


SCIENTIFIC REPORTS



OPEN

On neighbourhood degree sequences of complex networks

Keith M. Smith 

Network topology is a fundamental aspect of network science that allows us to gather insights into the complicated relational architectures of the world we inhabit. We provide a first specific study of neighbourhood degree sequences in complex networks. We consider how to explicitly characterise important physical concepts such as similarity, heterogeneity and organization in these sequences, as well as updating the notion of hierarchical complexity to reflect previously unnoticed organizational principles. We also point out that neighbourhood degree sequences are related to a powerful subtree kernel for unlabeled graph classification. We study these newly defined sequence properties in a comprehensive array of graph models and over 200 real-world networks. We find that these indices are neither highly correlated with each other nor with classical network indices. Importantly, the sequences of a wide variety of real world networks are found to have greater similarity and organisation than is expected for networks of their given degree distributions. Notably, while biological, social and technological networks all showed consistently large neighbourhood similarity and organisation, hierarchical complexity was not a consistent feature of real world networks. Neighbourhood degree sequences are an interesting tool for describing unique and important characteristics of complex networks.

Contemplating the roles of components in natural and man-made systems, we begin to realise their diversity. Take for example, the structure of an organisation. At face value, employees are assigned titles and pay-scales which place the workforce in a convenient hierarchy with each level comprising of equivalencies based on the competitive value of the work done. However, in large and multifaceted organisations the work done is often highly variable and it is beneficial to have employees with a diverse range of skills and talents interacting in different ways. Network science provides a natural framework to understand relationship patterns of such complex systems and we shall here formulate and study hierarchical equivalency in terms of neighbourhood degree sequences of complex networks. Figure 1A provides an illustration of how neighbourhood degree sequences intuitively help to understand global hierarchical patterns.

The distribution of connections among nodes in complex networks, known as the degree distribution, is a key consideration of its topology. Predated by the study of degree sequences¹, interest in degree distributions arose from the study of real-world networks, where it was noted that they approximated various statistical distributions with heavy tails², being particularly driven by the prevalence of strong hubs in real-world networks which are not present, for example, in random graphs³, random geometric graphs⁴ and small-world models⁵. Pertinent random null models, called configuration models, have since been developed in which the degree distribution is fixed, allowing unbiased random controls for studying network topologies^{6,7}.

Although often explicitly mentioned with regard to real-world networks, what is meant by concepts such as organisation and complexity has largely been left to intuition. In seeking to understand the complexity of real world networks, Smith & Escudero⁸ recently proposed to look at neighbourhood degree sequences. For a given node, its neighbourhood degree sequence was defined as the ordered degrees of nodes in its neighbourhood. This was based on observations that ordered networks such as regular networks, quasi-star networks, grid networks and highly patterned networks shared the common feature of highly homogeneous neighbourhood degree sequences for nodes of the same degree. Conceptualising the degree distribution as a hierarchy of nodes, they proposed an index called hierarchical complexity to characterise the heterogeneity of hierarchically equivalent (i.e. same degree) nodes. Note, the term 'hierarchy' in networks is also associated with the scaling of community structure^{9,10}. Here, it is used— in the more lexically familiar sense— with respect to levels of importance, where nodes of higher degree are often considered of higher importance in the network topology¹¹. Hierarchical

Usher Institute of Population Health Science and Informatics, University of Edinburgh, 9 BioQuarter, Little France, Edinburgh, EH16 4UX, UK. Correspondence and requests for materials should be addressed to K.M.S. (email: k.smith@ed.ac.uk)

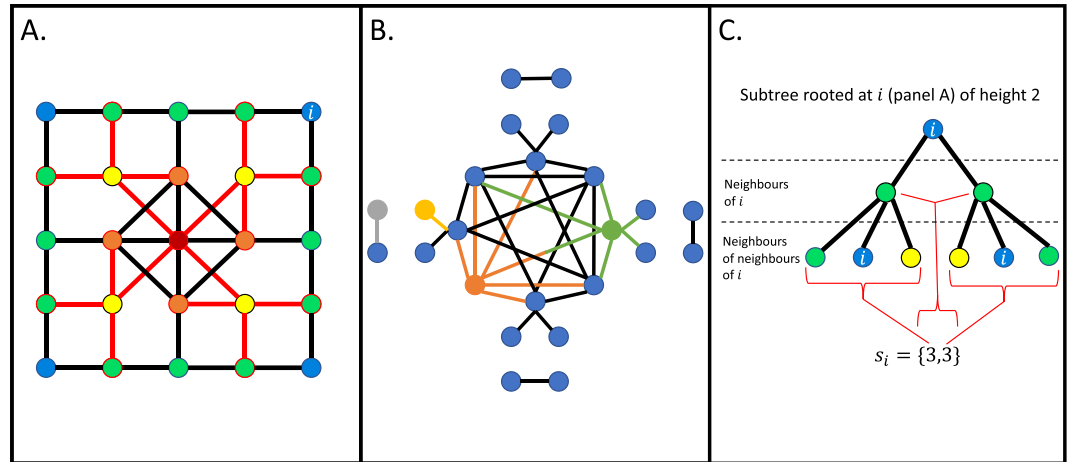


Figure 1. (A) How can we efficiently capture the organisation of this graph mathematically without reference to node placements on the plane? We can note that the neighbourhoods of nodes of a given degree are equivalent with respect to the degrees of nodes they connect to— e.g. all yellow nodes (degree 4) connect to the same number of green (degree 3), orange (degree 5) and red (degree 8) nodes. Thus neighbourhood degree sequences appear as a promising avenue. (B) Illustration of a multi-ordered degree graph whose equal-degree nodes are organised into two distinct classes with different degree sequences. (C) Illustration of a subtree of height 2 for node i in panel A. The number of nodes at height 1 is the degree of i , while for a node at height 1, its degree is the number of nodes at height 2 extending from it, all captured by i 's neighbourhood degree sequence, s_i .

complexity was developed in the context of electroencephalogram functional connectivity, which, in contrast to ordered and random systems, was found to have inordinately high levels of heterogeneity amongst its neighbourhood degree sequences⁸. This concept has since been utilised to help understand how best to binarise EEG functional connectivity for topological analysis¹² and has been validated in structural MRI networks¹³. However, the prevalence of such topology amongst complex networks in general is unknown. In pure mathematics, Barrus & Donovan independently initiated study of neighbourhood degree lists as a topological invariant more refined than both the degree sequence and joint degree graph matrix¹⁴, while Nishimura & Subramanya proposed to study neighbourhood degree lists for the combinatorial problem of changing a graph into one with given neighbourhood degrees¹⁵.

That is as far as has been done with neighbourhood degree sequences to date. Yet, the intriguing insights provided by hierarchical complexity in brain networks makes a broader study of neighbourhood degree sequences across a broader range of domains worthwhile. This work comes after work done involving neighbouring degrees and centralities such as the eigenvector centrality, a centrality index which is larger depending on the centralities of the nodes a node is connected to¹⁶; assortativity, an index of degree-degree correlation between connected nodes¹⁷; and network entropy, a measure of edgewise node degree eccentricity¹⁸. Neighbourhood degree sequences, however, are a completely separate consideration of networks. Most notably, rather than comparing nodes which are connected to each other, we compare nodes which have the same degree, irrespective of whether they are connected or not, regarding such nodes as hierarchically equivalent within the network topology.

In this study a number of ways to analyse neighbourhood degree sequences are proposed. Notably, indices of node heterogeneity and neighbourhood similarity are introduced. We also consider a new notion of multi-orderedness in a network. This is based on the observation that nodes of a given degree in an ordered network may have several distinct neighbourhood degree sequences. This gives rise to another index defined as neighbourhood organisation which measures the extent to which such multi-orderedness is present in the network. We then show that the existence of multi-ordered degrees can artificially raise the network's hierarchical complexity. Thus, we utilise the formulation of neighbourhood organisation to provided a version of hierarchical complexity which corrects for multi-ordered degrees. We also described how neighbourhood degree sequences have clear links with powerful and efficient subtree kernels for graph classification. The proposed indices are then applied to a range of network models and compared with existing classical network indices, the aim of which is to ascertain to what extent these indices explain unique topological properties in complex networks. They are also applied to 215 real world networks from various disciplines of study in order to assess the characteristics of neighbourhood degree sequences in the world around us and the insights these new indices offer.

Neighbourhood Degree Sequences

For k_i the degree of node i , the neighbourhood degree sequence, s_i , of node i is

$$s_i = \{k_1^i, k_2^i, \dots, k_{k_i}^i\}, \quad (1)$$

where the k_j^i 's are the degrees of the nodes to which i is connected and such that $k_1^i \leq k_2^i \leq \dots \leq k_{k_i}^i$. For example, the graph in Fig. 1A has four degree 4 nodes (yellow) all with neighbourhood degree sequence $\{3, 3, 5, 8\}$ and four

degree 5 nodes (orange) all with neighbourhood degree sequence $\{3, 4, 5, 5, 8\}$. In the following we shall consider a number of ways to study these sequences.

Node heterogeneity. One way to characterise neighbourhood degree sequences would be to employ the same methods to characterise degree distributions and then average over all nodes. As a pertinent example of this, a common index of graph heterogeneity is the degree variance $v = \text{var}(k)$ ¹⁹. We can then define node heterogeneity, V_n , as the average variance of neighbourhood degree sequences of a graph for all nodes of degree greater than 1:

$$V_n(G) = \frac{1}{n} \sum_{i \text{ s.t. } k_i > 1} \text{var}(s_i). \quad (2)$$

Of course, it is then interesting to understand how average node heterogeneity compares to graph heterogeneity, i.e. comparing local and global heterogeneities of a graph. To do this we can simply divide (3) by v , giving

$$\hat{V}_n(G) = \frac{1}{n \text{var}(k)} \sum_{i \text{ s.t. } k_i > 1} \text{var}(s_i). \quad (3)$$

High values of this measure tell us that nodes tend to be connected to nodes of homogeneous degrees, given the degree distribution, and low values tell us the opposite. Specifically, if this value is below 1, the degree variance within the neighbourhoods is on average less than the global degree variance, indicating that the nodes have more homogeneous neighbourhood degrees. It is worth highlighting the distinction between this and assortativity, which seeks to measure the similarity of degrees of connected nodes. Node heterogeneity is a measure of the similarity of the degrees of all neighbouring nodes, irrespective of the degree of the node itself.

Note that v is clearly minimal for regular graphs and is known to be maximal for quasi-star and quasi-complete graphs for any given number of nodes and edges²⁰. On the other hand V_n is zero for regular graphs but is also small for quasi-star and quasi-complete graphs. For instance, the star graph consists of one node connected to all other nodes and no other edges. Thus it has one $n - 1$ degree node with degree sequence $\{1, 1, \dots, 1\}$ and $n - 1$ 1 degree nodes with degree sequence $\{n - 1\}$. Clearly, these all have zero variance, giving $V_n = 0$ for the star graph. This is interesting because, while some believe star graphs should have maximum heterogeneity²¹, V_n points at a possible different view. The degree distribution of a star graph is just 1 node away from being completely regular—take the dominant node out and you have an empty graph (redundantly regular). Heterogeneity could perhaps be alternatively formulated in the sense that removing or adding nodes does not relegate the graph to being regular.

Neighbourhood similarity. The other way of characterising neighbourhood degree sequences we shall consider is to compare all neighbourhood degree sequences of equal length. Indeed, this is the perspective employed to formulate hierarchical complexity, looking at the element-wise variance of equal-length neighbourhood degree sequences. Another, fairly more simple characteristic can be posed by considering the number of nodes in the network whose neighbourhood degree sequence matches that of another node in the graph. We call this neighbourhood similarity (reflecting the concept of geometric similarity) and, using the Kronecker delta function $\delta(x, y)$ which is 1 if $x = y$ and 0 otherwise, write

$$S(G) = \frac{\sum_{i=1}^n \left(1 - \delta \left(\sum_{j=1}^n \delta(s_i, s_j), 0 \right) \right)}{n}. \quad (4)$$

Notice, this uses the δ function twice. The first time is to find the number of matching neighbourhood degree sequences for node i . The second delta is used to determine if there are *any* matching sequences, i.e. seeing if the sum of the first δ s is different from 0. Since this is a negation (δ returns 0 if there are any matches), we then have to subtract the answer from 1 to provide the answer to whether any match exists for node i . Summing over all i and dividing by n provides the proportion of nodes which have at least one matching neighbourhood degree sequence. It is clear that $0 \leq S \leq 1$ for all graphs, since it concerns a fraction of the network nodes. It certainly attains 1 for regular graphs. However, we prove the following result with respect to graph symmetry on the plane, establishing the link between neighbourhood similarity and graph symmetry.

Proposition 1: *Let G be a graph which can be arranged on the plane such that G has mirror or rotational symmetry whose axis does not pivot on any node. Then $S(G) = 1$.*

Proof: Let s_i be a neighbourhood degree sequence for general node i . Then the node, j , in the position symmetric to i with respect to the axis of symmetry has neighbourhood degree sequence s_j and has the same degree as i . Further, each node in the neighbourhood of i , p_i , also has a node in position symmetric to p_i with respect to the axis of symmetry, p_j , and these nodes are connected to j and such that $k_{p_i} = k_{p_j}$, by symmetry. Thus $s_i = s_j$ and since s_i was arbitrary and no nodes lie on the axis of symmetry itself, $S(G) = 1$, as required.

Thus, neighbourhood similarity of a graph is indeed related to the planar symmetry of a graph. That being said, the opposite is not true— not all values $S(G) = 1$ are attained by planar symmetric graphs, as can be quickly seen by regarding non-symmetric regular graphs such as the Frucht graph²².

Hierarchical complexity: oversights of multi-ordered degree graphs. Hierarchical complexity is an index developed with the aim to be low for all highly ordered graphs and graphs with simple generative

mechanisms. Simple in the sense that one needs only a few rules to compute the graph such as in random graphs (edges exist with uniformly random probabilities) or random geometric graphs (nodes are randomly sampled on a n -D Euclidean space and then connected based on distances in the space). In this sense, one can describe precisely how one can expect the graph and subsamples of the graph to behave. On the other hand, attempts to model real world networks indicates that a larger and more a complicated set of rules would be required to generate complex network-like topologies where subsamples of the graph (such as node neighbourhoods) would be less likely to show similar behaviours¹³. The hypothesis is that nodes of a given degree in highly ordered graphs play equivalent roles in the topology, which implies that they have the same or similar neighbourhood degree sequences. However, what fails to be taken account of in its formulation is the possibility to have a high degree of order in which nodes of a given degree can be split into different groups of identical sequences. For example, Fig. 1B shows a graph with degree 1 and 6 nodes. The six-degree nodes fall into one of two sequences $\{1, 1, 6, 6, 6, 6\}$ and $\{6, 6, 6, 6, 6, 6\}$, as illustrated by the green and orange nodes, respectively. One-degree nodes are connected to either one- or six-degree nodes, as illustrated by the grey and yellow nodes, respectively. We call such a graph here a *multi-ordered degree graph*.

Definition 1: Let q_p be the number of all p -length neighbourhood degree sequences and $\sigma_p = \{s_i\}_{k=1}^p$ be the set of (unique) p -length neighbourhood degree sequences. Then p is a multi-ordered degree of the graph if $1 < |\sigma_p| \ll q_p$. A graph for which $1 < |\sigma_p| \ll q_p$ or, otherwise, $|\sigma_p| = 1$ for all p is called a multi-ordered degree graph.

Neighbourhood organisation. We can pose a measure for this sense of multi-ordered degrees using neighbourhood degree sequences. We could simply divide the number of unique p -length sequences by the total number of p -length sequences, giving

$$\frac{|\sigma_p|}{q_p}, \quad (5)$$

however this is the same no matter how many unique degree sequences occur more than once. Consider the following. Let c_{pj} denote the number of neighbourhood degree sequences of length p in G that have equivalency to $s_j \in \sigma_p$. Then, for example, take $q_p = 5$ and $|\sigma_p| = 3$. We could have $c_{p1} = 1, c_{p2} = 1$ and $c_{p3} = 3$ or $c_{p1} = 1, c_{p2} = 2$ and $c_{p3} = 2$. Both of these options would have the same value of (5), yet the latter has better qualities of being multiply ordered than the former since there are two distinct sequences which occur more than once, rather than just the one in the former case. We can offset (5) by considering the differences between the number of p -length sequences, q_p , and the number of occurrences of each (unique) neighbourhood degree sequence in σ_p . Then $\sum_{j=1}^{|\sigma_p|} c_{pj} = 1$ and we consider the entity

$$\sum_{j=1}^{|\sigma_p|} (q_p - c_{pj}). \quad (6)$$

This is maximal, $q_p(q_p - 1)$, when all p -length neighbourhood degree sequences are unique and zero (i.e. minimal) when all p -length neighbourhood degree sequences are equal. We can thus normalise this term as

$$\frac{\sum_{j=1}^{|\sigma_p|} (q_p - c_{pj})}{q_p(q_p - 1)}. \quad (7)$$

Just taking (6) would also not reflect the multi-order requirement. It is really the combination of (5) and (6) that is required to realise a measure of multi-ordered degrees—elements of σ_p should occur frequently and at the same time the number of unique sequences should be as large as possible. Combining (5) and (6), then, we get

$$\omega_p = \frac{|\sigma_p| \sum_{j=1}^{|\sigma_p|} (q_p - c_{pj})}{q_p^2 (q_p - 1)} \quad (8)$$

Taking the mean of this over all degrees and subtracting from 1, we have the neighbourhood organisation coefficient

$$\Omega(G) = 1 - \frac{1}{|\mathcal{D}_2|} \sum_{p=1}^{n-1} \omega_p, \quad (9)$$

where \mathcal{D}_2 is the set of degrees of the graph taken by at least 2 nodes.

Updated hierarchical complexity. Given the above consideration of multi-ordered degrees and the neighbourhood organisation index, we can formulate an update to hierarchical complexity that takes into account multi-ordered degrees. In the terminology of this paper, hierarchical complexity can be written

$$R(G) = \frac{1}{|\mathcal{D}_2|} \sum_{p \in \mathcal{D}_2} \frac{1}{p(q_p - 1)} \sum_{j=1}^p \sum_{i=1}^p (s_i^p(j) - \mu^p(j))^2 \quad (10)$$

where \mathcal{V}_p is the set of nodes of degree p and $\mu^p(j)$ is the mean of the j th entries of all p length neighbourhood degree sequences.

To correct for multi-ordered degrees in this index, we can implement the term ω_p inside the first summand in to give

$$R_{\Omega}(G) = \frac{1}{|\mathcal{D}_2|} \sum_{p \in \mathcal{D}_2} \frac{\omega_p}{p(q_p - 1)} \sum_{j=1}^p \sum_{i \in \mathcal{V}_p} (s_i^p(j) - \mu^p(j))^2. \quad (11)$$

When ω_p is small, multi-orderedness is present in the p degree nodes and thus the value of hierarchical complexity for these degrees is suppressed and vice versa. Computing this for the example in Fig. 1A we obtain $R_{\Omega} = 0.0029$ — a 65 fold decrease from R and a more reasonable expected value of neighbourhood degree sequence diversity.

Link to the graph isomorphism problem. The Weisfeiler-Lehman graph isomorphism test²³ is a powerful method for distinguishing labelled graph topologies which holds for almost all graphs²⁴. Based on this test, subtree kernels have been produced for assessing graph similarity in machine learning approaches which are highly efficient compared to other successful kernels²⁵. Indeed, these subtree kernels have been shown to outperform the competition when implemented into a graph neural network approach while mapping similar graph topologies to similar embeddings in a low-dimensional space²⁶.

The subtree of node i of height h constructs a tree rooted at i which extends out to i 's neighbours and then out again to i 's neighbours' neighbours and so on for h steps, see Fig. 1C. The kernel is a reduction of these subtrees to identifying labels which are then compared between two graphs to check their similarity. Subtrees of height $h = 2$ or 3 have been shown to achieve best performance in most cases²⁵.

The link to neighbourhood degree sequences then can be established by realising that the information in a subtree of height 2 in an unlabelled graph is completely captured by the node's neighbourhood degree sequence. The length of the neighbourhood degree sequence tells us how many nodes are at height 1 of its subtree kernel (i.e. the degree of the node), while the entries of the sequence tell us how many nodes at height 2 are linked to each node at height 1 (the degrees of each neighbouring node).

Methods

Real-world networks. Thirty networks were obtained from the network repository²⁷ from different research domains. Descriptions are kept to a minimum. For further details, we refer the reader to the references.

Social networks. The classical Zachary's karate club network²⁸, a dolphin social network²⁹, the Advogato network³⁰; the anybeat network; the Hamsterster network³¹; and a wikivote network³².

Biological networks. The macaque cortex network freely available from the BCT was used³³. This comes as a binary, directed network. To make this undirected we simply took all connections as undirected connections to signify whether or not any connection exists between two regions. We also look at the undirected c. elegans metabolic network³⁴; bioGRID protein networks of the fruitfly, mouse and a plant; a yeast protein interaction network³⁵; and a mouse brain network³⁶.

Ecological networks. The everglades, florida and mangwet ecosystems networks³⁷.

Economic networks. The global city network is a network of economic ties between cities³⁸. This is a weighted network which was binarised at 20% density (20% of largest weights kept) for our analysis. We also used the beacx and beaflw economic networks.

Interaction networks. A university email network³⁹; a Dublin infection network⁴⁰; and an enron email network⁴¹.

Infrastructure networks. A US and Canada airport network found in the Graph Algorithms in Matlab Code toolbox⁴²; the euroroad network⁴³; and a grid power network⁵.

Web networks. The EPA hyperlink network⁴⁴; the edu hyperlink network⁴⁵ and the indochina 2004 hyperlink network⁴⁶.

Technological networks. A router network.

In addition, we study a benchmark dataset of 406 real world networks used in⁴⁷ from the Colorado Index of Complex Networks⁴⁸. This includes 186 static networks of which just 3 overlap with the above (dolphin social network, Macaque cortex and the uni email network). It also includes two temporal networks relating to the same data of organisation affiliations each with 111 samples taken monthly from May 2002 until August 2011⁴⁹. The first of these is a network of organisation co-affiliations of directors while the other is a network of co-directorship among organisations.

Models. *Configuration models.* Random graphs with fixed degree distributions⁷ were generated using a freely available algorithm in the Brain Connectivity Toolbox³³. Fifty randomisations were computed for each real world network.

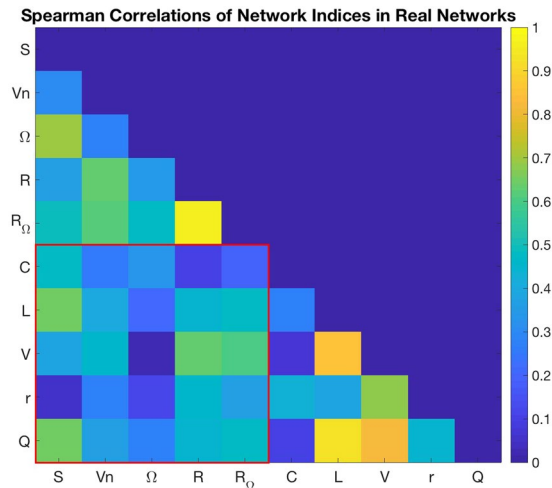


Figure 2. Absolute values of index correlations (Spearman's correlation coefficient) for combined values of small-world, scale-free, random, random geometric, and random hierarchy network models and the thirty real world networks considered (a total of 150 samples). Within the red square are the correlations between neighbourhood degree indices S - neighbourhood similarity, V_n - node heterogeneity, Ω - neighbourhood organisation, R - hierarchical complexity, R_Ω - hierarchical complexity corrected for multi-ordered degrees, and classical network indices C - transitivity, v - degree variance, L - characteristic path length, r - assortativity, and Q - modularity.

Classical global network indices. *Clustering coefficient.* The global clustering coefficient, C , measures the ratio of closed to open triples in the network. A triple is a path of length two, $\{(i, j), (j, k)\}$, where it is closed if (k, i) also exists in the network and open otherwise. It is a measure of network segregation.

Degree variance. The degree variance, $v = \text{var}(k)$, is a measure of network heterogeneity¹⁹. Here we use the normalised version⁵⁰.

Characteristic path length. The characteristic path length, L , is the average of the shortest paths existing between all pairs of nodes in the network. It is known as a measure of network integration.

Assortativity. Assortativity, r , is a correlation of the degrees of nodes which are connected in the network. It is positive if similar degree nodes are generally connected to one another, negative if similar degree nodes are generally not connected to one another and zero if there is no pattern of correlation¹⁷.

Modularity. Modularity, Q , measures the propensity of nodes to form into highly connected communities which are less connected to the rest of the network⁵¹.

Experiments

The supplementary material contains results of indices of a variety of different models- random graphs³, random geometric graphs⁴, small-world models⁵, scale-free models⁵² and random hierarchy models⁸. The main article shall focus on experiments using the most relevant data of all- over 200 real world networks.

Index correlations. Spearman correlations were computed between the proposed indices alongside classical network indices across all real networks, Fig. 2. We used Spearman's correlation since the values clearly did not follow a normal distribution (i.e. Pearson's correlation would not have been valid). The red box contains all correlations between neighbourhood degree sequence indices and classical network indices. It is clear that there are no observable high correlations between proposed indices and classical indices, providing strong evidence that indeed these new indices explain previously unrealised properties of network topology. Unsurprisingly, R and R_Ω were highly correlated, although the correlation between Ω and R_Ω was only low to moderate. But the fact there were no strong correlations other than between R and R_Ω (>0.8) suggests there is a rich amount of information to be obtained from neighbourhood degree sequences.

On the other hand, among classical network indices, strong correlations were found to exist between the L , V and Q , indicating that these indices all pointed mostly towards a single topological property of the networks. We suggest that this property is likely to be about the dominance of hub nodes, since these nodes are those which enable general short path lengths, while Newman's modularity is known to be confounded by hubs⁵³.

Although high correlations which are above the standard of 0.8 have been highlighted, there are notable moderate correlations between L and S (0.6477), Q and S (0.6419) and V and R (0.6274). However, the average correlation across all metric pairs has a magnitude of 0.4283, which would be regarded as a low-to moderate correlation. We then have to expect that measurements of a network will likely have some degree of correlation simply due to the fact that they are enacted in measuring the same topologies and since complex networks tend to show broadly consistent features in comparison with random null models. Nonetheless, the standard deviation of the metric

Type	Name	S	V_n	Ω	R	$R\Omega$	Size	Density
s	karate club	0.324 (0.062)	1.714 (2.081)	0.279 (0.062)	0.296 (0.318)	0.190 (0.269)	34	0.1390
	hi-tech firm	0.111 (0.044)	0.811 (0.993)	0.119 (0.066)	0.181 (0.183)	0.176 (0.157)	36	0.1444
	dolphins	0.113 (0.111)	0.669 (0.736)	0.076 (0.062)	0.045 (0.044)	0.040 (0.027)	62	0.0841
	wikivote	0.245 (0.249)	4.299 (4.594)	0.044 (0.042)	0.155 (0.145)	0.139 (0.121)	889	0.0074
	hamsterer	0.455 (0.124)	3.237 (4.581)	0.153 (0.012)	0.218 (0.129)	0.156 (0.120)	2426	0.0057
	advogato	0.394 (0.176)	16.977 (20.231)	0.016 (0.011)	0.607 (0.457)	0.569 (0.420)	6551	0.0019
	anybeat	0.593 (0.566)	333.632 (418.489)	0.017 (0.014)	15.561 (10.581)	11.164 (7.276)	12645	0.0006
	enron email	0.042 (0.025)	1.657 (1.711)	0.049 (0.019)	0.176 (0.151)	0.168 (0.145)	143	0.0614
	dublin contact	0.015 (0.011)	0.960 (1.137)	0.014 (0.008)	0.074 (0.044)	0.071 (0.044)	410	0.0330
	uni email	0.148 (0.144)	1.339 (1.559)	0.029 (0.027)	0.033 (0.027)	0.030 (0.023)	1133	0.0085
b	mouse brain	0.000 (0.000)	0.937 (0.930)	0.630 (0.597)	0.024 (0.016)	0.024 (0.016)	213	0.7160
	macaque cortex	0.050 (0.003)	1.116 (1.291)	0.039 (0.003)	0.436 (0.302)	0.354 (0.317)	242	0.1047
	celegans metabolic	0.265 (0.015)	18.957 (17.647)	0.209 (0.017)	2.410 (2.513)	1.757 (2.502)	453	0.0198
	mouse bioGRID	0.849 (0.792)	4.652 (11.464)	0.271 (0.173)	0.229 (0.206)	0.111 (0.133)	1455	0.0015
	plant bioGRID	0.675 (0.579)	1.363 (3.135)	0.122 (0.064)	0.050 (0.027)	0.040 (0.022)	1745	0.0020
	yeast protein	0.677 (0.664)	2.603 (3.694)	0.153 (0.136)	0.014 (0.016)	0.007 (0.013)	2114	0.0010
	fruitfly bioGRID	0.404 (0.384)	3.798 (3.775)	0.039 (0.024)	0.022 (0.016)	0.019 (0.014)	7282	0.0009
	everglades eco	0 (0)	1.300 (1.286)	0.174 (0.178)	0.323 (0.164)	0.323 (0.165)	69	0.3762
	mangwet eco	0.062 (0)	1.433 (1.323)	0.189 (0.133)	0.402 (0.281)	0.357 (0.239)	97	0.3106
florida eco	0.070 (0)	1.569 (1.642)	0.087 (0.029)	0.396 (0.192)	0.362 (0.190)	128	0.2553	
tr	US airports	0.022 (0)	0.976 (0.834)	0.076 (0.048)	1.578 (0.350)	1.552 (0.347)	456	0.3658
	euroroad	0.906 (0.912)	1.169 (1.327)	0.509 (0.550)	0.001 (0.001)	0.000 (0.000)	1174	0.0021
en	global cities	0.309 (0.249)	0.620 (0.606)	0.417 (0.356)	0.029 (0.034)	0.025 (0.035)	55	0.2000
	beacxc	0.010 (0.009)	1.332 (1.289)	0.008 (0.008)	1.238 (0.216)	1.238 (0.200)	506	0.332
	beaflw	0 (0)	1.316 (1.285)	0 (0)	0.839 (0.226)	0.839 (0.221)	507	0.352
in	EPA hyperlink	0.958 (0.767)	12.764 (7.723)	0.568 (0.113)	0.054 (0.045)	0.053 (0.070)	3031	0.0014
	edu hyperlink	0.538 (0.524)	10.752 (10.903)	0.056 (0.039)	0.073 (0.086)	0.029 (0.031)	4772	0.0008
	indochina hyperlink	0.940 (0.262)	8.608 (5.344)	0.509 (0.029)	0.029 (0.014)	0.016 (0.012)	11358	0.0007
t	techrouters	0.442 (0.376)	2.069 (3.254)	0.053 (0.028)	0.077 (0.040)	0.071 (0.034)	2113	0.0030
	power grid	0.862 (0.861)	1.314 (1.514)	0.310 (0.310)	0.001 (0.001)	0.001 (0.000)	4941	0.0005

Table 1. Neighbourhood degree sequence characteristics of 30 real-world networks from the Network Repository. Bracketed values are the means for ten realisations of configuration models. Underneath are the p -values for Wilcoxon signed rank tests and their effect sizes between real and edge-randomised values for each index. Legend: s- social, b-biological, tr- transportation, en- economic, in- informational, t- technological. S - neighbourhood similarity, V_n - relative node heterogeneity, Ω - neighbourhood organisation, R - hierarchical complexity, $R\Omega$ - hierarchical complexity update.

correlation magnitudes is 0.2269, putting one standard deviation above the mean at 0.6552 of which none of the moderate correlations previously mentioned lie above. Thus, although in usual terms these are moderate correlations, with respect to complex network metrics they appear to be within reasonable limits to suggest they broadly measure different network properties.

It is also worth recalling that correlation does not mean causation. This means that the general tendency of complex networks to exhibit correlated metrics does not necessarily mean they are measuring the same or similar property in the network, as it may be that networks which have greater modularity have greater characteristic path lengths by virtue of an underlying joint causation.

Characteristics of real-world networks. All proposed indices were applied to the thirty real-world networks of the Network Repository and the 181 non-overlapping static networks of the ICON, alongside median values taken over the two temporal networks. In addition, ten realisations of configuration models with fixed degree distributions were generated for each real-world network and we compared the neighbourhood indices of the real networks with the average values obtained from configuration models. The results are described for each Network Repository network in Table 1. Scatter plots of all real network values against configuration model values are shown in Fig. 3.

Although all indices found significant differences between real networks and configuration models, Table 2, first row, the greatest general differences found were in neighbourhood similarity, $p = 4.74 \times 10^{-28}$ with a paired ranked effect size of 0.5320, and in neighbourhood organisation, $p = 1.66 \times 10^{-23}$ with a paired ranked effect size of 0.4841. This was clearly observed in Fig. 3, first and centre plots, respectively. On the other hand, hierarchical complexity was only weakly greater in real networks than their configuration models. This was even less convincing when we took account of multi-orderedness, increasing the p -value to just below 0.05. This is interesting in light of the work done on hierarchical complexity of the human brain function and structure. Hierarchical

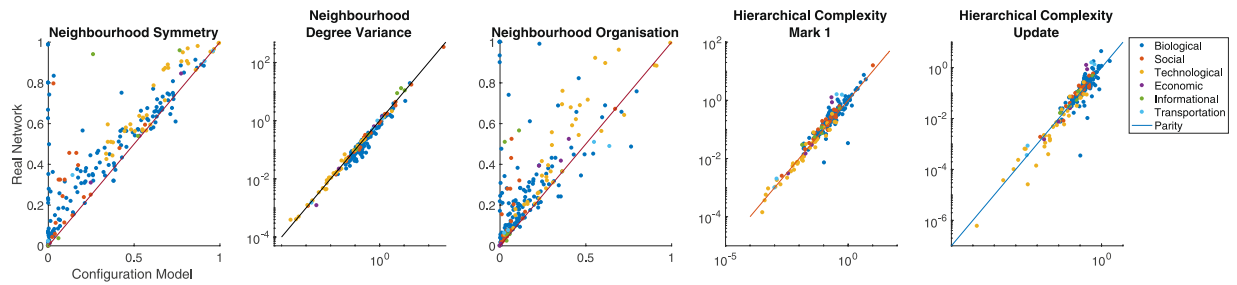


Figure 3. Scatter plots (log-log scale) of real network index values against average values for configuration models. Real world networks have generally well organised neighbourhoods for their degree distributions. They also show a tendency towards hierarchical complexity, although there is a strong drop in statistical significance when taking into account multi-ordered degrees.

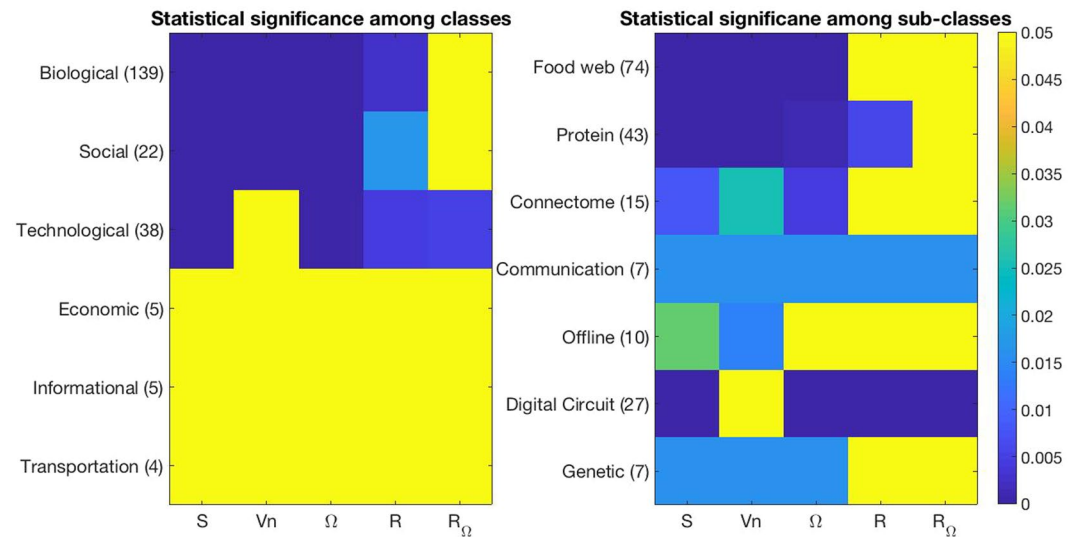


Figure 4. The *p*-values of Wilcoxon sign rank tests of the statistical difference between network indices for classes of real networks and their configuration models.

<i>p</i> -value	4.74×10^{-28}	6.07×10^{-19}	1.66×10^{-23}	9.25×10^{-4}	0.048
effect size	0.5320	-0.4308	0.4841	0.1605	0.0957

Table 2. Statistical differences between neighbourhood degree sequence characteristics of 213 real-world networks. Shown are the *p*-values for Wilcoxon signed rank tests and their effect sizes between real and edge-randomised values for each index.

complexity was not a consistent feature of real world networks and can thus be conjectured as a special feature of brain networks, where a great diversity of functional roles is present¹³.

Tentatively, hierarchical complexity also appears to be a strong property of ecological networks. We only studied three such networks here, but all had substantially higher hierarchical complexity than expected for their degree distributions, while other characteristics are not notably different from the expected values, Table 1.

We then looked at neighbourhood degree sequence properties among different network classes. We applied Wilcoxon sign rank tests, as before, but this time restricted to classes and subclasses of networks, see⁴⁷ for more details. Results are shown in Fig. 4. Greater neighbourhood organisation and similarity were found consistently among all classes with a high enough statistical power. On the other hand, technological networks, including digital circuit networks failed to find any difference in neighbourhood heterogeneity between real networks and their configuration models, suggesting a general topological difference between technological networks and biological and social networks, particularly. Interestingly, technological networks (including digital circuit networks) were found to have less hierarchical complexity than their configuration models. We expect that this is to do with a higher degree of order present in digital circuit networks, where different components connect in limited ways, constricted by the logical ordering of electronics. It was also very noticeable that the difference of hierarchical complexity in biological and social networks dropped away when updating for multi-orderedness, suggesting that

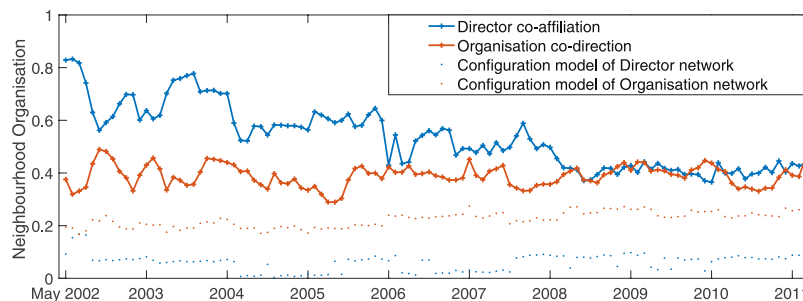


Figure 5. Temporal progression of neighbourhood organisation in the director-organisation affiliation networks.

multi-orderedness is a distinct feature of biological and social networks. In biological networks, this appeared to be driven by protein networks, where food webs and connectomes were not found to be more hierarchically complex than configuration models even from the original definition. The fact that connectomes of animals (3 cat, 5 primate, 2 macaque, 2 nematode, 2 visual cortical neuron level networks in human) were not found to have a general property of hierarchical complexity again suggests the specialness of this feature in the macro-level human brain particularly¹³ and hints towards possible links with intelligence.

Neighbourhood organisation in Norwegian director co-affiliation temporal networks. In a specific example of revealing new insights into networks using these methods, we undertook an analysis of the two temporal networks included in the ICON corpus. These were monthly sampled social networks of Norwegian company directors, where edges between directors appeared where the two were affiliated with at least one company, and concurrently sampled Norwegian company networks where edges existed where those companies shared a director⁴⁹. Both spanned the same time period from May 2002 to August 2011 and the significance of the data was that, during this time period, legislation was passed to ensure proportional representations of women in directorships to counteract structural inequalities⁴⁹. From an organisational standpoint, it stands to reason that this may cause a fairly dramatic disruption to these networks. Figure 5 shows neighbourhood organisation over time for both networks alongside that of their configuration models constructed at each time point.

It is striking that while the company network maintained similar levels of neighbourhood organisation throughout the period, the neighbourhood organisation of the director network steadily decreased throughout the period from roughly 0.8 down to around 0.4 (coinciding with company network levels) by mid 2008 where it stayed until the end of the sampling. No particular trends were noticed in either of the configuration models. Looking more closely at the director network trend, it was apparent that the decrease in neighbourhood organisation appeared almost stepwise in two year cycles with steps down around May 2004, 2006 and 2008. This validates the hypothesis that the overhaul in directorships in a short space of time contributed to a substantial disruption to the neighbourhood organisation of the network. Although it is beyond the scope of this study, it would be of interest to seek out explanations for this trend as well as possible correlations with this phenomenon and other factors.

Limitations and Future Work

There is significant scope to extend and improve on these proposed methods. A lot of the methods developed here depend on comparing nodes of the same degree, however it would be of great relevance to have this property more relaxed so that comparisons can be done across nodes of similar but not necessarily identical degrees. This is particularly the case for real-world and configuration models where the greater spontaneity of connections means that nodes which exhibit similar properties may differ in degree by one or two connections. Furthermore, this may help to create more reliable indices with less variability within populations.

We demonstrated a link between neighbourhood degree sequences and Weisfeiler-Lehman graph subtree kernels²⁵ which provide powerful graph learning results²⁶ based on long-standing graph isomorphism results²³. It would be of high interest to undertake a detailed study of the relevance of the neighbourhood degree sequence analyses for interpreting the embedding space of these graph classification approaches as network phenomena. At the same time, this link hints that analysing the diversity and structure of neighbourhood degree sequences within a network— such as hierarchical complexity and neighbourhood organisation— is indeed a very powerful and efficient way to describe the topological similarity within a network. Further detailed work is required to substantiate this conjecture.

Conclusion

We introduced several methods to understand complex networks through neighbourhood degree sequences. These targeted key concepts such as similarity and symmetry, organisation, complexity and heterogeneity. The developed network indices were not found to be strongly correlated with each other nor with classical network indices over 215 real world networks, indicating that neighbourhood degree sequences offer a rich and unique branch of analysis. We found that neighbourhood similarity and neighbourhood organisation were consistent general characteristics of complex networks. Evidence suggested that the hierarchical complexity evident in the human brain was not a general property of animal connectomes. Also, neighbourhood organisation was found

to decrease over time in a company director network where the composition of directors went through major alterations, while neighbourhood organisation in the company network remained steady. It is expected that this study will act as a springboard for new methods and applications relating to neighbourhood degree sequences, revealing important insights into networks across various disciplines.

Data Availability

The real data used in the manuscript were obtained freely online as noted in section III.A. Code for computing the network models and novel indices are available on the Open Science Framework at <https://doi.org/10.17605/OSF.IO/W7BK6>.

References

- Bollobás, B. Degree sequences of random graphs. *Discret. Math.* **33**, 1–19 (1981).
- Strogatz, S. H. Exploring complex networks. *Nat.* **410**, 268–276 (2001).
- Erdős, P. & Rényi, A. On random graphs. *Publicaciones Math. Debrecen* **6**, 290–297 (1959).
- Dall, J. & Christensen, M. Random geometric graphs. *Phys. Rev. E* **66**, 016121 (2002).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of small-world networks. *Nat.* **393**, 440–442 (1998).
- Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **6402**, 6118 (2001).
- Maslov, S. & Sneppen, K. Specificity and Stability in Topology of Protein Networks. *Sci.* **296**, 910 LP–913 (2002).
- Smith, K. & Escudero, J. The complex hierarchical topology of EEG functional connectivity. *J. Neurosci. Methods* **276**, 1–12 (2017).
- Ravasz, E. & Barabasi, A. L. Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 26112 (2003).
- Kaiser, M., Hilgetag, C. C. & Kötter, R. Hierarchy and dynamics of neural networks. *Front. Neuroinformatics* **4**, 112 (2010).
- Barthélemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92**, 178701 (2004).
- Smith, K., Abásolo, D. & Escudero, J. Accounting for the Complex Hierarchical Topology of EEG Phase-based Functional Connectivity in Network Binarisation. *PLOS One* **12**, e0186164 (2017).
- Smith, K. *et al.* Hierarchical Complexity of the Adult Human Structural Connectome. *Neuroimage* **191**, 205–215 (2019).
- Barrus, M. & Donovan, E. Neighbourhood degree lists of graphs. *Discret. Math.* **341**, 175–183 (2018).
- Nishimura, N. & Subramanya, V. Graph editing to a given neighbourhood degree list is fixed-parameter tractable. In Gao, X., Du, H. & Han, M. (eds) *COCOA 2017: Combinatorial optimization and applications*, vol. 10628 of *Lecture Notes in Computer Science*, 138–153 (Springer, Cham, 2017).
- Bonacich, P. Factoring and weighting approaches to clique identification. *J. Math. Sociol.* **2**, 113–120 (1972).
- Newman, M. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Solé R. & Valverde, S. Complex Networks. vol. 650 of *Lecture Notes in Physics*, chap. Informatio, 189–207 (Springer, 2004).
- Snijders, T. A. B. The degree variance: an index of graph heterogeneity. *Soc. Networks* **3**, 163–174 (1981).
- Bell, F. K. A note on the irregularity of graphs. *Lin. Alg. Appl.* **161**, 45–64 (1992).
- Estrada, E. Quantifying network heterogeneity. *Phys. Rev. E* **82**, 066102 (2010).
- Frucht, R. Herstellung von Graphen mit vorgegebener abstrakter Gruppe. *Compos. Math.* **6**, 239–250 (1939).
- Weisfeiler, B. & Lehman, A. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Tekhnicheskaya Informatsiya* **2**, 12–16 (1968).
- Babai, L. & Kucera, L. Canonical labelling of graphs in linear average time. In *Proceedings Symposium on Foundations of Computer Science*, 39–46 (1979).
- Shervashidze, N., Schweitzer, P., van Leeuwen, E., Mehlhorn, K. & Borgwardt, K. Weisfeiler-lehman graph kernels. *J. Mach. Learn. Res.* **12**, 2539–2561 (2011).
- Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? <https://arxiv.org/abs/1810.00826> (2018).
- Rossi, R. A. & Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015).
- Zachary, W. W. An Information Flow Model for Conflict and Fission in Small Groups. *J. Anthro. Research* **33**, 452–473 (1977).
- Lusseau, D. *et al.* The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology* **54**, 396–405 (2003).
- Massa, P., Salvetti, M. & Tomasoni, D. Bowling alone and trust decline in social network sites. In *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*, 658–663 (IEEE, 2009).
- Hamsterster. Hamsterster social network, <http://www.hamsterster.com>.
- Leskovec, J., Huttenlocher, D. & Kleinberg, J. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1361–1370 (ACM, 2010).
- Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).
- Duch, J. & Arenas, A. Community identification using extremal optimization. *Phys. Rev. E* **72**, 027104 (2005).
- Jeong, H., Mason, S., Barabasi, A. & Oltvai, Z. Lethality and centrality in protein networks. *arXiv preprint cond-mat/0105306* (2001).
- Amunts, K. *et al.* Bigbrain: An ultrahigh-resolution 3d human brain model. *Sci.* **340**, 1472–1475 (2013).
- Melián, C. J. & Bascompte, J. Food web cohesion. *Ecol.* **85**, 352–358 (2004).
- Taylor, P. Specification of the world city network. *Geogr. Analysis* **33**, 181–194 (2001).
- Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103 (2003).
- SocioPatterns. Infectious contact networks, <http://www.sociopatterns.org/datasets/>. Accessed 09/12/12.
- Cohen, W. Enron email dataset. <http://www.cs.cmu.edu/enron/>. Accessed in 2009.
- The US airport network. https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/24134/versions/1/previews/gaimc/demo/html/airports.html?access_key=
- Bader, D. A., Meyerhenke, H., Sanders, P. & Wagner, D. Graph partitioning and graph clustering. In *10th DIMACS Implementation Challenge Workshop* (2012).
- De Nooy, W., Mrvar, A. & Batagelj, V. *Exploratory social network analysis with Pajek*, vol. 27 (Cambridge University Press, 2011).
- Gleich, D., Zhukov, L. & Berkhin, P. Fast parallel pagerank: A linear system approach. *Yahoo! Research Technical Report YRL-2004-038* **13**, 22 (2004).
- Boldi, P., Rosa, M., Santini, M. & Vigna, S. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *WWW*, 587–596 (2011).
- Ghasemian, A., Hosseinmardi, H. & Clauset, A. Evaluating overfit and underfit in models of network community structure, <https://arxiv.org/abs/1802.10582>.
- Clauset, A., Tucker, E. & Sainz, M. The colorado index of complex networks.

49. Seierstad, C. & Opsahl, T. For the few not the many? the effects of affirmative action on presence, prominence, and social capital of women directors in norway. *Scand. J. Manag* **27**, 44–54 (2011).
50. Smith, K. & Escudero, J. Normalised degree variance, <https://arxiv.org/abs/1803.03057>.
51. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 26113 (2004).
52. Barabási, A.-L. & Albert, R. Emergence of Scaling in Random Networks. *Sci.* **286**, 509 LP–512 (1999).
53. Yang, J. & Leskovec, J. Overlapping communities explain core-periphery organization of networks. *Proceedings of the IEEE* **102**, 1892–1902 (2014).

Acknowledgements

We would like to thank Aaron Clauset for helpful discussions and provision of the data from the Colorado Index of Complex Networks. This work was supported by Health Data Research UK (MRC ref Mr/S004122/1), which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, National Institute for Health Research (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. A version of this article has been made available on an online preprint server at <https://arxiv.org/abs/1901.02353>.

Author Contributions

K.S. is the sole author and did all the work.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-44907-8>.

Competing Interests: The author declares no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019