

# Reducing module size bias of participation coefficient

Mangor Pedersen<sup>1\*</sup>, Amir Omidvarnia<sup>1</sup>, James M. Shine<sup>2</sup>,  
Graeme D. Jackson<sup>1,3</sup>, and Andrew Zalesky<sup>4</sup>

<sup>1</sup>The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, VIC, Australia

<sup>2</sup>Brain and Mind Center, The University of Sydney, Sydney, NSW, Australia

<sup>3</sup>Department of Neurology, Austin Health, Melbourne, VIC, Australia

<sup>4</sup>School of Engineering, The University of Melbourne, VIC, Australia

\* Corresponding author: Mangor Pedersen, The Florey Institute of Neuroscience and Mental Health, University of Melbourne, 3084, VIC, Australia. E-mail [mangor.pedersen@florey.edu.au](mailto:mangor.pedersen@florey.edu.au)

**Key words:** complex networks, graph theory, participation coefficient, fMRI, airports, *C.elegans*

## Abstract

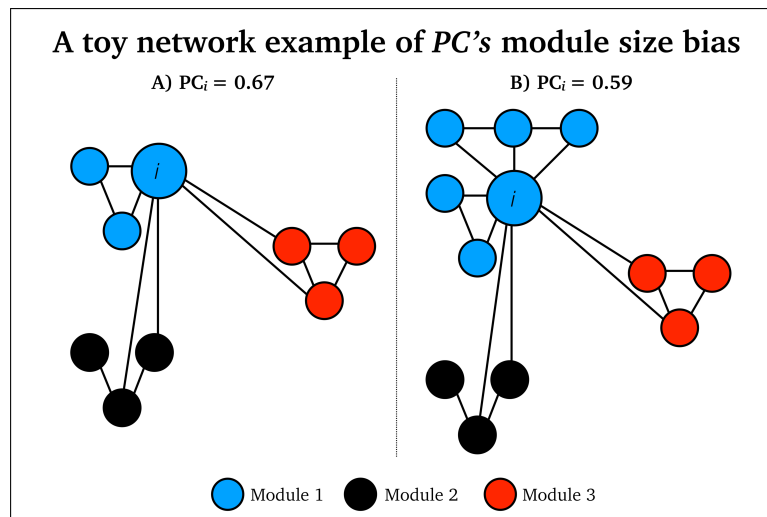
Both natural and engineered networks are often modular. Whether a network node interacts with only nodes from its own module or nodes from multiple modules provides insight into its functional role. The participation coefficient ( $PC$ ) is typically used to measure this attribute although its value also depends on the size of the module it belongs to, often leading to non-intuitive identification of highly connected nodes. Here, we develop a normalized  $PC$  that overcomes the module size bias associated with the conventional  $PC$ . Using brain, *C.elegans*, airport and simulated networks, we show that our measure of participation alleviates the module size bias, while preserving conceptual and mathematical properties, of the classic formulation of  $PC$ . Unlike the conventional  $PC$ , we identify London and New York as high participators in the air traffic network and demonstrate stronger associations with working memory in human brain networks, yielding new insights into nodal participation across network modules.

## Introduction

Many natural and engineered networks are modular [1–3]. Networks that are highly modular can be partitioned into communities of nodes, or modules, such that the density of connections is greater between the nodes within modules, relative to the density between nodes in different modules. Some nodes have connections that are distributed across many modules, whereas others are only connected with other nodes in their own module. This distinction can provide important insight into a node’s functional role in a modular architecture.

A node’s inter-modular connectivity is typically quantified with the participation coefficient ( $PC$ ) [4].  $PC$  provides insights into how specific nodes communicate between modules in a range of real-world networks, including air traffic and brain networks [5–13]. To compute a node’s  $PC$ , the proportion of a node’s connections to each module is first determined, yielding a proportion for each module. These proportions are then squared, summed across all modules and the resulting summand is subtracted from one to yield the node’s  $PC$ .  $PC$  of zero indicates a node that only connects with other nodes in its own module, whereas nodes with connections that are uniformly distributed across all modules have  $PC$  that approaches one.

$PC$  tacitly assumes that all modules in a network are equally sized. This is, however, rarely the case in real-world networks.  $PC$  is consequently susceptible to a bias wherein nodes in small modules often have high  $PC$  and nodes in large modules often have low  $PC$  [14]. This bias is exemplified in the two networks shown in Fig. 1, where node  $i$  in network B, by the virtue of belonging to a larger



**Figure 1: Toy network examples of  $PC$ 's module size bias:** Node  $i$  has identical intermodule connectivity in network A and B, but it has lower  $PC_i$  in network B as it belongs to a larger blue module than network A.

module, has a lower  $PC$  than node  $i$  in network A, even though node  $i$  has the same inter-modular connectivity in both networks. We refer to this as the module size bias of  $PC$ , which we show can yield biased inference about a node's participation.

The module size bias of  $PC$  was first highlighted by Klimm and others [14]. These authors proposed a dispersion index as an alternative to  $PC$ . In this paper, we propose a modification of the original  $PC$  measure rather than constructing a new measure of network integration. Our aim is to reduce the module size bias of  $PC$  while retaining the underlying mathematical assumptions, and numerical range, of the original  $PC$  measure. To achieve this, we developed a normalized  $PC$  in which a node's participation is benchmarked to an ensemble of random networks matched in node degree and connection density [15]. Our normalized participation coefficient ( $PC_{norm}$ ) accounts for the (intra-modular) connectivity expected by chance, given the size of network modules. To validate  $PC_{norm}$ , we used three real-world networks: i) undirected functional MRI (fMRI) brain network data with 100 brain regions from 1003 healthy adults (ages 22-35) participating in the Human Connectome Project [16]; ii) a single directed neural network from the *C.elegans* nematode containing a total of 277 neurons [17, 18]; and iii) a single directed flight network with 500 airports [19], as well as simulated networks [4].

Firstly, simulations were undertaken to test the hypothesis that  $PC_{norm}$  preserves the network features and numerical range of  $PC$ , in networks where all modules were equally sized. We then hypothesized that  $PC_{norm}$  would reduce the module size bias inherent to  $PC$  in real-world networks across a range of spatial resolution parameters and network density thresholds. In addition to our main hypotheses, we assessed nodal traits of  $PC_{norm}$ , to evaluate their relevance in neurobiological and air-traffic systems. We also investigated whether  $PC_{norm}$  measured in fMRI networks associated with behavior more strongly than  $PC$ . Together, our results demonstrate a simple refinement of  $PC$  that retains the interpretation of  $PC$  in real-world networks but alleviates the module size bias. Unlike the conventional  $PC$ , we identify London (LHR) and New York (JFK), among other airports, as high participators in the air traffic network, and demonstrate stronger associations with working memory in human brain networks. We hope  $PC_{norm}$  will lead to a more reliable estimation of nodal integration in complex networks.

## Results

### Normalized participation coefficient ( $PC_{norm}$ )

The participation coefficient ( $PC$ ) measures whether a node interacts with only nodes from its own module or nodes from multiple modules [4]. Formally,  $PC$  of node  $i$  is given by,

$$PC_i = 1 - \sum_{m \in M} \left( \frac{k_i(m)}{k_i} \right)^2, \quad (1)$$

where  $M$  is the set of network modules,  $k_i(m)$  is the degree between node  $i$  and all nodes in module  $m$  and  $k_i$  is the degree between node  $i$  and all other nodes in the entire network. A node with  $PC$  of zero only interacts with nodes comprising its own module, while nodes with connections uniformly distributed across all modules have a  $PC$  that approaches one. As demonstrated below and discussed elsewhere [14],  $PC$  is limited by an inherent module size bias, which may lead to inaccurate inference in networks with modules that vary in size.

To alleviate the module size bias, we propose the normalized participation coefficient ( $PC_{norm}$ ),

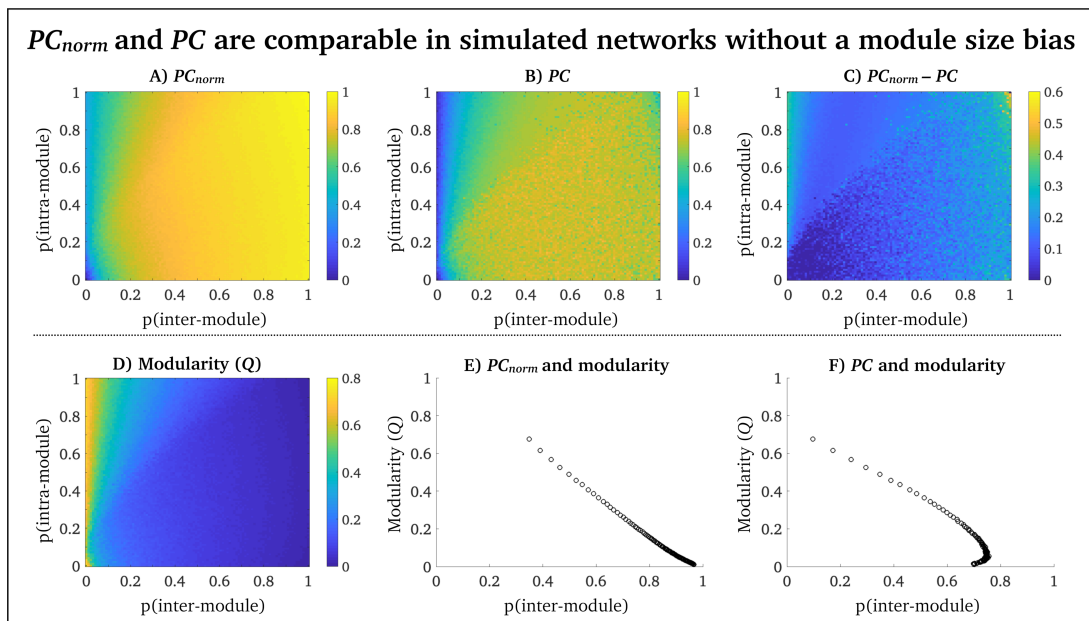
$$PC_{norm_i} = 1 - \sqrt{B_0 \sum_{m \in M} \left( \frac{k_i(m) - k_i(m)_{rand}}{k_i} \right)^2}. \quad (2)$$

The key difference between  $PC$  and  $PC_{norm}$  of node  $i$  is subtraction of the normalization factor,  $k_i(m)_{rand}$ , from  $k_i(m)$ . This normalization factor denotes the median intra-modular degree for node  $i$  across randomized networks generated with an established network rewiring algorithm that preserves connection density and node degree (i.e.,  $k_i$ ). This also means that  $k_i$  is the same for original and randomized networks [15]. We found that 1000 network randomizations were adequate to return a stable estimate of  $PC_{norm}$  (see Supplementary Fig 1). In each randomized network, all edges were rewired five times. We used the same underlying modular network structure ( $M$ ) for original  $k_i(m)$  and random  $k_i(m)_{rand}$ . To constrain the range of  $PC_{norm}$  between 0 (low network integration) and 1 (high network integration), we add the multiplicative term  $B_0 = 0.5$ , and we also calculate the square root of the difference of participation between original and randomized networks.

### $PC$ and $PC_{norm}$ are comparable in simulated networks without a module size bias

Firstly, we aimed to establish that our normalization process does not alter the conceptual basis of the  $PC$ , or its mathematical properties and relationships with other topological attributes. To discount this possibility, we implemented the network simulation described by Guimerà and Amaral [4] and verified that  $PC$  and  $PC_{norm}$  yielded comparable values in a network comprising equally-sized modules. Specifically, we simulated binary networks with 100 nodes, and four equal-sized modules (each containing 25 nodes). Given that the modules were equally sized, module size bias was absent by design, and thus  $PC$  and  $PC_{norm}$  should not markedly differ. We added connections either within or between the different modules through a range of probabilities (0 to 1, in increments of 0.01), and calculated  $PC_{norm}$  (Fig. 2A),  $PC$  (Fig. 2B),  $PC_{norm}$  minus  $PC$  (Fig. 2C) and modularity Q-score (Fig. 2D), for all possible connectivity probability values.

Figure 2A and B suggest that  $PC_{norm}$  and  $PC$  yield highly comparable estimates across a range of simulated networks with equally sized modules. More specifically, the relationship between  $PC_{norm}$  and  $PC$  was strong across all simulated networks (Pearson's  $r = 0.91$ ). This suggests that  $PC_{norm}$  remains consistent with  $PC$  in networks with no module size bias, despite our algorithmic changes to node participation. Our simulation additionally shows that  $PC_{norm}$  values reside within the same range as  $PC$  (i.e., [0,1]), although  $PC_{norm}$  tends to have higher values than  $PC$  (see Fig 2, 3 and 8).  $PC_{norm}$  is also shows a more linearly inverse relationship between modularity and inter-module connectivity than  $PC$  (Fig. 2E-F). This phenomena was also outlined by Guimerà and Amaral in their seminal  $PC$  paper [4], where the original  $PC$  algorithm may be unstable if a node has more than 50% inter-modular connections. This suggests that  $PC_{norm}$  may be an appropriate metric in networks with numerous inter-modular connections.



**Figure 2: Simulations in networks without module size bias:** Simulation results for  $PC_{norm}$  and  $PC$  in networks with equally sized modules, where the heat maps show that average  $PC$  across all nodes. (A)  $PC_{norm}$  across probabilities of intra- ( $y$ -axis) and inter-module ( $x$ -axis) connectivity. (B)  $PC$  across probabilities of intra- ( $y$ -axis) and inter-module ( $x$ -axis) connectivity. (C)  $PC_{norm}$  minus  $PC$  across probabilities of intra- ( $y$ -axis) and inter-module ( $x$ -axis) connectivity. (D) modularity ( $Q$ -score) across probabilities of intra- ( $y$ -axis) and inter-module ( $x$ -axis) connectivity. (E) average modularity for  $PC_{norm}$  across inter-module ( $x$ -axis) connectivity. (F) average modularity for  $PC$  across inter-module ( $x$ -axis) connectivity. This simulation suggests that  $PC_{norm}$  preserves the mathematical, and conceptual, features of the original  $PC$  measure.

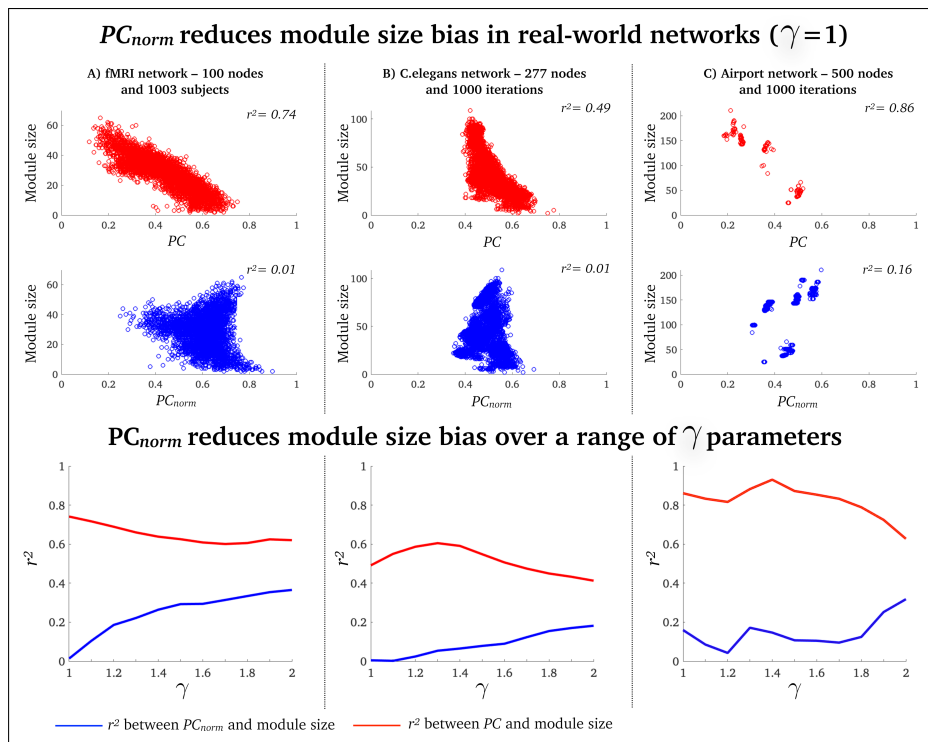
### $PC_{norm}$ reduces $PC$ 's module size bias in real-world modular networks

Next, we aimed to test whether  $PC_{norm}$  alleviates the module size bias of  $PC$  and thus yields more intuitive conclusions about the importance of single nodes in three real-world networks. Using a Louvain modularity decomposition method with a spatial resolutions of  $\gamma = 1$  [21] ( $\gamma = 1$  yields few and large modules), we found that fMRI human brain networks had an average modularity  $Q$ -score of  $0.35 \pm 0.03$  [SD]<sup>1</sup>, and an average of  $3.52 \pm 0.63$  [SD] modules, across 1003 healthy adults; the *C.elegans* network had an average modularity  $Q$ -score of  $0.42 \pm 0.01$  [SD], and an average of  $6.02 \pm 0.55$  [SD] modules, across 1000 Louvain modularity iterations; the airport network had an average modularity  $Q$ -score of  $0.50 \pm 0.01$  [SD] and an average of  $4.01 \pm 0.05$  [SD] modules, across 1000 Louvain modularity iterations (see Supplementary Fig. 3, for results across a range of  $\gamma$  parameters). This resulted in a total of 3003 networks for analysis, where we for each network calculated the average  $PC_{norm}$  and  $PC$  across all (non-zero) nodes in each module. In our main analysis we included 11 values for the  $\gamma$ -parameters (1 to 2, in increments of 0.1). This was done to verify that  $PC_{norm}$  reduces  $PC$ 's module size bias across several resolution parameters and module sizes.

To enable statistical inference between the two squared correlation coefficients of interest (correlation value #1 =  $r^2$  between  $PC$  and module size; correlation value #2 =  $r^2$  between  $PC_{norm}$  and module size), we employed a bootstrapping approach using 10000 bootstrap samples with replacement [22]. Here, we aimed to test the null hypothesis of equality in the two correlation coefficients. For each sample, we computed squared correlation coefficients based on the pooled bootstrap samples of participation ( $PC_{norm}$  and  $PC$ ) and module size, as well as its 95% confidence intervals (95% CI).

In line with our hypothesis, we found that the variance explained ( $r^2$ ) between the participation coefficient and module size was substantially lower for  $PC_{norm}$  relative to  $PC$ , where a module's size was determined by the number of nodes it comprised. This indicates that  $PC_{norm}$  alleviates the module size bias of  $PC$ . We display correlation patterns between  $PC_{norm}$ ,  $PC$  and module size, in Fig. 3, and we summarize the statistical differences between  $PC_{norm}$ ,  $PC$  and module size in Table 1.

<sup>1</sup> $Q$ -scores range between 0 and 1 where values proximate to 1 indicate that the network is highly modular, with a predominance of intra-modular connections.



**Figure 3:  $PC_{norm}$  reduces  $PC$ 's module size bias:** Relationship between  $PC$  and module size (top row),  $PC_{norm}$  and module size (middle row), for  $\gamma = 1$  (few and large module). Each data point is the average participation of all nodes within a module. Displayed are (A) 3357 modules (data points) across 1003 fMRI networks; (B) 6020 modules (data points) across 1000 *C.elegans* networks and (C); 3998 modules (data points) across 1000 airport networks. The scatter plots for the airport network appear sparser due to greater consistency across runs of the modular decomposition algorithm, relative to the other networks (see Supplementary Fig. 2). The bottom row shows the mean difference between  $PC$ ,  $PC_{norm}$  and module size for  $\gamma$  between 1 and 2 in 0.1 increments. It is worth noting that  $PC$ 's module size bias is reduced for high  $\gamma$  values ( $\gamma > 1.5$ ), where networks have numerous small modules with less variation in size, although  $\gamma > 1.5$  is rarely used in complex network analysis [20].

Given that connection density –i.e., the proportion of edges in a network– can markedly impact a network's modular architecture as well as other topological properties [23], we also confirmed that  $PC_{norm}$  reduces  $PC$ 's module size bias in fMRI networks, across multiple connection densities (Supplementary Fig. 4). From now on, we report results associated with  $\gamma = 1$ .

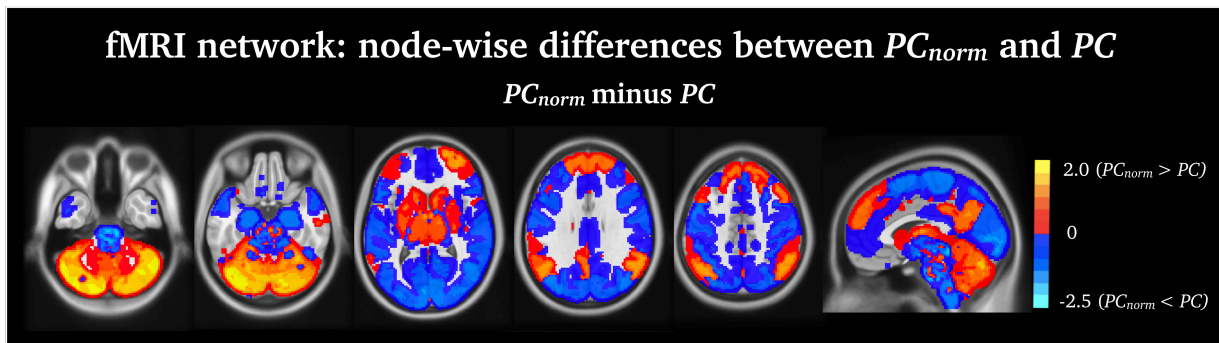
### Node-wise features of $PC_{norm}$ in real-world modular networks

Next, we investigated the distinct inferences and conclusions that can be reached about the role of specific nodes when alleviating the module size bias. For this purpose, we subtracted average  $PC$  from  $PC_{norm}$  across 1003 fMRI network subjects, and 1000 modularity runs for *C.elegans* and airport networks. In the next few paragraphs we have summarized node-wise results for fMRI, *C.elegans* and airport networks. We have also provided a full set of node-wise results for  $PC_{norm}$  and  $PC$  in Supplementary Table 1, 2 and 3, including  $z$ -score differences between these two measures. A  $z$ -score was used to calculate the relative differences between  $PC_{norm}$  and  $PC$  because  $PC_{norm}$  tends to yield higher values than  $PC$  (Fig. 2 and 3). In Supplementary Fig. 5, we display the average modularity structure for fMRI, *C.elegans* and airport networks.

In the brain, the largest difference between  $PC_{norm}$  and  $PC$  was observed in cerebellum subregions crus I ( $z = 1.76$ ), crus II ( $z$ -score = 1.56) and lobule VI ( $z$ -score = 1.42), but also in cortical areas such as inferior parietal cortex ( $z$ -score = 1.31) and superior frontal gyrus ( $z = 1.12$ ) (Fig. 4A), areas that are known to connect to a variety of sub-networks in the brain [24], and that subserve higher order cognitive function [11]. Brain regions with the highest  $PC_{norm}$  were cerebellum lobule VI ( $PC_{norm} = 0.73$ ), cerebellum crus I ( $PC_{norm} = 0.72$ ;  $PC = 0.49$ ), inferior parietal cortex ( $PC_{norm} = 0.71$ ;  $PC = 0.46$ ), middle frontal gyrus ( $PC_{norm} = 0.71$ ;  $PC = 0.50$ ) and precuneus ( $PC_{norm} = 0.70$ ;  $PC = 0.52$ ) (Supplementary Table 1).

**Table 1: Mean  $r^2$  bootstrap difference, and 95% CI, between PC and  $PC_{norm}$  (correlation with module size)**

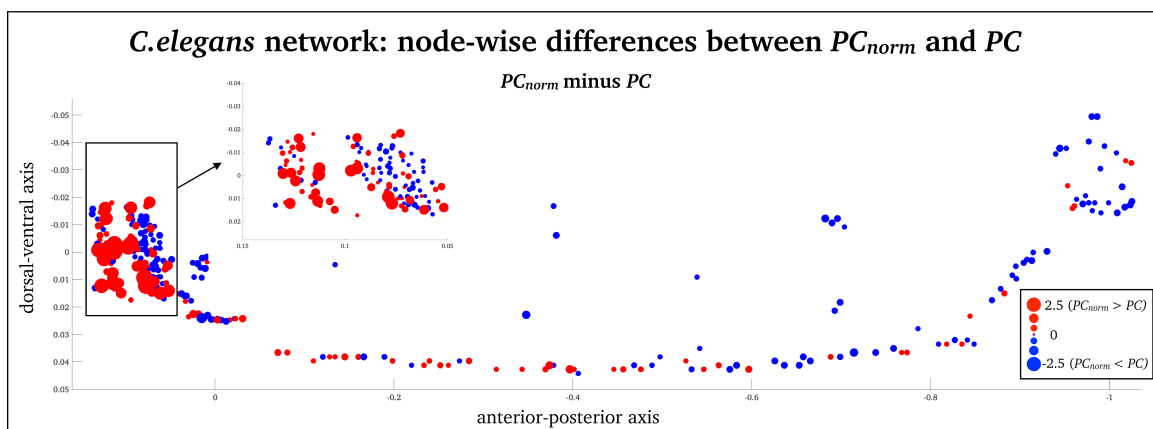
	$\gamma = 1$	$\gamma = 1.1$	$\gamma = 1.2$	$\gamma = 1.3$	$\gamma = 1.4$	$\gamma = 1.5$	$\gamma = 1.6$	$\gamma = 1.7$	$\gamma = 1.8$	$\gamma = 1.9$	$\gamma = 2$
<b>fMRI network</b>	0.73 [0.71-0.74]	0.60 [0.57-0.63]	0.49 [0.46-0.52]	0.42 [0.39-0.45]	0.36 [0.23-0.28]	0.32 [0.33-0.38]	0.30 [0.27-0.33]	0.28 [0.25-0.30]	0.27 [0.25-0.28]	0.26 [0.25-0.28]	0.25 [0.23-0.27]
<b>C.elegans network</b>	0.49 [0.46-0.50]	0.54 [0.53-0.56]	0.56 [0.55-0.57]	0.55 [0.54-0.56]	0.53 [0.51-0.54]	0.47 [0.46-0.48]	0.42 [0.40-0.43]	0.35 [0.34-0.36]	0.29 [0.28-0.30]	0.26 [0.25-0.27]	0.23 [0.22-0.24]
<b>Airport network</b>	0.70 [0.69-0.71]	0.74 [0.73-0.76]	0.77 [0.76-0.78]	0.71 [0.69-0.73]	0.78 [0.77-0.80]	0.76 [0.75-0.78]	0.74 [0.73-0.76]	0.73 [0.72-0.75]	0.66 [0.65-0.68]	0.47 [0.45-0.49]	0.31 [0.30-0.32]



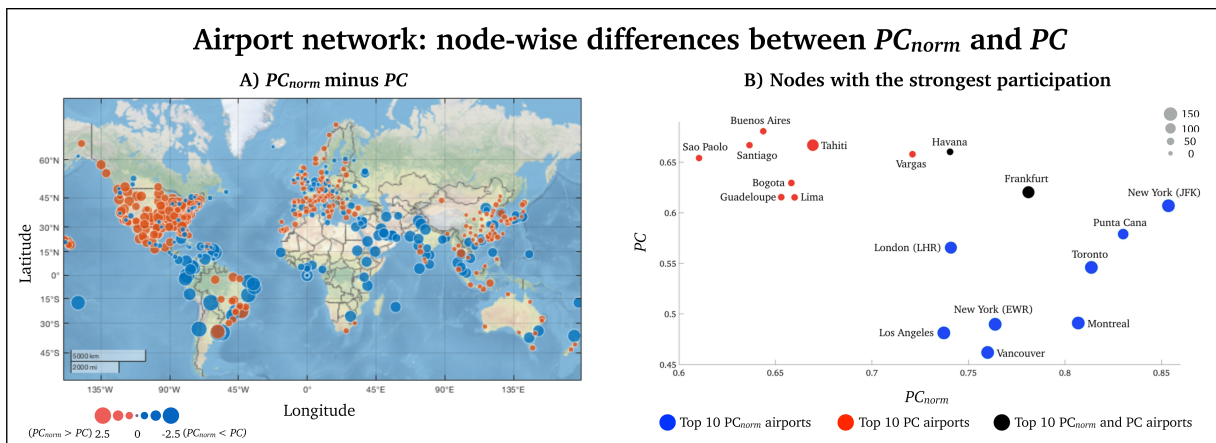
**Figure 4: Node-wise fMRI differences between  $PC_{norm}$  and  $PC$ :** Displayed are the subtracted group-average  $PC$  from  $PC_{norm}$  in fMRI networks (z-score). The largest difference between  $PC_{norm}$  and  $PC$  was observed in cerebellum subregions lobule VI, crus I and II, fronto-parietal and insular cortex.

For the *C.elegans* network, the greatest nodal differences between  $PC_{norm}$  and  $PC$  was observed in the largest module of the nematode encompassing inhibitory GABAergic neurons involved in head movement (RMDDL – z-score = 2.48; RICR – z-score = 2.35; RMDVL – z-score = 2.27; RMDDR – z-score = 2.26; and RMDVR z-score = 2.23 – Fig 5 and Supplementary Table 2). These were not the same subset of neurons with highest node participation. The strongest  $PC_{norm}$  was observed in several interneurons, also located in the head of the nematode including AVER ( $PC_{norm}$  = 0.87;  $PC$  = 0.75); AVAR ( $PC_{norm}$  = 0.85;  $PC$  = 0.75); AVAL ( $PC_{norm}$  = 0.84;  $PC$  = 0.75); AVBR ( $PC_{norm}$  = 0.83;  $PC$  = 0.74) and AVBL ( $PC_{norm}$  = 0.80;  $PC$  = 0.74), responsible for locomotor behavior, which is an important function for the *C.elegans* nematode. These locomotor interneurons have been found to be densely inter-connected across a range of sub-networks forming a selective ‘rich-club’ responsible for a bulk of neural signalling in the nervous system of the *C.elegans* nematode [13, 25–27]. See <https://www.wormatlas.org/neurons/Individual%20Neurons/Neuronframeset.html>, for full naming, definition and function of specific *C.elegans* neurons.

In the airport network, the greatest difference between  $PC_{norm}$  and  $PC$  was observed in South American airports (Montevideo – z-score = -2.26; Sao Paulo – z-score = -2.22; Lima – z-score = -2.20; Rio de Janeiro – z-score = -2.19; and Buenos Aires – z-score = -2.18 – see Fig. 6A and Supplementary Table 3), which was the smallest module in this network. Notably, 8 of the 10 airports with highest  $PC$  belonged to the module encompassing Central and South America (Fig 6B – red circles). This was in contrast to the nodes with strongest  $PC_{norm}$ , which included major airport hubs, including New York-JFK ( $PC_{norm}$  = 0.85;  $PC$  = 0.61), Punta Cana ( $PC_{norm}$  = 0.83;  $PC$  = 0.58), Toronto ( $PC_{norm}$  = 0.81;  $PC$  = 0.55), Montreal ( $PC_{norm}$  = 0.81;  $PC$  = 0.49) and Frankfurt ( $PC_{norm}$  = 0.78;  $PC$  = 0.62) (see blue circles in Fig. 6). These airports are known to have many intercontinental



**Figure 5: Node-wise *C.elegans* differences between  $PC_{norm}$  and  $PC$ :** Z-score difference between  $PC$  and  $PC_{norm}$ , in the *C.elegans* network (z-score). The largest difference between  $PC_{norm}$  and  $PC$  was observed in inhibitory GABAergic neurons that belong to the largest module of the nematode encompassing RMDDL, RICR, RMDVL, RMDDR and RMDVR neurons.



**Figure 6: Node-wise difference between  $PC_{norm}$  and  $PC$  in the network of airports:** (A) A world map displaying all airports included in the study. As highlighted in the manuscript, we observed the greatest difference between the two metrics in Central- and South American airports (red nodes). (B) A visual representation of the ten nodes with strongest  $PC_{norm}$  on the  $x$ -axis (blue nodes) and  $PC$  (red nodes) on the  $y$ -axis. The two black nodes were high for  $PC_{norm}$  and  $PC$ . Here, the size of each node is proportional to its module size (number of nodes in modules).

flights. This node-wise analysis suggests that  $PC$  overestimates the integrative nature of nodes within relatively small modules and that  $PC_{norm}$  can alleviate systematic module size bias, thereby enabling clearer and more intuitive conclusions to be drawn about the role of nodes.

### **$PC_{norm}$ is more strongly correlated with working-memory performance than $PC$**

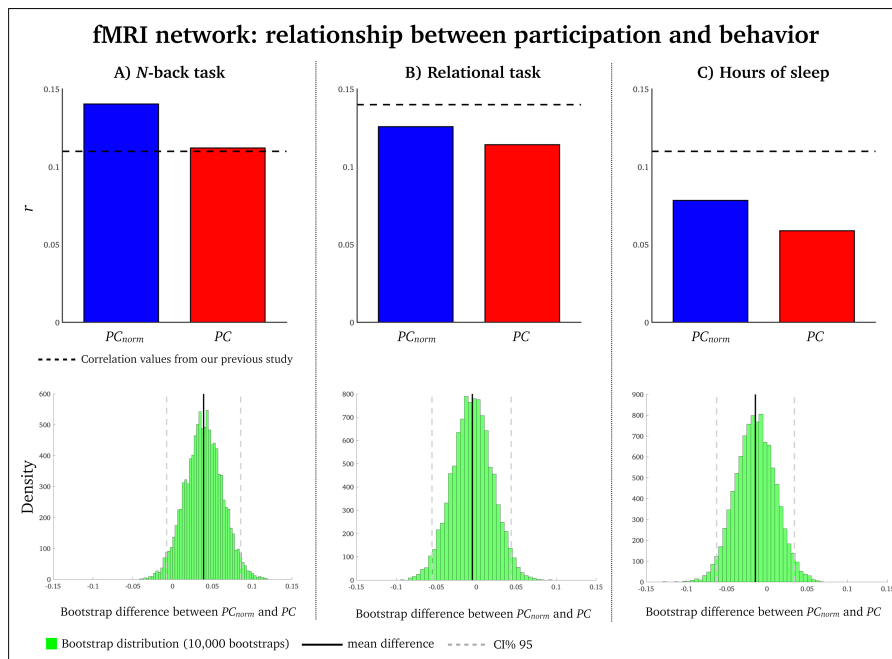
Having established that  $PC_{norm}$  alleviates the module size bias of  $PC$  and yields a more parsimonious interpretation of node roles, we next aimed to test whether inter-individual variation in  $PC_{norm}$  measured in brain networks would associate more strongly with behavioral measures than  $PC$  [28].

We conducted an analysis informed by our previous study where we showed that inter-modular network switching was related to the following behavioral domains [8]: i)  $N$ -back task, calculated as the average of 0- and 2-back task, important for working memory performance; ii) a relational task central for planning and reasoning; iii) a sleep index averaging hours of sleep the month before the fMRI scan. Each subject had a single behavioral score for each of the three behavioral domains. We therefore averaged  $PC_{norm}$  across all nodes in the network to yield a single  $PC_{norm}$  value, for each subject. We show that, using the same bootstrapping procedure outlined previously, mean  $PC_{norm}$  associated with functional brain networks was correlated more strongly with the  $N$ -back task compared to  $PC$ , but not the sleep index and relational task (Fig. 7). This suggests that working memory performance relates to the extent of inter-modular connectivity, rather than the size, of modules.

### **$PC_{norm}$ , $PC$ and within-module degree $z$ -score in real-world networks**

$PC$  is often interpreted alongside within-module degree  $z$ -score, a metric that quantifies the normalized degree of intra-modular nodes. First of all, we observed almost no correlation between within-module degree  $z$ -score and module size (Pearson's  $r^2 = 0.01$ ,  $\gamma = 1$ ). We next projected corresponding values between  $PC$  and within-module degree  $z$ -score in a joint histogram which categorizes the intra- and inter-modular statuses of nodes (see Fig 8). Similarly to previous work, most network nodes are peripheral nodes classified as nodes that mainly connect to a single module [6]. However, our joint histogram analysis showed that  $PC_{norm}$  detected more non-hub connector nodes than  $PC$  in fMRI networks (Fig. 8A). Non-hub connector nodes are classified as nodes that connect to a variety of modules. Airport networks also showed increased  $PC_{norm}$ , compared to  $PC$ , but remained as peripheral nodes (Fig. 8B). On the other hand,  $PC$  and  $PC_{norm}$  yielded comparable inference about the *C.elegans* network, which primarily comprised peripheral nodes (see Fig. 8C). This suggests that  $PC_{norm}$  may alter the topological structure, and interpretation, of networks where the module size bias is prominent.



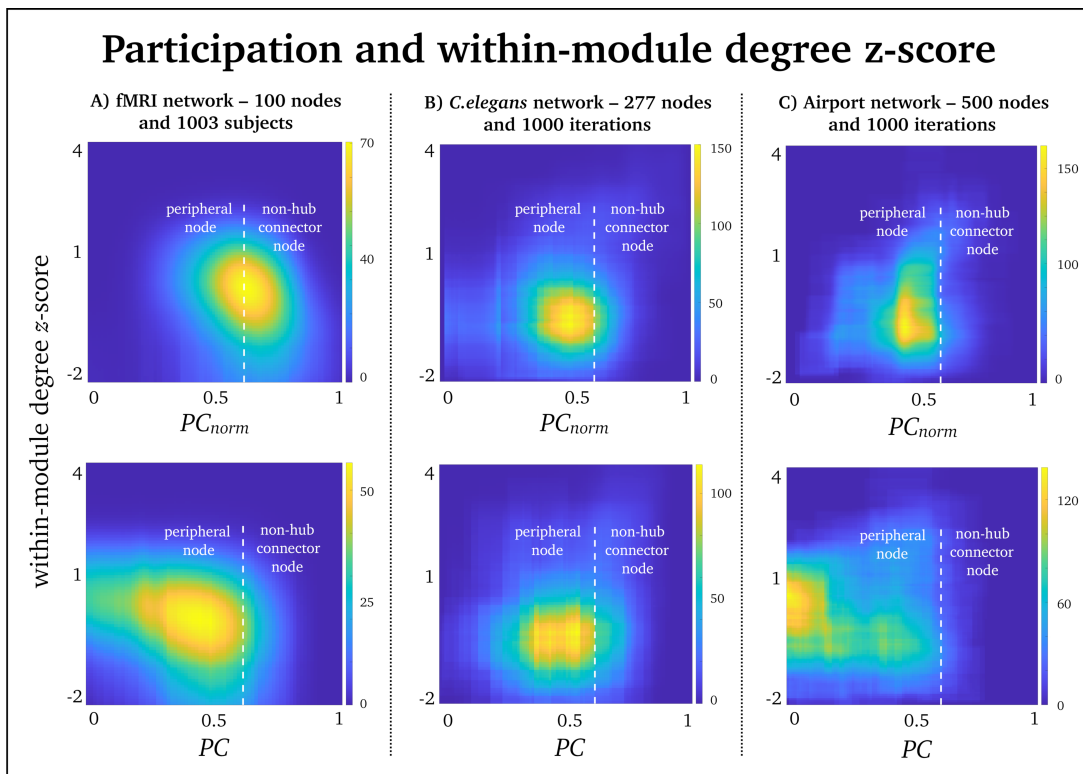


**Figure 7: Correlation between  $PC_{norm}$ ,  $PC$  and behavior:** Inter-individual variation in  $PC_{norm}$  (blue) and  $PC$  (red) was associated with three behavioral measures (A)  $N$ -back task, (B) sleep index and (C) relational task. In each panel, the dashed line represents the correlation between modular network switching and behavior obtained in our previous study [8]. Pearson’s  $r$  between  $N$ -back task and  $PC_{norm}$  was 0.14 ( $df = 1001$ ,  $p < 0.00001$ );  $r$  between  $N$ -back task and  $PC$  was 0.11 ( $df = 1001$ ,  $p < 0.001$ ). Bootstrap difference between  $PC_{norm}$  and  $PC$  for  $N$ -back task was 0.04 [95% CI = -0.01 - 0.08].  $r$  between sleep and  $PC_{norm}$  was 0.12 ( $df = 1001$ ,  $p < 0.0001$ );  $r$  between sleep and  $PC$  was 0.13 ( $df = 1001$ ,  $p < 0.0001$ ). Bootstrap difference between  $PC_{norm}$  and  $PC$  for sleep was 0.00 [95% CI = -0.05 - 0.05].  $r$  between relational task and  $PC_{norm}$  was 0.06 ( $df = 1001$ ,  $p = 0.07$ );  $r$  between relational task and  $PC$  was 0.07 ( $df = 1001$ ,  $p = 0.04$ ). Bootstrap difference between  $PC_{norm}$  and  $PC$  for relational task was 0.01 [95% CI = -0.04 - 0.05].

## Discussion

We proposed a normalized participation coefficient ( $PC_{norm}$ ) to alleviate the module size bias inherent to the established formulation of  $PC$  developed by Guimerà and Amaral [4]. Using brain, *C.elegans*, airport and simulated networks, we demonstrated that our proposed measure of participation: i) alleviates module size bias to a significant extent; ii) preserves the conceptual and mathematical properties of the classic formulation of  $PC$ ; and, iii) yields new insights into the role and function of nodes in the context of their participation across modules, thereby enabling clearer and more intuitive conclusions to be drawn about the role of nodes. Our work should be conceptualized as an extension to Guimerà and Amaral’s classic formulation of  $PC$  [4], rather than a new measure to include in the network scientist’s toolbox. We advocate use of  $PC_{norm}$  in favour of  $PC$ , particularly for networks with significant variation in module size. These are the networks for which the module size bias is most detrimental to inference performed based on the classic participation coefficient.

It is important to clarify and highlight the crux of  $PC$ ’s module size bias, namely network modularity. Modularity decomposition methods are designed to search for a collection of nodes that have stronger and more widespread inter-connectedness than we expect by chance [21]. Then, a sizeable collection of inter-connected nodes forming a large module will, more often than not, have low  $PC$  values without necessarily reflecting the node’s inter-connectedness with other modules. An example of this is the case of airport networks, which showed the strongest module size bias. We observed maximal  $PC$  in a subset of airports located in the smallest module comprising Central- and South America, that consequently led to high  $PC$  values (Fig. 6). After correcting this module size bias,  $PC_{norm}$  suggested that the airports with strongest inter-modular connectivity were located in metropolitan cities such as New York, London and Paris where over one third of flights were cross-continental. It is also important to note that  $PC_{norm}$ , like  $PC$ , appears independent from node degree. This is evident as the two airports with highest  $PC_{norm}$  had highly divergent node degree: i) New York had a total of 218 connections (127 intra-modular connections); and ii) Punta Cana in the Dominican Republic had



**Figure 8: Participation and within-module degree z-score:** Joint histograms containing the number of simultaneous occurrences between  $PC_{norm}$  (top row),  $PC$  (bottom row) and within-module degree z-score in (A) fMRI networks, (B) *C.elegans* networks and (C) airport networks. Within-module degree z-score is shown on the vertical axis, while  $PC$  and  $PC_{norm}$  is shown on the horizontal axis. In accordance with [6],  $PC$  of 0.62 was used to distinguish between peripheral nodes ( $PC < 0.62$ ) and non-hub connector nodes ( $PC > 0.62$ ).

only 52 connections (34 intra-modular connections). Although Punta Cana is not a densely connected airport, it belongs to the large module encompassing North America. This airport has a substantial number of cross-continental flights, mainly to South America, again demonstrating how  $PC$  may underestimate the integrative nature of nodes within large modules. A specific example where  $PC$  was less affected by module size was the RMGR neuron in the *C.elegans* nematode. The RMGR neuron of the *C.elegans* nematode, which showed similar values between  $PC$  and  $PC_{norm}$  (Supplementary Table 2), despite belonging to a spatially large module. This finding may be due to the high fidelity of the *C.elegans* connectome, which was mapped using electron microscopy, and also the inherent function of this neuron. RMGR is an interneuron located at the head of the *C.elegans* nematode, and it makes gap junctions to multiple other neuron classes [29]. Its main function is to integrate information between extrinsic and intrinsic brain activity that leads to behavioral responses [30]. The function of RMGR therefore provides a plausible explanation why it connects to several modules.

An example of the utility of  $PC_{norm}$  in real-world networks was seen in fMRI brain networks, where we observed increased inter-subject correlation with the  $N$ -back task, compared to  $PC$  (Fig. 7A). This task is important for working memory performance [31], a higher-order cognitive process that engages a distributed brain network comprising symmetric and bilateral frontal and parietal cortices [32–34]. We found that  $PC_{norm}$  was relatively high in these regions compared to other areas of the cortex, but this was not as prominent as  $PC$ . Several reports also suggest that  $PC$  is correlated with working memory performance [5, 11, 35, 36]. For example, Shine et al. [11] reported increased  $PC$  in fronto-parietal networks during the  $N$ -back task, compared to the resting state. This suggests that the human brain may allocate more inter-modular connectivity to meet demands of this cognitively strenuous task. Other brain regions that expressed relatively high values of  $PC_{norm}$  were located in the cerebellum, mainly sub-region Crus I and Lobule VI (see Fig. 4A). These cerebellar sub-regions are also involved in working memory [37–39]. They display connectivity with frontal and parietal cortices [40, 41], and have preferentially expanded over recent evolutionary time [42].

Unlike  $PC$ , nodes that only have intra-modular connections will not have a  $PC_{norm}$  value of zero. This is because our randomization approach always returns a 'random' number to subtract from the original  $PC$  algorithm. Nevertheless, our simulations show that  $PC_{norm}$  return values close to zero for nodes that only have intra-modular connections, unless these are high-degree nodes (see upper left corner of Fig. 2C). We also found that, in real-world networks,  $PC_{norm}$  remains low when  $PC$  is zero (Supplementary Fig. 6). An alternative is to conduct a post-hoc correction setting  $PC_{norm}$  nodes to zero if corresponding  $PC$  values are zero.

We have released a user-friendly MATLAB code for calculation of  $PC_{norm}$ , available at [GitHub](#). The current version of  $PC_{norm}$  works for undirected/directed and weighted/binary networks. A limitation of  $PC_{norm}$  is the computational time needed for network randomizations, which can be particularly time consuming for large networks (i.e.  $> 1000$  nodes). To circumvent this issue, we added a parallel computing option in our code that allows for faster computation time, ensuring this process is tractable for all but the largest networks. Nevertheless, as network randomization procedures are becoming increasingly popular, it is imperative to continue optimizing computational capabilities when randomizing edges in large and complex networks.

## Methods

### Modular decomposition of networks

Before estimating  $PC$  and  $PC_{norm}$  (defined in the Results section - Eqs. 1 and 2), we parsed fMRI, *C.elegans* and airport networks into modules using a Louvain modularity optimization procedure [21] written as:

$$[Q, M] = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \quad (3)$$

where  $A_{ij}$  represents the binary connectivity edge between nodes  $i$  and  $j$ . Louvain modularity optimisation procedure finds nodes with greater intra-modular connectivity that is expected by chance using a null model where  $k_i k_j$  (the sum of the weights of the edges attached to nodes  $i$  and  $j$ ) is divided by  $2m$  (the sum of all of the edge weights in the graph).  $c_i$  and  $c_j$  are the modules associated with node  $i$  and  $j$ , and the module vector output is  $M$ .

$\gamma$  controls the spatial resolution of modules. A low  $\gamma$  ( $\gamma < 1.5$ ) results in a few large modules, whereas a high  $\gamma$  ( $\gamma > 1.5$ ) returns numerous small modules [20]. We validated our findings across a variety of  $\gamma$  parameters between 1 and 2 in 0.1 increments, [43]. The Louvain clustering method also contain heuristics that may cause run-to-run variability due to the degeneracy issue in complex network [44]. This provided us with a good opportunity to assess  $PC$  across a range of modular structures in the *C.elegans* and airport network where we only had a single network available for analysis. In these two networks, we obtained 1000 different modular outputs. In fMRI networks we computed a single modular output of each of the 1003 subjects using a consensus clustering approach with 1000 iterations. For each modularity run in *C.elegans* and airport networks, and fMRI network subject, we obtained a  $1 \times N$  vector where  $N$  is total number of nodes (i.e., 100 nodes for fMRI networks, 277 nodes for the *C.elegans* neuronal system and 500 nodes for the airports network) comprising  $m$ -numbers of modules ( $M = \text{all modules}$ ).

### Within-module degree $z$ -score

To compare  $PC$  and  $PC_{norm}$  with intra-modular connectivity we computed *within-module degree z-score* [4], i.e., the normalized connectivity of a node within its own module [4]. It is given by:

$$z_i = \frac{k_i(m_i) - \bar{k}(m_i)}{\sigma^{k(m_i)}}, \quad (4)$$

where  $m_i$  denotes the module containing  $i$ ;  $k_i(m)$  is the overall degree of node  $i$  in module  $m$  (where node  $i$  is located).  $\bar{k}(m_i)$  is the mean degree of all nodes in module  $m$  and  $\sigma^{k(m_i)}$  is the standard deviation of all nodes in module  $m$ .

### **Dataset 1: fMRI network**

We used 57.6 min (4800 time-points  $\times$  0.72s repetition time) of resting state fMRI data from 1003 healthy adults between ages of 22 and 35 years, from the Human Connectome Project [16], that we filtered between frequencies of 0.01 and 0.1 Hz. The fMRI data was distortion, motion, and field bias corrected and normalized into a common Montreal Neurological Institute space. An independent components analysis was conducted to parcellate the brain into 100 regions of interests (55 sub-cortical and 45 cortical brain regions). At the individual level, we generated an undirected binary brain network with 100 nodes using pair-wise Pearson correlation analysis of subject-specific fMRI time series over 100 brain regions. For our main analysis we thresholded networks at a sparsity level 20% (990 edges). We also derived networks at additional sparsity levels of 10% (495 edges), 30% (1485 edges) and 40% (1980 edges – Supplementary Fig. 4)

### **Dataset 2: *C.elegans* network**

We analyzed two-dimensional spatial representations of the global neuronal network (277 neurons, or nodes, and 2105 connections, or edges) of the nematode *C.elegans*, previously identified by electron microscope reconstructions [18]. This network is directed and binary – i.e., it contains both incoming and outgoing neural connections between brain areas of the *C.elegans*. In this study we only used outgoing connections.

### **Dataset 3: Airport network**

Airport network was constructed based on flights between the 500 busiest airports (nodes) in the world. Edges were based on the total passenger volume between airports. In this network there were a total of 24009 flights (edges). The existence of flight connections between airports is based on flights within one year from 1 July 2007 to 30 June 2008 [19]. An edge means that there is at least one flight between two airports. This network is directed and binary, meaning it contains incoming and outgoing flights connections between airports. In this study we only used outgoing connections.

## Acknowledgements

This study was supported by the National Health and Medical Research Council (NHMRC) of Australia (no 628952). The Florey Institute of Neuroscience and Mental Health acknowledges the strong support from the Victorian Government and in particular the funding from the Operational Infrastructure Support Grant. We also acknowledge the facilities, and the scientific and technical assistance of the National Imaging Facility (NIF) at the Florey node and The Victorian Biomedical Imaging Capability (VBIC). GJ is supported by an NHMRC practitioner's fellowship (no 1060312).

The primary fMRI data in this study was provided by the Human Connectome Project, WUMinn Consortium (1U54MH091657; Principal Investigators: David Van Essen and Kamil Ugurbil) funded by the 16 National Institutes of Health (NIH) institutes and centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We also thank Marcus Kaiser for releasing network datasets including *C.elegans* and airport networks.

## Author contributions

M.P., A.O., J.M.S., G.D.J., and A.Z. designed research; M.P., A.O., J.M.S., A.Z. performed research; M.P., A.O., analyzed data; and M.P., A.O., J.M.S., G.D.J., and A.Z. wrote the paper.

## Competing interests

The authors declare no conflict of interest.

## Data availability

All fMRI network data can be downloaded from: <https://www.humanconnectome.org>.

*C.elegans* and airport network data can be downloaded from <https://www.dynamic-connectome.org>.

$PC_{norm}$  can be downloaded from [https://github.com/omidvarnia/Dynamic\\_brain\\_connectivity\\_analysis](https://github.com/omidvarnia/Dynamic_brain_connectivity_analysis).

## References

- [1] Sporns, O. Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology* **23**, 162–171 (2013).
- [2] Tononi, G., Sporns, O. & Edelman, G. M. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5033–5037 (1994).
- [3] Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).
- [4] Guimerà, R. & Nunes Amaral, L. A. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
- [5] Bertolero, M. A., Yeo, B. T. T., Bassett, D. S. & D’Esposito, M. A mechanistic model of connector hubs, modularity and cognition. *Nature Human Behaviour* **2**, 765 (2018).
- [6] Guimerà, R., Mossa, S., Turtschi, A. & Amaral, L. a. N. The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7794–7799 (2005).
- [7] Kim J. S. & Kaiser M. From *Caenorhabditis elegans* to the human connectome: a specific modular organization increases metabolic, functional and developmental efficiency. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130529 (2014).
- [8] Pedersen, M., Zalesky, A., Omidvarnia, A. & Jackson, G. D. Multilayer network switching rate predicts brain performance. *Proceedings of the National Academy of Sciences* **115**, 13376–13381 (2018).
- [9] Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N. & Petersen, S. E. Evidence for hubs in human functional brain networks. *Neuron* **79** (2013).
- [10] Sethi, M., Pedersen, M. & Jackson, G. D. Polymicrogyric Cortex may Predispose to Seizures via Abnormal Network Topology: An fMRI Connectomics Study. *Epilepsia* **57**, e64–68 (2016).
- [11] Shine, J. M. *et al.* The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* **92**, 544–554 (2016).
- [12] Sohn, Y., Choi, M.-K., Ahn, Y.-Y., Lee, J. & Jeong, J. Topological Cluster Analysis Reveals the Systemic Organization of the *Caenorhabditis elegans* Connectome. *PLoS Computational Biology* **7** (2011).
- [13] Towilson, E. K., Vértés, P. E., Ahnert, S. E., Schafer, W. R. & Bullmore, E. T. The Rich Club of the *C. elegans* Neuronal Connectome. *The Journal of Neuroscience* **33**, 6380–6387 (2013).
- [14] Klimm, F., Borge-Holthoefer, J., Wessel, N., Kurths, J. & Zamora-López, G. Individual node’s contribution to the mesoscale of complex networks. *New Journal of Physics* **16**, 125006 (2014).
- [15] Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science (New York, N.Y.)* **296**, 910–913 (2002).
- [16] Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79 (2013).
- [17] Kötter, R. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* **2**, 127–144 (2004).
- [18] White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The Structure of the Nervous System of the Nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **314**, 1–340 (1986).

- [19] Marcelino, J. & Kaiser, M. Critical paths in a metapopulation model of H1n1: Efficiently delaying influenza spreading through flight cancellation. *PLoS currents* **4**, e4f8c9a2e1fca8 (2012).
- [20] Bassett, D. S. *et al.* Robust detection of dynamic community structure in networks. *Chaos (Woodbury, N.Y.)* **23**, 013142 (2013).
- [21] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).
- [22] Wilcox, R. R. Comparing dependent robust correlations. *British Journal of Mathematical and Statistical Psychology* **69**, 215–224 (2016).
- [23] Zalesky, A. *et al.* Whole-brain anatomical networks: does the choice of nodes matter? *NeuroImage* **50**, 970–983 (2010).
- [24] van den Heuvel, M. P. & Sporns, O. Rich-Club Organization of the Human Connectome. *Journal of Neuroscience* **31**, 15775–15786 (2011).
- [25] Arnatkeviciūtė, A., Fulcher, B. D., Pocock, R. & Fornito, A. Hub connectivity, neuronal diversity, and gene expression in the caenorhabditis elegans connectome. *PLoS computational biology* **14**, e1005989 (2018).
- [26] Bentley, B. *et al.* The multilayer connectome of caenorhabditis elegans. *PLoS computational biology* **12**, e1005283 (2016).
- [27] Bertolero, M. A., Yeo, B. T. T. & Desposito, M. The diverse club. *Nature Communications* **8**, 1277 (2017).
- [28] Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013).
- [29] Macosko, E. Z. *et al.* A Hub-and-Spoke Circuit Drives Pheromone Attraction and Social Behavior in *C. elegans*. *Nature* **458**, 1171–1175 (2009).
- [30] Lockery, S. R. Neuroscience: A social hub for worms. *Nature* **458**, 1124–1125 (2009).
- [31] Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology* **55**, 352–358 (1958).
- [32] Borst, J. P. & Anderson, J. R. Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences* **110**, 1628–1633 (2013).
- [33] LaBar, K. S., Gitelman, D. R., Parrish, T. B. & Mesulam, M. M. Neuroanatomic Overlap of Working Memory and Spatial Attention Networks: A Functional MRI Comparison within Subjects. *NeuroImage* **10**, 695–704 (1999).
- [34] Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E. & Lieberman, M. D. Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences* **109**, 1883–1888 (2012).
- [35] Cohen, J. R. & D’Esposito, M. The Segregation and Integration of Distinct Brain Networks and Their Relationship to Cognition. *Journal of Neuroscience* **36**, 12083–12094 (2016).
- [36] Stanley, M. L., Dagenbach, D., Lyday, R. G., Burdette, J. H. & Laurienti, P. J. Changes in global and regional modularity associated with increasing working memory load. *Frontiers in Human Neuroscience* **8** (2014).
- [37] Desmond, J. E. & Fiez, J. A. Neuroimaging studies of the cerebellum: language, learning and memory. *Trends in Cognitive Sciences* **2**, 355–362 (1998).

- [38] Marvel, C. L. & Desmond, J. E. Functional Topography of the Cerebellum in Verbal Working Memory. *Neuropsychology Review* **20**, 271–279 (2010).
- [39] Schmahmann, J. D. & Caplan, D. Cognition, emotion and the cerebellum. *Brain* **129**, 290–292 (2006).
- [40] Bernard, J. A. *et al.* Resting state cortico-cerebellar functional connectivity networks: a comparison of anatomical and self-organizing map approaches. *Frontiers in Neuroanatomy* **6**, 31 (2012).
- [41] Allen, G. *et al.* Magnetic resonance imaging of cerebellar-prefrontal and cerebellar-parietal functional connectivity. *NeuroImage* **28**, 39–48 (2005).
- [42] Shine, J. M. & Shine, R. Delegation to automaticity: the driving force for cognitive evolution? *Frontiers in Neuroscience* **8** (2014).
- [43] Lancichinetti, A. & Fortunato, S. Consensus clustering in complex networks. *Scientific Reports* **2** (2012).
- [44] Good, B. H., de Montjoye, Y.-A. & Clauset, A. Performance of modularity maximization in practical contexts. *Physical Review E* **81**, 046106 (2010).