

## RESEARCH ARTICLE

# Consistency and differences between centrality measures across distinct classes of networks

Stuart Oldham<sup>1\*</sup>, Ben Fulcher<sup>1,2</sup>, Linden Parkes<sup>1</sup>, Aurina Arnatkevičiūtė<sup>1</sup>, Chao Suo<sup>1</sup>, Alex Fornito<sup>1</sup>

**1** The Turner Institute for Brain and Mental Health, School of Psychological Sciences and Monash Biomedical Imaging, Monash University, Clayton, Victoria, Australia, **2** School of Physics, The University of Sydney, Sydney, New South Wales, Australia

\* [stuart.oldham@monash.edu](mailto:stuart.oldham@monash.edu)



## OPEN ACCESS

**Citation:** Oldham S, Fulcher B, Parkes L, Arnatkevičiūtė A, Suo C, Fornito A (2019) Consistency and differences between centrality measures across distinct classes of networks. PLoS ONE 14(7): e0220061. <https://doi.org/10.1371/journal.pone.0220061>

**Editor:** Satoru Hayasaka, University of Texas at Austin, UNITED STATES

**Received:** February 6, 2019

**Accepted:** July 8, 2019

**Published:** July 26, 2019

**Copyright:** © 2019 Oldham et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data generated or analysed in the current study are available in the figshare repository, [<https://figshare.com/s/22c5b72b574351d03edf>]. Code to process this data and reproduce all figures and analyses presented here is on github (<https://github.com/BMHLab/CentralityConsistency>). There are no restrictions on any of the datasets used, all can be freely accessed by other researchers.

**Funding:** BF was supported by a National Health and Medical Research Council (<https://www.nhmrc.gov.au/>).

## Abstract

The roles of different nodes within a network are often understood through centrality analysis, which aims to quantify the capacity of a node to influence, or be influenced by, other nodes via its connection topology. Many different centrality measures have been proposed, but the degree to which they offer unique information, and whether it is advantageous to use multiple centrality measures to define node roles, is unclear. Here we calculate correlations between 17 different centrality measures across 212 diverse real-world networks, examine how these correlations relate to variations in network density and global topology, and investigate whether nodes can be clustered into distinct classes according to their centrality profiles. We find that centrality measures are generally positively correlated to each other, the strength of these correlations varies across networks, and network modularity plays a key role in driving these cross-network variations. Data-driven clustering of nodes based on centrality profiles can distinguish different roles, including topological cores of highly central nodes and peripheries of less central nodes. Our findings illustrate how network topology shapes the pattern of correlations between centrality measures and demonstrate how a comparative approach to network centrality can inform the interpretation of nodal roles in complex networks.

## Introduction

Connections are often distributed heterogeneously across the elements of many real-world networks, endowing each node with a specific pattern of connectivity that constrains its role in the system. One popular way of characterizing the role of a node in a network is by using one or more measures of centrality. These measures aim to quantify the capacity of a node to influence, or be influenced by, other system elements by virtue of its connection topology [1–4]. Accordingly, centrality measures are often used to identify highly central or topologically important nodes, commonly referred to as hubs, that play a key role in many diverse kinds of networks. Examples include individuals who enhance the spread of disease in a population [5],

nhmrc.gov.au/) Early Career Fellowship (ID: 1089718); AF was supported by the Australian Research Council <http://www.arc.gov.au/> (ID: FT130100589) and National Health and Medical Research Council <https://www.nhmrc.gov.au/> (IDs: 1146292, 1050504, 1104580 and 1066779). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

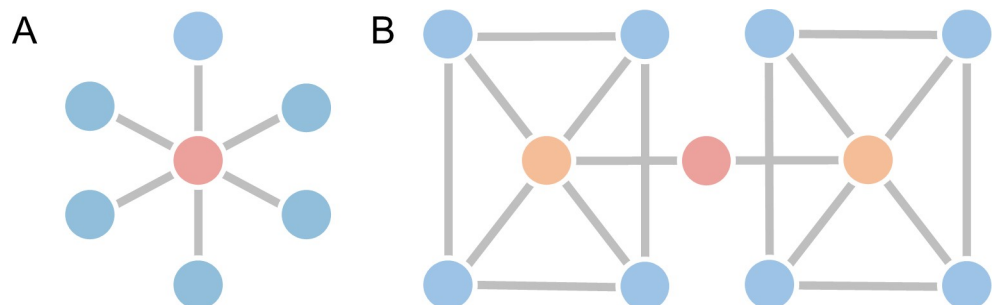
**Competing interests:** The authors have declared that no competing interests exist.

proteins that are indispensable for an organism's survival [6], researchers that are frequent collaborators in scientific collaboration networks [7], and brain regions thought to be important for regulating consciousness in functional brain networks [8,9].

Whether a node is ranked highly on a given centrality measure depends on the dynamical processes that are assumed to take place on the network [1]. For instance, nodes that are ranked as highly central according to measures that assume routing of information along shortest paths may not be ranked as highly by measures that assume diffusive dynamics [10,11]. Over 200 centrality measures have been proposed to date [12], each making different assumptions about network dynamics and the topological properties that are important for driving those dynamics. In addition, some centrality measures capture local information (e.g., with respect to immediate nodal neighbours), whereas others quantify how a node is situated within the global network context [13–15]. In theory, these measures should capture different aspects of network topology, and thus identify different kinds of node roles and, accordingly, different highly-central hub nodes. However, theoretical and conceptual differences between centrality measures do not always translate into empirical differences in real-world networks. For example, two different centrality measures may behave similarly on real-world networks, thus being practically redundant despite their distinct theoretical foundations.

The extent to which different centrality measures offer unique or redundant information depends on the topology of the network (e.g., see Fig 1). Past empirical work has investigated correlations between the scores assigned by different centrality measures in a number of real-world networks, such as scientific collaboration networks, airline networks, and internet routing networks, finding that the correlations between centrality measures—while typically moderate to high—can vary substantially from one network to another [16,17]. As an example, closeness and eigenvector centrality were very highly correlated in a network of collaborations between high-energy physicists ( $r = 0.91$ ), but not in a power grid network ( $r = -0.04$ ) [17]. The specific reasons for these variations in correlations between centrality measures, hereafter referred to as centrality measure correlations (CMCs), in different networks remains unclear.

What are the topological properties that influence the CMC structure of a network? Recent theory, developed in the analysis of social networks, has pointed to the *neighbourhood inclusion preorder* of a network as being a major determinant of CMCs (for a more detailed description, see *Methods*) [18–20]. This property can be quantified using the *majorization gap*, which measures the topological distance of a network from a threshold graph, a type of network in which all centrality measures should rank nodes the same way [18]. Networks that have a low majorization gap, and which are thus topologically similar to a threshold graph, exhibit higher



**Fig 1. Basic properties of topological centrality.** Panel A shows an example of a star network. The red node has maximal degree (greatest number of connections), closeness (is a short distance from other nodes) and betweenness (lies on many shortest-paths between nodes) in this network. In this case these three centrality measures are perfectly concordant. Panel B shows a network in which centrality measures are not concordant. The red node has the highest betweenness and closeness, but it has the lowest degree in the network.

<https://doi.org/10.1371/journal.pone.0220061.g001>

correlations between centrality measures [20]. Another body of work has shown that networks with a large spectral gap, quantified as the difference between the first and second eigenvalues of the adjacency matrix, have very high correlations between centrality measures that quantify walks between nodes [21–23] (for example, subgraph and eigenvector centrality). Clustered, modular networks can reduce CMCs by dissociating measures that quantify centrality within local neighbourhoods of nodes (e.g., degree, leverage) from those that index centrality across the entire network (e.g., betweenness, closeness). This is because a node may have high local centrality (highly connected with nodes in the same module) but low global centrality (unconnected to nodes in other modules), or vice-versa [24]. Other studies have examined the role of network edge density and the impact of specific node or edge removals on the network [14,25–28].

While numerous studies have investigated how different centrality measures are related [16,17,20,27,29–33], the extent to which any association between topology and CMC structure generalizes beyond this past work is unclear, as these studies have typically focused only on specific network classes (e.g., social, synthetic), used networks varying within a limited range of sizes and densities, explored just a few types of network organization, or examined a small subset of centrality measures. A systematic evaluation of CMCs, quantified across a broad array of centrality metrics and in a large set of different classes of networks, has not been performed. Furthermore, given the abundance of centrality measures proposed, many of which are highly correlated to each other when applied to real-world networks, it is important to understand whether there are benefits to using multiple centrality measures, or whether there is a reduced, canonical set of measures for capturing nodal roles in most applications. Past research has found that using multiple centrality measures to define multivariate profiles can offer a better description of nodal roles in the network [34,35]. Broad, comparative studies—such as those performed recently for time-series analysis [36]—allow us to uncover empirical relationships between the large and interdisciplinary literature on centrality measures for network data. While the selection of which centrality measure to apply to a given network analysis task is typically done subjectively, the combination of many centrality measures together can offer a more systematic and comprehensive framework in which the most useful measures can be informed more objectively from the empirical structure of a given network.

In this article, we evaluate 17 different centrality measures across 212 networks. We examine how CMCs vary across the networks and characterize the association between global topological properties of each network and CMC variation. We also examine how multivariate profiling of nodal centrality can be used to gain insight into the roles that different nodes play a given network.

## Methods

### Centrality measures

We used 17 different centrality measures to analyse each network, focusing on centrality measures that are commonly used in the network science literature, or which have received recent interest. Each measure used is listed in [Table 1](#); definitions and further details are in [S1 Text](#). Analysis was performed in MATLAB 2017a. The code for all centrality measures were either obtained from the Brain Connectivity Toolbox (BCT) [37], MatlabBGL library, or were written in custom code, available at [<https://github.com/BMHLab/CentralityConsistency>]. All data generated or analysed in the current study are available in the figshare repository, [<https://figshare.com/s/22c5b72b574351d03edf>].

Centrality measures are often defined in relation to the different ways in which information is thought to propagate across nodes, which can occur through: (1) *walks*, which follow an

**Table 1. Definitions for centrality measures.**

Centrality name	Characteristics of a central node	Equation
Degree (DC)	Connected to many other nodes [3]	$DC_i = d_i = \sum_{j \neq i} A_{ij}$
Eigenvector (EC)	Connected to many other nodes and/or to other high-degree nodes [40]	$EC_i = \frac{1}{\lambda_1} \sum_j A_{ji} v_j$
Katz (KC)	Connected to many other nodes and/or connected to other high-degree [41]	$KC_i = \alpha \sum_j A_{ji} v_j + \beta$
PageRank (PR)	Connected to many other nodes and connected to other high-degree nodes [42]	$PR_i = \alpha \sum_j A_{ji} \frac{v_j}{k_j} + \beta$
Leverage (LC)	Has a higher degree than its neighbours [43]	$LC_i = \frac{1}{d_i} \sum_{j \in h(i)} \frac{d_i - d_j}{d_i + d_j}$
H-index (HC)	Connected to many other high-degree nodes [44]	$HC_i = \max_{1 \leq h \leq d_i} \min( \mathcal{N}_{\geq h}(i) , h)$
Laplacian (LAPC)	Removal of this node would greatly impair the network [45,46]	$LAPC_i = d_i^2 + d_i + 2 \sum_{j \in \mathcal{N}(i)} d_j$
Shortest-path Closeness (CC)	Low average shortest path length to other nodes in the network [47]	$CC_i = \frac{1}{\sum_{j \neq i} l_{ij}}$
Subgraph (SC)	Involved in many closed short-range walks [48]	$SC_i = [e^A]_{ii}$
Participation coefficient (PC)	Connections distributed across different topological modules [24]	$PC_i = 1 - \sum_{m=1}^M \left( \frac{d_i(m)}{d(i)} \right)^2$
Total Communicability (TCC)	Can be easily reached by a walk from any other node [21]	$TCC_i = \sum_j [e^A]_{ji}$
Random-walk Closeness (RWCC)	Can be easily reached by a random-walk from any other node [49,50]	$RWCC_i = \frac{1}{\sum_j H_{ji}}$
Information (IC)	Can be easily reached by paths from other nodes [51]	$IC_i = \left( C_{ii} + \frac{\sum_{j \neq i} c_{ij}^{-2} \sum_{k \neq i} c_{kj} \right)^{-1}$
Shortest-path Betweenness (BC)	Lies on many shortest topological paths linking other node pairs [3]	$BC_i = \sum_{p \neq i, q \neq i} \frac{g_{pq}(i)}{g_{pq}}$
Communicability betweenness (CBC)	Takes part in many walks between pairs of other nodes [52]	$CBC_i = \frac{1}{\dot{C}} \sum_p \sum_q \frac{G_{piq}}{G_{pq}}, p \neq q, q \neq i$
Random-walk Betweenness (RWBC)	Takes part in many random walks between pairs of other nodes [53]	$RWBC_i = \frac{\sum_{p \neq q} I_i^{(pq)}}{\frac{1}{2} N(N-1)}$
Bridging (BridC)	Forms key links between high degree nodes [54]	$BridC_i = BC_i \times BC_i$

$A$  = adjacency matrix;  $d_i$  = degree of node  $i$ ;  $\lambda_1$  = leading eigenvalue of  $A$ ;  $v$  = leading eigenvector of  $A$ ;  $\alpha$  = penalty on distant connections to a node's centrality score;  $\beta$  = preassigned centrality constant;  $h(i)$  = the neighbours of node  $i$ ;  $\mathcal{N}_{\geq h}(i)$  = neighbours of node  $i$  which have at least a degree of  $h$ ;  $N$  = number of nodes in a network;  $l_{ij}$  = length of the shortest between nodes  $i$  and  $j$ ;  $e^A$  = matrix exponential of  $A$ ;  $M$  = number of modules in a network;  $d_i(m)$  = neighbours of node  $i$  which are part of module  $m$ ;  $H$  = the matrix of mean-first passage times between nodes in a network;  $C = (L+J)^{-1}$  where  $L$  is the Laplacian of  $A$  and  $J$  is a  $N \times N$  matrix with all elements equal to one;  $g_{pq}$  = the number of shortest-paths between nodes  $p$  and  $q$ ;  $g_{pq}(i)$  = the number of shortest-paths between nodes  $p$  and  $q$  which pass through  $i$ ;  $G_{pq}$  = number of walks between nodes  $p$  and  $q$ ;  $G_{piq}$  = number of walks between nodes  $p$  and  $q$  involving node  $i$ ;  $\dot{C} = (N-1)^2 - (N-1)$  which is a normalisation term;  $I_i^{(pq)}$  = current flowing through nodes  $p$  and  $q$  which passes through node  $i$ ;  $BC_i = d_i^{-1} / \sum_{j \in \mathcal{N}(i)} d_j^{-1}$ . All measures here are defined for unweighted networks, see [S1 Text](#) for information on weighted versions.

<https://doi.org/10.1371/journal.pone.0220061.t001>

unrestricted trajectory through the network; (2) *trails*, which can return to a visited node but cannot reuse an edge; and (3) *paths*, which cannot visit a node or edge more than once [1]. Thus, paths are a subset of trails which, in turn, are a subset of walks. We sought to include measures based on these different propagation approaches, although most centrality measures developed to date have focused on walks and paths.

While not typically thought of as a centrality index, the *participation coefficient* was also included in our set of centrality measures for comparison, as it is frequently used as a measure of nodal roles in networks with community structure [4,24]. The participation coefficient quantifies the distribution of a node's connections across different topological modules of the

network, where the modules are defined using a specific community detection algorithm (for a review of community detection algorithms see [38]). The participation coefficient was first introduced to distinguish between different types of network hubs [24] and has been proposed as a singular measure for defining hubs in some classes of networks, such as those based on correlations [39].

## Network data

Nearly all networks were obtained from freely-available sources. We examined 107 networks compiled by Ghasemian and colleagues [55] from the Index of Complex Networks (ICON) [56], together with a further 104 networks sourced by searching ICON for networks of varying sizes and domains. An additional network, the human structural brain network, was generated from diffusion-weighted magnetic resonance imaging data from the Human Connectome Project [57] (see [S1 Text](#) for details). Thus, we considered a total of 212 networks. Each network, comprising  $N$  nodes and  $E$  edges, was represented as an  $N \times N$  adjacency matrix. For the main analysis, each network was treated as unweighted (any edge weight information was removed) and undirected (any unidirectional edges were made bi-directional). If the network was comprised of multiple components, only the largest connected component was considered. In addition, weighted analysis was performed for 39 networks for which edge-weight information was available.

To examine the extent to which simple network properties—such as number of nodes, edges, and degree/strength distribution—contribute to the CMCs for a network, we compared the empirical networks to a set of matched surrogate networks. For each empirical network, we generated 100 unconstrained and 100 constrained surrogate networks. Unconstrained surrogate networks were created using a variant of the Erdős-Rényi generative model [58] which guaranteed the network would be non-fragmented, while preserving the number of nodes, number of edges, and the distribution of edge weights of the original network. Constrained surrogate networks were generated using the Maslov-Sneppen algorithm [59] for unweighted networks and a modified version for weighted networks [37]. The constrained surrogates preserve the number of nodes and edges, in addition to the degree sequence and approximate node strength (weighted degree) distributions. See [S1 Text](#) for more on the surrogate generation algorithms. Due to the computational complexity of calculating random-walk betweenness centrality and communicability betweenness centrality, we did not compute these measures for the surrogate networks.

## Centrality Measure Correlations (CMCs)

We used Spearman's  $\rho$  to calculate the correlation between the nodal scores assigned by any two centrality measures. This statistic was used to quantify CMCs because many such relationships were nonlinear yet almost always monotonic, and many centrality metrics have a non-Gaussian distribution [20]. CMCs were computed in every network for all pairs of centrality metrics. To find which centrality measures were consistently highly correlated across networks (indicating redundancy), we took the mean CMC for each pair of metrics across all networks, which we term the *mean between-network CMC*. We also quantified the variability of CMCs across networks as the *between-network CMC standard deviation*.

As an additional supporting analysis, we conducted a Principal Component Analysis (PCA) on the centrality data. While centrality measures often have non-linear relationships and contain outliers—properties not ideally suited to PCA [60,61]—we conducted this analysis to evaluate, in a preliminary way, how the different measures grouped together based on linear covariance. In line with previous work, a PCA was run separately for each network on the z-scored centrality measures [13].

## Assessing the relationship between network topology and CMCs

Given the assumed relationship between network topology and CMCs (e.g., Fig 1), we examined how CMCs vary as a function of eight different global network properties: connection density, assortativity, clustering, connection density, global efficiency, diffusion efficiency, modularity, majorization gap, and spectral gap. Further details on how these global topological properties were calculated can be found in S1 Text. Briefly, the connection density of a network,  $\kappa$ , is the proportion of connections that are present in a network relative to the total number of possible connections. Previous work has shown that networks with higher density show higher CMCs [27]. In the limit of  $\kappa = 1$ , the network is fully connected and all nodes are identical. As the density decreases, there is more variability in how the connections in the network can be arranged, and this is likely to result in centrality measures diverging and thus becoming less correlated.

Assortativity, clustering and global efficiency are commonly used descriptors of global network topology. Assortativity measures the extent to which nodes preferentially connect to other nodes with similar degree [62]. Clustering measures the proportion of closed triangles present in the network and is often taken as a measure of cliquish connectivity [63]. Global efficiency is inversely related to the characteristic path length of a network and is thus a useful descriptor for networks characterized by shortest-path routing [64]. Diffusion efficiency is an analogous measure that captures the efficiency of a network in supporting communication governed by a diffusion process [11].

Modularity is the extent to which a network contains groups of nodes that are densely interconnected with each other but sparsely connected to nodes outside the group [62]. Prior work has indicated that networks with stronger modularity show weaker CMCs [24]. Modules can enhance topological heterogeneity in a network, dissociating centrality metrics that favour high within-module connectivity (high local neighbourhood connectivity) from high between-module connectivity (globally integrative connectivity). We quantified modularity using the widely-used  $Q$  metric [65], and modules were identified using the Louvain algorithm [66] combined with a consensus clustering procedure (50 iterations with  $\tau = 0.4$ ) [67,68] to address algorithmic degeneracy [69] (see S1 Text).

The majorization gap quantifies the distance between an empirical network and an idealized network, called a threshold graph [20]. Threshold graphs are formed by adding nodes to a network, one at a time, such that the new node either connects to all existing nodes or connects to no other nodes (see S1 Fig for an example). Threshold graphs preserve a property known as the neighbourhood-inclusion preorder, which is argued to form the basis of centrality rankings [18,19]. If the neighbours of node  $j$  are a subset of the neighbors of node  $i$ , then node  $i$  is said to dominate node  $j$ , and must have a greater or equivalent level of centrality. The neighbourhood inclusion preorder is the rank ordering of nodes in terms of these dominance relationships, such that nodes that are not dominated by any others are ranked first and are thus more central. Nodes that are dominated by many others are ranked last, and are thus least central (e.g., S2 Fig). As this preorder is complete in threshold graphs—i.e., a dominance relationship can be established for every pair of nodes—the centrality rankings of all nodes across different measures in these networks is perfectly concordant. Thus, networks with a larger majorization gap will be more topologically distant from a threshold graph and should have lower CMCs.

The final property investigated was the spectral gap. This property quantifies the quality of a network's 'expansion properties'; namely, whether a network is simultaneously sparse and well-connected. A large spectral gap is indicative of a network being a good expander. Such networks lack bottlenecks—nodes/edges that, if removed, will fragment the network. A larger

spectral gap has been associated with higher correlations between walk-based centrality measures [21–23].

To combine the overall similarity of all pairs of centrality measures into a single value for a network, we took the mean of every CMC within each network to obtain the *mean within-network CMC*. A higher mean within-network CMC indicates that, on average, centrality measures are highly correlated in a network. This value was then correlated with each global topological descriptor. To determine which specific topological descriptor was the best predictor of variations in mean CMC across networks, we used multiple linear regression. In secondary analyses, we examined whether specific CMCs correlated with variations in global topology across networks.

As simple network properties like edge density and the degree/strength distribution can account for many higher-order features network topology, we compared the CMCs of empirical networks to matched surrogate networks. The unconstrained model can be used to determine whether the relationship is explained simply by variations in size and density across networks, while the constrained surrogates can be used to examine the impact of degree sequence and strength distribution in driving this relationship. To allow comparison between different networks and their associated surrogates, we calculated the difference of the empirical network properties/mean within-network CMCs compared to the mean value obtained in each of the surrogates.

**Clustering nodes based on their centrality profiles.** Finally, we investigated whether combining multiple centrality measures into a multivariate ‘centrality profile’ for each node could be used to meaningfully cluster nodes into groups with distinct topological roles. Centrality scores were converted to ranks and hierarchical clustering was performed using Ward’s minimum variance method [70] for Euclidean distances between pairs of ranked centrality metrics. For visualization, the Davies-Bouldin (DB) index [71] was used to determine a specific resolution to cut the dendrogram and investigate the resulting clusters. The DB index is a ratio of intra-cluster similarity to inter-cluster differences for a given clustering solution; lower values of the DB indicate a better clustering solution. We note that there are many different algorithms for clustering data (including alternative heuristics for forming clusters from a dendrogram) and for dendrogram cutting [72]. Our goal is not to determine any particular clustering solution or approach as robust or optimal, but rather to demonstrate how clustering of centrality profiles may aid in identifying subsets of nodes with distinct topological roles. A forced-directed algorithm was used to visualize node roles in the context of the broader topology of the network [73].

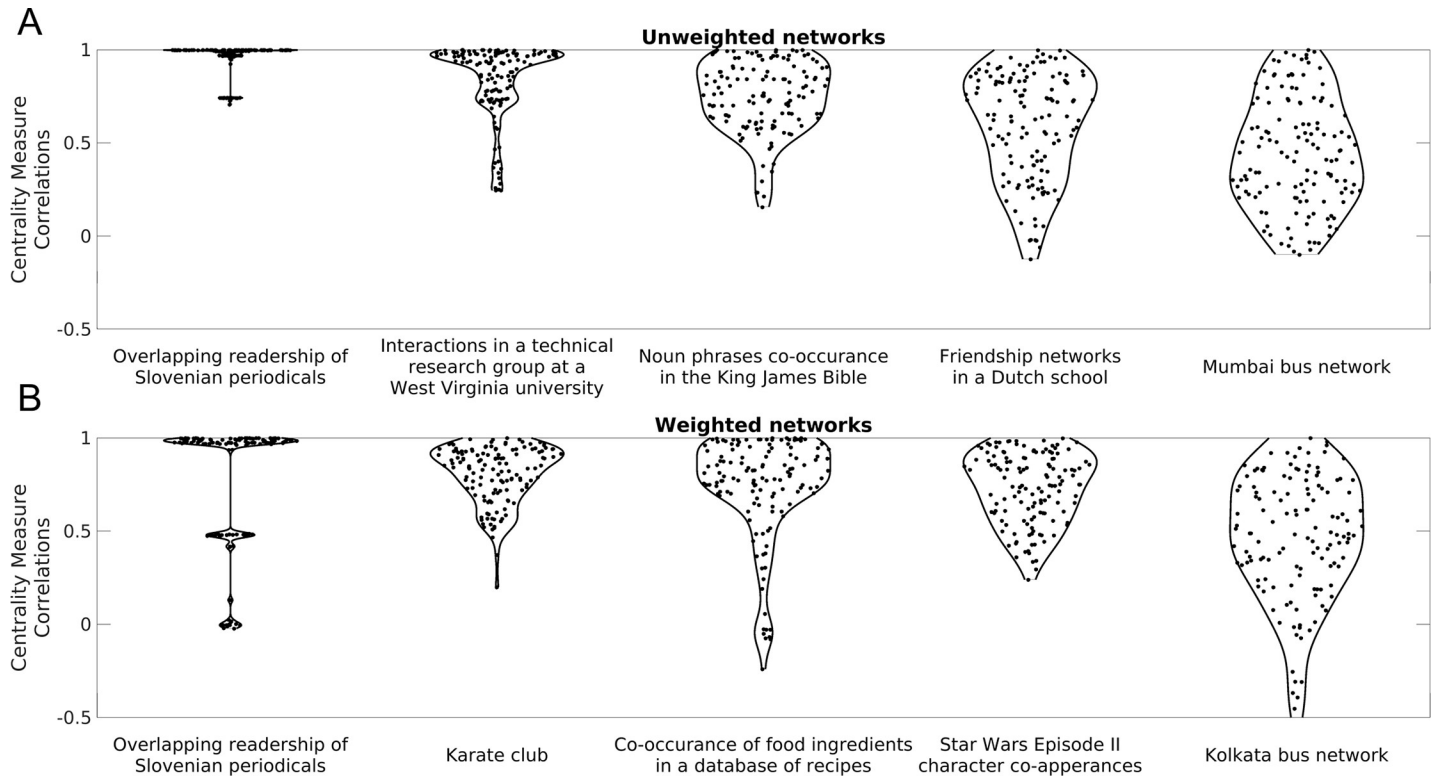
## Results

### Correlations between centrality measures

First, to examine the similarity of centrality measures across different networks, we calculated Spearman correlations between each of the 17 measures listed in Table 1 across each of the 212 networks. All 212 networks were analysed in unweighted form. A separate weighted centrality analysis was performed for 39 of these networks with edge-weight information.

Fig 2 shows the distribution of CMCs of five example unweighted and weighted networks. The distributions of CMCs for all networks are shown in S3 Fig. These results indicate that, despite a general trend for most networks to have high and mostly positive CMCs, there is considerable heterogeneity in CMC patterns across different networks, as previously reported [16,17]. This variability did not clearly map on to the natural class of the network (i.e., whether the network is social, biological technological, etc; S3 Fig).

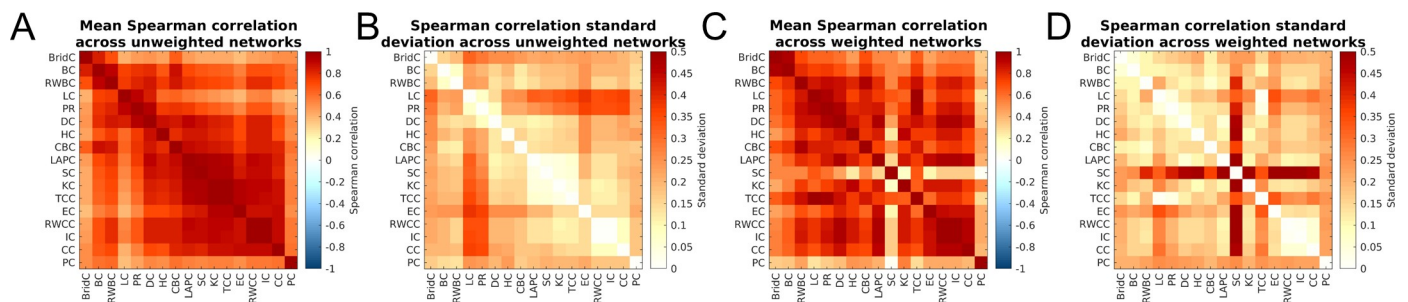
To determine which pairs of centrality measures were consistently correlated across networks, we calculated the mean between-network CMC (the mean CMC for each pair of



**Fig 2. Distributions of Centrality Measure Correlations (CMCs) for example unweighted and weighted networks.** Distributions of CMCs for every pair of centrality measures for five example unweighted (panel A); and weighted networks (panel B). Networks have been ordered from highest (left) to lowest (right) median CMC.

<https://doi.org/10.1371/journal.pone.0220061.g002>

measures across all networks) and standard deviation (standard deviation of CMCs across networks) for each pair of metrics in unweighted (Fig 3A and 3C) and weighted (Fig 3B and 3D) networks. Most measures show moderate-to-high correlations across all networks, with 97% of all mean CMCs exceeding 0.5 in unweighted networks and 80% in weighted networks. Weighted CMCs were slightly weaker than their unweighted counterparts. The PCA also indicated that centrality measures are highly interrelated, with the first principal component (PC1)—on which nearly all measures uniformly loaded—explaining 45–93% of the variance across different networks. More heterogeneous loadings were observed for the second and third components (see S1 Text; S4 Fig). For the 39 networks with edge weight information, we compared the unweighted and weighted centrality measures. Individual unweighted and



**Fig 3. Mean and standard deviation of between-network CMCs.** Panels A and B show the between-network CMC mean and standard deviation for unweighted measures, respectively. Panels C and D show the between-network CMCs mean and standard deviation for weighted measures, respectively.

<https://doi.org/10.1371/journal.pone.0220061.g003>



weighted measures were highly correlated (S5 Fig), as were the weighted and unweighted mean within-network CMCs for each network (S6 Fig).

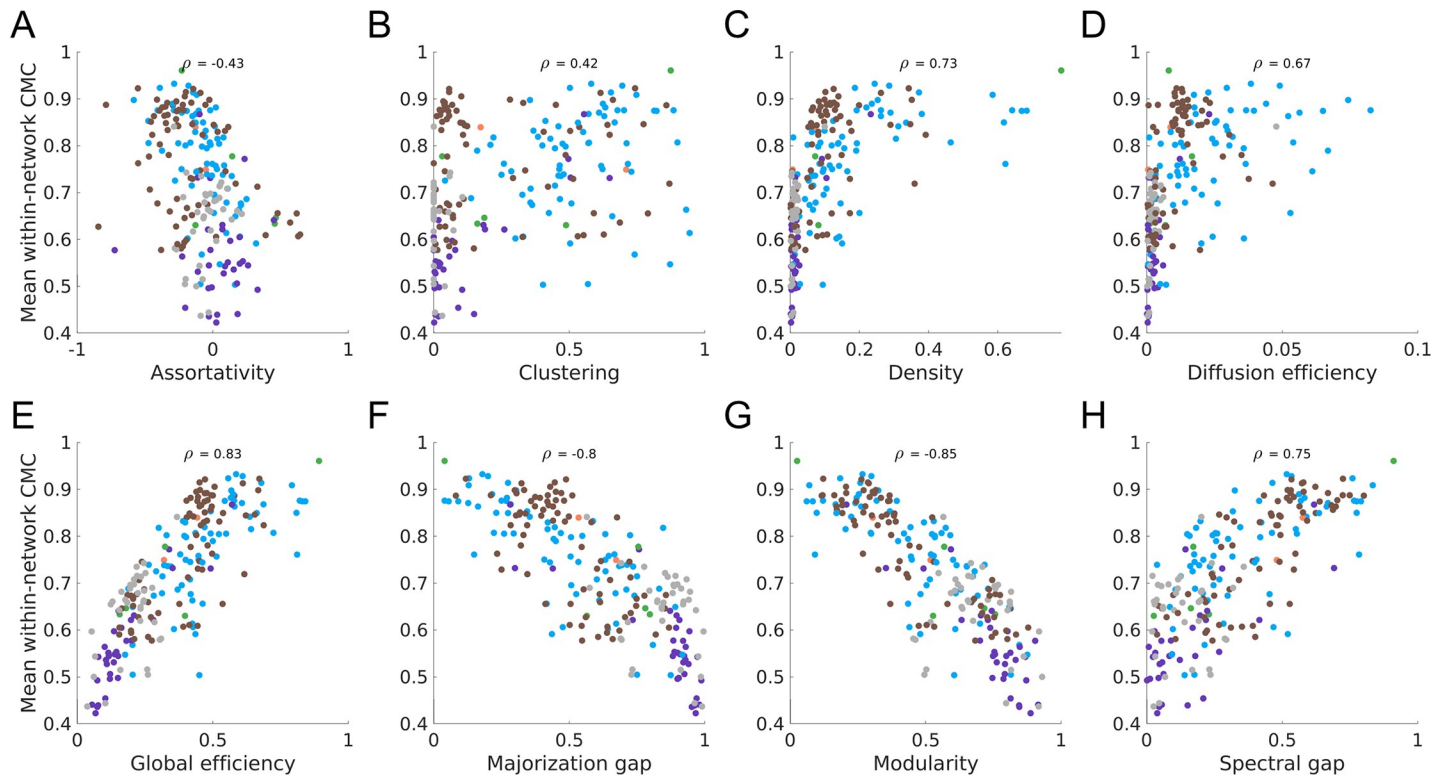
Several pairs of centrality measures displayed notable relationships. First, random-walk closeness centrality (RWCC) and information centrality (IC) were very highly correlated across networks (ranging from 0.88–1 with a mean correlation of 0.998 in unweighted networks and ranging from 0.937–1 with a mean correlation of 0.996 in weighted networks). Thus, these two theoretically-related measures [74] are practically redundant in most real-world scenarios. Other pairs, like Katz centrality (KC) and total communicability centrality (TCC), were also highly correlated across the wide range of unweighted networks analysed (all  $\rho > 0.98$ ). The participation coefficient and bridging centrality generally had the lowest average correlation with other measures, likely because they are conceptually distinct, and in the case of the participation coefficient, depend on a modular decomposition of the network. Subgraph centrality in weighted networks showed low correlations with other measures, suggesting it may be capturing a unique aspect of node centrality.

### Network topology and CMCs

We now examine how variations in CMCs across different networks relate to differences in the global topological properties of those networks. Specifically, we consider how the mean within-network CMC (the average of all pairwise CMCs within a network) relates to the following eight global network properties: connection density, assortativity, clustering, global efficiency, diffusion efficiency, modularity, majorization gap, and spectral gap.

In unweighted networks, higher mean within-network CMC was correlated with lower values of assortativity, majorization gap, and modularity, and higher values of clustering, density, diffusion efficiency, global efficiency, and spectral gap (Fig 4). Similar results were obtained for weighted networks (S7 Fig), with some exceptions. First, the correlation between global efficiency and mean within-network CMC was among the strongest for unweighted networks but among the weakest for weighted networks. Conversely, the correlation between assortativity and mean within-network CMC was strong for weighted networks, but weak for unweighted networks. Weighted clustering showed no relationship with CMCs once outliers were removed. Post-hoc analyses indicated that many individual pairs of CMCs correlated with network properties, showing that the relationship between network properties and mean CMCs is representative of a general trend across most pairs of centrality measures, and not driven by a small subset of CMCs (S8 Fig for unweighted and S9 Fig for weighted). However, CMCs involving bridging centrality or the participation coefficient had weak correlations with nearly all global properties in both unweighted and weighted networks, further suggesting that these measures may capture a unique aspect of nodal centrality. We also compared the amount of variance explained by PC1 (as a proxy for the unidimensional nature of centrality) in each network to each network property. These results were highly similar to those observed when using the mean within-network CMCs (S10 and S11 Figs).

We used multiple linear regression to quantify the unique contributions of each topological descriptor to CMC variability across networks (note: network density and diffusion efficiency were excluded due to strong non-linear associations with CMCs). In unweighted networks, modularity was the only significant predictor of mean within-network CMCs (Table 1 in S1 Text). As modularity and the majorization gap were highly correlated (S12 Fig), we reran the model excluding one of these properties each time, and found that only modularity was a significant predictor of network CMCs (Table 1 in S1 Text). In weighted networks, weighted assortativity explained the most variance in network CMCs. Due to collinearity, modularity and majorization gap were included in separate models. Both were significant predictors in



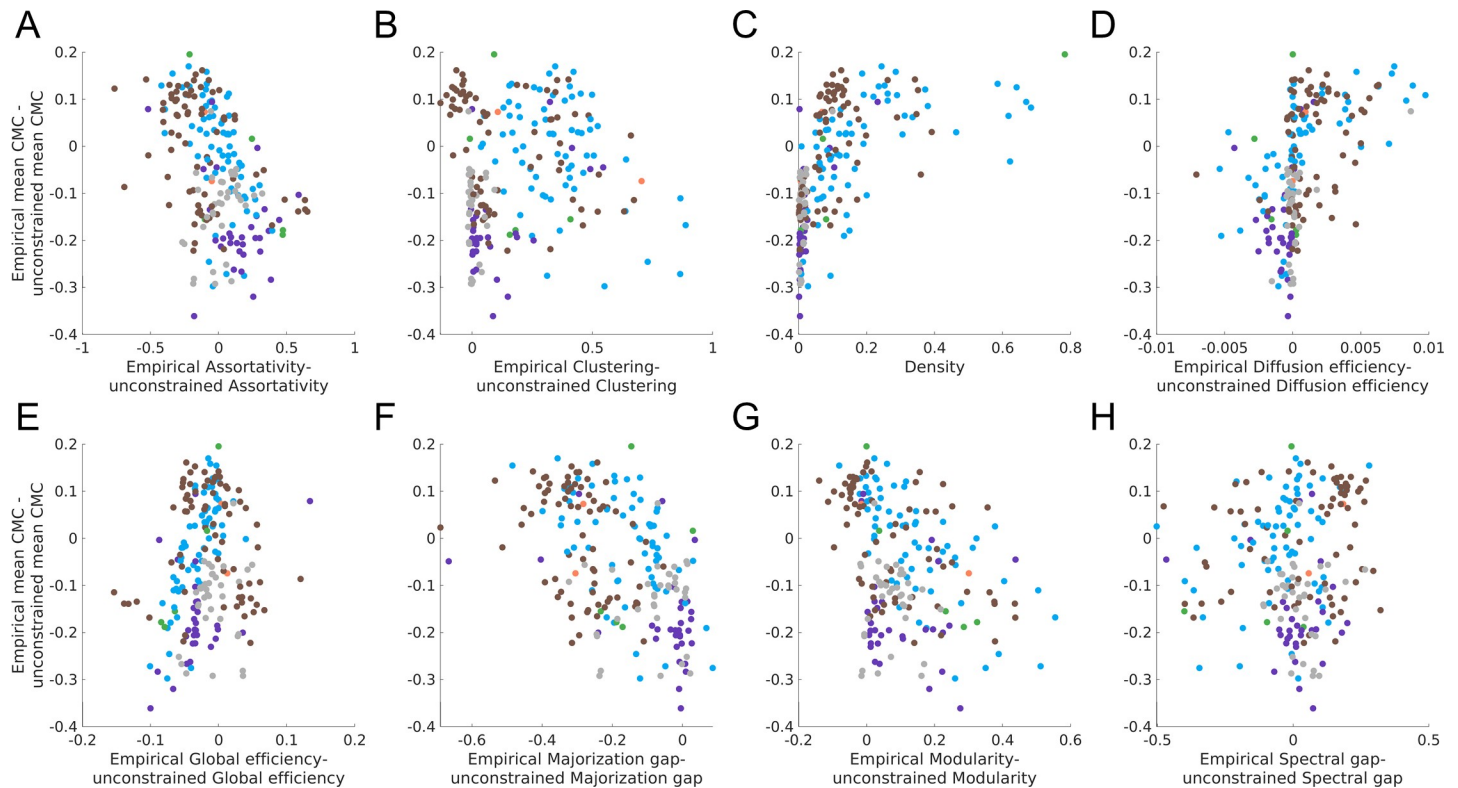
**Fig 4. Association between mean within-network CMC and network properties in unweighted networks.** The association between the mean within-network CMC (the average of all CMCs within a single network) and each of the global topological properties. Networks are coloured by their natural category (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).

<https://doi.org/10.1371/journal.pone.0220061.g004>

these models, with the former accounting for slightly less variance than the latter (49% vs 55%) (Table 2 in [S1 Text](#)).

To ensure that the associations between the mean within-network CMC and global topology could not be explained by lower-order features (e.g., density of the network or degree sequence), we examined these associations in surrogate networks matched for number of nodes, number of edges, edge weight distribution (unconstrained surrogate), and degree sequence and strength distribution (constrained surrogate). We compared the mean within-network CMCs and each network property in empirical networks to those obtained in the surrogates. Specifically, we calculated the difference between the mean within-network CMC/network property in the empirical network and the corresponding mean values of the surrogates. A difference greater than zero means the property was higher in the empirical network than the surrogates; conversely, if it was less than zero it was higher in the surrogate networks. A difference close to zero indicates the property is simply a side-effect of the network's density (for unconstrained surrogates) or degree/strength distribution (for constrained surrogates). These results are shown in [Figs 5 and 6](#) for unweighted network while results for weighted network surrogates are presented in [S12 and S13 Figs](#) respectively.

There are three major results from this comparison to the surrogates. First, for most networks, the mean within-network CMC of the surrogate networks (both constrained and unconstrained) was higher or equivalent to the respective matched empirical network ([Figs 5 and 6](#)). Second, unconstrained surrogates also had a higher majorization gap than the empirical networks. Finally, despite the empirical networks and constrained surrogates having the exact same majorization gap (due to the majorization gap being solely determined by the



**Fig 5. Difference between unweighted empirical and unconstrained surrogates in mean within-network CMC and network properties.** The y-axis of each plot shows the difference between the empirical networks and unconstrained surrogates mean within-network CMC. The x-axis shows the difference between the empirical networks and unconstrained surrogates on a particular property (except for panel C as the unconstrained surrogates have the same density as the empirical network). On both axis, except for the x-axis in panel C, a negative value indicates the empirical network had a lower value than the mean value of the surrogates, while a positive value indicates the empirical networks had a larger value. Points are coloured by the natural category of the empirical network (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).

<https://doi.org/10.1371/journal.pone.0220061.g005>

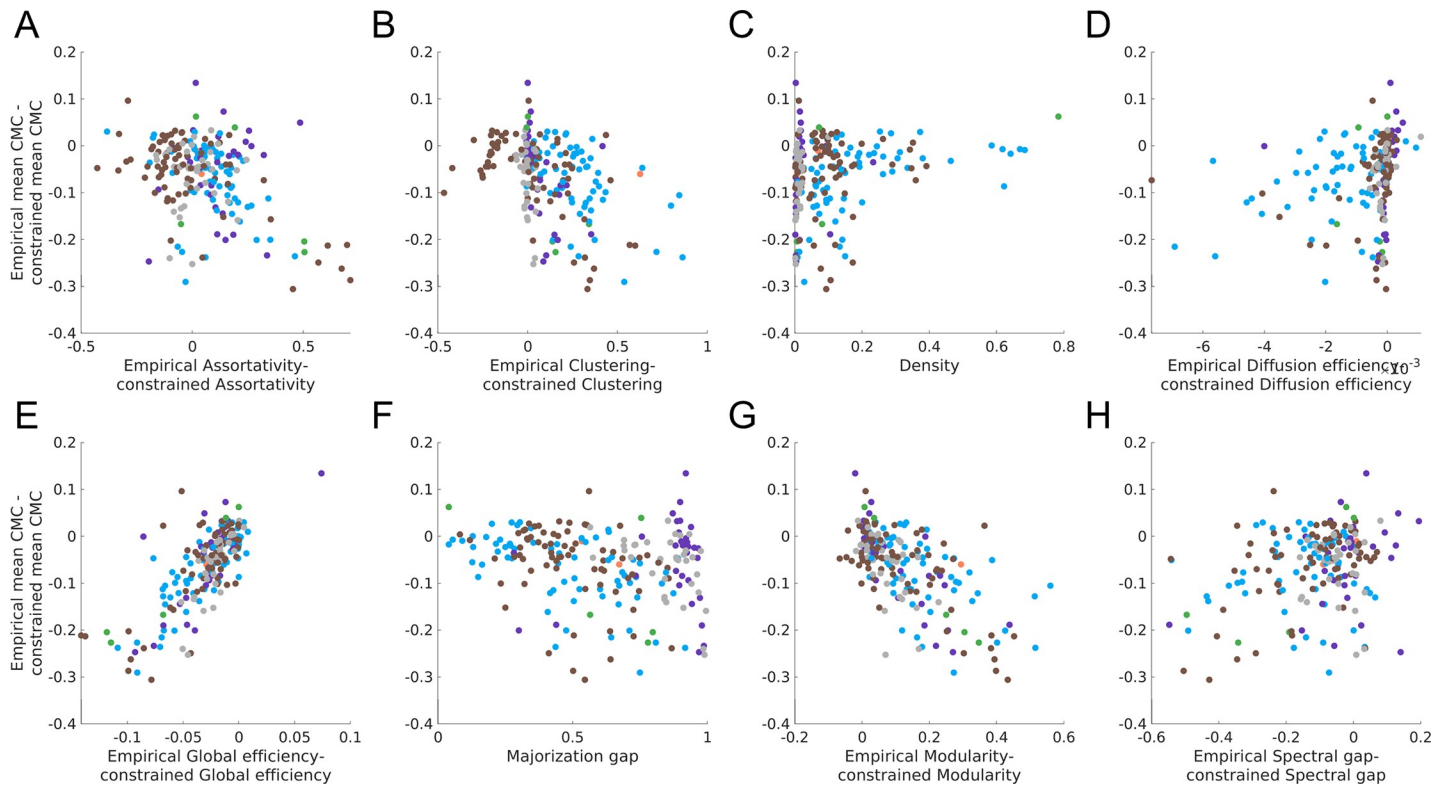
degree sequence of a network), empirical networks often had lower CMCs. Together, these results counter theoretical expectations that a higher majorization gap should be associated with lower CMCs.

### Centrality-based clustering of nodes

We now use hierarchical clustering to investigate whether multiple centrality measures can be used in combination to identify distinct roles for nodes. Due to the consistent high correlations ( $\rho > 0.99$ ) between random-walk closeness and information centrality, we excluded random-walk closeness from this analysis.

In most networks, the Davies-Bouldin (DB) criterion, a measure of the quality of a given clustering solution, suggested a two-cluster solution. Nearly all networks contained a subset of nodes with high scores across most measures, and another subset with low scores across most measures. The two-cluster solution often favoured one of these groups, such that either all nodes with low centrality were grouped in one cluster and the remaining nodes in the other (e.g., Fig 7), or vice-versa (e.g., Fig 8). Such subsets were also apparent when examining finer-grained clustering solutions.

While a putative core of high-scoring nodes and a periphery of low-scoring nodes was consistently found across nodes and clustering resolutions, distinct patterns were found for nodes interposed between these two subsets across different networks. Broadly these patterns can be



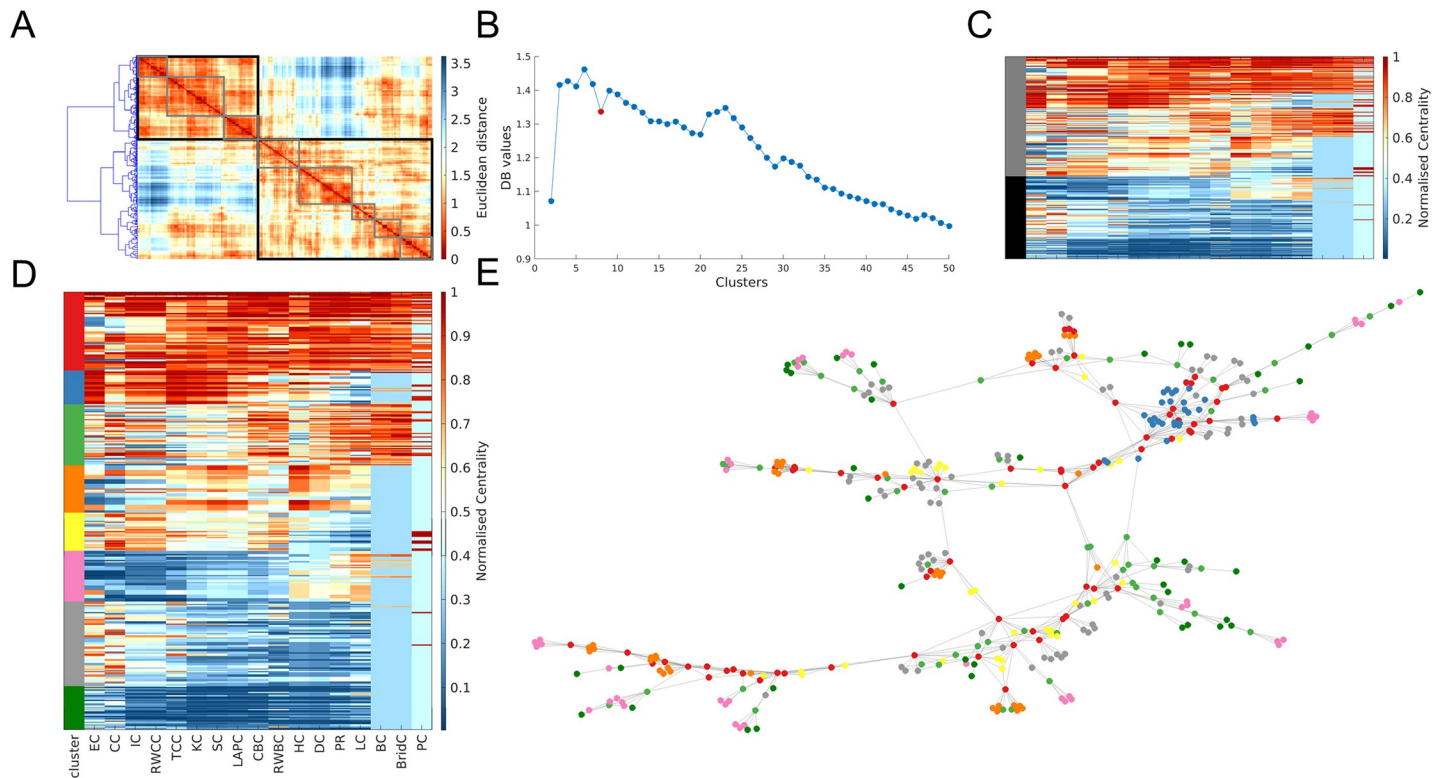
**Fig 6. Difference between unweighted empirical and constrained surrogates in mean within-network CMC and network properties.** The y-axis of each plot shows the difference between the empirical networks and constrained surrogates mean within-network CMC. The x-axis shows the difference between the empirical networks and constrained surrogates on a particular property (except for panels C and F as the constrained surrogates have the same density and majorization gap as the empirical network). On both axis, except for the x-axis in panels C and F, a negative value indicates the empirical network had a lower value than the mean value of the surrogates, while a positive value indicates the empirical networks had a larger value. Points are coloured by the natural category of the empirical network (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).

<https://doi.org/10.1371/journal.pone.0220061.g006>

classified into two types, characterized by either (a) a gradual progression from high-scoring core nodes to low-scoring periphery nodes (Fig 8A, see also S15–S17 Figs), or (b) a semi-discrete cluster structure observable at different resolutions (Fig 7A, see also S15–S17 Figs), in which each cluster has a distinctive profile of scores across different centrality measures. An example of one such intermediate cluster present in several networks comprises nodes that score highly on closeness (e.g., shortest-path closeness, total communicability, subgraph, information) and eigenvector-like (e.g., eigenvector, Katz) measures of centrality, but low on betweenness-based (shortest-path, random-walk, communicability) measures (e.g. Fig 7 blue cluster; S16 Fig purple cluster). These nodes were thus topologically positioned within a central core of the network (accounting for their high closeness) and were connected to other nodes with high degree (accounting for their high eigenvector values), yet lacked connections to nodes outside of the main cluster (thus having low betweenness and participation coefficient scores). Other intermediate clusters varied depending on the network and may thus define nodes serving unique roles within each specific system.

### Discussion

We evaluated CMCs between 17 different centrality measures in 212 networks to determine how variation in the strength of CMCs across networks tracks differences in global topological properties. We also investigated whether subsets of nodes with consistent topological roles,

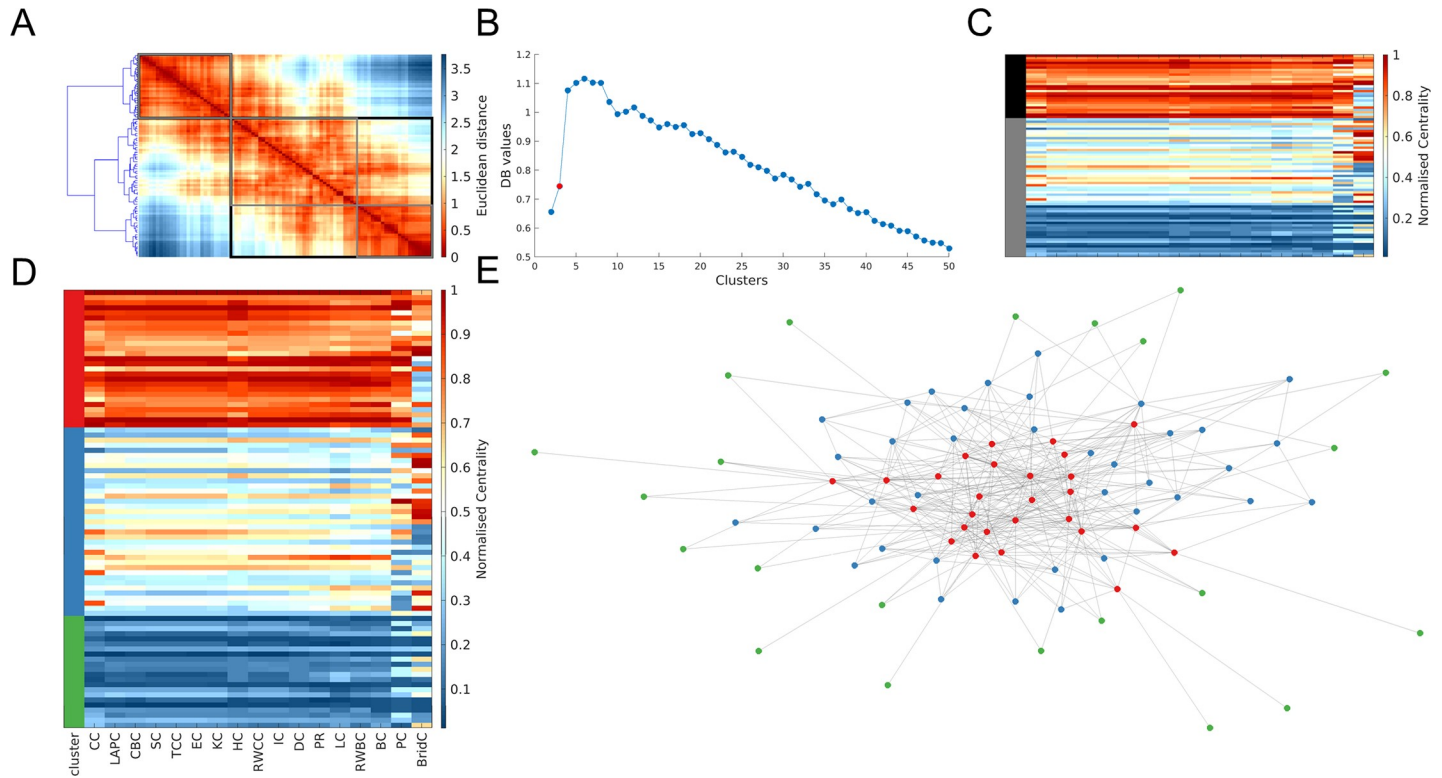


**Fig 7. Multivariate centrality profiling of the network science author collaboration network.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and eight-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.

<https://doi.org/10.1371/journal.pone.0220061.g007>

including network hubs, could be identified based on their multivariate centrality profiles. We found that centrality measures show moderate-to-high positive correlations across most networks; modularity is the strongest predictor of mean CMC variability across unweighted networks; and most networks contain a subset of nodes with consistently high scores across nearly all centrality measures and another subset with consistently low scores.

Consistent with past findings [13,16,17], most CMCs were high, although there was considerable variability across networks (Fig 1 and S3 Fig). This finding is also supported by the PCA results, which showed that the dimensionality of centrality correlations varies from one network to another. CMCs in weighted networks were only slightly weaker than their unweighted forms. Notably, the simplest and most popular measure of centrality, node degree, showed high correlations with most other centrality metrics, likely because a highly connected node is likely to be rated as central by other metrics. Degree may thus act as a useful first approximation of node centrality. Despite generally high CMCs, some measures showed low correlations with other metrics. For instance, Leverage and PageRank centrality were both highly correlated with each other but less so with other measures in both weighted and unweighted networks, possibly because these measures scale a node's importance in relation to the importance of its immediate neighbours, unlike other centrality measures. Bridging centrality and the participation coefficient also demonstrated weaker correlations with other measures, likely because these metrics are conceptually different to standard centrality measures.



**Fig 8. Multivariate centrality profiling of trophic-level species interactions in a New Zealand stream.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and three-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.

<https://doi.org/10.1371/journal.pone.0220061.g008>

We found density, global efficiency, modularity, majorization gap, and spectral gap were correlated with CMCs, which is in line with past findings [20,21,23,27]. Of these, the majorization has been most clearly linked to CMCs by theory [18–20]. However, our regression analysis revealed that the majorization gap was not a significant predictor of the unweighted mean within-network CMCs. The weak association between majorization gap and CMCs was confirmed by the analysis of surrogate data—while we predicted that a lower gap should be associated with higher CMCs, our surrogates were characterized by higher CMCs despite having a comparable or larger gap relative to the observed networks. Recent work has noted that in networks where there are fewer dominance relationships (i.e. the neighbourhood inclusion preorder is less complete), there is more freedom in how different centrality measures can rank nodes. Our findings fit within this interpretation, namely that a larger majorization gap (which is indicative of a less complete neighbourhood inclusion preorder) does not necessarily mean centrality measures must be discordant (where nodes will be ranked differently on different measures), but rather there is more variability in the possible ranks a node can achieve on different centrality measures [19,20]. Our regression analysis also indicated that modularity was the only topological property to make a significant, unique contribution to mean CMC variation across networks. Networks with higher modularity than their matched surrogates also had weaker CMCs (and vice-versa). Modular networks provide greater opportunities to decouple local from global measures of centrality; they can also result in bottlenecks that can

dissociate path-based from degree-based measures (e.g., Fig 1B). The net effect will be a reduction in mean CMCs.

We note that our empirical analysis measured global properties of network topology using methods that may only approximate the actual topology. For example the modularity of a network is highly dependent on the decomposition algorithm used [55], it is not clear how large the spectral gap needs to be for a network to be a good expander [21,23], and the majorization gap is a heuristic for quantifying the distance of a network from a threshold graph, which itself is itself a heuristic to generate a network with perfect neighbourhood-inclusion preorder [20]. Thus, these approximations may partially obscure the relationship between topology and centrality.

Hierarchical clustering of multivariate nodal centrality profiles indicated that two general clusters are present in nearly all networks: a subset of nodes scoring highly on nearly all centrality measures, representing a putative core, and a subset of nodes with low scores on nearly all measures, representing a putative periphery. Beyond these clusters, networks fell into one of two classes, such that they either shows a gradual progression moving from highly central core nodes to peripheral nodes, or a more clustered structure in which subsets of nodes had distinct centrality profiles. These intermediate clusters may define distinct nodes roles that cannot be identified through reliance on a single centrality measure. Networks with this structure tended to have higher modularity or formed a ring with “tendrils” of nodes (i.e. S15 Fig). Together, these results suggest that multivariate centrality profiles may be particularly useful in characterizing nodes roles in networks with modular structure.

An unresolved question concerns the optimal set of centrality measures for such centrality profiling. We focused on a small subset of the >200 metrics that have been proposed, and a wider investigation of this issue is required. We note however, that a limitation of using hierarchical clustering to group nodes is that this approach is unlikely to place individual nodes (or small subsets of nodes) with a distinctive centrality profile within a separate cluster. Indeed, we did find that some networks do contain a small number of nodes with highly discrepant scores across centrality measures (e.g., Fig 7 and S15–S17 Figs). Alternative clustering approaches may be better placed to delineate such nodes, which may play an important role in shaping network dynamics. Nonetheless, our basic approach demonstrates how a comparative approach to centrality analysis, as has been employed in other domains [36], can yield useful insights into the roles of different nodes within a network.

## Supporting information

**S1 Fig. A threshold graph.** A threshold graph is formed by adding in nodes one at a time in one of two ways: a node can either be added in forming no connections (blue nodes) or a node can be added forming connections to all existing nodes (red nodes). The number in each node is the order in which it was added into the network.

(TIF)

**S2 Fig. The neighbourhood-inclusion pre-order and centrality ranks.** Panels A and D shows a network which demonstrates the neighbourhood-inclusion pre-order and panels B and E shows the dominance relation between nodes (a directed edge indicates that the source node is dominated by the target, that is all the neighbours of the source node are a subset of the neighbours of the target node) for the respective network. The network in panel A has no dominance relationship thus has an inconsistent ranking of nodes by four different centrality measures: degree (DC), closeness (CC), eigenvector (EC), and betweenness (BC), as shown in panel C. Conversely the network in panel D has a complete neighbourhood-inclusion pre-order (each node either dominates or is dominated any other node in the network) and thus

all centrality measures will give the same ranking to a node, as shown in panel F. Note that the network in panel D is a threshold graph. A similar figure is presented in Schoch and Brandes [18].

(TIF)

**S3 Fig. Distributions of Centrality Measure Correlations (CMCs) in each network.** The distribution of Spearman correlation coefficient between every pair of centrality measures, CMCs, in each network is represented as a boxplot. Distributions for unweighted networks are split across panels A and B, while distributions for weighted networks are shown in panel C. Networks are coloured by their natural category (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic). Networks have been ordered from highest (left) to lowest (right) median CMC.

(TIF)

**S4 Fig. PCA loadings on the first three principal components for several example networks.** A PCA was run on the (z-scored) centrality scores for each node in each network. The proportion of variance explained by each component is given in the brackets. Nearly all centrality measures loaded onto the first component. In PC2 and PC3 we could observe other types of structure including betweenness measures loading together (panel D-E), bridging centrality/participation coefficient loading together/uniquely on a component (panel A-C), PageRank and leverage loading together (panel D-E), and closeness measures loading with bridging centrality and the participation coefficient (panel F).

(TIF)

**S5 Fig. Mean and standard deviation of between-network CMCs for unweighted and weighted measures.** This figure shows the between-network CMC mean (panel A) and standard deviation (panel B) for unweighted and weighted measures, as calculated on the 40 weighted networks. Both weighted and unweighted measures were generally highly correlated with each other, and unweighted measures were more highly intercorrelated than weighted measures. DC = Degree centrality; EC = Eigenvector centrality; KC = Katz centrality; PR = PageRank centrality; LC = Leverage Centrality; HC = H-index centrality; CC = Shortest-path closeness centrality; SC = Subgraph centrality; PC = Participation coefficient; TCC = Total communicability centrality; RWCC = Random-walk closeness centrality; BC = Shortest-path betweenness centrality; CBC = Communicability betweenness centrality; RWBC = Random-walk betweenness centrality; LAPC = Laplacian centrality; BridC = Bridging centrality. A “w” next to the abbreviated name for the centrality measure indicates it is the weighted version.

(TIF)

**S6 Fig. Scatter plot of the mean within-network CMCs for unweighted and weighted measures.** For the 40 networks with edge weights the mean within-network CMC was calculated separately for unweighted and weighted measures, and then these values were plotted against each other. There is a strong relationship between the two indicating higher correlations between unweighted measures is mirrored by higher correlations in weighted measures.

(TIF)

**S7 Fig. Association between mean within-network CMC and network properties in weighted networks.** The association between the mean within-network CMC (the average of all CMCs within a single network) and each of the global topological properties. Networks are coloured by their natural category (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).

(TIF)



**S8 Fig. Association between each CMC and global topology in unweighted networks.** The lower matrix indicates the value of the Spearman correlation between a CMC and a network property. The upper matrix indicates if this correlation was significant (grey) or not (white) when Bonferroni corrected for 136 combinations of centrality measures. This result shows the strength of individual CMCs was correlated with specific network properties.  
(TIF)

**S9 Fig. Association between each CMC and global topology in weighted networks.** The lower matrix indicates the value of the Spearman correlation between a CMC and a network property. The upper matrix indicates if this correlation was significant (grey) or not (white) when Bonferroni corrected for 136 combinations of centrality measures. This result shows the strength of individual CMCs was correlated with specific network properties.  
(TIF)

**S10 Fig. Association between variance explained by PC1 and network properties in unweighted networks.** The association between the variance explained by PC1 and each of the global topological properties. Networks are coloured by their natural category (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).  
(TIF)

**S11 Fig. Association between variance explained by PC1 and network properties in weighted networks.** The association between the variance explained by PC1 and each of the global topological properties. Networks are coloured by their natural category (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).  
(TIF)

**S12 Fig. Correlations between network properties in unweighted and weighted networks.** Panel A shows the Spearman correlations between each network property in the unweighted networks, while panel B shows the Spearman correlations between each network property in the weighted networks.  
(TIF)

**S13 Fig. Difference between weighted empirical and unconstrained surrogates in mean within-network CMC and network properties.** The y-axis of each plot shows the difference between the empirical networks and unconstrained surrogates mean within-network CMC. The x-axis shows the difference between the empirical networks and unconstrained surrogates on a particular property (except for panel C as the unconstrained surrogates have the same density as the empirical network). On both axis, except for the x-axis in panel C, a negative value indicates the empirical network had a lower value than the mean value of the surrogates, while a positive value indicates the empirical networks had a larger value. Points are colored by the natural category of the empirical network (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).  
(TIF)

**S14 Fig. Difference between weighted empirical and constrained surrogates in mean within-network CMC and network properties.** The y-axis of each plot shows the difference between the empirical networks and constrained surrogates mean within-network CMC. The x-axis shows the difference between the empirical networks and constrained surrogates on a particular property (except for panels C and F as the constrained surrogates have the same density and majorization gap as the empirical network). On both axis, except for the x-axis in

panels C and F, a negative value indicates the empirical network had a lower value than the mean value of the surrogates, while a positive value indicates the empirical networks had a larger value. Points are colored by the natural category of the empirical network (blue = social, grey = technological, brown = biological, orange = informational, purple = transportation; green = economic).

(TIF)

**S15 Fig. Multivariate centrality profiling of the Berlin subway network.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and six-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.

(TIF)

**S16 Fig. Multivariate centrality profiling of the network of drug user acquaintanceships in Hartford, UK.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and nine-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.

(TIF)

**S17 Fig. Multivariate centrality profiling of the Balerna irrigation water distribution network.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and seven-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.

(TIF)

**S18 Fig. Multivariate centrality profiling of the network of noun phrases (places and names) in the King James Version of the Bible.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and three-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.  
(TIF)

**S19 Fig. Multivariate centrality profiling of the network of adjacent adjective and nouns in David Copperfield.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and three-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.  
(TIF)

**S20 Fig. Multivariate centrality profiling of the network of food ingredients and flavours.** Panel A shows the dendrogram projected alongside the distance matrix of node pairs (ranks scores were normalised to be in the range 0–1 with 1 indicating the highest rank). The black and grey boxes and indicate the clusters when a two-cluster and three-cluster solution is used, respectively. Panel B displays the results for the Davies-Bouldin (DB) criterion. A lower DB value represents a better clustering solution. The solution shown in panels D and E is labelled in red. Only the first 50 clustering solutions are shown for ease of visibility. Panel C shows the matrix of nodal centrality scores (each row is a node and each column is a measure) and how these are clustered in a two-cluster solution (the black and grey represent the two different clusters). Panel D shows the matrix of nodal centrality scores as well as the clusters each node was assigned to. Panel E shows a topological representation of the network, produced using the force-directed layout algorithm, where each node is coloured according to the cluster it was allocated to in panel D.  
(TIF)

**S1 Text. Supplementary methods and results.**  
(DOCX)

## Author Contributions

**Conceptualization:** Stuart Oldham, Ben Fulcher, Alex Fornito.

**Data curation:** Chao Suo.

**Formal analysis:** Stuart Oldham, Linden Parkes, Aurina Arnatkevičiūtė.

**Investigation:** Stuart Oldham.

**Methodology:** Stuart Oldham, Ben Fulcher, Alex Fornito.

**Resources:** Chao Suo.

**Supervision:** Ben Fulcher, Alex Fornito.

**Writing – original draft:** Stuart Oldham, Ben Fulcher, Alex Fornito.

**Writing – review & editing:** Stuart Oldham, Ben Fulcher, Linden Parkes, Aurina Arnatkevičiūtė, Chao Suo, Alex Fornito.

## References

1. Borgatti SP. Centrality and network flow. *Soc Networks*. 2005; 27: 55–71. <https://doi.org/10.1016/j.socnet.2004.11.008>
2. Borgatti SP, Everett MG. A Graph-theoretic perspective on centrality. *Soc Networks*. 2006; 28: 466–484. <https://doi.org/10.1016/j.socnet.2005.11.005>
3. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks*. 1978; 1: 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
4. Fornito A, Zalesky A, Bullmore E. *Fundamentals of Brain Network Analysis*. London: Academic Press; 2016.
5. Bell DC, Atkinson JS, Carlson JW. Centrality measures for disease transmission networks. *Soc Networks*. 1999; 21: 1–21. [https://doi.org/10.1016/S0378-8733\(98\)00010-0](https://doi.org/10.1016/S0378-8733(98)00010-0)
6. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001; 411: 41–42. <https://doi.org/10.1038/35075138> PMID: 11333967
7. Yan E, Ding Y. Applying centrality measures to impact analysis: A coauthorship network analysis. *J Am Soc Inf Sci Technol*. 2009; 60: 2107–2118. <https://doi.org/10.1002/asi.21128>
8. Achard S, Delon-Martin C, Vertes PE, Renard F, Schenck M, Schneider F, et al. Hubs of brain functional networks are radically reorganized in comatose patients. *Proc Natl Acad Sci*. 2012; 109: 20608–20613. <https://doi.org/10.1073/pnas.1208933109> PMID: 23185007
9. Gili T, Saxena N, Diukova A, Murphy K, Hall JE, Wise RG. The thalamus and brainstem act as key hubs in alterations of human brain network connectivity induced by mild propofol sedation. *J Neurosci*. 2013; 33: 4024–31. <https://doi.org/10.1523/JNEUROSCI.3480-12.2013> PMID: 23447611
10. Avena-Koenigsberger A, Misić B, Sporns O. Communication dynamics in complex brain networks. *Nat Rev Neurosci*. Nature Publishing Group; 2017; 19: 17–33. <https://doi.org/10.1038/nrn.2017.149> PMID: 29238085
11. Gofñi J, Avena-Koenigsberger A, Velez de Mendizabal N, van den Heuvel MP, Betzel RF, Sporns O. Exploring the Morphospace of Communication Efficiency in Complex Networks. *PLoS One*. 2013; 8. <https://doi.org/10.1371/journal.pone.0058070> PMID: 23505455
12. Jalili M, Salehzadeh-Yazdi A, Asgari Y, Arab SS, Yaghmaie M, Ghavamzadeh A, et al. CentiServer: A comprehensive resource, web-based application and R package for centrality analysis. *PLoS One*. 2015; 10: 1–8. <https://doi.org/10.1371/journal.pone.0143111> PMID: 26571275
13. Estrada E. Characterization of topological keystone species. Local, global and “meso-scale” centralities in food webs. *Ecol Complex*. 2007; 4: 48–57. <https://doi.org/10.1016/j.ecocom.2007.02.018>
14. Kim PJ, Jeong H. Reliability of rank order in sampled networks. *Eur Phys J B*. 2007; 55: 109–114. <https://doi.org/10.1140/epjb/e2007-00033-7>
15. del Rio G, Koschützki D, Coello G. How to identify essential genes from molecular networks? *BMC Syst Biol*. 2009; 3: 102. <https://doi.org/10.1186/1752-0509-3-102> PMID: 19822021
16. Li C, Li Q, Van Mieghem P, Stanley HE, Wang H. Correlation between centrality metrics and their application to the opinion model. *Eur Phys J B*. 2015; 88: 1–13. <https://doi.org/10.1140/epjb/e2015-50671-y>
17. Ronqui J, Travieso G. Analyzing complex networks through correlations in centrality measurements. *J Stat Mech Theory Exp*. 2015; 9. <https://doi.org/10.1088/1742-5468/2015/05/P05030>
18. Schoch D, Brandes U. Re-conceptualizing centrality in social networks. *Eur J Appl Math*. 2016; 19: 1–15. <https://doi.org/10.1017/S0956792516000401>

19. Schoch D. Centrality without indices: Partial rankings and rank probabilities in networks. *Soc Networks*. Elsevier B.V.; 2018; 54: 50–60. <https://doi.org/10.1016/j.socnet.2017.12.003>
20. Schoch D, Valente TW, Brandes U. Correlations among centrality indices and a class of uniquely ranked graphs. *Soc Networks*. Elsevier B.V.; 2017; 50: 46–54. <https://doi.org/10.1016/j.socnet.2017.03.010>
21. Benzi M, Klymko C. Total communicability as a centrality measure. *J Complex Networks*. 2013; 1: 124–149. <https://doi.org/10.1093/comnet/cnt007>
22. Estrada E. Network robustness to targeted attacks. the interplay of expansibility and degree distribution. *Eur Phys J B*. 2006; 52: 563–574. <https://doi.org/10.1140/epjb/e2006-00330-7>
23. Estrada E. Spectral scaling and good expansion properties in complex networks. *Europhys Lett*. 2006; 73: 649–655. <https://doi.org/10.1209/epl/i2005-10441-3>
24. Guimerà R, Amaral LAN. Functional cartography of complex metabolic networks. *Nature*. 2005; 433: 895–900. <https://doi.org/10.1038/nature03288> PMID: 15729348
25. Frantz TL, Cataldo M, Carley KM. Robustness of centrality measures under uncertainty: Examining the role of network topology. *Comput Math Organ Theory*. 2009; 15: 303–328. <https://doi.org/10.1007/s10588-009-9063-5>
26. Bloch F, Jackson MO, Tebaldi P. Centrality Measures in Networks. 2016.
27. Valente TW, Coronges K, Lakon C, Costenbader E. How Correlated Are Network Centrality Measures? *Connections*. 2008; 28: 16–26. <https://doi.org/10.1016/j.bbi.2008.05.010> PMID: 20505784
28. Iyer S, Killingback T, Sundaram B, Wang Z. Attack Robustness and Centrality of Complex Networks. *PLoS One*. 2013; 8. <https://doi.org/10.1371/journal.pone.0059613> PMID: 23565156
29. Batool K, Niazi MA. Towards a methodology for validation of centrality measures in complex networks. *PLoS One*. 2014; 9. <https://doi.org/10.1371/journal.pone.0090283> PMID: 24709999
30. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identifying influential spreaders in complex networks. *Nat Phys*. Nature Publishing Group; 2010; 6: 36. <https://doi.org/10.1038/nphys1746>
31. Lozares C, López-Roldán P, Bolibar M, Muntanyola D. The structure of global centrality measures. *Int J Soc Res Methodol*. 2015; 18: 209–226. <https://doi.org/10.1080/13645579.2014.888238>
32. Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, Hennig H, Wolkenhauer O, Mirzaie M, et al. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Syst Biol*. *BMC Systems Biology*; 2018; 12: 1–17. <https://doi.org/10.1186/s12918-017-0484-3>
33. Koschützki D, Schreiber F. Comparison of Centralities for Biological Networks. *Proc Ger Conf Bioinforma*. 2004; 199–206.
34. Andreotti J, Jann K, Melie-Garcia L, Giezendanner S, Abela E, Wiest R, et al. Validation of network communicability metrics for the analysis of brain structural networks. *PLoS One*. 2014; 9: 1–26. <https://doi.org/10.1371/journal.pone.0115503> PMID: 25549088
35. Wang X, Lin Q, Xia M. Differentially categorized structural brain hubs are involved in different micro-structural, functional, and cognitive characteristics and contribute to individual identification. *Hum Brain Mapp*. 2018; 1–17. <https://doi.org/10.1002/hbm.23941> PMID: 29314415
36. Fulcher BD, Little MA, Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. *J R Soc Interface*. 2013; 10: 20130048. <https://doi.org/10.1098/rsif.2013.0048> PMID: 23554344
37. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*. Elsevier Inc.; 2010; 52: 1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003> PMID: 19819337
38. Fortunato S. Community detection in graphs. *Phys Rep*. Elsevier B.V.; 2010; 486: 75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>
39. Power JD, Schlaggar BL, Lessov-Schlaggar CN, Petersen SE. Evidence for hubs in human functional brain networks. *Neuron*. Elsevier Inc.; 2013; 79: 798–813. <https://doi.org/10.1016/j.neuron.2013.07.035> PMID: 23972601
40. Bonacich Phillip. Factoring and weighting approaches to status scores and clique identification. *J Math Sociol*. 1972; 2: 113–120.
41. Katz L. A new status index derived from sociometric analysis. *Psychometrika*. 1953; 18: 39–43.
42. Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. *World Wide Web Internet Web Inf Syst*. 1998; 54: 1–17. <https://doi.org/10.1.1.31.1768>
43. Joyce KE, Laurienti PJ, Burdette JH, Hayasaka S. A new measure of centrality for brain networks. *PLoS One*. 2010;5. <https://doi.org/10.1371/journal.pone.0012200> PMID: 20808943

44. Lü L, Zhou T, Zhang Q-M, Stanley HE. The H-index of a network node and its relation to degree and coreness. *Nat Commun.* 2016; 7: 10168. <https://doi.org/10.1038/ncomms10168> PMID: 26754161
45. Qi X, Duval RD, Christensen K, Fuller E, Spahiu A, Wu Q, et al. Terrorist Networks, Network Energy and Node Removal: A New Measure of Centrality Based on Laplacian Energy. *Soc Netw.* 2013; 02: 19–31. <https://doi.org/10.4236/sn.2013.21003>
46. Qi X, Fuller E, Wu Q, Wu Y, Zhang CQ. Laplacian centrality: A new centrality measure for weighted networks. *Inf Sci (Ny)*. Elsevier Inc.; 2012; 194: 240–253. <https://doi.org/10.1016/j.ins.2011.12.027>
47. Sabidussi G. The centrality index of a graph. *Psychometrika.* 1966; 31: 581–603. <https://doi.org/10.1007/BF02289527> PMID: 5232444
48. Estrada E, Rodriguez-Velazquez J a. Subgraph Centrality in Complex Networks. *Phys Rev E.* 2005; 71: 29. <https://doi.org/10.1103/PhysRevE.71.056103> PMID: 16089598
49. Noh JD, Rieger H. Random Walks on Complex Networks. *Phys Rev Lett.* 2004; 92: 1–4. <https://doi.org/10.1103/PhysRevLett.92.118701> PMID: 15089179
50. Blöchl F, Theis FJ, Vega-Redondo F, Fisher EON. Vertex centralities in input-output networks reveal the structure of modern economies. *Phys Rev E—Stat Nonlinear, Soft Matter Phys.* 2011; 83: 1–9. <https://doi.org/10.1103/PhysRevE.83.046127> PMID: 21599260
51. Stephenson K, Zelen M. Rethinking centrality: Methods and examples. *Soc Networks.* 1989; 11: 1–37. [https://doi.org/10.1016/0378-8733\(89\)90016-6](https://doi.org/10.1016/0378-8733(89)90016-6)
52. Estrada E, Higham DJ, Hatano N. Communicability betweenness in complex networks. *Phys A Stat Mech its Appl.* 2009; 388: 764–774. <https://doi.org/10.1016/j.physa.2008.11.011>
53. Newman MEJ. A measure of betweenness centrality based on random walks. *Soc Networks.* 2005; 27: 39–54. <https://doi.org/10.1016/j.socnet.2004.11.009>
54. Hwang W, Cho Y, Zhang A, Remanathan M. Bridging Centrality: Identifying Bridging Nodes In Scale-free Networks. *Proc 14th ACM SIGKDD Int Conf Knowl Discov data Min.* 2008; 336–344. <https://doi.org/10.1145/1401890.1401934>
55. Ghasemian A, Hosseinmardi H, Clauset A. Evaluating Overfit and Underfit in Models of Network Community Structure. *arXiv.* 2018; 1–17. arXiv:1802.10582v2
56. Clauset A, Tucker E, Sainz M. The Colorado Index of Complex Networks [Internet]. 2016 [cited 5 Aug 2018]. Available: <https://icon.colorado.edu/>
57. Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, et al. The Human Connectome Project: A data acquisition perspective. *Neuroimage.* 2012; 62: 2222–2231. <https://doi.org/10.1016/j.neuroimage.2012.02.018> PMID: 22366334
58. Erdős P, Rényi A. The Evolution of Random Graphs. *Publ Math Debrecen.* 1959; 6: 290–297. <https://doi.org/10.2307/1999405>
59. Maslov S, Sneppen K. Specificity and Stability in Topology of Protein Networks. *Science.* 2002; 296: 910–913. <https://doi.org/10.1126/science.1065103> PMID: 11988575
60. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans R Soc A Math Phys Eng Sci.* 2016; 374. <https://doi.org/10.1098/rsta.2015.0202> PMID: 26953178
61. Tabachnick BG, Fidell LS. Using multivariate statistics. 6th ed. Harlow, Essex: Pearson Education Limited; 2014.
62. Newman MEJ. *Networks: An Introduction.* New York: Oxford University Press; 2010.
63. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature.* 1998; 393: 440–442. <https://doi.org/10.1038/30918> PMID: 9623998
64. Latora V, Marchiori M. Efficient behavior of small-world networks. *Phys Rev Lett.* 2001; 87: 198701. <https://doi.org/10.1103/PhysRevLett.87.198701> PMID: 11690461
65. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E.* 2004; 69: 026113. <https://doi.org/10.1103/PhysRevE.69.026113> PMID: 14995526
66. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008; 2008: P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
67. Lancichinetti A, Fortunato S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys Rev E—Stat Nonlinear, Soft Matter Phys.* 2009; 80: 1–8. <https://doi.org/10.1103/PhysRevE.80.016118> PMID: 19658785
68. Lancichinetti A, Fortunato S. Consensus clustering in complex networks. *Sci Rep.* 2012; 2. <https://doi.org/10.1038/srep00336> PMID: 22468223
69. Good BH, De Montjoye YA, Clauset A. Performance of modularity maximization in practical contexts. *Phys Rev E—Stat Nonlinear, Soft Matter Phys.* 2010; 81: 1–19. <https://doi.org/10.1103/PhysRevE.81.046106> PMID: 20481785

70. Ward JH. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. 1963. pp. 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
71. Davies DL, Bouldin DW. A Cluster Separation Measure. *IEEE Trans Pattern Anal Mach Intell*. 1979; PAMI-1: 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
72. Hancer E, Karaboga D. A comprehensive survey of traditional, merge-split and evolutionary approaches proposed for determination of cluster number. *Swarm Evol Comput*. Elsevier; 2017; 32: 49–67. <https://doi.org/10.1016/j.swevo.2016.06.004>
73. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp*. 1991; 21: 1129–1164. <https://doi.org/10.1002/spe.4380211102>
74. Brandes U, Fleischer D. Centrality measures based on current flow. *Lect Notes Comput Sci*. 2005; 533–544. [https://doi.org/10.1007/978-3-540-31856-9\\_44](https://doi.org/10.1007/978-3-540-31856-9_44)