runtime. For this work the optimization procedure was performed using the Gnu Linear Problem Kit[1] which can be used for linear programming as well as mixed integer programming.

EXAMPLE

Consider the example in **Figure 1**. We are looking for a hierarchy function h: $V \to \mathbb{R}$ with $h(v_1) = 0$ and for all $(u, v) \in E$ with edge value $[x, y]$ the conditions $h(u) + y + \Delta_{(u,v)} \ge h(v)$ and $h(u) + x - \Delta_{(u,v)} \le h(v)$ should hold (compare to Eq. 2).

From these inequalities we receive the following two conditions for every edge:

$$h(v) - h(u) - \Delta_{(u,v)} \le y \text{ and } h(v) - h(u) + \Delta_{(u,v)} \ge x. \tag{4}$$

Since we want to calculate the values of the hierarchy function h for every $v \in V$ we also include $h(v)$ as a variable in the linear program. To make the notation easier these variables will receive the names of the vertices in the linear program. So for every $v \in V$ there is a variable v that represents the function value $h(v)$ in the program.

[1] http://www.gnu.org/software/glpk/

With this replacement the conditions in the program look like this

$$v - u - \Delta_{(u,v)} \leq y \text{ and } v - u + \Delta_{(u,v)} \geq x. \quad (5)$$

With those inequalities we can create a linear program to calculate an optimal hierarchy for the example in **Figure 1**. The objective of the program is to minimize the sum of all deviations. (see **Figure 3**, left: the variable sumΔ is the sum of all deviations.)

The node v_1 is supposed to be the starting point of the hierarchy, and therefore its value is fixed at 0. The variables v_1 to v_7 are not limited by the optimization objective (c20), and thus can become as big or small as necessary to attain the optimal value for sumΔ.

ADDITIONAL OBJECTIVES

So far our objective for finding an optimal hierarchy is to keep the sum of deviations as small as possible for the given edge values. But since there is normally more than one hierarchy that fulfils this objective, it can be useful to employ additional secondary objectives. We will illustrate this by an example.

Consider graph A of **Figure 4**. Since the sum of all edge values in that circle is 1 the minimal sum of deviations of any hierarchy for this graph is 1. The question is how that sum is best distributed on the involved edges. One option is to concentrate the deviation on as few edges as possible. That does not change the sum of deviations but it keeps the number of violated conditions small (see **Figure 4B**). Another possibility can be seen in **Figure 4C**, there the deviation is distributed on all edges. This keeps the maximum deviation small and therefore all distances imposed by the hierarchy close to the original edge weights. **Figures 4B,C** are not the only options for an optimal hierarchy. Note that there is an infinite number of other hierarchies with a minimal sum of deviations for this example.

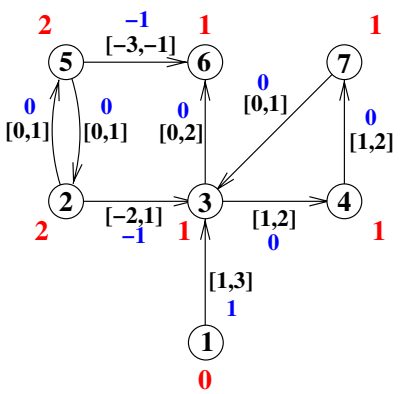
The question is: which hierarchy describes the information given by the edge values best? The advantage of minimizing the number of violations is that most of the original distance information is preserved in the hierarchy, and we ideally disregard only the information that fits the model the least. Minimizing the maximum deviation, alternatively, has the advantage that all the information is treated equally and non is disregarded completely. The rationale for this is that it is better to change many distances a little than few distances a lot.

To implement these additional objectives, additional variables are needed in the program. In the first case it is necessary to count and minimize the number of violations, which can not be done with just linear programming, but rather requires the use of mixed integer programming. For the second option we need to minimize the maximal deviation, which can easily be integrated in the linear program, so it will be discussed first.

Maximal deviation

To calculate the maximal deviation of the hierarchy we introduce a variable Δ_{max} which measures the maximal deviation from any constraints. To implement this every constraint of the original

Minimize
objective: sumΔ
Subject to
c01: $v_3 - v_1 - \Delta_1 \leq 3$
c02: $v_3 - v_1 + \Delta_1 \geq 1$
c03: $v_3 - v_2 - \Delta_2 \leq 1$
c04: $v_3 - v_2 + \Delta_2 \geq -2$
c05: $v_5 - v_2 - \Delta_3 \leq 1$
c06: $v_5 - v_2 + \Delta_3 \geq 0$
c07: $v_4 - v_3 - \Delta_4 \leq 2$
c08: $v_4 - v_3 + \Delta_4 \geq 1$
c09: $v_6 - v_3 - \Delta_5 \leq 2$
c10: $v_6 - v_3 + \Delta_5 \geq 0$
c11: $v_7 - v_4 - \Delta_6 \leq 2$
c12: $v_7 - v_4 + \Delta_6 \geq 1$
c13: $v_2 - v_5 - \Delta_7 \leq 1$
c14: $v_2 - v_5 + \Delta_7 \geq 0$
c15: $v_6 - v_5 - \Delta_8 \leq -1$
c16: $v_6 - v_5 + \Delta_8 \geq -3$
c17: $v_3 - v_7 - \Delta_9 \leq 1$
c18: $v_3 - v_7 - \Delta_9 \leq 0$
c19: $v_1 = 0$
c20: -sumΔ $+ \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Delta_5 + \Delta_6 + \Delta_7 + \Delta_8 + \Delta_9 = 0$



sum of deviations: 2
maximal deviation: 1
number of deviations: 2

FIGURE 3 | Left: The linear program to calculate an optimal hierarchy for the graph from Figure 1. The objective is to minimize the sum of all deviations (defined by constraint c20). Right: The resulting hierarchy. Red numbers are the hierarchy levels of nodes, blue numbers the actual distances in the hierarchy.

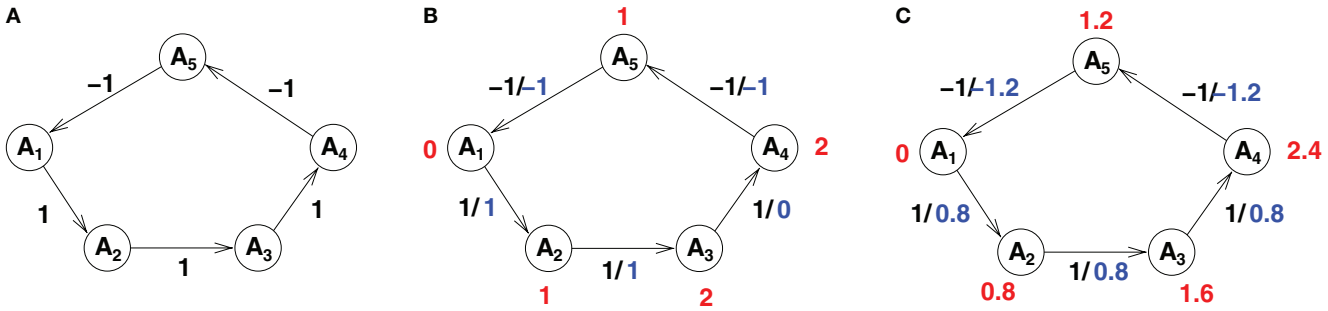


FIGURE 4 | (A) A graph with edge values. **(B)** One possible hierarchy obtained by optimizing for the number of constraint violations. **(C)** Another possible hierarchy obtained by optimizing for the maximum deviation. Both **(B)** and **(C)** red numbers are the hierarchy levels of nodes, blue numbers the actual distances in the hierarchy.

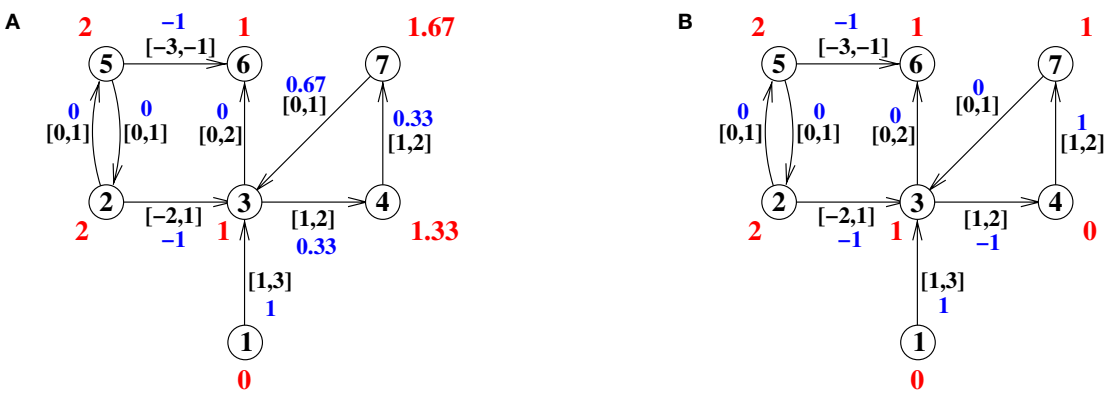


FIGURE 5 | (A) An optimal hierarchy for the graph from Figure 1 given the objective to minimize primarily the sum of all deviations and in addition the maximal deviation. Red numbers are the hierarchy levels of nodes, blue numbers the actual distances in the hierarchy. **(B)** An optimal hierarchy for the graph from **Figure 1** given the objective is to minimize primarily the sum of all deviations and in addition the number of deviations. Again, red numbers are the hierarchy levels of nodes, blue numbers the actual distances in the hierarchy.

program is doubled and in the second version the specific $\Delta_{(u,v)}$ for the edge is replaced by the global $\Delta_{\max}$. So we now have four conditions in the program for an edge (u,v) with edge value $[x,y]$ (compare Eq. 5):

$$v-u-\Delta_{(u,v)} \le y, \quad v-u-\Delta_{\max} \le y,$$
$$v-u+\Delta_{(u,v)} \ge x, \quad v-u+\Delta_{\max} \ge x. \tag{6}$$

If one of the original constraints only holds if $\Delta_{(u,v)}$ is greater than 0, then $\Delta_{\max}$ needs to be as least as big as $\Delta_{(u,v)}$ for the doubled constraints to hold as well. Since this is true for every constraint it means $\Delta_{\max}$ needs to be at least as large as the largest edge specific deviation. If $\Delta_{\max}$ is included in the objective to be minimized it will take exactly the value of the maximal edge specific deviation $\Delta_{(u,v)}$.

Since the prime objective is still to minimize the sum of all deviations we introduce a factor in the objective to give the sum of deviations a bigger weight than the maximal deviation $\Delta_{\max}$. The factor needs to be large enough that the smallest edge-specific deviation $\Delta_{(u,v)}$ multiplied by the factor is considerably larger than $\Delta_{\max}$. This ensures that $\Delta_{\max}$ is not of the same magnitude as the sum of deviations with regard to the optimization and therefore does not influence the primary optimization goal. The result is a hierarchy with a minimal sum of deviations, but among those hierarchies one with the smallest possible $\Delta_{\max}$ is chosen.
As factor we used 100 which proved to be large enough that the resulting hierarchy is still one with a minimal sum of deviations, so it fulfilled the requirements outlined in the previous paragraph. The results for the graph from **Figure 1** are shown in **Figure 5A** (compare to **Figure 3**).

Number of violations

To minimize the number of violations in a hierarchy it is necessary to count them within the program. To accomplish this we introduce an edge-specific integer variable $B_{(u,v)}$ which assumes the value 1 if $\Delta_{(u,v)} > 0$ and 0 if $\Delta_{(u,v)} = 0$. The sum of these violation counters can then be included in the objective to be minimized along with the sum of the deviations. These variables are not reals but integers, so it is necessary to use mixed integer programming. Mixed integer problems look like the linear problems we have seen so far (compare again to Eq. 3), but some of the variables can only take integer values. This gives the optimization problem a higher complexity: other than linear optimization problems, mixed integer problems can in general not be solved efficiently, they are NP-hard. While there are well-known algorithms to solve these problems, one of the first was developed by Land and Doig (1960), this does not mean that solutions can actually be found for all mixed integer problems as there are limitations of computer memory and time.

A problem with the implementation is that an integer variables $B_{(u,v)}$ could take any non-negative integer value, but we want them to only take the values 0 and 1. To ensure this we use a trick: For the program we double every constraint from the original program and get four conditions for an edge (u,v) with edge value $[x,y]$ (compare Eqs 5 and 6):

$$v - u - \Delta_{(u,v)} \le y, \quad v - u - 100 \cdot B_{(u,v)} \le y,$$
$$v - u + \Delta_{(u,v)} \ge x, \quad v - u + 100 \cdot B_{(u,v)} \ge x. \tag{7}$$

The two original conditions measure once again the size of a deviation from a constraint and the two new conditions test if there actually is a deviation. The factor 100 in front of the $B_{(u,v)}$ is much bigger than any number that actually occurs in the calculation. If we look at one of the conditions from Eq. 7, say $v - u - 100 \cdot B_{(u,v)} \le y$ then if $v - u \le y$ the variable $B_{(u,v)}$ can assume the value 0. If on the other hand $v - u > y$ then the variable $B_{(u,v)}$ needs to be at least 1 for the second inequality in Eq. 7 to hold. (Remember, unlike $\Delta_{(u,v)}$ the variable $B_{(u,v)}$ is an integer and does not take values between 0 and 1.) Since the factor 100 is chosen to be a lot bigger than $|v - u - y|$ the inequality $v - u - y - 100 \cdot 1 \le 0$ always holds. Therefore, with $B_{(u,v)} = 1$ the constraint is fulfilled, and there is no need for any $B_{(u,v)}$ to be bigger than 1. By the same argument as above for $\Delta_{\max}$ we get

$$B_{(u,v)} = 0 \Leftrightarrow \Delta_{(u,v)} = 0 \text{ and } B_{(u,v)} = 1 \Leftrightarrow \Delta_{(u,v)} > 0$$

if the sum of all $B_{(u,v)}$ is included in the objective to be minimized. The results for the graph from **Figure 1** is shown in **Figure 5B** (compare to **Figure 3**).

Note that all three examples have the same sum of all deviations since it is primarily being optimized in all three cases. The differences lie in the second part of the objective, i.e. the maximal deviation and the number of deviations. Also note that the hierarchies shown here are again not the only optimal hierarchies that can be found for these criteria.

EMPIRICAL DATA

We use the data set described by Felleman and Van Essen (1991) for the visual system of the macaque monkey (FV91). As regions MDP and MIP had no constraints defined, they are not included here. The projections in this network were assigned ranges according to a modification of the original relationship classification scheme, which incorporates ideas presented in subsequent publications (Kennedy and Bullier, 1985; Barone et al., 2000; Batardiere et al., 2002), and permits a richer representation of hierarchical distance and the assignment of refined ranges (compare to Reid et al., 2009). As in Reid et al. (2009) we use the notation A+ for strongly ascending, A for ascending, L for lateral, D for descending and D+ for strongly descending projections. To investigate the effect of range sizes upon the resulting optimal hierarchies, ranges were varied from disjoint intervals with clear gaps between the five connection types to intervals that overlap so that the connection types are merging into one another. To implement this ranges were systematically expanded by 0.1 at each limit, resulting in 10 range sets, as presented in **Table 1**. The outer bounds for these hierarchies were chosen as 32, which is the total number of regions; this ensures that no individual projection can have a hierarchical distance greater than the total number of regions.

We employed two different optimization methods: For the first we minimized the objective 1,000·sum $\Delta_{(u,v)}$ + sum$B_{(u,v)}$. For the second we minimized the objective 1,000,000·sum$\Delta_{(u,v)}$ + 1,000$\Delta_{\max}$ + sum$B_{(u,v)}$.

Table 1 | The borders of the ranges for the different constraint sets.

	D+		D		L		A		A+	
	From	**To**	**From**	**To**	**From**	**To**	**From**	**To**	**From**	**To**
Set 0	–32	–2.0	–1.0	–1.–0	0.0	0.0	1.0	1.0	2.0	32
Set 1	–32	–1.9	–1.1	–0.9	–0.1	0.1	0.9	1.1	1.9	32
Set 2	–32	–1.8	–1.2	–0.8	–0.2	0.2	0.8	1.2	1.8	32
Set 3	–32	–1.7	–1.3	–0.7	–0.3	0.3	0.7	1.3	1.7	32
Set 4	–32	–1.6	–1.4	–0.6	–0.4	0.4	0.6	1.4	1.6	32
Set 5	–32	–1.5	–1.5	–0.5	–0.5	0.5	0.5	1.5	1.5	32
Set 6	–32	–1.4	–1.6	–0.4	–0.6	0.6	0.4	1.6	1.4	32
Set 7	–32	–1.3	–1.7	–0.3	–0.7	0.7	0.3	1.7	1.3	32
Set 8	–32	–1.2	–1.8	–0.2	–0.8	0.8	0.2	1.8	1.2	32
Set 9	–32	–1.1	–1.9	–0.1	–0.9	0.9	0.1	1.9	1.1	32

RESULTS

The two optimization methods produced similar, but not identical results. **Table 2** shows the values for the sum of deviations, the maximal deviation and the number of violations for the optimal hierarchies. For comparison the values from Reid et al. (2009) are also included. There only the sum of deviations was minimized and the optimization was performed using the QS-Opt Linear Problem Solver[2]. Note that the sum of deviations is the same for each set for all optimizations, since this was always the first objective.

With the exception of number of violations for set 0 the optimized values are getting smaller when the size of the constraining intervals is getting bigger. The sum of deviations goes down from 60.0 for set 0 to 8.8 for set 9, the maximal deviation is reduced from 3.0 for set 0 to 0.7 for set 9 with optimization method 2, the number of violations decreases from 49 for set 1 to 13 for set 9 with optimization method 1 and from 50 for set 1 to 17 for set 9 with optimization method 2.

For three sets (0, 5 and 6) we get exactly the same optimal values for both optimization methods, but only for one of these sets (set 0) are the calculated hierarchies identical. For set 5 and set 6 the two calculated hierarchies are not identical. However, for each set both hierarchies have the same number of violations and maximal deviations, which means that they are both optimal hierarchies for both optimization methods. The corresponding hierarchies from Reid et al. (2009) all have the same maximal deviation but larger numbers of violations, so regarding the number of violations these hierarchies are not optimal. In general the number of violations for the hierarchies from Reid et al. (2009) are higher than for the new optimization methods, but since the violations were not minimized before this is not surprising. However the maximal deviations tend to be small, often minimal (as seen in comparison with optimization methods 2), for these hierarchies, even though they were not minimized.

For all other sets than 0, 5 and 6, we see differences in values between the two new optimization methods, therefore there is a trade-off between the number of violations and the maximal deviations. Therefore these values cannot both be at their minimum for 7 of the 10 sets.

When we look at the violated constraints we find only a total number of 54 of the 386 constraints being violated by either of the two optimization methods. Of those nine constraints were violated by every set for both methods (see **Table 3**). Note that everything is counted as a constraint violation that does not exactly fit the classification of a connection. For example, if a connection is classified as ascending (A) but ends up being strongly ascending (A+) in the calculated hierarchy this counts as a violation. Also note that for all the connections listed in **Table 3** there are reciprocal connections

[2]http://www.isye.gatech.edu/~wcook/qsopt/

Table 2 | The results of the optimization.

Set	Sum of deviations (1st obj.)	Optimization method 1		Optimization method 2		Data from Reid et al. (2009)	
		Number of violations (2nd obj.)	Maximal deviation (not opt.)	Number of violations (3rd obj.)	Maximal deviation (2nd obj.)	Number of violations (not opt.)	Maximal deviation (not opt.)
0	60.0	42	3.0	42	3.0	49	3.0
1	51.8	49	2.7	50	2.6	57	2.6
2	43.8	42	2.4	45	2.2	54	2.2
3	36.2	42	2.1	45	1.8	53	1.8
4	29.2	38	1.8	40	1.6	42	1.6
5	22.5	25	1.5	25	1.5	27	1.5
6	18.0	26	1.4	26	1.4	27	1.4
7	14.4	19	1.3	20	1.2	22	1.3
8	11.6	13	1.4	18	0.9	20	1.0
9	8.8	13	1.1	17	0.7	19	1.0

Table 3 | The 9 constraints that are violated by all 10 sets for both objectives. The number 1 means the corresponding connection is classified as D+, D, L, A or A+.

Origin	V4	AITd	STPp	STPp	V2	V2	PO	PO	FST
Termination	V1	7a	FEF	FST	V3	VP	MSTd	LIP	TF
D+	–	–	–	1	–	–	–	–	–
D	1	1	1	–	–	–	–	–	–
L	–	1	–	–	–	–	–	–	1
A	–	1	–	–	–	–	–	–	–
A+	–	–	–	–	1	1	1	1	–

that were not consistently violated. This means that the classification of these reciprocal connections is not complementary. For example the connection V2 to V3 is classified A+, but the connection V3 to V2 is classified D (not D+).

The two optimization methods generally violated the same constraints: For the first optimization method 53 and for the second optimization method 52 different constraints (of 386) were violated over the 10 sets and of those 51 constraints were violated by both methods. In addition constraint V1 to V3 (A+), was violated by the second optimization in four sets and constraints LIP to V4 (D), and 46 to TH (L), were each violated by the first optimization in one set.

For comparison between sets we normalized the resulting hierarchies: by definition V1 always had the hierarchical value 0, the region with the highest hierarchical level was assigned the value 1, the hierarchical levels of the other areas were transformed accordingly. **Figure 6** shows the average hierarchical levels over the 10 sets for the first optimization method. The equivalent figure for the second optimization method looks very similar and is therefore not included here.

DISCUSSION

While the parameters for the hierarchies do show some differences between the optimization methods, the resulting hierarchies were remarkably similar. This does not mean that all hierarchies that are optimal for the two optimizations are similar: since we only have one possible solution per optimization method per set there might also be solutions that differ substantially. What it does mean is that there are examples for optimal hierarchies for the two optimization methods for which the differences between the optimized values seem to be accomplished through minor changes within the hierarchies.

The boundaries chosen for the optimization constraints (set 0 to 9) seem to have a big influence on the optimization results. All optimized values for set 0 are several times as big as the corresponding values for set 9. The "looser" the boundaries are (i.e., the larger the defining intervals) the smaller the optimized values. This is because in the bigger intervals for the connection classes more values can be assumed in the optimization without violating the constrains. For example, if an ascending connection is assigned the value 1.5, this is a violation in sets 0 through 4, but not in sets 5 through 9. Therefore there are fewer violations and, as a result, smaller optimized values for sets that use bigger intervals for connection classes. However in Reid et al. (2009) we found that the resulting hierarchies are very comparable across sets after normalization. There were only minimal changes in the order of the areas in the calculated hierarchies for the different conditions.

The similarity in the violated constraints across methods suggest that these constraints may be erroneous. In particular, of the nine constraints that are violated in this study by all constraint sets, and for both objectives, eight were violated in the calculations of Reid et al. (2009) for all constraint sets as well. [The last of the nine constraints, PO to LIP (A+), was violated by eight sets in Reid et al. (2009)]. While it is possible that a solution exists where

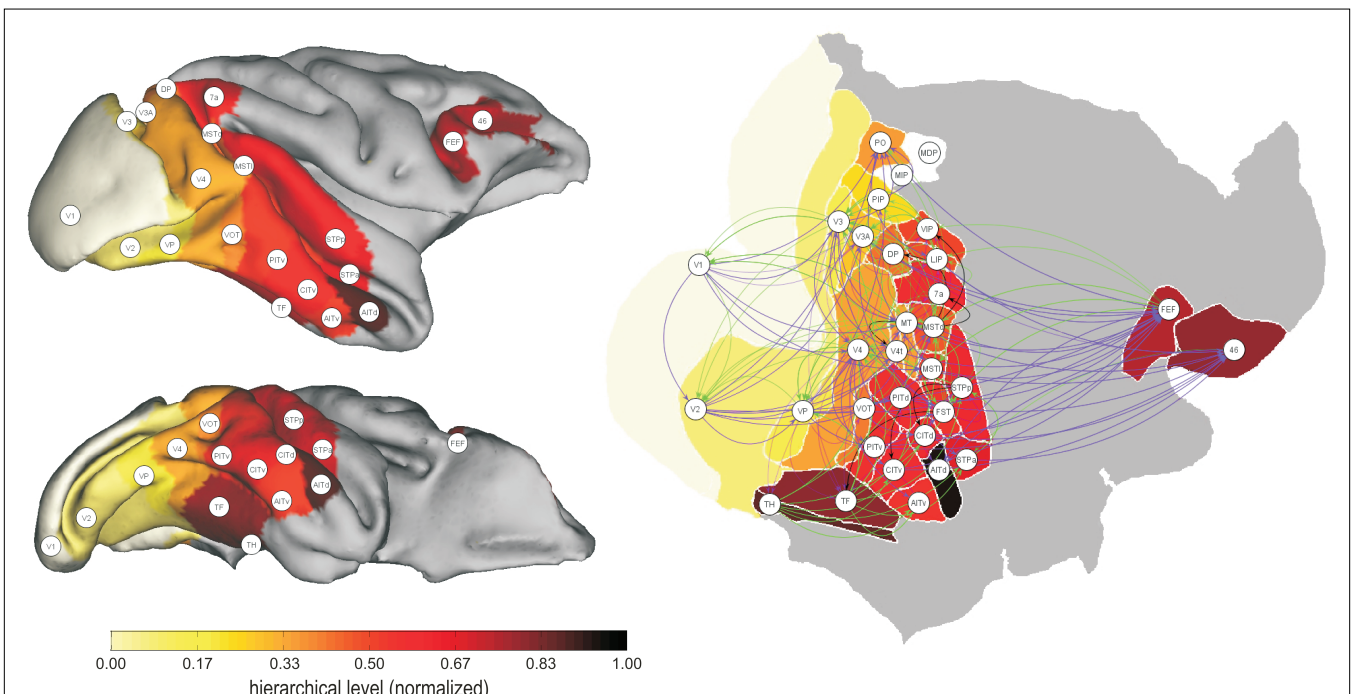


FIGURE 6 | Geometrical distribution of hierarchy levels in the FV91 visual network of the macaque, both as three-dimensional cortical surface renderings (left), and as a two-dimensional "flat map" representation of the cortical sheet (right). Regions are coloured by their mean normalized hierarchical position, obtained by the first optimization method over the 10 constraint sets described in **Table 1**. Directed edges in the flat map illustration represent interregional connections, and are coloured according to projection class: A+ (purple, opaque), A (purple, transparent), L (black), D (green, transparent), and D+ (green, opaque). Compare this representation to Figures 2 and 4 in Felleman and Van Essen (1991).

none of these constraints are violated, this consistency suggests a pattern which is worth further investigation. The fact that all of these connections have reciprocal connections whose constraints are not consistently violated seems to suggest that the classification of the reciprocal connections fits an optimal hierarchy better than the classification of the nine violated connections. Since the classification of the reciprocal connections is not complementary at most one of the classifications (the one for the connection or the one for the reciprocal connection) can be correct. This does not necessarily mean that the classification is better for the reciprocal connection since in some cases this classification is very broad, spanning several of our five classes.

The classical optimization problem of minimizing the number of violations [as done by Felleman and Van Essen (1991) as well as Hilgetag et al. (1996)] can theoretically be solved by the method presented here, where it was only used as an secondary objective. In practical terms, solving for number of violations as a primary objective has proven too computationally expensive, whereas using it only as a secondary objective made the calculation easier, since the solutions were already limited by the minimal sum of deviations. However, we were able to minimize just the number of violations for sets 8 and 9, which produced the smallest optimal values for all the other optimization methods. The results (12 violated constraints with a sum of deviation of 13.4 for set 8, and 11 violated constraints with a sum of deviations of 10.7 for set 9) are only slightly below the number of violated constraints for optimization method 1.

The hierarchies calculated here are of course also solutions for the original optimization problem from Reid et al. (2009), since they have the same minimal sum of deviations. Additionally, however, they also have either a minimal number of violations or a minimal maximum deviation, and in some cases even both. This added optimality does not seem to be especially advantageous, however, given that the resulting hierarchies remained for the most part unchanged, which suggests that they are not highly sensitive to the addition of these criteria. Other objectives are also conceivable, of course, which may result in more strongly altered hierarchies than we report here. For instance, it is conceivable that further knowledge about the reliability of the anatomical data underlying our constraints may yield more informative objective criteria, that would allow a Bayesian approach to this problem. Another option is not to allow any deviation for connections that are classified with a great certainty, ensuring that the connection appears as classified in the hierarchy. However, this cannot be done for all connections since the resulting constraints are not consistent and some deviation needs to be allowed to find an hierarchy.

In the choice of the optimization method the data quality should be considered. If we expect that most of the connections are correctly classified but there might be some classifications that are erroneous then minimizing the number of violations is the better option. Most of the distance information of the classification should then be preserved in the hierarchy. Assuming that the wrongly classified connections are the ones that do not fit the hierarchy these would be the ones that are disregarded. They would be "taken out" of the hierarchy. If on the other hand, we expect all the classifications to be equally correct or faulty (or just an approximation of the true value), then we want to consider all the information to the same degree. This can imply changing many classifications a little to "squeeze" the information into a hierarchy, such that, ideally, none of the information is disregarded completely, while many of the classifications might be "corrected" a little.

Since our results produced highly similar hierarchies across a variety of constraint sets for the two optimization methods presented here, both appear equally suited for the calculation of optimal hierarchies. It is thus a matter of personal preferences which one to employ, and this choice is one that can conceivably be expanded to accommodate alternative definitions of optimality. This leads us to conclude that there is no unique optimal hierarchy, not only because of the quality of the empirical data and the freedom to choose boundaries for the defining intervals, but also because there is more than one way to define optimality. While the method described does not provide a unique optimal hierarchy, it can produce hierarchies that are optimal in more than one way.

ACKNOWLEDGMENTS

Rolf Kotter and Egon Wanke received funding from the DFG (KO 1560/6-2, WA 674/10-2). Andrew T. Reid, Gleb Bezgin and Rolf Kotter were funded by a collaborative network grant from the McDonnell Foundation. The authors acknowledge funding received by the Initiative and Networking Fund of the Helmholtz Association within the Helmholtz Alliance on Systems Biology.

REFERENCES

Barone, P., Batardiere, A., Knoblauch, K., and Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas v1 and v4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci.* 20, 3263–3281.

Batardiere, A., Barone, P., Knoblauch, K., Giroud, P., Berland, M., Dumas, A. M., and Kennedy, H. (2002). Early specification of the hierarchical organization of visual cortical areas in the macaque monkey. *Cereb. Cortex* 12, 453–465.

Dantzig, G.B. (1963). *Linear Programming and Extensions*. Princeton, NJ: Princeton University Press.

Felleman, D. J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Hilgetag, C.C., O'Neil, M. A., and Young, M. P. (1996). Indeterminate organization of the visual system. *Science* 271, 776–777.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica* 4, 373–395.

Kennedy, H., and Bullier, J. (1985). A double-labelling investigation of the afferent connectivity to cortical areas v1 and v2 of the macaque monkey. *J. Neurosci.* 5, 2815–2830.

Land, A.H., and Doig, A.G. (1960). An automatic method of solving discrete programming problems. *Econometrica* 28, 497–520.

Papadimitiou, C. H., and Steiglitz, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Mineola, NY: Dover Publications Inc.

Reid, A. T., Krumnack, A., Wanke, E., and Kötter, R. (2009). Optimization of cortical hierarchies with continuous scales and ranges. *Neuroimage* 47, 611–617.

Vezoli, J., Falchier, A., Jouve, B., Knoblauch, K., Young, M., and Kennedy, H. (2004). Quantitative analysis of connectivity in the visual cortex: extracting function from structure. *Neuroscientist* 10, 476–482.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 August 2009; paper pending published: 22 September 2009; accepted: 09 March 2010; published online: 31 March 2010.
*Citation: Krumnack A, Reid AT, Wanke E, Bezgin G and Kötter R (2010) Criteria for optimizing cortical hierarchies with continuous ranges. Front. Neuroinform. **4**:7. doi: 10.3389/fninf.2010.00007*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 30 October 2009
doi: 10.3389/neuro.11.037.2009

Hierarchical modularity in human brain functional networks

David Meunier[1,2], Renaud Lambiotte[3], Alex Fornito[1,2,4], Karen D. Ersche[1,2] and Edward T. Bullmore[1,2,5]*

[1] *Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, UK*
[2] *Behavioural and Clinical Neurosciences Institute, University of Cambridge, Cambridge, UK*
[3] *Institute for Mathematical Sciences, Imperial College, London, UK*
[4] *Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne, VIC, Australia*
[5] *GSK Clinical Unit Cambridge, Addenbrooke's Hospital, Cambridge, UK*

Edited by:
Marcus Kaiser,
Newcastle University, UK

Reviewed by:
Roger Guimera,
Northwestern University, USA
Pedro Valdes-Sosa, Cuban
Neuroscience Center, Cuba

****Correspondence:***
Edward T. Bullmore, Brain Mapping Unit, Herchel Smith Building, Robinson Way, Cambridge CB2 0SZ, UK.
e-mail: etb23@cam.ac.uk

The idea that complex systems have a hierarchical modular organization originated in the early 1960s and has recently attracted fresh support from quantitative studies of large scale, real-life networks. Here we investigate the hierarchical modular (or "modules-within-modules") decomposition of human brain functional networks, measured using functional magnetic resonance imaging in 18 healthy volunteers under no-task or resting conditions. We used a customized template to extract networks with more than 1800 regional nodes, and we applied a fast algorithm to identify nested modular structure at several hierarchical levels. We used mutual information, $0 < I < 1$, to estimate the similarity of community structure of networks in different subjects, and to identify the individual network that is most representative of the group. Results show that human brain functional networks have a hierarchical modular organization with a fair degree of similarity between subjects, $I = 0.63$. The largest five modules at the highest level of the hierarchy were medial occipital, lateral occipital, central, parieto-frontal and fronto-temporal systems; occipital modules demonstrated less sub-modular organization than modules comprising regions of multimodal association cortex. Connector nodes and hubs, with a key role in inter-modular connectivity, were also concentrated in association cortical areas. We conclude that methods are available for hierarchical modular decomposition of large numbers of high resolution brain functional networks using computationally expedient algorithms. This could enable future investigations of Simon's original hypothesis that hierarchy or near-decomposability of physical symbol systems is a critical design feature for their fast adaptivity to changing environmental conditions.

Keywords: graph theory, brain, network, modularity, hierarchy, near-decomposability, information

INTRODUCTION

Almost 50 years ago, Herbert Simon wrote an essay entitled "The architecture of complexity" (Simon, 1962). In this prescient analysis, he argued that most complex systems, such as social, biological and physical symbolic systems, are organized in a hierarchical manner. He introduced the notion of "nearly-decomposable systems", i.e. systems where elements have most of their interactions (of any kind) with a subset of elements in some sense close to them, and much less interaction with elements outside this subset. In mainstream contemporary parlance, Simon's near-decomposability is closely analogous to the concept of topological modularity: nodes in the same module have dense intra-modular connectivity with each other and sparse inter-modular connectivity with nodes in other modules (Newman, 2004, 2006). Simon argued that near-decomposability was a virtually universal property of complex systems because it conferred a very important evolutionary or adaptive advantage. Decomposability, or modularity, accelerates the emergence of complex systems from simple systems by providing stable intermediate forms (component modules) that allow the system to adapt one module at a time without risking loss of function in other, already-adapted modules.

Our understanding of complexity has progressed considerably since that time, partly due to the availability of large data-sets that now allow us to explore empirically the architecture of complex systems and thereby to feedback on theoretical considerations (Strogatz, 2001; Amaral and Ottino, 2004). Many complex systems can be represented using tools drawn from graph theory as networks of nodes linked by edges. Such networks have been used to represent a broad variety of systems, ranging from genetic and protein networks to the World Wide Web. The huge size of some of these systems (~10 billion nodes in the WWW) has driven the development of new statistical tools in order to characterize their topological properties (Newman, 2003).

A quantity called modularity has been introduced in order to measure the decomposability of a network into modules (Guimerà et al., 2004; Newman and Girvan, 2004). Modularity can be used as a merit function to find the optimal partition of a network. The resulting partition has been shown to reveal important network community structures in a variety of contexts, e.g. the global air transportation network (Guimerà et al., 2005) and gene expression interactomes (Oldham et al., 2008) are two diverse examples of complex systems with topological modularity. However, in systems having an intrinsic hierarchical structure, finding a single partition is not satisfactory. Several approaches have therefore been proposed in order to allow for more flexibility and to uncover communities at different hierarchical levels. Among those multi-scale approaches, there are algorithms searching for local minima of

the modularity landscape (Sales-Pardo et al., 2007) or modifying the adjacency matrix of the graph in order to change its typical scale (Arenas et al., 2008). Another class of methods consists in modifying modularity by incorporating in it a resolution parameter (Reichardt and Bornholdt, 2006). This allows one to "zoom in and out" of a modular hierarchy in order to find communities on different levels; for example, the resolution parameter can be interpreted as the time scale of a dynamical process unfolding on a network (Lambiotte et al., 2009).

There is already strong evidence that brain networks have a modular organization; see Bullmore and Sporns (2009) for review. Some support comes from non-human data, like the anatomical networks in felines and primates (Hilgetag et al., 2000) or functional networks in rodents (Schwarz et al., 2008). Recently, human neuroimaging studies have also provided evidence for comparable modular organization in both anatomical (Chen et al., 2008) and functional (Ferrarini et al., 2009; Meunier et al., 2009) brain networks. However, a limitation of these previous neuroimaging studies has been the computational time required to derive a modular decomposition (Brandes et al., 2006), thus limiting the size of the networks under study. In addition, these studies were limited to studying modularity at one particular level of community structure, neglecting consideration of possible sub-modular communities at lower levels. Finally, it has been a taxing problem to quantify the topological similarity between two or more modular decompositions, with most investigators simply examining modularity on the basis of an averaged connectivity matrix estimated from a group of individuals.

In this study, we report on progress towards addressing each of these issues. We applied a recently developed, computationally efficient algorithm (Blondel et al., 2008) to derive a hierarchical, modular decomposition of human brain networks measured using functional magnetic resonance imaging (fMRI) in 18 healthy volunteers. By providing rapid decomposition, the algorithm enabled us to study the modular structure of whole brain networks on a larger scale (thousands of equally sized nodes) than previously possible (tens of differently sized nodes), with concomitant improvements in the spatial or anatomical resolution of the network, while simultaneously avoiding biases associated with using *a priori* anatomical templates that are inevitably somewhat arbitrary in their definition of regions-of-interest (Tzourio-Mazoyer et al., 2002). Thus, the method enabled rapid, high-resolution, hierarchical modular decomposition of brain functional networks constructed from individual fMRI datasets. In addition, we present a method for comparing the similarity or mutual information between two modular community structures obtained for different subjects, and use it to identify the single, "most representative" subject whose brain network modularity was most similar to that of all the other networks in a sample of 18 healthy participants.

MATERIALS AND METHODS

EXPERIMENTAL DATA

Study sample

Eighteen right-handed healthy volunteers (15 male, 3 female) were recruited from the GlaxoSmithKline (GSK) Clinical Unit Cambridge, a clinical research facility in Addenbrooke's Hospital, Cambridge, UK. All volunteers (mean age 32.7 years ± 6.9 SD) had a satisfactory medical examination prior to study enrolment and were screened for any other current Axis I psychiatric disorder using the Structured Clinical Interview for the DSM-IV-TR Axis I Disorders (SCID). Participants were also screened for normal radiological appearance of structural MRI scans by a consultant neuroradiologist, and female participants were screened for pregnancy. Urine samples were used to confirm abstinence from illicit drugs and breath was analysed to ensure that no participant was under the influence of acute alcohol intoxication. All volunteers provided written informed consent and received monetary compensation for participation. The study was reviewed and approved by the Cambridge Local Research Ethics Committee (REC06/Q0108/130; PI: TW Robbins).

Functional MRI data acquisition

Whole-brain echoplanar imaging (EPI) data depicting BOLD contrast were acquired at the Wolfson Brain Imaging Centre, University of Cambridge, UK, using a Siemens Magnetom Tim Trio whole body scanner operating at 3 T with a birdcage head transmit/receive coil. Gradient-echo, EPI data were acquired for the whole brain with the following parameters: repetition time = 2000 ms; echo time = 30 ms, flip angle = 78°, slice thickness = 3 mm plus 0.75 mm interslice gap, 32 slices parallel to the inter-commissural (AC-PC) line, image matrix size = 64 × 64, within-plane voxel dimensions = 3.0 mm × 3.0 mm.

Participants were asked to lie quietly in the scanner with eyes closed during the acquisition of 300 images. The first four EPI images were discarded to account for T1 equilibration effects, resulting in a series of 296 images, of which the first 256 images were used to estimate wavelet correlations.

Functional MRI data preprocessing

The images were corrected for motion and registered to the standard stereotactic space of the Montreal Neurological Institute EPI template image using an affine transform (Suckling et al., 2006). Time series were then extracted using a whole brain, high resolution, regional parcellation of the images, implemented in the following manner; see **Figure 1A**. First, a binarized version of a commonly used template image (Tzourio-Mazoyer et al., 2002) was used as a broad grey matter mask. Second, each 8 mm^3 voxel in this mask was downsampled by a factor of 4 such that each equally sized region in the parcellation comprised 4 × 4 × 4 voxels of the original image. This initial parcellation included some regions of the image which were not largely representative of grey matter: these were excluded from further analysis by applying the criteria that each region must be at least 50% overlapping with the grey matter mask and must contain at least 80% voxels having BOLD signal (defined operationally as mean signal intensity >50). To be included in the definitive parcellation scheme (which comprised 1808 regional nodes), a region had to satisfy these two inclusion criteria for every individual dataset in the sample.

The mean time series of each region was extracted and wavelet-filtered using Brainwaver R package[1] (Achard et al., 2006; Achard and Bullmore, 2007). The wavelet correlation coefficient

[1]http://cran.r-project.org/web/packages/brainwaver/index.html

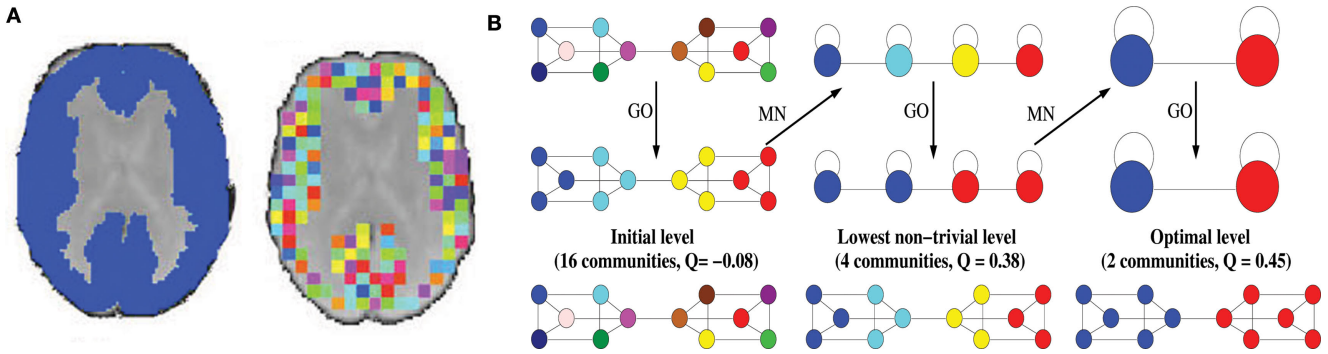


FIGURE 1 | Methods. (A) Downsampled template. Starting from a binary version of the AAL template (left), the downsampling procedure will produce a template of small (64 voxels), equal size regions covering the original template (right). **(B)** Illustration of the Louvain method on a simple hierarchical graph. The algorithm starts by assigning a different module to each node (16 modules of single nodes). The method then consists of two phases that are repeated iteratively. The first phase is a greedy optimization (GO) where nodes adopt the community of one of their neighbours if this action results in an increase of modularity (typically, the community of the neighbour for which the increase is maximal is chosen). The second phase builds a meta-network (MN) whose nodes are the communities found in the first phase. We denote by "pass" a combination of these two phases. The passes are repeated until no improvement of modularity is possible and the optimal partition is found. When applied on this graph, the algorithm first finds a lowest non-trivial level made of four communities. The next level is the optimal level and is made of two communities.

was estimated for each of four wavelet scales between each pair of nodes, resulting in a $\{1808 \times 1808\}$ association matrix, or frequency-dependent functional connectivity matrix, for each wavelet scale in the overall frequency range 0.25–0.015 Hz. In what follows, we will focus on results at wavelet scale 3, subtending a frequency interval of 0.06–0.03 Hz.

This choice of frequency interval was guided by the fact that prior work on resting-state fMRI functional connectivity has found that the greatest power in connectivity occurs in frequency bands lower than 0.1 Hz (Cordes et al., 2001). However, analysing very low frequency scales in limited time series such as those acquired with fMRI can reduce precision in estimating inter-regional wavelet correlations (Achard et al., 2006). So scale 3 was chosen for the focus of this study as representing a reasonable compromise between retaining sufficient estimation precision while measuring low frequency network properties.

Each association matrix was thresholded to create an adjacency matrix A, the $a_{i,j}$th element of which is either 1, if the absolute value of the wavelet correlation between nodes i and j, $w_{i,j}$, exceeds a threshold value τ; or 0, if it does not. We have chosen here to take a high threshold, leading to very sparse networks comprising 8000 edges, i.e. with a connection density of 0.5% of all possible edges in a network of this size. Modularity of neuroimaging networks is typically greater (Meunier et al., 2009), and computational costs are lower, when the networks are more sparsely thresholded. Up to 10% of nodes were disconnected from the rest of the network at this threshold.

GRAPH THEORETICAL ANALYSIS

Hierarchical modularity

In recent years, many methods have been proposed to discover the modular organization of complex networks. A key step was taken when Girvan and Newman popularized graph-partitioning problems by introducing the concept of modularity. Modularity is by far the most widespread quantity for measuring the quality of a partition $\mathcal{P}$ of a network. In its original definition, an unweighted and undirected network that has been partitioned into communities has modularity (Newman and Girvan, 2004):

$$Q = \frac{1}{2m} \sum_{C \in P} \sum_{i,j \in C} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \quad (1)$$

where A is the adjacency matrix of the network; m is the total number of edges; and $k_i = \sum_j A_{ij}$ is the degree of node i. The indices i and j run over the N nodes of the graph. The index C runs over the modules of the partition $\mathcal{P}$. Modularity counts the number of edges between all pairs of nodes belonging to the same community or module, and compares it to the expected number of such edges for an equivalent random graph. Modularity therefore evaluates how well a given partition concentrates the edges within the modules.

A popular method for discovering the modules of a network consists in optimizing modularity, namely in finding the partition having the largest value of Q. However, it is typically impossible computationally to sample modularity exhaustively by enumerating all the possible partitions of a network into communities. Several heuristic algorithms have therefore been proposed to provide good approximations, and so to allow for the analysis of large networks in reasonable times. The computational expediency of the algorithm has become a crucial factor due to the increasing size of the networks to be analysed.

More recently, methods to study hierarchical modularity, also called nested modularity, have been introduced (Sales-Pardo et al., 2007; Arenas et al., 2008; Rosvall and Bergstrom, 2008). In this case, each module obtained at the partition of the highest level can further be decomposed into sub-modules, which in turn can be decomposed into sub-submodules, and so on. Here, we will use a multi-level method which was introduced very recently in order to optimize modularity (Blondel et al., 2008); see **Figure 1B**. The primary advantages of this method are that it unfolds a complete

hierarchical community structure for the network and outperforms previous methods with respect to computation time. This so-called "Louvain method" takes advantage of the hierarchical organization of complex networks in order to facilitate the optimization. The algorithm starts by assigning a different module to each node of the network. The initial partition of the network is therefore made of N communities. It then consists of two phases that are repeated iteratively. The first phase consists in a greedy optimization where nodes are selected sequentially in an order that has been randomly assigned. When a node is selected, it may leave its community and adopt a community which is in its direct neighbourhood, but only if this change of community leads to an increase of modularity (GO on **Figure 1B**). The second phase builds a new network whose meta-nodes are the communities found in the first phase (MN on **Figure 1B**). Let us denote by "pass" a combination of these two phases. These passes are repeated iteratively until a maximum of modularity is attained and an optimal partition of the network into communities is found. By construction, the meta-nodes, or intermediate communities, are made of more nodes at subsequent passes. The optimization is therefore done in a multi-scale way: among adjacent nodes at the first pass, among adjacent meta-nodes at the second pass, etc. The output of the algorithm is a set of partitions, one for each pass. The optimal partition is the one found at the last pass. It has been shown on several examples that modularity estimated by this method is very close to the optimal value obtained from slower methods (Blondel et al., 2008). Intermediate partitions can also be shown to be meaningful and to correspond to communities at intermediate resolutions (see Section "Discussion"). In the following, we will call "lowest non-trivial level" the partition found after the first pass.

Node roles

Once a maximally modular partition of the network has been identified, it is possible to assign topological roles to each node based on its density of intra- and inter-modular connections (Guimerà and Amaral, 2005a,b; Guimerà et al., 2005; Sales-Pardo et al., 2007).

Intra-modular connectivity is measured by the normalized within-module degree,

$$z_i = \frac{\kappa_{n_i} - \overline{\kappa}_n}{\sigma_{\kappa_n}} \tag{2}$$

where κ_{n_i} is the number of edges connecting the ith node to other nodes in the nth module, $\overline{\kappa}_n$ is the average of κ_{n_i} over all nodes in the module n, and σ_{κ_n} is the standard deviation of the intra-modular degrees in the nth module. Thus z_i will be large for a node that has a large number of intra-modular connections.

Inter-modular connectivity is measured by the participation coefficient,

$$P_i = 1 - \sum_{n=1}^{N} \left(\frac{\kappa_{n_i}}{k_i} \right)^2 \tag{3}$$

where κ_{n_i} is the number of edges linking the ith node to other nodes in the nth module, and k_i is the total degree of the ith node. Thus P_i will be close to 1 if the ith node is extensively linked to all other modules in the community and 0 if it is linked exclusively to other nodes in its own module.

The two-dimensional space defined by these parameters, the $\{P, z\}$ plane, can be partitioned to assign categorical roles to the nodes of the network. Contrarily to our previous study (Meunier et al., 2009), where we used a simplified definition of node roles, the higher number of nodes examined in the current study allowed us to adopt the original definitions of node roles as described for large metabolic (Guimerà et al., 2005) and transportation networks (Guimerà and Amaral, 2005b):

- The hubness of a node can be defined by its within-module degree: If a given node i has a value of $z_i > 2.5$. It is classified as a hub, otherwise as a non-hub.
- The limits for the participation coefficient are different for hubs and non-hubs. For non-hubs, if a given node has value $0 < P_i < 0.05$, the node is classified as an *ultra-peripheral node*, $0.05 < P_i < 0.62$ corresponds to a *peripheral node*, $0.62 < P_i < 0.80$ corresponds to a *connector node*, and $0.80 < P_i < 1.0$ is a *kinless node*. For hubs, $0 < P_i < 0.30$ corresponds to a *provincial hub*, $0.30 < P_i < 0.75$ corresponds to a *connector hub*, and $0.75 < P_i < 1.0$ is a *kinless hub*.

These different categories allowed us to classify the nodes according to their topological functions in the network. For example, a provincial hub is a hub with greater intra- vs inter-modular connectivity, thus having a pivotal role in the function realized by its module, whereas a connector hub will play a central role in transferring information from its module to the rest of the network.

The results of modular decomposition were visualized in anatomical space using Caret software for cortical surface mapping[2], and in topological space using Pajek software[3].

Similarity measure

To compare the different modularity partitions obtained at different hierarchical levels in the same subject, or at the same hierarchical level in different subjects, we used the normalized mutual information, as defined in Kuncheva and Hadjitodorov (2004). For two given partitions A and B, with a number of communities denoted C_A and C_B:

$$I(A,B) = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} N_{ij} \log\left(\frac{N_{ij}N}{N_{i.}N_{.j}}\right)}{\sum_{i=1}^{C_A} N_{i.} \log\left(\frac{N_{i.}}{N}\right) + \sum_{j=1}^{C_B} N_{.j} \log\left(\frac{N_{.j}}{N}\right)} \tag{4}$$

where N_{ij} is the number of nodes in common between modules i and j, the sum over row i of matrix N_{ij} is denoted N_i, and the sum over column j is denoted $N_{.j}$. If the two partitions are identical then $I(A,B)$ takes its maximum value of 1. If the two partitions are totally independent, $I(A,B) = 0$.

The initial application of this quantity was to evaluate different modularity partition algorithms (Danon et al., 2005). The similarity index was used to compute how closely the partitions obtained

[2]http://brainmap.wustl.edu/register.html

[3]http://vlado.fmf.uni-lj.si/pub/networks/pajek

from different algorithms matched the "target" partition of a given test network, i.e. a network whose modular structure was known *a priori*. Here the application was different, since we wanted to compare partitions obtained for different subjects in a group. Since the equation is symmetric in A and B, it is however possible to use the index without a target partition.

The networks constructed for each individual had the same number of nodes N, so the partitions of each subject have the same number of nodes. However, due to the high threshold applied to construct the adjacency matrix, the number of disconnected nodes in the networks can be different for each subject. One solution is to consider each disconnected node as a single module. In this case, each node (disconnected or not) of the network will be in the set of modules of each subject. However, it introduces artificially high values in the similarity values, especially if the networks of two subjects have similar sets of disconnected edges. So we have chosen to remove the disconnected nodes from the partitions and study only the partitions obtained on the giant component of each network, but keeping the value of N in the equation as the total number of nodes. This leads to a value of similarity slightly lower than if the disconnected nodes were included in the partitions, but is more representative of the relevant set of connected modules.

RESULTS

SIMILARITY AND VARIABILITY OF MODULAR DECOMPOSITIONS

It was possible to define a hierarchical modular decomposition for each of the 18 subjects in the sample. At the highest hierarchical level, the mean brain functional network modularity for the group was 0.604, with SD = 0.097. By comparison, modularity at the highest level for 18 random networks with an equivalent number of nodes (1808) and edges (8000) was 0.303 (SD = 0.003). There was a significant increase in brain network modularity compared to random network modularity (Kolmogorov–Smirnov test, $D = 1$, $P \sim 2^{-10}$).

The similarity of network community structure between each pair of subjects, at each level of the hierarchy, was calculated using Eq. 4. The resulting similarity matrices for level 3 (the highest level) and level 1 (the lowest non-trivial level) are shown in **Figure 2**.

The average pairwise similarity was 0.57 at level 3 and 0.63 at level 1, indicating a reasonable degree of consistency between subjects in modular organization of functional networks. The similarity between subjects was highly correlated over levels of the modular hierarchy: for example, if a pair of networks had a similar modular partition at the highest level, the sub-modular organization at lower levels was also similar.

Simply by summing the pairwise similarity scores for each row of the similarity matrix, it was possible to identify the individual subject (number 2) that was most similar to all other subjects in the sample, i.e. the most representative subject, and the subject (4) that was least similar to the rest of the sample. In what follows, we will focus attention on the modular decomposition of the most representative subject.

HIERARCHICAL MODULARITY

The hierarchical modular decomposition of the most representative subject's brain functional network is shown in **Figure 3**. At the highest level of the hierarchy (level 3), there were eight large modules, each comprising more than 10 nodes. At the lowest level of the hierarchy (level 1), there were 57 sub-modules. The largest five modules (with putative functional interpretations) and their sub-modular decomposition are briefly described below; some additional details are provided in **Table 1**.

- Central module (somatosensorimotor): The largest high level module comprised extensive areas of lateral cortex in premotor, precentral and postcentral areas, extending inferiorly to superior temporal gyrus, as well as to premotor and dorsal cingulate cortex medially. At a lower hierarchical level, medial and

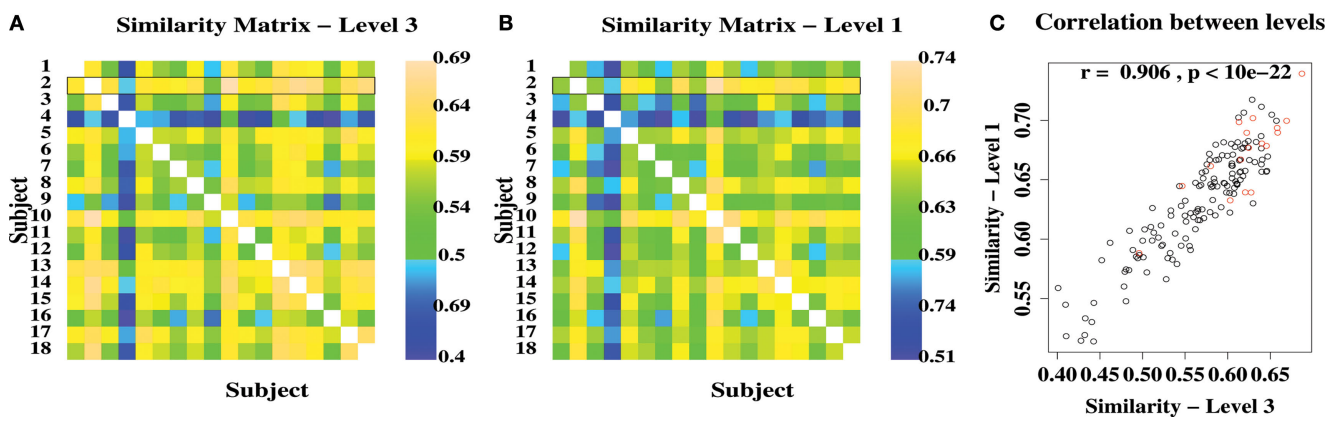


FIGURE 2 | Variability and similarity of brain functional network community structure between 18 different subjects. (A) Matrix showing the between-subject similarity measure for community structure at the highest level of the modular hierarchy. The pairwise similarity scores for the most representative subject are highlighted by a black rectangle. **(B)** Matrix showing the between-subject similarities for community structure at the lowest level of the modular hierarchy. **(C)** Scatter plot showing strong correlation of between-subject similarities at high and low levels of the modular hierarchy. Red points are similarities for the most representative subject.

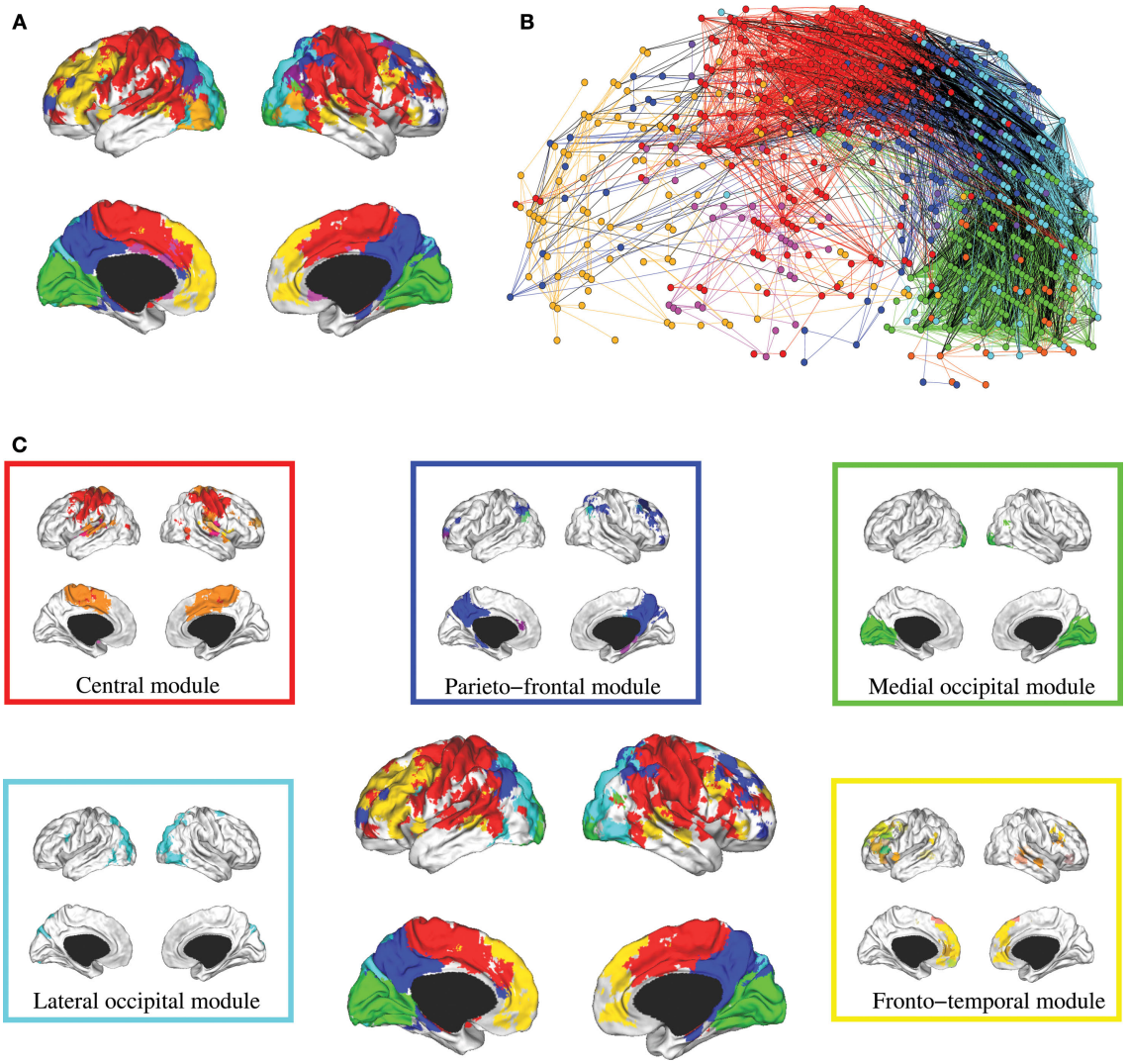


FIGURE 3 | Hierarchical modularity of a human brain functional network. **(A)** Cortical surface mapping of the community structure of the network at the highest hierarchical level of modularity, showing all modules that comprise more than 10 nodes. **(B)** Anatomical representation of the connectivity between nodes in colour-coded modules. The brain is viewed from the left side with the frontal cortex on the left of the panel and the occipital cortex on the right of the panel. Intra-modular edges are coloured differently for each module; inter-modular edges are drawn in black. **(C)** Sub-modular decomposition of the five largest modules (shown centrally) illustrates that the medial occipital module has no major sub-modules whereas the fronto-temporal modules has many sub-modules.

Table 1 | The five largest modules of the human brain functional network in a representative normal volunteer, indicating the number and type of nodes and sub-modules.

Module description	# Nodes	Connector nodes	Provincial hubs	Connector hubs	Sub-modules	Size of sub-modules
Central (sensorimotor)	239	8	1	4	11	115, 96, 8, 4, 3 *(2)*, 2 *(5)*
Parieto-frontal (default/attention)	138	10	1	0	10	115, 3 *(5)*, 2 *(4)*
Medial occipital (primary visual)	132	3	0	0	1	132
Lateral occipital (secondary visual)	101	7	0	1	1	101
Fronto-temporal (symbolic)	89	0	2	3	24	19, 8, 6, 5 *(2)*, 4, 3 *(6)*, 2 *(12)*

lateral cortex were segregated in different sub-modules and, within lateral cortex, precentral and postcentral areas were segregated from superior temporal cortex.

- Parieto-frontal module (default/attentional): This module comprised medial posterior parietal and posterior cingulate cortex, extending to medial temporal lobe structures inferiorly, and

areas of inferior parietal and dorsal prefrontal cortex laterally.
- Medial occipital module (primary visual): This module comprised medial occipital cortex and occipital pole, including primary visual areas.
- Lateral occipital (secondary visual): This module comprised dorsal and ventral areas of lateral occipital cortex, including secondary visual areas.
- Fronto-temporal module (symbolic): This module comprised dorsal and ventral lateral prefrontal cortex, medial prefrontal cortex, and areas of superior temporal cortex. It was less symmetrically organized than most of the other high level modules and was decomposed to a larger number of sub-modules at lower levels.

Note that most high level modules are bilaterally symmetrical, comprise both lateral and medial cortical areas, and tend to be spatially concentrated in an anatomical neighbourhood. Sub-modular decomposition sometimes resulted in a dominant sub-module, comprising most of the nodes in the higher level module, with some much smaller sub-modules each comprising a few peripheral nodes. For example, this was the pattern for the occipital modules. An alternative result was a more even-handed decomposition of a high level module into multiple sub-modules; this was the pattern for the prefronto-temporal module. In Simon's terminology, the number of sub-modules into which a module can be decomposed is its span of control, and so we can describe occipital modules as having a greater span of control than, say, the fronto-temporal module.

NODE ROLES

On the basis of the highest level (level 3) of modular decomposition, we assigned topological roles to each of the regional nodes. A node was defined as a hub or non-hub (more or less highly connected) with a provincial, connector or kinless role (depending on its balance of intra- vs inter-module connectivity). Provincial hubs will play a key role in intra-modular processing; connector hubs will play a key role in inter-modular processing.

Figure 4 displays an example of the node roles obtained from the most representative subject. **Figure 4A** shows the participation coefficient (P, our measure of inter-modular connectivity) vs the intra-modular degree (z, our measure of hubness) for each regional node in the network. Most nodes (416, 53%) have no inter-modular connections $P = 0$, but some (28, 4%) have a high proportion of inter-modular connections, qualifying for connector status. **Figure 4B** is a spatial representation of the node roles, the locations of the nodes corresponding to their position in three-dimensional stereotactic space. **Figure 4C** is a topological representation obtained by applying the Fruchterman–Reingold algorithm (Fruchterman and Reingold, 1991) to the network displayed in **Figure 4B**. In this representation, the distances between the nodes are not related to their spatial location, but to how strongly linked connected they are to their neighbours. The main idea is to start from an initial random placement of the nodes, and replace the edges by springs, letting the equivalent mechanical system evolve until it reaches a stable mechanical state. Thus, this representation locates nodes with similar connectivity patterns closer together in space.

We can see that most nodes (743, i.e. 95% of the nodes) have either the role of ultra-peripheral nodes or peripheral nodes and a small minority (39, i.e. 5% of the nodes) have the topologically important roles of hubs and/or connector status. Inter-modular connections, and the connector nodes and hubs which mediate them, are most numerous in posterior modules containing regions of association cortex; the fronto-temporal module is sparsely connected to other modules and the medial occipital module also has relatively few connector nodes.

METHODOLOGICAL ISSUES

This work is a first attempt to uncover the hierarchical organization of brain functional networks and to compare the stability of hierarchical modular decompositions across individuals. There are, however, three possible weaknesses in our analysis that we would like to address in this section.

Validation of the algorithm

A first consideration concerns the choice of the Louvain method (LM) in order to uncover nested modules in the brain networks. LM was first proposed in order to uncover optimal partitions of a graph by maximising modularity. This is a greedy method which is known to be very fast and very precise (Blondel et al., 2008), albeit less precise than much slower methods such as simulated annealing (SA). It is interesting to note, however, that this lack of precision may be an advantage, in practice, as it may avoid some of the pitfalls of modularity analysis such as its resolution limit (Fortunato and Barthélemy, 2007). For instance, it has been recently shown that LM performs much better than SA when applied to benchmark networks with unbalanced modules comprising different numbers of nodes (Lancichinetti and Fortunato, 2009). We therefore believe that there is good evidence that the top level partitions uncovered by LM are valid. The validity of the intermediate hierarchical levels identified by the algorithm is, however, more arguable, as it has not been studied in detail yet. In order to show the validity of these intermediate levels, we need to verify that the method uncovers all the significant partitions present in the network and only those.

To do so, we have tested LM on a benchmark network with known hierarchical structure (Sales-Pardo et al., 2007); **Figure 5A**). This benchmark network is made of 640 nodes with three levels of organization: small modules comprising 10 nodes, medium-size modules comprising 40 nodes and large modules comprising 160 nodes. The cohesiveness of the hierarchy between levels is tuned by a single parameter σ, i.e. the larger the value of σ, the more difficult it is to find the sub-modules. When applied on this benchmark network, the algorithm finds with an excellent precision the first two levels (16 modules and 64 modules), but does not uncover the partition into 4 modules. This result is to be expected because this partition into four modules is sub-optimal in terms of modularity and can therefore not be uncovered by an aggregative method. This shows that the method can at best uncover the partition optimising modularity and finer partitions. In order to uncover coarser partitions, one needs to decrease the resolution of the method, which can be done by following Reichardt and Bornholdt (2006), or Sales-Pardo et al. (2007), for instance.

On the same benchmark network, the algorithm typically finds two levels (one corresponding to 64 modules and one corresponding to 16 modules) but it may occasionally find three levels (one level corresponding to 64 modules and two levels similar to the partition into 16 modules). When $\sigma = 1.0$, for instance, over 100 realizations

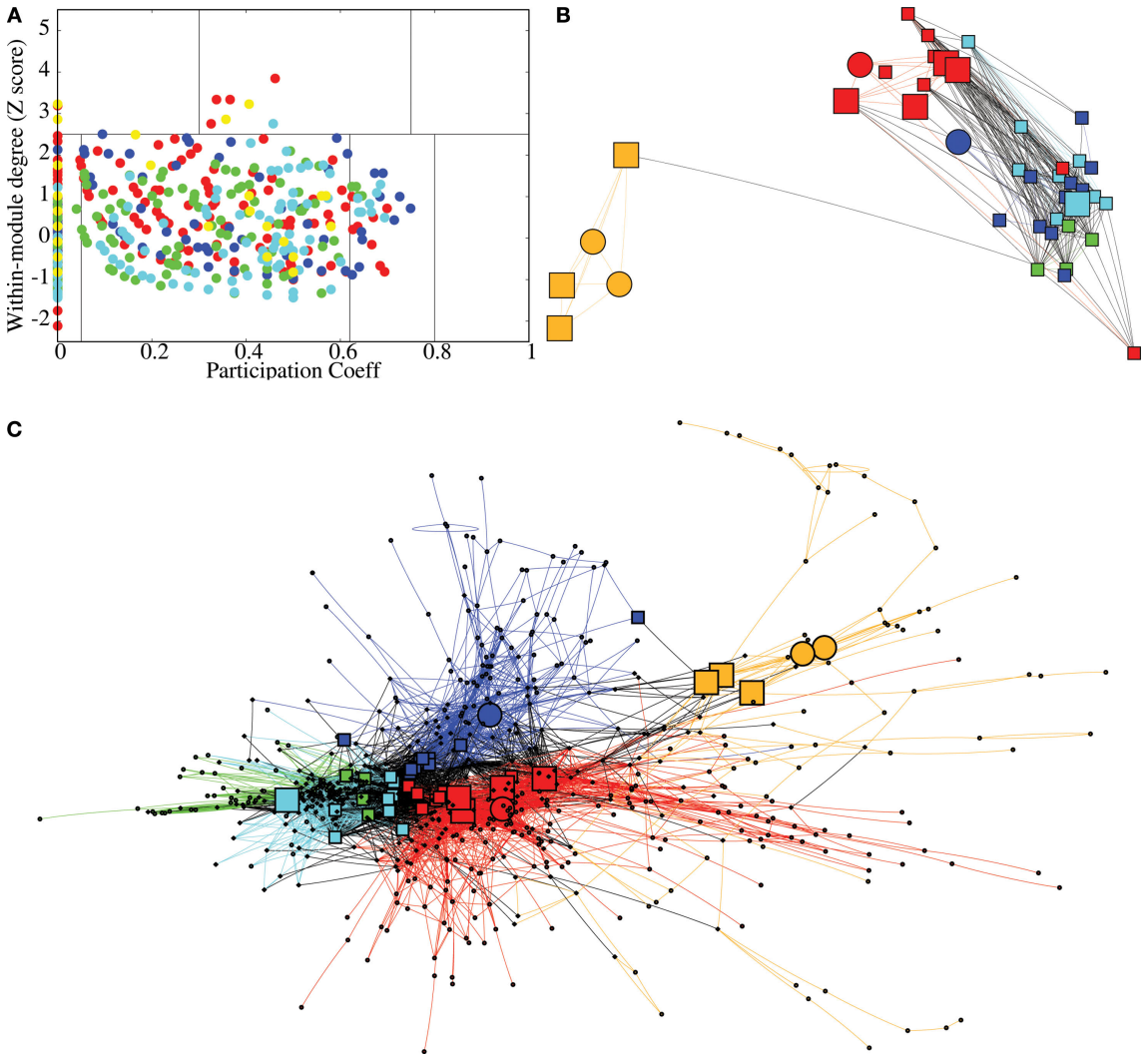


FIGURE 4 | Topological roles of network nodes in intra- and inter-modular connectivity. (A) All nodes are plotted in the {$P - z$} plane of intra-modular degree z vs participation coefficient P; the solid lines partition the plane according to criteria for hubs vs non-hubs and connector, provincial, peripheral or kinless nodes. **(B)** Anatomical representation of the provincial hubs (circles), connector hubs (large squares) and connector nodes (small squares) of each of each of the five largest modules at the highest level of the modular hierarchy. **(C)** Topological representation of the network in using Fruchterman–Reingold algorithm (Fruchterman and Reingold, 1991) to highlight topological proximity of highly connected nodes; colour and shape of the nodes represent their modular assignment and topological role as above and in **Figure 2**.

of the graph, the algorithm finds two levels on 86 realizations, and three levels on 14 realizations. This result is encouraging as it suggests that the algorithm only produces significant partitions. However, it is possible to find situations where it is not the case, e.g. random graphs. It is therefore still necessary to verify the significance of intermediate partitions, as we will discuss below.

Comparison with a random graph

A second consideration concerns the comparison of the partition of the original network with randomized data, as the algorithm also gives a hierarchical decomposition for randomly generated networks. To show that the representative brain network under study (subject ID 2) displays a non-random hierarchical modular structure, we have randomized the original data and processed the hierarchical structure of randomized networks, with two kind of randomization. First, by computing 100 randomizations of the time points in the original time-series (in green on **Figure 5B**) and, second, by randomising the original adjacency matrix 100 times (in blue on **Figure 5B**). Note that the two kinds of randomization lead to networks with different sizes: in the randomized time-series networks, almost all the nodes are connected, thus leading to networks with 1808 nodes and 8000 edges. Whereas starting from the original adjacency matrix leads to networks of 844 nodes and 8000 edges. The modularity obtained for the lowest and highest partitions of the original network are displayed in **Figure 5B**. The modularity values are clearly reduced in the randomized networks, relative to the original data, indicating that our results on real brain networks are not trivially reproduced in random networks.

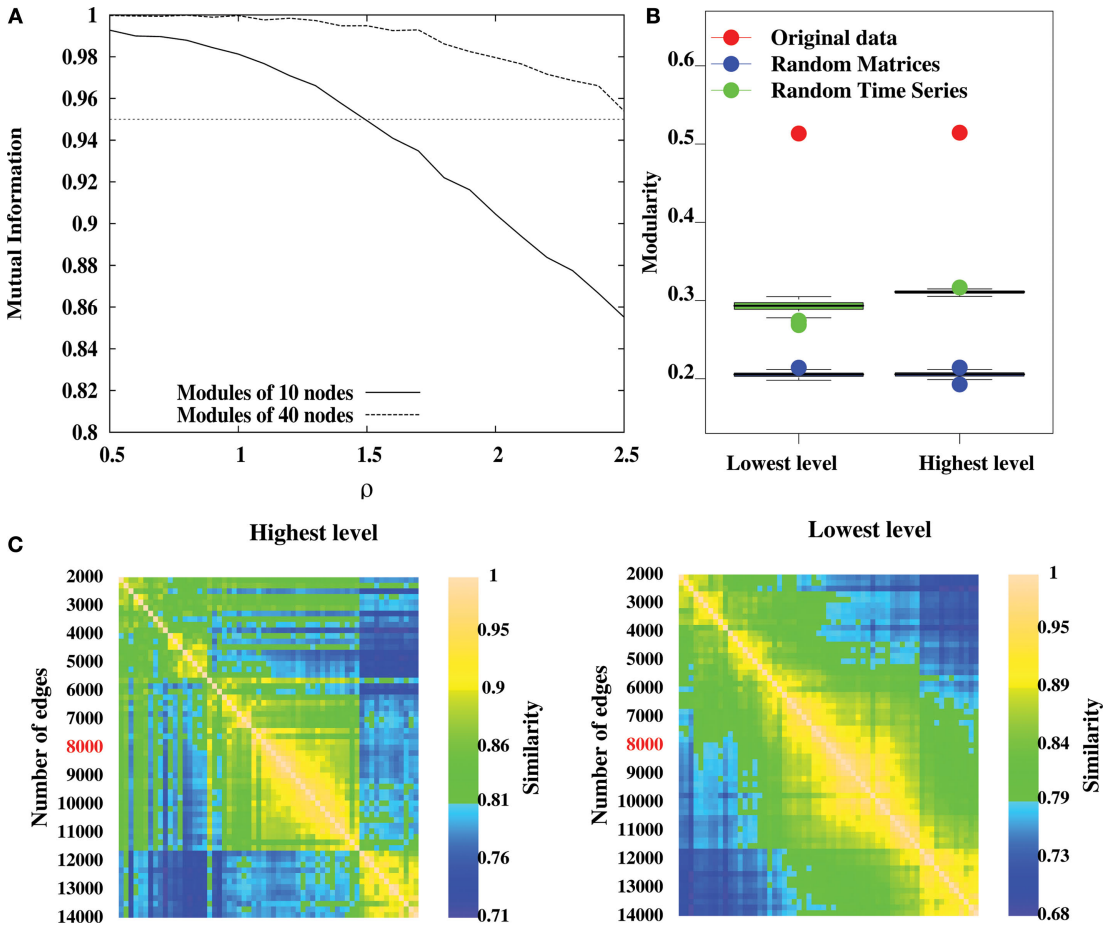


FIGURE 5 | Methodological issues in analysis of hierarchical modularity. **(A)** Validation of the Louvain method for hierarchical decomposition on a benchmark network defined in Sales-Pardo et al. (2007). The network is naturally made of 64, 16 and 4 modules of 10, 40 and 160 nodes respectively. The separability of different levels of the benchmark network is controlled by the parameter ρ. We calculate the normalized information between the lowest non-trivial level partition and the natural partition of 64 modules (solid curve), and between the second level partition and the natural partition of 16 modules (dashed curve). After averaging over 20 different realizations of the network, our simulations show an excellent agreement as mutual information is above 0.95 for values of ρ up to 1.5 for the lowest non-trivial and intermediate levels. **(B)** Modularity values at the highest and lowest levels of hierarchical community structure in a representative brain network (Subject ID 2, in red) and for networks obtained from 100 randomizations of the original time-series (in green), and for networks obtained by 100 randomizations of the original adjacency matrix. **(C)** Similarity measures between highest level partitions (left) and non-trivial lowest level partitions (right) obtained by thresholding the original network to retain different number of highest correlations as edges.

In order to show that the intermediate levels considered in this paper are significant, we have followed the argument that significant partitions should be robust, in the sense that they should only be weakly altered by a modification of the optimization algorithm. As argued by Ronhovde and Nussinov (2009), comparing the optimal partitions found by the algorithm for different orders of the nodes is a way to test their robustness and therefore their validity. We have therefore optimized the modularity of the representative brain network 100 times by choosing the nodes in a different order, and focused on the first non-trivial partition found by the algorithm. The mutual information between pairs of partitions obtained for each different order is then computed. The average mutual information among those pairs is very high (0.89) compared to what is obtained for a comparable random network (0.44), thereby suggesting that partitions obtained at the lowest non-trivial levels are relevant for the network under study.

Dependence on the number of edges

A third consideration concerns the number m of edges that we have chosen in order to map the correlation matrices onto unweighted graphs. This is a known problem when dealing with fMRI data and building brain networks. If m is too small, i.e. keeping the top most significant links, the network will be so sparsely connected that it will be made of several disconnected clusters. If m is too large, in contrast, the network will be very densely connected, but mainly made of unsignificant links. In these two extremes, the network structure is a bad representation of the correlation matrix. This is still an open problem that requires the right trade-off between these two competing factors. In order to show the robustness of our results, we propose to look at the resilience of the hierarchical modular organization under the tuning of the value of m. Meaningful values of m are identified by intervals over which the structure of the network is preserved. We have applied this scheme

to the optimal partitions of the most representative subject (Subject 2), over a wide range of threshold (2000–14000 edges, with a step of 200 edges). Our results show that partitions are very similar (in terms of mutual information) over the range (6000–11000) for both highest level (left on **Figure 5C**) and non-trivial lowest level (right on **Figure 5C**), indicating our results are robust to the specific choice of threshold.

DISCUSSION

In this study, we have applied recently developed tools for characterizing the hierarchical, modular structure of complex systems to functional brain networks generated from human fMRI data recorded under no-task or resting state conditions. Where previous comparable work was limited by the computational expense of available modularity algorithms, meaning that only one or a few relatively low resolution networks (comprising 10 s of nodes) could be analysed, here we were able to obtain modular decompositions on a larger number of higher resolution networks (each comprising 1000s of nodes). In addition, we used an information-based measure to quantify the similarity of community structure between two different networks and so to find a principled way of focusing attention on a single network that is representative of the group.

HIERARCHICAL MODULARITY

There was clear evidence for hierarchical modularity in these data and the community structure of the networks at all levels of the hierarchy was reasonably similar across subjects ($I \sim 0.6$), suggesting that brain functional modularity is likely to be a replicable phenomenon. This position is further supported by the qualitative similarity between the major modules identified at the highest level of the hierarchy in this study and the major modules or functional clusters identified in comparable prior studies on independent samples (Salvador et al., 2005; Meunier et al., 2009). As previously, the major functional modules comprised functionally and/or anatomically related regions of cortex and this pattern was also evident to some extent at sub-modular levels of analysis. For example, the central module comprising areas of somatosensorimotor and premotor cortex was segregated at a sub-modular level into a medial component, comprising supplementary motor area and cingulate motor area, and a lateral component, comprising precentral and postcentral areas of primary motor and somatosensory cortex.

Another plausible aspect of the results was the clear evidence for a symmetrical posterior-to-anterior progression of cortical modules. This was seen most clearly on the medial surfaces of the cerebral hemispheres in terms of their division into medial occipital, parieto-frontal and central modules. A posterior-to-anterior organization of cortical modules in adult brain functional networks is arguably compatible with the abundant evidence from neurodevelopmental studies which have shown rostro-caudal modularity of the spinal cord, brain stem, hind brain and diencephalon defined by segmented patterns of gene expression (Redies and Puelles, 2001). This speculative link between the topological modularity of adult brain networks and the embryonic modularity of the developing nervous system presents an interesting focus for future studies.

NODE ROLES IN INTER-MODULAR CONNECTIVITY

One important potential benefit of a modular analysis of complex networks is that it allows us to be more precise about the topological role of any particular node in the network. For example, rather than simply saying that a particular region has a high degree we may be able to say that it has a disproportionately important role in transfer of information between modules, rather than within a module. In these data, the location of connector nodes and hubs with a prominent role in inter-modular communication was concentrated in posterior areas of association cortex. The fronto-temporal module, on the other hand, was rather sparsely connected to other modules. One possible explanation for these anatomical differences in inter-modular communication may relate to the stationarity of functional connectivity between brain regions. Our measure of association between brain regions (the wavelet correlation corresponding to a frequency interval of 0.03–0.06 Hz) provides an estimate of functional connectivity "on average" over the entire period of observation (8 min 35 s). If there is significant variability over time in the strength of functional connections between modules this may be manifest in terms of reduced connectivity on average over a prolonged period. Thus one possible explanation for the sparser inter-modular connections of the fronto-temporal module is that the interactions of this system with the rest of the brain network may be more non-stationary or labile over time. This interpretation could be tested by future studies using time-varying measures of functional connectivity, such as phase synchronization (Kitzbichler et al., 2009).

DEALING WITH MORE THAN ONE SUBJECT

One of the challenges in analysis of network community structure is the richness of the results (every node will have a modular assignment and a topological role) and the difficulties attendant on properly managing inter-individual variability in such novel metrics. In previous work, we estimated a functional connectivity matrix for each subject, then thresholded the group mean association, and explored the community structure of the group mean network. This allows us to focus attention on a single network but it neglects between-subject variability and, like any use of the mean in small samples, it is potentially biased by one or more outlying values for the functional connectivity. Here we have explored an alternative approach, using an information-based measure of similarity to quantify between subject differences in network organization and to identify the most representative subject in the sample. One can imagine that this measure could be combined with resampling based approaches to statistical inference in order to estimate, for example, the probability that the community structure identified in a single patient is significantly dissimilar to a reference group of brain networks in normal volunteers. However, it fair to say that there are a number of technical challenges to be addressed in using modularity measures for statistical comparisons between different groups.

RETURNING TO SIMON'S HYPOTHESIS

As this is the first study to attempt a hierarchical modular decomposition of human brain functional networks, there is little guidance in the existing literature as to what the correct structure of the network should resemble. Our results are encouraging in that they have

been able to identify well defined neuroanatomical systems, but they remain empirical and require further validation in appropriate animal models. However, our analysis of simulated data (Section "Discussion") indicates that our algorithm does indeed identify the correct structure of a hierarchical, modular network, which lends confidence to our results.

In Simon's theoretical analysis, near-decomposability was considered to be a ubiquitous property of complex systems because it conferred advantages of adaptive speed in response to evolutionary selection pressures as well as shorter-term developmental or environmental contingencies. In relation to the modularity of human brain systems, this view prompts a number of questions. Perhaps the most immediately addressable, at least by functional neuroimaging, is the question of how the modularity of brain network organization relates to cognitive performance and the capacity to shift attention rapidly between different stimuli or tasks. According to Simon's theory, this key aspect of the brain's cognitive function should depend critically on modular or sub-modular components and the rapid reconfiguration of inter-modular connections between them. Future studies, applying graph theoretical techniques to modularity analysis of fMRI data recorded during task performance (rather than in no-task state) may be important in testing this prediction.

CONCLUSION

We have described graph theoretical tools for analysis of hierarchical modularity in human brain functional networks derived from fMRI. Our main claims are that these techniques are computationally feasible and generate plausible and reasonably consistent descriptions of the brain functional network community structure in a group of normal volunteers. The potential importance theoretically of this analysis has been highlighted by reference to Simon's seminal theory of hierarchy and decomposability in the design of information processing systems.

ACKNOWLEDGMENTS

This research was supported by a Human Brain Project grant from the National Institute of Mental Health and the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD, USA. Renaud Lambiotte was supported by UK EPSRC. Alex Fornito was supported by a National Health and Medical Research Council CJ Martin Fellowship (ID: 454797). Karen D. Ersche is a recipient of the Betty Behrens Research Fellowship at Clare Hall, University of Cambridge. The experiment was sponsored by GlaxoSmithKline and conducted at the GSK Clinical Unit Cambridge.

REFERENCES

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3, e17. doi:10.1371/journal.pcbi.0030017.

Achard, S., Salvador, R., Witcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72.

Amaral, L., and Ottino, J. (2004). Complex networks: augmenting the framework for the study of complex systems. *Eur. Phys. J. B* 38, 147–162.

Arenas, A., Fernandez, A., and Gomez, S. (2008). Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* 10, 053039.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory E* 10, P10008.

Brandes, U., Delling, D., Gaertler, M., Goerke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. (2006). Maximizing modularity is hard. *arXiv:physics*, 0608255.

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

Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., and Evans, A. C. (2008). Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb. Cortex* 18, 2374–2381.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., Quigley, M. A., and Meyerand, M. E. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am. J. Neuroradiol.* 22, 1326–1333.

Danon, L., Duch, J., Diaz-Guilera, A., and Arenas, A. (2005). Comparing community structure identification. *J. Stat. Mech. Theory E* 9, P09008.

Ferrarini, L., Veer, I. M., Baerends, E., van Tol, M.-J., Renken, R. J., Van Der Wee, N. J. A., Veltman, D. J., Aleman, A., Zitman, F. G., Penninx, B. W. J. H., Van Buchem, M. A., Reiber, J. H. C., Rombouts, S. A. R. B., and Milles, J. (2009). Hierarchical functional modularity in the resting-state human brain. *Hum. Brain Mapp.* 30, 2220–2231.

Fortunato, S., and Barthélemy, M. (2007). Resolution limit in community detection. *Proc. Natl. Acad. Sci. U.S.A.* 104, 36–41.

Fruchterman, T., and Reingold, E. (1991). Graph drawing by force-directed placement. *Software* 21, 1129–1164.

Guimerà, R., and Amaral, L. A. N. (2005a). Cartography of complex networks: modules and universal roles. *J. Stat. Mech. Theory E* 2, P02001.

Guimerà, R., and Amaral, L. A. N. (2005b). Functional cartography of complex metabolic networks. *Nature* 433, 895–900.

Guimerà, R., Mossa, S., Turtschi, A., and Amaral, L. A. N. (2005). The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7794–7799.

Guimerà, R., Sales-Pardo, M., and Amaral, L. A. N. (2004). Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* 70, 025101.

Hilgetag, C.-C., Burns, G. A., O'Neill, M. A., and Scannell, J. W. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 91–110.

Kitzbichler, M., Smith, M., Christensen, S., and Bullmore, E. (2009). Broadband criticality of human brain network synchronization. *PLoS Comput. Biol.* 5, e1000314. doi:10.1371/journal.pcbi.1000314.

Kuncheva, L., and Hadjitodorov, S. (2004). Using diversity in cluster ensembles. *IEEE Int. Conf. Syst. Man Cybern.* 2, 1214–1219.

Lambiotte, R., Delvenne, J. C., and Barahona, M. (2009). Laplacian dynamics and multiscale modular structure in networks. arXiv:0812.1770.

Lancichinetti, A., and Fortunato, S. (2009). Community detection algorithms: a comparative analysis. arXiv:0908.1062.

Meunier, D., Achard, S., Morcom, A., and Bullmore, E. (2009). Age-related changes in modular organization of human brain functional networks. *Neuroimage* 44, 715–723.

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

Newman, M. E. J. (2004). Detecting community structure in networks. *Eur. Phys. J. B* 38, 321–330.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8577–8582.

Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69, 026113.

Oldham, M. C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D. H. (2008). Functional organization of the transcriptome in human brain. *Nat. Neurosci.* 11, 1271–1282.

Redies, C., and Puelles, L. (2001). Modularity in vertebrate brain development and evolution. *Bioessays* 23, 1100–1111.

Reichardt, J., and Bornholdt, S. (2006). Statistical mechanics of community detection. *Phys. Rev. E* 74, 016110.

Ronhovde, P., and Nussinov, Z. (2009). Multiresolution community detection for megascale networks by information-based replica correlations. *Phys. Rev. E* 80, 016109.

Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1118–1123.

Sales-Pardo, M., GuimerÃ, R., Moreira, A. A., and Amaral, L. A. N. (2007). Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15224–15229.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., and Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* 15, 1332–1342.

Schwarz, A. J., Gozzi, A., and Bifone, A. (2008). Community structure and modularity in networks of correlated brain activity. *Magn. Reson. Imaging* 26, 914–920.

Simon, H. (1962). The architecture of complexity. *Proc. Am. Philos. Soc.* 106, 467–482.

Strogatz, S. H. (2001). Exploring complex networks. *Nature* 410, 268–276.

Suckling, J., Davis, M. H., Ooi, C., Wink, A. M., Fadili, J., Salvador, R., Welchew, D., Sendur, L., Maxim, V., and Bullmore, E. (2006). Permutation testing of orthogonal factorial effects in a language-processing experiment using fMRI. *Hum. Brain Mapp.* 27, 425–433.

Tzourio-Mazoyer, N., Landeau, N., Papathanassiou, B., Crivello, D., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.

Conflict of Interest Statement: Edward Bullmore is employed half-time by GlaxoSmithKline and half-time by University of Cambridge.

Received: 20 March 2009; paper pending published: 25 June 2009; accepted: 02 October 2009; published online: 30 October 2009.

Citation: Meunier D, Lambiotte R, Fornito A, Ersche KD and Bullmore ET (2009) Hierarchical modularity in human brain functional networks. Front. Neuroinform. ***3****:37. doi: 10.3389/neuro.11.037.2009*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 19 March 2010
doi: 10.3389/neuro.11.001.2010

Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks

Gorka Zamora-López[1]*, Changsong Zhou[2,3] and Jürgen Kurths[4,5]

[1] *Interdisciplinary Center for Dynamics of Complex Systems, University of Potsdam, Potsdam, Germany*
[2] *Department of Physics, Centre for Nonlinear Studies, Hong Kong Baptist University, Hong Kong, China*
[3] *The Beijing-Hong Kong-Singapore Joint Centre for Nonlinear and Complex Systems, Hong Kong Baptist University, Hong Kong, China*
[4] *Transdisciplinary concepts and methods, Potsdam Institute for Climate Impact Research, Potsdam, Germany*
[5] *Institute of Physics, Humboldt University, Berlin, Germany*

Edited by:
Claus C. Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Steven Bressler, Florida Atlantic University, USA
David Meunier, University of Cambridge, UK

****Correspondence:***
Gorka Zamora-López, Interdisciplinary Center for Dynamics of Complex Systems, University of Potsdam, Komplex II – Golm (Haus 28) , Karl-Liebknecht-Str. 24, D-14476 Potsdam, Germany.
e-mail: gorka@agnld.uni-potsdam.de

Sensory stimuli entering the nervous system follow particular paths of processing, typically separated (segregated) from the paths of other modal information. However, sensory perception, awareness and cognition emerge from the combination of information (integration). The corticocortical networks of cats and macaque monkeys display three prominent characteristics: (i) modular organisation (facilitating the segregation), (ii) abundant alternative processing paths and (iii) the presence of highly connected hubs. Here, we study in detail the organisation and potential function of the cortical hubs by graph analysis and information theoretical methods. We find that the cortical hubs form a spatially delocalised, but topologically central module with the capacity to integrate multisensory information in a collaborative manner. With this, we resolve the underlying anatomical substrate that supports the simultaneous capacity of the cortex to segregate and to integrate multisensory information.

Keywords: corticocortical networks, cortical hubs, multisensory integration, segregation, integration

INTRODUCTION

The mammalian nervous system is responsible for collecting and processing of information, and for providing adaptive responses which permit the organism to survive in a permanently changing environment. Sensory neurones encode environmental information into electrical signals which propagate in a "bottom-up" manner through different processing stages (Kandel et al., 2000; Bear et al., 2006). Each level provides responses of increasing complexity and at different time scales, e.g. reflex arcs, emotional responses and more elaborate cognitive responses. Information of the same modality (e.g. visual, auditory, somatosensory, etc.) traverses the body together, typically separated from the processing paths of other modalities. This permits that particular regions of the cortex specialise in detecting different features of the sensory stimuli, e.g. orientation, velocity and colour of the visual input; or frequency and pitch of the auditory stimuli. However, in order to generate a coherent perception of the reality, the brain needs to combine (integrate) this multisensory information at some place (Robertson, 2003) and during some time (Fahle, 1993; Singer and Gray, 1995; Engel and Singer, 2001). For that, the paths of information need to converge.

It has been argued that the functional capacity of the NS to balance between segregation (specialisation) and integration might be facilitated by its structural organisation (Sporns and Tononi, 2001). Analysis of the connectivity between regions of the cerebral cortex in macaque monkeys and cats has revealed the following characteristics: (i) clustered organisation of the cortical areas (Scannell and Young, 1993; Scannell et al., 1995; Hilgetag et al., 2000; Hilgetag and Kaiser, 2004) (see **Figure 2**), (ii) a large density of connections, and (iii) a broad degree distribution containing highly connected areas which are referred as *hubs* (Zemanová et al., 2006; Sporns et al., 2007; Zhou et al., 2007; Hagmann et al., 2008). While the organisation of cortical areas into clusters permits the segregated processing of information of different modality, the large number of connections involves that sensory information is highly accessible to all cortical areas, regardless of its modal origin. A detailed analysis of the corticocortical communication substrate has revealed the central role of the cortical hubs, by facilitating the communication between the different sensory modalities (Zamora-López et al., 2009).

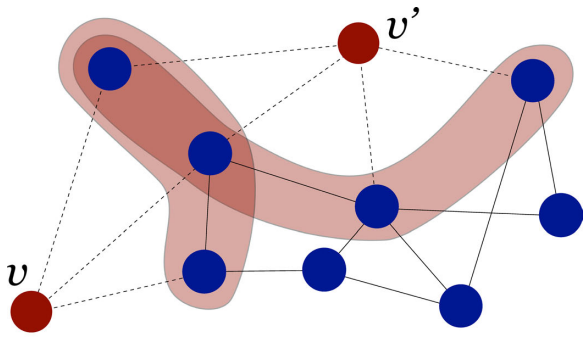
Whether the cortical hubs act as passive transmitters of information, or they perform a more active function is a relevant question that we try to answer in the present paper. We start by summarising principles of complex network analysis and information theory in Section "Materials and Methods". Section "Topological Capacity of Integration" contains a thorough application of graph theoretical measures which reveal that the cortical hubs form an additional module, expressed as a higher hierarchical level. In Section "Functional Capacity of Integration", we challenge the intuitively assigned integrative properties of this central module by means of dynamical and information theoretical measures. On the one hand, we find that only simultaneous lesion of particular hubs leads to a dynamical segregation of the sensory modules (visual, auditory, somatosensory-motor and frontolimbic). On the other hand, the same hubs form a dynamical cluster after simultaneous excitation of primary sensory areas, a clear sign of their integrative capacities.

MATERIALS AND METHODS

GRAPH ANALYSIS

We first introduce basic concepts of graph theory. A network is an abstract manner to represent different aspects of a real system, providing it with a form (topology) which can be mathematically

tractable. A network $\mathcal{G}\ (N, L)$, composed of N nodes interconnected by L links, is described by an adjacency matrix $\mathbf{A}$ with entries $A_{ij} = 1$ when there is a link pointing from node i to node j, and $A_{ij} = 0$ otherwise. The density of $\mathcal{G}$ is the fraction between the number of links L and the total number of links possible: $\rho = \frac{L}{N(N-1)}$. In order to characterise the topological scales of networks, there exist many statistical descriptors, all measurable from the information encoded in the adjacency matrix. The output degree $k_o(i) = \sum_{j=1}^{N} A_{ij}$ of a node i is the number of *efferent* connections that it projects to other nodes, and its input degree $k_i(i) = \sum_{j=1}^{N} A_{ji}$, is the number of the *afferent* connections it receives. The degree distribution $p(k)$ is the probability that a randomly chosen node has degree k. One of the key discoveries that triggered a renewed interest in graph theory is that the distribution $p(k)$ of many empirical networks approximately follows a power-law $p(k) \sim k^{-\gamma}$ (Newman, 2003), where γ is the degree exponent. In such scale-free(-like) networks the majority of nodes possess a small number of neighbours, and few nodes (the hubs) are highly connected.



$$MI(v, v') = \frac{2}{5} = 0.4$$

FIGURE 1 | Schematic representation of the normalised matching index, computed as in Eq. 2. For proper comparison between pairs, the measure is normalised by the number of different neighbours of v and v'.

Distance and centrality

The distance d_{ij} between two nodes i and j is the length of the shortest path between them, i.e. the minimal number of links crossed to travel from i to j. If there is a link $i \to j$, then $d_{ij} = 1$. If there is no other choice than going through an intermediate node k such that $i \to k \to j$, then $d_{ij} = 2$, and so on. When there exists no path from i to j then $d_{ij} = \infty$. The average pathlength l is the average distance between all pairs of nodes. The shortest path between two nodes is usually not unique and there are several alternative shortest paths. In order to characterise the influence of individual nodes on the flow and the spread of information through a network, the betweenness centrality $C_B(i)$, is defined as the fraction of all shortest paths passing through i (Anthonisse, 1971; Freeman, 1977):

$$C_B(i) = \sum_{s \neq i, t \neq i}^{N} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{1}$$

where $\sigma_{st}(i)$ is the number of shortest paths starting in s, running through i and finishing in t, and σ_{st} is the number of all shortest paths from s to t.

Matching index

The topological similarity of two nodes can be characterised as the number of common neighbours they share. In the extreme case, two nodes are topologically identical if both have the same set of connections. The neighbourhood of node i is defined as the set of nodes it connects with, $\Gamma(i) = \{j : A_{ij} = 1\}$. In graphs without multiple links the size of the neighbourhood $|\Gamma(i)|$ equals the degree of i. The matching index of two nodes i and j is thus the overlap of their neighbourhoods: $MI(i,j) = |\Gamma(i) \cap \Gamma(j)|$. Defined in this manner $MI(i,j)$ depends on the degrees of i and j, and the values for different pairs are not comparable. Imagine two nodes with degrees $k(i) = k(j) = 3$ which are connected to the same neighbours. As $\Gamma(i) = \Gamma(j)$ their matching is $MI(i,j) = 3$. Imagine other two nodes with degrees $k(i') = 3$ and $k(j') = 4$. Maximally, they could share three neighbours and have $MI(i',j') = 3$ as well, despite i and j are topologically equivalent but i' and j' are not. In order to compare the values for different pairs the measure can be normalised by the number of distinct neighbours of the two nodes, i.e. the union of the two neighbourhoods $|\Gamma(i) \cup \Gamma(j)|$ as illustrated in **Figure 1**. The normalised matching index can be computed as:

$$MI(i,j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} = \frac{\sum_{n,m=1}^{N} A_{in} A_{jm}}{k(i) + k(j) - \sum_{n,m=1}^{N} A_{in} A_{jm}} \tag{2}$$

Now, $MI(i,j) = 1$ only if i and j are connected exactly to the same nodes, $\Gamma(i) = \Gamma(j)$, and $MI(i,j) = 0$ if they have no common neighbours.

Reference surrogate networks

Graph theoretical measures help understand the topological organisation of networks. Equally relevant is to uncover the features which are characteristic to the underlying system and the fundamental properties of its development. In this sense, the question is not whether a graph measure takes a specific numerical value, but whether this value distinguishes the empirical network $\mathcal{G}_{emp}$ from others of similar characteristics. For that, the formulation of appropriate null-models is required. A typical such null case is to generate surrogate networks with the same size N, number of links L and degree distribution $p(k)$ as in $\mathcal{G}_{emp}$. The *link switching method* (Katz and Powell, 1957; Holland and Leinhardt, 1977; Rao et al., 1996; Kannan et al., 1999; Roberts, 2000) consists of the following iterative process: starting from $\mathcal{G}_{emp}$, at each iteration two links are chosen at random $(i_1 \to j_1)$ and $(i_2 \to j_2)$. The links are rewired as $(i_1 \to j_2)$ and $(i_2 \to j_1)$ provided that the new links do not already exist and do not introduce self-loops, i.e. $i \to i$. Repeating the process sufficient times the resulting surrogate network conserves the initial degree distribution but any higher order structure is destroyed.

DATA

The classical textbook illustration of the cerebral cortex as a surface (grey matter) which can be subdivided into functional or cytoarchitectonic regions is only a limited picture. Additionally, long-range fibres link the cortical areas via the white matter forming

a complex network which is neither regular nor completely random. This intricated structure enhances the richness and complexity of information processing capabilities of the cerebral cortex. In this paper we focus on the analysis of the cortical connectivity of the cat because it is, up to date, the most complete and reliable dataset of this kind.

Corticocortical connectivity of the cat

The dataset of the corticocortical connections within the cortex of cats was created after an extensive collation of literature reporting anatomical tract-tracing experiments (Scannell and Young, 1993; Scannell et al., 1995). It consists of a parcellation into 53 cortical areas and 826 fibres of axons between them as summarised in

Figure 2. The connections are weighted according to the axonal density of the projections. After application of data mining methods (Scannell and Young, 1993; Hilgetag and Kaiser, 2004), the network was found to be organised into four distinguishable clusters which closely follow functional subdivisions: visual (V), auditory (A), somatosensory-motor (SM) and frontolimbic (FL).

Surrogate data

In order to perform significance tests of the graph measures, an ensemble of 1000 surrogate networks has been created following the *link switching method* (see Section "Graph Analysis"). All the resulting networks have the same size $N = 53$, the same number of links $L = 826$ and the same degree distribution as the corticocortical

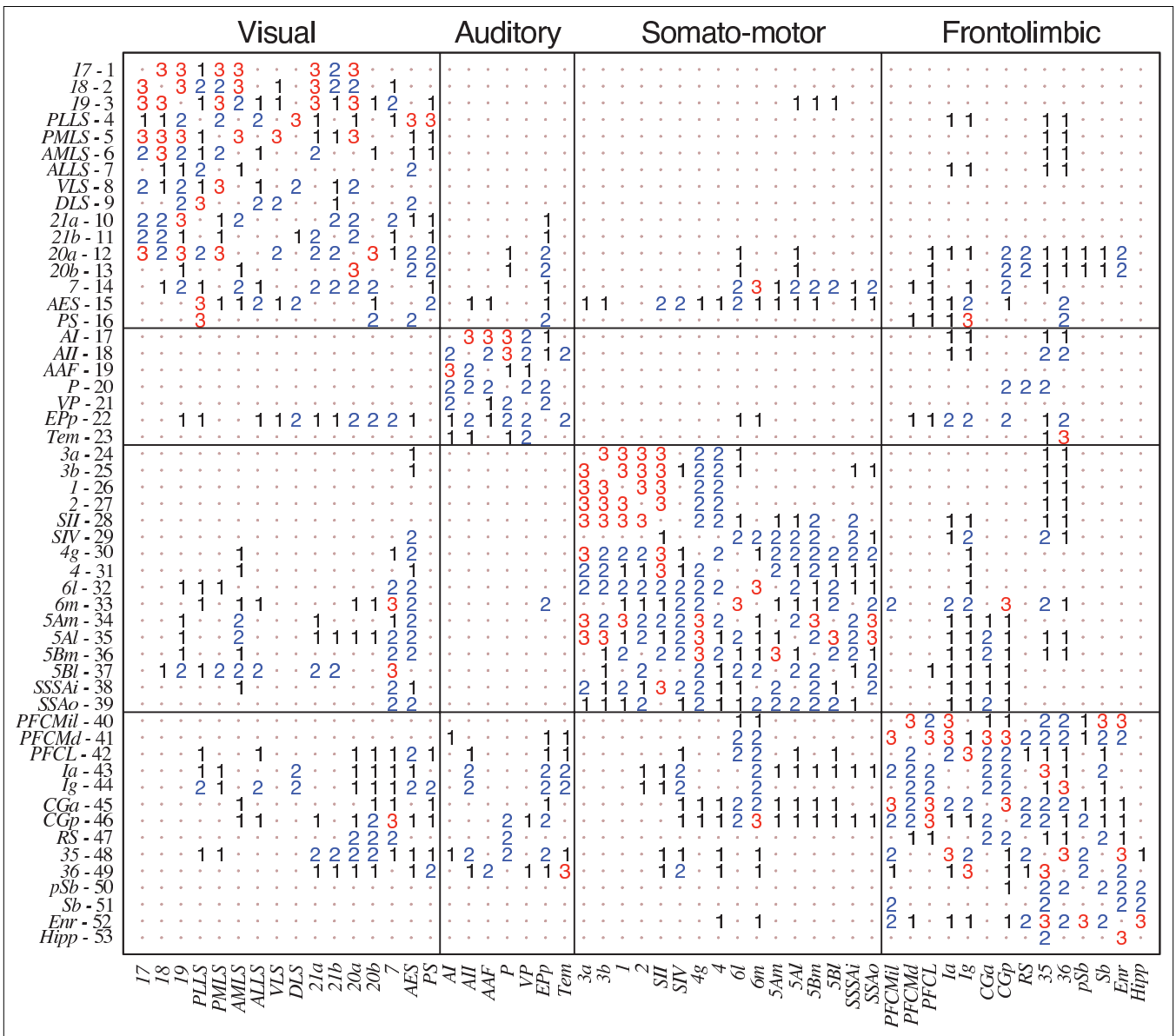


FIGURE 2 | Weighted adjacency matrix W of the corticocortical connectivity of the cat comprising of $L = 826$ directed connections between $N = 53$ cortical areas (Scannell and Young, 1993; Scannell et al., 1995). For visualisation purposes, the non-existing connections (0) have been replaced by dots. The network has clustered organisation, reflecting four functional subdivisions: visual (V), auditory (A), somatosensory-motor (SM) and frontolimbic (FL).

network of the cat. To assure that any further internal structure is destroyed, each surrogate network is the product of $10 \times L$ iterations. In the following, this set will be referred as the rewired ensemble $\{\mathcal{G}_{1n}\}$, and the original corticocortical network of the cat as $\mathcal{G}_{cat}$. The ensemble average of graph measures applied on the surrogate set $\{\mathcal{G}_{1n}\}$ will be considered as the expected values.

INFORMATION THEORY AND INTEGRATION

Information theory has been very successful to describe transmission of information, encoding and channel capacity. At the root of this success lies the original idea of Shannon to apply concepts of statistical physics to represent the nature of communication. Consider a system A with M possible states. That is, a measurement made on A yields the values $a_1, a_2, \ldots, a_M$, with a probability $p(a_i)$. The average amount of information gained from a measurement that specifies one particular value a_i is given by the entropy of the system (Shannon, 1948; Cover and Thomas, 1991):

$$H(A) = -\sum_{i=1}^{M} p(a_i) \log p(a_i). \tag{3}$$

The entropy can be interpreted as the *amount of surprise one should feel upon reading the result of a measurement* (Faser and Swinney, 1986). It vanishes when the system has only one accessible state because the value a is always obtained, i.e. there is no surprise. $H(A)$ is maximum when all the states are equally likely, i.e. there are no preferred states.

The *statistical dependence* between two systems x_1 and x_2 is quantified by their mutual information:

$$MI(x_1, x_2) = H(x_1) + H(x_2) - H(x_1, x_2). \tag{4}$$

By definition, the joint entropy is $H(x_1,x_2) \leq H(x_1) + H(x_2)$. The equality is only fulfilled if x_1 and x_2 are statistically *independent*, hence $MI(x_1,x_2) = 0$, and otherwise $MI(x_1,x_2) > 0$.

Integration

In a series of papers Tononi and Sporns proposed a particular measure of integration (Tononi and Sporns, 1994; Tononi et al., 1996, 1998). Given a system X composed of N subsystems x_1, integration is defined as:

$$I(X) = \sum_{i=1}^{N} H(x_i) - H(X) \tag{5}$$

where $H(x_i)$ is the entropy of one subsystem and $H(X) = H(x_1, x_2, \ldots, x_n)$ is the joint entropy of the system considered as a whole. $I(X) = 0$ only if all $x_i \in X$ are statistically independent of each other, and positive otherwise. After this definition, integration is the extension of mutual information for more than two systems. In other words, $I(X)$ measures the internal level of statistical dependence among all the subsystems $x_i \in X$.

Linear dynamical systems

The *steady-state* of a linear system whose N subsystems $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are driven by a Gaussian noise $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$, is described by $x_i = g \sum_j A^t_{ij} x_j + \xi_i$, where g is the coupling strength and A^t is the transpose of the adjacency matrix. Otherwise the dynamics of x_i would be characterised by its own outputs, not by the inputs it receives. Written in matrix form:

$$\mathbf{x} = g\mathbf{A}^{\mathrm{t}}\mathbf{x} + \boldsymbol{\xi}. \tag{6}$$

In practical terms the variable x_i might be interpreted as the *activity level* of the cortical area i (Kötter and Sommer, 2000; Young et al., 2000), or as the *mean firing rate* of the neurones in the area i (Graben et al., 2007). The entropy of such a multivariate Gaussian system can be analytically calculated out of its covariance matrix such that $H(X) = \frac{1}{2}\log\left[(2\pi e)^N |COV(X)|\right]$, where $|\cdot|$ stands for the determinant (Papoulis, 1991; Tononi and Sporns, 1994). The entropy of an individual Gaussian process is $H(x_i) = \frac{1}{2}\log(2\pi e \nu_i)$, where ν_i is the variance of x_i, say, the i^{th} diagonal element of the $COV(X)$ matrix. Replacing $H(X)$ and $H(x_i)$ into Eq. 5 and applying basic algebra, we reduce the integration of such a multivariate Gaussian system as:

$$I(X) = \frac{1}{2}\log\left[\frac{\prod_{i=1}^{N}\nu_i}{|COV(X)|}\right]. \tag{7}$$

This expression shows that $I(X)$ of the linear system is properly normalised and is independent of system size N. The covariance matrix can be analytically computed by solving the system such that $\mathrm{x} = \frac{1}{1-g\mathrm{A}^{\mathrm{t}}} \cdot \xi = Q \cdot \xi$, and averaging over the states produced by successive values of ξ one finds: $COV(X) = \langle \mathrm{x} \cdot \mathrm{x}^t \rangle = \langle (Q \cdot \xi) \cdot (\xi^t \cdot Q^t) \rangle = Q \cdot Q^t$.

Comparing different systems

To compare $I(X)$ of different systems, the matrix A^t needs to be adequately normalised because application of the same coupling strength g to different networks might set them into different dynamical states. Hence, they might not be comparable. The linear System (6) has several poles depending on g. The smallest pole corresponds to $g_1 = \frac{1}{\lambda_{\max}}$ where $\lambda_{\max}$ is the largest eigenvalue of the transposed adjacency matrix $\mathbf{A^t}$. The solutions only have physical meaning for $g < g_1$, otherwise the stationarity condition does not hold. In **Figure 3** the poles corresponding to the corticocortical network of the cat are shown. Notice that at the poles, both entropy and integration diverge. To make the comparison of the dynamics of different networks possible, we normalise the adjacency matrices as $\hat{\mathrm{A}} = g_1 \mathrm{A}^{\mathrm{t}} = \frac{\mathrm{A}^{\mathrm{t}}}{\lambda_{\max}}$. In this manner, all systems have the smallest pole at $g = 1$.

Finally, a proper coupling strength g needs to be chosen. For that, we have estimated the covariance matrices of the cat cortical network under different coupling strengths (**Figure 4**). They are similar to the correlation patterns arising from more complex models (Zemanová et al., 2006; Honey et al., 2007; Zhou et al., 2006, 2007). This similarity indicates the validity of the simple linear System (6) for the exploratory purposes here intended. All networks considered in Section "Functional Capacity of Integration" are normalised by their first pole and a coupling strength of $g = 0.5$ is applied. Unless otherwise stated, the noise level is set to $\xi_i = 1.0$.

Subsets of elements

The entropy of a subset of systems $S \subseteq X$ can be obtained by first computing $COV(X)$ as indicated above, and then extracting the covariance submatrix $COV(S)$ out of $COV(X)$ by considering only the elements $x_i \in S$. The entropy of the subset is then $H(S) = \frac{1}{2}\log\left[(2\pi e)^{N_S}\,|COV(S)|\right]$, and its integration $I(S)$ is:

$$I(S) = \sum_{x_j \in S} H\left(x_j\right) - H(S) = \frac{1}{2}\log\left[\frac{\prod_{j=1}^{N_S} \nu_j}{|COV(S)|}\right]. \quad (8)$$

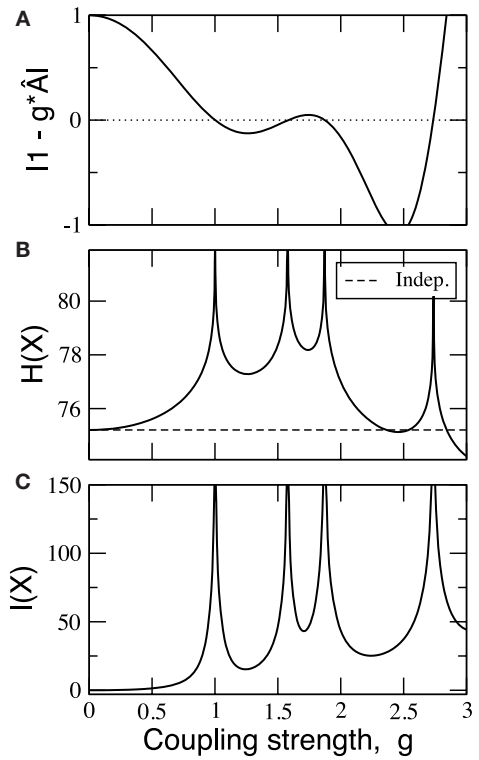


FIGURE 3 | Parametric study of the linear System (6) using the cortical network of the cat. (A) Integrability range. When the determinant $|1 - g\hat{A}| = 0$ the system has a pole. Negative values lead to non-physical solutions. **(B)** Entropy and **(C)** Integration diverge around the poles.

RESULTS

TOPOLOGICAL CAPACITY OF INTEGRATION

In order to characterise the connectional organisation of the nervous system and to understand its functional implications, the complex network approach has been applied in the recent years, particularly at the level of the cerebral cortex. This analysis has revealed several organisation properties, e.g. the clustering of cortical areas according to their sensory modality (visual, auditory, somatosensory-motor and frontolimbic). Recently, it has been reported that communication paths between cortical areas in different sensory modules are not random, but mediated by the hubs of the network (Zamora-López et al., 2009). In this section we present a more detailed graph analysis aiming to characterise the potential function of the cortical hubs.

Inter-modal communication

The betweenness centrality $C_B(\nu)$ quantifies the relevance of a node ν within the communication paths in a network. As represented in **Figure 5A**, we observe that within each of the sensory systems, few cortical areas possess a large betweenness. With $C_B(\nu) > 500$ we find: visual areas 20a, 7 and AES; auditory area EPp; somatosensory-motor areas 6m and 5Al; and frontolimbic areas Ia, Ig, CGp, 35 and 36. On the contrary, only the visual primary cortex (area 17) and the hippocampus have $C_B(\nu) = 0$. In general, we observe that cortical regions known to perform highly specialised sensory function have few connections and very low centrality, e.g. primary and secondary visual or auditory areas, and early somatosensory-motor areas. These areas typically contain ordered mappings of the sensory stimuli such as retinotopic or tonotopic maps, see Appendix of Scannell et al. (1995).

The centrality of a node usually correlates with its degree, hence, it is trivial to find out that precisely the hubs have larger centrality. Drawing any further conclusion requires performing a proper significance test. For comparison, the average $C_B(\nu)$ of the nodes in all the 1000 rewired networks of the surrogate ensemble $\{\mathcal{G}_{1n}\}$ has been computed. The ascending line in **Figure 5B** shows the expected dependence of the betweenness centrality on the degree of the nodes. As a node receives $k_i(\nu)$ inputs and projects $k_o(\nu)$ outputs, the number of shortest paths passing through ν is linearly proportional to $k_i(\nu)k_o(\nu)$ in the surrogate networks. The most prominent observation is that, while C_B of the low degree areas follow the expected centrality, the centrality

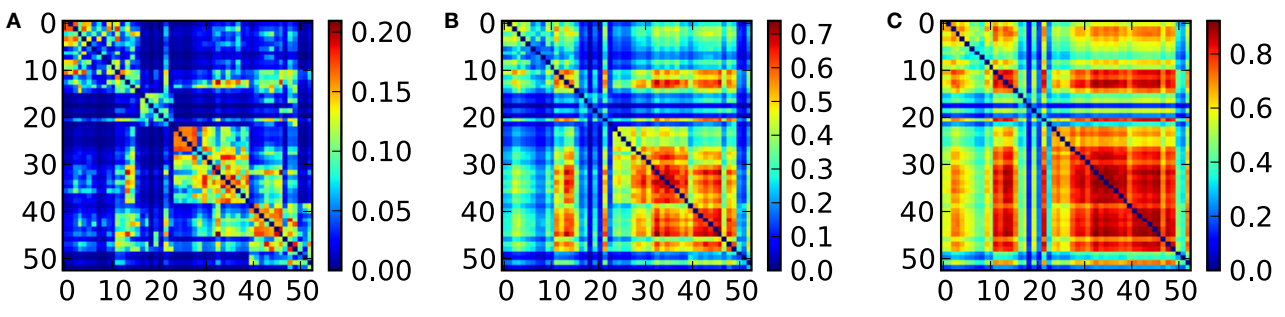


FIGURE 4 | Covariance matrix of the cat cortical network as a linear system. The adjacency matrix has been previously normalised by $1/\lambda_{max}$ and the noise level set to $\xi_i = 1.0$. Coupling strengths are: **(A)** $g = 0.52$, **(B)** $g = 0.84$ and **(C)** $g = 0.92$.

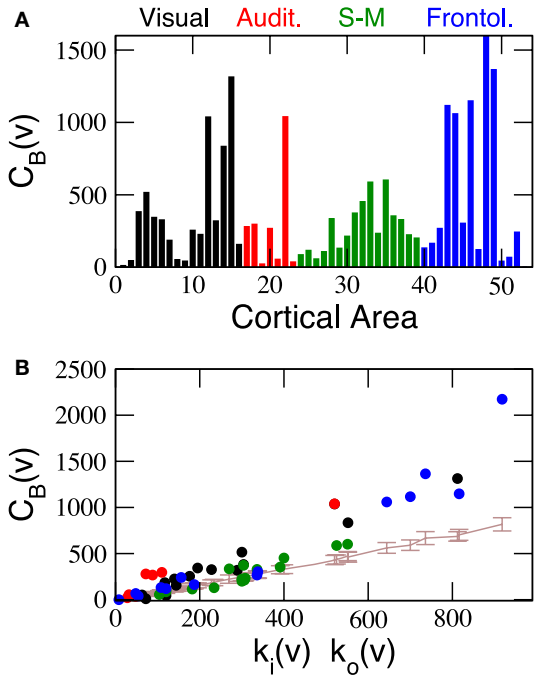


FIGURE 5 | Centrality of cortical areas. (A) Betweenness of cortical areas shows that at each sensory system few areas are very central. **(B)** Comparison between C_B of cortical areas and the expected centrality due to their degree (brown line). As a consequence of the modular and hierarchical organisation of the network, low degree areas closely follow the expected centrality but hubs are significantly more central than expected. Communication paths between sensory systems are centralised through the hubs.

of the hubs is largely significant. This is an evident consequence of the modular organisation of the network and the particular role of the cortical hubs for the inter-modal communication. Communication paths running between low-degree areas of different modules are usually mediated through the hubs (Zamora-López et al., 2009).

This significance test permits us to uncover the most likely candidates to be a hub of the network, not only in terms of their number of links, but considering their contribution for the corticocortical communications. The hubs found here are potential candidates to perform high level integration because they have access to the information of different modalities. However, with the current results we can only affirm with certainty that the hubs are useful for the *transmission* of information from one modality to another. Concluding whether they perform any further function or not, it requires a more careful analysis.

Collective organisation of cortical hubs

A relevant question is now whether the cortical hubs are functionally independent of each other, i.e. each hub has a specialised function, or they perform some collaborative function. A graph measure to characterise the relation between the hubs of a network is the *rich-club phenomenon*. The k-density $\phi(k)$, is defined as the internal density of links between the nodes with degree larger than k':

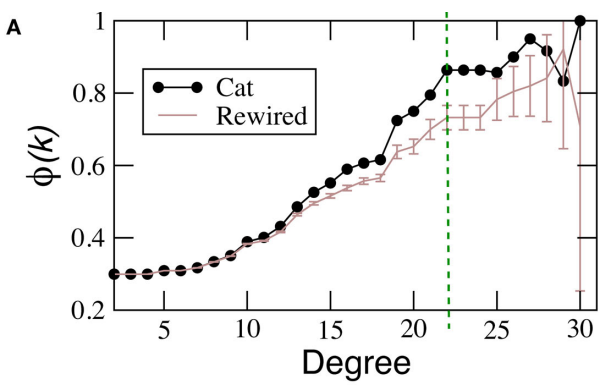
$$\phi(k') = \frac{L_{k'}}{N_{k'}\left(N_{k'}-1\right)}, \tag{9}$$

where $N_{k'}$ is the number of nodes with $k(v) \geq k'$ and $L_{k'}$ is the number of links between them. Notice that $\phi(k)$ is an increasing function of k. As $\phi(0) = \frac{L}{N(N-1)}$ is the density of the network, after the nodes with low degrees are removed the remaining reduced network contains more links per node. Thus, a plain measure of $\phi(k)$ is not very informative because hubs have a higher intrinsic chance of being connected to each other. Again, a conclusive interpretation requires the comparison to random networks with the same degree distribution. The question is then whether $\phi(k)$ of the real network grows faster or slower with k than the expected k-density $\phi_{1n}(k)$ out of the surrogate networks $\{\mathcal{G}_{1n}\}$. If $\phi(k)$ grows faster than $\phi_{1n}(k)$, it means that the hubs are more connected than expected and form a dense module (a *rich-club*). On the contrary, if $\phi(k)$ grows slower than $\phi_{1n}(k)$, the hubs are more independent of each other than expected.

In **Figure 6A** the k-density $\phi_{cat}(k)$ of $\mathcal{G}_{cat}$ is presented together with the ensemble average $\phi_{1n}(k)$. For low degrees, $\phi_{cat}(k)$ follows very close the expectation, but for degrees $k(v) > 15$, $\phi_{cat}(k)$ starts to grow faster showing that the hubs of the network form a *rich-club*. The largest difference occurs for $k = 23$, comprising of 11 cortical hubs from all the four sensory systems (**Figure 6B**). Compared to the internal density of the four modules of the network, we find that the hubs form an even denser module (**Table 1**).

Topological similarity of cortical hubs

A central assumption in systems neuroscience is that the function of brain regions are specified by their afferents and efferents (Passingham et al., 2002). Under this assumption, it is to be expected that cortical areas of similar function, i.e. specialised in the processing of same modal information, should display a similar pattern of



B

Visual	Auditory	Somato-Motor	Frontolimbic
20a, 7, AES	EPp	6m, 5Al	Ia, Ig, CGp, 35, 36

FIGURE 6 | Rich-club organisation. (A) k-density of the corticocortical network of the cat ϕ_{cat}, compared to the expectation out of the surrogate ensemble $\{\mathcal{G}_{1n}\}$. The largest difference occurs at $k = 23$ (vertically dashed line) giving rise to **(B)** a rich-club composed of 11 areas.

projections. In the case of the cortical hubs, it has been shown in the previous section that they form a tightly connected module. Whether this module could be regarded as a functional module, at least from a topological point of view, is the goal of the following analysis. The matching index $MI(v,v')$ is a graph measure to estimate the topological similarity of two nodes, by counting the number of common neighbours of v and v' (Section "Graph Analysis"). In order to compare the values obtained for different pairs, the measure is normalised such that $MI(v,v') = 1$ only if all the neighbours of node v are also all the neighbours of v'. See the example in **Figure 1**.

We have computed the matching index for all pairs of cortical areas and the result is shown in matrix form, **Figure 7A**. Visual inspection reveals the modular organisation of the network. This is reflected by the fact that $MI(v,v')$ is typically larger if both v and v' belong to the same anatomical module, than if they belong to different modules. To highlight this difference, in **Figure 7B** the distribution of the matching values is shown: when the areas belong to the same module (internal matching), or to different modules (external matching). The external matching has a broad skewed distribution but peaking near $MI = 0.15$. The internal matching displays a more constrained distribution with maximum at approximately $MI = 0.55$. In **Table 1** the average matching of the network is compared to the average internal matching for each of the anatomical modules V, A, SM and FL. The internal averages are always larger than the global average despite the broad deviations, confirms the expected functional cohesiveness of the modules; not only in terms of their internal density of connections, but also in terms of their common connectivity.

As pointed out, the distribution of external matching is skewed and contains some larger values up to $MI \approx 0.6$. We find that most of these larger values are contributed precisely by the links between the cortical hubs which lie in different modules. We have recomputed the distribution of external matching, but ignoring the matching between the cortical hubs (solid line in **Figure 7C**). The distribution decays now faster than in **Figure 7B**. Finally, the distribution of the internal matching for the 11 hubs forming the rich-club is displayed (dotted line of **Figure 7C**). It appears clearly separated from that of the distribution of external matching and peaking near $MI = 0.55$. Its average is 0.52 ± 0.10, comparable to, or larger than, the internal matching of the anatomical modules, **Table 1**. These observations support the idea that the cortical hubs form a functional module on their own, as the anatomical modules do.

Hierarchical organisation and integration capacity

The two structural properties of the cortical hubs here presented, (*i*) hubs are densely connected with each other and (*ii*) they are functionally interrelated in terms of their inputs and outputs, extend the current understanding of cortical networks by uncovering that the

Table 1 | Comparison between the anatomical modules and the *Rich-Club*. Both the internal density of links and the average matching of the areas in each of the functional modules V, A, SM and FL are larger than the whole network averages. The same happens for the areas in the *Rich-Club*, with values comparable to, or larger than those for the anatomical modules.

	Whole net	Visual	Auditory	S.-M.	Frontol.	Rich-Club
Density	0.30	0.58	0.81	0.74	0.65	0.86
Matching	0.28 ± 0.18	0.39 ± 0.13	0.47 ± 0.16	0.58 ± 0.13	0.40 ± 0.17	0.52 ± 0.10

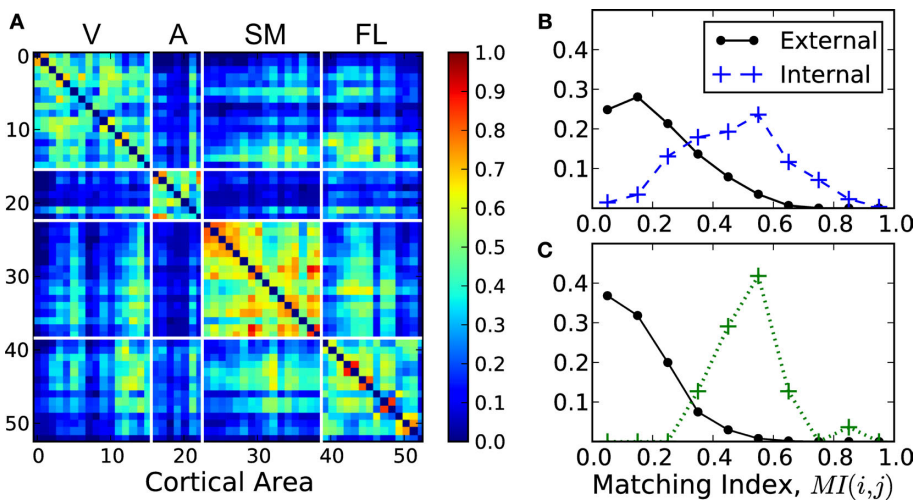


FIGURE 7 | Topological similarity of cortical areas. (A) Pairwise matching index $MI(v,v')$ for all areas summarised in matrix form. Self-matching $MI(v,v)$ is ignored for visualisation. **(B)** Distribution of the MI values in **(A)** if the areas v and v' are in the same anatomical module V, A, SM or FL (dashed line), and if they belong to different modules (solid line). **(C)** Recomputed distribution of MI if the areas belong to different modules, but cortical hubs are discarded (solid line). And distribution of $MI(v,v')$ only if v and v' are hubs in the *Rich-Club* (dotted line).

multisensory hubs form yet another module which lies at a higher level in the hierarchical organisation. In the complex networks literature one finds two types of hierarchical topologies. The model after Arenas et al. (2006) considers hierarchies as the agglomeration of modules, say, small modules join to form larger modules, **Figure 8A**. Another type of hierarchy after Ravasz and Barabási (2003) can be regarded as a tree-like fractal structure which produces modular networks with scale-free degree distribution. At each level, there is a central community connecting to all the modules at the same level, and to all modules in the hierarchies below. Such centralised patterns are repeated through different scales, **Figure 8B**.

The organisation that we uncover here is none of these two, but it might be regarded as a combination of them. Notice that in the model by Arenas et al. (2006), the small communities are randomly linked to each other such that their union forms a larger community. In the present case, the inter-community links are not random, but centralised. Therefore, the highest hierarchical level is formed by a *partial overlap* of the underlying modules. See **Figure 8C** for a schematic representation.

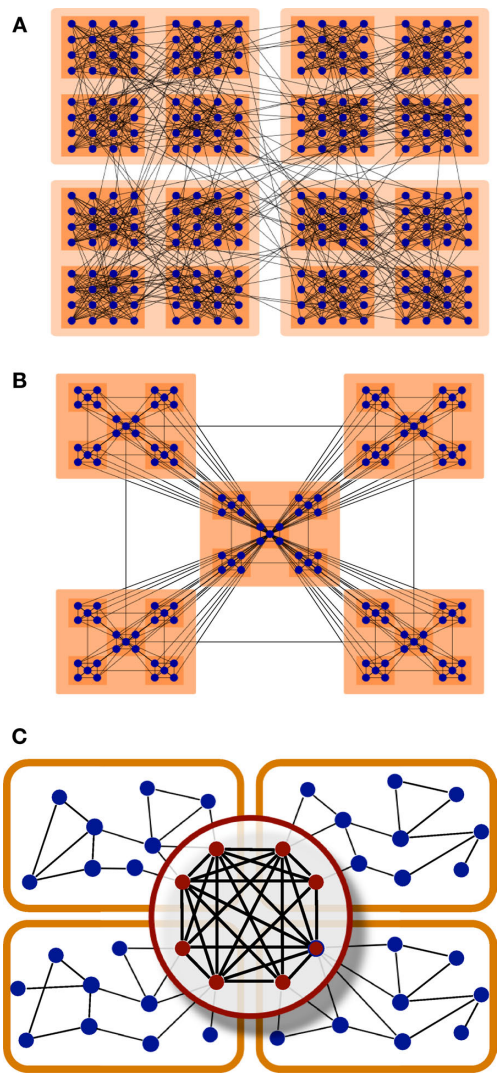


FIGURE 8 | Hierarchical organisation of complex networks. (A) Hierarchies as agglomeration of modules (Arenas et al., 2006). **(B)** Centralised and fractal hierarchical model (Ravasz and Barabási, 2003). **(C)** Illustrative representation of the modular and hierarchical structure found in the corticocortical connectivity of the cat. The highest hierarchical level is formed by a densely interconnected overlap of the modules.

The functional implications of the topological findings described in this section, necessarily arise from intuitive interpretation of the intrinsic relationship between structure and function in neural systems. To provide a more solid ground to these intuitive interpretations, in the following section we challenge them by means of dynamical and information theoretical measures. We focus in a very simple dynamical model which has the benefit of being analytically solvable, although its validity for our purposes is confirmed by comparison to the dynamical output of more complex models, see Section "Information Theory and Integration".

FUNCTIONAL CAPACITY OF INTEGRATION

The structural organisation described in the previous section supports the idea that the cortical hubs might be responsible for combining the multisensory information hence facilitate the emergence of a global (integrated) perception. In this sense, we aim for a definition of integration which characterises *the capacity of one or more nodes to receive information of different character and combine it to produce new useful information*. Certainly, this definition involves crucial theoretical problems, e.g. what the *character* of information is, or what are the rules under which information is *combined*. Nevertheless, within a networked system, the nodes with a capacity to integrate information should obey certain measurable conditions. We propose the following:

1) Accessibility to information: A node can perform an integrative function only if it has general access to the information contained within the system.
2) Sharing of information: Two or more nodes can perform integrative function in a collaborative manner only if they are sufficiently connected with each other.
3) Segregation after selective damage: If a node has an integrative function, its removal should lead to a decrease of the integrative capabilities of the whole system.

From the structural point of view, the hubs listed in **Figure 6B** obey these three conditions. They are the most central areas and they are densely connected to each other. Besides, robustness studies (Kaiser et al., 2007) have shown that intentional lesion of the highly connected cortical areas largely affect the communication within the network. In the following, we introduce a framework to characterise the integrative function of the hubs by means of dynamical systems and information theory. Additionally, we perform a probabilistic analysis of the composition of the dynamical core, rather than a deterministic one. The reason is that even if the corticocortical networks of the cat is the most complete and reliable dataset of its kind up to date, it is not free of experimental errors. For example, some of the real connections might still be absent in the data. We aim to discriminate those hubs which, grouped together, possess a larger potential to integrate multisensory information from those groups which might have lesser capacities. For that, we

arbitrarily choose all the areas with output degree $k_o(v) \geq 20$ as potential members of the integrator module giving rise to a set of $N_S = 19$ areas:

$\mathcal{S}_{\text{hubs}}$ = {20a, 7, AES, EPp, 6l, 6m, 5Am, 5Al, 5Bm, 5Bl, SSSAi, SSAo, PFCL, Ia, Ig, CGa, CGp, 35, 36}.

The statistical analysis consists in measuring the integrative capacities of all the 524,097 combinations of sizes $N_S = 1$ to $N_S = 19$ out of the 19 hubs in $\mathcal{S}_{\text{hubs}}$.

Integration capacity after sensory stimulation

Consider the linear System (6) with $\hat{\mathbf{A}}$ being the transposed and normalised adjacency matrix of the cat $\mathcal{G}_{\text{eat}}$. All areas are driven by a small Gaussian noise level $\xi_i = 1.0$ and coupled by $g = 0.5$. This case might be regarded as the activity of the network in the *resting-state* because all x_i are driven by noise of small intensity and there is no sensory input. Now, we intend to illustrate the joint capacity of a group of areas to integrate information of different character. Even if it is unclear how to define the character of information, in the case of cortical networks it is known that sensory information enters the cortex through specific regions termed as primary sensory areas: primary visual cortex (area 17), primary auditory cortex (area AI) and primary somatosensory cortex (areas 1, 2 and 3b). According to Scannell et al. (1995) the cortical areas 1, 2 and 3b are subregions of the primary somatosensory area, named by some authors as SI. Hence, we simultaneously excite all the primary sensory areas {17, AI, 1, 2, and 3b} by assigning them a larger noise level $\xi_j = 10.0$) and we measure the integration $I(S)$ of all the subsets S of hubs out of $\mathcal{S}_{\text{hubs}}$. Because of the excited condition, we denote the integration of the subsets as $I^e(S)$.

The results depicted in **Figure 9A** show that $I^e(S)$ can largely differ. For example, among all the subsets of size $N_s = 10$, the integration of some of them is very small, $I^e(S) \sim 0.1$, while the integration of others becomes much larger, $I^e(S) \sim 0.5$. These differences permit us to identify those cortical hubs which, grouped together, become more statistically dependent among them as a consequence of the multisensory stimulation. Considering only those subsets whose $I^e(S)$ lies within the largest 10% (red crosses in **Figure 9A**) a co-participation matrix **C** is constructed such that C_{ij} is the number of times (given in frequency) that two cortical hubs participate together in one of the maximal sets, **Figure 9B**. It is observed that areas {7, AES; EPp; 6m; Ia, Ig, CGp, 35, 36} participate together in over 75% of all the maximal sets. Visual area 20a and the somatosensory-motor area 6l participate only in 50% of the occasions with those areas in the core. The remaining areas, {5Am, 5Al, 5Bm, 5Bl, SSSAi, SSSAo and PFCL}, can be discarded as members of the dynamical core.

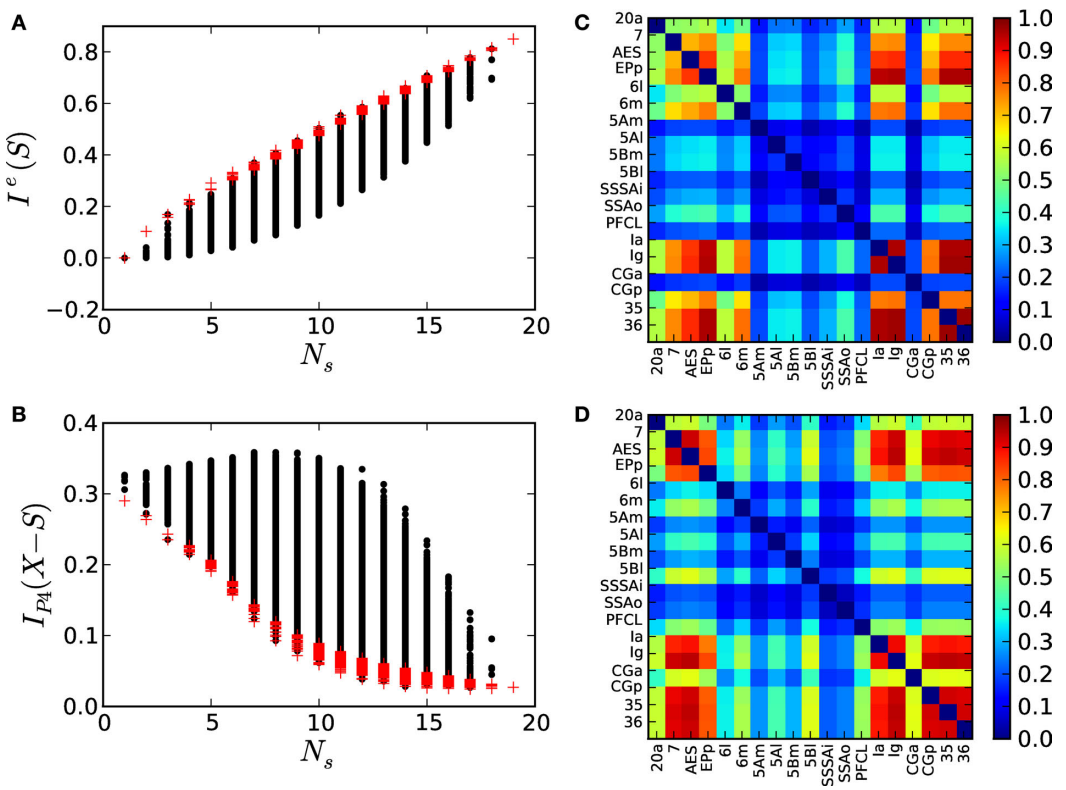


FIGURE 9 | Functional segregation and integration. (A) Local integration $I(S)$ of cortical hubs after stimulation of the primary sensory areas. **(B)** Co-participation matrix of cortical hubs within the subsets leading to large $I^e(S)$ (red dots). **(C)** Modular integration I_{P_4} of the sensory modules V, A, SM and FL after simultaneous lesion of cortical hubs. N_S is the number of hubs removed. **(D)** Co-participation matrix of the hubs within the subsets S which lead to a larger decrease in the dynamical dependence $\left(I_{P_4}(S)\right)$ of the sensory modules (marked by red dots).

Dynamical segregation after multiple lesions

Within a networked system the removal of critical nodes should lead to a decrease in its integrative capacities. In the following, we study the impact of targeted lesions of the corticocortical network of the cat, $\mathcal{G}_{\text{cat}}$. For all the possible subsets S composed of hubs in $\mathcal{S}_{\text{hubs}}$, we perform a lesion to the network by simultaneously removing the nodes $x_i \in S$ and characterise the consequent functional segregation of the network $\mathcal{G}_S = \mathcal{G}_{\text{cat}} - S$ as the change in statistical dependence between the four modules (V, A, SM and FL). Lesion of areas critical for the integration capacities of the system should lead to a dynamical segregation of the modules, i.e. a decrease in their statistical dependence.

Recall that integration $I(X)$ as defined in Eq. 7 is an extension of the mutual information for more than two systems. It represents the limit case in which the statistical dependence among all the elements x_i in the system X is quantified. To cover different scales of organisation we propose to characterise the statistical dependence between groups of elements. Imagine a partition $\mathcal{P} = \{S_1, S_2, \ldots, S_n\}$ into n groups (modules) of the elements x_i such that $X = S_1 \cup S_2 \cup \ldots \cup S_n$. Then, we define the modular integration of the partition $\mathcal{P}$ as:

$$I_{\mathcal{P}}(X) = \sum_{j=1}^{n} H\left(S_j\right) - H(X). \tag{10}$$

Note that when $n = N$, then $I_{\mathcal{P}}(X) = I(X)$.

Considering the partition $\mathcal{P}_4 = \{\text{V, A, SM, FL}\}$ and the corticocortical network of the cat, then $I_{\mathcal{P}_4}\left(\mathcal{G}_{\text{cat}}\right) = 0.292$. The modular integration of each lesioned network $\mathcal{G}_S$ is computed for the partition $\mathcal{P}_4$. Notice that (*a*) the nodes are also removed from the partition and (*b*) every $\mathcal{G}_S$ is adequately normalised by its largest eigenvalue such that the measured observables are comparable across realisations (see Section "Information Theory and Integration"). The results in **Figure 9C** permit us again to discriminate between subsets of hubs whose simultaneous removal lead to a large segregation of the network, while removal of other subsets has barely no effect. For example, among all the possible lesions of size $N_S = 10$, some trigger a large segregation of the modules, $I_{\mathcal{P}_4}\left(\mathcal{G}_S\right) \sim 0.05$ while other lesions do even increase their dynamical dependence: $I_{\mathcal{P}_4}\left(\mathcal{G}_S\right) \sim 0.35 > I_{\mathcal{P}_4}\left(\mathcal{G}_{\text{cat}}\right)$.

Selecting only those subsets whose lesion leads to a larger segregation of the modules, i.e. $I_{\mathcal{P}_4}\left(\mathcal{G}_S\right)$ lies among 10% of the minimal modular integration for each size N_S (red dots in **Figure 9A**), a co-participation matrix **C** is constructed, **Figure 9D**. The entries C_{ij} are the number of times (given in frequency) that two areas participate together in one of the minimal subsets. A core of cortical areas is found which participate together in over 70% of these cases: {7, AES; EPp; Ia, Ig, CGp, 35, 36}. Somatosensory-motor areas 6m, 5Al and 5Bl join them in over 50% of the cases.

In summary, both the multiple lesion and the multisensory excitation analysis performed in this section lead to the identification of the same cortical hubs as responsible for the integration of multisensory information in the corticocortical network of the cat. Moreover, this set largely coincides with the top hierarchical level found by the graph analysis in Section "Topological Capacity of Integration", corroborating the integrative function assigned to the hubs by intuitive interpretation of their topological characteristics.

SUMMARY AND DISCUSSION

In this paper we have analysed the modular and hierarchical organisation of the corticocortical network of the cat and its relationship to the intrinsic necessities of the brain to simultaneously segregate and integrate multisensory information. From the topological point of view, we have extended the current understanding of cortical organisation with the finding that the cortical hubs form a central module on top of the cortical hierarchy; which is expressed as the partial overlap of the four anatomical modules (visual, auditory, somatosensory-motor and frontolimbic). By means of dynamical and information theoretical measures, we have corroborated its capacity to integrate multisensory information, i.e. after simultaneous excitation of visual, auditory and somatosensory primary areas, a particular set of hubs becomes statistically dependent forming a dynamical cluster. Additionally, the simultaneous lesion of these hubs leads to a largest decrease in the integrative capacities of the network. Both structural and functional results indicate that visual areas 7 and AES, auditory area EPp and frontolimbic areas Ia, Ig, CGp, 35 and 36 are the most likely candidates to form the top hierarchical module. The participation of somatosensory-motor areas is less clear, although area 6m is the strongest candidate of them. Visual area 20a and somatosensory-motor areas 5Al and 5Bl are also potential candidates.

The modular and hierarchical organisation here detected agrees with the behaviour observed in dynamical simulations of cortical networks. The resting state dynamics are typically governed by the formation of dynamical clusters which closely relate to the anatomical modules, but the influence of the hierarchical organisation is also expressed. In Zemanová et al. (2006) and Zhou et al. (2006, 2007) it was shown that the correlation between the dynamical clusters is mediated by the cortical hubs. In Honey et al. (2007) the centrality of the hubs was found to oscillate in time. Simulation of excitable dynamics on hierarchical networks (Müller-Linow et al., 2008) has shown that the dynamical behaviour of the cortical network of the cat may be dominated either by the modular structure or by the hubs, depending on the time scales.

SEGREGATION, INTEGRATION AND LOCALISATION

The separation of modal information paths is a relevant characteristic of organisation in the nervous system that permits simultaneous (parallel) processing of sensory input and detection of its features. Cortical regions containing neurones specialised in similar function, e.g. in processing information of the same sensory modality, lie geographically close to each other (**Figure 10A**). However, a coherent perception and the emergence of mental states such as awareness and consciousness require that information is integrated at different levels: the binding of sensory features into entities, the combination of entities with memories (personal experiences) into events, etc. While experimental techniques have led to a deep understanding about the basis of sensory perception, the nature of integration and the localisation of brain regions involved in it, is still under the subject of debate.

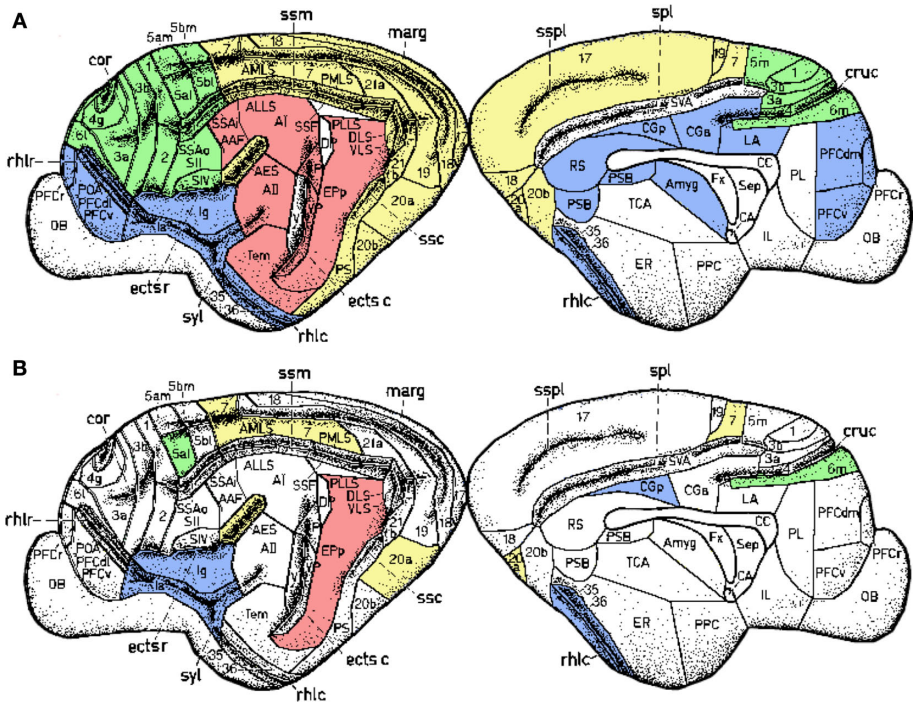


FIGURE 10 | Spatial location of the areas according to their modality: visual (yellow), auditory (red), somatosensory-motor (green) and frontolimbic (blue). While areas of similar modality tend to lie close to each other **(A)**, the hubs form a topological cluster which is spatially delocalised **(B)**.

As stated by Fuster (2003), simple extrapolation of the principles of sensory organisation do not lead to the identification of the substrate for cognition.

Several models have proposed that high-level functions are represented by distributed, interactive and overlapping networks of neurones, which transcend any of the traditional subdivisions of the cortex by structural (cytoarchitecture) or functional criteria (Damasio, 1989; Fuster, 2003, 2006; Tononi, 2004). During the recent years increasing experimental evidence has confirmed this hypothesis and the networked perspective has gained the favour against the assumption of a single brain region fully responsible for integration (Stam and Reijneveld, 2007; Bullmore and Sporns, 2009; Knight, 2009). The anatomical networked connectivity may serve as the basis in which localised and distributed functional networks rapidly emerge and dissolve governed by coordination dynamics according to the sensory stimulation and the ongoing activity (Bressler and Kelso, 2001).

As a further evidence, our results resolve the anatomical organisation substrate that supports the capacity of the cerebral cortex to simultaneously segregate and integrate information. In the light of this organisation, it could be envisioned that multisensory integration emerges from the collaborative function of the cortical hubs. While *early* sensory cortical regions perform specialised processing of the sensory input, the hubs of the network may work together to combine the multisensory information. A relevant organisation difference is that the cortical hubs form a module which is densely connected by axonal paths through the white matter, but is *geographically delocalised* (**Figure 10B**).

LIMITATIONS AND OUTLOOK

The current paper focuses in the corticocortical connectivity of cats because it is, up to date, the most complete and reliable dataset of its kind. Hence, it is the most suitable for a detailed and statistically consistent analysis. The main limitation is that it comprises of interconnection between cortical areas in only one cerebral hemisphere. Because of the known inter-hemisfere differences in many mammals, particularly in humans, it will be very valuable in the future to acquire the connectivity within and between both hemispheres in animal and human models. Based on current literature in which the cortical networks of the macaque and cat models display similar features, we expect that the general organisation principles here exposed to be valid in a wide range of mammals.

An interesting challenge is now to explain the emergence of this modular and hierarchical organisation in terms of evolution and development, in particular how the delocalised cluster of hubs could have evolved if, apparently, areas of similar function tend to be grouped close to each other. Very likely, the balancing between short wiring requirements (leading to minimisation of energy costs) and short processing paths allowing for robustness and fast responses (Kaiser and Hilgetag, 2006) plays a major role. It would also be of relevance to find out whether similar hierarchical patterns are repeated across smaller scales within the cortex, i.e. the interconnections between cortical columns and microcolumns. This would imply an underlying fractal-like complex architecture which can emerge from simple rules of assembly during development.

Finally, we should remind that current non-invasive techniques such as EEG, MEG and fMRI reveal only the presence of brain activity. They permit to identify which brain regions are *associated* with certain experimental condition. However, at the current stage it is very difficult, if not impossible, to understand *what* is exactly an activated region doing. Is it filtering a signal? Is it integrating information? Is an activation detected only because that particular region contains memories which are being retrieved and passed to other regions for processing? In our opinion, it would be highly interesting to further develop concepts of information theory as the modular and local capacity of integration here presented which applied to the time series of regional activity might help understand the particular function of individual brain regions within a given experimental task.

ACKNOWLEDGMENTS

We thank Lucia Zemanová, Claus-C. Hilgetag and Werner Sommer for valuable discussions. Gorka Zamora-López and Jürgen Kurths are supported by the Deutsche Forschungsgemeinschaft, research group FOR 868 (contract No. KU 837/23-1) and by the BioSim network of excellence (contract No. LSHB-CT-2004-005137 and No. 65533).

REFERENCES

Anthonisse, J. M. (1971). The rush in a directed graph. Technical Report BN9/71. Amsterdam, Stichting Mathematisch Centrum.

Arenas, A., Díaz-Guilera, A., and Pérez-Vicente, C. (2006). Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.* 96, 114102.

Bear, M. F., Connors, B. W., and Paradiso, M. A. (2006). Neuroscience: Exploring the Brain. Baltimore, MD, Lippincott Williams and Wilkins.

Bressler, S. L., and Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends Cogn. Sci.* 5, 26–36.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 1.

Cover, T. M., and Thomas, J. A. (1991). Elements of Information Theory. New York, Wiley.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neuronal substrates of recall and recognition. *Cognition* 33, 25–62.

Engel, A. K., and Singer, W. (2001). Temporal binding and neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25.

Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proc. R. Soc. Lond., B, Biol. Sci.* 254, 199–203.

Faser, A. M., and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* 33, 2318–2321.

Freeman, L. C. (1977). A set of measures of centrality based upon betweenness. Sociometry 40, 35–41.

Fuster, J. M. (2003). Cortex and Mind: unifying cognition. New York, Oxford University Press.

Fuster, J. M. (2006). The cognit: a network model of cortical representation. *Int. J. Psychophysiol.* 60, 125–132.

Graben, P., Zhou, C. S., Thiel, M., and Kurths, J. (eds). (2007). Lectures in Supercomputational Neuroscience. Berlin, Springer-Verlag.

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

Hilgetag, C. C., Burns, G. A., O'neill, M. A., Scannell, J. W., and Young, M. P. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 91–110.

Hilgetag, C. C., and Kaiser, M. (2004). Clustered organisation of cortical connectivity. *Neuroinformatics*. 2, 353–360.

Holland, P. W., and Leinhardt, S. (1977). The statistical analysis of local structure in social networks. In Sociological Methodology, D. R. Heise, ed. (San Francisco, Jossey-Bass), pp. 1–45.

Honey, C. J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10240–10245.

Kaiser, M., and Hilgetag, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* 2, e95. doi: 10.1371/journal.pcbi.0020095.

Kaiser, M., Martin, R., Andras, P., and Young, M. P. (2007). Simulation of robustness against lesions of cortical networks. *Eur. J. Neurosci.* 25, 3185–3192.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). Principles of Neural Science. New York, McGraw-Hill.

Kannan, R., Tetali, P., and Vempala, S. (1999). Simple Markov-chain algorithms for generating bipartite graphs and tournaments. *Random Struct. Algorithms* 14, 293–308.

Katz, L., and Powell, J. H. (1957). Probability distributions of random variables associated with a structure of the sample space of sociometric investigations. *Ann. Math. Stat.* 28, 442–448.

Knight, R. T. (2009). Neural networks debunk phrenology. *Science* 316, 1578.

Kötter, R., and Sommer, F. T. (2000). Global relationship between anatomical connectivity and activity propagation in the cerebral cortex. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 127–134.

Müller-Linow, M., Hilgetag, C. C., and Hütt, M. T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS Comput. Biol.* 4, e1000190. doi: 10.1371/journal.pcbi.1000190.

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

Papoulis, A. (1991). Probability, Random Variables and Stochastic Processes. New York, McGraw-Hill.

Passingham, R. E., Stephan, K. E., and Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nat. Rev.* 3, 606.

Rao, A. R., Jana, R., and Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhya A* 58, 225–242.

Ravasz, E., and Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E* 67, 026112.

Roberts, J. M. (2000). Simple methods for simulating sociomatrices with given marginal totals. *Soc. Netw.* 22, 273–283.

Robertson, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nat. Rev.* 4, 93–102.

Scannell, J. W., Blakemore, C., and Young, M. P. (1995). Analysis of connectivity in the cat cerebral cortex. *J. Neurosci.* 15, 1463–1483.

Scannell, J. W., and Young, M. P. (1993). The connectional organization of neural systems in the cat cerebral cortex. *Curr. Biol.* 3, 191–200.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.

Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586.

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

Sporns, O., and Tononi, G. M. (2001). Classes of network connectivity and dynamics. *Complexity* 7, 28–38.

Stam, C. J., and Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed. Phys.* 1, 3.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5, 42.

Tononi, G., McIntosh, A. R., Russell, D. P., and Edelman, G. M. (1998). Functional clustering: identifying strongly interactive brain regions in neuroimaging data. *Neuroimage* 7, 133.

Tononi, G., and Sporns, O. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5033.

Tononi, G., Sporns, O., and Edelman, G. M. (1996). A complexity measure for selective matching of signals by the brain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 3422.

Young, M. P., Hilgetag, C.-C., and Scannell, J. W. (2000). On imputing function to structure from the behavioural effects of brain lesions. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 147–161.

Zamora-López, G., Zhou, C. S., and Kurths, J. (2009). Graph analysis of cortical networks reveals complex anatomical communication substrate. *Chaos* 19, 015117.

Zemanová, L., Zhou, C. S., and Kurths, J. (2006). Structural and functional clusters of complex brain networks. *Phy. D* 224, 202–212.

Zhou, C. S., Zemanová, L., Zamora-López, G., Hilgetag, C. C., and Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity

in complex brain networks. *Phys. Rev. Lett.* 97, 238103.

Zhou, C. S., Zemanová, L., Zamora-López, G., Hilgetag, C. C., and Kurths, J. (2007). Structure-function relationship in complex brain networks expressed by hierarchical synchronization. *New J. Phys.* 9, 178.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 April 2009; paper pending published: 10 June 2009; accepted: 02 February 2010; published online: 19 March 2010.

Citation: Zamora-López G, Zhou C and Kurths J (2010) Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. Front. Neuroinform. ***4****:1. doi: 10.3389/neuro.11.001.2010*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 14 May 2010
doi: 10.3389/fninf.2010.00008

Optimal hierarchical modular topologies for producing limited sustained activation of neural networks

Marcus Kaiser*[1,2,3] *and Claus C. Hilgetag*[4,5]

[1] *School of Computing Science, Newcastle University, Newcastle Upon Tyne, UK*
[2] *Institute of Neuroscience, Newcastle University, Newcastle Upon Tyne, UK*
[3] *Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Korea*
[4] *School of Engineering and Science, Jacobs University Bremen, Bremen, Germany*
[5] *Department of Health Sciences, Boston University, Boston, MA, USA*

Edited by:
Rolf Kötter, Radboud University Nijmegen, Netherlands

Reviewed by:
Olaf Sporns, Indiana University, USA
Ingo Bojak, Radboud University Nijmegen, Netherlands

****Correspondence:***
Marcus Kaiser, School of Computing Science, Newcastle University, Claremont Tower, Newcastle Upon Tyne NE1 7RU, UK.
e-mail: m.kaiser@ncl.ac.uk

An essential requirement for the representation of functional patterns in complex neural networks, such as the mammalian cerebral cortex, is the existence of stable regimes of network activation, typically arising from a limited parameter range. In this range of limited sustained activity (LSA), the activity of neural populations in the network persists between the extremes of either quickly dying out or activating the whole network. Hierarchical modular networks were previously found to show a wider parameter range for LSA than random or small-world networks not possessing hierarchical organization or multiple modules. Here we explored how variation in the number of hierarchical levels and modules per level influenced network dynamics and occurrence of LSA. We tested hierarchical configurations of different network sizes, approximating the large-scale networks linking cortical columns in one hemisphere of the rat, cat, or macaque monkey brain. Scaling of the network size affected the number of hierarchical levels and modules in the optimal networks, also depending on whether global edge density or the numbers of connections per node were kept constant. For constant edge density, only few network configurations, possessing an intermediate number of levels and a large number of modules, led to a large range of LSA independent of brain size. For a constant number of node connections, there was a trend for optimal configurations in larger-size networks to possess a larger number of hierarchical levels or more modules. These results may help to explain the trend to greater network complexity apparent in larger brains and may indicate that this complexity is required for maintaining stable levels of neural activation.

Keywords: modularity, functional criticality, brain connectivity, neural dynamics, neural networks, cerebral cortex

INTRODUCTION

Complex systems operate within a critical functional range (Bak et al., 1987), sustaining diverse dynamical states on the basis of their intricate system architecture. Criticality is associated with the phase transition between ordered and chaotic dynamics, and systems tuned to the critical point produce power-law distributions in their dynamics. Recent studies indicate that brain networks also operate close to a critical point. Evidence for this comes, for example, from the observation of neuronal avalanches (i.e., bursts of activity separated by longer periods of relative rest) with a power-law size distribution in cortical slices (Beggs and Plenz, 2003), and from time series analysis of EEG data (Freeman et al., 2000) showing that the power spectral density of background activity follows a power law. Critical activity has also been demonstrated in human brain functional networks (Kitzbichler et al., 2009). While its functional significance is still not well understood, it has been suggested that critical dynamics may enhance information processing capabilities of neuronal networks (e.g., Bertschinger and Natschläger, 2004). This idea is supported by work showing that the dynamic range in an excitable network is optimized at criticality (Kinouchi and Copelli, 2006). Given these findings, it is desirable to obtain a better understanding of the conditions for criticality in complex excitable networks.

A necessary precondition for attaining a critical point in intricate neural systems, such as the mammalian cerebral cortex, is that initial network activations result in neuronal activation patterns that neither die out too quickly nor spread across the entire network. Without this feature, activation patterns would not be stable, or would lead to a pathological excitation of the whole brain. What are the essential structural and functional parameters that allow complex neural networks to maintain such a dynamic balance of sustained yet limited activity? Most current models of neural network dynamics focus on maintaining the right balance of network activation and rest through functional interactions among populations of inhibitory and excitatory nodes (Beggs and Plenz, 2003). Alternative balancing mechanisms may be provided by broad external input from neuromodulatory systems or self-sustained neuronal activity (Muresan and Savin, 2007). However, the topology of neural networks may also contribute to critical network dynamics, even in the absence of explicit inhibition. For this reason, we are particularly interested in the relationship of different kinds of neural network topology to the condition of limited sustained activation (LSA). The involvement of inhibitory neuronal populations and other dynamic control mechanisms may then further extend the parameter range for LSA that is provided by principal topological features of the neural network architecture.

Two central topological features of brain networks, in particular of the cerebral cortex, are their modular and hierarchical organization. A modular hierarchical organization of cortical architecture and connections is apparent across many scales, from cellular microcircuits in cortical columns (Mountcastle, 1997; Binzegger et al., 2004) at the lowest level, via cortical areas at the mesoscopic scale, to clusters of highly connected brain regions at the global systems level (Hilgetag et al., 2000; Breakspear and Stam, 2005; Kaiser, 2007). The precise organization of these features at each level is still sketchy, and there is exists controversy about the exact organization or existence of modules even at the level of cortical columns (Rakic, 2008; Smith, 2010). Nonetheless, current data and concepts suggest that at each level of neural organization clusters arise, with denser connectivity within than between the modules. This means that neurons within a column, area or cluster of areas are more frequently linked with each other than with neurons in the rest of the network.

The spreading of activity has been modeled for cortical networks (Kötter and Sommer, 2000) and other complex networks with a non-random organization (Pastor-Satorras and Vespignani, 2001). In a previous study of activity spreading through different topologies of excitable networks (Kaiser et al., 2007a), we showed that patterns of limited but sustained activity are well supported by the organization of hierarchical multi-modular networks, but not random or simple small-world networks (Watts and Strogatz, 1998) of the same size. In addition, such properties arose without the need for explicit inhibitory feedback or external input, demonstrating the significant role of network topology in sustaining and limiting neural activation (Latham and Nirenberg, 2004; Roxin et al., 2004).

While our previous study (Kaiser et al., 2007a) demonstrated that a wide range of initial parameters in hierarchical modular networks could result in LSA, it did not clarify whether this range was due mainly to the multi-modular organization of the network or its hierarchical structure. In the present study, we investigated the relation of different hierarchical network configurations to the range of LSA more extensively. The principal type of hierarchical structure, an interconnected set of modules with encapsulated sub-modules without explicit hub nodes, as well as the settings for the dynamic mechanisms were preserved from our previous model. A fixed number of nodes was activated at the beginning of each dynamical simulation. Other nodes became activated when at least k of their directly connected node neighbors were active at the same time. Each active node deactivated in the following time step with probability v. Note that this model only assumes initial activation at time step 0, but no ongoing external input or internal random activation.

Our hierarchical topological model reflects general features of brain connectivity at the large and mesoscopic scale, in particular the modularity of neural networks across scales. Nodes in the model are intended to represent cortical columns (Mountcastle, 1997) rather than individual neurons. Connections between columns are modeled as exclusively excitatory, since it is appears to hold that there are no long-distance inhibitory connections within the cerebral cortex (Latham and Nirenberg, 2004). However, nodes can also deactivate (controlled via the model's deactivation probability) due to intrinsic inhibition or exhaustion, as observed for cortical tissue after prolonged firing, for instance in epileptic seizures (Milton and Jung, 2003).

Two main parameters were explored in the hierarchical networks, the number of hierarchical levels and the number of sub-modules at each level (cf. **Figure 1**). These parameters were varied, while other topological features, such as the probability that any two nodes are connected, or alternatively, the average number of connections per node, were kept constant. We explored whether optimal hierarchical configurations existed, in which the proportion of tested cases with LSA was at a maximum. In addition, we tested whether the parameters for such optimal configurations changed with network size; that is, whether small networks, representing the approximate number of columns as in a small rodent (rat) brain, had different optimal settings than larger cortical networks that might reflect the number of column nodes in larger mammalian (cat) and primate (macaque) brains.

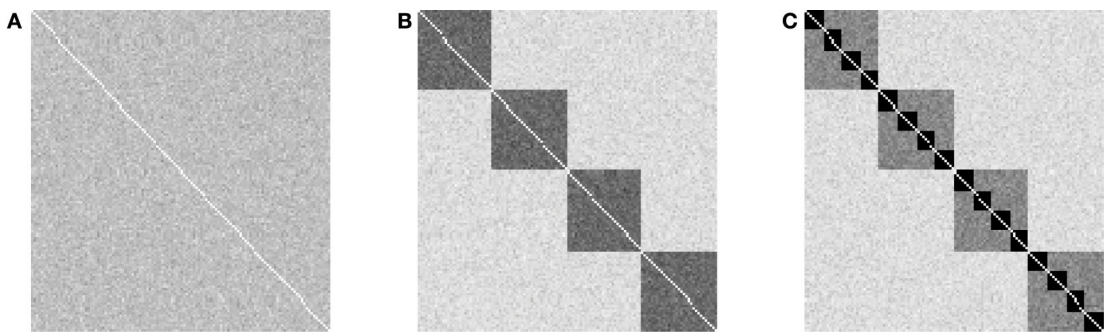


FIGURE 1 | Overview of variation of granularity and scales in the explored hierarchical modular networks. The plots show the outcome of 100 realizations of networks with 128 nodes and 4,096 directed edges. Gray level shading of the adjacency matrix indicates the frequency with which an edge was established (white: never established; black: established in all 100 generated networks). **(A)** Random networks without hierarchical structure, resulting from $h = 0$ (number of hierarchical levels) and $m = 0$ (number of sub-modules); **(B)** Flat modular networks with four modules, resulting from $h = 1$, $m = 4$; **(C)** Hierarchical modular networks with $h = 2$, $m = 4$. Note that each hierarchical level contains the same number of edges, resulting in 16 modules at the lowest hierarchical level in **(C)**, which possess the highest edge density.

MATERIALS AND METHODS

Calculations were performed on a 16-core HP ProLiant server using the Linux version of Matlab R2009a (Mathworks Inc., Natick, MA, USA). Scripts are available at http://www.biological-networks.org and are part of CARMEN (http://www.carmen.org.uk).

ANATOMICAL CONSTRAINTS

We investigated if the topology of optimal hierarchical networks, leading to a maximum parameter range of LSA, varied with brain size. For this approach, the number of nodes was set to the number of columns estimated to exist in one cortical hemisphere in different species. The number of columns was estimated from the surface size of one cortical hemisphere in rat (Nieuwenhuys et al., 1998), cat (Nieuwenhuys et al., 1998), and macaque (Felleman and van Essen, 1991) under the assumption of each (macro-)column occupying 1 mm^2. Real columns might be smaller and we elaborate on the role of column size differences across areas and species in the Section "Discussion". We explored three networks with different surface sizes for one hemisphere, rat-like (300 nodes; 3 cm^2 surface), cat-like (4,150 nodes; 41.5 cm^2 surface), and macaque-like (11,000 nodes; 110 cm^2 surface). Note that these are very simple estimates based on the assumption that columns in different species are comparable in the basic circuit layout even though the absolute number of neurons may vary (Herculano-Houzel et al., 2008). Either edge density or average number of edges per node (average degree $\langle k \rangle$) was kept constant across network sizes. The edge density was set to 1.2%, corresponding to the one chosen in a previous study (Kaiser et al., 2007a) and is close to values of 0.48% for a model of the rat cortex.

The constraint of a constant average node degree was motivated by comparative studies showing largely constant numbers of connections per neuron across many species (Hellwig, 2000; Schüz and Braitenberg, 2002; Binzegger et al., 2004; Striedter, 2004; Changizi and Shimojo, 2005; Schüz et al., 2006). At the column level, it can also be reasoned that columns are mostly connected with adjacent columns on the cortical surface (short-distance; intra-areal) and only a few columns in different areas (long-distance; inter-areal connections). Due to this presumed homogenous arrangement of cortical networks, the number of connections per column should be independent of the total number of columns in the network. The average degree was set to the arbitrary but fixed number of 50. Note that the actual values for edge density or average degree might differ from the ones chosen here without changing the principal conclusions of this study, as results across different networks were compared qualitatively.

GENERATING HIERARCHICAL NETWORKS

Alternative approaches exist for generating a hierarchical network with m sub-modules per module and a total number of levels h. As a default, we settled on a strategy in which the total number of edges E was distributed to the different levels (see **Figure 1**) with E_i edges on level i, so that each level received the same number of edges: $E_i = E/(h+1)$. This model, which was used throughout the study, preserved a constant number of edges when the number of levels or sub-modules within modules was varied.

The network was generated beginning with the highest level and adding modules to the next lower level with random connectivity within modules. The resulting networks were similar to the ones produced by an alternative procedure (Sporns, 2006), but differed in the generating algorithm (we present pseudocode here, the actual Matlab algorithm is available online at http://www.biological-networks.org):

```
for i from 0 to h − 1                    for all hierarchical levels
  A_i = (m − 1)/m^(i+1)                  proportion of the adjacency matrix occupied by one module at the current hierarchical level
  p_i = E_i/(N^2 A_i)                    edge density within a module
  N_i = N × (1/m)^i                      number of nodes in a module
  N_c = N/N_i                            number of modules at the current level

  for j from 1 to N_c
    c_i = random graph with N_i          random N_i×N_i graph with
          and edge density p_i           edge density p_i
    r_0 = 1 + (j − 1) N_i                first node of the module
    r_1 = r_0 + N_i − 1                  last node of the module
    CIJ(r_0 to r_1, r_0 to r_1) = c_i    part of the matrix CIJ is replaced with c_i

for i from 1 to N
  CIJ(i, i) = 0                          remove connections across the diagonal (loops)
```

The algorithm involves the following steps: Starting with an empty adjacency matrix CIJ, modules at hierarchical level i are added starting with modules at the hierarchical level $h = 0$ (global network). The ratio of matrix elements that represent potential connections within any module A_i depends on the hierarchical level i and the number of sub-modules per module m. Based on A_i and the number of edges at that hierarchical level, E_i, we can determine the probability p_i that any two nodes within a module are connected (edge density within modules). Next, the number of nodes in each module N_i is calculated leading to the total number of modules N_c at that level. Then, each module given as a random graph with edge density p_i is inserted in the adjacency matrix *CIJ*. Finally, edges along the diagonal (loops) are removed from the network. Due to this removal, the total number of edges might be slightly lower than the desired number of edges. In these cases, additional edges are added randomly to the network to generate the desired total number of edges and edge density (not shown in the pseudocode; however, see Matlab routine online).

We explored hierarchical networks with different numbers of hierarchical levels, h (scales), and numbers of sub-modules at each level, m (granularity). A network without hierarchical levels forms a random network, with one level a "flat" modular network, two levels a network with modules and sub-modules, and so on (**Figure 1**).

SPREADING MODEL

A basic spreading model (Newman, 2005) was modified to simulate the propagation of activity through the network. This dynamic model was identical to the one used in Kaiser et al. (2007a).

The simulation operated in discrete time steps, with nodes being in one of two states, active or inactive.

We used a simple threshold model for activity spreading where a number i of randomly selected nodes was activated in the first step. At each subsequent time step, inactive nodes became activated if at least k neighbors were currently active (neighbors of a node are nodes to which direct connections exist). Activated nodes could become inactive with probability v in the next time step, or otherwise stayed active.

An additional parameter was the extent of localization of the initial activation, i_0. For initialization, i ($i \leq i_0$) nodes among the nodes 1 to i_0 were selected randomly and activated in the first time step. The networks nodes were numbered consecutively. For instance, for a network where the largest modules at the highest level contained 100 nodes and where each module contained 10 sub-modules with 10 nodes each, by setting i_0 to 10, 20 or 100, the first sub-module, the first two sub-modules, or the first module, respectively, were activated during initialization. Thus, i determined the number of initially active nodes while i_0 controlled the localization of initial activations, with smaller values resulting in more localized initial activity.

CALCULATING THE AVERAGE RANGE OF LIMITED SUSTAINED ACTIVITY

We systematically explored the network activation resulting from different settings of the initial node activation and localization parameters. Persistent contained activity in hierarchical networks (e.g., intermediate-level trace in **Figure 2A**) existed for a wide range of initial localization and activation parameters (indicated by gray filled circles in **Figure 2B**).

We also explored if the results were robust for variations in the dynamic model parameters k and v, by using a Monte Carlo approach in which, for each pair of k and v, spreading simulations with randomly chosen parameters i (number of initially activated nodes) and i_0 (localization) were tested (**Figure 2B**). A trial was considered to show sustained activity if at least one but at most 50% of the nodes were activated at the end of the simulation (after 200 steps). In our experience, activity did not further die out or spread through the whole network if such an activity level was reached at the end of the simulation. For each pair of spreading parameters k and v, the average proportion of cases for which sustained activity occurred was charted (**Figure 2C**). This proportion is specified by the ratio of gray filled circles relative to all data points in **Figure 2B**. The threshold k ranged from 1 to 9 (step size 2), while the deactivation probability v ranged from 10 to 90% (step size 20%). Therefore, the average ratio over all entries in **Figure 2C** reflected the size of the parameter space for a given network topology which could give rise to LSA, taking into account the initialization parameters i and i_0 as well as the dynamic model parameters v and k.

We tested the proportion of cases with sustained activity for different hierarchical configurations. These configurations varied in the number of hierarchical levels, from 0 for random networks to 4, and the number of sub-modules into which each module was divided for creating the next-lower hierarchical level. For each configuration, different values for the threshold k and the deactivation probability v were tested, in that for each (k, v) pair, 200 runs were performed and the network state was observed after 200 time steps leading to a classification as dying-out, sustained, or spreading activity. For these 200 runs, the number of initially activated nodes i and the localization parameter i_0 was chosen randomly (see range in **Figure 2B**). The average proportion of sustained activity cases for each configuration was plotted as gray-scale value for **Figure 4** and the subsequent figures.

INACCESSIBLE PARAMETER RANGE

For all network sizes, a variety of hierarchical configurations could not be realized, due to the limited network size (regions indicated by horizontal lines in **Figure 3** and subsequent figures). Inadmissible configurations were those where the smallest module at the bottom level would have contained more edges than there were edges possible between module members; that means where $N_c\ (N_c - 1) < E_c$ (N_c: number of nodes; E_c: number of edges in the smallest module). Note that the number of sub-modules per module was varied in steps of two (2, 4, 6, 8,…). Variations by a different step size might have produced more clearly apparent differences in inaccessible configurations between small (300 nodes) and large (11,000 nodes) networks.

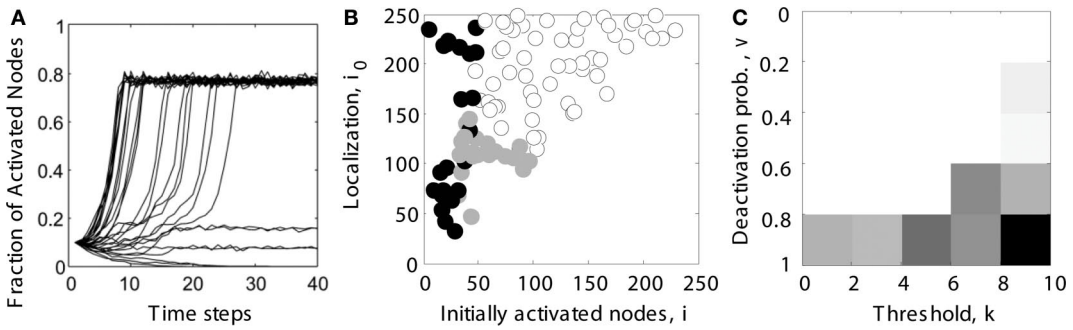


FIGURE 2 | Determining the parameter range of limited sustained activity (schematic overview). (A) For several trials (shown here: 30 runs), it was tested whether activity spread through the whole network (here: activating 80% of all nodes), died out (all nodes becoming inactive), or was sustained at an intermediate level (here: activating 10 or 20% of all nodes). Note that even during complete spreading, not all the nodes were constantly active, due to the inactivation probability v specified in the dynamic model. **(B)** Simulations were run for different combinations of the number of initially activated nodes i and the localization parameter i_0. For each run, the simulated activity died out (●), spread through the whole network (o) or was sustained within a limited compartment of the network (●). **(C)** The parameter space of simulations was further explored for different combinations of deactivation probability v and activation threshold k. Gray levels for each parameter combination in the diagram reflect the percentage of cases giving rise to LSA (from subplot **B**). The average value across all entries was taken as the final measure of the parameter range of LSA for a particular network topology. It reflects the average proportion of limited sustained activation cases obtained across all parameter settings for a given hierarchical modular network.

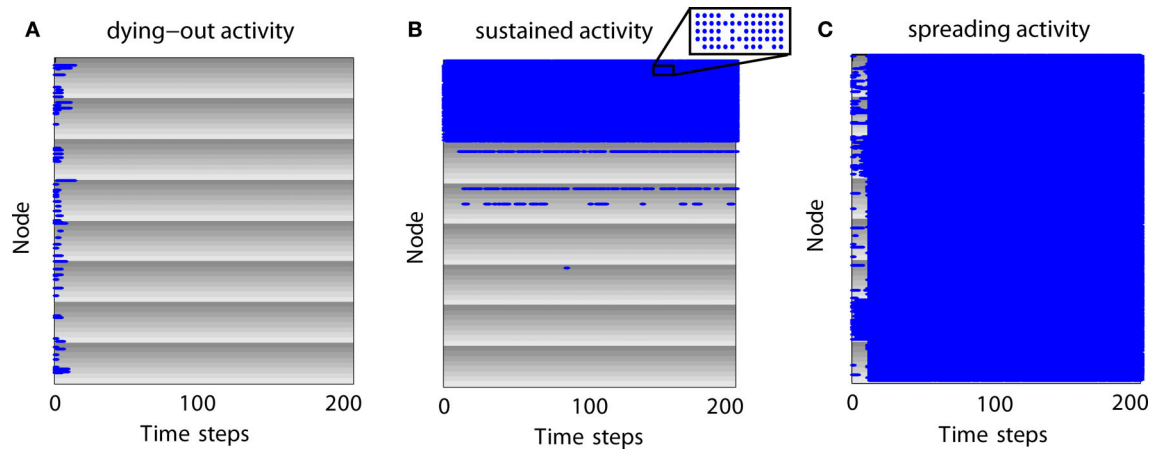


FIGURE 3 | Examples of neural dynamics for different simulation outcomes. Gray shading represents modules and individual gray levels represent different sub-modules. Nodes which are active at a time step are represented as blue dots. **(A)** Expiring (dying-out) activity. **(B)** Limited sustained activity. Although some modules appear completely activated, nodes can be inactive at various time steps due to the inactivation probability (inset). **(C)** Completely spreading activity.

RESULTS

EXPIRING, LIMITED SUSTAINED AND COMPLETELY SPREADING ACTIVITY PATTERNS

How does activity change over time for different parameter settings? In **Figure 3** we give examples for different outcomes in a network with 512 nodes, two hierarchical levels, and eight modules with eight sub-modules per module. Each sub-module contains eight nodes. Modules are represented by gray shading where the individual gray levels represent sub-modules. Blue dots indicate that a node is active at a certain time step.

For expiring activity (**Figure 3A**), initial activity quickly died out as active nodes became de-activated and not enough active neighbors existed to sustain the activity. For LSA (**Figure 3B**), modules and sub-modules became activated indicating that a critical number of neighbors of a node were active and able to (re-) activate a node. For completely spreading activity (**Figure 3C**), activity that was initially contained in one module or several sub-modules managed to spread to other parts of the network and quickly led to complete network activation. This time-course of an early focus of activity with a rapid spread to the whole network may be compared to the generalizations of seizures in epilepsy patients. Note that the blue lines in **Figure 3A** as well as the large blue areas in **Figures 3B,C** also contain nodes which are not active (see inset of **Figure 3B**); however, these nodes are not visible in the figure due to the dot size and image resolution.

TOPOLOGICAL AND SMALL-WORLD PROPERTIES OF HIERARCHICAL NETWORKS

For all tested network sizes, the generated hierarchical networks ($h \geq 1$) possessed characteristics of small-world networks (Watts and Strogatz, 1998), in that the clustering coefficients (the average frequency with which neighbors of a node are directly connected) were much higher than for same-size Erdös–Rényi random networks (Erdös and Rényi, 1960), whereas the characteristic path lengths (the average number of connections on the shortest path between any two nodes) remained comparable to those for random networks of the same size (**Tables 1 and 2**). Note that networks with only one hierarchical level represent the special case of simple modular networks.

Table 1 | Graph and small-world characteristics of hierarchical networks with constant edge density.

N	*E*	*C*	C_{rand}	*L*	L_{rand}	SW
300	1,080	0.025	0.012	23.6	25.0	2.21
512	3,146	0.024	0.012	5.1	4.6	1.80
4,150	206,670	0.023	0.012	2.6	2.5	1.84
11,000	1,452,000	0.023	0.012	2.3	2.2	1.83

The number of edges E for a given number of nodes N was chosen such that the edge density remained 1.2%. Networks with two hierarchical levels and four sub-modules per module are shown (cf. **Figure 1C***). C and C_{rand}: clustering coefficients of the hierarchical and random networks. L and L_{rand}: Characteristic path lengths of the hierarchical and random networks. SW: small-world index $(C/C_{rand})/(L/L_{rand})$.*

Table 2 | Graph and small-world characteristics of hierarchical networks with constant node degree.

N	*E*	*C*	C_{rand}	*L*	L_{rand}	SW
300	15,000	0.227	0.167	1.8	1.8	1.36
512	25,600	0.163	0.098	1.9	1.9	1.66
4,150	207,500	0.023	0.012	2.6	2.5	1.84
11,000	550,000	0.009	0.005	2.8	2.8	1.80

The number of edges E for a given number of nodes N was chosen such that the average number of connections (node degree) was 50. C and C_{rand}: clustering coefficients of the hierarchical and random networks. L and L_{rand}: Characteristic path lengths of the hierarchical and random networks. SW: small-world index $(C/C_{rand})/(L/L_{rand})$.

The characteristic path length for the case of constant edge density (**Table 1**) was particularly high for the 300-node network. This is due to the low edge density of 1.2%; small networks with low edge density exhibit fewer alternative pathways than larger random networks with the same edge density. Therefore, the path length decreases when more edges are added, as for

the larger networks. This behavior resembles the behavior of random networks where the characteristic path length $L \sim \ln N / \ln \langle k \rangle$, where N is the number of nodes and $\langle k \rangle$ is the average node degree (Albert and Barabási, 2002; Costa et al., 2007). All networks, however, show features of small-world networks (Watts and Strogatz, 1998). The clustering coefficient for random networks, C_{rand}, was the same for all network sizes. For random networks, the clustering coefficient is the same as the edge density; that means, the probability that neighbors of a node are connected is the same as the probability that any two nodes are connected. As the edge density is kept constant for all network sizes, C_{rand} remains constant at that value as well. The extent of a small-world organization can be characterized by the small-world coefficient $SW = (C/C_{rand})/(L/L_{rand})$ (Humphries et al., 2006; Humphries and Gurney, 2008). The index SW is around 2 indicating a small-world organization of these networks. Whereas SW is 2.2 for a small network size of 300 nodes, it remains at a lower level of 1.8 for larger networks.

For constant average node degree $\langle k \rangle$ (**Table 2**), the average path length increases with network size. Smaller networks with 300 and 512 nodes show a considerably lower path length compared to constant edge density. Again, all networks displayed features of small-world networks (Watts and Strogatz, 1998). The small-world index SW, however, was lower for small network sizes of 300 and 512 nodes compared with the scenario of constant edge density.

OPTIMAL HIERARCHICAL CONFIGURATIONS FOR LSA IN A SMALL NETWORK

Variation of sustained activity and topological measures

As a first test, we explored the link between hierarchical organization and the parameter range of LSA for different configurations of a network with 512 nodes and, on average, 50 connections per node (**Figure 4A**). The parameter range for LSA tended to increase with the number of sub-modules at each hierarchical level (along rows in **Figure 4A**). The maximum range of LSA occurred for one hierarchical level and the largest possible number of sub-modules per module.

In this and all following plots (**Figures 5 and 6**), regions with horizontal lines indicate hierarchical configurations that cannot be realized, as some modules would need to contain more edges than can be fitted between members of that module. These cases were detected whenever $N_c (N_c - 1) < E_c$ (N_c: number of nodes; E_c: number of edges in the smallest module); that means the number E_c of edges that needed to be established was higher than the number of possible edges in a module, $N_c (N_c - 1)$. Note also that the gray levels indicating proportion of cases were normalized so that white regions represent the minimum and black regions the maximum value for each plot.

Due to the network generation algorithm, modules at the lowest level of the hierarchy had the largest edge density (cf. **Figure 1**). We used this effect to test if LSA patterns were facilitated by more densely connected bottom modules in the network (**Figure 3B**). Interestingly, there existed no clear relation of the density with sustained activation: whereas both maximum edge density and sustained activity probability increased with the number of sub-modules for a network with two hierarchical levels (**Figures 4A,B**), the relation was less clear for larger numbers of hierarchical levels.

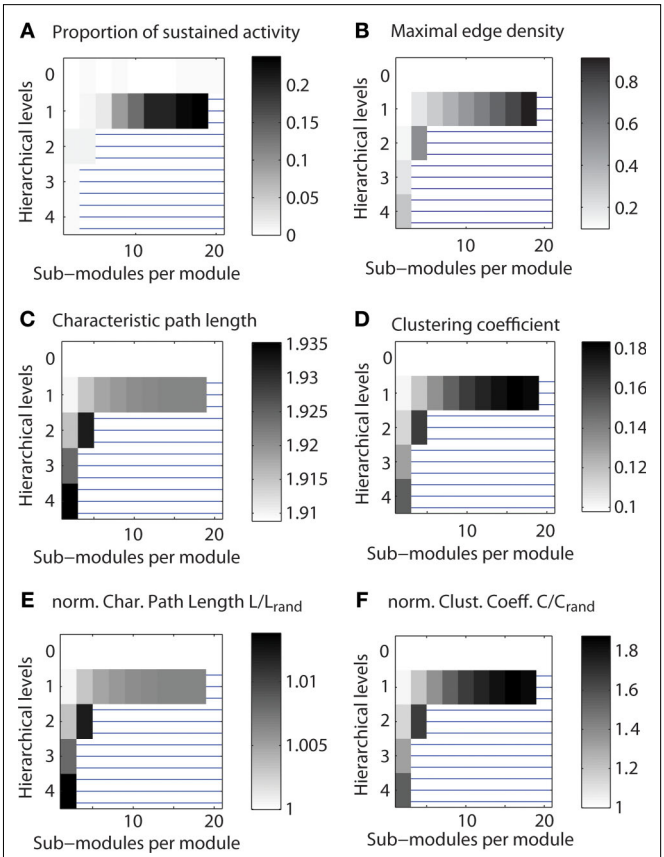


FIGURE 4 | Range of limited sustained activity for different hierarchical configurations of a small network. Shown is the parameter range of limited sustained activation and of topological features for a network with 512 nodes and average node degree of 50. Regions blocked by horizontal lines indicate configurations that were not admissible (see Materials and Methods). Parameters were explored for 1,000 runs of each set of spreading parameters k and v, while the number of initially activated nodes i and the localization parameter i_0 varied for each run. **(A)** Average of the number of parameter settings leading to LSA. **(B)** Maximum edge density based on the most highly connected modules (modules at the lowest level of the respective hierarchy). **(C)** Characteristic path length of the networks. **(D)** Average clustering coefficient of the networks. **(E)** Normalized characteristic path length (divided by the value for Erdös–Rényi random networks). **(F)** Normalized average clustering coefficient (divided by the value for Erdös–Rényi random networks).

How are small-world properties linked to the different hierarchical configurations? The characteristic path length (**Figure 4C**) appeared to show lower values when two or more hierarchical levels existed in the network, but the values were in a narrow range of 1.91–1.92 for one hierarchical level. The clustering coefficient (**Figure 4D**) increased with the number of levels and the number of sub-modules per module. The characteristic path lengths of the hierarchical networks were comparable to those of Erdös–Rényi random networks (**Figure 4E**) whereas the clustering coefficient was higher than in random networks (**Figure 4F**). As the normalized path length is around 1, the SW index has a similar value as the normalized clustering coefficient. Given large SW indices, the networks possessed features of small-world networks (Watts and Strogatz, 1998).

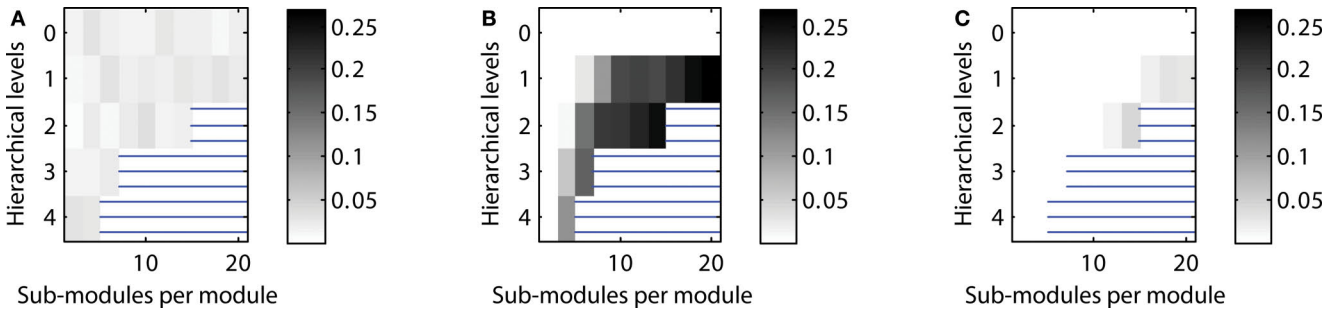


FIGURE 5 | Scaling of optimal configurations with network size for constant global edge density. Proportion of cases showing LSA (averaged over 200 generated networks for each configuration) in **(A)** "rat-size" networks with 300 columns, **(B)** "cat-size" networks with 4,150 columns, **(C)** "macaque-size" networks with 11,000 columns.

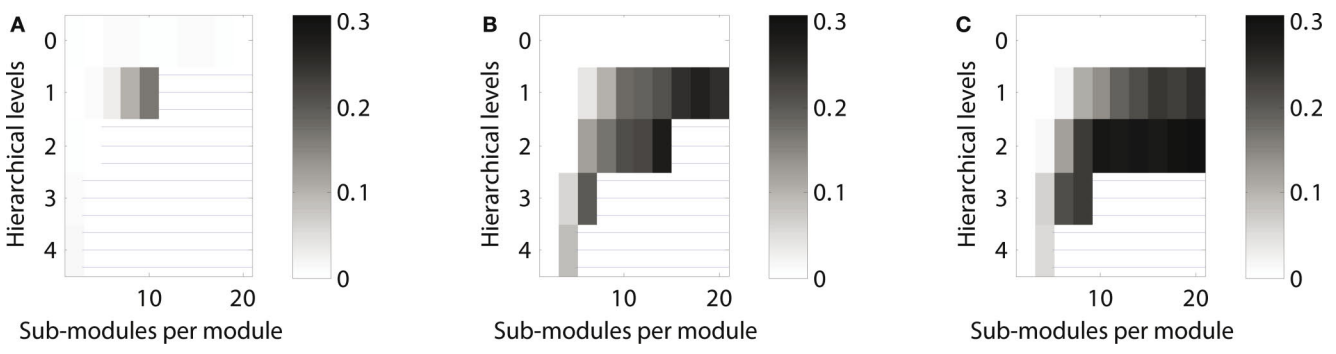


FIGURE 6 | Scaling of optimal configurations with network size for constant average node degree ($\langle k \rangle = 50$). Proportion of cases showing LSA (averaged over 200 networks generated for each configuration) in **(A)** "rat-size" networks with 300 columns, **(B)** "cat-size" networks with 4,150 columns, and **(C)** "macaque-size" networks with 11,000 columns.

Control calculations

We tested several parameters that were used for generating hierarchical networks. Networks consisted of 512 nodes and, on average, 50 connections per node. The Appendix contains a full description of these control calculations including additional figures.

Varying the number of edges for different hierarchical levels. By default, the number of edges for each hierarchical level was set to be equal, that means, E_i was the same for each hierarchical level i. Here, we tested sustained activity patterns for varying numbers of edges per level. We considered two cases: (a) a decrease of the number of edges with each hierarchical level or (b) an increase of the number of edges with each hierarchical level. The absolute level of sustained activity was lower in case (a) and higher in case (b) compared to the original setting (cf. **Figure 7**).

Varying the parcellation for different hierarchical levels. By default, we used the same parcellation of modules at each level; that means if a module consisted of two sub-modules for the highest level, this condition would be the same for all other levels of the network hierarchy as well. Here, we also tested varying the parcellation into sub-modules depending on the hierarchical level. Again, we tested two cases: (a) creating fewer sub-modules for higher hierarchical levels ("decreasing parcellation") or (b) creating more sub-modules for higher hierarchical levels ("increasing parcellation"). Configurations with a high proportion of LSA cases for decreased as well as increased numbers of sub-modules for each hierarchical level remained comparable with the original calculation. However, for the "increasing parcellation" type, the overall proportion of LSA cases increased, extending the maximum probability of sustained activity from 0.23 to 0.42 (cf. **Figure 8**).

Number of cases close to 50% activation threshold for classification as sustained. In additional simulations, we tested how close the final activity was to the threshold used for classification as a case of LSA. Indeed, final activity levels close to the 50% threshold could occur. However, final activity levels were around 10–20% for most configurations producing a high number of LSA cases. This indicates that configurations leading to a high proportion of LSA cases were not substantially affected by the threshold (cf. **Figure 9**).

Topologies leading to expiring, limited sustained, and completely spreading activity. As a default, we investigated the distribution of LSA cases depending on the hierarchical network organization. In additional simulations, we also explored the distribution of the other two possible simulation outcomes: activity dying out before

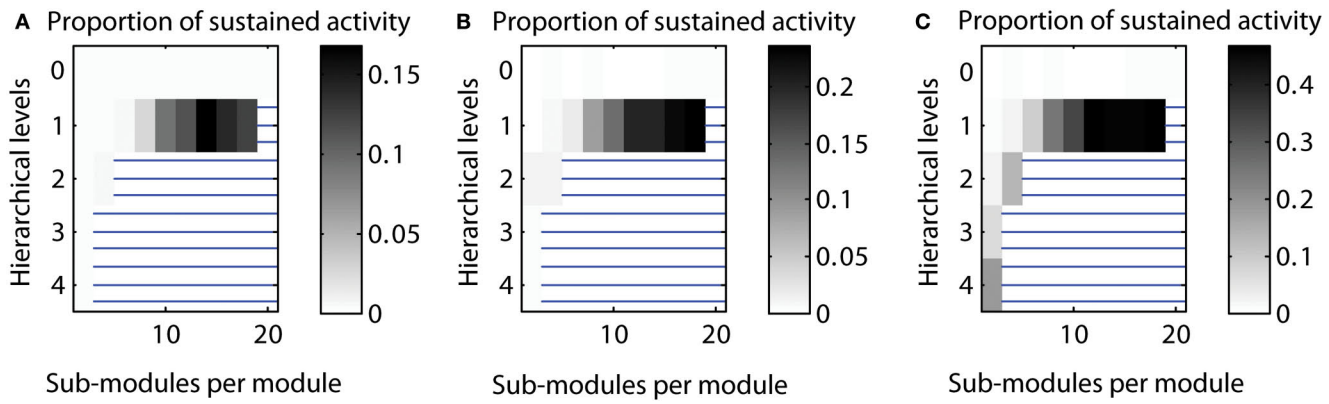


FIGURE 7 | Varying the number of edges per hierarchical level. (A) Decreasing number of edges for higher hierarchical levels [$E_i \sim (2/3)^i$]. **(B)** Number of edges independent from hierarchical level (E_i = const.). **(C)** Increasing number of edges for higher hierarchical levels [$E_i \sim (3/2)^i$].

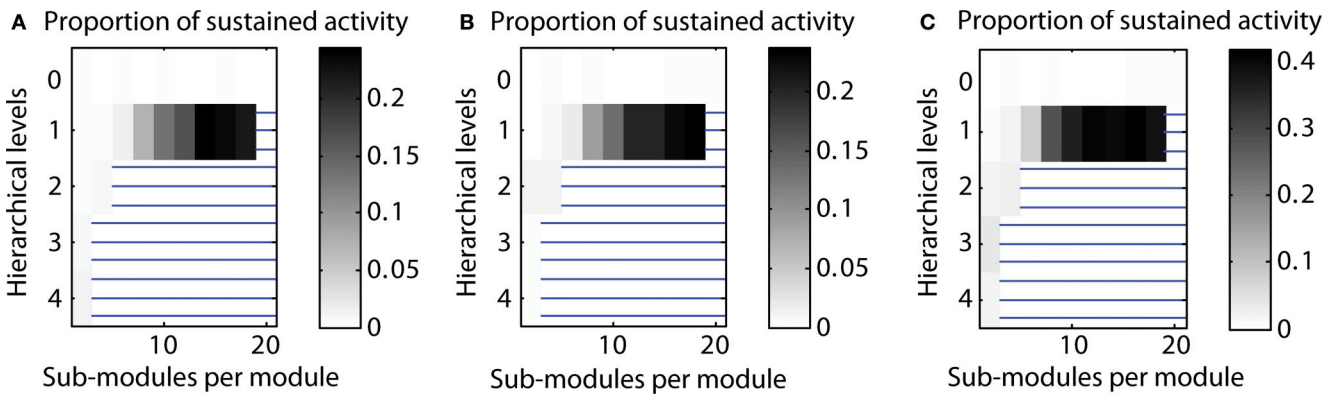


FIGURE 8 | Varying the parcellation (number of sub-modules per module) for hierarchical levels. (A) Decreasing number of sub-modules m_i for higher hierarchical levels ($m_i \sim 0.9^i$). **(B)** Parcellation into sub-modules independent from hierarchical level ($m_i = m$ = const.; see main text). **(C)** Increasing number of sub-modules m_i for higher hierarchical levels ($m_i \sim 1.1^i$).

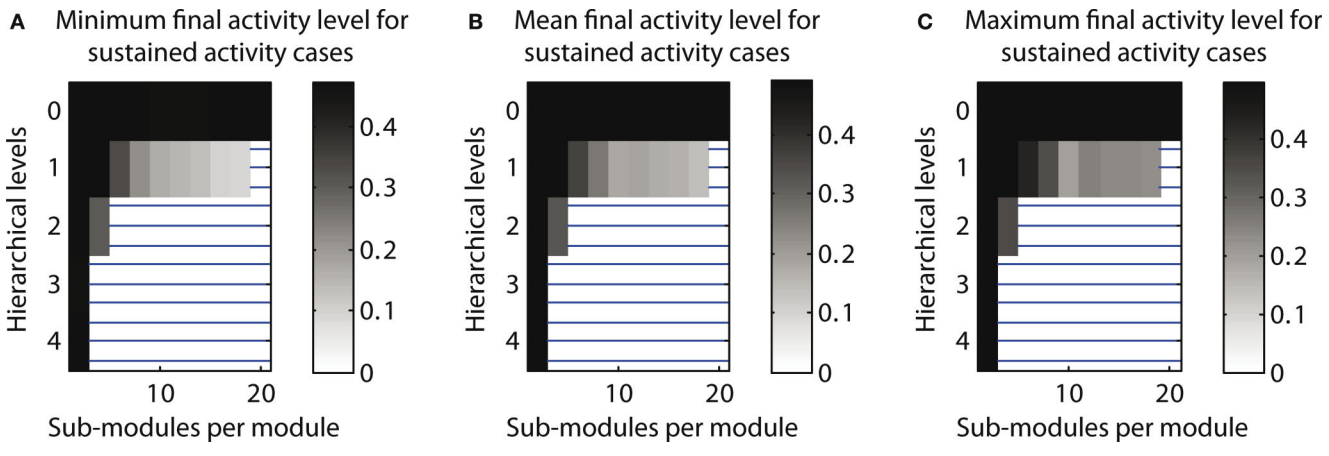


FIGURE 9 | Final activity for *n* = 20 runs classified as sustained activity. (A) Minimal final activity level. **(B)** Average final activity level. **(C)** Maximum final activity level.

the end of the simulation and final activation of more than 50% of the nodes, which was classified as complete spreading. For a random network (zero hierarchical levels), both dying-out and complete spreading occurred in 50% of the cases. However, when more than one hierarchical level was present, complete spreading activity occurred more often than dying-out. For one hierarchical level, outcomes depended strongly on the number of sub-modules per module. The cases with expiring activity formed 50% of the cases for two sub-modules, but decreased with the number of sub-modules (cf. **Figure 10**).

Varying the edge density. The pattern of sustained activity remained comparable to the default settings even when the edge density differed from the original value of 10% for a network with 512 nodes and an average node degree of 50. The maximum proportion of cases with LSA varied between 0.172 for decreased edge density (5%) to 0.238 for increased edge density (20%). The relative distribution of case for networks with one hierarchical level was similar across edge densities (cf. **Figure 11**).

SCALING OF OPTIMAL HIERARCHICAL CONFIGURATIONS FOR LSA WITH NETWORK SIZE

Two different scaling scenarios were explored. In the first one, the global edge density of the networks was kept constant (at 1.2%) while the average number of connections per node varied; in the second scenario, the average node degree was kept constant (at 50 connections per node) while the networks' edge density varied.

Constant edge density

In the first approach, the probability that any two nodes (representing cortical columns) in the network were connected was, on average, 1.2%. Connection density was larger within modules and lower between modules; however the global average remained constant, independent of the hierarchical configuration.

Given a constant setting for testing neural activation across network sizes, the "rat-size" network showed sustained activity for a wide variety of hierarchical configurations (**Figure 5A**). Surprisingly, for the larger "cat-size" network, a smaller variety of hierarchies existed that could generate LSA (**Figure 5B**). However,

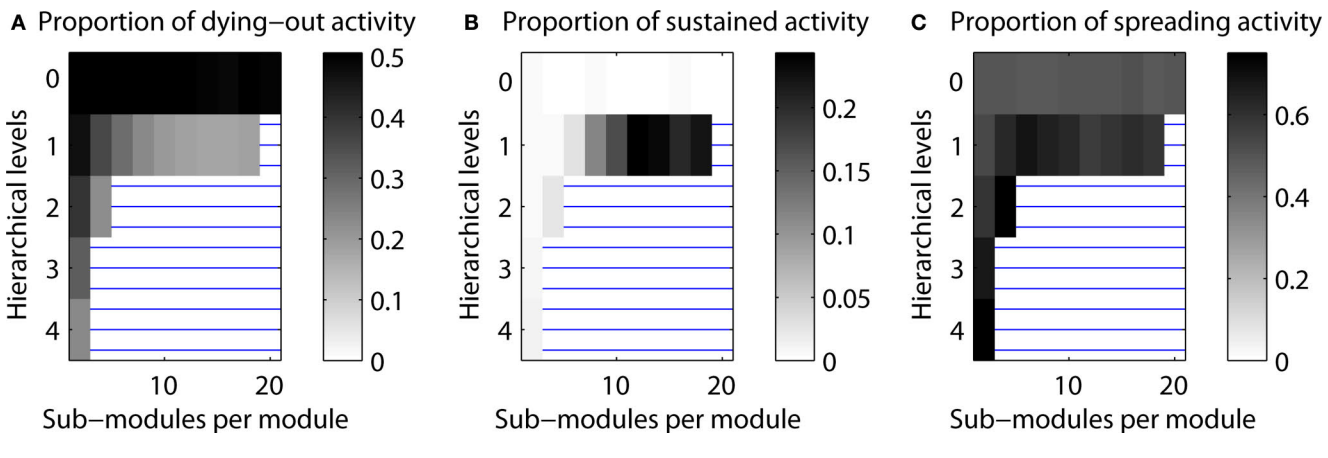


FIGURE 10 | Proportion of cases resulting in one of three scenarios of final activity level. (A) Activity dying out. **(B)** Limited sustained activity. **(C)** Activity spreading through the network (above 50% activation threshold).

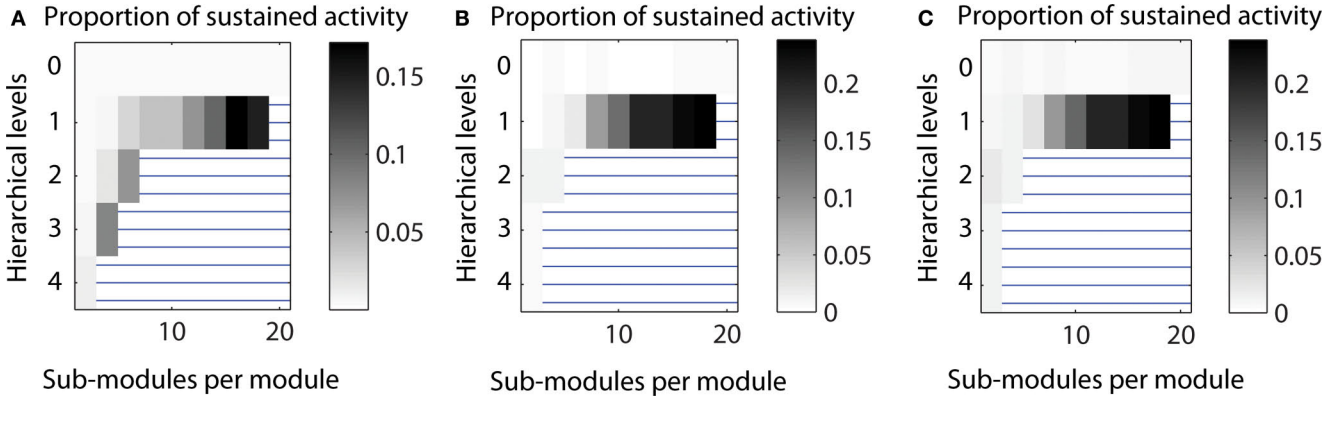


FIGURE 11 | Varying the edge density in a network with 512 nodes. (A) Decreased edge density d = 5%, average node degree $\langle k \rangle$ = 25. **(B)** Original edge density d = 10%, average node degree $\langle k \rangle$ = 50. **(C)** Increased edge density d = 20%, average node degree $\langle k \rangle$ = 100.

for these configurations, sustained activity occurred in up to 25% of the explored parameter settings, whereas it occurred only in up to 3% of the tested settings for the rat-like network (cf. scale of activation range). For the even larger "macaque-size" network, the maximum range of LSA (up to 4% of tested parameter settings) was as low as for the "rat-size" network (**Figure 5C**), while overall, the variety of hierarchical configurations that resulted in LSA was also lowest for the macaque-like network.

These results indicate that the number of possible hierarchical configurations (resulting from combinations of the number of levels and number of sub-modules) leading to LSA decreased with increasing network size. Only a few hierarchical configurations appeared suitable for producing LSA in all network sizes. Such configurations typically combined an intermediate number of levels with a large number of sub-modules (**Figures 4C and 5B**). Interestingly, the combination of a large number of hierarchical levels with a small number of sub-modules proved ineffective for supporting LSA in the larger-size network (**Figure 5C**). For large networks the best strategy for achieving sustained activity was provided by an arrangement of two hierarchical levels containing the largest possible number of modules and sub-modules.

Constant average node degree

The results obtained under the constraint of a constant edge density (see section Constant edge density) suggested that configurations for LSA were harder to attain in large as well as small networks. Only networks of an intermediate size appeared to result in a large variety of hierarchical networks possessing a wide parameter range for LSA. In an alternative approach, we also tested optimal configurations of networks of different sizes under the constraint that the average number of connections per column, rather than the probability that any two columns are connected (edge density), was kept constant. For this approach, the number of edges was set to 50 times the number of nodes, leading to an average node degree of 50 in all networks, albeit with variation for individual nodes.

Under these conditions, LSA in the "rat-size" networks arose mostly in networks with one hierarchical level, and for an increasing number of sub-modules (**Figure 6A**). Both the "cat-size" and the "macaque-size" networks possessed a similar, large range of hierarchical configurations showing sustained activity (**Figures 6B,C**). All networks demonstrated that cases of LSA increased with the number of sub-modules per module. Whereas the range of hierarchical configurations differed between the "rat-size" and the "cat-size" or "macaque-size" networks, the maximum range of sustained activity was comparable for all sizes with 15–30%. A constant number of connections per node, therefore, permitted a wide range of optimal hierarchical configurations for LSA even if the network size increased.

DISCUSSION

This study investigated an essential precondition of criticality in neural systems, the capability of neural networks to produce LSA patterns following an initial activation. We addressed this question by simulating the spreading of neural activity and systematically varying model parameters and network topology in hierarchical modular networks, which are inspired by the organization of biological neural networks across scales. Our previous study (Kaiser et al., 2007a) demonstrated that hierarchical cluster networks possess a large parameter range leading to LSA, in contrast to random and non-hierarchical small-world networks. Here we expanded this analysis by varying the number of levels and sub-modules in hierarchical networks and scaling their size within two alternative scenarios, constant edge density or constant average node degree. This study demonstrated, first, that LSA patterns are supported by a variety of parameter settings for hierarchical modular networks, combining different numbers of hierarchical levels with varying numbers of sub-modules per level; second, that for the same network size and the same number of sub-modules, networks with a larger number of levels resulted in a wider range of LSA, while for the same number of hierarchical levels a larger activity parameter range was produced by increasing the number of sub-modules; and third, that a high level of sustained activity was attainable across network sizes for a constant average node degree, but not for constant edge density.

The present results provide a proof of concept for three points. First, hierarchical network configurations lead to different levels of sustained activity independent of global topological properties, such as characteristic path length or clustering coefficient. Therefore, the identification of an optimal network configuration associated with a maximum level of LSA is a suitable target for evolutionary graph optimization. Second, only specific hierarchical network arrangements can be realized for a limited network size. Even for the human brain with an estimated number of 125,000 columns per hemisphere (Jones and Peters, 1984) under the assumptions made in section "Anatomical constraints", only a small fraction of potential hierarchies can be realized within the current framework. For the "human-size" network, three hierarchical levels with up to 18 sub-modules and four levels with up to 8 sub-modules are possible. These limits are beyond the ones of the "macaque-size" network which maximally allowed six and four sub-modules for three and four hierarchical levels, respectively. Therefore, simple combinatorics suggest that it is easier to vary the number of modules at each level than to increase the number of levels for larger brains. Third, the number of configurations which lead to sustained activity decreases with network size if the edge density remains constant, but remains large even for large network sizes if the average number of connections per node is kept constant. This model finding corresponds to the observation from comparative studies that the number of connections of a neural node (e.g., the number of synapses of a neuron) rather than the ratio of connections (e.g., being connected to 10% of all neighbors) largely remains constant across species with different brain sizes (Ringo, 1991; Schüz and Demianenko, 1995; Zhang and Sejnowski, 2000; Striedter, 2004). Moreover, for constant average node degree, the optimal configurations in larger networks tended to possess more hierarchical levels, suggesting a beneficial contribution of the more intricately structured topology in larger neural networks to dynamical stability. There are indications from compilations of biological neural connectivity (e.g., www.cocomac.org) that support this model finding. For example, cluster analyses suggest that primate cortical connectivity is structured on more levels than connectivity in smaller brains, if one considers that there exist primate visual "streams" (Young, 1992; Hilgetag et al., 2000), that is, subdivisions of the visual network module that are apparently absent

in the rat (Burns and Young, 2000) or cat network (Scannell et al., 1999; Zamora-López et al., 2010) Moreover, there are generally more modules in larger brains, if cortical areas can be considered as modules.

The finding of increased hierarchical structure in larger networks may appear counterintuitive given that there are limits on the number of hierarchical levels even in large networks, as discussed above. However, an appropriately large number of levels may be a necessary constraint for sustaining activity. If the number of modules in a large network was increased without increasing the number of levels, then, in principle, it would be easy to activate each module. However, activation of the global network may be prevented by dispersion of the activity across the entire network, which means that there may not exist enough projections into each of the individual modules to activate them. Similarly, if there are few large modules, activity may be dispersed within the modules. In order to establish a balance between the number and size of modules in large networks, additional levels need to be created, as confirmed by the modeling results.

The hierarchical network topology we explored reflects the distributed multi-level modularity that is considered a central feature of biological neural networks. Neural networks show strong modularity across many levels of scale, ranging from cellular neuronal circuits and neural populations organized in cortical columns (Mountcastle, 1997) to communities of closely linked areas at the systems level (Hilgetag et al., 2000; Breakspear and Stam, 2005). Smaller modules are nested within larger ones, such as columns within an area, which itself is a module in a large-scale brain division, such as the visual system. Another important feature of complex networks that has been discussed widely is the existence of hub, that is, nodes possessing a significantly larger number of links than the majority of nodes in a network (Albert and Barabási, 2002; Ravasz et al., 2002). However, it is difficult to identify nodes in the brain that integrate modules across scales (with the potential exception of unspecific neuromodulatory systems, such as the serotonergic system) and act as global hubs. While there are hub-like nodes in neural networks (Kaiser et al., 2007b; Sporns et al., 2007), they may not act globally, such that most projections in the network originate from, or converge on, a central node. This topology is different from "centralistic" networks where most nodes are linked to hubs (Ravasz et al., 2002) and which may be more suitable for representing large-scale biochemical networks. However, the detailed investigation of biological neural topologies needs to be continued, since modeling studies have shown a strong impact of topology on network dynamics. For instance, networks which contain hubs may support different modes of activity propagation than hub-less modular networks (Müller-Linow et al., 2008; Hütt and Lesne, 2009).

The present study was set up under several simplifying assumptions, in order to provide general insights into the relationship between hierarchical neural topology and activation patterns. This approach resulted in a number of model limitations. First, nodes representing columns were assumed to be uniform building blocks, whereas actual column organization (layer structure and number of neurons) in the brain might differ across regions (Hutsler et al., 2005) as well as species (Herculano-Houzel et al., 2008), potentially leading to additional constraints on hierarchical network organization. Moreover, the internal organization of columns (self-loops) was not explicitly part of the modeled networks. However, it was represented through the node activation rule: an active node could remain active for the following time step given a sufficient number of active neighbors and potential collaterals going back to the neuron itself. For each time step, the deactivation probability determined whether an active node became inactive. The strength of self-loops was therefore implicitly represented in this deactivation probability with lower likelihood of deactivation for more frequent self-loops. Second, specifically organized populations of inhibitory neurons within columns might additionally influence global network dynamics. Thus, future models could incorporate more detailed biologically realistic mechanisms for reducing activity at the neuronal level, instead of the presently employed phenomenological deactivation probability. Third, the parcellation of modules into sub-modules for each level was treated as symmetric, that means, when a module is split into sub-modules, each module has the same size. It will be important to test asymmetric parcellation of a module into smaller and larger sub-modules in future studies. Finally, the model considered neural network behavior in the absence of external inputs, except for the initial activation. Therefore, the current findings may particularly apply to situations where there is limited external input to the brain, such as during sleep or early development. The results also relate to the organization of neural dynamics associated with the "default mode" or "resting state" of the brain (e.g., Raichle and Snyder, 2007). The role of external inputs should be addressed in future studies, which could also investigate if there is a difference in the processing of external stimuli by networks that are optimal for LSA and those that are not. In addition, the edge density at the lowest level (**Figure 4B**) could in some cases be higher than 50% which is unlikely in biological neuronal networks.

In this study, we varied the network size to represent networks of columns of a hemisphere in a "rat-size" (300 nodes), "cat-size" (4,150 nodes), and "macaque-size" brain (11,000 nodes). For large networks and constant edge density, two hierarchical levels with the maximum possible number of sub-modules per module appeared to provide the best strategy for achieving sustained activity (cf. **Figure 5**). These multi-level configurations often resulted in a high density of connectivity within modules at the lowest level. Such high edge densities are theoretically possible, but only realized to some extent in biological neural systems. At the global level of human fiber tract connectivity between brain regions, for example, edge densities around 46% can be reached (Honey et al., 2007). Within columns, the connection frequency between any two neurons is around 16% (Douglas and Martin, 2007) but around 35% for neurons from the same cell lineage (Yu et al., 2009). Given constant edge density, the range of feasible hierarchical configurations – that is, the degrees of freedom for evolving neural network architectures – appeared to decrease with larger network sizes. This was due to the fact that the number of neighbors of a node increased with network size. Since the probability for connecting a node to other nodes (the edge density) remained constant, nodes were connected to a larger number of nodes when the number of potential neighbors increased with network size. The larger number of

connected neighbors in larger networks also meant that the (absolute) threshold for activating a node was more easily reached. In such cases, activity was harder to contain and more likely spread through the whole network.

Using the constraint of a constant average degree, on the other hand, enabled a wider range of hierarchical configurations with up to 30% sustained activity cases across network sizes (cf. **Figure 6**). Such scalability with network size might be beneficial both for ontogenetic and phylogenetic development. Using a constant number of connections per node, rather than a constant edge ratio, across species appears to have several benefits (Changizi, 2001). First, reducing edge density is necessary due to the limited volume available for white matter fiber tracts (Ebbesson, 1980; Karbowski, 2001; Striedter and Northcutt, 2006). For constant edge density, a brain with two times as many columns would contain four times as many connections, quickly increasing brain volume. Second, the setting of a constant number of connections per node provided a setup for sustained activity in different brain network sizes. This might mean that sustained activity can occur for different brain sizes during evolution, if they are appropriately hierarchically structured. Third, hierarchical structuring may also provide the functional stability of LSA in the developing brain. At early stages of ontogenetic development, neural networks generally have few modules and few nodes. During development, more modules, nodes, and hierarchical levels are established (Robinson et al., 2009). Therefore, sustained activity can occur continuously through different stages of development and brain network growth. However, the hierarchical organization is not the only mechanism that can sustain activity during development; early neuronal mechanisms include, for example, spontaneous activity such as retinal waves (Sernagor et al., 2006; Hennig et al., 2009) or the early excitatory function of gamma aminobutyric acid (GABA).

Finally, whereas several earlier studies have explored spatial (e.g., brain volume) or topological (e.g., characteristic path length, Kaiser and Hilgetag, 2006) constraints on brain organization, the present study focused on dynamic constraints, specifically the necessity of brain dynamics to subsist at a sustained yet limited level of activity. Alternative or additional dynamic constraints that may be relevant for this phenomenon could be synchronous activity (König et al., 1995; von der Malsburg, 1995; Masuda and Aihara, 2004), or functional attributes such as multi-modal integration, functional complexity (Sporns et al., 2000), information propagation, or processing speed. An accessible parameter range for sustained limited activity is a necessary condition for criticality, but does not in itself guarantee it. Criticality has been interpreted as an abolishing of length scales, that is, the coexistence of dynamical processes at all scales. We saw examples for this phenomenon in activation patterns at LSA where modules and sub-modules of different sizes were activated together. Non-LSA conditions, by contrast, produced only the trivial states of activating all or none of the network nodes. It will be particularly interesting to see how networks optimized with respect to functional diversity are related to networks having optimal range for LSA. A possible link between these two properties was suggested by an earlier analysis showing that the number of significantly repeating activation patterns is maximized at the critical point (Haldeman and Beggs, 2005).

APPENDIX

CONTROL EXPERIMENTS

We tested several additional simulation parameters in the following sections. As in section "Optimal hierarchical configurations for LSA in a small network", networks consisted of 512 nodes and 50 connections per node.

Varying the number of edges for different hierarchical levels

In the default settings, the number of edges E_i was set to be the same for each hierarchical level i. Additionally, we tested sustained activity patterns where the number varied. We considered two cases: (a) a decrease of the number of edges with each hierarchical level or (b) an increase of the number of edges with each hierarchical level. The change followed a function where the number of E_i at level i was given by $E_i = s^i E_c/C$, where s is a scaling factor of 2/3 for decreased and 3/2 for increased number of edges per level. The parameter $E_c = E/L$ is the number of edges in a network with E edges and L levels, which was used for the original calculation and $C = L\ (s^{L+1} - s)/(s - 1)$ is a normalization factor to ensure that the total number of edges remains E.

As shown in **Figure 7**, configurations with a high proportion of LSA for decreased as well as increased numbers of edges per hierarchical level remained comparable with the original calculation. The absolute level of sustained activity was lower in case (a) and higher in case (b) compared to the original settings. In addition, sustained activity cases also occurred for two or more hierarchical levels when the number of edges was increased (**Figure 7C**).

Varying the parcellation for different hierarchical levels

In the default settings, we used the same split-up of modules for each level; that means, if a module consisted of two sub-modules at the highest network level, this condition would be the same for all other levels of the network hierarchy as well. Here, we tested varying the parcellation into sub-modules depending on the hierarchical level. Again, we tested two cases: (a) creating fewer sub-modules for higher hierarchical levels ("decreasing parcellation") or (b) creating more sub-modules for higher hierarchical levels ("increasing parcellation"). The change followed a function where the number of parcellations, sub-modules per module, m_i at level i was given by $m_i = s^i m$ where s is a scaling factor of 0.9 for decreased and 1.1 for increased parcellation. The parameter m was the same as for the original calculation.

As can be seen from **Figure 8**, configurations with a high probability of sustained activity for decreased as well as increased number of sub-modules for each hierarchical level remained comparable with the original calculations. However, whereas the absolute proportions for decreased parcellation were similar, for increased parcellation the overall probabilities increased, extending the maximum proportion of sustained activity cases from 0.23 to 0.42. Therefore, the absolute parameter range of sustained activity was the same for case (a) and almost twice as high for case (b), compared to the original setting.

Note that both an increased parcellation and a larger number of edges led to a higher edge density of modules at the highest (fine-grained) level (maximum edge density, **Figure 3B**). This could explain why the ratio of sustained activity cases was higher for these configurations.

Number of cases close to 50% activation threshold for classification as sustained

In these simulations, we tested how close the final activity after 200 time steps came to the threshold used for classification as sustained activity case.

As seen in **Figure 9**, levels close to the 50% threshold did occur. However, final activity levels were around 10–20% in all cases for which a high number of sustained activity cases was reported. This observation indicates that configurations with high proportions of sustained activity cases were not affected by the threshold. On the other hand, configurations with a high proportion of cases with complete spreading only had few cases of sustained activity. Due to higher activity levels for such configurations, sustained activity cases were close to the 50% threshold for classifying sustained activity (black regions in **Figure 9**).

Topologies leading to expiring, limited sustained and completely spreading activity patterns

The main simulations of this project investigated the proportion of LSA cases depending on hierarchical network organization. Here, we also considered the distribution of the other two possible simulation outcomes: activity dying out before the end of the simulation and final activity in more than 50% of the nodes, which was classified as complete spreading. **Figure 10** shows the dependence of all three outcomes on hierarchical network organization.

For configurations resulting in a small number of LSA cases (white regions in **Figure 10B**), both dying-out and complete spreading occurred in 50% of the cases (note the different gray level setting due to re-scaling). However, when more than one hierarchical level was present, complete spreading occurred more often than dying-out. For one hierarchical level, outcomes depended strongly on the number of sub-modules per module. The cases in which activity expired formed 50% of cases for two sub-modules, but decreased with the number of sub-modules. The maximum proportion of complete spreading, around 70% of the cases, occurred for 4–10 sub-modules per module. The maximum values for LSA occurred for 12–20 sub-modules per module.

Varying the edge density

In the main simulations, the edge density for a network with $N = 512$ nodes and an average node degree $\langle k \rangle$ of 50 was 0.098, that means, around 10%. Both parameters, edge density and average node degree are related: the edge density in a directed network is given by $d = E/[N\ (N-1)]$ whereas the average node degree is $\langle k \rangle = E/N$, meaning that $d = \langle k \rangle/(N-1)$. How does variation of edge density for the same network size influence the range of LSA? To answer this question, we compared edge densities which were half (5%) of or twice (20%) of the value of the original calculations.

As shown in **Figure 11**, the pattern of sustained activity remained comparable to the original calculations even when the edge density varied. The maximum level of LSA varied between 0.172 (for decreased edge density) to 0.238 (for increased edge density; original maximum level: 0.238). The relative distribution for one hierarchical level was similar across different edge densities. For two and three hierarchical levels, two additional configurations with high levels of sustained activity occurred for decreased edge density. These additional configurations were impossible for higher numbers of edges to be realized when edge densities were around 10% or above.

ACKNOWLEDGMENTS

Marcus Kaiser was supported by WCU program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R32-10142), the Royal Society (RG/2006/R2) and the CARMEN e-science project (www.carmen.org.uk) funded by EPSRC (EP/E002331/1).

REFERENCES

Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.

Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of the 1/f noise. *Phys. Rev. Lett.* 59, 381–384.

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177.

Bertschinger, N., and Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural. Comput.* 16, 1413–1436.

Binzegger, T., Douglas, R. J., and Martin, K. A. C. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* 24, 8441–8453.

Breakspear, M., and Stam, C. J. (2005). Dynamics of a neural system with a multiscale architecture. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360, 1051–1074.

Burns, G. A. P. C., and Young, M. P. (2000). Analysis of the connectional organization of neural systems associated with the hippocampus in rats. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 55–70.

Changizi, M. A. (2001). Principles underlying mammalian neocortical scaling. *Biol. Cybern.* 84, 207–215.

Changizi, M. A., and Shimojo, S. (2005). Parcellation and area-area connectivity as a function of neocortex size. *Brain Behav. Evol.* 66, 88–98.

Costa, L. F., Rodrigues, F. A., Travieso, G., and Villas Boas, P. R. (2007). Characterization of complex networks: a survey of measurements. *Adv. Phys.* 56, 167–242.

Douglas, R. J., and Martin, K. A. C. (2007). Mapping the matrix: the ways of neocortex. *Neuron* 56, 226–238.

Ebbesson, S. O. E. (1980). The parcellation theory and its relation to interspecific variability in brain organization, evolutionary and ontogenetic development, and neuronal plasticity. *Cell Tissue Res.* 213, 179–212.

Erdös, P., and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17–61.

Felleman, D. J., and van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Freeman, W. J., Rogers, L. J., Holmes, M. D., and Silbergeld, D. L. (2000). Spatial spectral analysis of human electrocorticograms including the alpha and gamma bands. *J. Neurosci. Methods* 95, 111–121.

Haldeman, C., and Beggs, J. M. (2005). Critical branching captures activity in living neural networks and maximizes the number of metastable States. *Phys. Rev. Lett.* 94, 058101.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol. Cybern.* 82, 111–121.

Hennig, M. H., Adams, C., Willshaw, D., and Sernagor, E. (2009). Early-stage waves in the retinal network emerge close to a critical state transition between local and global functional connectivity. *J. Neurosci.* 29, 1077–1086.

Herculano-Houzel, S., Collins, C. E., Wong, P., Kaas, J. H., and Lent, R. (2008). The basic nonuniformity of the cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12593–12598.

Hilgetag, C. C., Burns, G. A. P. C., O'Neill, M. A., Scannell, J. W., and Young, M. P. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 91–110.

Honey, C. J., Koetter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10240–10245.

Humphries, M. D., and Gurney, K. (2008). Network 'Small-World-Ness': a quantitative method for determining canonical network equivalence. *PLoS ONE* 3, e0002051. doi: 10.1371/journal.pone.0002051.

Humphries, M. D., Gurney, K., and Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not scale-free, network. *Proc. Biol. Sci.* 273, 503–511.

Hutsler, J. J., Lee, D.-G., and Porter, K. K. (2005). Comparative analysis of cortical layering and supragranular layer enlargement in rodent carnivore and primate species. *Brain Res.* 1052, 71–81.

Hütt, M., and Lesne, A. (2009). Interplay between topology and dynamics in excitation patterns on hierarchical graphs. *Front. Neuroinform.* 3:28. doi: 10.3389/neuro.11.028.2009.

Jones, E. G., and Peters, A. (eds) (1984). *Cerebral Cortex*: *Functional Properties of Cortical Cells*, *Volume 2 of Cerebral Cortex*. New York: Plenum Press.

Kaiser, M. (2007). Brain architecture: a design for natural computation. *Philos. Trans. R. Soc. Lond., A* 365, 3033–3045.

Kaiser, M., Görner, M., and Hilgetag, C. C. (2007a). Functional criticality in clustered networks without inhibition. *New J. Phys.* 9, 110.

Kaiser, M., Martin, R., Andras, P., and Young, M. P. (2007b). Simulation of Robustness against Lesions of Cortical Networks. *Eur. J. Neurosci.* 25, 3185–3192.

Kaiser, M., and Hilgetag, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* 2, e95. doi: 10.1371/journal.pcbi.0020095.

Karbowski, J. (2001). Optimal wiring principle and plateaus in the degree of separation for cortical neurons. *Phys. Rev. Lett.* 86, 3674–3677.

Kinouchi, O., and Copelli, M. (2006). Optimal dynamical range of excitable networks at criticality. *Nat. Phys.* 2, 348–351.

Kitzbichler, M. G., Smith, M. L., Christensen, S. R., and Bullmore, E. (2009). Broadband criticality of human brain network synchronization. *PLoS Comput. Biol.* 5, e1000314. doi: 10.1371/journal.pcbi.1000314.

König, P., Engel, A. K., Roelfsema, P. R., and Singer, W. (1995). How precise is neuronal synchronization? *Neural Comput.* 7, 469–485.

Kötter, R., and Sommer, F. T. (2000). Global relationship between anatomical connectivity and activity propagation in the cerebral cortex. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 355, 127–134.

Latham, P. E., and Nirenberg, S. (2004). Computing and stability in cortical networks. *Neural Comput.* 16, 1385–1412.

Masuda, N., and Aihara, K. (2004). Global and local synchrony of coupled neurons in small-world networks. *Biol. Cybern.* 90, 302–309.

Milton, J., and Jung, P. (2003). *Epilepsy as a Dynamic Disease*. Berlin: Springer.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain* 120(Pt. 4), 701–722.

Müller-Linow, M., Hilgetag, C. C., and Hutt, M. T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS Comput. Biol.* 4, e1000190. doi: 10.1371/journal.pcbi.1000190.

Muresan, R. C., and Savin, C. (2007). Resonance or integration? Self-sustained dynamics and excitability of neural microcircuits. *J. Neurophysiol.* 97, 1911–1930.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* 46, 323–351.

Nieuwenhuys, R., Donkelaar, H. J., and Nicholson, C. (1998). *The Central Nervous System of Vertebrates*. Berlin: Springer.

Pastor-Satorras, R., and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200.

Raichle, M. E., and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage* 37, 1083–1090.

Rakic, P. (2008). Confusing cortical columns. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12099–12100.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555.

Ringo, J. L. (1991). Neuronal interconnection as a function of brain size. *Brain Behav. Evol.* 38, 1–6.

Robinson, P. A., Henderson, J. A., Matar, E., Riley, P., and Gray, R. T. (2009). Dynamical reconnection and stability constraints on cortical network architecture. *Phys. Rev. Lett.* 103, 108104.

Roxin, A., Riecke, H., and Solla, S. A. (2004). Self-sustained activity in a small-world network of excitable neurons. *Phys. Rev. Lett.* 92, 198101.

Scannell, J. W., Burns, G. A., Hilgetag, C. C., O'Neil, M. A., and Young, M. P. (1999). The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex* 9, 277–299.

Schüz, A., and Braitenberg, V. (2002). "The human cortical white matter: quantitative aspects of cortico-cortical long-range connectivity," in *Cortical Areas: Unity and Diversity*, eds A. Schuez and R. Miller (London: CRC Press).

Schüz, A., Chaimow, D., Liewald, D., and Dortenman, M. (2006). Quantitative aspects of corticocortical connections: a tracer study in the mouse. *Cereb. Cortex* 16, 1474–1486.

Schüz, A., and Demianenko, G. P. (1995). Constancy and variability in cortical structure. A study on synapses and dendritic spines in hedgehog and monkey. *J. Hirnforsch.* 36, 113–122.

Sernagor, E., Eglen, S., Harris, B., and Wong, R. (2006). *Retinal Development*. Cambridge: Cambridge University Press.

Smith, C. U. M. (2010). Does history repeat itself? Cortical columns. *Cortex* 46, 279–280.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems* 85, 55–64.

Sporns, O., Honey, C. J., and Kotter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE* 2, e1049. doi: 10.1371/journal.pone.0001049.

Sporns, O., Tononi, G., and Edelman, G. M. (2000). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex* 10, 127–141.

Striedter, G. F. (2004). *Principles of Brain Evolution*. Sunderland, MA: Sinauer.

Striedter, G. F., and Northcutt, R. G. (2006). Head size constrains forebrain development and evolution in ray-finned fishes. *Evol. Dev.* 8, 215–222.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Curr. Opin. Neurobiol.* 5, 520–526.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-World' networks. *Nature* 393, 440–442.

Young, M. P. (1992). Objective analysis of the topological organization of the primate cortical visual system. *Nature* 358, 152–155.

Yu, Y.-C., Bultje, R. S., Wang, X., and Shi, S.-H. (2009). Specific synapses develop preferentially among sister excitatory neurons in the neocortex. *Nature* 458, 501–504.

Zamora-López, G., Zhou, C., and Kurths, J. (2010). Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.* 4:1. doi: 10.3389/neuro.11.001.2010.

Zhang, K., and Sejnowski, T. J. (2000). A universal scaling law between gray matter and white matter of cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5621–5626.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 May 2009; paper pending published: 07 June 2009; accepted: 14 March 2010; published online: 14 May 2010.
Citation: Kaiser M and Hilgetag CC (2010) Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. Front. Neuroinform. ***4****:8. doi: 10.3389/fninf.2010.00008*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 07 July 2010
doi: 10.3389/fninf.2010.00011

Extending stability through hierarchical clusters in Echo State Networks

Sarah Jarvis[1,2]*, Stefan Rotter[1,3] and Ulrich Egert[1,2]

[1] *Bernstein Center Freiburg, Freiburg, Germany*
[2] *Biomicrotechnology, Department of Microsystems Engineering, Faculty of Engineering, University of Freiburg, Freiburg, Germany*
[3] *Computational Neuroscience, Faculty of Biology, University of Freiburg, Freiburg, Germany*

Edited by:
Claus C. Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
John M. Beggs, Indiana University, USA

***Correspondence:**
Sarah Jarvis, Biomicrotechnology, Department of Microsystems Engineering, Faculty of Engineering, Albert-Ludwig University, Georges-Koehler-Allee 102, 79110 Freiburg im Breisgau, Germany.
e-mail: jarvis@bcf.uni-freiburg.de

Echo State Networks (ESN) are reservoir networks that satisfy well-established criteria for stability when constructed as feedforward networks. Recent evidence suggests that stability criteria are altered in the presence of reservoir substructures, such as clusters. Understanding how the reservoir architecture affects stability is thus important for the appropriate design of any ESN. To quantitatively determine the influence of the most relevant network parameters, we analyzed the impact of reservoir substructures on stability in hierarchically clustered ESNs, as they allow a smooth transition from highly structured to increasingly homogeneous reservoirs. Previous studies used the largest eigenvalue of the reservoir connectivity matrix (spectral radius) as a predictor for stable network dynamics. Here, we evaluate the impact of clusters, hierarchy and intercluster connectivity on the predictive power of the spectral radius for stability. Both hierarchy and low relative cluster sizes extend the range of spectral radius values, leading to stable networks, while increasing intercluster connectivity decreased maximal spectral radius.

Keywords: reservoir networks, feedforward, clustered networks

INTRODUCTION

Echo State Networks (ESN) are a type of reservoir networks that have been demonstrated to be successful at predicating non-linear signals, especially those with strong spatiotemporal components (Skowronski and Harris, 2007; Tong et al., 2007; Verstraeten et al., 2007). Proposed by Jaeger (2001), they exploit a reservoir of analog units with random but fixed connections where training affects only weights that project from the reservoir to the output population units. This makes reservoir networks computationally cheap to train in comparison to methods such as backpropagation. The stability of ESNs is assured by fulfilling the *echo state property*, which ensures that the initial conditions are washed out at a rate independent of the input and prevents the accumulation of activity (Jaeger, 2001). Criteria for fulfilling the echo state property are outlined for a specific subclass of ESNs in Buehner and Young (2006) and for leaky integrator units (Jaeger et al., 2007).

Although their strength derives from the homogeneity of the reservoir and its ability to generate rich non-linear dynamics, the connectivity of ESN reservoirs is random and therefore suboptimal. Previous reports investigated the idea of optimizing the network by modifying the reservoir connections, either by pruning connections (Dutoit et al., 2009) or training connections within the reservoir (Sussillo and Abbott, 2009). Other work has focused instead on modifying network architecture by either including a hierarchy of ESNs to extract features (Jaeger, 2007) or introducing reservoir substructures (Deng and Zhang, 2007; Xue et al., 2007). Both of the latter papers reported an improvement in performance against conventional ESNs for specific tasks with interesting additional properties: For the decoupled ESNs (DESN) presented by Xue et al., an input signal composed of sinusoidal terms was presented to a network where the reservoir consisted of laterally inhibited clusters. They observed that each cluster attenuated to a subcomponent of the input signal, improving performance. The Scale-Free Hierarchical ESNs (SHESN) of Deng and Zhang lack explicit decoupling and instead employ a hierarchically clustered reservoir architecture. These SHESNs perform better in predicting the Mackey-Glass sequence, which is often used as a benchmark for non-linear systems.

Deng and Zhang further noted that SHESNs appeared to be more robust to the choice of spectral radius as they remained stable for a larger range of spectral radius values. They observed stable behavior for $\rho = 6$, much higher than in similar ESNs that only displayed stable responses for $\rho < 1$. While this does not imply that the echo state property has been violated by increasing $\rho > 1$, it does, however, indicate that clustered ESNs may extend the range of permissible values for stable and transient states. In their report, only one set of reservoir parameters was examined for SHESNs and so how ρ varied with reservoir architecture was not quantified. Thus, it is unclear how the transition from a homogeneous reservoir to a hierarchically clustered network affects the stability of the system; in particular, is stability affected more by the presence of hierarchy or of clusters, or by a combination of both? Furthermore, is the transition from stable to unstable behavior independent of the degree of clustering and does the activity transition through an intermediate regime, such as those observed elsewhere (Ozturk and Principe, 2005)?

To address these issues, we use hierarchical ESNs (HESNs), a modified version of SHESNs, as a case study. We begin with a conventional ESN and introduce hierarchical clusters, charting how stability depends on both reservoir architecture and the spectral radius

to determine the relative influence of each. To identify the relative contribution of clusters and hierarchy, we compare reservoirs that are clustered but lack hierarchy against those that have both.

MATERIALS AND METHODS

NETWORK MODEL

We first consider a generic ESN (Jaeger, 2001), which consists of an input population u, reservoir population x and output population y. These populations are coupled by connection weight matrices W^{in} for input to reservoir units, sparse connection matrix W^{res} for connections between reservoir units and W^{out} for reservoir to output units. Note that all matrices are directed, so generally $W_{ij} \neq W_{ji}$. Additionally, ESNs can include feedback connections projecting from y back to x as well as connections directly from the input to output populations. Here, we chose to exclude both as we wanted to consider only the dynamics of the reservoir and their impact on the output population.

The governing equations, modified from the original description in Jaeger (2001), are given by

$$x(n+1) = f\left(\mathbf{W}^{in}u(n+1) + \mathbf{W}^{res}x(n) + n_a(n)\right) \quad (1)$$

$$y(n) = g\left(\mathbf{W}^{out}x(n)\right) \quad (2)$$

where f and g represent activation functions which are typically monotonic and bounded, such as sigmoidals. We chose both f and g as tanh and scaled and shifted the incoming signal from the input population in order to place it into the optimal operating range of the network. Here, the input signal always consisted of a sequence of impulses, each of unit amplitude. Noise n_a is added to the reservoir units in order to stabilize the network during the training process. Unless otherwise stated, noise was uniformly distributed for $|n_a| \leq 10^{-6}$.

The input and output populations were chosen to consist of 5 units each. The reservoir populations were set to be of size 50–1000 units (see following subsection for more details). These sizes were chosen for computational tractability. Connections in both W^{in} and W^{res} were randomly and independently assigned with connection probabilities $conn_{in}$ and $conn_{res}$, respectively. The connectivities were set as W^{in} being fully connected with weights drawn with uniform probability from $[-1, 1]$. Since this study is concerned with reservoirs for feedforward ESNs, training was only performed for tasks where the output units were relevant, such as the calculation of memory capacity. As for all ESNs, W^{in} and W^{res} remain fixed for the duration of the network simulation so training, when performed, did not affect the structure of the reservoir.

INTRODUCTION OF HIERARCHY

Different models of hierarchical networks exist which focus on different aspects of their structure: a clustered or modular architecture (Ravasz et al., 2002), existence of hubs (Sporns et al., 2007; Mueller-Linow et al., 2008), repetition of motifs across different scales (Ravasz and Barabasi, 2003; Sporns, 2006) or a combination of these features. For this study, we focus on a network model based on the SHESNs (Deng and Zhang, 2007), where sigmoidal units were replaced by ESN reservoirs. The resulting reservoir architecture differs from flat modular networks due to the presence of *backbone* units, which are units within a cluster that connect to backbone units in other clusters, in contrast to intracluster (*local*) units that only make connections within their own cluster (**Figure 1A**). Although no universally accepted definition of hierarchy exists, the majority of models

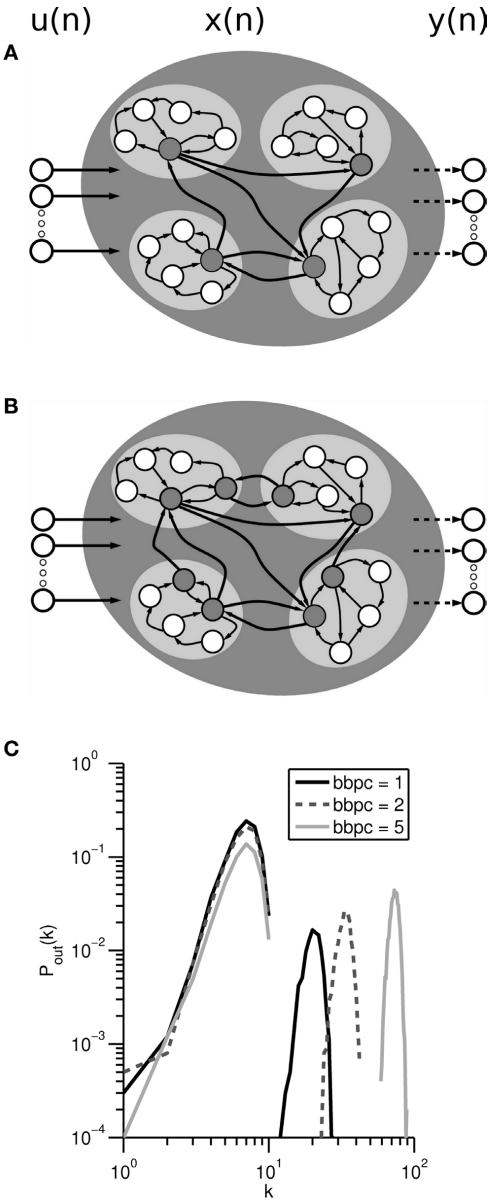


FIGURE 1 | (A) Example of a HESN network with four clusters (light gray). Each cluster contains four local units (white circles) that only project within the cluster and one backbone unit (dark gray circles) that provide connections between clusters. Dashed lines indicate trainable weights. Each unit has a sigmoidal transfer function, bounded on [–1, 1]. **(B)** Non-hierarchical HESN with similar topology and two backbone units per cluster. By increasing the number of backbone units per cluster, the HESN becomes non-hierarchical and increasingly more homogeneous. **(C)** Average out-degree distribution $P_{out}(k)$ for HESN with arbitrary configuration (200 units, 20 clusters, $conn_{inter} = conn_{intra} = 0.7$) against different values for backbone unit per cluster (*bbpc*), calculated over 50 independent realizations of reservoir. The connectivity does not follow a power law distribution and so is not scale-free, unlike the original SHESN model.

examined in the literature focus on structures with a minimum of two levels of hierarchy. However, our aim was to establish the impact of hierarchy within a reservoir and contrast this effect against a structured but non-hierarchical architecture. Thus our modified SHESNs are minimally hierarchical networks, a property that is lost when the number of backbone units per cluster exceeds 1.

Four important changes were made in our networks as compared to the generation process outlined originally by Deng and Zhang to avoid some of the limitations present in their model with respect to the purpose of this case study: (1) Backbone units were originally detailed as being fully connected to one another. To enable us to decouple the intercluster and intracluster connectivities and examine the effect of varying them independently, we allowed the network of backbone units to have sparse connectivity; (2) To ensure that our network was topology independent, we did not assign a spatial location to the reservoir units; (3) To predetermine the network connectivities and cluster sizes, we kept the number of units per cluster constant within a network, using a seeding method outlined in more detail below; (4) To examine the transition from highly clustered to homogeneous architectures, we allowed more than one backbone unit per cluster which additionally allowed us to inspect non-hierarchically clustered networks (**Figure 1B**). These changes introduced three new parameters: intracluster connectivity $conn_{intra}$, intercluster connectivity $conn_{inter}$ and number of backbone units per cluster *bbpc*.

The degree-distributions $P_{out}(k)$ and $P_{in}(k)$ (Albert and Barabási, 2002) followed a bimodal (**Figure 1C**) rather than the scale-free distribution observed for SHESNs. This can be attributed to our restriction of uniform cluster sizes and the fixed values for $conn_{intra}$ and $conn_{inter}$. For this reason, we differentiate from SHESNs and instead refer to our model as hierarchically clustered ESNs (HESN).

A network was generated by first specifying the total reservoir size R, the number of clusters n, and the number of backbone units per cluster b. Each cluster was then created for size R/n and connectivity probability $conn_{intra}$, followed by the generation of the intercluster connectivity matrix. This was performed by identifying the first b units within each cluster as backbone units and then defining a matrix of size (bR/n) with connection probability $conn_{inter}$. The connectivity weights were drawn randomly from a uniform probability on [−1, 1]. The connectivity matrix of the reservoir W^{res} was then rescaled to set the largest eigenvalue to be equal to the defined spectral radius.

To establish the effect of reservoir size on cluster configuration, we defined several configurations of cluster number and size while keeping the total number of units within the reservoir constant. Reservoir size was limited to 1000 units for computational tractability. Note that a conventional ESN is obtained when cluster size is equal to either 1 or the total reservoir size. As a default configuration for HESNs, we assumed $bbpc = 1$ and $conn_{inter} = conn_{intra} = 0.7$. Connection probabilities were always incremented in steps of 0.05. *bbpc* was generally set to be multiples of 5, except if cluster size was below 10. In these instances, all possible values for *bbpc* were tested. These additional criteria allowed us to manipulate the reservoir structure and independently vary its degree of hierarchy and clustering. Importantly, we were able to test hierarchically clustered reservoirs against non-hierarchically clustered reservoirs by retaining a fixed total number of intercluster connections while changing their distribution. This was possible by specifying a larger number of backbone units per cluster and then decreasing the intracluster connectivity to compensate.

In conventional ESNs, the largest eigenvalue of W^{res} (spectral radius ρ) is used as an indicator of stability of network dynamics. Traditionally, ρ less than unity is sufficient for zero-input networks to ensure stable dynamics; Buehner and Young (2006), however, outlines an additional criterion necessary to ensure stability, as stability is also observable for ESNs when $\rho > 1$ (Ozturk and Principe, 2005; Verstraeten et al., 2007). We examined the effect of increasing ρ on stability and during our simulations the spectral radius was varied from 0.05 to 2 in increments of 0.05. For each cluster configuration, we determined ρ_{max}, which we defined as the largest permissible ρ for which the network remained stable under the criteria outlined in the following subsection.

IDENTIFICATION OF NETWORK BEHAVIOR

Different possibilities exist for characterizing a specific reservoir configuration (see Lukosevicius and Jaeger, 2007; Schrauwen et al., 2007 for discussion). Of specific interest to us was how sensitive the system was to noise and how reproducible its dynamics were. Here, we considered an approximation of the maximal Lyapunov exponent to determine stability, and the distribution of reservoir activity values upon multiple stimulation with the same sparse stimulus to assess reproducibility. We also calculated the network memory capacity to provide a measure for the functional performance.

A measure of stability is the Lyapunov exponent λ, which relates the divergence Δ between two trajectories after introducing a small perturbation δ to the elapsed time via $\Delta = e^{\lambda t}$. For stable systems, the deviation between the two trajectories should converge to 0, corresponding to a negative Lyapunov exponent. The existence of a positive Lyapunov exponent signifies that the trajectories exponentially diverge from each other with time, indicating that the system is sensitive to initial conditions and is thus likely to be chaotic. While the Lyapunov exponent cannot be easily calculated analytically for non-autonomous systems, different approaches for determining approximations of Lyapunov exponents for reservoir networks have been outlined, either based on experimental observation of a trajectory's divergence following a perturbation (Skowronski and Harris, 2007) or theoretically by considering the deformation of an infinitesimally small hypersphere as the center follows a trajectory (Verstraeten et al., 2007). Here, we implemented an approximate numerical method, using the former approach (Wolf et al., 1985; Skowronski and Harris, 2007) and estimated the pseudo-Lyapunov exponent by introducing a perturbation δ to the reservoir state at time t_δ. Two identical networks, f and p, were created and driven with identical input and $n_a = 0$. At t_δ, only the second network p was perturbed, after which the simulation continued to run until t_{end}. The resulting divergences was then measured as the Euclidean distance $d = \sqrt{\sum_i (x_i^f(t_{end}) - x_i^p(t_{end}))^2}$. The Lyapunov exponent $\hat{\lambda}$ was calculated as $\hat{\lambda} = 1/(t_{end} - t_\delta)\log_2 d$. The networks were always driven by input consisting of a train of impulses, whose average rate was set to be equal to the rate used for the reservoir activity distribution task below. The perturbation δ was applied to the entire reservoir state, uniformly distributed with $|\delta| = 10^{-6}$ with $t_{end} = 100$ timesteps. As this technique does not lend itself to estimating the

Lyapunov exponent by combining estimates of the same system (Wolf et al., 1985), we classified each network configuration by calculating the Lyapunov exponent for 100 independent realizations and determining the mean.

To relate this somewhat abstract notion of stability to a quantitative description of reservoir dynamics, the distribution of reservoir activity values in response to a sparse stimulus was also considered. This was determined by first defining a stimulus consisting of fixed pattern of input impulses presented multiple times at a constant interstimulus interval of either 10, 50, 100, or 1000 timesteps. We calculated the total reservoir activity for each timestep and normalized it by the number of reservoir units before computing the mean and standard deviation. This was repeated for 100 independent realizations of the same network configuration with $n_a = 0$.

The memory capacity *MC* for different HESN configurations was determined by presenting input signals to the network to be reproduced with increasingly longer delays between presentation and retrieval. *MC* is obtained by summing the correlation coefficient between the output signal and the input signal for each delay k, $MC = \sum_{k=1} MC_k$, where $MC_k = \frac{\mathrm{cov}^2(x(t-k),y(t))}{\mathrm{var}(x(t))\mathrm{var}(y(t))}$, as described in Jaeger (2002) and Verstraeten et al. (2007). Here, we calculated *MC* for $k = 1$–20 timesteps for an input signal consisting of train of exactly one delta input present at each timestep, distributed equally across all five input channels.

RESULTS

The aim of the current study was to establish the impact of reservoir substructures – particularly hierarchical clustering – on the stability of ESNs. We begin by analyzing the effect of hierarchical clustering within reservoirs for a fixed spectral radius, considering underlying factors such as the number of backbone units per cluster and intercluster connectivity. We then test whether clustering by itself is sufficient to alter the upper bound of permissible values for the spectral radius. Lastly, we consider the memory capacity for various network configurations.

IMPACT OF CLUSTERS ON STABILITY

We aim to determine the effect of hierarchical clustering on the limits for spectral radius values for stability and began by systematically examining the impact of cluster on dynamics for different cluster configurations while the spectral radius remained constant. We considered a simple HESN with a fixed $\rho = 1.2$, $conn_{inter} = conn_{intra} = 0.7$ and assumed only one backbone unit per cluster.

Various network configurations were examined by first fixing the total reservoir size and then increasing the number of clusters (**Table 1**). We determined the reservoir activity distribution, using very sparse input sequence repeated every 50 timesteps, and observed three response types across all simulations: (i) the network showed some activation, but this quickly died with exponential decay (**Figure 2A**). The corresponding reservoir activity values had both low mean and low standard deviation. (ii) The network either never stabilized or was unstable after the presentation of the first stimulus (**Figure 2C**), with large mean reservoir activity values with low standard deviation; or (iii) transient behavior, which was similar to stable behavior except

Table 1 | Cluster configurations tested for different reservoir sizes.

Reservoir size	Number of clusters
50	1, 2, 5, 10, 25
100	1, 2, 4, 5, 10, 20, 25, 50
200	1, 2, 4, 5, 8, 10, 20, 25, 40, 50, 100
500	1, 2, 4, 5, 10, 20, 25, 50, 100, 125, 250
750	1, 3, 5, 6, 10, 15, 25, 30, 50, 75, 125, 250
1000	1, 2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 125, 200, 250, 500

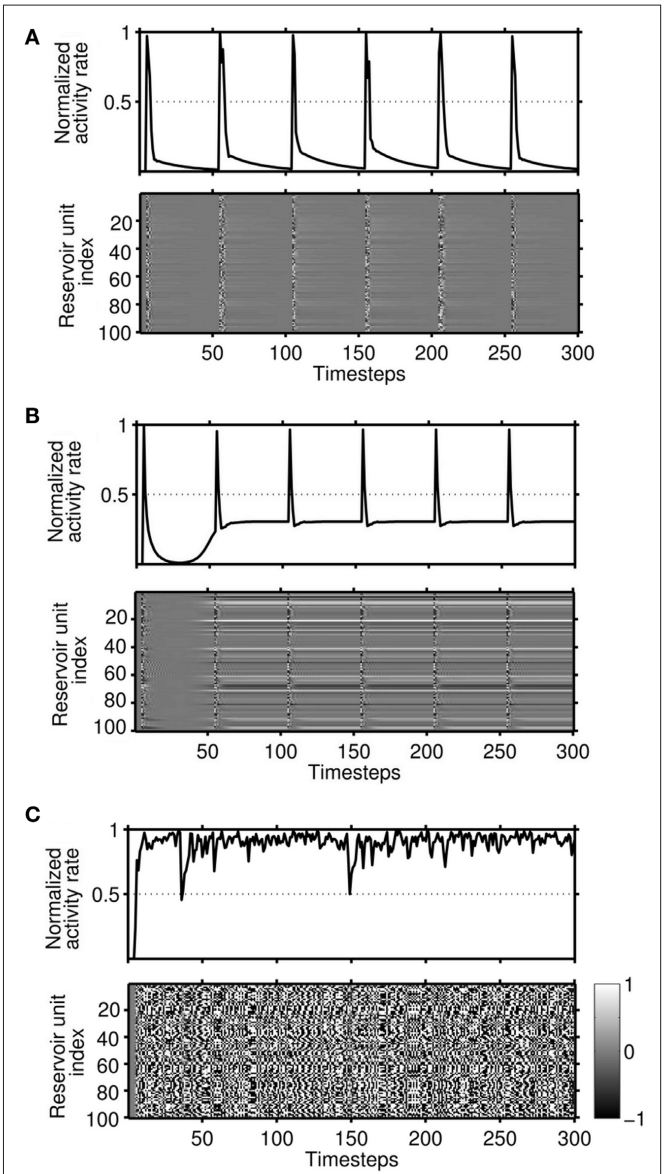


FIGURE 2 | Responses of reservoirs to a repeated stimulus pattern for networks with $R = 100$ and different values of ρ. (A) For $n = 2$ clusters, stable dynamics occur where responses fade within a highly reproducible and fixed time-period, illustrating that the dynamics are input driven only. Shown is for both the normalized total activity rate (top) and reservoir activity (bottom). **(B)** With $n = 20$, transient dynamics were similar to stable but were characterized by activity decaying to a non-zero baseline. Baseline activity values for identical network configurations displayed high variability across trials. **(C)** Unstable dynamics for $n = 50$, where the system never returns to zero activity after the first input.

that the reservoir activity does not return to 0, but rather to some non-zero baseline activity (**Figure 2B**). Although the non-zero baseline remained more or less constant within a trial, its value varied between trials, leading to a standard deviation of normalized reservoir activity values that was far higher than of either stable or unstable reservoirs, where the standard deviation of reservoir activity was <0.05.

The stability of a network did not depend on absolute reservoir size, absolute cluster size or the number of clusters alone, but rather on the combination of these factors (**Figures 3A,B**). Unstable responses occurred in large networks with a small number of larger clusters, while stable responses were observed for configurations with smaller clusters. Transient responses occurred in a broad region between stable and unstable responses and for configurations of intermediate cluster size.

The pseudo-Lyapunov exponent $\hat{\lambda}$ was also calculated for the same set of network configurations (**Figure 3C**). Only negative values were observed, indicating that all configurations were non-chaotic for $\rho = 1.2$. However, the magnitude of the exponent varied and revealed that configurations of intermediate cluster size had the lowest value for $\hat{\lambda}$. This result also demonstrates that the responses of networks with a negative maximum pseudo-Lyapunov exponent do not necessarily decay to 0, and that ongoing activity in feedforward ESNs (**Figures 2B,C**) does not necessarily imply a chaotic regime.

To examine how these trends for stability and chaoticity developed as the spectral radius increased, we analyzed different cluster configurations for a HESN with a fixed reservoir size and considered both distribution of reservoir activity values and the largest pseudo-Lyapunov

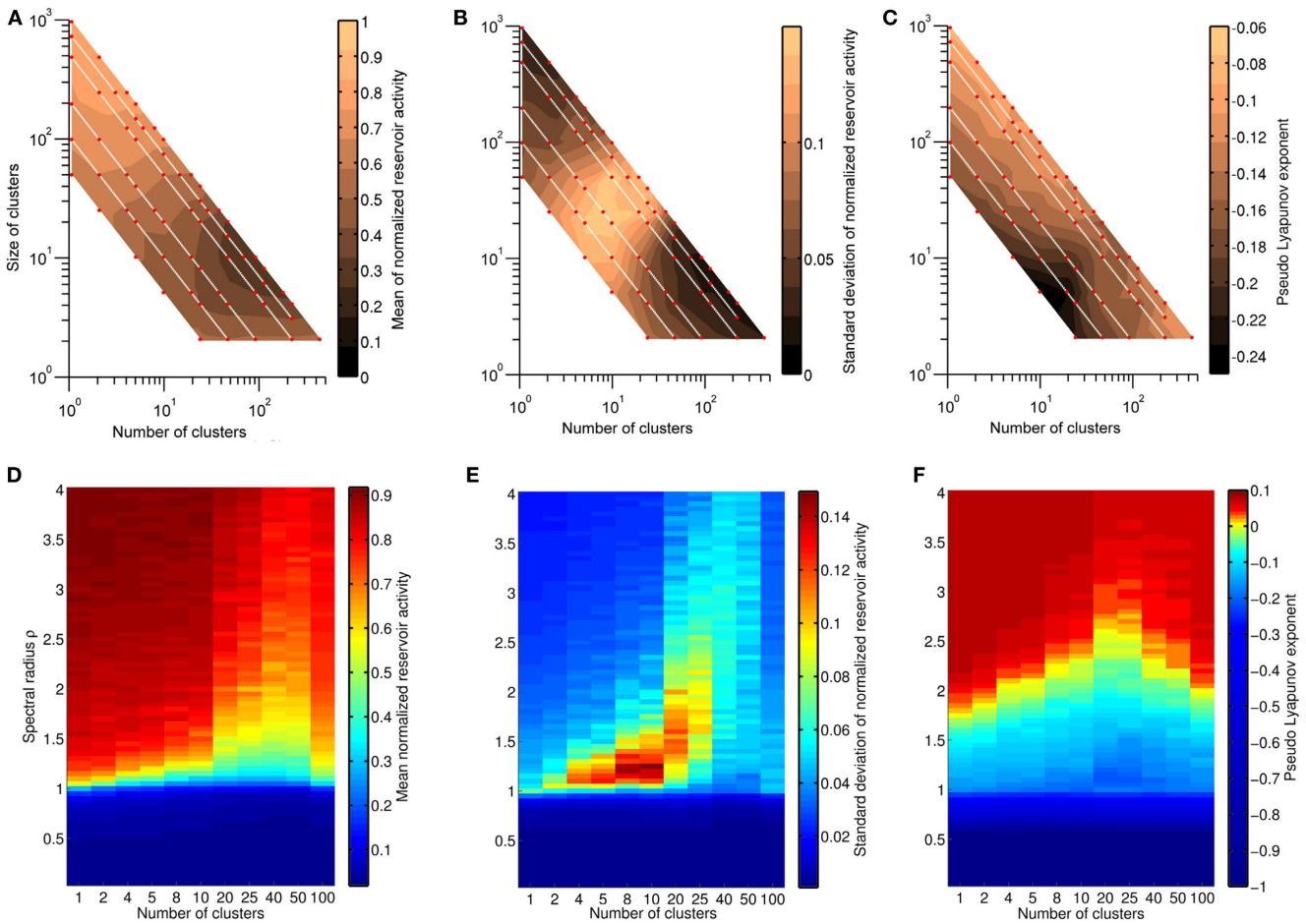


FIGURE 3 | (A–C) Six different reservoir sizes were examined (R = 50, 100, 200, 500, 750, 1000) with varying number of clusters. All other network parameters were kept constant with $\rho = 1.2$, $conn_{intra} = conn_{inter} = 0.7$, and $bbpc = 1$. Mean and standard deviation of normalized reservoir activity **(A,B)** were determined for 100 realizations of each network configuration, with red dots indicating the location of each network configuration in parameter space and white lines indicating networks with identical reservoir size. The area between tested network configurations was interpolated to visualize the trend. Unstable responses are characterized by a high value for the mean reservoir activity and occurred for networks with a small number of large clusters. Transient responses can be distinguished by large fluctuations of normalized reservoir activity, while stable responses have low mean activity. The same network configurations were also used to calculate the pseudo-Lyapunov exponent $\hat{\lambda}$ **(C)**. Negative values were observed for all networks, indicating that networks tested were non-chaotic. **(D–F)** Distributions of normalized reservoir activity values and the pseudo-Lyapunov exponent were calculated as the spectral radius ρ was varied between 0.05 and 4 in HESNs for $R = 200$ with different cluster configurations. ρ_{max} increased with increasing number of clusters up to $n = 40$, while the region of networks that generate transient responses also increases with increasing n.

exponent $\hat{\lambda}$ (**Figures 3D–F**). We set $R = 200$ and kept all other parameters identical to those used in the previous task, but increased ρ from 0.05 to 4 in 0.05 increments. Our results clearly demonstrate that both the mean reservoir activity value and the pseudo-Lyapunov exponent were lower for the same spectral radius value when clusters were present, with the lowest values occurring for reservoirs with 20–50 clusters. This confirms that the addition of hierarchical clusters extends the permissible range of spectral radius values and is consistent with the trend observed in **Figure 3C**, with cluster configurations of intermediate size displaying the largest range of negative Lyapunov exponents.

Increasing the input population from 5 to 50 units for *bbpc* = 1 led to more network configurations with unstable responses (not shown), corresponding to an extension of the region of instability in **Figure 3A** toward networks with a larger number of clusters. We hypothesize that the larger input population provides more activation to the reservoir, which saturates network activity. Thus, networks with larger clusters are able to resist network activity saturation better than networks with smaller clusters.

SPECTRAL RADIUS RANGE

To establish the effect of other network parameters on the range of permissible spectral radius values, we determined the maximal spectral radius value that resulted in a negative pseudo-Lyapunov exponent ρ_{max} for three different parameters related to reservoir architecture: reservoir size *R*, the number of backbones per cluster *bbpc* and intercluster connectivity $conn_{inter}$.

To determine how the span of ρ for stable networks was affected by total reservoir size, the simulation was repeated for $R = 100$, 200, and 500 (**Figure 4A**). As the reservoir size increased, ρ_{max} strongly decreased. The location of the peak value for ρ_{max}, however, remained relatively constant when plotted against relative cluster size (n/R) and occurred for clusters sizes 2–5% of the total reservoir size.

Increasing the number of backbone units per cluster makes the reservoir more homogeneous, which should have a greater impact on the spectral radius range for networks with smaller clusters. Increasing the number of backbone units per cluster to *bbpc* = 5 for $R = 200$ (**Figure 4B**) yielded a decrease in ρ_{max} and caused the distribution to become progressively more uniform. The location of the peak value for ρ_{max} appeared to remain unchanged for $bbpc \leq 2$. To test whether increasing *bbpc* affects reservoir activity and the pseudo-Lyapunov exponent, we examined the various size networks previously tested using a fixed ρ = 1.2 and *bbpc* = 5 (**Figures 5A–C**). We observed that many previously stable networks now displayed higher values for both mean reservoir activity and the pseudo-Lyapunov exponent, confirming that this effect was a general property of increasing *bbpc*.

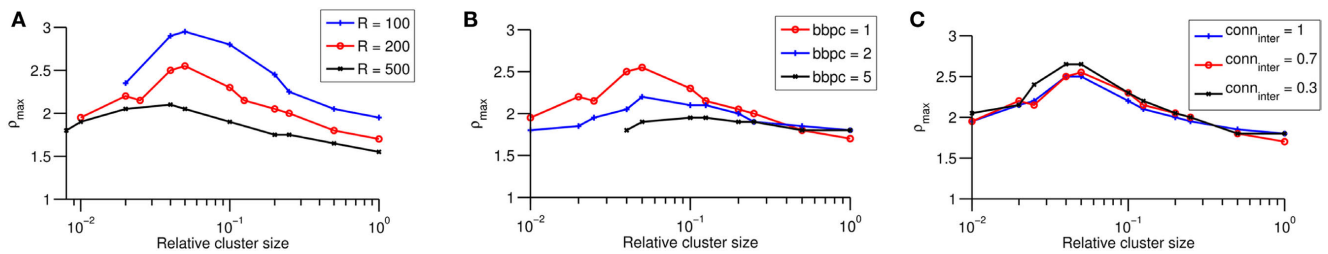


FIGURE 4 | (A–C) Range of ρ_{max} as a function of relative cluster size for three different reservoir sizes **(A),** different number of backbone units per cluster **(B)** and different intercluster connectivities **(C). (A)** Increasing the reservoir size decreases ρ_{max}, while the location of the peak value remains unchanged. **(B)** Increasing *bbpc* drastically decreases the peak value of ρ_{max}, while **(C)** decreasing $conn_{inter}$ only slightly increases ρ_{max} across different relative cluster sizes. Note that the default network configuration for this task is signified by red in all plots.

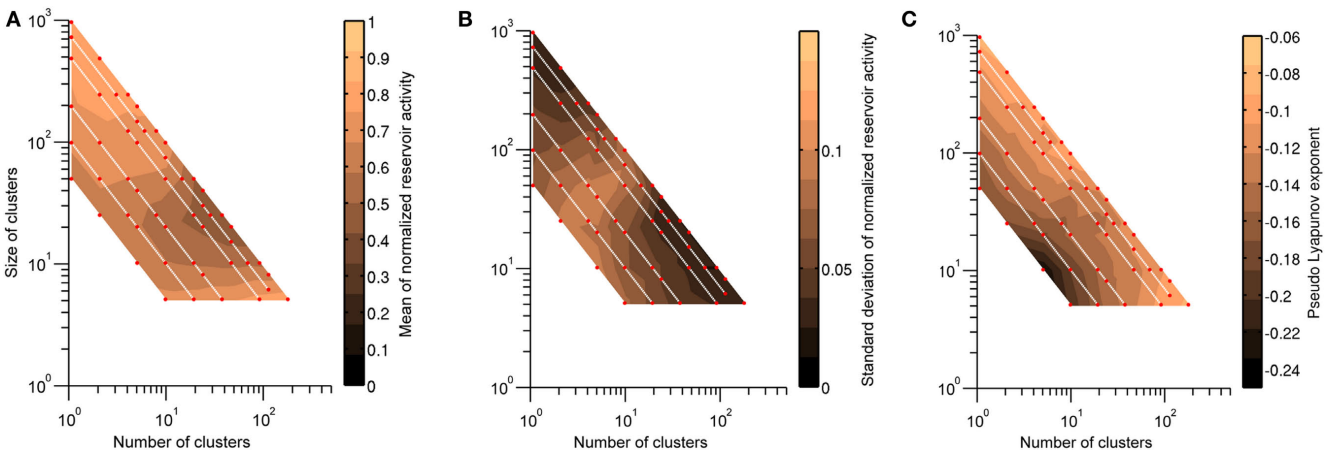


FIGURE 5 | Mean and standard deviation of normalized reservoir activity values (A,B) and the pseudo-Lyapunov exponent $\hat{\lambda}$ (C) when the number of backbone units per cluster *bbpc* is increased to 5. This leads to an increase in the number of configurations with unstable responses. Details as for **Figure 3.**

It has been observed for ESNs that sparser connectivity results in lower error rates. If we treat each cluster as being comparable to a single unit within an ESN reservoir, then the performance of HESNs should increase with lower intercluster connectivity $conn_{inter}$. Since high performance requires stable network configurations, we expect some dependence of ρ_{max} on $conn_{inter}$. Our findings reflect this, with ρ_{max} increasing for network configurations with larger numbers of clusters as $conn_{inter}$ was decreased (**Figure 4C**). The location of peak ρ_{max} at $n = 20$, however, remained unchanged by decreasing $conn_{inter}$.

As varying the spectral radius only changed the values but not the structure of the distribution of eigenvalues, we expected that any change in ρ_{max} for increasingly clustered networks is likely due to differences in the eigenvalue distribution, rather than the spectral radius itself. We therefore analyzed the eigenvalue spectra of 20 realizations of each cluster configuration previously examined. Eigenvalue spectra for configurations with a small number of clusters were homogeneous (**Figure 6A**), similar to those obtained for ESNs. The spectra became increasingly more focused around the origin as clustering was increased (**Figure 6B**). However, the eigenvalue spectra partially dispersed again for reservoirs with a high number of clusters, e.g., $n = 500$, and appears similar to **Figure 6A**, but now with approximately 40% of the eigenvalues located at the origin. The distribution of eigenvalues sorted by distance from the origin summarizes this for all cluster configurations (**Figure 6C**). The distributions increasingly deviated from a homogeneous distribution as the number of clusters increased, due to the progressively more significant contribution of the intercluster connections.

HIERARCHICAL CLUSTERS

Hierarchy is determined by the degree distribution while clustering is determined by the degree count. To establish the relative influence of each on network dynamics, we compared four network configurations with varying *bbpc* and $conn_{inter}$ values. Specifically, we chose values for both parameters that would lead to identical degree counts but different degree distributions (**Table 2**).

Our results in the previous section demonstrate that ρ_{max} is decreased by increasing *bbpc* (**Figure 4B**) or decreasing the intercluster connectivity. These trends were confirmed while testing the degree counts of networks (**Figure 7**). While we observed no significant difference for reservoirs with five clusters or less, as the number of clusters increased, so did the difference between ρ_{max} for hierarchical and non-hierarchically clustered networks. The range of ρ_{max} was higher for the hierarchically clustered networks than for the non-hierarchically clustered networks. We can conclude therefore that it is the degree distribution and thus hierarchy, rather than degree count and clusters, that exerts the larger influence on the range of ρ_{max}.

MEMORY CAPACITY

So far, we considered only measures related to the stability using measures related to reservoir dynamics. To reconcile the affect of clustering with measures related to the performance of a ESN, we calculated the memory capacity *MC* of HESNs while varying the number of clusters *n*, intercluster connectivity $conn_{inter}$, and *bbpc* separately to establish their individual influence as ρ is increased ($R = 200$, $conn_{intra} = conn_{inter} = 0.7$, and $bbpc = 1$, unless otherwise specified). For each network configuration, *MC* was calculated as the mean of twenty independent realizations.

The cumulative memory capacity $\sum_{k=1}^{t} MC_k$ was determined for increasing delays *t* from 1 to 20 while ρ was fixed at 0.7 (**Figure 8**). *MC* increased with decreasing the number of clusters, with a maximum of 8.04 for $n = 1$. The higher plateau level indicates that there was no contribution to *MC* for delays higher than 10 timesteps.

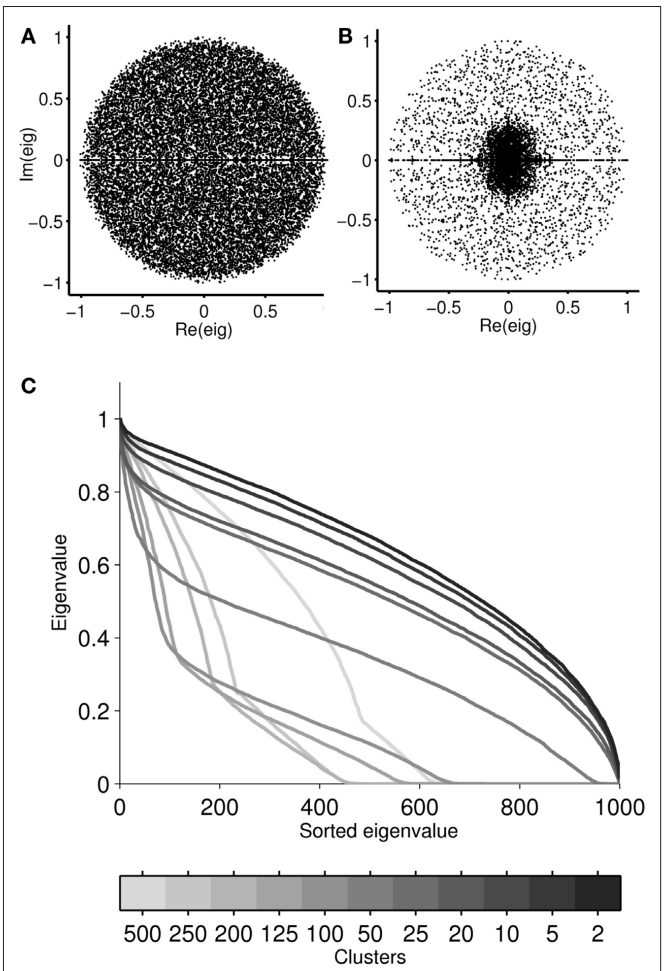


FIGURE 6 | Eigenvalue spectra for 20 realizations of a reservoir with $R = 1000$ and 2 clusters (A) or 125 clusters (B). To visualize how the distribution of eigenvalues varied as clustering was increased, eigenvalues were sorted by their distance from the origin and plotted for $\rho = 1$ **(C)**, with each curve representing the mean of 20 realizations for each cluster configuration. Reservoirs with $n > 25$ have a non-uniform eigenvalue distribution with more than 30% of their eigenvalues concentrated at the origin.

Table 2 | Network configurations tested for the effect of varying both intercluster connectivity $conn_{inter}$ and number of backbones per cluster *bbpc*. Degree count is calculated as the product of *bbpc* and $conn_{inter}$.

bbpc	$conn_{inter}$	Degree count	
1	0.2	0.2	Hierarchical
1	1	1	Hierarchical
5	0.2	1	Non-hierarchical
5	1	5	Non-hierarchical

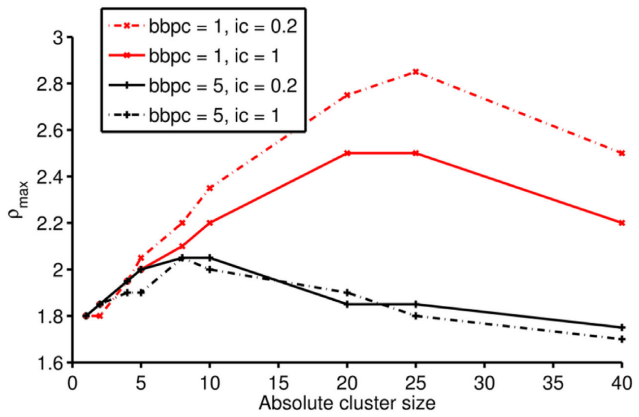


FIGURE 7 | ρ_{max} for increasing number of clusters for four different combinations of values for *bbpc* and $conn_{inter}$. The corresponding increases in ρ_{max} are consistent with trends described above. The relative influence of hierarchy and clustering can be observed by considering the two networks with the same degree count but different distributions: $bbpc = 1$, $conn_{inter} = 1$, and $bbpc = 5$, $conn_{inter} = 0.2$. For $n > 5$ clusters, the former configuration has higher values of ρ_{max}, indicating a greater dependence of ρ_{max} on hierarchy rather than clustering.

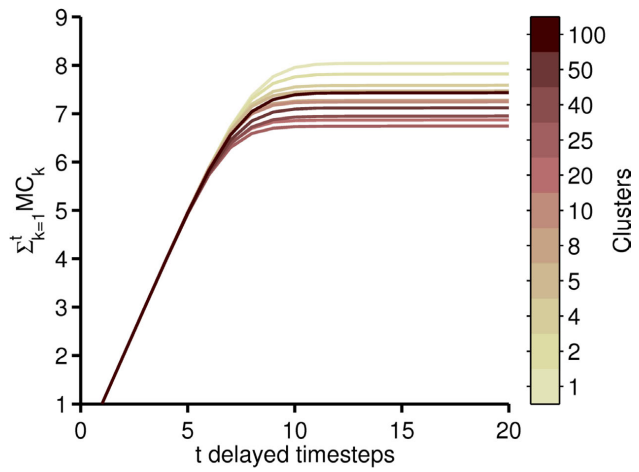


FIGURE 8 | $\sum_{k=1}^{t} MC_k$ was calculated for a collection of clustered reservoirs with total reservoir size $R = 200$ and for delay t ranging from 1 to 20 timesteps.

We examined the effect of varying *bbpc* and $conn_{inter}$ for four cluster configurations with $n \leq 40$ to ensure we could use the same cluster configurations with $bbpc = 5$. Based on curves obtained for **Figure 8**, we selected $n = 1$, 5, 25, and 40 as a representative cross-section. For $0.05 \leq \rho \leq 4$ using $bbpc = 1$, 2, and 5 and the cluster configurations as described above (**Figure 9A**), increasing *bbpc* had no difference for networks where $n \leq 5$. The effect of increasing *bbpc*, however, was most obvious for networks with larger n, with the maximum *MC* increasing from 7.4 ($bbpc = 1$) to 9.8 ($bbpc = 5$) for $n = 40$. Overall, increasing *bbpc* led to a more homogeneous distribution of *MC* values, reflecting that a network with $n = 1$ can be approximated by networks with high n and high *bbpc* values.

MC was compared for $conn_{inter} = 0.2$, 0.3, 0.7, and 1, which were chosen for comparison with previous simulations. Here we examined *MC* across the four cluster configurations with ρ increasing from 0.05 to 4 in increments of 0.05 (**Figure 9B**). *MC* was maximal for both $n = 1$ and 5 when ρ was slightly larger than 1. As expected, *MC* was independent of $conn_{inter}$, since $conn_{inter}$ has little to no influence in networks with few clusters. While *MC* was low in networks with $n = 25$ and 40, their *MC* was also maximal for $\rho > 1$ but, importantly, decreased as $conn_{inter}$ increased. This is likely because intercluster connections became more significant as the number of clusters increased and thus the dominance of intracluster connections decreased in favor of intercluster connections.

DISCUSSION

ESNs with structured reservoirs have been suggested to perform better in predicting non-linear signals (Deng and Zhang, 2007; Xue et al., 2007). Furthermore, clustered reservoirs appear to extend the range of permissible spectral radius values and result in more robust reservoir networks. Our results demonstrate that the addition of clusters within a reservoir does increase the range of spectral radius values that can be chosen, but that this increase does not occur uniformly. Specifically, we charted how the upper limit for the range of the spectral radius ρ that resulted in stable dynamics depended on the reservoir structure and showed: (1) ρ_{max} varied with the number and size of clusters; (2) the range of ρ_{max} was flattened as the number of backbones per cluster was increased and the reservoir becomes more homogeneous; (3) decreasing the intercluster connection probability $conn_{inter}$ increased ρ_{max} but did not affect the location of the peak value for ρ_{max}. These findings demonstrate that several factors interact to set the upper limit for spectral radius within HESNs. As we could examine only a small subset of all possible combinations of parameter values, however, varying multiple parameters simultaneously may have unpredicted effects on reservoir dynamics.

The criterion for stability within ESNs is that the echo state property is fulfilled; for purely feedforward ESNs, $\rho < 1$ is a sufficient but not necessary condition as it ensures that the eigenvalues are scaled such that the system always contracts for any possible input (Jaeger, 2001). This property is evident in our plots, as the estimated Lyapunov exponent and the activity levels are independent of clusters when $\rho < 1$. However, increasing ρ beyond 1 does not directly violate the echo state property and there has been recent interest in methods to identify the upper bound (Buehner and Young, 2006). As our results and those from others (i.e., Ozturk and Principe, 2005; Verstraeten et al., 2007) demonstrate, it is easily possible to choose ρ to be larger than 1 and still avoid unstable states. Understanding the interaction between different structural properties should assist in the identification by quantifying bounds for the spectral radius for different reservoir structures, although we were unable to determine a universal criterion for the echo state property in the networks we examined,

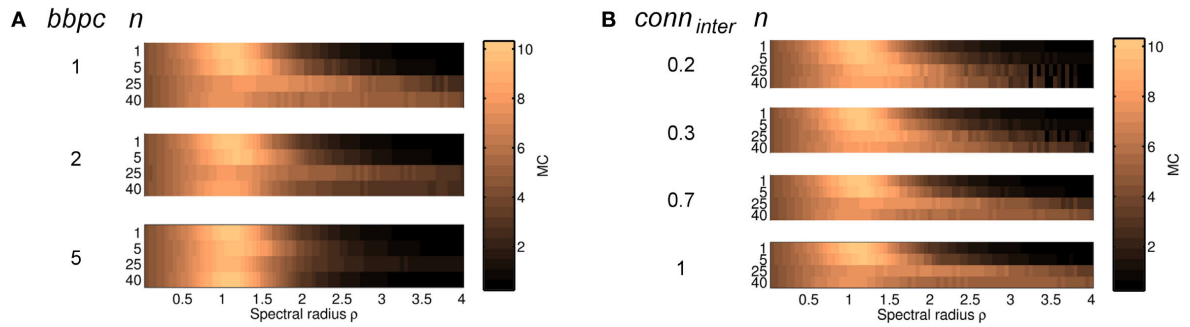


FIGURE 9 | *MC* was calculated for four different cluster configurations as ρ was increased from 0.05 to 4 ($R = 200$, $conn_{intra} = 0.7$ with $conn_{inter} = 0.7$ and *bbpc* = 1, unless otherwise stated). The effect of increasing *bbpc* **(A)** and varying $conn_{inter}$ **(B)** was most noticeable for networks where $n \geq 25$ where *MC* was high over a broader range of ρ.

In homogeneous ESNs, larger ρ values result in a slower decay of the network's response to an impulse (Jaeger, 2001) and strongly affect performance (Venayagamoorthy and Shishir, 2009). It is thus important to choose ρ appropriately. For example, a HESN with 200 units and four clusters was unstable when ρ = 1.2, but was stabilized when the number of clusters was increased. Thus, configurations that would previously have been dismissed due to instability can now be reconsidered, allowing greater flexibility in the design of clustered reservoirs. We were also able to demonstrate that the presence of clusters is reflected in the memory capacity *MC*. Clustered reservoirs displayed lower peak *MC* values; importantly however, *MC* values decreased more slowly as ρ increased for clustered reservoirs when compared to non-clustered reservoirs for ρ > 1.5. These factors together imply that clustered ESNs are more robust over a broader range of spectral radius values than traditional ESNs.

We were also able to map the transition between stable to unstable responses through an intermediate regime, where the reservoir activity decayed to a non-zero baseline after the first stimulus was presented, resembling the transient activity identified by Ozturk and Principe (2005). The baseline of reservoir activity was generally constant for each realization of an HESN. For all reservoir configurations, the mean reservoir activity increased as ρ grew larger. The level of ongoing activity can be thought of as corresponding to the elasticity of the system: as the ongoing activity increases, the potential energy of the system decreases. The gradient at which this transition to instability occurs as the spectral radius is increased, therefore, greatly affects the maximal spectral radius that can be chosen. Our findings indicate that this transition was slowest for 40–50 clusters in reservoirs with 200 units (**Figures 4D,E**), which is reflected in the slow decrease in *MC* for the corresponding network configurations.

EFFECT OF HIERARCHICAL CLUSTERS ON DYNAMICS

One question raised implicitly following the original observation of an extended range of ρ_{max} was whether it was increased by the introduction of hierarchy, or whether merely a clustered topology was sufficient. As hierarchically clustered reservoir are characterized by having only one backbone unit per cluster, we compared hierarchically clustered reservoirs against cluster configurations that had the same number of intercluster connections that were, however, distributed over more than one backbone unit per cluster. Our findings indicate that, while the choice of cluster configuration does influence the value of ρ_{max}, the range of spectral radius values for stable ESNs is affected more by the distribution of intercluster connections onto cluster units, rather than their number. Therefore, we conclude that the presence of hierarchy, rather than clustering *per se*, is responsible for the extended range of permissible ρ values.

This characteristic is likely due to the ability of hierarchically clustered networks to reduce the impact of unstable behavior: if one cluster becomes unstable, its influence on other clusters is minimized by a low number of backbone units, which act as bottlenecks. This behavior has been observed for tri-state networks (Kaiser et al., 2007). Consistent with this, increasing the number of intercluster connections decreased network stability. This is supported by analysis of reservoir unit activity traces in networks with transient responses, where the activity trace showed one cluster that was strongly active.

One question of special interest was whether there was "golden ratio" of cluster sizes: Given a total reservoir size, is there an optimal cluster size for that will maximize ρ_{max}? For the reservoir sizes we examined, the configuration that resulted in the maximal ρ_{max} for $R = 200$ was 25 clusters of 8 units each, corresponding to cluster sizes of 4% of reservoir size. For $R = 100$ and 500, these values were 20 clusters of 5 each (5%) and 50 clusters of 10 units (2%), respectively. These results suggest that cluster sizes of 2–5% of the total reservoir size are optimal. The benefit of having such small cluster sizes can be seen by considering a network with only a few large clusters: if one cluster becomes unstable, the instability of that cluster drives the baseline activity rate higher, reducing the elasticity of the reservoir. As successive clusters become unstable, the baseline rises and the computational potential of the reservoir decreases. For large cluster sizes, this increase happens quite quickly; in contrast, the loss of a small cluster results in smaller increments and therefore a slower transition to instability. Importantly, this interpretation holds true only when clusters have little impact on the activity of other clusters and so is

only observed in hierarchically clustered networks, due to the presence of bottlenecks which allow the HESN's performance to degrade gracefully.

QUANTIFYING ESNs

One problem that has often been discussed is how to measure the "goodness" of a given ESN (Lukosevicius and Jaeger, 2007). While more primitive measures, such as the error rate, can be easily applied, they are task dependent and therefore highly subjective. Most importantly, they do not supply any additional information as to how the reservoir structure impacts on performance. Other measures that examine reservoir activity, such as cross-correlation of reservoir unit activity or entropy of reservoir states (Ozturk et al., 2007), have been used to relate performance to reservoir dynamics. Here, we used the normalized reservoir activity, a metric based on mean reservoir activation following a sparse input applied at a fixed frequency. While this measure has limited use, it was well-suited to indicate the level of ongoing activity, which was of particular interest to us.

A more sophisticated measure is the Lyapunov exponent, which indicates how chaotic a system is. However, this measure has strong shortcomings associated with it. Generally, Lyapunov exponents are difficult to estimate for non-autonomous systems, such as stable feedforward ESNs, which must be externally driven. Correspondingly, any estimate of the Lyapunov exponent is sensitive to the input used, making generalization difficult. On the other hand, there is no clearly better method to estimate the Lyapunov exponent: while the theoretical approach used by Verstraeten et al. (2007) allows for simultaneous consideration of all trajectories in the vicinity of the operating point, it can only estimate positive Lyapunov exponents. As we explicitly wanted to observe how quickly a network configuration moved from stable to chaotic behavior, we chose to approximate the Lyapunov exponent numerically. However, the disadvantage of numerically approximating the Lyapunov exponent is that it is inexact and time-consuming to calculate. Furthermore, determining the Lyapunov exponent by applying a reservoir perturbation is also subjective to the magnitude of the applied perturbation. We were able to demonstrate that for perturbations both two orders of magnitude larger and smaller than $\delta = 10^{-6}$, the same dependency of stability on the number of clusters was still present and with the same structure. This observation allows us to conclude that, although the location of the transition to a chaotic regime depends on input and perturbation magnitude, reservoir stability remains strongly determined by not only the presence but also the size of clusters within the reservoir.

CONCLUSION

The total reservoir size, number of clusters, backbone clusters per unit, and intercluster connectivity all affect the range of permissible spectral radius values and we demonstrated that their interplay determines the upper boundary for the spectral radius. Specifically, the range of permissible spectral radius values strongly depended on the ratio of cluster size to total reservoir size. Hierarchy, rather than clustering alone, had the largest impact on the range of spectral radius values for stable networks. Furthermore, increasing the intercluster connectivity extended the range of spectral radius values for stable ESNs, while increasing the number of backbone units per cluster had the opposite effect.

The transition from stable to unstable dynamics was characterized by responses with varying levels of ongoing activity, even in the absence of any stimulus. The amount of ongoing activity increased as the spectral radius was raised, leading to progressively more unstable reservoir dynamics. Importantly, the rate at which this transition occurred was slowest for hierarchically clustered reservoirs with clusters of size in the range 2–5% of the total reservoir size. We suggest that this is due to backbone units acting as bottlenecks that partition the reservoir, which minimizes the influence of any unstable units by limiting their impact to their own cluster. This effect is optimal when clusters sizes are small enough that the loss of any single cluster does not greatly affect the overall reservoir dynamics, leading to a graceful degradation of reservoir performance, but large enough for them to still contribute to reservoir dynamics. We conclude that hierarchy is a crucial feature for extending the range of permissible spectral radius values within feedforward ESNs.

ACKNOWLEDGMENTS

We would like to thank H. Jaeger for extensive and valuable feedback and T. Gürel for helpful comments and discussion. This work was supported by the German BMBF (FKZ 01GQ0420, FKZ 01GQ0830) and by the EC (NEURO, No. 12788).

REFERENCES

Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.

Buehner, M., and Young, P. (2006). A tighter bound for the echo state property. *IEEE Trans. Neural Netw.* 17, 820–824.

Deng, Z., and Zhang, Y. (2007). Collective behavior of a small-world recurrent neural system with scale-free distribution. *IEEE Trans. Neural Netw.* 18, 1364–1375.

Dutoit, X., Schrauwen, B., Van Campenhout, J., Stroobandt, D., Van Brussel, H., and Nuttin, M. (2009). Pruning and regularization in reservoir computing. *Neurocomputing* 72, 1534–1546.

Jaeger, H. (2001). *The "Echo State" Approach to Analysing and Training Recurrent Neural Networks*, Technical Report 148. Bremen: German National Research Center for Information Technology.

Jaeger, H. (2002). *Short Term Memory in Echo State Networks*, Technical Report 152. Bremen: German National Research Center for Information Technology.

Jaeger, H. (2007). *Discovering Multiscale Dynamical Features with Hierarchical Echo State Networks*, Technical Report 10. Bremen: Jacobs University Bremen.

Jaeger, H., Lukosevicius, M., Popovici, D., and Siewert, U. (2007). Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Netw.* 20, 335–352.

Kaiser, M., Görner, M., and Hilgetag, C. C. (2007). Criticality of spreading dynamics in hierarchical cluster networks without inhibition. *New J. Phys.* 9, 110.

Lukosevicius, M., and Jaeger, H. (2007). *Overview of Reservoir Recipes*, Technical Report 11. Bremen: Jacobs University Bremen.

Mueller-Linow, M., Hilgetag, C. C., and Huett, M. T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS Comput. Biol.* 4, e1000190. doi: 10.1371/journal.pcbi.1000190.

Ozturk, M., and Principe, J. (2005). "Computing with transiently stable states," in *Proceedings of the IEEE International Joint Conference on Neural Networks, 2005* (IJCNN '05) (Houston: IEEE), Vol. 3. 1467–1472.

Ozturk, M., Xu, D., and Principe, J. (2007). Analysis and design of echo state networks. *Neural Comput.* 19, 111–138.

Ravasz, E., and Barabasi, A. L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 67, 026112.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabasi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555.

Schrauwen, B., Verstraeten, D., and Van Campenhout, J. (2007). "An overview of reservoir computing: theory,

applications and implementations," in *Proceedings of the 15th European Symposium on Artificial Neural Networks* (Bruges: ESANN). 471–482.

Skowronski, M. D., and Harris, J. G. (2007). Automatic speech recognition using a predictive echo state network classifier. *Neural Netw.* 20, 414–423.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *BioSystems* 85, 55–64.

Sporns, O., Honey, C. J., and Kotter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE* 2, e1049. doi: 10.1371/journal.pone.0001049.

Sussillo, D., and Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557.

Tong, M. H., Bickett, A. D., Christiansen, E. M., and Cottrell, G. W. (2007). Learning grammatical structure with Echo State Networks. *Neural Netw.* 20, 424–432.

Venayagamoorthy, G. K., and Shishir, B. (2009). Effects of spectral radius and settling time in the performance of echo state networks. *Neural Netw.* 22, 861–863.

Verstraeten, D., Schrauwen, B., D'Haene, M., and Stroobandt, D. (2007). An experimental unification of reservoir computing methods. *Neural Netw.* 20, 391–403.

Wolf, A., Swift, J., Swinney, H., and Vastano, J. (1985). Determining Lyapunov exponents from a time series. *Physica D* 16, 285.

Xue, Y., Yang, L., and Haykin, S. (2007). Decoupled echo state networks with lateral inhibition. *Neural Netw.* 20, 365–376.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 March 2009; paper pending published: 14 May 2009; accepted: 10 June 2010; published online: 07 July 2010.

*Citation: Jarvis S, Rotter S and Egert U (2010) Extending stability through hierarchical clusters in Echo State Networks. Front. Neuroinform. **4**:11. doi: 10.3389/fninf.2010.00011*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 15 September 2009
doi: 10.3389/neuro.11.028.2009

Interplay between topology and dynamics in excitation patterns on hierarchical graphs

***Marc-Thorsten Hütt*[1]* *and Annick Lesne*[2,3]**

[1] *School of Engineering and Science, Jacobs University Bremen, Bremen, Germany*
[2] *Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France*
[3] *Université Pierre et Marie Curie, Paris, France*

Edited by:
Marcus Kaiser, Newcastle University, UK

Reviewed by:
Anirban Banerjee, Max Planck Institute of Molecular Genetics, Germany
Changsong Zhou, Hong Kong Baptist University, China

Correspondence:
Marc-Thorsten Hütt, Computational Systems Biology, School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, D-28759 Bremen, Germany.
e-mail: m.huett@jacobs-university.de

In a recent publication (Müller-Linow et al., 2008) two types of correlations between network topology and dynamics have been observed: waves propagating from central nodes and module-based synchronization. Remarkably, the dynamic behavior of hierarchical modular networks can switch from one of these modes to the other as the level of spontaneous network activation changes. Here we attempt to capture the origin of this switching behavior in a mean-field model as well in a formalism, where excitation waves are regarded as avalanches on the graph.

Keywords: complex networks, patterns, excitable dynamics, mean-field, hierarchical networks, avalanches, modules

PACS: 89.75.Fb, 89.75.Kd, 87.85.Xd, 64.60.aq

INTRODUCTION

The question, how network architecture systematically shapes dynamic processes on the network, has become one of the key topics of research in a range of disciplines – from systems biology (Barabási and Oltvai, 2004; Alon, 2007; Brandman and Meyer, 2008) and ecology (Uchida et al., 2007) to logistics (Armbruster et al., 2005; Guimera et al., 2005) and sociology (Kearns et al., 2006). In neuroscience this question is of particular importance, as functional properties of the brain can be expected to emerge from the organization of (essentially excitable) dynamics on the network of neurons.

Network research employs the formal view of graph theory to understand the design principles of complex systems. For many biological and technical networks, a large-scale system-wide perspective of the network architecture (the 'topology' of such graphs) has yielded some unexpected universal features, e.g., the ubiquitousness of heavy-tail degree distributions (Barabási and Albert, 1999; Barabási and Oltvai, 2004), the presence and possible functions of modularity in enhancing the robustness of a network and in organizing network tasks (Ravasz et al., 2002; Guimera and Amaral, 2005) and a similarity in motif content of functionally similar networks (Milo et al., 2004; Alon, 2007), where motifs are groups of few nodes with a specific link pattern.

Discrete dynamics, and in particular binary and three-state dynamics, have proven helpful in the past for studying, how dynamical processes depend on graph topology (Bornholdt, 2005; Marr and Hütt, 2005; Müller-Linow et al., 2006; Drossel, 2008).

Here we follow this line of thought and study the interplay of topology and dynamics for an important class of dynamical processes, namely excitable dynamics currently used as a minimal model of neuron firing, and for an important graph class, namely hierarchical networks. A hierarchical organization is an important feature of many complex networks in biology. Recently, synchronization on hierarchical networks has been studied (Arenas et al., 2006, 2008). In particular, it has been shown that the hierarchical levels of nodes in the network prescribe a cascade-like sequence toward a fully synchronized state (Arenas et al., 2006).

In general, the shaping of dynamic processes by network topology can also be characterized as a correlation between network properties and properties of the dynamics (Müller-Linow et al., 2008). Qualitatively speaking, a dynamic process typically groups statistically identical nodes into different functional categories. Understanding the impact of the network topology on the functioning of a dynamic process therefore starts by explaining the topological systematics of these node categories induced by the dynamical process.

Within the simple model of excitable dynamics on graphs, which we also use here (see Materials and Methods), two types of correlation between network topology and dynamics have been analyzed by numerical simulation: waves propagating from central nodes and module-based synchronization (Müller-Linow et al., 2008). These two dynamic regimes represent a graph-equivalent to classical spatiotemporal pattern formation.

In order to quantify such modes of pattern formation, one can analyze properties of the matrix of simultaneous excitations, which for example can be studied using a clustering analysis. When analyzing these dynamics on a modular graph, the clusters obtained from the matrix of simultaneous excitations coincide well with the topological modules of the graph. Similarly, when analyzing the dynamics on a scalefree graph, the clusters essentially match groups of nodes with the same distance to the main hub of the graph.

Surprisingly, certain networks, e.g. hierarchical modular networks, contain enough topological cues for allowing both types of patterns to emerge: The dynamic behavior of such networks can switch from one of these modes to the other as the level of spontaneous node activation increases (Müller-Linow et al., 2008).

Here we attempt to understand this switching behavior analytically using a combination of mean-field techniques with the notion of an effective network accessible to excitation.

In Section 'Materials and Methods' we briefly describe the hierarchical network model we use and the model of excitable dynamics, together with a short summary of previous results obtained within the same model. In Section 'Results' we first reproduce the findings from Müller-Linow et al. (2008) in a simpler context, then we discuss the failure of the straightforward mean-field as well as pair-correlation descriptions to account for the organization of the dynamics on graphs (see Materials and Methods). A suitable incorporation of graph topology into a mean-field framework is proposed with the notion of effective networks, which at each moment in time are accessible to the dynamics due to an interplay of spontaneous activity and recovery rate. Excitation patterns can then be viewed as avalanches comprising the accessible effective network. This avalanche model is described in Section 'Results'.

MATERIALS AND METHODS

NETWORK ARCHITECTURE

A hierarchical system is intuitively defined by a multi-layered organization, where few top-level elements are related to several elements on intermediate levels, which are in turn related to a large number of bottom-level elements. Several parameterizations and generative rules of hierarchical graphs coexist in the literature. Typical variants rely on a modules-within-modules view (Ravasz et al., 2002; Kaiser et al., 2007), others focus on the coexistence of modules and central nodes (hubs) (Han et al., 2004; Guimera and Amaral, 2005) or relate the concept of hierarchies to fractality (Sporns, 2006).

Even though our formalism is applicable to any network, here we analyze a specific model of hierarchical graphs, namely the one introduced by Ravasz et al. (2002) and Ravasz and Barabási (2003). In each iteration step four copies of the current network are set up and interlinked in a specific pattern: the central node is linked to all outside nodes; then the local hubs are interlinked among themselves (see **Figures 1 and 2**). The first iteration step leads to four fully connected nodes. At the fourth iteration, the network has $N = 256$ nodes linked by 780 edges, parted in 4 shells according to their distance from the hub, containing respectively $S_0 = k_{\mathrm{hub}} = 120$, $S_1 = 54$, $S_2 = 72$ and $S_3 = 9$ nodes.

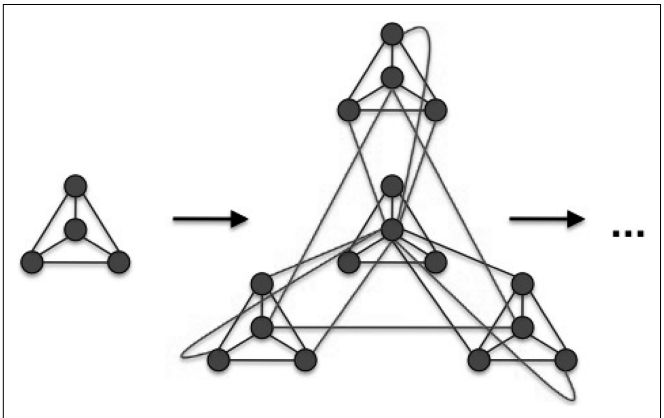


FIGURE 1 | Iteration step in the recursive construction of the hierarchical graphs from Ravasz et al. (2002) and Ravasz and Barabási (2003).

DYNAMICS

For discussing the link between topology and dynamics we use a simple three-state model of an excitable medium. The model consists of three discrete states for each node (susceptible *s*, excited *e*, refractory *r*), which are updated synchronously in discrete time steps according to the following rules: (1) A susceptible node becomes an excited node, if there is at least one excited node in its direct neighborhood. If not, spontaneous firing occurs with the probability f, which is the rate of spontaneous excitation; (2) an excited node enters the refractory state; (3) a node regenerates with the recovery probability p (the inverse of which is the average refractory time of a node). This minimal model of an excitable system has a rich history in biological modeling. It has been first introduced in a simpler variant under the name 'forest fire model' (Bak et al., 1990) and subsequently expanded by Drossel and Schwabl (1992) who also introduced the rate of spontaneous excitation (the 'lightning probability' in their terminology). In this form it was originally applied on regular architectures in studies of self-organized criticality. Other variants of three-state excitable dynamics have been used to describe epidemic spreading (see, e.g., Bailey, 1975; Anderson and May, 1991; Hethcote, 2000; Moreno et al., 2002) or to investigate the impact of topology on the dynamics (Carvunis et al., 2006). Note that in Carvunis et al. (2006) the recovery is deterministic ($p = 1$) and there is no spontaneous excitation ($f = 0$). In contrast, there is no refractory state in the SIS model of epidemic spreading, and no recovery ($p = 0$) in the SIR model (Pastor-Satorras and Vespignani, 2004). As discussed previously (Graham and Matthai, 2003; Müller-Linow et al., 2006), this general model can readily be implemented on arbitrary network architectures. In Graham and Matthai (2003) it has been shown that short-cuts inserted into a regular (e.g., ring-like) architecture can mimic the dynamic effect of spontaneous excitation. Using a similar model setup we have shown (Müller-Linow et al., 2006) that the distribution pattern of excitations is regulated by the connectivity as well as by the rate of spontaneous excitations. An increase of either of the two quantities leads to a sudden increase in the excitation density accompanied by a drastic change in the distribution pattern from a collective, synchronous firing of a large number of nodes in the graph (spikes) to more local, long-lasting and propagating excitation patterns (bursts). Further studies on the activity of integrate-and-fire neurons in the classical small-world model from Watts and Strogatz (1998) also revealed a distinct dependency of the dynamic behavior on the connectivity of the system (Roxin et al., 2004).

In order to study the pattern of joint excitations on a hierarchical graph, we consider for all nodes the individual time series describing their successive states and for each pair of nodes (i, j) compute the number C_{ij} of simultaneous firing events. When applied to the whole network the resulting matrix C (which in the following will be called the *similarity matrix*, as it captures the similarity of time series of two nodes) essentially represents the distribution pattern of excitations. This pattern can now be compared with a corresponding distribution pattern of some topological property.

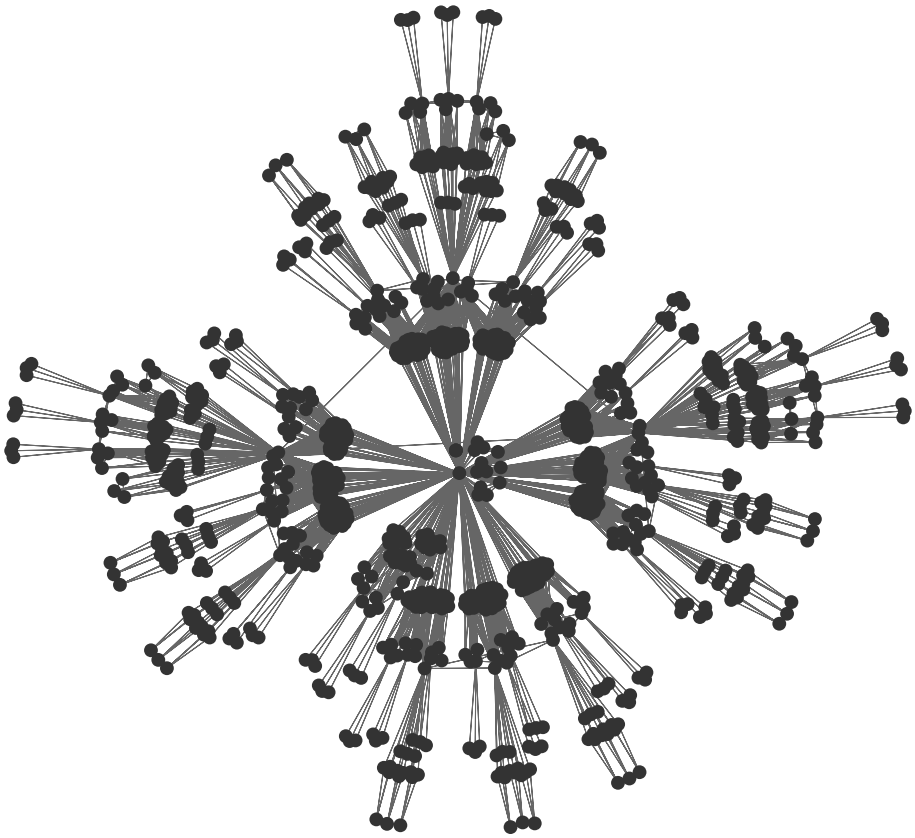


FIGURE 2 | Hierarchical graph with $N = 1024$ nodes obtained from the iterative scheme described in Ravasz et al. (2002) and sketched on Figure 1.

Previous numerical studies (Müller-Linow et al., 2008) have shown that different topological features of complex networks, such as node centrality and modularity, organize the synchronized network function at different levels of spontaneous activity. These findings serve as a starting point of our investigation.

RESULTS

CORRELATIONS OF THE SIMILARITY MATRIX WITH GRAPH TOPOLOGY

In Müller-Linow et al. (2008) clustering trees derived from the matrix of simultaneous excitations (the similarity matrix C) have been compared with groups of nodes derived from topological properties. These properties are modularity and node centrality and they have been represented by a topological-modularity-based reference and a central-node-based reference, respectively. Here we analyze the correlations more directly by comparing the similarity matrix C with matrices designed to capture the topological feature of interest, because the quantities introduced by Müller-Linow et al. (2008), while suitable for experimental studies of dynamics on graphs, have no direct analytical counterpart.

Two topological features are discussed here: (i) modularity, represented by the topological overlap matrix (Ravasz et al., 2002) $T_{ij} = (N_{ij} + A_{ij})/\min(k_i, k_j)$, where N_{ij} is the number of common neighbors of nodes i and j, A is the graph's adjacency matrix and k_i is the degree of node i; (ii) the distance d_i of node i from the graph's central hub; for comparison with the similarity matrix C this distance is cast into a matrix D, where $D_{ij} = 1$, if nodes i and j have the same distance to the hub and $D_{ij} = 0$ otherwise.

As a measure of correlation of two matrices (where one typically is the similarity matrix and the other one of the matrices derived from topology) we use the mutual information $M(m^{(1)}, m^{(2)})$ of the two binary matrices $m^{(1)}$ and $m^{(2)}$:

$$M\left(m^{(1)}, m^{(2)}\right) = \sum_{a,b=0,1} p_{ab} \log\left(\frac{p_{ab}}{p_a^{(1)} p_b^{(2)}}\right), \tag{1}$$

where p_{ab} denotes the relative frequency of encountering the element a in matrix $m^{(1)}$ and b at the same position in matrix $m^{(2)}$ (with $a,b = 1,0$) and $p_a^{(k)}$ is the relative frequency of a in matrix $m^{(k)}$. An alternative quantification is the correlation coefficient $Corr(m^{(1)}, m^{(2)})$ of the two matrices:

$$Corr(m^{(1)}, m^{(2)}) = \sum_{ij} \frac{\left[m_{ij}^{(1)} - av(m^{(1)})\right]\left[m_{ij}^{(2)} - av(m^{(2)})\right]}{\sigma(m^{(1)})\sigma(m^{(2)})}, \tag{2}$$

where $av(m^{(k)})$ and $\sigma(m^{(k)})$ denote the average and standard deviation over all values in matrix $m^{(k)}$, respectively. However, neither for the topological case nor for a typical similarity matrix at intermediate f the distribution of matrix elements follows a Gaussian distribution, so that mutual information better captures the correlation than the correlation coefficient $Corr$. It should be noted

that the quantitative comparison of graphs (and matrices derived from graphs) is, in itself, a broad area of research. Distances between graphs can differ in the topological feature they emphasize or the length scale(s) considered. One alternative to the correlation coefficient and the mutual information used here could be the spectral distance method from Ipsen and Mikhailov (2002). Here we preferred the other two methods, because the dynamics are not directly coupled to spectral features (although this in itself may be an interesting investigation).

Figure 3 shows the correlation of the similarity matrix with the two matrices derived from topology, namely T (measuring modularity; full curves) and D (measuring similarity in their distance to the hub; dashed curves) as a function of the rate of spontaneous excitations f. These curves reproduce the finding from Müller-Linow et al. (2008): at low f the distribution of excitations is predominantly explained by ring structures around the hub (CN, central node reference), while at higher f the distribution becomes dominated by the modular structure of the graph (TM, topological module reference).

It should be noted that these results differ from Müller-Linow et al. (2008, Figure 8A), because in Müller-Linow et al. (2008) a sparser graph has been used in order to keep the average number of excitations at a comparable level for all the graphs discussed there (see Methods section in Müller-Linow et al., 2008 for a detailed description of this procedure). Here we wanted to use the original graph from Ravasz et al. (2002) instead. **Figure 4** shows the same mutual information curves as in **Figure 3B**, but with the 1024-node network depicted in **Figure 2**. With increasing network size the two patterns seem to become more pronounced. In both cases, **Figures 3 and 4**, the two domains in f are clearly visible, each of which is dominated by a specific type of correlation between topology and dynamics.

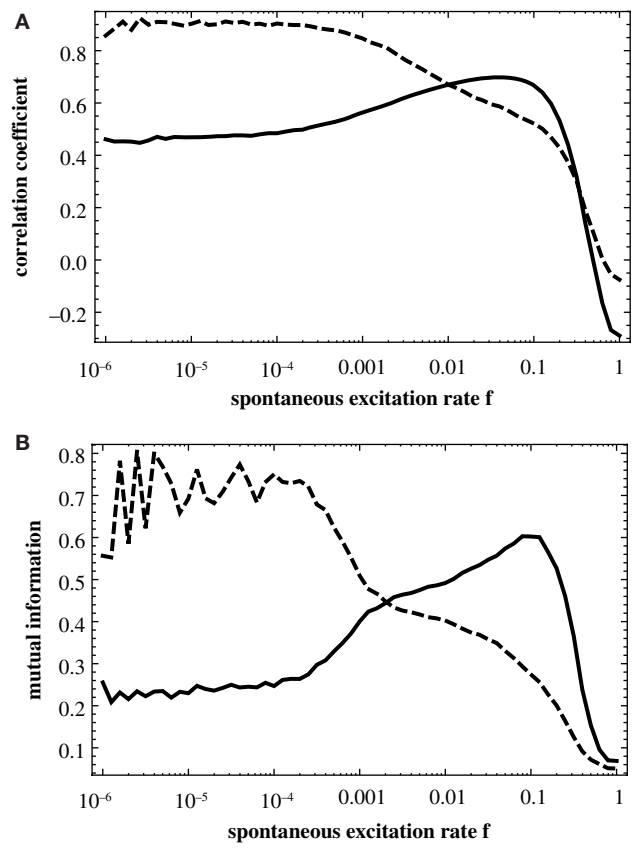


FIGURE 3 | Correlation between the similarity matrix $C(f)$ and the topological matrices T (full curve) and D (dashed curve) as a function of f for the hierarchical graph with $N = 256$ nodes. The correlation is measured **(A)** by the correlation coefficient of the two matrices, Eq. 2, and **(B)** by the mutual information, Eq. 1.

MEAN-FIELD MODEL

Following the lines of Graham and Matthai (2003) and Müller-Linow et al. (2006) we formulate a mean-field description of the three-state excitable dynamics introduced in Section 'Materials and Methods'. As we are interested in the rate of simultaneous excitations of two nodes, we include pair correlations in this model.

Denoting $p_\alpha(t)$ the density of nodes in state $\alpha = e,r,s$ and $q_{\alpha,\beta}(t)$ the probability that a pair of nodes is in state (α,β) at time t [obviously $\sum_\alpha p_\alpha(t) = 1$ and $\sum_\alpha q_{\alpha,\beta}(t) = p_\alpha(t)$], mean-field evolution equations write:

$$p_r(t+1) = p_e(t) + (1-p)p_r(t) \tag{3}$$

$$p_e(t+1) = \left[f + (1-f)\inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] p_s(t) \tag{4}$$

$$\begin{aligned} q(e,s,t+1) = {} & fpq(s,r,t) + (1-f)p\frac{q(s,r,t)q(e,s,t)}{p_s(t)} \\ & + (1-f)fq(s,s,t)\left[1 - \inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] \\ & + (1-f)^2 q(s,s,t)\left[1 - \inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] \\ & \times \inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right] \end{aligned} \tag{5}$$

$$\begin{aligned} q(e,r,t+1) = {} & \left[(1-p)f + (1-p)(1-f)\inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] q(s,r,t) \\ & + \left[f + (1-f)\inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] q(s,e,t) \end{aligned} \tag{6}$$

$$\begin{aligned} q(s,r,t+1) = {} & (1-f)(1-p)\left[1 - \inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] q(s,r,t) \\ & + (1-p)pq(r,r,t) + pq(r,e,t) \\ & + (1-f)\left[1 - \inf\left[1, \frac{\langle k\rangle q(e,s,t)}{p_s(t)}\right]\right] q(s,e,t). \end{aligned} \tag{7}$$

We can first make use of the mean-field description by discussing the average excitation density, which has been in the focus in Müller-Linow et al. (2006), as a function of the connectivity c (i.e., the number of links divided by the number $N(N-1)/2$ of possible

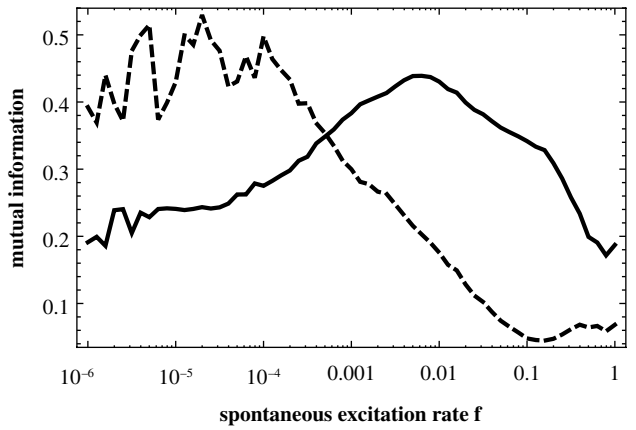


FIGURE 4 | Same as Figure 3B, but for *N* = 1024. The full curve shows the mutual information obtained from the similarity matrix $C(f)$ and the matrix T; the dashed curve is the mutual information for $C(f)$ and D.

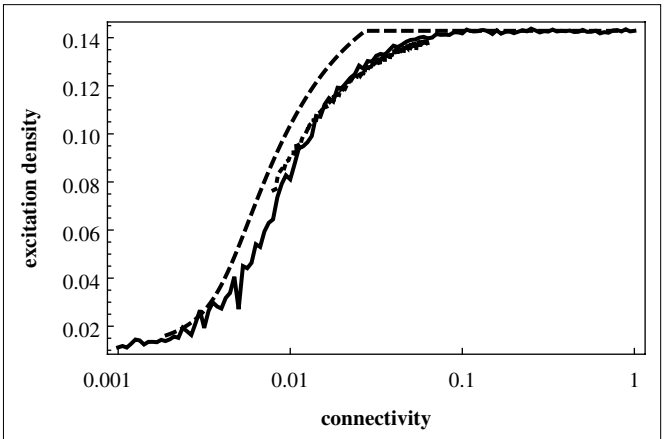


FIGURE 5 | Mean-field excitation density p_e as a function of the connectivity c. Dashed curve: prediction of p_e obtained from the mean-field model including pair correlations; full curve: results from numerical simulations for Erdös–Rényi (ER) graphs; dotted curve: results from numerical simulations for Barabasi–Albert (BA) scalefree graphs; the general setup of the numerical simulations, including the model graph architectures, is the same as in Müller-Linow et al. (2006).

links). **Figure 5** shows the excitation density p_e obtained in the mean-field description, Eqs 3–7, as a function of the connectivity c. It is seen that this refined mean-field model (i.e., including pair correlations) nicely reproduces the increase of average excitation with connectivity.

The joint excitation matrix (i.e., our prediction for the similarity matrix C) is then computed as the conditional correlation function $Q(f, p)$ that two nodes are simultaneously excited knowing that they belong to the same topological class – shells around a central node, or modules. We performed the computation in two limiting regimes, first assuming that the dominant contribution comes from excitation of the central node (CN case) or from excitation from the middle node of even shortest paths (of length 2) within modules (TM case). In this computation, we can ignore the contribution of independent sources of excitation of the two nodes, and consider only the joint excitation initially caused by a unique remote firing event.

In the first case, the probability of excitation of the central node cannot be computed within the above mean-field approach due to the specific hub status of this node. Excitation of the hub comes from either the spontaneous firing of this node or the propagation of an excitation occurring at some node, with stationary probability p_e, leading to:

$$p_{\mathrm{hub},e} = \frac{\inf\left[1, f + k_{\mathrm{hub}} p_e^*(1-f)\right]}{1+\left(\frac{p+1}{p}\right)\inf\left[1, f + k_{\mathrm{hub}} p_e^*(1-f)\right]} \tag{8}$$

where k_{hub} is the degree of the hub; it has to be multiplied by $(1-f)$ to ensure that the hub has not fired at the previous step and the excitation does not encounter a refractory hub. We compute the contribution $Q_n^{(\mathrm{CN})}(f)$ of shell n to $Q^{(\mathrm{CN})}(f)$

$$Q_n^{(\mathrm{CN})}(f) = p_{\mathrm{hub},e} p_s^{2n} = p_{\mathrm{hub},e}\left[1 - \frac{(p+1)p_e}{p}\right]^{2n} \tag{9}$$

and get

$$Q^{(\mathrm{CN})}(f) = \sum_{n\geq 1} \frac{S_n(S_n-1)}{\sum_{j\geq 1} S_j(S_j-1)} Q_n^{(\mathrm{CN})}(f) \tag{10}$$

where S_n is the number of nodes in shell n.

In the second case, all the shortest even paths between two nodes in the same module play the same role; the 2-step paths presumably give the dominant contribution. The network topology is involved through the average number ν of such paths, with roughly, $\nu = C(\langle k\rangle - 1)$ where C is the clustering coefficient. The analog quantity q_{ee} obtained for two nodes chosen at random has then to be subtracted. It then comes

$$Q^{(\mathrm{TM})}(f) = \frac{\nu p_e q_{ss} - q_{ee}}{q_{ee}} \tag{11}$$

Figure 6 shows $Q^{(\mathrm{CN})}$ and $Q^{(\mathrm{TM})}$ as a function of f. Comparing these curves with **Figures 3 and 4**, it is obvious that we fail to account for any of the important features of the numerical results. In particular, $Q^{(\mathrm{CN})}$ does not decrease rapidly enough with f and $Q^{(\mathrm{TM})}$ does not show a peak at higher values of f. The switching of the two modes of organizing excitations on graphs is absent in this Figure.

These two expressions $Q^{(\mathrm{CN})}$ and $Q^{(\mathrm{TM})}$ could be refined by introducing the precise number of loopless paths computed from reduced iterates of the adjacency matrix, but this does not change significantly the results.

Qualitatively, we can already say that the presence of large correlated events (dynamic heterogeneities in space and time) contradicts a mean-field view on the problem. We expect the approach to be fully valid only at large f, when the system reaches a stationary state (the behavior is then trivial, consisting in uncorrelated dynamics, with characteristic time scale $1 + 1/f + 1/p \approx 2 + 1/p$ and no impact of the network topology).

The above equations account for pair correlations in the dynamic state (e.g. $q_{ee} \neq p_e^2$). This improves the plain mean-field approach

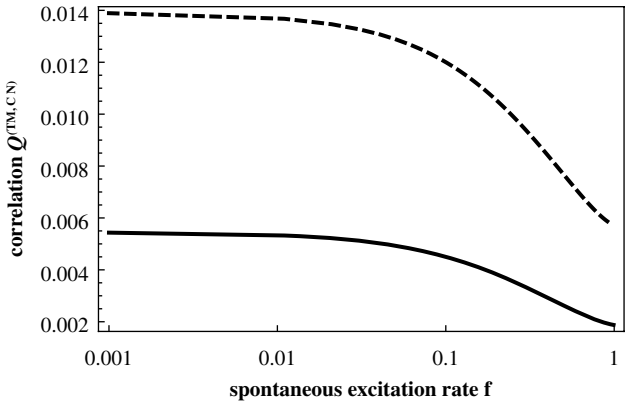


FIGURE 6 | $Q^{(CN)}$ (from Eq. 10; dashed curve) and $Q^{(TM)}$ (from Eq. 11; full curve) as a function of *f*.

and leads to a proper prediction of the average excitation density. But this model dramatically fails to reproduce the actual excitable dynamics and excitation patterns because it ignores large correlated spatio-temporal fluctuations, mainly transient and synchronized waves of excitation ('avalanches') spanning the effective network of susceptible nodes and shaped by network topology. Note that our approach does not account for pair correlations in the degree: Introducing the conditional degree distribution $p(k'|k)$ and considering degree-dependent densities (averages over subsets of nodes with a given degree) in order to account for degree heterogeneity and correlation between neighbors (see e.g. Pastor-Satorras and Vespignani, 2004), instead of the approximation using $\langle k \rangle$, would yield more complicated (analytically untractable) equations but still miss the essential dynamic correlation between neighboring nodes. It would not cure the failure of the mean-field approach to reproduce the correlation between excitation pattern and network topology. We rather turn to another formalism, explicitly describing the avalanches of excitation and how they reflect network topology.

AVALANCHE MODEL

The challenge is to combine a mean-field description of the dynamics with a suitable representation of topology in order to identify the typical *effective network* accessible to the dynamics as a function of the parameters *f* and *p* and the architecture of the network itself. This is the key idea proposed here for understanding the distribution of excitations as pattern on the graph. For random graphs (Erdös–Renyi graphs) the fluctuations in p_e are distributed randomly across the graph. As soon as we provide a topological feature the fluctuations can exploit, these fluctuations lock onto the topological features. This is clearly seen when looking at the correlation between the above topological matrices *T* and *D* and the matrix *C* of simultaneous node excitations (**Figure 3**).

We thus propose to understand the distribution of excitations as avalanches on the accessible parts of the network. A competition between *p* and *f* regulates the size of the effective networks and therefore the avalanche size and, more importantly, the topological features available to the dynamics. Avalanches have been discussed on abstract graphs (Lee et al., 2004) and, in the context of self-organized criticality, in neural networks (see, e.g. Levina et al., 2007b). While we are not discussing self-organized criticality here, we nevertheless employ the formal concept of avalanches as groups of excitation events correlated by the graph's topology. In particular when the graph has a very heterogeneous degree distribution (and, more specifically, contains hubs) an avalanche of excitations propagating in the graph can be regarded as a two-step process: (1) the transportation of the excitation to the seed of the real avalanche, (2) the (amplified) spreading of the avalanche from the seed towards the accessible part of the graph.

Step (1), reminiscent of the coalescent view on a tree, is accounted for by considering the gathering of excitations at the seed *i* of an avalanche, yielding a factor equal to the degree k_i of the seed. Step (2), close to the viewpoint adopted in branching processes, is accounted for by considering nested shells around the avalanche seed. To delineate the seed nodes, we might compute for each node *i* the multiplicative factor $\mu_i = k(i)S_2(i)$: the probability that a node is a seed is identified with the weight $\mu_i / \sum_q \mu_q$. The density of susceptible nodes in shell *n* around *i* is computed in a mean-field approximation as:

$$\Theta(p,f,q) = p_s(1-f)^q + \frac{p_r\, p\left[(1-f)^q - (1-p)^q\right]}{p-f} \tag{12}$$

and the number of nodes in shell *n* is $S_n(i) = \sum_j B_{ij}^{(n)}$ where the matrices $B^{(n)}$ are computed recursively as $B^{(0)} = Id$, $B^{(1)} = A$, $B^{(n)} = H(A^n)\prod_{q=0}^{n-1}[1 - H(A^q)]$ where *H* is the Heaviside function. The overall excitation amplification factor at node *i* is thus:

$$M(i) = k_i \sum_{n\geq 1}\left[\sum_j B_{ij}^{(n)}\right]\prod_{q=1}^{n}\Theta(p,f,q). \tag{13}$$

The largest integer *n* for which $[\sum_j B_{ij}^{(n)}]\prod_{q=1}^{n}\Theta(p,f,q)$ is appreciable gives the depth $d_i(p,f)$ of an avalanche initiated at node *i*. The correlation (as regards their joint excitation) between two nodes *j* and *l* is finally given by:

$$Q_{jl} = \sum_i \frac{\mu_i}{\sum_q \mu_q} \sum_{n\geq 1}\left(\prod_{q=1}^{n}\Theta(p,f,q)\right)^2 B_{ij}^{(n)} B_{il}^{(n)} \tag{14}$$

and the average size of an avalanche can be computed as:

$$\bar{A} = \sum_i \frac{\mu_i}{\sum_q \mu_q} \sum_{n\geq 1}\sum_j B_{ij}^{(n)} \prod_{q=1}^{n}\Theta(p,f,q) \tag{15}$$

Avalanches size and duration reveal the effective connectivity, while their location and frequency depend on *f* and the actual connectivity.

The idea of an effective graph is relevant only if recovery time is smaller than spontaneous firing period, namely $p >> f$. Also, speaking of avalanches makes sense only at low and moderate *f* (typically $1/f > 3$): observing distinct avalanches (well-separated in space and time) requires slow driving: $f << 1$, a threshold dynamics (here all-or-none spontaneous excitation with probability *f*), and a rapid relaxation (diameter $D << 1/f$).

The distribution of weighting factors $\mu_i / \Sigma_q \mu_q$ depends strongly on the detailed topological features of the graph. Essentially, the

weighting factor $\mu_i/\Sigma_q \mu_q$ gives the probability that a node i can trigger an avalanche of excitations. The actual size of the avalanche will then be determined by the effective (i.e. accessible) network of susceptible elements.

Figure 7 shows the correlation between the predicted similarity matrix Q and the matrices T and D derived from topology, together with the corresponding curves from **Figure 3B**. The overall features of the numerical findings (red curves) are explained by the predictions (black curves), in particular the dominance of the hub distance captured in D at low f over the modules accounted for in T, the decrease of the correlation with D as a function of f and the increase of the correlation with T at high f.

When computing the matrix elements Q_{ij} giving the probability of the two nodes i and j being simultaneously excited, we assume that even for small f the observation time is long enough for the potential joint excitations of the two nodes to really unfold. Alternatively, we could have required some spontaneous excitation to have taken place (by inserting an additional factor of f). In the results from the numerical simulations we often observe a decrease of the correlation between the similarity matrix and topological matrices when going to small values of f. The above argument clearly identifies this behavior as a consequence of the finite simulation time. To support this interpretation we performed a longer simulation run for small f and found a systematic increase of the mutual information with simulation time, thus reducing the discrepancies between the numerical and analytical curves.

The results from **Figure 7** have been obtained for the 64-node hierarchical graph. **Figure 8** shows the corresponding results for the larger graph, $N = 256$. The lower two plots in **Figure 8** compare these predictions with the relevant segments of the numerical curves (dashed) from **Figure 4**. Again, the predictions clearly reproduce the increase of the one and the decrease of the other curve. In order to further assess, whether the predicted similarity matrices reproduce the actual distribution features of excitations on the graph, we compute the matrices $Q(f)$ also for a scalefree graph (for which a dominance of a central-node type organization of excitations throughout a large regime of f is known from Müller-Linow et al. (2008) and for a modular, non-hierarchical graph (for which a dominance of the module-based organization of excitations has been observed across a broad range in f). **Figure 9** summarizes these results. The predicted similarity matrices reproduce the numerical finding that on a scalefree graph (left) the central node-based pattern dominates and on a modular graph (right) the module-based pattern dominates across essentially the whole range of f.

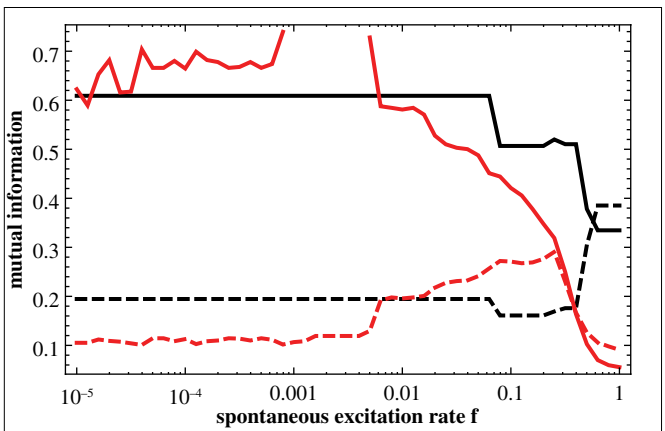


FIGURE 7 | Mutual information as a function of *f*, measuring the correlation of the predicted similarity matrix *Q* with modularity (matrix *T*; black; full) and with distance to the central node (matrix *D*; black; dashed), compared to similar correlation measure between the simulated similarity matrix *C* and matrices *T* and *D* (red; respectively full and dashed); in all cases the 64-node hierarchical graph has been used.

The systematic contribution to the correlations between topology and dynamics we observed comes from synchronized nodes. The amplification factor $M(i)$ given above regulates the amount of simultaneously excited nodes in a particular topological class (a ring around the hub or a module) and therefore the strength of the synchronous signal.

For a given topological class an optimal duration T^* of an avalanche as a function of f and p can be computed as the solution to $T^* mfR(p, f, T^*) = 1$ with $R(p, f, T) = \frac{p}{p-f}[(1-f)^T - (1-p)^T]$ (effective recovery rate, defined as the probability that a site excited at time $t = 0$ has recovered and not yet experienced a spontaneous excitation at time $T + 1$). This time $T^*(p, f)$ is expected to be the relevant 'recovery time', separating two avalanches. Comparing $1/f$ (i.e., the typical time scale available to an avalanche) with T^* should give the expected peak in the correlation of the similarity matrix to this topological class.

Our main finding is that we can explain the qualitative feature of the switching from central node dominated to a module dominated organization of the dynamics, as the rate of spontaneous excitation f is increased.

It should be pointed out that this switching is not a dynamic phase transition. In particular, we do not observe the switching to become sharper with increasing graph size.

The main motivation for our use of avalanches is that excitations spread on the *effective* network (consisting of the nodes in the susceptible state at a given moment in time), whose topological properties depend on f and in turn on the previous history of the dynamics.

A random excitation will trigger a certain cascade of joint excitations of other nodes. The sizes of these contributions to the similarity matrix (and the nodes involved in these contributions) depend strongly on the *effective* network. At large f, random excitations follow so rapidly one after another that only few nodes in the graph are susceptible, leading to only few and small-scale contributions to the similarity matrix. At small f, on the other hand, the graph has enough time to recover between consecutive random excitations, allowing each random excitation to essentially trigger a whole avalanche of joint excitations.

In **Figure 10** the average shortest path length of the effective networks as a function of f. According to our avalanche concept, we believe that this decrease of the average path length with f, which is clearly discernible in **Figure 10**, is the main driving force of the switching behavior.

DISCUSSION

We are mainly interested in understanding the global interplay and network effect, not only the local consequences of a node

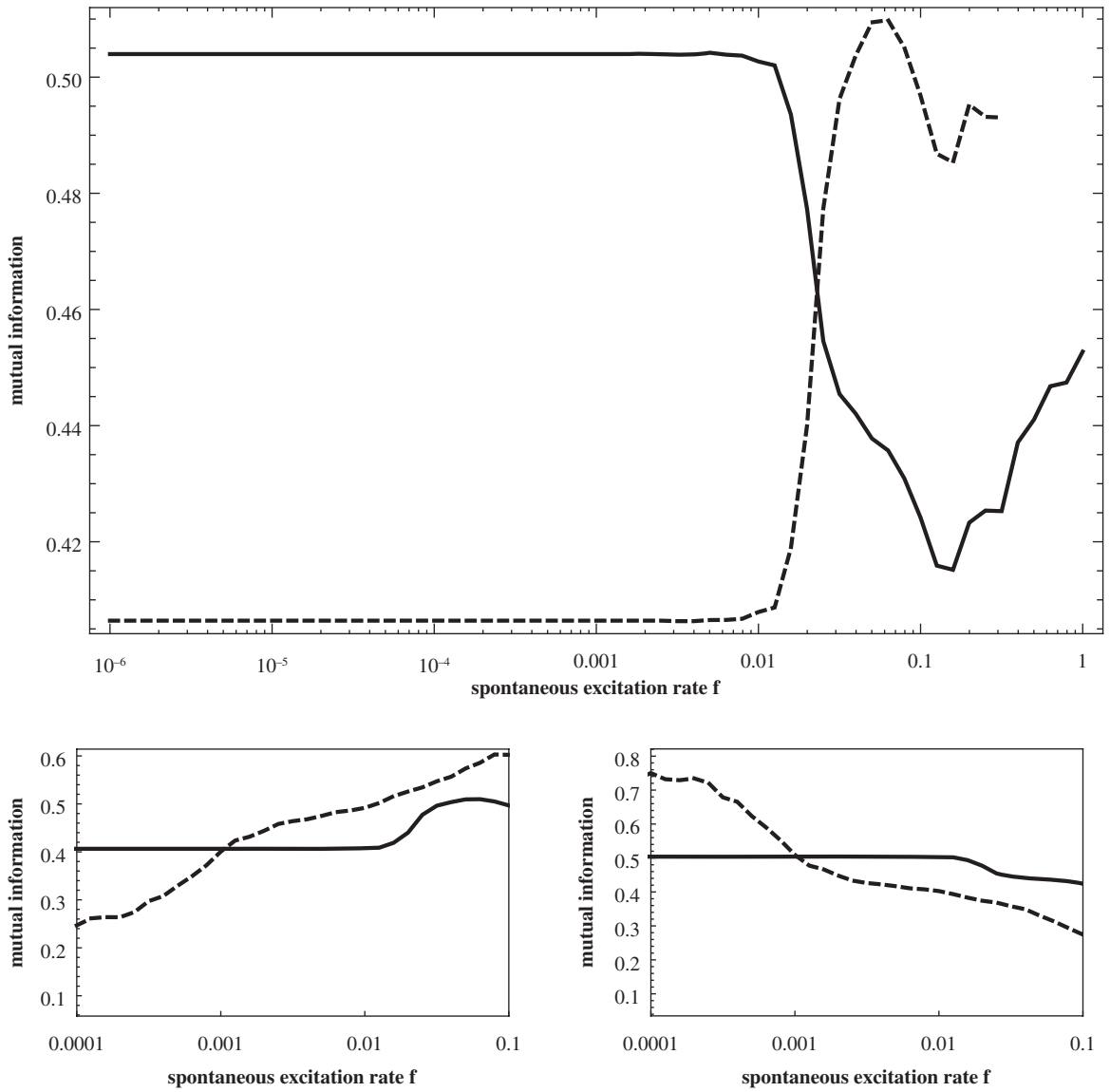


FIGURE 8 | Upper figure: Mutual information as a function of *f*, measuring the correlation of the predicted similarity matrix Q_{ij} with modularity (matrix *T*; dashed curve) and with distance to the central node (matrix *D*; full curve). Compared to **Figure 7** ($N = 64$) the graph size is now $N = 256$. The lower two figures compare each of these curves (left-hand side: modularity; right-hand side: central node; full curve: prediction; dashed curve: result from numerical simulation).

or motif topological specificity. Accordingly, we analyzed the relation between topology and dynamics with methods from spatiotemporal pattern formation. When the carrier space is no longer a regular lattice, it becomes relevant to investigate the origin of the pattern and the shaping of the patterns by network topology. This can be achieved by studying correlations between topological properties and properties of the dynamics. The results reflect the exploitation of certain topological features by dynamics. In particular, by computing the mutual information between the respective matrices, we show the *f*-dependence of the correlation between topology and dynamics and re-discover the dependence as discussed previously with other means (Müller-Linow et al., 2008).

Here we have been able to understand the dominant features of these patterns from Müller-Linow et al. (2008) analytically using a combination of a mean-field approach and avalanche viewpoint: on a hierarchical graph, the core feature is the switching from a 'central-node' to a 'topological-module' mode of organization of simultaneous excitations as a function of *f*; on scalefree and modular graphs, it is the dominance of one of the modes across the whole range of *f*.

We have also shown that the mean-field model reasonably well explains the increase of the average excitation density as a

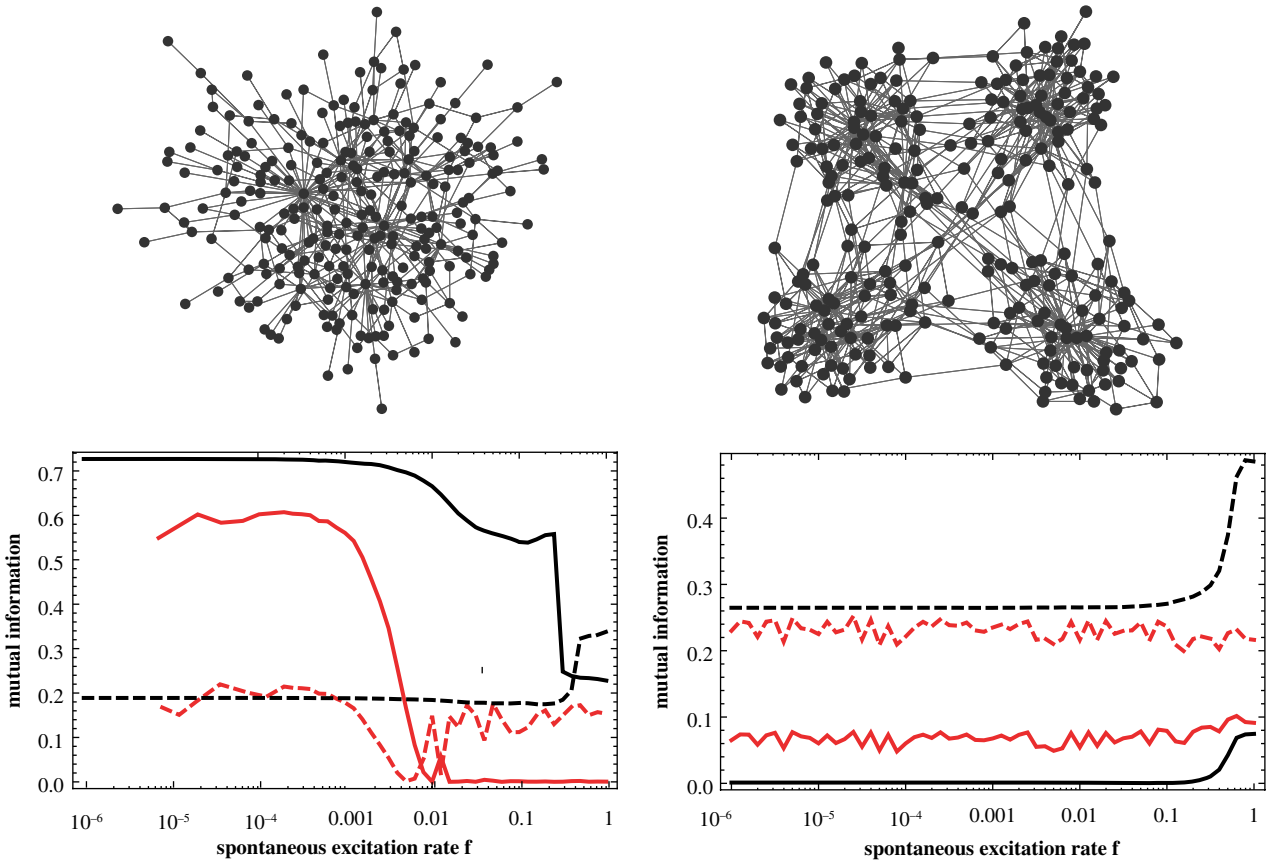


FIGURE 9 | Top row: scalefree (BA) graph (left) and modular graph (right), each consisting of 256 nodes. Bottom row: For each of the graphs, mutual information as a function of *f*, measuring the correlation of the predicted similarity matrix Q_{ij} with modularity (matrix *T*; dashed curve) and with distance to the central node (matrix *D*; full curve). In both cases, numerical results are shown as well (red curves) based on a single 5000 time step simulation of the system on the graphs shown in the top row. In the case of the numerical result for the scalefree graph, two additional points should be noted: (1) as in Müller-Linow et al. (2008) due to the high connectivity, the dynamics are re-scaled by requiring a certain percentage of neighbors (20%) to be active in order to activate a node; (2) at such small size, the population of scalefree graphs is very heterogeneous and therefore the curve can only be seen as a single-graph representative.

function of connectivity, which was reported in Müller-Linow et al. (2006).

Due to its very principles and associated approximations, the mean-field model can describe average properties of the excitations, but not the interplay between topology and dynamics (i.e., how topological features regulate the distribution patterns of excitations on the graph). In fact, the validity of the mean-field approach depends on the investigated feature. For instance, when looking at the curve of average excitation as a function of connectivity, only a very small dependence on graph topology is seen (Müller-Linow et al., 2006), and the mean-field description is here in good agreement with the numerical data. But it fails to account for the two regimes of correlation between topology and dynamics observed as a function of *f*, whose origin is unraveled and quantitatively captured in our avalanche description.

Several important steps are left for future work: In order to develop a global picture of this interplay between topology and dynamics one needs to study the topological properties of the effective networks and, therefore, the size distribution of the avalanches as a function of topology and system parameters. As discussed at

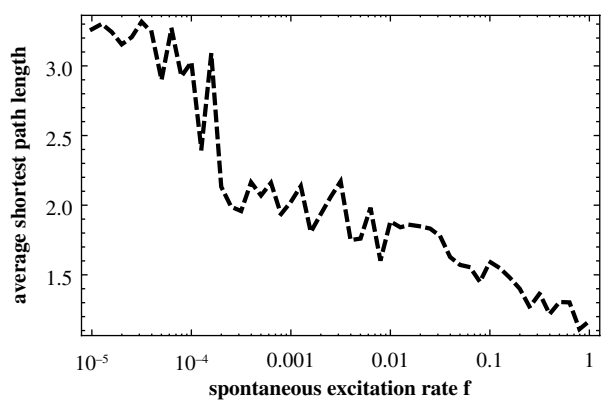


FIGURE 10 | Average shortest path length for effective networks formed by susceptible elements at each time step of the numerical simulation of the dynamics. In order to compute this curve for each value of *f* we used a 10,000 time step simulation of the dynamics on the 256-node hierarchical graph, for each effective network averaged over the shortest paths between all nodes and, finally, averaged over the effective networks from 1000 time steps.

the end of the results section, our current formalism already leads to some predictions for the time scales relevant for avalanches of excitations, as well as for the avalanche size and duration distributions. An extension of this formalism may help us link the observations from this simple model of excitations to the formalisms discussed in Levina et al. (2007a,b).

We also believe that the distribution of weighting factors μ_i and amplification factors $M(i)$ may provide interesting characterizations of a network for dynamical purposes. This certainly needs further exploration.

ACKNOWLEDGMENTS

We are indebted to Mark Müller-Linow (Darmstadt, Germany) for providing the first set of numerical simulations we here compare our analytical predictions with. We wish to thank Claus Hilgetag (Bremen, Germany) for fruitful discussions on this topic. M.H. acknowledges generous support from the IHÉS guest program for the purposes of this work; A.L. has been a visitor of the ICTS program at Jacobs University during the completion of this work. Publication of this work has been supported by the Agence Nationale de la Recherche (ANR 05-NANO-062-03 'CHROMATOPINCES').

REFERENCES

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.

Anderson, R. M., and May, R. M. (1991). Infectious Diseases of Humans. Oxford, Oxford University Press.

Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., and Zhou, C. (2008). Synchronization in complex networks. *Phys. Rep.* 469, 93–153.

Arenas, A., Diaz-Guilera, A., and Perez-Vicente, C. J. (2006). Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.* 96, 114102.

Armbruster, D., Mikhailov, A. S., and Kaneko, K. (eds) (2005). Networks of Interacting Machines. Singapore, World Scientific Publishing.

Bailey, N. T. J. (1975). The Mathematical Theory of Infectious Diseases and its Applications. London, Charles Griffin.

Bak, P., Chen, K., and Tang, C. (1990). A forest-fire model and some thoughts on turbulence. *Phys. Lett. A* 147, 297.

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

Barabási, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113.

Bornholdt, S. (2005). Systems biology: less is more in modeling large genetic networks. *Science* 310, 449.

Brandman, O., and Meyer, T. (2008). Feedback loops shape cellular signals in space and time. *Science* 322, 390.

Carvunis, A. R., Latapy, M., Lesne, A., Magnien, C., and Pezard, L. (2006). Dynamics of three-state excitable units on poisson vs. power-law random networks. *Physica A* 367, 595–612.

Drossel, B. (2008). Random boolean networks. In Reviews of Nonlinear Dynamics and Complexity, Chapter 3, Vol. 1, H. G. Schuster, ed. (Weinheim, Wiley VCH), p. 69.

Drossel, B., and Schwabl, F. (1992). Self-organized critical forest-fire model. *Phys. Rev. Lett.* 69, 1629.

Graham, I., and Matthai, C. C. (2003). Investigation of the forest-fire model on a small-world network. *Phys. Rev. E* 68, 036109.

Guimera, R., and Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895.

Guimera, R., Uzzi, B., Spiro, J., and Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 697–702.

Han, J. D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., and Roth, F. P. (2004). Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 430, 88–93.

Hethcote, H. W. (2000). Mathematics of infectious diseases. *SIAM Rev.* 42, 599.

Ipsen, M., and Mikhailov, A. (2002). Evolutionary reconstruction of networks. *Phys. Rev. E* 66. 046109.

Kaiser, M., Görner, M., and Hilgetag, C. (2007). Criticality of spreading dynamics in hierarchical cluster networks without inhibition. *New J. Phys.* 9, 110.

Kearns, M., Suri, S., and Montfort, N. (2006). An experimental study of the coloring problem on human subject networks. *Science* 313, 824–827.

Lee, D. S., Goh, K. I., Kahng, B., and Kim, D. (2004). Sandpile avalanche dynamics on scale-free networks. *Physica A* 338, 84–91.

Levina, A., Ernst, U., and Herrmann, J. M. (2007a). Criticality of avalanche dynamics in adaptive recurrent networks. *Neurocomputing* 70, 1877–1881.

Levina, A., Herrmann, J. M., and Geisel, T. (2007b). Dynamical synapses causing self-organized criticality in neural networks. *Nat. Phys.* 3, 857–860.

Marr, C., and Hütt, M. T. (2005). Topology regulates pattern formation capacity of binary cellular automata on graphs. *Physica A* 354, 641–662.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., and Alon, U. (2004). Superfamilies of evolved and designed networks. *Science* 303, 1538–1542.

Moreno, Y., Pastor-Satorras, R., and Vespignani, A. (2002). Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* 26, 521.

Müller-Linow, M., Hilgetag, C. C., and Hütt, M. T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS Comput. Biol.* 4, e1000190. doi: 10.1371/journal.pcbi.1000190

Müller-Linow, M., Marr, C., and Hütt, M. T. (2006). Topology regulates the distribution pattern of excitations in excitable dynamics on graphs. *Phys. Rev. E* 74, 016112.

Pastor-Satorras, R., and Vespignani, A. (2004). Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge, Cambridge University Press.

Ravasz, E., and Barabási, A. L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E* 67, 1–7.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555.

Roxin, A., Riecke, H., and Solla, A. (2004). Self-sustained activity in a small-world network of excitableneurons. *Phys. Rev. Lett.* 92, 198101.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems* 85, 55–64.

Uchida, S., Drossel, B., and Brose, U. (2007). The structure of food webs with adaptive behaviour. *Ecol. Modell.* 206, 263–276.

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

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 March 2009; paper pending published: 14 May 2009; accepted: 07 August 2009; published online: 15 September 2009. Citation: Hütt M-T and Lesne A (2009) Interplay between topology and dynamics in excitation patterns on hierarchical graphs. Front. Neuroinform. 3:28. doi: 10.3389/neuro.11.028.2009

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 20 July 2009
doi: 10.3389/neuro.11.024.2009

Signal propagation in cortical networks: a digital signal processing approach

Francisco Aparecido Rodrigues[1] and Luciano da Fontoura Costa[1,2]*

[1] *Universidade de São Paulo, Instituto de Física de São Carlos, São Carlos, São Paulo, Brazil*
[2] *National Institute of Science and Technology for Complex Systems, Brazil*

Edited by:
Marcus Kaiser, Newcastle University, UK

Reviewed by:
Petra Ritter, Berlin NeuroImaging Center, Germany
Richard Gray, The University of Sydney, Australia

****Correspondence:***
Luciano da Fontoura Costa, Instituto de Fisica de Sao Carlos, Universidade de Sao Paulo, P.O. Box 369, 13560-970, Sao Carlos, São Paulo, Brazil.
e-mail: luciano@if.sc.usp.br

This work reports a digital signal processing approach to representing and modeling transmission and combination of signals in cortical networks. The signal dynamics is modeled in terms of diffusion, which allows the information processing undergone between any pair of nodes to be fully characterized in terms of a finite impulse response (FIR) filter. Diffusion without and with time decay are investigated. All filters underlying the cat and macaque cortical organization are found to be of low-pass nature, allowing the cortical signal processing to be summarized in terms of the respective cutoff frequencies (a high cutoff frequency meaning little alteration of signals through their intermixing). Several findings are reported and discussed, including the fact that the incorporation of temporal activity decay tends to provide more diversified cutoff frequencies. Different filtering intensity is observed for each community in those networks. In addition, the brain regions involved in object recognition tend to present the highest cutoff frequencies for both the cat and macaque networks.

Keywords: cortical networks, digital signal processing, graphs, networks

PACS numbers: 89.75.Fb, 87.18.-H, 02.10.Ox

INTRODUCTION

Brains are modular, interconnected structures optimized for transmission and processing of information at a level compatible with the survival and reproduction of each particular species (Hilgetag and Kaiser, 2004; Koch and Laurent, 1999; Sporns, 2002; Sporns et al., 2004). Information is progressively altered as it flows through the brain as a consequence of: (i) the processing performed by each individual neuron; (ii) the interconnection between the neurons along the path of the information flow, which implements the mixture of different signals; and (iii) interferences at the neurons or interconnection links (e.g. noise and cross-talk). While great attention has been focused on information processing in the brain, specially at the neuronal level, relatively fewer investigations have addressed the equally important issue of how signals are disseminated, while being integrated, through the several brain areas. Indeed, a great deal of the brain hardware (Sporns and Kötter, 2004), especially the white matter, is responsible for conveying signals along considerable distances from their origin, typically to several destinations, where they are modified, blended, and transmitted further.

Brain connectivity can be effectively represented, modeled and simulated in terms of graphs (e.g. Barabási and Albert, 2002). More specifically, each neuron or cortical region can be mapped as a node of a graph, while the synaptic or inter-regional connections are represented as directed links. Though a more complete understanding of information processing in the brain ultimately requires the integration of the non-linear processing taking place at each neuron, valuable insights can be nevertheless obtained by adopting some simpler (e.g. linear) dynamics and focusing on the interconnectivity and signal modification between neurons or cortical areas, as represented by graphs and networks (Watts and Strogatz, 1998).

Several linear approximations to non-linear problems have been reported in the literature, most of which related to linear synchronization dynamics (e.g. Zemanova et al., 2008; Zhou et al., 2006, 2007) and active media (e.g. Biktasheva et al., 2009; Ermentrout and Edelstein-Keshet, 1993; Hramov et al., 2005), yielding valuable insights about information transmission and processing. One element of particular importance in such investigations regards the interplay between structure and function. For instance, it has been established that the neural systems seem to form networks whose structures lie at the critical regime between local and global synchrony (Percha et al., 2005). In this way, the appearance of connections in damaged regions may lead to the onset of epileptic seizures (Nadkarni and Jung, 2003). It should be also observed that, as in the present work, linear approaches can be applied to model the collective dynamics of whole cortical regions, defining a more macroscopic investigation. In such cases, the explicit non-linearity of individual neuronal firing are averaged among several cells, yielding signals which are more graded and more propitious to being represented by linear approximations, especially during short periods of time. This property is ultimately one of the main justifications for the relatively large number of works in the literature in which brain activity is approached in terms of linear synchronization. It is also possible that the macroscopic propagation of cortical activation amongst different cortical regions could exhibit dynamics similar to traditional diffusion. Nevertheless, it should always be borne in mind that linear models of cortical activations may not reflect all the important dynamical features, especially those involving longer time intervals.

While synchronization is inherently important in the sense of being related to the brain workings, other linear approaches can be equally applied in order to reveal complementary aspects of the relationship between structure and function in the brain. One

particularly interesting possibility which has been mostly overlooked is the investigation of signal transmission and processing in terms of *signal processing* approaches. Founded on a well-established, sound mathematical framework, signal processing research (e.g. McClellan et al., 2002; Proakis and Manolakis, 2006) focuses on the representation and analysis of signals and systems in terms of frequencies and filters. The generality of such an approach stems from the fact that any real-world signal can be represented in terms of its respective Fourier series, namely a linear combination of basic harmonic components (sines and cosines) with different, well-defined frequencies. Linear systems typically modify such signals by changing the intensity of each component, such as in filters. For instance, a low-pass filter will attenuate the high-frequency harmonic components, while allowing the lower frequency components to pass with little or no alteration. The application of such an approach to signal transmission and processing in the brain paves the way to a series of promising possibilities. For instance, the alterations undergone by information as it proceeds from a specific origin neuron or cortical region to a specific target can be approximated as a kind of filter. Provided the properties of such a filter can be obtained, this approach allows modeling of the alterations undergone by the information while going from origin to destination. In other words, such a filter would replicate the functionality of the whole portion of brain hardware comprised between the origin and destination. In addition to its simplicity and elegance, such a filter modeling approach would also clearly characterize the way in which the information is altered in an intuitive and meaningful way, i.e. in terms of the alterations of the magnitudes of specific harmonic components. For instance, in case a specific portion of the brain is found to correspond to a low-pass filter, it becomes immediately clear that the high-frequency content of the signal is being attenuated, which corresponds to a smoothing operation implying loss of its details, therefore suggesting that that particular processing is focusing on the slower variations of the signal. In addition, low-pass filters are immediately related to the operation of integrating signals along time. In this respect, the smoothing could be a consequence of too intense mixing of several delayed versions of the signal, therefore providing valuable information about the level of blending of the signals as they passed through the network topology. Interestingly, the dynamical effect of low- and high-pass filters can be to a large extent summarized in terms of their respective *cutoff frequency*, namely the frequency where the attenuation reaches $1/\sqrt{2}$ of the amplitude of the largest harmonic component. Such an approach allows the function of the whole portion of brain in question to be effectively summarized in terms of a single real value. In the case of a low-pass filter, the higher the cutoff frequency, the smaller the alteration and intermixing undergone by the signal.

The current work describes a signal processing approach to the integration of brain structure and functionality which relies on the adoption of linear dynamics, namely diffusion. This type of dynamics underlies several natural systems and also participates in a large variety of non-linear dynamics (e.g. reaction-diffusion Giordano and Nakanishi, 2005). More specifically, at each time step, the signals arriving at each cortical region are added and redistributed among the respective outgoing links. The specific way in which such alterations take place are intrinsically related to the specific topology of the portion of the network comprised between the origin and destination nodes. In this way, the signals are blended as they are propagated along the brain in a way that is analogue to several sources of sound going through an environment as the sound signals reverberate and intermix, giving rise to constructive and destructive interferences. As such, this approach provides a nice complementation of other linear approximations to brain functionality, such as synchronization, by emphasizing the intermixing of signals as they progress through specific pathways along the intricate brain topology. Though we focus on cortical networks, this approach is immediately extensible to neuronal networks.

As reported recently (Rodrigues and da Fontoura Costa, 2009), non-conservative diffusion dynamics, more specifically the situation where each outgoing edge produces unit activation, in cortical networks can be effectively modeled in terms of finite impulse response digital filters (FIR). Interestingly, the coefficients of the FIR associated to a given network undergoing that type of dynamics are completely defined by the number of walks between the origin and destination areas, therefore establishing a clear-cut relationship between network structure and dynamics. The present work extends and explores these possibilities much further by assuming conservative diffusion with and without time decay in the cat and macaque cortical networks.

The manuscript starts by presenting the concepts of complex networks and digital signal processing, as well as the adopted cortical databases. Next, results obtained with respect to macaque and cat cortical networks are presented and discussed. The text concludes by reviewing the main contributions and identifying possibilities for further investigations.

CONCEPTS AND METHODS

NETWORKS AND DIFFUSION

A directed complex network, composed by a set of N nodes connected by E edges, can be represented by its adjacency matrix A, whose elements a_{ij} are equal to unity if the node j sends a connection to node i, and equal to zero otherwise. Two nodes i and j are said to be adjacent or neighbors if $a_{ij} \neq 0$. Two non-adjacent nodes i and j can be connected through a sequence of m edges $(i,n_1),(n_1,n_2),\ldots,(n_{m-1},j)$. Such a set of edges between i and j is called a *walk* of length m. The special case of a walk where no nodes are repeated is called a *path*.

A particular property of most complex networks is their community or modular structure. Communities are modules of densely interconnected nodes (Girvan and Newman, 2002). There are many methods for community identification and their choice depends on specific needs, e.g. accuracy against fast execution (da Fontoura Costa et al., 2007). In the current work, we considered the method based in random walks called *Walktrap* (Pons and Latapy, 2005), because such approach is intrinsically related to diffusion dynamics, which also underlies our modeling approach.

The characterization of the properties of a given network can be performed in terms of structural (e.g. da Fontoura Costa et al., 2007) and dynamical measurements (e.g. da Fontoura Costa and Rodrigues, 2008). Structural measurements includes, for instance, the node degree, clustering coefficient, average shortest path length and assortativity coefficient (da Fontoura Costa et al., 2007). On the other hand, dynamical measurements depend on the specific

dynamic process that is being executed in the network, such as synchronization (e.g. Pikovsky et al., 2002), random walks (e.g da Fontoura Costa and Sporns, 2006; da Fontoura Costa and Travieso, 2003), opinion formation (e.g Rodrigues and da Fontoura Costa, 2005) and epidemic spreading (e.g Newman, 2002). The structure and dynamics of complex networks are intrinsically inter-related (Boccaletti et al., 2006).

As shown in the current work, the relationship between structure and function of networks can also be addressed by using signal processing approaches. Signals are assumed to spread throughout the network by random walks initiating from a given source node (e.g. Barber and Ninham, 1970; Giordano and Nakanishi, 2005). Such a dynamical process, which is inherently related to diffusion (Barber and Ninham, 1970), involves the progressive dissemination and intermixing of the signals along time and network space, closely reflecting the specific topology of the network. Therefore, the signal arriving at a given destination node depends strongly on the structure of the portion of the network comprised between the source and destination nodes (da Fontoura Costa and Rodrigues, 2008). It has been shown recently (Rodrigues and da Fontoura Costa, 2009), with respect to a specific non-conservative diffusion dynamics, that such a strong interplay between network structure and dynamics can be fully modeled in terms of finite impulse response filters (FIRs). More specifically, the coefficients of such digital filters are given by the number of walks between the source and destination nodes.

The diffusion of activations in complex networks can be obtained by considering the transition matrix S, which can be calculated from the adjacency matrix A as

$$S(i,j) = \frac{A(i,j)}{\sum_{j=1}^{N} A(i,j)} \quad (1)$$

Each element $S(i,j)$ gives the probability of moving from the node i to node j. In this way, if a given signal is injected into a network, we can determine its diffusive propagation by repeatedly applying the transition matrix. More specifically, the probability of transition between the source and a destination at n edges of distance can be immediately obtained from the matrix

$$H_n = S^n \quad (2)$$

Similarly, the number of walks of length n between two nodes can be determined by the elements of the matrix $D_n = A^n$. The above dynamics is conservative, as there is no loss in the activations (all signals in the present work are formed by zeroes and ones). On the other hand, it is also possible to adopt a decay parameter that reduces the amplitude of the received activation along time. In this case, the matrix H_n, which gives the probability of transition between the source and destination nodes separated by walks of length n, is given as

$$H_n = \varepsilon(n+1)S^n \quad (3)$$

where $\varepsilon(n+1) = (1-\alpha)\varepsilon(n)$, $\varepsilon(1) = 1$ and $0 \le \alpha \le 1$. The coefficient α can be understood as the rate of decay according to the distance from the source of signal propagation. Such a dynamics, which is no longer conservative, has biological backing in the sense that sensory brain activations tend to diminish with time. In the current work, we show that the FIR approach to modeling the cortical networks can easily incorporate time decay, allowing the investigation of the diffusion dynamics without and with decay.

DISCRETE SIGNALS AND THEIR PROCESSING

A discrete-time signal is a time series consisting of a sequence of discrete values. The process of converting a continuous-valued discrete-time signal into a digital (discrete-valued discrete-time) signal is known as quantization (Orfanidis, 1996). A time-invariant system is a system that remains unchanged along time. This implies that if a given input is inserted into the system and causes a definite output, if we repeat the same process at another time, an equally delayed version of the previous output will be obtained. A linear, time-invariant system (LTI) can be fully classified in terms of its finite impulse response (FIR) and infinite impulse response (IIR), depending on whether the inserting signal has finite or infinite duration. More specifically, given the impulse response, the output produce for any input signal can be immediately calculated in terms of the convolution between the input signal and the impulse response.

A filter can be defined as any medium that can modify the signal in some way (Smith, 2007). A digital filter operates on discrete-time signals by taking a sequence of values (the input signal) and producing a new discrete-time signal (the filtered output signal). The main objective underlying the current work is to model the dynamics of signal transmission and integration between pairs of nodes in terms of digital signal processing concepts (McClellan et al., 2002; Orfanidis, 1996; Proakis and Manolakis, 2006). More specifically, the signal processing between pairs of nodes (source and destination of signal) is modeled as a FIR digital filter structure whose coefficients correspond to the total probability of transition of walks of different lengths between the source and destination (see **Figure 1**). This approach extends and complements a preliminary investigation assuming non-conservative diffusion dynamics, where the signals were propagated by using the adjacency, instead of transition, matrix (Rodrigues and da Fontoura Costa, 2009).

If a signal is injected into a network from a given source node i, the activation of each node at time t implied by the diffusion dynamics can be represented in terms of the system state vector (da Fontoura Costa, 2008)

$$\vec{y}(t) = (y_1(t), y_2(t), \ldots, y_N(t)) \quad (4)$$

where $y_j(t)$ represents the state of the node j at the time t. Given the state of a network at time t, the subsequent state can be calculated by

$$\vec{y}(t+1) = S\vec{y}(t) + \vec{s}(t) \quad (5)$$

where $\vec{s}(t) = (s_1(t), s_2(t), \ldots, s_N(t))$ is the vector representing the forcing signal injected at each node i. Note that in the case where only one node i receives activation at each time t, we have $s_i(t) = 1$, while all the other elements of $\vec{s}$ are equal to zero. The forcing signal $\vec{x}$ injected into the network is assumed to have length L and be composed of elements which are equal to zero or one. At each time step, one element of this vector is injected into the source node i, i.e. $s_i(t) = x(t)$. Thus, by considering this dynamics, signals

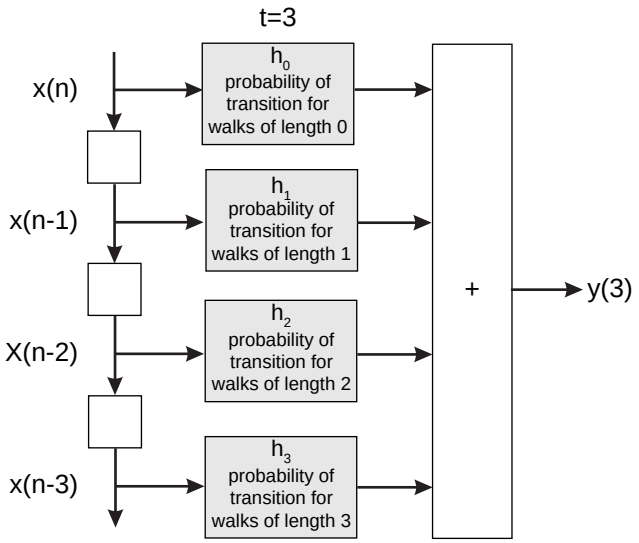


FIGURE 1 | The FIR model of the dynamics between a pair of nodes in a given network. The signal $\vec{x}(t)$ is injected into the network until $t = 3$. The signal at each time instant in the destination $\vec{y}(t)$ corresponds to a linear combination of the input signal and the coefficients $h(n)$ given by the probability of transition for walks of varying lengths between the source and destination.

are diffused, distributed and intermixed along the network. More specifically, the signal arriving at a node j after t time steps is a linear combination of the original signal values after undergoing all possible delays (smaller or equal to t) and combinations along the portion of the network comprised between the source and destination nodes, i.e.

$$y_j(t) = h(0)x(t) + h(1)x(t-1) + \ldots + h(t)x(0) \tag{6}$$

where $y_j(t)$ is the activation of the node j at the time t and the elements $h(n)$ represent the probability of transition between the source i and the destination j, considering all random walks of length t between the source and destination.

It can be easily verified that Eq. 6 corresponds to a FIR filtering structure, such as that illustrated in **Figure 1**. Indeed, this equation is equivalent to the convolution between the injected signal and the finite impulse response sequence $\vec{h}$ of a digital filter considering some initial period of time. Therefore, the dynamic of signal transmission between each pair of nodes i and j is effectively summarized, for a given finite period of time, by the respective FIR structure, which is completely specified in terms of the coefficients of $\vec{h}$, respectively given by Eq. 2. Thus, the coefficients defining the FIR structure are fully specified by the transition matrix describing the diffusion dynamics for each specific network topology.

The convolution above can be conveniently evaluated in terms of the z-transform. The z-transform converts a discrete time-domain signal, which is a sequence of real or complex numbers, into a complex domain representation. The z-transform is closely related to the Laplace transform, from which it can be obtained through the variable change $z = e^{st}$ (McClellan et al., 2002). As a consequence, the z-transform is also related to the Fourier transform. Given a discrete time signal $\vec{x}(t)$, its z-transform is defined as corresponding to the following series (Sirovich, 1988),

$$X(z) = \sum_{n=0}^{L} x(n) z^{-n} \tag{7}$$

We can recover $\vec{x}(t)$ from $X(z)$ by extracting the coefficient of the n-th power of z^{-1} and placing that coefficient in the t-th position in the sequence $x(t)$. Note that the inverse z-transform may not be unique unless its region of convergence is specified. The inverse z-transform can be computed using the contour integral (McClellan et al., 2002)

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} \mathrm{d}z \tag{8}$$

Among the main features of the z-transform that facilitate the analysis of linear systems we have: (i) linearity, (ii) delay representation, and (iii) the convolution property. In the case of the cortical networks, the FIR representation makes it clear how the existence of several paths of different lengths between the source and destination nodes, by defining distinct transition probabilities, completely specifies the functionality of the FIR as well as of the respective cortical network (Rodrigues and da Fontoura Costa, 2009). The convolution property is fundamental in FIR analysis, since it can be easily calculated by a simple multiplication in the transformed space, i.e.

$$y(t) = h(t) * x(t) \Rightarrow Y(z) = H(z)X(z) \tag{9}$$

where $h(t)$ is the impulse response sequence of a digital filter.

As the z-transform of a time-delay function $\delta(t-p)$ is known to be z^{-p}, we have that the system function $H(z)$ (i.e. the z-transform of the finite impulse response) for the network modeled as a FIR structure is given as

$$H(z) = \sum_{n=0}^{\Gamma} h(n) z^{-n} \tag{10}$$

where Γ is the FIR size and $h(n)$ represents the probability of transition between the source and the destination for walks of length n, i.e. the elements of $H_n = S^n$. **Figure 2** illustrates the dynamics in a network as modeled by the FIR approach.

Mathematically, the numerator of $H(z)$ has M roots (corresponding to the zeros of H) and the denominator has Q roots (corresponding to poles). The roots and poles of the system function $H(z)$ determine it to within a constant. In particular, as the FIR always have multiple poles at zero, the system is always stable (McClellan et al., 2002). The poles are the values of z at which $H(z)$ is undefined (infinite). The transfer function can be written in terms of poles and zeros

$$H(z) = \frac{(1-q_1 z^{-1})(1-q_2 z^{-1}) \cdots (1-q_M z^{-1})}{(1-p_1 z^{-1})(1-p_2 z^{-1}) \cdots (1-p_Q z^{-1})} \tag{11}$$

where q_n is the n-th zero and p_n is the n-th pole. The zeros and poles are commonly complex and, when plotted on the complex plane (z-plane), they define the so-called pole-zero plot (see **Figure 5**, for instance).

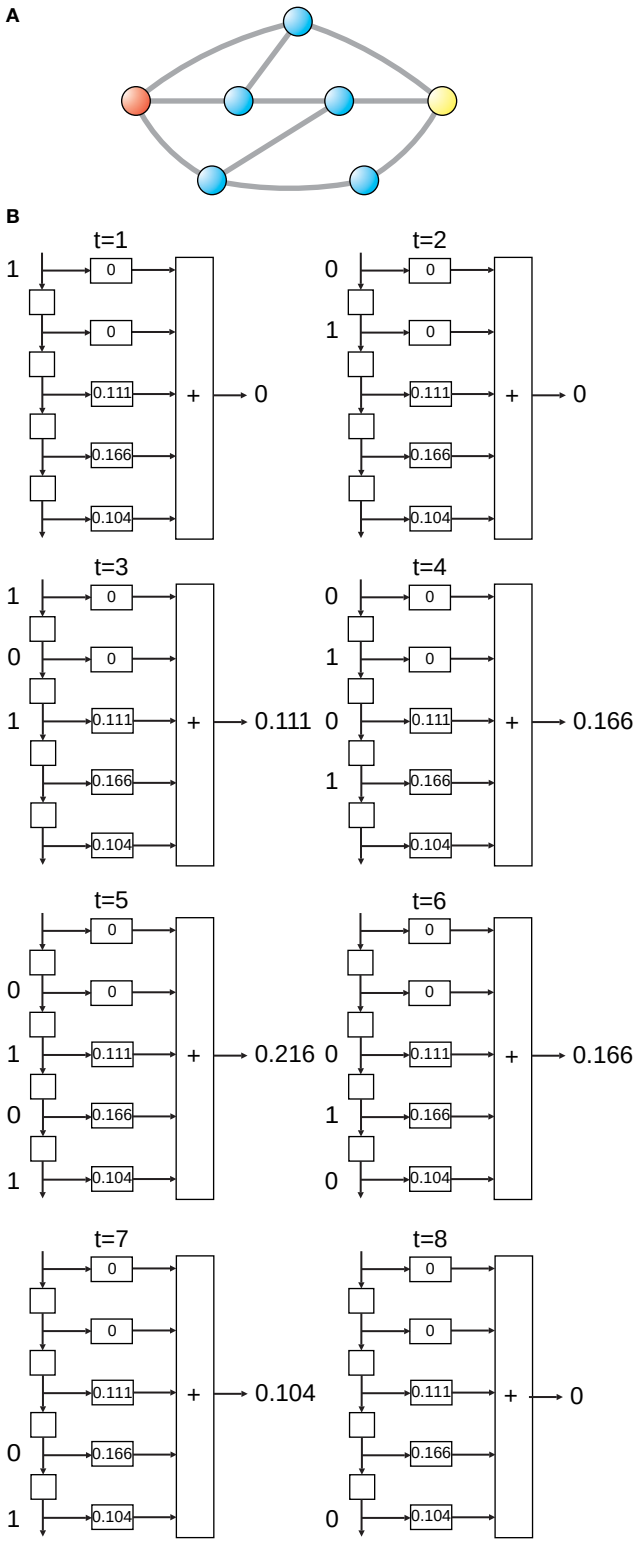


FIGURE 2 | Example of FIR modeling of the dynamics between a pair of nodes in the network (A). The signal $\vec{x} = (1, 0, 1, 0)$ is injected into the red node and the output is observed at the yellow node along the first 8 time steps. The output, $\vec{y} = (0, 0, 0.111, 0.166, 0.216, 0.166, 0.104, 0)$, can be obtained by considering the FIR structure, as illustrated in **(B)**.

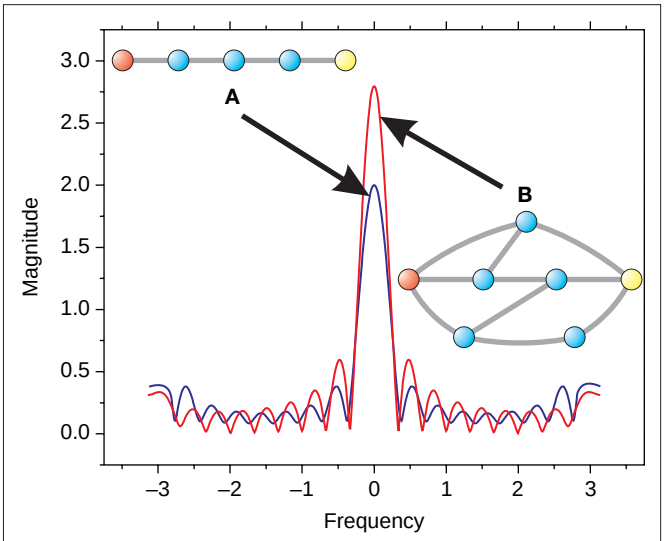


FIGURE 3 | Two different topologies that result in specific frequency responses. The input and output nodes are represented in red and yellow, respectively. In case (A), the sequence of nodes defining a chain present a low-pass filter with cutoff frequency value equal to $f_c = 0.18$. In the case (B), the low-pass filter presents $f_c = 0.13$. Therefore, the first case leads to less intermixing of differently delayed versions of the signal, and consequently a less intense alteration of the original signal.

It follows from the above results that all the intricacies of the diffuse dynamics in complex networks can be summarized in terms of the respective poles and zeroes of the system function. Moreover, the frequency response of the system, as well as the respective cut-off frequencies, can be immediately obtained from the pole-zero representation. The cutoff frequency, summarizes to a great extent the overall function of the respective low- or high-pass filter. The frequency response is highly dependent of the network structure, as shown in **Figure 3**. Indeed, the mixture of signals tends to reduce the amplitude of the frequency response and therefore the cut-off frequency. The frequency response is defined as the spectrum of the output signal divided by the spectrum of the input signal (Orfanidis, 1996; Smith, 2007). Observe that such formulations refer to the stationary state of the system, which is henceforth approximated by using several periods of a given input signal. The frequency response is typically characterized by the magnitude and phase of the system's response in terms of frequency. The frequency response magnitude is given by the transfer function $H(z)$ evaluated along the unit circle in the z-plane. In other words, the frequency response of a linear time-invariant system is equal to the Fourier transform of the impulse response (Smith, 2007).

CORTICAL NETWORKS

We investigate the cortical networks of macaque and cat, which contain predominantly isocortical brain regions (Sporns et al., 2007). All data sets consist of binary matrices describing the interconnectivity between the brain regions given by inter-regional pathways. The macaque network, including 47 nodes connected by 505 links, incorporates the visual, somatosensory and motor cortical regions (Felleman and Van Essen, 1991). The cat cortical network is derived

from the matrix published by Scannell et al. (1999), and excludes the hippocampus, all thalamic regions and the thalamo-cortical pathways (Sporns et al., 2007). This network contains 52 nodes and 818 links. All connections in these networks are directed.

The communities were identified by using the *Walktrap* method (Pons and Latapy, 2005), which is founded on random walk dynamics. Both cat and macaque cortical networks were split into four communities characterized by modularity (e.g Newman and Girvan, 2004) equal to $Q = 0.25$ for the cat, and $Q = 0.28$ for the macaque. The adjacency matrices of cat and macaque with the highlighted community connections are shown in **Figure 4**. We named the communities in terms of their main functions. In the case of the cat, the identified communities are formed by the following cortical regions:

- *Cognitive*: area 20b, area 7, anterior ectosylvian sulcus, posterior part of the posterior ectosylvian gyrus, medial area 6, lateral area 5B, infralimbic medial prefrontal cortex, dorsal medial prefrontal cortex, lateral prefrontal cortex, agranular insula, granular insula, anterior cingulate cortex, posterior cingulate cortex, retrosplenial cortex, area 35 of the perirhinal cortex, area 36 of the perirhinal cortex, presubiculum, parasubiculum and postsubicular cortex, subiculum, and entorhinal cortex;
- *Visual*: area 17, area 18, area 19, posterolateral lateral suprasylvian area, posteromedial lateral suprasylvian area, anteromedial lateral, surpasylvian area, anterolateral lateral suprasylvian area, ventrolateral suprasylvian area, dorsolateral suprasylvian area, area 21a, area 21b, area 20a, and posterior suprasylvian area;
- *Auditory*: primary auditory field, secondary auditory field, anterior auditory field, posterior auditory field, ventroposterior auditory field, and temporal auditory field;
- *Sensory system*: area 3a, area 3b, area 1, area 2, second somatosensory area, fourth somatosensory area, area 4γ, areas 4f, 4sf and 4d; lateral area 6, medial area 5A, lateral area 5A, medial area 5B, inner (deep) suprasylvian sulcal region of area 5, outer suprasylvian sulcal region of area 5.

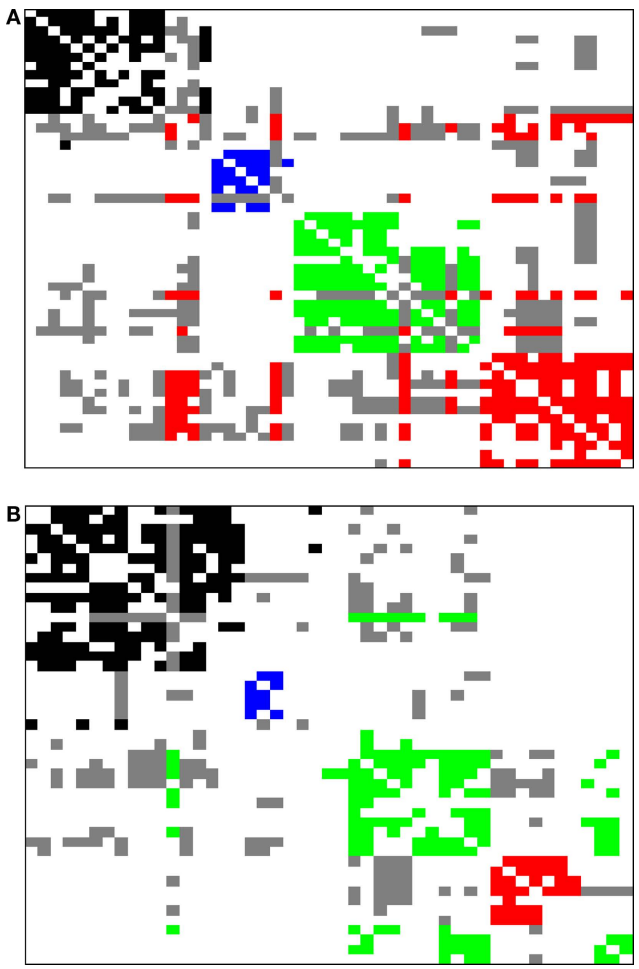


FIGURE 4 | The adjacency matrix of (A) cat and (B) macaque. For the cat cortical network, the colors represent the following communities: (i) black: cognitive, (ii) blue: visual, (iii) green: auditory, and (iv) red: sensory system. For the macaque, (i) black: memory, (ii) blue: visual, (iii) green: motor, and (iv) red: detection of movement. Connections between communities are shown in gray.

The communities identified in the macaque cortical network are:

- *Memory*: area 35, area 36, area 46, area 5, insular cortex, area 6, area 7a, area 7b, anterior inferotemporal (dorsal), frontal eye field, insular cortex (granular), medial dorsal parietal, medial intraparietal, retroinsular cortex, superior temporal polysensory (anterior), superior temporal polysensory (posterior), TF, and TH;
- *Visual*: anterior inferotemporal (ventral), central inferotemporal (dorsal), central inferotemporal (ventral), posterior inferotemporal (dorsal), posterior inferotemporal (ventral), visual area 4, and ventral occipitotemporal;
- *Motor*: area 1, area 2, area 3a, area 3b, area 4, secondary somatosensory area, and supplemental motor area;
- *Detection of movement*: dorsal preluneate, floor of superior temporal, lateral intraparietal, medial superior temporal (dorsal), medial superior temporal (lateral), middle temporal, posterior intraparietal, parieto-occipital, visual area 1, visual area 2, visual area 3, visual area V3A, v4 transitional, and ventral intraparietal.

RESULTS AND DISCUSSION

We start by illustrating the several concepts of the digital signal processing approach to the cortical networks. A signal of length 20 was injected into the largest hub of the cat (posterior cingulate cortex, CGp) and macaque (Visual area 4, V4) networks. Such hubs were chosen for this first experiment because they tend to act as connectors between different cortical regions (Sporns et al., 2007). The length of the signal was chosen so as to be larger than the diameter of the cortical networks, which is equal to four in both cat and macaque. **Figure 5** presents the zeros and poles obtained with respect to having the destination at each of the nodes in the cat and macaque, given this specific signal length, without time decay. The nodes in **Figure 5** are color-coded according to the respective community to which they belong. Recall that the values of the zeros

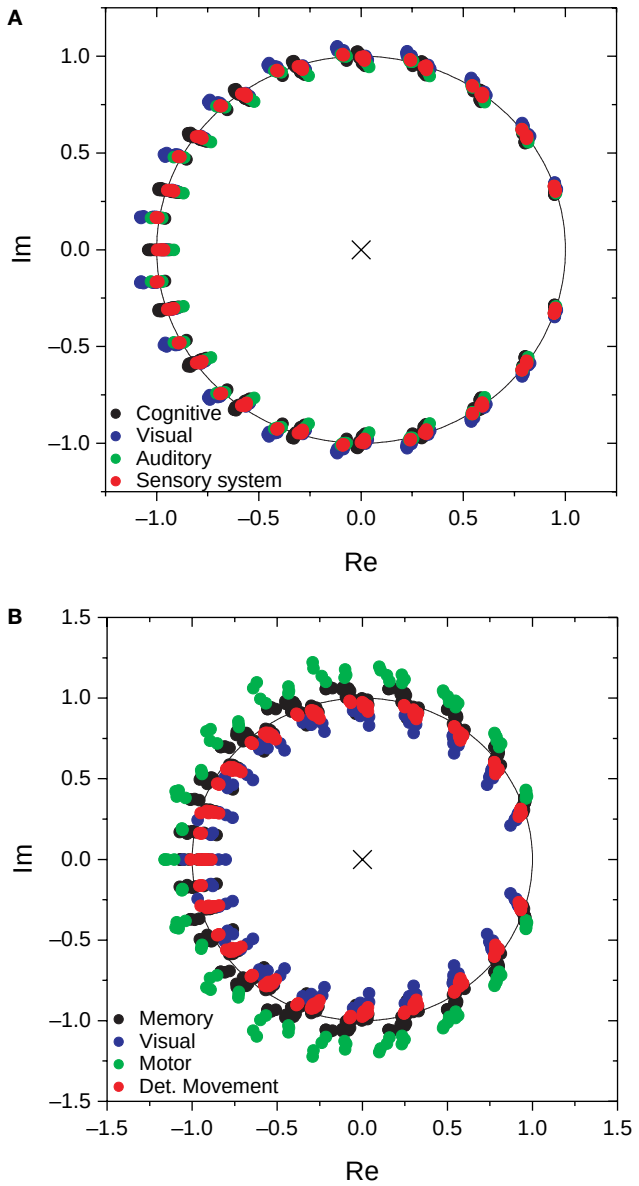


FIGURE 5 | The zeroes and poles obtained for the cat (A) and macaque (B) cortical networks. The colors identify the communities respective to each zero. The poles is represented by ×. The unit radius circle is shown by the continuum line.

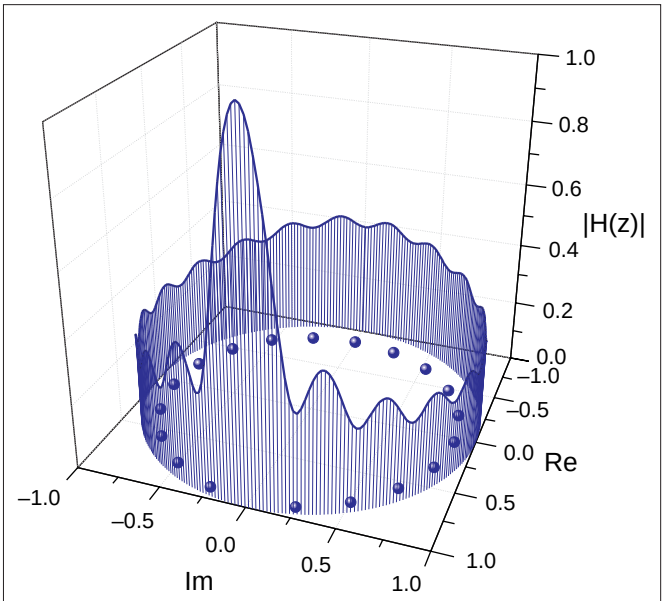


FIGURE 6 | Relationship between the zeroes of the system function and the frequency response (magnitude). The balls represent the zeroes of the system function obtained from a signal transmitted between the region v4 and central inferotemporal regions, in the macaque cortical network. The continuum line represents the *z*-transform evaluated over the unit radius circle.

and poles depend only of the FIR size and the network topology between the source and destination, and not of the specific content of the injected signal. Though the positions of the zeroes are similar in both cat and macaque, a wider dispersion is observed for the latter case. Interestingly, the zeroes found for each community tend to have similar positions in the complex plan, implying that those regions receive versions of the original signal modified in similar ways.

Figure 6 shows the frequency response curve (magnitude) obtained for one of the pairs of nodes of the macaque network, together with the respective zeroes, which are internal to the unit radius circle. The fact that such curves are determined by the zeroes and poles is clear from this figure, where each low valley along the frequency response is associated to a respective zero in the complex plan (Im, Re). After all, by definition the zeroes are the values of z for which the system function $H(z)$ becomes zero. The maximum magnitude along the unit circle is obtained between the two most spatially separated zeroes, at the lowest frequencies.

Figure 7 shows the frequency response curves (magnitude) obtained for the previous configuration, i.e. with the signal injected at the largest hubs of each network. The frequency responses obtained for the cat (**Figure 7A**) are remarkably similar to one another. This is to a great extent a consequence of the intense uniformity and high density of the connections characterizing this specific network. A much more varied set of curves is observed for the macaque (**Figure 7B**), suggesting a greater diversity of cortical organization and functioning. The curves obtained for each community tend to appear clustered, reflecting their similar zeroes positions. Both networks are characterized by intense low-pass filtering, revealing strong smoothing and mixing of the original signal.

In order to analyze a more realistic situation, we considered a decay parameter that reduces the amplitude of the system state along time. In this case, the matrix H_n, which gives the probability of transition between the source and destination separated by walks of length n, and therefore defines the coefficient of the system function $H(z)$, is given by Eq. 3. The parameter α specifies the intensity of the decay. It is henceforth adopted that $\alpha = 0.25$. **Figure 8** presents the zeroes and poles obtained for the cat and macaque after injecting a signal of length 20 into the CGp and V4 regions of cat and macaque, respectively. Observe that this corresponds to the same situation as above, but now with time decay. The obtained results are similar

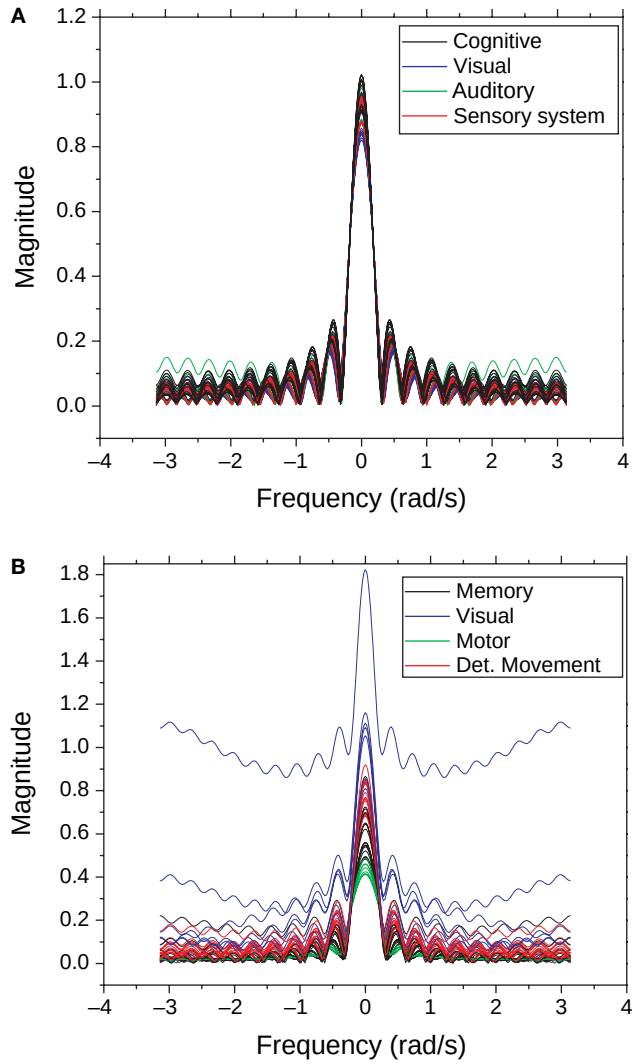


FIGURE 7 | The frequency response without cutoff for the cat (A) and macaque (B). Each line represents the frequency response of a destination node, and each color represents a different community.

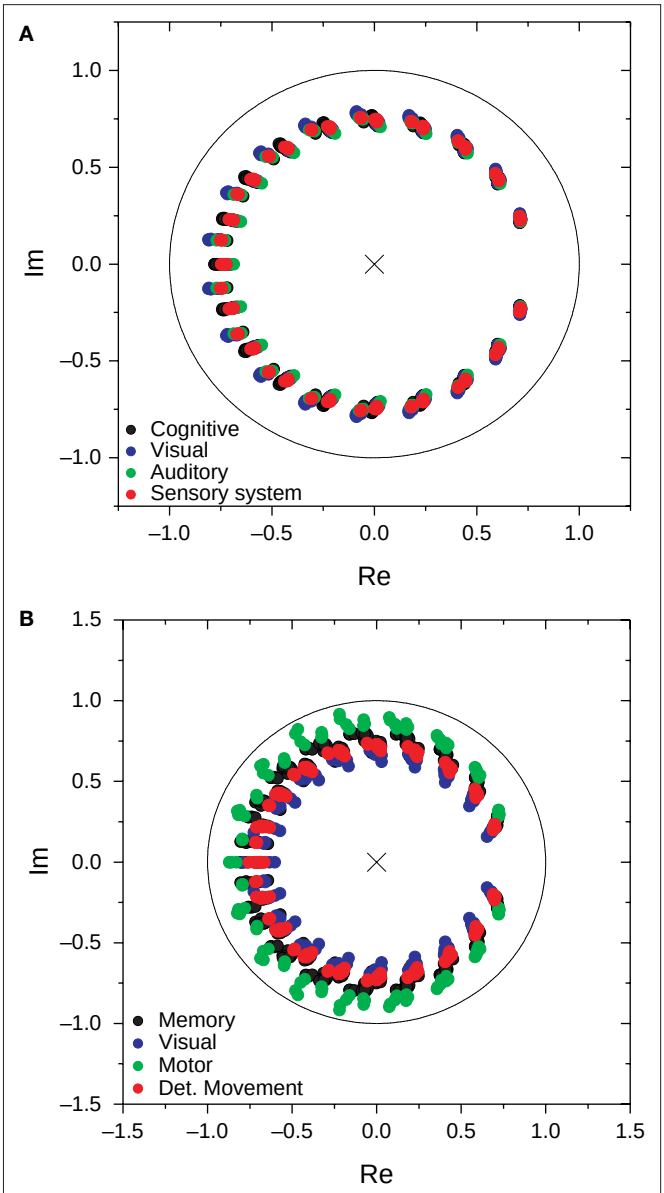


FIGURE 8 | The zeroes and poles obtained for the cat (A) and macaque (B) cortical networks considering time decay $\alpha = 0.25$. The communities are identified by different colors.

to those obtained previously (e.g. **Figure 5**), except that the zeroes were displaced towards the center of the complex plan. **Figure 9** shows the obtained frequency response curves (magnitude) for the cat and macaque cortical networks. It is clear from these curves that the time decay promotes diversity of filter action, as revealed by the more diverse curve shapes. Note that the time decay makes the zeroes to move away from the unit circle towards the origin in the zero-pole plot (compare **Figures 5 and 8**), therefore changing the respective gain magnitude.

The outlier in **Figure 9B**, which presents the highest magnitudes, corresponds to the central inferotemporal (dorsal) region, which happens to be connected to the input region (v4) and two other hubs, i.e. posterior inferotemporal (ventral) and anterior inferotemporal (dorsal).

From each of the frequency response (magnitude) curves, we can determine the respective cutoff frequency. For generality's sake, we assume signals being injected from all vertices of the networks (one at each simulation), instead of only from the largest hubs. **Figures 10 and 11** present the distribution of the cutoff frequencies without and with decay, respectively. The cutoff frequencies obtained for both the cat and macaque networks without decay are rather similar, agreeing with the similar zeroes positions identified previously. However, the cutoff frequencies obtained in presence of time decay exhibit greater diversity, which is a consequence of the displacement of the zeros inwards the unit circle. As shown in the insets of **Figure 11**, although the shape of the cutoff distributions obtained for the cat and macaque are visually similar, only a clear power-law degree distribution has been verified only for the cat. This result was obtained by applying the method proposed by Clauset et al.

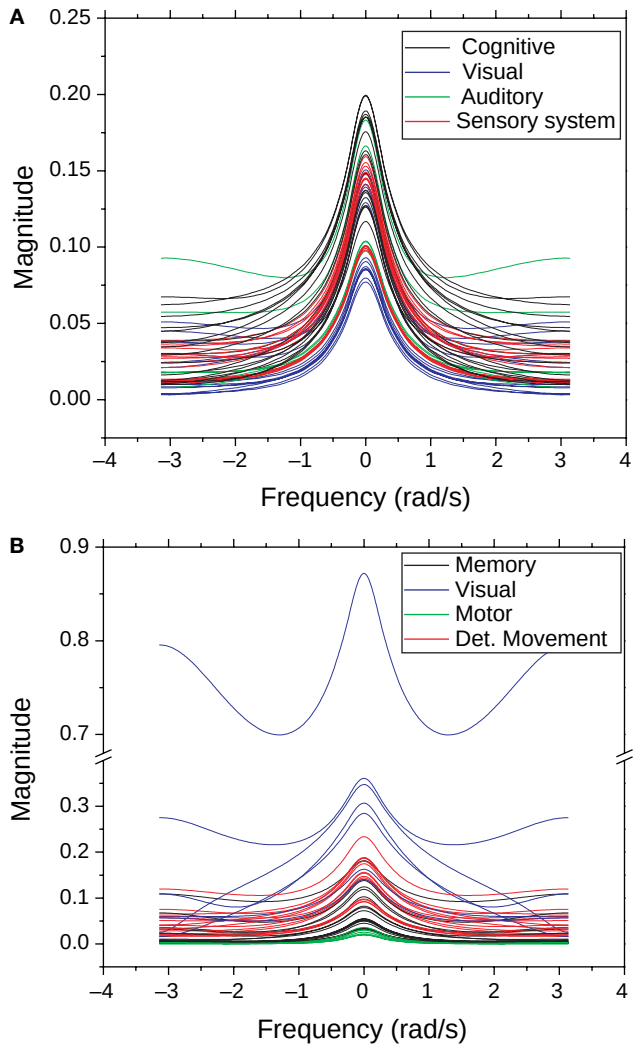


FIGURE 9 | The frequency response considering cutoff $\alpha = 0.25$ for the cat (A) and macaque (B). Each curve represents the frequency response (magnitude) of a node, while each color identifies the respective community.

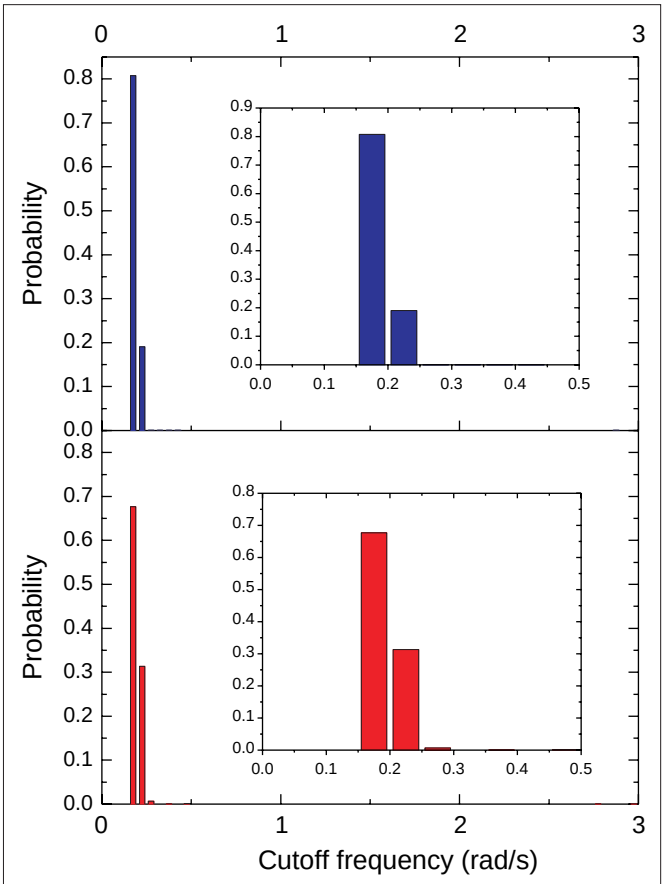


FIGURE 10 | The distribution of the cutoff frequencies for the cat and macaque networks without decay. The inset represents the zoom of the region that concentrates the greatest part of the distribution.

(2009), which combines maximum-likelihood fitting methods with goodness-of-fit tests based on the Kolmogorov–Smirnov statistic and likelihood ratios. In this way, we used a maximum likelihood estimators for fitting the power-law distribution of the cutoff frequencies obtained for the cat and macaque networks and calculated the p-value through the Kolmogorov–Smirnov test. The obtained p-value for the cat was equal to $p = 0.2$ and for the macaque, $p = 0$ (values larger than 0.05 indicates a power-law distribution). Thus, while the cat cortical networks present cutoff frequencies that follows a power law ($P(x) \sim x^{-\gamma}$) with coefficient $\gamma = 3.72$), the macaque does not present such feature. In fact, we tested other distributions, including exponential, log-normal, stretched exponential, and power law with cutoff, and none of the revealed to be a suitable fitting to the distribution of the cutoff frequencies of the macaque. Though the macaque and cortical networks differ with respect to the distribution of the cutoff frequencies, both networks present a high variability in the cutoff frequency values. More specifically, most nodes present small cutoffs, while a few of the present high cutoffs. Therefore, the majority of the cortical areas receive signals with a high degree of modification.

In order to perform a more detailed analysis of the cutoff frequencies characterizing each community, we determined their cumulative distribution in the macaque and cat networks. Several simulations were performed while injecting signals from all nodes in the network and monitoring the response for nodes inside each community. **Figure 12A** shows the cumulative distributions of cutoff frequencies obtained for the cat cortical network with respect to each community, considering time decay (the situation without decay is not discussed here because of its uniform response). The signals arriving at the sensory system, which include the somatosensory areas, are heavily filtered, indicating greater respective modifications and blending of differently delayed versions of the original signal. This suggests that signals coming from sensory modalities such as touch, temperature, proprioception (body position), and nociception (pain) are strongly intermixed, eliminating higher frequencies. The community involved in cognition, which includes the ectosylvian gyrus, prefrontal cortex, insula, cingulate cortex, perirhinal cortex and entorhinal cortex also presents small cutoff frequency and therefore receives strongly mixed versions of the

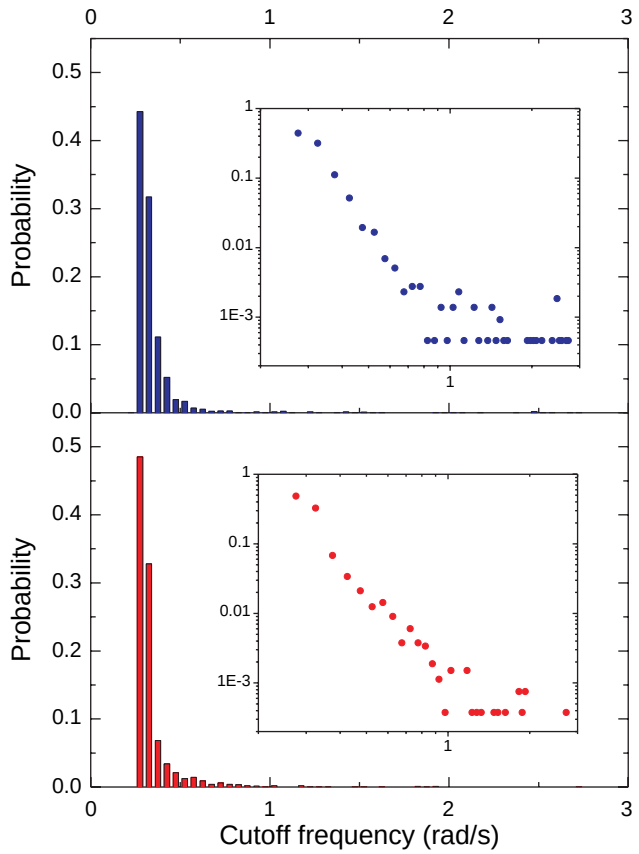


FIGURE 11 | The distribution of the cutoff frequencies for the cat and macaque in presence of decay with $\alpha = 0.25$. The insets show the respective log–log curves.

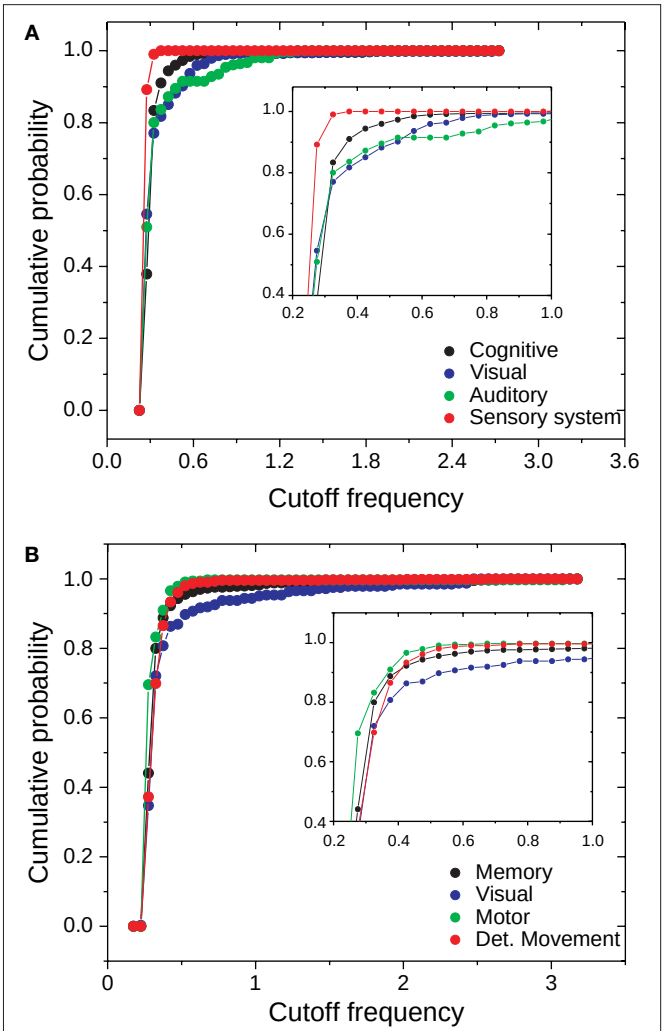


FIGURE 12 | The cumulative distribution of the cutoff frequencies for the cat (A) and macaque (B) communities considering time decay with $\alpha = 0.25$.

original signal. On the other hand, communities 2 and 3, involved in perception of complex motion (Rudolph and Pasternak, 1996) and representing the auditory field, respectively, tend to receive signals with the overall smallest modifications, and therefore smallest degrees of modifications and blending.

In the case of the macaque cortical network, shown in **Figure 12B**, the visual community, representing the inferotemporal and ventral occipitotemporal areas, receives signals with the lowest level of changes, with the highest cutoff frequencies. The inferotemporal area is thought to be the final visual area in the ventral stream of cortical areas responsible for object recognition (Tanaka, 1996). The same effect is observed in the occipitotemporal cortical areas of the macaque, which are known to be important for normal object recognition and for selective attention (Walsh and Perrett, 1994). On the other hand, the motor area community, i.e. motor and somatosensory areas, which are highly integrated one another (Kaas, 2004), tends to receive highly modified versions of the original signal. The movement detection community, which incorporates the somatosensory cortex, perirhinal cortex, insular cortex, parietal cortex, intraparietal cortex, polysensory and frontal eye field, also receives signals with high levels of alterations and intermixing.

In addition to the analysis of signal transmissions and intermixing with respect to communities, we can also systematically investigate the transmission of signals between different cortical areas. The cutoff frequencies were determine between all pairs of nodes in the cat and macaque networks. We determined the 20 connections that result in the highest cutoff frequencies for the cat and macaque cortical networks, therefore corresponding to the smallest levels of signal alternations and intermixing. **Table 1** presents the input and output nodes, as well as the respective communities that they are included, yielding the highest cutoff frequencies among all combinations of nodes for the cat cortical network. The suprasylvian area (localized in the visual community) is the dominant motion-processing region of the parietal cortex (Shen et al., 2006), being sensitive to texture and the distance between edges defined by motion (Robitaille et al., 2008). Areas in the cognitive community are involved in cognitive performance (infralimbic medial prefrontal cortex (van Aerde et al., 2008)), specific roles in the cognitive functions and pathological deficits of the hippocampal formation (subiculum area de la Prida et al., 2006), as well as spatial memory which helps to reduce errors when navigating

Table 1 | The 20 connections with the highest cutoffs (*cf*) in the cat cortical network.

Input	Com.	Output	Com.	*cf*
Posterolateral lateral suprasylvian area	Visual	Dorsolateral suprasylvian area	Visual	2.74
Presubiculum, parasubiculum and postsubicular cortex	Cognitive	Subiculum	Cognitive	1.94
Subiculum	Cognitive	Area 35 of the perirhinal cortex	Cognitive	1.94
Presubiculum, parasubiculum and postsubicular cortex	Cognitive	Entorhinal cortex	Cognitive	1.86
Ventrolateral suprasylvian area	Visual	Dorsolateral suprasylvian area	Visual	1.84
Posterior auditory field	Auditory	Retrosplenial cortex	Cognitive	1.84
Dorsolateral suprasylvian area	Visual	Ventrolateral suprasylvian area	Visual	1.64
Subiculum	Cognitive	Infralimbic medial prefrontal cortex	Cognitive	1.54
Area 21b	Visual	Dorsolateral suprasylvian area	Visual	1.45
Ventroposterior auditory field	Auditory	Anterior auditory field	Auditory	1.34
Dorsolateral suprasylvian area	Visual	Anterolateral lateral suprasylvian area	Visual	1.25
Secondary auditory field	Auditory	Temporal auditory field	Auditory	1.24
Anterior auditory field	Auditory	Primary auditory field	Auditory	1.16
Anterior auditory field	Auditory	Secondary auditory field	Auditory	1.16
Anterior auditory field	Auditory	Ventroposterior auditory field	Auditory	1.16
Ventroposterior auditory field	Auditory	Primary auditory field	Auditory	1.16
Anterior auditory field	Auditory	Posterior auditory field	Auditory	1.04
Ventroposterior auditory field	Auditory	Posterior auditory field	Auditory	1.04
Secondary auditory field	Auditory	Primary auditory field	Auditory	1.04
Secondary auditory field	Auditory	Ventroposterior auditory field	Auditory	1.04

Table 2 | The 20 connections with the highest cutoffs (*cf*) in the macaque cortical network.

Input	Com.	Output	Com.	*cf*
Insular cortex	Memory	Secondary somatosensory area	Motor	3.14
Medial superior temporal (lateral)	det. movement	Superior temporal polysensory (posterior)	Memory	3.14
Anterior inferotemporal (dorsal)	Memory	Area 7a	Memory	2.74
Anterior inferotemporal (dorsal)	Memory	Area 46	Memory	2.66
Dorsal preluneate	det. movement	Posterior intraparietal	det. movement	2.56
Central inferotemporal (ventral)	Visual	Superior temporal polysensory (posterior)	Memory	2.54
Ventral Posterior	det. movement	Posterior inferotemporal (dorsal)	Visual	2.46
Posterior inferotemporal (ventral)	Visual	Central inferotemporal (ventral)	Visual	2.46
Visual area 4	Visual	Posterior inferotemporal (dorsal)	Visual	2.46
Ventral occipitotemporal	Visual	Posterior inferotemporal (ventral)	Visual	2.46
Central inferotemporal (dorsal)	Visual	Anterior inferotemporal (ventral)	Visual	2.36
Anterior inferotemporal (ventral)	Visual	TH	Memory	2.16
Area 35	Memory	Insular cortex	Memory	2.06
Ventral occipitotemporal	Visual	Ventral Posterior	det. movement	2.04
Anterior inferotemporal (dorsal)	Memory	Central inferotemporal (ventral)	Visual	1.96
Anterior inferotemporal (dorsal)	Memory	Posterior inferotemporal (ventral)	Visual	1.94
Ventral occipitotemporal	Visual	Posterior inferotemporal (dorsal)	Visual	1.64
Posterior inferotemporal (dorsal)	Visual	Anterior inferotemporal (dorsal)	Memory	1.56
Ventral Posterior	det. movement	Ventral occipitotemporal	Visual	1.54
Medial dorsal parietal	Memory	Area 7a	Memory	1.54

in the dark (retrosplenial cortex Cooper et al., 2001). The highest cutoffs are therefore observed for transmissions between regions involved mainly in visual, cognitive and audio processing.

Table 2 shows the 20 connections that result in the highest cutoff frequencies for the macaque cortical network. These regions are mainly related to object recognition and movement detection. Areas in communities 1 and 2, which include the visual, occipitotemporal and inferotemporal areas, are involved in image processing and object detection and recognition (DiCarlo and Maunsell, 2003; Felleman et al., 1997). The areas in the movement detection community are also related to visual tasks. For instance, the posterior intraparietal area is involved in visually guided, object-related and

hand movements (Shikata et al., 2003). The ventral posterior areas are also related to visual processing. At the same time, areas in memory community participate in object recognition and memory (e.g. the insular cortex (Bermudez-Rattoni et al., 2005)). Therefore, interconnections between visual processing-related regions in the macaque cortex tend be characterized by the smallest low-pass filtering modifications. This fact is related to a relatively small mixture of signals between the source and visual reals. A given cortical region presents a high cutoff frequency because there are little alterations of signals going from the source to such region. The alterations are mainly caused by dependent walks, where signals tend to mix. Therefore, the more independent the walks between the source and destination, higher the frequency cutoff.

CONCLUSIONS

The relationship between brain organization and function corresponds to one of the most fundamental and challenging issues in neuroscience currently. Linear dynamics approaches, such as synchronization (e.g. Zemanova et al., 2008; Zhou et al., 2006, 2007), have been extensively considered in order to investigate the structure-function paradigm in the brain. For instance, it has been observed that the onset of epileptic seizures can be induced by addition of random connections that tend to decrease the small-world character of the brain (e.g. Nadkarni and Jung, 2003; Percha et al., 2005).

In the current work, we described a methodology to investigate diffusive signal propagation and blending between pairs of areas in cortical networks in terms of digital signal processing concepts and methods. Under these assumptions, the whole dynamics of brain propagation between each pair of nodes (source and destination) can be described by the convolution between the input signal and the probabilities of transition for walks of different lengths between the respective source and destination, a processing which can be neatly summarized in terms of finite-impulse-response filters (FIRs). We applied the z-transform in order to effectively perform these convolutions in terms of products. This approach also paves the way to the recovery, under certain conditions, of the original signal given the respective FIR structure. In addition, the z-transform approach allows the identification of the zeroes and poles of the system function (the z-transform of the finite impulse response). This is important because the zeroes and poles define completely the system response, and therefore can be used for the characterization of the functionalities implemented by the diffusion in the respective cortical topologies.

The obtained dynamics for the cat and macaque cortical networks was found to correspond to low-pass filtering, which tends to attenuate high-frequency harmonic components and allow the lower frequency components to pass with little or no alteration. In this way, the signal alterations undergone between the source and destination node can be summarized in terms of their respective FIR cutoff frequency. By analyzing signals received at each community, it was found that the areas involved in object recognition tended to suffer the smallest modifications in both the cat and macaque networks. In addition, in the cat, the areas related to sound processing were also verified to receive signals with smaller modifications than the other regions.

The extension of the current work to other cortical networks, such as human and rat, is immediate. In addition, it would be interesting to investigate how failures and attacks to the original networks induce changes in the respective filtering. The validation of the proposed approach involves monitoring several specific brain regions while a known input is fed into a giver region. For instance, a known stimulus can be applied in the auditory system of a macaque and be measured at different brain regions by electrode insertion. A comparison between the input and output signals in such experiments could be used to validate our theory.

ACKNOWLEDGMENTS

Luciano da Fontoura Costa is grateful to FAPESP (05/00587-5) and CNPq (301303/06-1 and 573583/2008-0) for financial support. Francisco A. Rodrigues acknowledges FAPESP sponsorship (07/50633-9).

REFERENCES

Barabási, A. L., and Albert, R. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.

Barber, M. N., and Ninham, B. W. (1970). Random and Restricted Walks: Theory and Applications. New York, Gordon & Breach Science Publishers.

Bermudez-Rattoni, F., Okuda, S., Roozendaal, B., and McGaugh, J. L. (2005). Insular cortex is involved in consolidation of object recognition memory. *Learn. Mem.* 12, 447–449.

Biktasheva, I. V., Barkley, D., Biktashev, V. N., Bordyugov, G. V., and Foulkes, A. J. (2009). Computation of the response functions of spiral waves in active media. *Phys. Rev. E* 79, 56702.

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

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* arXiv:0706.1062, in press.

Cooper, B. G., Manka, T. F., and Mizumori, S. J. Y. (2001). Finding your way in the dark: the retrosplenial cortex contributes to spatial memory and navigation without visual cues. *Behav. Neurosci.* 115, 1012–1028.

da Fontoura Costa, L. (2008). Diffusion of time-varying signals in complex networks: a structure-dynamics investigation focusing the distance to the source of activation. *arXiv:0811.3783.*

da Fontoura Costa, L., and Rodrigues, F. A. (2008). Superedges: connecting structure and dynamics in complex networks. arXiv:0801.4068.

da Fontoura Costa, L., Rodrigues, F. A., Travieso, G., and Villas Boas, P. R. (2007). Characterization of complex networks: a survey of measurements. *Adv. Phys.* 56, 167–242.

da Fontoura Costa, L., and Sporns, O. (2006). Correlating thalamocortical connectivity and activity. *Appl. Phys. Lett.* 89, 013903.

da Fontoura Costa, L., and Travieso, G. (2003). Exploring complex networks through random walks. *Phys. Rev. E* 75, 016102.

de la Prida, L. M., Totterdell, S., Gigg, J., and Miles, R. (2006). The subiculum comes of age. *Hippocampus* 16.

DiCarlo, J. J., and Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* 89, 3264–3278.

Ermentrout, G. B., and Edelstein-Keshet, L. (1993). Cellular automata approaches to biological modeling. *J. Theor. Biol.* 160, 97.

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Felleman, D. J., Xiao, Y., and McClendon, E. (1997). Modular organization of occipito-temporal pathways: cortical connections between visual area 4 and visual area 2 and posterior inferotemporal ventral area in macaque monkeys. *J. Neurosci.* 17, 3185–3200.

Giordano, N., and Nakanishi, H. (2005). Computational Physics, 2nd Edn. Upper Saddle River, NJ, Prentice Hall.

Girvan, M., and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7821–7826.

Hilgetag, C. C., and Kaiser, M. (2004). Clustered organization of cortical connectivity. *NeuroInformatics* 2, 353–360.

Hramov, A. E., Koronovskii, A. A., Popov, P. V., and Rempen, I. S. (2005). Chaotic synchronization of coupled electron-wave systems with backward waves. *Chaos* 15, 013705.

Kaas, J. H. (2004). Evolution of somatosensory and motor cortex in primates. *Anat. Rec.* 281, 1.

Koch, C., and Laurent, G. (1999). Complexity and the nervous system. *Science* 284, 96–98.

McClellan, J. H., Schafer, R. W., and Yoder, M. A. (2002). Signal Processing First. Upper Saddle River, NJ, Prentice Hall.

Nadkarni, S., and Jung, P. (2003). Spontaneous oscillations of dressed neurons: a new mechanism for epilepsy? *Phys. Rev. Lett.* 91, 268101.

Newman, M. E. J. (2002). The spread of epidemic disease on networks. *Phys. Rev. E* 66, 016128.

Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69.

Orfanidis, S. J. (1996). Introduction to Signal Processing. Upper Saddle River, NJ, Prentice Hall.

Percha, B., Dzakpasu, R., Żochowski, M., and Parent, J. (2005). Transition from local to global phase synchrony in small world neural network and its possible implications for epilepsy. *Phys. Rev. E* 72, 31909.

Pikovsky, A., Rosenblum, M., Kurths, J., and Hilborn, R. C. (2002). Synchronization: a universal concept in nonlinear science. *Am. J. Phys.* 70, 655.

Pons, P., and Latapy, M. (2005). Computing communities in large networks using random walks. *Lect. Notes Comput. Sci.* 3733, 284.

Proakis, J. G., and Manolakis, D. K. (2006). Digital Signal Processing, 4th Edn. Prentice Hall.

Robitaille, N., Lepore, F., Bacon, B. A., Ellemberg, D., and Guillemot, J. P. (2008). Receptive field properties and sensitivity to edges defined by motion in the postero-lateral lateral suprasylvian (PLLS) area of the cat. *Brain Res.* 1187, 82–94.

Rodrigues, F. A., and da Fontoura Costa, L. (2005). Surviving opinions in Sznajd models on complex networks. *Int. J. Mod. Phys. C* 16, 1785–1792.

Rodrigues, F. A., and da Fontoura Costa, L. (2009). Diffusion of time-varying signals in cortical networks. In 16th International Conference on Digital Signal Processing, Santorini, Greece, July 5–7 2009. IEEE Conference Proceedings, pp. 1–5.

Rudolph, K. K., and Pasternak, T. (1996). Lesions in cat lateral suprasylvian cortex affect the perception of complex motion. *Cereb. Cortex* 6, 814–822.

Scannell, J. W., Burns, G., Hilgetag, C. C., O'Neil, M. A., and Young, M. P. (1999). The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex* 9, 277–299.

Shen, W., Liang, Z., Chen, X., and Shou, T. (2006). Posteromedial lateral suprasylvian motion area modulates direction but not orientation preference in area 17 of cats. *Neuroscience* 142, 905–916.

Shikata, E., Hamzei, F., Glauche, V., Koch, M., Weiller, C., Binkofski, F., and Buchel, C. (2003). Functional properties and interaction of the anterior and posterior intraparietal areas in humans. *Eur. J. Neurosci.* 17, 1105.

Sirovich, L. (1988). Introduction to Applied Mathematics. Springer.

Smith, J. O. (2007). Introduction to Digital Filters: with Audio Applications. Center for Computer Research in Music and Acoustics, Stanford University.

Sporns, O. (2002). Network analysis, complexity, and brain function. *Complexity* 8, 56–60.

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425.

Sporns, O., Honey, C. J., and Kötter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE* 2, e1049. doi: 10.1371/journal.pone.0001049.

Sporns, O., and Kötter, R. (2004). Motifs in brain networks. *PLoS Biol.* 2, 1910–1918. doi: 10.1371/journal.pbio.0020369.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.

van Aerde, K. I., Heistek, T. S., and Mansvelder, H. D. (2008). Prelimbic and infralimbic prefrontal cortex interact during fast network oscillations. *PLoS ONE* 3, 7. doi: 10.1371/journal.pone.0002725.

Walsh, V., and Perrett, D. I. (1994). Visual attention in the occipitotemporal processing stream of the macaque. *Cogn. Neuropsychol.* 11, 243–263.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature* 393, 440–442.

Zemanova, L., Zamora-Lopez, G., Zhou, C., and Kurths, J. (2008). Complex brain networks: from topological communities to clustered dynamics. *Pramana – J. Phys.* 70.

Zhou, C., Zemanova, L., Zamora, G., Hilgetag, C. C., and Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.* 97, 238103.

Zhou, C., Zemanova, L., Zamora-Lopez, G., Hilgetag, C. C., and Kurths, J. (2007). Structure–function relationship in complex brain networks expressed by hierarchical synchronization. *New J. Phys.* 9, 178.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 April 2009; paper pending published: 13 May 2009; accepted: 30 June 2009; published online: 20 July 2009.

*Citation: Rodrigues FA and da Fontoura Costa L (2009) Signal propagation in cortical networks: a digital signal processing approach. Front. Neuroinform. **3**:24. doi: 10.3389/neuro.11.024.2009*

frontiers in
NEUROINFORMATICS

REVIEW ARTICLE
published: 20 July 2009
doi: 10.3389/neuro.11.020.2009

Perception and hierarchical dynamics

***Stefan J. Kiebel*[1]*, Jean Daunizeau*[2] *and Karl J. Friston*[2]**

[1] *Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2] *Wellcome Trust Centre for Neuroimaging, UCL, London, UK*

Edited by:
Marcus Kaiser, Newcastle University, UK

Reviewed by:
Viktor Jirsa, Movement Science Institute CNRS, France
Michael Breakspear, The University of New South Wales, Australia

***Correspondence:**
Stefan J. Kiebel, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany.
e-mail: kiebel@cbs.mpg.de

In this paper, we suggest that perception could be modeled by assuming that sensory input is generated by a hierarchy of attractors in a dynamic system. We describe a mathematical model which exploits the temporal structure of rapid sensory dynamics to track the slower trajectories of their underlying causes. This model establishes a proof of concept that slowly changing neuronal states can encode the trajectories of faster sensory signals. We link this hierarchical account to recent developments in the perception of human action; in particular artificial speech recognition. We argue that these hierarchical models of dynamical systems are a plausible starting point to develop robust recognition schemes, because they capture critical temporal dependencies induced by deep hierarchical structure. We conclude by suggesting that a fruitful computational neuroscience approach may emerge from modeling perception as non-autonomous recognition dynamics enslaved by autonomous hierarchical dynamics in the sensorium.

Keywords: dynamic systems theory, recognition, perception, birdsong, speech, biological movement, environment, Bayesian inversion

INTRODUCTION

Although there have been tremendous advances in the development of algorithms and devices that can extract meaningful information from their environment, we seem still far away from building machines that perceive as robustly and as quickly as our brains. For example, in artificial speech recognition, Deng et al. (2006) summarize current technology with: 'The machine would easily break if the users were to speak in a casual and natural style as if they were talking with a friend.' The situation is similar in machine vision: Although highly specialized recognition devices exist; e.g., for face recognition (Tan et al., 2006; Zhao et al., 2003), there is no generally accepted computational principle for robust perception.

In artificial speech recognition, the conventional approach is to approximate the acoustic expression of speech by hidden Markov models (Bilmes, 2006; O'Shaughnessy, 2008). This scheme and its variants do not seem, by construction, to capture efficiently the long-range temporal and contextual dependencies in speech (O'Shaughnessy, 2008). However, a novel approach is emerging that suggests a fundamental computational principle: the idea is to model fast acoustic features of speech as the expression of comparatively slow articulator movement (Deng et al., 2006; King et al., 2007; McDermott and Nakamura, 2006). These models describe speech as a hierarchy of dynamic systems, where the lowest (fastest) level generates auditory output. Although this approach, due to its complexity, is still at an early stage of development, the premise is that hierarchical dynamics may provide valuable constraints on speech recognition. These could make artificial speech recognition systems more robust, in relation to conventional approaches, which do not embody hierarchical constraints efficiently. In the visual domain, similar hierarchical models have been considered for making inference on dynamic human behavior, such as those used in robot-human interaction or surveillance technology (Kruger et al., 2007; Moeslund et al., 2006; Oliver et al., 2004; Robertson and Reid, 2006; Saenko et al., 2005; Yam et al., 2004).

The question we address in this paper is whether these developments in hierarchical, trajectory-based perception models point to a computational principle which can be implemented by the brain. In Kiebel et al. (2008) we developed a simple recognition system, based on a specific functional form of hierarchical dynamics. We reprise the approach here to show it affords schemes for perception that are both robust to noise and can represent deep hierarchical structure in the sensory streams.

We consider three constraints on perception that the brain has to contend with. The first is that our environment and sensations are dynamic processes. This places computational demands on the speed of recognition and makes perception, at first glance, more formidable than recognizing static scenes or objects. However, a dynamic environment has temporal structure and regularities, which can be learned and may be beneficial for robust perception.

The second constraint is that the brain performs perception online, because it has no access to future sensory input and cannot store the details of past sensations (we assume here that the brain does not have the equivalent of computer memory, which could faithfully store the sensory stream for off-line processing). This means that transient sensory information must be used to represent the dynamic state of the environment. This constraint renders perception distinct from other analyses of time-series data, where timing is not critical and stored data can be analyzed off-line.

The third constraint is that we assume that the perception conforms to the free-energy principle (FEP); i.e., the perceptual system dynamically minimizes its free-energy and implicitly makes inferences about the causes of sensory input (Friston et al., 2006). To minimize its free-energy, the agent uses a generative model of how the environment produces sensory input. This formulation leads to the question 'what generative model does the brain use?' (Dayan et al., 1995; Lee and Mumford, 2003). Here,

we will review and discuss a hierarchical model for perception, where higher levels (further away from sensory input) encode the shape of attractors which contain faster, lower level dynamics (Kiebel et al., 2008). Previously we have shown in simulations, that this hierarchical model enables agents to recognize states causing sensory input, at two time scales. In this paper, we focus on the implications of hierarchical attractor models for artificial agents, for example speech recognition devices, and real brains. In particular, we introduce neurocomputational models of perception that emerge when one describes the dynamics of two systems (the environment and the agent) that are coupled via sensory input.

THEORY

In the following, we summarize a generative model based on a hierarchy of attractors and its variational inversion. In Kiebel et al. (2008), we used simulations to show that the inversion of these models shows a range of features which reproduce experimental findings in systems neuroscience. Here, we relate this model to research in artificial speech recognition.

A MODEL OF PERCEPTUAL INFERENCE

Human speech perception has been construed as the output of a multi-level hierarchical system, which must be decoded at different time-scales (Chater and Manning, 2006; Poeppel et al., 2008). For example, while a spoken sentence might only last for seconds, it also conveys information about the speaker's intent (an important environmental cause) that persists over much longer time-scales. To illustrate these points, we will simulate the recognition of birdsongs. We use this avian example to illustrate that communication entails (i) embedding information at various time-scales into sound-waves at a fast time-scale and (ii) that the recipient must invert a hierarchical dynamic model to recover this information. Our objective is to show that communication can be implemented using hierarchical models with separation of temporal scales. In the following, we describe a two-level system that can generate sonograms of synthetic birdsong and serves as a generative model for perception of these songs.

There is a large body of theoretical and experimental evidence that birdsongs are generated by dynamic, nonlinear and hierarchical systems (Glaze and Troyer, 2006; Sen et al., 2001; Vu et al., 1994; Yu and Margoliash, 1996). Birdsong contains information that other birds use for decoding information about the singing bird. It is unclear which features birds use to extract this information; however, whatever these features are, they are embedded in the song, at different time-scales. For example, at a long time-scale, another bird might simply register the duration of a song, which might belie the bird's fitness. At short time-scales, the amplitude and frequency spectrum of the song might reflect attributes of the bird or imminent danger.

A GENERATIVE BIRDSONG MODEL

In Kiebel et al. (2008), we described a system of two coupled Lorenz attractors, whose output was used to construct a sonogram and associated sound wave, which sounds like a series of chirps. The key point of this model is that, when generating output, the states of a Lorenz attractor at a slower time scale act as control parameters for another Lorenz attractor at a faster time scale. The model can be expressed as

$$\begin{aligned}\dot{x}^{(2)} &= f\left(x^{(2)}, v^{(2)}, T^{(2)}\right) + w^{(2)} \\ v^{(1)} &= x_3^{(2)} - 4 + z^{(2)} \\ \dot{x}^{(1)} &= f\left(x^{(1)}, v^{(1)}, T^{(1)}\right) + w^{(1)} \\ y &= \begin{bmatrix} x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} + z^{(1)}\end{aligned} \tag{1}$$

where, $v^{(i)}$ represent inputs to level i (or outputs from level $i+1$), which perturb the possibly autonomous dynamics among that level's states $x^{(i)}$. The nonlinear function f encodes the equations of motion of the Lorenz system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = f(x, v, T) = \frac{1}{T}\begin{pmatrix} -a & a & 0 \\ v - x_3 & -1 & 0 \\ x_2 & 0 & -c \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{2}$$

For both levels, we used $a = 10$ (the Prandtl number) and $c = 8/3$. The parameter T controls the speed at which the Lorenz attractor evolves; here we used $T^{(1)} = 0.25$ s and $T^{(2)} = 2$ s; so that the dynamics at the second level are an order of magnitude slower than at the first. At the second-level we used a Rayleigh number; $v^{(2)} = 32$. We coupled the fast to the slow system by making the output of the slow system $v^{(1)} = x_3^{(2)} - 4$ the Rayleigh number of the fast system. The Rayleigh number is effectively a control parameter that determines whether the autonomous dynamics supported by the attractor are fixed point, quasi-periodic or chaotic (the famous butterfly shaped attractor). The sensory signals generated are denoted by y, which comprises the second and third state of $x^{(1)}$ (Eq. 1). We will call the vectors $x^{(i)}$ hidden states, and the scalar $v^{(1)}$ the causal state, where superscripts indicate model level and subscripts refer to elements. At each level we modeled Gaussian noise on the causes and states ($w^{(i)}$ and $z^{(i)}$) with a log-precision (inverse variance), of eight (except for observation noise $z^{(1)}$, which was unity). We constructed the sonogram (describing the amplitude and frequency of the birdsong) by making $|y_1|$ the amplitude and y_2 the frequency (scaled to cover a spectrum between 2 and 5 kHz). Acoustic time-series (which can be played) are constructed by an inverse windowed Fourier transform. An example of the system's dynamics and the ensuing sonogram are shown in **Figures 2A,B**. The software producing (and playing) these dynamics and the sonogram can be downloaded as Matlab 7.7 (Mathworks) code (see software note).

This model can be regarded as a generative or forward model that maps states of the singing bird to sensory consequences (i.e., the sonogram). For human listeners, the resulting song sounds like a real birdsong. Given a generative model of birdsong, we can generate (different) songs and ask: How could a synthetic bird recognize these songs?

The online inversion of this forward model; i.e., the online reconstruction of the hidden and causal states, corresponds to perception or mapping from the sonogram to the underlying states of the singing bird. In this example, perception involves the online estimation of states at the fast and slow level. Although, at the fast first-level, two of the states (those controlling amplitude and frequency of the acoustic input) are accessed easily, the third $x_1^{(1)}$

describes a completely hidden trajectory. It is important to estimate this state correctly because it determines the dynamics of the others (see Eq. 2). Model inversion also allows the listening bird to perceive the slowly varying hidden states at the second level, $x^{(2)}$, which cannot be heard directly but must be inferred from fast sensory input. The second-level hidden states encode the high-order structure of the song by specifying the shape of the attractor at the first level. The ensuing inversion problem is difficult to solve because the bird can only infer states at both levels through the nonlinear, continuous and stochastic dynamics of the Lorenz attractor at the first level.

PERCEPTION USING VARIATIONAL INVERSION

In Kiebel et al. (2008), we showed how inversion of this hierarchical model can be implemented using the free-energy principle (Friston et al., 2006). This variational online inversion can be conceptualized as shown in **Figure 1**. The environment, here a synthetic bird, generates output using a hierarchical system with coupled slow and fast dynamics (Eqs 1 and 2). This generates sensory input that is recognized by the receiving bird. It does this by passing top-down messages (predictions) and bottom-up messages (prediction errors) between the levels of its generative model. When top-down messages from the first level predict sensory input, the hidden and causal states of the generative model become representations of the corresponding states of the singing bird and perceptual inference is complete. For mathematical details, we refer the interested reader to Friston et al. (2008).

SIMULATIONS OF BIRDSONG PERCEPTION

Here, we describe the result of a single simulation to show that the online inversion can successfully recognize songs and track the trajectories of the states at all levels. In Friston and Kiebel (2009) and Kiebel et al. (2008) we present more simulations, and discuss and relate them to perception, categorization and omission responses in the brain. In **Figure 2A** we plot the hidden and causal states, which produce sensory output corresponding to synthetic birdsong generation. One can see immediately that the two levels have different time-scales due to their different rate constants (Eqs 1 and 2). The resulting sonogram is shown in **Figure 2B**.

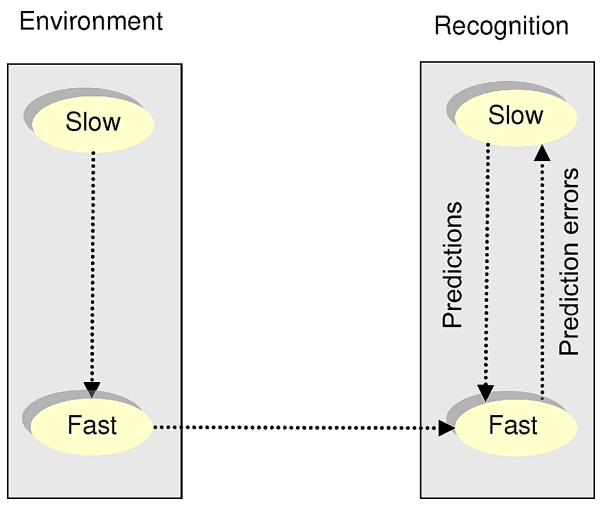


FIGURE 1 | Birdsong generation and its recognition using variational inversion. Environment (left): In this two-level birdsong model, sonograms are generated by the autonomous, coupled dynamics of two Lorenz attractors (see Eqs 1 and 2). The states of the first Lorenz attractor evolve at a slow time scale and act as control parameters for the faster Lorenz attractor. Perception system (right): The implicit variational dynamic inversion is a recurrent message passing scheme, where top-down predictions are sent from the slow level to the fast level, while the fast level receives sensory input and generates bottom-up prediction errors. The resulting recognition dynamics are non-autonomous and try to 'mirror' the environmental dynamics.

The results of online inversion (i.e., song recognition) are shown in **Figure 3**. At the first level, the uncertainty about the states was small, as indicated by narrow 90% confidence intervals, shown in grey. At the second level, the system tracks the hidden and causal states veridically. However, as these variables are inferred through the sensory data, uncertainty about the hidden state reaches, intermittently, high values. The uncertainty about the hidden states at the second-level is very high, because these variables can only be inferred via the causal state $v^{(1)}$. In particular, note the increased period of uncertainty at about 0.3 s, at both levels. This uncertainty is caused by the hidden state of the first-level switching between the 'wings' of the Lorenz attractor. At this point, the hidden state at the first level is less identifiable than when it is on the outer reaches of a wing. This is because of nonlinearities in the generative model, which mean, at this point, the motion of the state is a weaker function of the states *per se*. This uncertainty (i.e., will the state cross to the other wing or not?) is part of inference.

In summary, these results show that the hierarchical model can not only generate birdsong dynamics but, using the free-energy principle, it can be used as a generative model to decode incoming sensory input with relatively high precision. Critically, at the second level, the decoding (listening) bird infers hidden states that evolve slowly over time. This is an important result because the values of the hidden states at the second level specify the attractor manifold, and therefore the trajectory of states at the first. In other words, one location in state space at the higher level specifies a sequence of states at the lower. Moreover, because the states at the second level also follow a slowly varying trajectory, the attractor manifold at the first level keeps changing slowly over time. It is this slow modulation of the first-level manifold that expresses itself in the variations of the fast moving first-level state, which enable the perception to track hidden trajectories at the second level.

A key aspect of this model rests on the nonlinearity of the generative model. This is because the only way for slowly varying causes to be expressed as faster consequences is through nonlinear mechanisms (Eq. 2). It is this nonlinearity that allows high-level states to act as control parameters to reconfigure the motion of faster low-level states. If the equations of motion at each level were linear in the states, each level would simply convolve its supraordinate inputs with an impulse response function. This precludes the induction of faster dynamics because linear convolutions can only suppress various frequencies. However, the environment is nonlinear, where long-term causes may disclose themselves through their influence on the dynamics of other systems. To predict the ensuing environmental trajectories accurately, top-down effects in the agent's generative model must be nonlinear too. We suggest that this principle of separation of time scales in a nonlinear hierarchy is not only used in avian but also in human communication, because both birdsong and speech share the common feature that information is transmitted

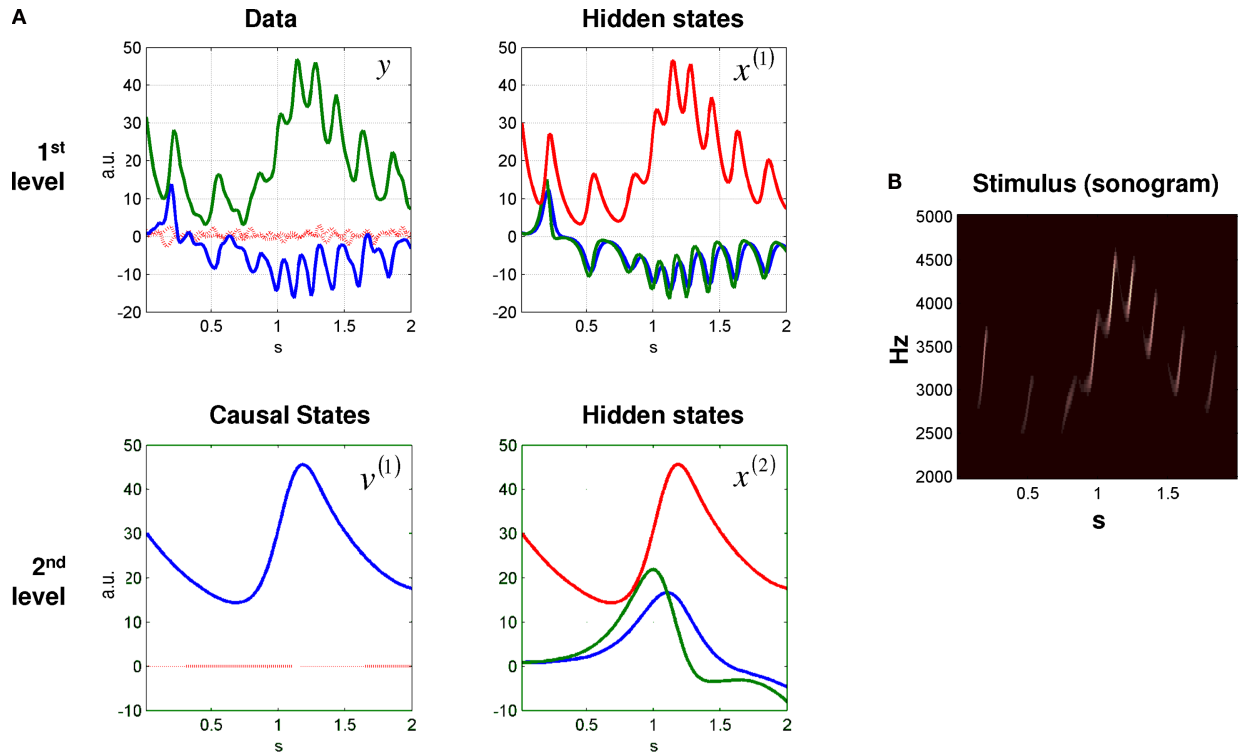


FIGURE 2 | Data and states, over 2 s, generated by a two-level birdsong model. (A) At the first level, there are two outputs (i.e., sensory data) (left: blue and green solid line) and three hidden states of a Lorenz attractor (right: blue, green, and red solid line). The second level is also a Lorenz attractor that evolves at a time-scale that is one magnitude slower than the first. At the second level, the causal state (left: blue solid line) serves as control parameter (Rayleigh number) of the first-level attractor, and is governed by the hidden states at the second level (right: blue, green, and red solid line). The red dotted lines (top left) indicate the observation error on the output. **(B)** Sonogram (time-frequency representation) constructed from model output. High intensities represent time-frequency locations with greater power.

via rapidly modulated sound waves. In the following, we will review evidence which suggests that human speech can be appropriately modeled and recognized by a hierarchy of attractors.

ARTIFICIAL SPEECH RECOGNITION

How are our simulations related to artificial perception systems that solve 'real-world' recognition tasks? Here, we focus on artificial speech recognition (ASR) but note that there are similar modeling initiatives in other areas of artificial perception; e.g., vision (Moeslund et al., 2006; Oliver et al., 2004; Yam et al., 2004).

An intuitive approach to speech recognition is to consider speech as a sequence of phonemes; i.e., speech sounds are like 'beads on a string', which form syllables, words and sentences (Ostendorf, 1999). The idea here is that when one knows the 'single beads', one just needs to read out the sentence. This intuition leads naturally to models that treat speech as a sequence of states, which can be recognized, given the auditory input, using hidden Markov models (Bilmes, 2006; O'Shaughnessy, 2008). However, speech does not seem to work like this: Speech exhibits all kinds of contextual effects, at various time-scales, leading to cross-temporal dependencies. For example, co-articulation induces a dependence of the acoustic expression of speech-sounds on the sound's temporal neighbors (Browman and Goldstein, 1992). These temporal dependencies introduce a tremendous amount of variations in the 'single beads'. In conventional hidden Markov models these can be modeled by increasing the number of states and parameters, which can lead to serious model identification issues: Various reviews discuss why the hidden Markov model and its extensions, as conventionally used in ASR, are probably not appropriate to model and recognize speech with human-like performance (Bilmes, 2006; King et al., 2007; O'Shaughnessy, 2008).

Although ignored as a main-stream modeling assumption in the ASR field, the acoustic stream is the consequence of hidden state-space trajectories: the vocal tract (VT) dynamics, i.e. tongue, mouth and lips and other VT components, generate articulatory gestures, which are understood to be the basic elements of speech (Browman and Goldstein, 1997; Deng et al., 2006; Liberman and Whalen, 2000; McDermott and Nakamura, 2006). A novel modeling approach, which seems to be emerging from the ASR field, focuses on two crucial points: First, the specification of a generative hierarchical speech model for recognition, which models VT dynamics as hidden trajectories. Second, these VT dynamics form speech 'gestures', whose perception is the goal of artificial speech recognition. There are many interesting variants of this approach, e.g. (Deng et al., 2006, 2007; Hofe and Moore, 2008; King et al., 2007; Livescu et al., 2007; McDermott and Nakamura, 2006; Rose et al., 1996; Saenko et al., 2005).

Such hierarchical generative models place fast acoustics at the lowest level, whereas (various levels of) VT dynamics causing the acoustics through top-down influences (Deng et al., 2006).

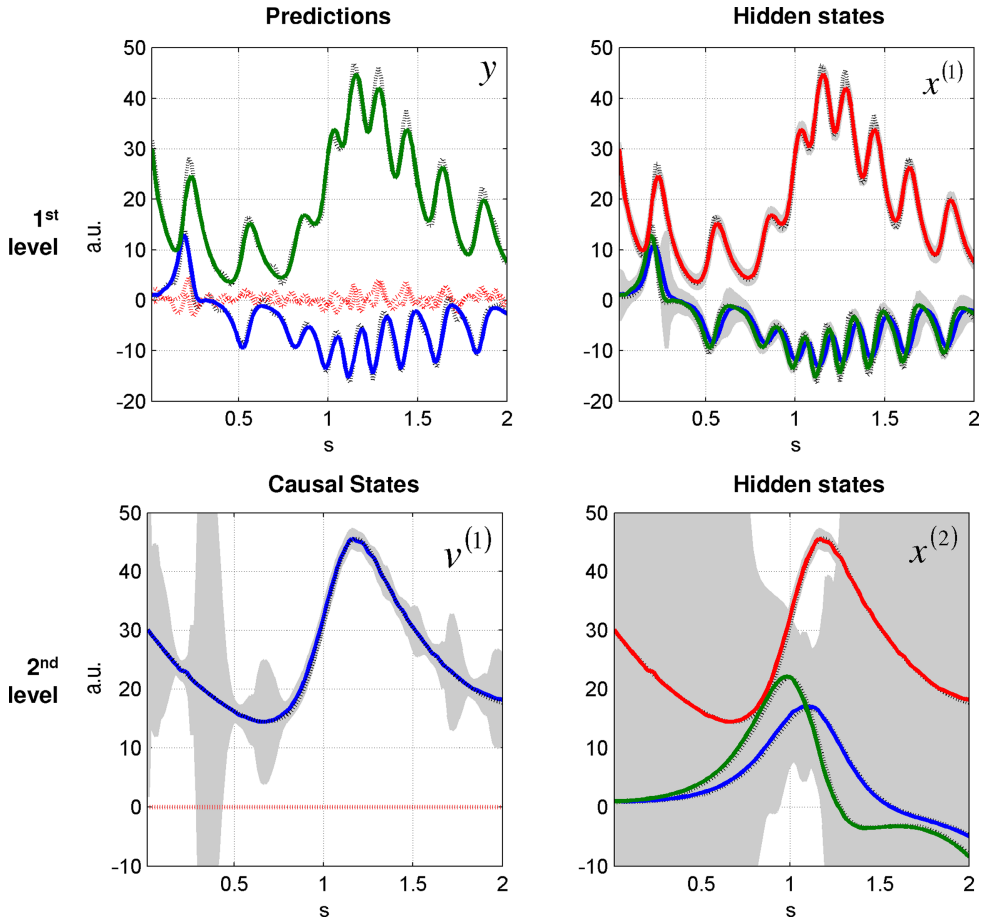


FIGURE 3 | Dynamic online inversion of the data presented in Figure 2. Observed data (see **Figure 2**) are now shown as black, dotted lines, and the model predictions as solid, coloured lines. The 90% confidence interval around the conditional means is shown in grey. The prediction error (i.e. difference between observation and model prediction) is indicated by red dotted lines.

Importantly, VT dynamics tend to be slower than the changes in acoustics they cause and the function which maps VT to acoustic dynamics can be highly nonlinear. Naturally, development of these generative models is slow because of their complexity and the ongoing development of novel schemes for inverting dynamic nonlinear hierarchical models. It may be that recent developments (Friston et al., 2008) in the inversion of these models, particularly in a neurobiological setting (Friston, 2008a), may play a useful role in the recognition of generative speech models used in ASR.

DISCUSSION

We have suggested that a simple model of birdsong perception, motivated by computational neuroscience and ongoing developments in artificial speech recognition share a critical feature: Generative models for human and avian communication seem to be based on a hierarchy of dynamical systems, where high levels display slow variations and provide contextual guidance for lower faster levels. The principle of hierarchical inference, using appropriate inversion schemes, with separation of time-scales, could be an inherent part of the computations that underlie successful artificial recognition of human action and behavior.

A hierarchical inference has several implications for cortical structure as well as for artificial and human perception. For cortical structure, these are:

- Cortical areas are organized hierarchically (Felleman and Van Essen, 1991; Fuster, 2004).
- Macroscopic neuroanatomy recapitulates hierarchical separation of time-scales; see Kiebel et al. (2008) for a discussion of the evidence that the cortex is organized as an anatomic-temporal hierarchy.
- Extrinsic forward connections convey prediction error (from superficial pyramidal cells) and backward connections mediate predictions, based on hidden and causal states (from deep pyramidal cells) (Friston, 2005; Mumford, 1992; Sherman and Guillery, 1998).

In the following we discuss the implications for artificial and human perception.

A COMPUTATIONAL PRINCIPLE FOR PERCEPTION

The conjecture that the brain inverts hierarchical generative models may lead to a deeper understanding of the computational principles

behind perception. As described above, a hierarchical approach has also been adopted in the engineering and artificial perception literature (Deng et al., 2006; Kim et al., 2008; Moeslund et al., 2006; Yam et al., 2004). It is worth noting that these developments seem to have made minimal reference to neuroscience but were driven by the insight that conventional non-hierarchical models do not capture the deep hierarchical structure of sensory data (Bilmes, 2006; Deng et al., 2006; Oliver et al., 2004).

What are the advantages and disadvantages of using hierarchical models as the basis of artificial perception? A clear disadvantage is that, for real-world applications like speech recognition, the dynamics of movements may take complicated forms, at various time scales. It is therefore not surprising that the best working solutions for artificial speech recognition rather rely on learning large numbers of free parameters in less constrained models (McDermott and Nakamura, 2006). In addition, the inversion of nonlinear stochastic hierarchical dynamic models is a non-trivial challenge (Budhiraja et al., 2007; Friston et al., 2008; Judd and Smith, 2004). However, in principle, hierarchical dynamics can be parameterized by rather low-dimensional systems, in comparison to the high-dimensional sensory stream. This means that relatively few parameters are required to track acoustic trajectories. This might make dynamic speech identifiable, leading to robust perception schemes. Interestingly, for speech, prior research has already investigated the dynamics of articulation but is embraced with reluctance by the artificial speech recognition field (McDermott and Nakamura, 2006).

A hierarchical model may also be useful in robust perception of motor behavior, because human movements seem to be more invariant than the sensory features which they cause (Todorov and Jordan, 2002). This means that movements, which are on a comparatively slower time-scale than their sensory expressions, may be expressed naturally at a higher level in hierarchical models. This is consistent with neuroscience findings that higher cortical levels show invariance over greater time scales than lower levels (Giese and Poggio, 2003; Hasson et al., 2008; Koechlin and Jubault, 2006). Furthermore, the relative slowness of human movements, in comparison to consequent variations in the sensory stream, may also enable the prediction of fast sensory features, increasing the robustness of perception (King et al., 2007; Yam et al., 2004). We have demonstrated this by showing that a hierarchical scheme can out-perform a non-hierarchical scheme, see Figure 5 in Kiebel et al. (2008).

In addition, speech trajectories could be modelled at time-scales beyond single speech-sounds and syllables, e.g. covering words and sentences. At this level, long-range hierarchical and cross-temporal dependencies are subject of active research in computational linguistics and natural language (Bengio et al., 2003; Huyck, 2009; Smits, 2001). The inversion of models with temporal hierarchies may provide a framework for computational models of language processing. For example, they are in a position to explain how uncertainty about the meaning of the early part of a sentence is resolved on hearing the end: i.e., increases in conditional certainty about hidden states, based on current sensory input confirms their predictions. In other words, the long-range or deep temporal dependencies in speech might lend themselves to hierarchical temporal modelling. The resulting inference, using serial speech input, may appear to be non-serial because decisive evidence for hidden states at different levels arrives at different times. To our knowledge, a fully dynamical hierarchical scheme that maps from sound waves to the semantics is still beyond the current abilities of artificial speech recognition (Deng et al., 2006).

SIMPLE NETWORK OPERATIONS

Although the variational inversion of hierarchical dynamic models might appear too unwieldy for a simple theory of perception, the actual operations needed to implement recognition dynamics are rather simple (Friston et al., 2008). By 'simple' we mean that all operations are instantaneous and just involve message-passing among neurons in a network and associative plasticity of their connections. This renders the approach neurobiologically plausible. The message-passing scheme is not the only possible implementation, there may be others, each with their own approximations and simplifications to compute the free energy. Irrespective of the optimization scheme used, the requisite update equations are determined by the generative model, which is specified by the likelihood and priors. This means that the identification of the brain's generative model of the environment is the key to understanding perception (Friston, 2008a; Rao and Ballard, 1999; Yuille and Kersten, 2006).

The variational inversion using generative models is just a recipe to construct a system of differential equations, which recognize sensory input, i.e., optimise a free-energy bound on the log evidence for some model. This means the scheme shares many formal similarities with dynamical systems used in computational neuroscience to describe neuronal systems (Rabinovich et al., 2006). As noted by one of our reviewers, it may be that such schemes have evolved to exploit natural or universal phenomena that appear when dynamical systems are coupled (Breakspear and Stam, 2005). Indeed, in an evolutionary setting, the emergence of efficient coupled dynamical systems that optimise free-energy may exploit these phenomena. For example, coupled nonlinear systems naturally evolve towards a synchronous state, even with relatively weak coupling. It would be very interesting if these synchronised states could be associated with optimised free-energy states that are mandated by perception in particular and the free-energy principle in general.

In short, the variational approach entails finding a dynamic system (the generative model), which describes the generation of sensory input. Variational learning principles are then applied to derive differential equations, which decode hidden states from sensory input. The use of generic inversion systems as proposed in Friston et al. (2008) enables one to focus on the first challenge, which may be informed by the study of coupled dynamical systems, in a more general setting.

COUPLING BETWEEN TIME-SCALES

The variational inversion of temporal hierarchies describes how fast sensory input can influence inferred states at slow time-scales. There are recent studies that suggest this sort of coupling may be a generic feature of coupled dynamical systems: Fujimoto and Kaneko describe how to exploit a bifurcation cascade to couple slow high-level states to fast low-level dynamics. Crucially, they find that coupling is seen only in a narrow regime of time-scale ratios, around two to three (Fujimoto and Kaneko, 2003a,b).

As shown in Kiebel et al. (2008), dynamical systems based on variational inversion schemes operate in a broader regime: one can construct systems where fast dynamics influence slow dynamics at much higher time-scale ratios. In the present work, we use a ratio of eight, which is beyond the limit identified by Fujimoto and Kaneko (2003b). However, dynamics based on variational inversion have a natural lower limit on the time-scale ratio: When the ratio approaches one, the changes in the manifold of the fast system, caused by the slow system, evolve nearly as fast as the states themselves. This means that the changes in the manifold cannot be separated from the dynamics of the states. This suggests that robust inversion of temporal hierarchies rests on a separation of temporal scales, which may impose a lower bound on the relative time-scales.

Although we have not emphasized it in this paper, the fact that one can formulate the inversion of dynamic models with deep or hierarchical temporal structure as a dynamical system rests on recent technical advances in Bayesian filtering (Friston, 2008b; Friston et al., 2008). In brief, these advances use generalised coordinates of motion to represent the trajectories of hidden states. Generalised coordinates cover position, velocity acceleration *etc.* Although this increases the number of implicit hidden states it greatly simplifies inversion, in comparison with conventional schemes like particle and extended Kalman filtering. This simplification reduces filtering (i.e., inversion) to a gradient descent, which can be implemented in a neurobiologically plausible fashion. The use of generalised coordinates is formally similar to temporal embedding in the characterisation of dynamical systems: Taken's theorem (Takens, 1981) states that it is possible to embed (i.e. geometrically represent) the structure of a vector-field in a higher dimensional space. This means that one can reconstruct the structure of the manifold, on which dynamics unfold, by using a Taylor expansion of the vector-field. This is very close to the idea of projecting the system into generalized coordinates. In essence, this projection allows the observer to encode the structure of the flow-field at each point in time.

A GENERAL MECHANISM FOR PERCEPTION AND ACTION IN THE BRAIN?

In a recent paper, we reviewed some compelling experimental evidence for temporal hierarchies in the brain. We argued that these hierarchies may reflect a general form of generative models that the brain uses to recognize causes beyond the temporal support of elemental percepts (e.g., formants in audition and biological motion in vision, Kiebel et al., 2008). We have shown previously that the inversion of these generative models lead to robust and accurate inferences about the causes of sensory input. Hierarchical models are approximations to the environmental processes that generate sensory data (Todorov et al., 2005); so one might ask why evolution selected temporal hierarchies? Intuitively, there is something fundamentally correct about generative models based on temporal hierarchies; in the sense that the content of our sensorium changes more quickly than its context. However, for communication and biological motion there may be additional reasons to suppose temporal hierarchies afford just the right model; this is because our brains may use the same architecture to generate and recognise movements (Kilner et al., 2007). This means that, during co-evolution with our conspecifics, temporal hierarchies may have been subject to selective pressure, precisely because they enable generation and recognition of communicative stimuli over multiple time-scales (i.e., with deep temporal structure) (Rauschecker and Scott, 2009; von Kriegstein et al., 2008).

PERCEPTION MIRRORS THE ENVIRONMENT

The role of non-autonomous recognition dynamics is to mirror or track autonomous dynamics in the environment. If this tracking is successful, the recognition system 'inherits' the dynamics of the environment and can predict its sensory products accurately. This inheritance is lost when the sensory input becomes surprising, i.e. is not predicted by perception. In this case, the recognition system uses prediction error to change the predictions and make sensory input unsurprising again. This heuristic explains how the agent's dynamics manage to switch rapidly between different attractor regimes. This switching, e.g. see **Figure 3** in Kiebel et al. (2008), is caused by the interplay between the system's attempt to minimize surprise (which is bounded by free-energy) and (surprising) sensory input.

IDENTIFICATION OF THE ENVIRONMENTAL MODEL

Explicit modeling of environmental dynamics and their inversion may be a useful approach to model perception for several reasons: most current research in computational neuroscience focuses on modeling a single neuronal system, which generates neuronal dynamics just as the brain does. This 'single system' approach, which does not model the environmental dynamics explicitly, is very useful for identifying neuronal mechanisms and relating them to applied sensory input and neuronal or behavioral observations (Rabinovich et al., 2006). However, this approach does not address how these neuronal mechanisms (and not others) come about in the first place.

An alternative approach may be to model neuronal dynamics 'from scratch': Such a full forward model would comprise three components: (i) A model of the environment with autonomous dynamics, which, using the free-energy principle, prescribes (ii) non-autonomous recognition dynamics, which are implemented by (iii) neuronal dynamics (**Figure 4**, left panel). In other words, appropriate models of the environment may be requisite to make strong predictions about observed neuronal dynamics. Given the complexity and detail of neuronal dynamics, one might argue that the identification of appropriate environmental models is a daunting task. However, the 'dual-system' approach of modeling both environment and the brain would essentially rephrase the question 'How does the brain work?' to 'What is a good model of the environment that discloses how the brain works?' (see, e.g., Chiel and Beer, 1997; Proekt et al., 2008). This approach has the advantage that environmental models, which cannot be inverted, disqualify themselves and are unlikely to be used as generative models by the brain. For example, in artificial speech recognition, the conventional hidden Markov model has been found difficult to invert for casual speech. Moreover, this model is also a poor generative model of speech, i.e. speech generated by this model yields barely intelligible speech (McDermott and Nakamura, 2006). Given that one can identify appropriate models of the environment; e.g., for audiovisual speech, the recognition performance can be directly compared to human performance. Furthermore, one could use established model selection schemes to evaluate

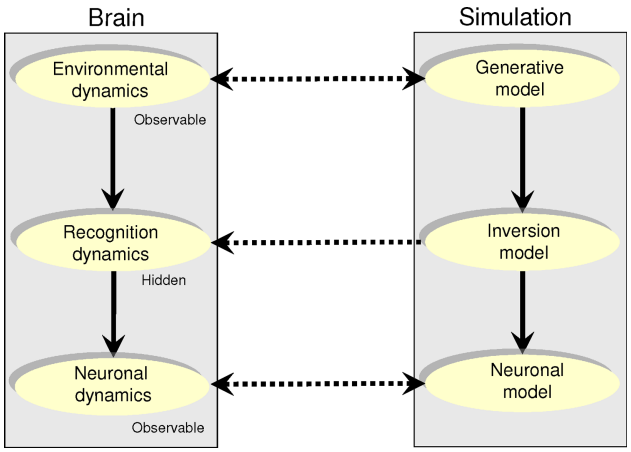


FIGURE 4 | Modeling neuronal dynamics caused by environmental dynamics. Brain system (left): In this dual-system model, neuronal dynamics (bottom) correspond to inversion or recognition dynamics (middle) induced by environmental dynamics (top). We assume that the environmental and neuronal dynamics can be partially observed, while the recognition dynamics are hidden. Simulated system (right): The full generative model of neuronal dynamics; starting with environmental dynamics, which specify recognition dynamics, which predict neuronal dynamics.

environmental models in the context of their neuronal inversion (Friston et al., 2008). This dual-system modeling approach may also allow one to ask whether simulated recognition produces the same kind of predictions and prediction errors as humans, e.g. when exposed to sensory input that induces the McGurk effect (Cosker et al., 2005). Such experiments would enable us to explain the McGurk effect and similar perception phenomena in a causal fashion, as the consequence of our brains' generative environmental model. In addition, one may be able to couple simulated recognition dynamics with models of neuronal dynamics and relate these to observed neuronal dynamics (**Figure 4**, right). This would enable us to make predictions about observed neuronal responses under specific assumptions about the generative model used by the brain, and how neuronal dynamics implement recognition.

The value of this dual-system approach is that neuroscience and artificial perception have a common interest in these models (Scharenborg, 2007). Not only would such an integrative approach provide a constructive account of brain function, at multiple levels of description, but also enable machines to do real-world tasks, (see, e.g., Rucci et al., 2007) for a spatial localization example at the interface between artificial perception, robotics and neuroscience.

CONCLUSIONS

We have demonstrated that the recognition of environmental causes from sensory input can be modeled as the inversion of dynamic, nonlinear, hierarchical, stochastic models. We have discussed relevant developments in artificial perception, which suggest that perception models the environment as a hierarchy of autonomous systems, evolving at various time-scales, to generate sensory input. In this view, the computational principles of perception may be accessed by considering variational inversion of these models.

ACKNOWLEDGMENTS

SJK is funded by the Max Planck Society. KJF and JD are funded by the Wellcome Trust. We thank Katharina von Kriegstein for valuable comments and discussions. We also thank the two reviewers for their helpful and constructive comments.

SUPPLEMENTARY MATERIAL

All procedures described in this note have been implemented as Matlab (MathWorks) code. The source code is freely available in the Dynamic Expectation Maximization (DEM) toolbox of the Statistical Parametric Mapping package (SPM8) at http://www.fil.ion.ucl.ac.uk/spm/.

REFERENCES

Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.* 3, 1137–1155.

Bilmes, J. A. (2006). What HMMs can do. *IEICE Trans. Inf. Syst.* E89d, 869–891.

Breakspear, M., and Stam, C. J. (2005). Dynamics of a neural system with a multiscale architecture. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 360, 1051–1074.

Browman, C. P., and Goldstein, L. (1992). Articulatory phonology – an overview. *Phonetica* 49, 155–180.

Browman, C. P., and Goldstein, L. (1997). The gestural phonology model. *Speech Prod. Motor Control Brain Res. Fluency Disord.* 1146, 57–71.

Budhiraja, A., Chen, L. J., and Lee, C. (2007). A survey of numerical methods for nonlinear filtering problems. *Physica D* 230, 27–36.

Chater, N., and Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends Cogn. Sci.* 10, 335–344.

Chiel, H. J., and Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci.* 20, 553–557.

Cosker, D., Paddock, S., Marshall, D., Rosin, P. L., and Rushton, S. (2005). Towards perceptually realistic talking heads: models, metrics, and McGurk. *ACM Trans. Appl. Percept.* 2, 270–285.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The Helmholtz machine. *Neural Comput.* 7, 889–904.

Deng, L., Yu, D., and Acero, A. (2006). Structured speech modeling. *IEEE Trans. Audio Speech Lang. Processing* 14, 1492–1504.

Deng, L., Lee, L. J., Attias, H., and Acero, A. (2007). Adaptive Kalman filtering and smoothing for tracking vocal tract resonances using a continuous-valued hidden dynamic model. *IEEE Trans. Audio Speech Lang. Processing* 15, 13–23.

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*360, 815–836.

Friston, K. (2008a). Hierarchical models in the brain. *PLoS Comput. Biol.* 4, e1000211.

Friston, K. J. (2008b). Variational filtering. *Neuroimage* 41, 747–766.

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87.

Friston, K. J., and Kiebel, S. J. (2009). Attractors in song. *New Math. Nat. Comput.* 5, 83–114.

Friston, K. J., Trujillo-Barreto, N., and Daunizeau, J. (2008). DEM: a variational treatment of dynamic systems. *Neuroimage* 41, 849–885.

Fujimoto, K., and Kaneko, K. (2003a). How fast elements can affect slow dynamics. *Physica D* 180, 1–16.

Fujimoto, K., and Kaneko, K. (2003b). Bifurcation cascade as chaotic itinerancy with multiple time scales. *Chaos* 13, 1041–1056.

Fuster, J. M. (2004). Upper processing stages of the perception-action cycle. *Trends Cogn. Sci.* 8, 143–145.

Giese, M. A., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4, 179–192.

Glaze, C. M., and Troyer, T. W. (2006). Temporal structure in zebra finch song: implications for motor coding. *J. Neurosci.* 26, 991–1005.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550.

Hofe, R., and Moore, R. (2008). Towards an investigation of speech energetics

using 'AnTon': an animatronic model of a human tongue and vocal tract. *Connect. Sci.* 20, 319–336.

Huyck, C. R. (2009). A psycholinguistic model of natural language parsing implemented in simulated neurons. *Cogn. Neurodyn.* (in press).

Judd, K., and Smith, L. A. (2004). Indistinguishable states II – the imperfect model scenario. *Physica D* 196, 224–242.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4, e1000209.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). The mirror-neuron system: a Baynesian perspective. *Neuroreport* 18, 619–623.

Kim, M., Kumar, S., Pavlovic, V., and Rowley, H. (2008). Face tracking and recognition with visual constraints in real-world videos. *2008 Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1–12, 1787–1794.

King, S., Frankel, J., Livescu, K., McDermott, E., Richmond, K., and Wester, M. (2007). Speech production knowledge in automatic speech recognition. *J. Acoust. Soc. Am.* 121, 723–742.

Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.

Kruger, V., Kragic, D., Ude, A., and Geib, C. (2007). The meaning of action: a review on action recognition and mapping. *Adv. Robot.* 21, 1473–1501.

Lee, T. S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448.

Liberman, A. M., and Whalen, D. H. (2000). On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196.

Livescu, K., Cetin, O., Hasegawa-Johnson, M., King, S., Bartels, C., Borges, N., Kantor, A., Lal, P., Yung, L., Bezman, A., Dawson-Haggerty, S., Woods, B., Frankel, J., Magimai-Doss, M., and Saenko, K. (2007). Articulatory feature-based methods for acoustic and audio-visual speech recognition: Summary from the 2006 JHU summer workshop. *2007 Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IV(Pts 1–3), 621–624.

McDermott, E., and Nakamura, A. (2006). Production-oriented models for speech recognition. *IEICE Trans. Inf. Syst.* E89d, 1006–1014.

Moeslund, T. B., Hilton, A., and Kruger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* 104, 90–126.

Mumford, D. (1992). On the Computational Architecture of the Neocortex.2. The Role of Corticocortical Loops. *Biol. Cybern.* 66, 241–251.

O'Shaughnessy, D. (2008). Invited paper: automatic speech recognition: history, methods and challenges. *Pattern Recognit.* 41, 2965–2979.

Oliver, N., Garg, A., and Horvitz, E. (2004). Layered representations for learning and inferring office activity from multiple sensory channels. *Comput. Vis. Image Underst.* 96, 163–180.

Ostendorf, M. (1999). Moving beyond the 'beads-on-a-string' model of speech. *Proc. IEEE Automat. Speech Recognit. Underst. Workshop* 1, 5.

Poeppel, D., Idsardi, W. J., and van, W. V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 363, 1071–1086.

Proekt, A., Wong, J., Zhurov, Y., Kozlova, N., Weiss, K. R., and Brezina, V. (2008). Predicting adaptive behavior in the environment from central nervous system dynamics. *PLoS ONE* 3, e3678.

Rabinovich, M. I., Varona, P., Selverston, A. I., and Abarbanel, H. D. I. (2006). Dynamical principles in neuroscience. *Rev. Mod. Phys.* 78, 1213–1265.

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.

Robertson, N., and Reid, I. (2006). A general method for human activity recognition in video. *Comput. Vis. Image Underst.* 104, 232–248.

Rose, R. C., Schroeter, J., and Sondhi, M. M. (1996). The potential role of speech production models in automatic speech recognition. *J. Acoust. Soc. Am.* 99, 1699–1709.

Rucci, M., Bullock, D., and Santini, F. (2007). Integrating robotics and neuroscience: brains for robots, bodies for brains. *Adv. Robot.* 21, 1115–1129.

Saenko, K., Livescu, K., Glass, J., and Darrell, T. (2005). Production domain modeling of pronunciation for visual speech recognition. *2005 Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* 1–5, V473–V476.

Scharenborg, O. (2007). Reaching over the gap: a review of efforts to link human and automatic speech recognition research. *Speech Commun.* 49, 336–347.

Sen, K., Theunissen, F. E., and Doupe, A. J. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* 86, 1445–1458.

Sherman, S. M., and Guillery, R. W. (1998). On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators". *Proc. Natl. Acad. Sci. U.S.A.* 95, 7121–7126.

Smits, R. (2001). Hierarchical categorization of coarticulated phonemes: a theoretical analysis. *Percept. Psychophys.* 63, 1109–1139.

Takens, F. (ed.) (1981). Detecting Strange Attractors in Turbulence. Berlin/Heidelberg, Springer.

Tan, X. Y., Chen, S. C., Zhou, Z. H., and Zhang, F. Y. (2006). Face recognition from a single image per person: a survey. *Pattern Recognit.* 39, 1725–1745.

Todorov, E., and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5, 1226–1235.

Todorov, E., Li, W., and Pan, X. (2005). From task parameters to motor synergies: a hierarchical framework for approximately-optimal control of redundant manipulators. *J. Robot. Syst.* 22, 691–710.

von Kriegstein, K., Patterson, R. D., and Griffiths, T. D. (2008). Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. *Curr. Biol.* 18, 1855–1859.

Vu, E. T., Mazurek, M. E., and Kuo, Y. C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* 14, 6924–6934.

Yam, C. Y., Nixon, M. S., and Carter, J. N. (2004). Automated person recognition by walking and running via model-based approaches. *Pattern Recognit.* 37, 1057–1072.

Yu, A. C., and Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science* 273, 1871–1875.

Yuille, A., and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.* 10, 301–308.

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: a literature survey. *ACM Comput. Surv.* 35, 399–459.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 March 2009; paper pending published: 20 May 2009; accepted: 23 July 2009; published online: 20 July 2009.
*Citation: Kiebel SJ, Daunizeau J and Friston KJ (2009) Perception and hierarchical dynamics. Front. Neuroinform. (2009) **3**:20. doi: 10.3389/neuro.11.020.2009*