

PUBLISHED VERSION

Mark D. McDonnell, Ömer Nebil Yaveroğlu, Brett A. Schmerl, Nicolangelo Iannella, Lawrence M. Ward **Motif-role-fingerprints: the building-blocks of motifs, clustering-coefficients and transivities in directed networks**

PLoS One, 2014; 9(12):e114503-1-e114503-25

2014 McDonnell et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PERMISSIONS

<http://www.plosone.org/static/license>

Open-Access License



No Permission Required

PLOS applies the [Creative Commons Attribution \(CC BY\) license](#) to all works we publish (read the [human-readable summary](#) or the [full license legal code](#)). Under the CC BY license, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLOS journals, so long as the original authors and source are cited. **No permission is required from the authors or the publishers.**

In most cases, appropriate attribution can be provided by simply citing the original article (e.g., Kaltenbach LS et al. (2007) Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. *PLOS Genet* 3(5): e82. doi:10.1371/journal.pgen.0030082). If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then please indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared. For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

This broad license was developed to facilitate open access to, and free use of, original works of all types. Applying this standard license to your own work will ensure your right to make your work freely and openly available. Learn more about [open access](#). For queries about the license, please [contact us](#).

9 June, 2015

<http://hdl.handle.net/2440/91743>

RESEARCH ARTICLE

Motif-Role-Fingerprints: The Building-Blocks of Motifs, Clustering-Coefficients and Transitivity in Directed Networks

Mark D. McDonnell^{1,3*}, Ömer Nebil Yaverolu², Brett A. Schmerl¹, Nicolangelo Iannella¹, Lawrence M. Ward³

1. Computational and Theoretical Neuroscience Laboratory, Institute for Telecommunications Research, University of South Australia, Mawson Lakes, South Australia, **2.** California Institute of Telecommunications and Information Technology (Calit2), University of California Irvine, Irvine, California, United States of America, **3.** Department of Psychology and Brain Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

*mark.mcdonnell@unisa.edu.au



CrossMark
click for updates

OPEN ACCESS

Citation: McDonnell MD, Yaverolu ÖN, Schmerl BA, Iannella N, Ward LM (2014) Motif-Role-Fingerprints: The Building-Blocks of Motifs, Clustering-Coefficients and Transitivity in Directed Networks. PLoS ONE 9(12): e114503. doi:10.1371/journal.pone.0114503

Editor: Michal Zochowski, University of Michigan, United States of America

Received: April 9, 2014

Accepted: November 10, 2014

Published: December 8, 2014

Copyright: © 2014 McDonnell 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: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper.

Funding: Mark D. McDonnell's contribution was by supported by an Australian Research Fellowship from the Australian Research Council (project number DP1093425) (www.arc.gov.au), an Endeavour Award from the Australian Government (<http://www.innovation.gov.au/InternationalEducation/EndeavourAwards/>) and the National Health and Medical Research Council (NHMRC) of Australia (project grant, APP1050832) (www.nhmrc.gov.au). Lawrence M. Ward's contribution was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada (www.nserc-crsng.gc.ca). Ömer Nebil Yaverolu's contribution was supported by the United States of America National Science Foundation (NSF) Cyber-Enabled Discovery and Innovation (CDI) grant OIA-1028394. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Mark D. McDonnell is a member of the Editorial Board of PLOS ONE.

Abstract

Complex networks are frequently characterized by metrics for which particular subgraphs are counted. One statistic from this category, which we refer to as *motif-role fingerprints*, differs from global subgraph counts in that the number of subgraphs in which each node participates is counted. As with global subgraph counts, it can be important to distinguish between motif-role fingerprints that are 'structural' (induced subgraphs) and 'functional' (partial subgraphs). Here we show mathematically that a vector of all functional motif-role fingerprints can readily be obtained from an arbitrary directed adjacency matrix, and then converted to structural motif-role fingerprints by multiplying that vector by a specific invertible conversion matrix. This result demonstrates that a unique structural motif-role fingerprint exists for any given functional motif-role fingerprint. We demonstrate a similar result for the cases of functional and structural motif-fingerprints without node roles, and global subgraph counts that form the basis of standard motif analysis. We also explicitly highlight that motif-role fingerprints are elemental to several popular metrics for quantifying the subgraph structure of directed complex networks, including motif distributions, directed clustering coefficient, and transitivity. The relationships between each of these metrics and motif-role fingerprints also suggest new subtypes of directed clustering coefficients and transivities. Our results have potential utility in analyzing directed synaptic networks constructed from neuronal connectome data, such as in terms of centrality. Other potential applications include anomaly detection in networks, identification of similar networks and identification of similar nodes within networks.

Matlab code for calculating all stated metrics following calculation of functional motif-role fingerprints is provided as [S1 Matlab File](#).

Introduction

Complex relational systems from different domains, such as biology, sociology or economics, can be systematically analyzed using their network representations. A *network* (also known as a *graph*) is composed of nodes and edges, where *nodes* represent the entities in the system and *edges* represent the relationships between these entities. Depending on the type of represented relations, the node pairs that form the edges can have a certain ordering, in which case the resulting network is called *directed*. For example, in networks of biological neurons and synapses (also known as *neuronal connectomes* [1]), the nodes correspond to individual neurons, while directed edges between the nodes (typically) represent the existence of chemical synapses that enable communications between neurons [2]. The wiring patterns of networks cast light on the functional mechanisms of the analyzed complex systems, and therefore, network structure analysis is gaining increasing interest from different disciplines.

However, many network analysis problems are computationally intractable [3]. Therefore, the only available solutions are based on approximations to the exact solutions of these problems. *Network properties* that describe different wiring characteristics of networks are used for this purpose. For example, given two networks without any labeling on the nodes, the problem of finding all the node pairs that have identical wiring patterns in the two networks is a computationally intractable problem. However, this problem can be simplified by computing the *degrees* (i.e., the number of neighbors a node has) of all nodes and using the degree statistics to compare the nodes. Even if the resulting matches are not guaranteed to have identical wiring patterns, these matches would extensively reduce the size of the search space. The search space can be reduced even further by computing other network properties that capture different types of interaction patterns; e.g., using the similarities of *clustering coefficients* that measure the tendency of nodes to form triangular interactions [4].

Different *subgraphs* of a network can be obtained from different subsets of its nodes and edges. Many of the network properties are indeed dependent on the subgraph properties of the networks; e.g., clustering coefficient is defined based on three-node subgraphs of a network in which all nodes are connected with each other forming a triangle. In a *connected* subgraph, all nodes are reachable from any of the other nodes in the subgraph. A subgraph is *induced* (also known as *node induced*) if it is enforced that all the edges between the chosen subset of nodes are included in the subgraph. The subgraphs that do not carry the induced property are called *partial* (also known as *edge induced*) subgraphs. For example, a 3-node *clique* contains 3 different two-path subgraphs (two-path subgraphs are those that

contain 3 nodes and 2 edges) when partial subgraph properties are considered. However, such a graph does not contain any two-path subgraphs when induced subgraph properties are considered.

Triangular patterns in networks are commonly utilized to analyze the network topology. In undirected networks, the *clustering coefficient* of a node is calculated by dividing the number of triangles around the node by the number of different pairs of its neighbors [5]. *Average clustering coefficient* explains the clustering (triangulation) within a network by averaging the clustering coefficients of all its nodes. Extension of clustering coefficient to directed networks is not trivial since there are two different types of triangular directed subgraphs; one being a cyclic subgraph ($m=5$ in Figs. 1 and 2) and the other being an acyclic subgraph ($m=9$ in Figs. 1 and 2). Based on the counts of the four distinct node roles on these two subgraphs (i.e., $r=11,14,17$ and 18 in Figs. 1 and 2), the definition of clustering coefficient has been extended to the directed case [4, 6]. A different metric for quantifying network clustering known as *transitivity* is calculated by considering every possible combination of three nodes in a network, and counting how many of these triads are mutually connected by three edges, normalized by the number of triads with at least two edges [7]. It is similar to clustering coefficient but unlike that metric, it is not an average of local node-specific clustering. Transitivity is typically used for undirected networks rather than directed ones, but an expression for directed transitivity is given in [8].

Recent work on network properties use the statistics of all observable connected subgraph configurations as detailed descriptors of the wiring in networks [9, 10]. *Network motifs* were originally defined as the partial subgraph patterns of a network that appears more frequently than expected from a 'null-hypothesis' network model that preserve the input network's degree distribution, or other statistical properties [9, 11–15]. Network motifs are defined for both directed and undirected networks, covering all observable subgraphs patterns on sets of nodes ranging in dimension from 2 to n . Network motifs have been used to analyze network structures of a wide-range of networks, such as those of the neuronal connectome of *C. elegans* [16–20]. Practically, network motif analyses are performed with 3-node subgraph patterns due to the high computational cost of null model generation step for larger subgraphs; all directed 3-node subgraph patterns are illustrated in Fig. 1.

Another group of network properties that are based on subgraph counts have been studied in the context of *graphlets*—these are small, connected, non-isomorphic and induced subgraphs of a large network [10]. There are three major differences between network motifs and graphlets:

1. network motifs account for partial subgraphs while graphlets are based on induced subgraphs;
2. network motifs are dependent on a given null network model while graphlets are completely independent from any null hypotheses; and
3. graphlets are defined only for undirected graphs while network motifs are defined also for directed graphs.

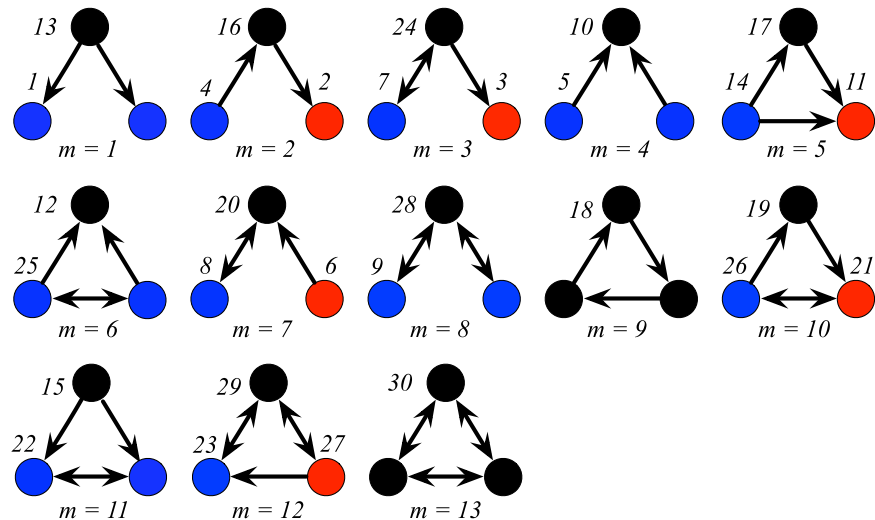


Fig. 1. All 13 three-node connected motifs and all 30 three-node connected motif-roles. A directed network is assumed. The numerical label for each motif (denoted with the label m) is identical to that used in [9]. Each distinct motif-role within each motif is denoted by different colours, and the numerical label next to each node. The numerical label provided for each motif-role is represented by the label r in the text and in Fig. 2, where $r = 1, \dots, 30$.

doi:10.1371/journal.pone.0114503.g001

The number of times that each graphlet appears in a network describes the network's topology [10]. Currently, the most advanced method for describing the topology of an undirected network is based on the dependencies between different graphlets [21].

Subgraph properties are not only useful for describing the topology of networks, but they can also be utilized for describing the local wiring around nodes. For instance, degree describes the wiring around a node by counting the number of edges touching the node. Replacing edges with subgraphs of each kind in this definition, the local wiring around a node can be described by the number of subgraph patterns that the node participates in. While these subgraph statistics on nodes can be computed without imposing any orientations on the subgraphs [8, 22], a node's *role* in the network can be characterized more accurately by introducing such orientation constraints based on the symmetries within the subgraphs [23–25]. For example, as illustrated in Figure 1 of [23], and Fig. 1 here, there are 30 unique motif-roles on the 3-node directed subgraph configurations. Przulj [25] identifies the *orbits* (i.e., the nodes that have identical wiring patterns within graphlets) of all 2- to 5-node graphlets and uses these orbits to describe the wiring around a node by defining *graphlet degree*, which is the number of graphlets that touch a node at an orbit. Furthermore, the vector containing the graphlet degrees of all 73 orbits of 2- to 5-node graphlets is termed the *graphlet degree vector* and successfully applied for identifying the wiring similarities between the nodes of a network, and also, between the nodes of different networks [26, 27]. It has been argued that analysis of neuronal connectome data will need to take into account node-referenced heterogeneity [28–30], such as measured by

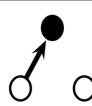
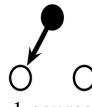
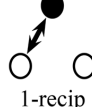
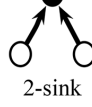
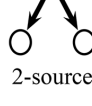
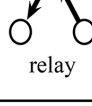
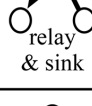
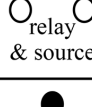
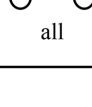
Role	Motif-roles			
 1-sink	$r = 1$ $m = 1$ $d_r = 2$ $\mathcal{A}^\top (\mathcal{A} - \mathcal{I}) \mathbf{1}$	$r = 2$ $m = 2$ $d_r = 1$ $((\mathcal{A}^\top)^2 - \mathcal{R}) \mathbf{1}$	$r = 3$ $m = 3$ $d_r = 1$ $(\mathcal{A}^\top \mathcal{R} - \mathcal{R} \circ \mathcal{A}) \mathbf{1}$	
 1-source	$r = 4$ $m = 2$ $d_r = 1$ $(\mathcal{A}^2 - \mathcal{R}) \mathbf{1}$	$r = 5$ $m = 4$ $d_r = 2$ $\mathcal{A} (\mathcal{A}^\top - \mathcal{I}) \mathbf{1}$	$r = 6$ $m = 7$ $d_r = 1$ $(\mathcal{A} \mathcal{R} - \mathcal{A} \circ \mathcal{R}) \mathbf{1}$	
 1-recip	$r = 7$ $m = 3$ $d_r = 1$ $(\mathcal{R} \mathcal{A} - \mathcal{R} \circ \mathcal{A}) \mathbf{1}$	$r = 8$ $m = 7$ $d_r = 1$ $(\mathcal{R} \mathcal{A}^\top - \mathcal{R} \circ \mathcal{A}) \mathbf{1}$	$r = 9$ $m = 8$ $d_r = 2$ $\mathcal{R} (\mathcal{R} - \mathcal{I}) \mathbf{1}$	
 2-sink	$r = 10$ $m = 4$ $d_r = 1$ $\frac{1}{2} \mathcal{A}^\top \mathbf{1} \circ (\mathcal{A}^\top - \mathcal{I}) \mathbf{1}$	$r = 11$ $m = 5$ $d_r = 1$ $((\mathcal{A}^\top \mathcal{A}) \circ \mathcal{A}^\top) \mathbf{1}$	$r = 12$ $m = 6$ $d_r = 1$ $\frac{1}{2} ((\mathcal{A}^\top \mathcal{R}) \circ \mathcal{A}^\top) \mathbf{1}$	
 2-source	$r = 13$ $m = 1$ $d_r = 1$ $\frac{1}{2} \mathcal{A} \mathbf{1} \circ (\mathcal{A} - \mathcal{I}) \mathbf{1}$	$r = 14$ $m = 5$ $d_r = 1$ $(\mathcal{A}^2 \circ \mathcal{A}) \mathbf{1}$	$r = 15$ $m = 11$ $d_r = 1$ $\frac{1}{2} ((\mathcal{A} \mathcal{R}) \circ \mathcal{A}) \mathbf{1}$	
 relay	$r = 16$ $m = 2$ $d_r = 1$ $\mathcal{A}^\top \mathbf{1} \circ \mathcal{A} \mathbf{1} - \mathcal{R} \mathbf{1}$	$r = 17$ $m = 5$ $d_r = 1$ $((\mathcal{A} \mathcal{A}^\top) \circ \mathcal{A}^\top) \mathbf{1}$	$r = 18$ $m = 9$ $d_r = 3$ $(\mathcal{A}^2 \circ \mathcal{A}^\top) \mathbf{1}$	$r = 19$ $m = 10$ $d_r = 1$ $((\mathcal{A} \mathcal{R}) \circ \mathcal{A}^\top) \mathbf{1}$
 relay & sink	$r = 20$ $m = 7$ $d_r = 1$ $(\mathcal{R} \mathbf{1}) \circ ((\mathcal{A}^\top - \mathcal{I}) \mathbf{1})$	$r = 21$ $m = 10$ $d_r = 1$ $((\mathcal{R} \mathcal{A}) \circ \mathcal{A}^\top) \mathbf{1}$	$r = 22$ $m = 11$ $d_r = 2$ $((\mathcal{A}^\top \mathcal{A}) \circ \mathcal{R}) \mathbf{1}$	$r = 23$ $m = 12$ $d_r = 1$ $(\mathcal{R}^2 \circ \mathcal{A}^\top) \mathbf{1}$
 relay & source	$r = 24$ $m = 3$ $d_r = 1$ $(\mathcal{R} \mathbf{1}) \circ ((\mathcal{A} - \mathcal{I}) \mathbf{1})$	$r = 25$ $m = 6$ $d_r = 2$ $((\mathcal{R} \mathcal{A}) \circ \mathcal{A}) \mathbf{1}$	$r = 26$ $m = 10$ $d_r = 1$ $(\mathcal{A}^2 \circ \mathcal{R}) \mathbf{1}$	$r = 27$ $m = 12$ $d_r = 1$ $((\mathcal{A} \mathcal{R}) \circ \mathcal{R}) \mathbf{1}$
 all	$r = 28$ $m = 8$ $d_r = 1$ $\frac{1}{2} \mathcal{R} \mathbf{1} \circ ((\mathcal{R} - \mathcal{I}) \mathbf{1})$	$r = 29$ $m = 12$ $d_r = 1$ $((\mathcal{R} \mathcal{A}) \circ \mathcal{R}) \mathbf{1}$	$r = 30$ $m = 13$ $d_r = 3$ $\frac{1}{2} (\mathcal{R}^2 \circ \mathcal{R}) \mathbf{1}$	

Fig. 2. Formulae for counting the three-node motif-role fingerprints. The first column depicts the 9 distinct roles on functional motifs. Each row shows each three-node motif in which the corresponding role appears (indexed by $m = 1, \dots, 13$), and the plurality d_r with which motif-role r appears within motif m (see [Methods](#)).

Black filled circles indicate the nodes in motif m that play motif-role r (see also Fig. 1). The equations shown for each role, r , are the entries of the functional motif-role fingerprint matrix, $\mathcal{F}_{R,(r,1:N)}$, where \circ denotes the Hadamard product, $\mathbf{1}$ is an $N \times 1$ unit column matrix, \mathcal{I} is the $N \times N$ identity matrix, and $\mathcal{R} := \mathcal{A} \circ \mathcal{A}^T$ is the matrix of reciprocal edges.

doi:10.1371/journal.pone.0114503.g002

graphlet degree. Another possible application is in the analysis of genetic networks [31].

The terminology on subgraph properties is not well-defined, with some studies using the terms “subgraphs”, “network motifs” and “graphlets” interchangeably. In order to avoid confusion, we use the term “*functional motifs*” to represent the partial subgraph properties (e.g., network motif properties defined in [9]), and “*structural motifs*” to represent the induced subgraph properties (e.g., graphlet properties defined in [10, 25]) in a consistent manner with [8]. Structural motifs quantify anatomical building blocks, whereas functional motifs represent elementary processing modes of the networks [22]. This distinction between structural and functional subgraph properties have different implications for neuronal networks: structural motifs describe all synapses amongst a specific subset of neurons. In contrast, functional motifs can describe, for example, potential patterns of actual synaptic activations occurring (near) simultaneously amongst a specific subset of neurons. It is expected to observe correlation between structural and functional subgraph properties to some extent. Even though this is the case, the wiring characteristics that can be captured by these two types of subgraphs differ. For example, a node’s importance in the networks as a ‘broker’ (e.g., $r = 16$ in Fig. 2) can only be captured by structural motifs since functional motifs consider also the cases that the node appears as roles $r = 17, 18$ or 19 (Fig. 2). In these cases, the reference node is not a broker because of the edge between the two other nodes.

For both structural and functional motifs, we consider four different types of subgraph frequency derived network properties, as follows:

- **Global Metrics:** These metrics aim to describe the topology of an entire network.
- **Motif Counts:** A network’s topology can be described by the number of subgraphs that appear in the network. We use the term *motif counts* to represent these networks statistics. Different from the original definition of network motifs [9] (but consistent with usage in [8]), our motif statistics are independent of any comparison to null-hypothesis network model. For a given network, the corresponding motif counts form a M dimensional vector, each value representing the count for one of the M subgraphs.
- **Motif-Role Counts:** A network’s topology can also be described in terms of the roles within subgraphs. We use the term *motif-role counts* to represent the number of times that a given motif role appears in a network. Motif-role counts can be directly obtained by scaling the motif counts depending on the number of times the motif-role appears within the corresponding subgraph. For a given

network, the corresponding motif-role counts form an L dimensional vector, each value representing the number of times one of the L node roles appears in the network.

- **Node-referenced Metrics:** These metrics aim to describe the local topology around a node in the network.
- **Motif Fingerprints:** The wiring around a node in a network can be described by the number of subgraph patterns that it participates in, independent of the position (i.e., the role) on these subgraphs. Such statistics have been termed *motif fingerprints* [8, 22]. For each of the N nodes in a given network, the corresponding motif fingerprints are M dimensional vectors, each value corresponding to count of one of the M subgraphs that the node participates in.
- **Motif-Role Fingerprints:** The wiring around a node in the network can be described at a finer detail by the number of subgraphs that touches the node at a specific orientation (i.e., on a node-role within the subgraph). We term such statistics as *motif-role fingerprints*. For each of the N nodes in a given network, the corresponding motif-role fingerprints are L dimensional vectors, each value corresponding to the number of subgraphs that touches a node at one of the L node-role positions.

In this study, we explore the relationships between all these different types of subgraph statistics (see Fig. 3). First, we present efficient ways of calculating the functional motif-role fingerprints of a given directed network. Second, we show that structural motif statistics can be derived from functional motif statistics and vice versa. This transformation enables efficient computation of structural motif-fingerprints which are computationally more expensive to obtain. Third, we show that the motif-role fingerprints are the most fundamental and informative of all the other subgraph metrics. We identify the transformations that derive all other subgraph statistics (i.e., motif fingerprints, motif-role counts, motif counts) from the motif-role fingerprints. Fourth, we discuss the relationships between motif-role fingerprints and directed clustering coefficients and transivities, and show how these can be derived from motif-role fingerprints. Finally, we illustrate applications of these transformations on the neuronal connectome of *c. elegans*.

Results and Discussion

While exploring the relationships between different subgraph properties, we assume a directed network with N nodes. The *adjacency matrix* representation of a network (\mathcal{A}) is an $N \times N$ matrix, where $\mathcal{A}[i,j]$ is 1 when there exists a directed edge from node i to node j , and otherwise 0. We label each of the $M = 13$ connected three-node motifs with the index $m = 1, \dots, 13$ according to the classification introduced by [9]—see Fig. 1. When structural motifs of a directed network are considered, there are $L = 30$ different motif-roles, which we label with the index $r = 1, \dots, 30$, as illustrated in Fig. 1. However, when considering the functional

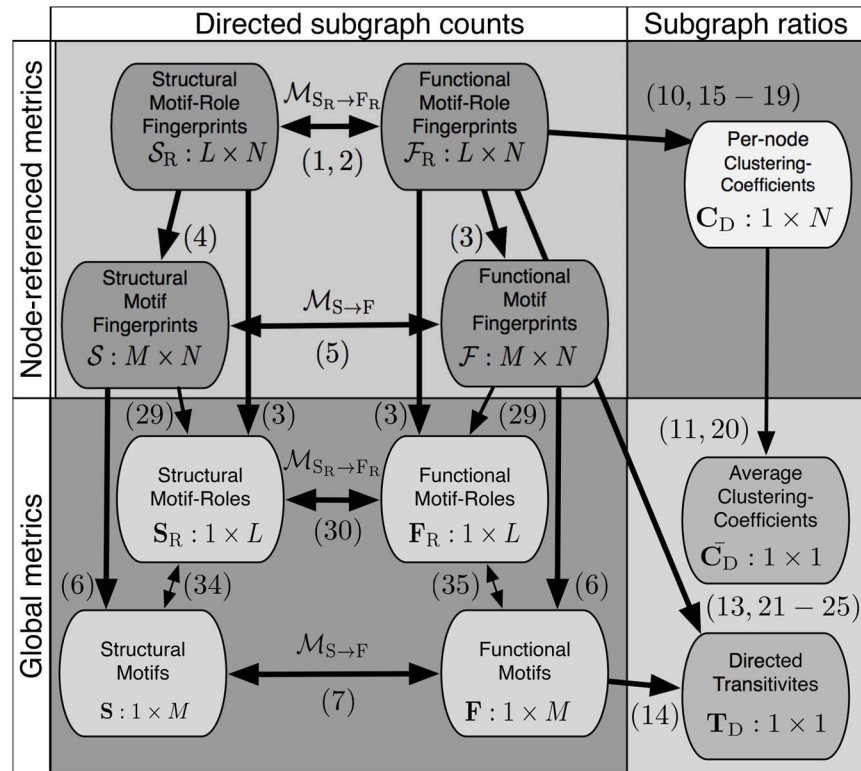


Fig. 3. Dependencies between metrics that count three-node directed subgraphs. Arrows indicate that metrics can be derived from other metrics and numbers in brackets refer to equations in the text that mathematically describe these dependencies. The left side of the figure lists metrics that count subgraphs, while the right side shows metrics that are ratios of subgraph counts. The top half of the figure shows metrics that are node-referenced subgraph counts, while the bottom half shows metrics that are global subgraph counts.

doi:10.1371/journal.pone.0114503.g003

motifs, these 30 motif-roles induces on 9 distinct roles—see Fig. 2. The ordering of our labels is determined by these roles, and hence is non-sequential when depicted in Fig. 1.

Calculating Functional Motif-Role Fingerprints

We introduce two $L \times N$ matrices, S_R and F_R , where the elements of the i -th column of these matrices is the transpose of the $1 \times L$ vector that denotes the structural motif-role fingerprints and functional motif-role fingerprints, respectively, in which node i participates. Fig. 2 lists equations that can be used to efficiently obtain all elements of the matrix F_R , in terms of the adjacency matrix, A . Further explanation on the computation of functional motif-role fingerprints is provided in the Methods section.

The Relationship Between Structural and Functional Motif-Role Fingerprints

Structural motifs (as counted for an overall network) can contain multiple functional motifs as illustrated in Fig. 4. We extend the distinction between structural and functional motifs, and show that the motif-role fingerprints of these two types of motifs can be derived from each other.

The mathematical relationship between structural and functional motif-role fingerprints can be conveniently expressed as

$$\mathcal{F}_R = \mathcal{M}_{S_R \rightarrow F_R} \mathcal{S}_R, \tag{1}$$

where $\mathcal{M}_{S_R \rightarrow F_R}$ is an invertible $L \times L$ upper-triangular matrix, in which element (i, j) indicates how many copies of functional motif-role i are contained in structural motif-role j (see Equation (27) in Methods).

The fact that this matrix is invertible is important for numerical calculation of structural motif-role fingerprints. Although expressions for functional motif-role fingerprints can be efficiently calculated (see above and Fig. 2), it is more difficult to derive simple expressions for structural motif-role fingerprints. Instead, the inverse relationship

$$\mathcal{S}_R = \mathcal{M}_{S_R \rightarrow F_R}^{-1} \mathcal{F}_R, \tag{2}$$

where $\mathcal{M}_{S_R \rightarrow F_R}^{-1}$ is given by Equation (28) in Methods, enables the structural motif-role fingerprint vector to be obtained without directly using the adjacency matrix. Moreover, the fact that $\mathcal{M}_{S_R \rightarrow F_R}$ is invertible means that a unique structural motif-role fingerprint vector exists for any given functional motif-role fingerprint vector.

Motif-Fingerprints and Global Motif Counts from Motif-Role Fingerprints

We now introduce the *motif-fingerprint* matrices, \mathcal{S} and \mathcal{F} , each of size $M \times N$, where the elements of the i -th column of these matrices denote the total number of structural motifs and functional motifs respectively in which node i participates [22]. The entries in the motif-fingerprints matrix can be trivially obtained from the motif-role fingerprints as follows:

$$\mathcal{F}_{(m, 1:N)} = \sum_{k \in Q_m} \mathcal{F}_{R, (k, 1:N)}, \quad m = 1, \dots, M, \tag{3}$$

$$\mathcal{S}_{(m, 1:N)} = \sum_{k \in Q_m} \mathcal{S}_{R, (k, 1:N)}, \quad m = 1, \dots, M, \tag{4}$$

where Q_m is the set of motif-role indices corresponding to motif index m . These sets can be readily identified in Fig. 1. The relationship between structural and functional motif fingerprints can be expressed as

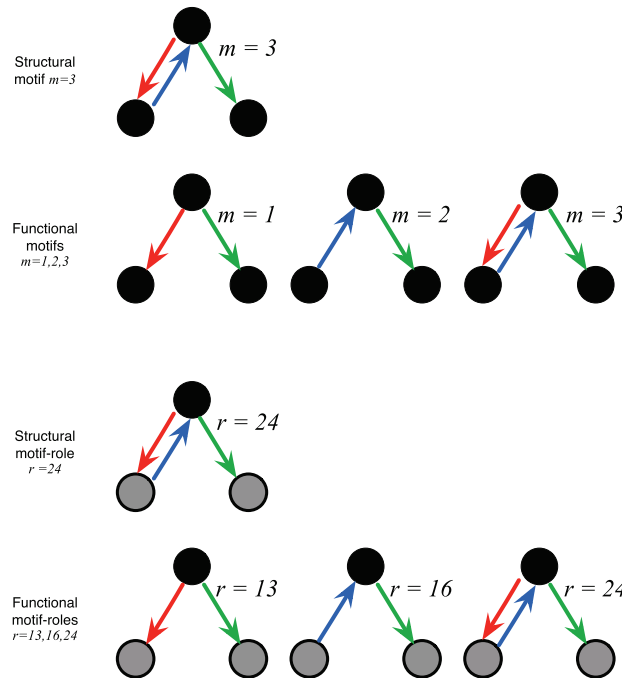


Fig. 4. Structural motifs and motif-roles decompose into functional motifs and motif-roles. Illustration of the difference between structural and functional motifs and motif-roles. When counting structural motifs in a network, the connectivity between each set of three nodes is considered. In this case, if the nodes form motif $m=3$, then this counts as one instance of structural motif $m=3$, and no instances of structural motifs 1 or 2. However, the same subgraph provides one instance each of functional motifs $m=1$, $m=2$, and $m=3$ (see also Fig. 1 in [22] for a similar illustration). Consequently, there are no more structural motifs in total than the number of combinations of three nodes. However, this is not the case for functional motifs, since the same set of three nodes can contain multiple functional motifs. The same decomposition occurs for motif-roles. In the example in this figure, a single instance of structural motif-role $r=24$ decomposes into one instance each of functional motif-roles $r=13$, $r=16$ and $r=24$.

doi:10.1371/journal.pone.0114503.g004

$$\mathcal{F} = \mathcal{M}_{\mathcal{S} \rightarrow \mathcal{F}} \mathcal{S}, \tag{5}$$

where $\mathcal{M}_{\mathcal{S} \rightarrow \mathcal{F}}$ is a 13×13 upper-diagonal invertible matrix in which element (i, j) indicates how many copies of functional motif i are contained in structural motif j (see Equation (31) in Methods).

Various methods exist for obtaining motif counts within networks, as reviewed by [32]. Here, we state how such counts for three-node motifs can be calculated from motif fingerprints. We introduce the length M vectors \mathbf{S} and \mathbf{F} , where the elements of each vector (S_m and $F_m, m=1, \dots, M$) denote the total number of structural motifs and functional motifs, respectively. Obtaining the global motif counts from the motif fingerprints is a simple matter of summing the fingerprints for all nodes, and dividing by three, since each global motif appears in the fingerprint of exactly three nodes:

$$\mathbf{S}^\top = \frac{1}{3} \mathcal{S} \mathbf{1}, \quad \mathbf{F}^\top = \frac{1}{3} \mathcal{F} \mathbf{1}, \tag{6}$$

where we also have

$$\mathbf{F}^\top = \mathcal{M}_{\mathcal{S} \rightarrow \mathcal{F}} \mathbf{S}^\top. \tag{7}$$

Similarly to motif-role fingerprints, the existence of an invertible matrix for converting between functional and structural motifs implies that a unique structural motif or motif-fingerprint vector exists for any given functional motif or motif fingerprint vector.

Directed Clustering Coefficients & Transitivityes from Motif-Role Fingerprints

We now consider directed clustering coefficients and directed transitivityes, and demonstrate how they are simple derivatives of motif-role fingerprints. We begin by defining two length N vectors; the first is the total number of closed directed triangles in which each node participates,

$$\begin{aligned} \mathbf{V} &= \mathcal{F}_{R,(11,1:N)} + \mathcal{F}_{R,(14,1:N)} + \mathcal{F}_{R,(17,1:N)} + \mathcal{F}_{R,(18,1:N)} \\ &= 3(\mathcal{F}_{5,(1:N)} + \mathcal{F}_{9,(1:N)}), \end{aligned} \tag{8}$$

and the second is the total number of potential triangles in which each node may participate,

$$\mathbf{T} = 2(\mathcal{F}_{R,(10,1:N)} + \mathcal{F}_{R,(13,1:N)} + \mathcal{F}_{R,(16,1:N)}). \tag{9}$$

The total directed clustering coefficient per node as derived by [4] may be expressed as the $1 \times N$ vector

$$\mathbf{C}_D = \mathbf{V} \oslash \mathbf{T}, \tag{10}$$

where \oslash indicates Hadamard division. In any instance where division by zero occurs, we set the corresponding term of the result vector to zero. Because \mathbf{V} cannot be written in terms of functional motif fingerprints (since roles are integral to the definition of the various directed clustering coefficients), it is clear that finding specific functional motif-roles is a necessary step in finding the directed clustering coefficient. The global mean directed clustering coefficient is

$$\bar{C}_D = \mathbf{C}_D \mathbf{1} / N_0 \tag{11}$$

$$= (\mathbf{V} \oslash \mathbf{T}) \mathbf{1} / N_0, \tag{12}$$

where N_0 is the count of all nodes i for which $T_i \neq 0$.

The transitivity of an undirected network is defined as the ratio of the total number of three-node subgraphs with three edges, to one third of the total number of pairs of edges that share a node [7]. Consequently, transitivity

measures the fraction of potential closed ‘triangles’ in a network that actually do form closed triangles.

Generalization to a definition of *directed transitivity* was given by [8]. This can be re-expressed in terms of elements from the functional motif-role matrices as

$$T_D = (\mathbf{V1})/(\mathbf{T1}), \tag{13}$$

or, unlike \bar{C}_D , in terms of functional motif counts as

$$T_D = \frac{3(F_5 + F_9)}{2(F_1 + F_2 + F_4)}. \tag{14}$$

In addition to the total directed clustering coefficient, [4] also described four sub-types of directed clustering coefficient, both on a per-node basis and as a global network average (see also [6]). These arise from the four motif-roles that exist within the two closed-triangle motifs with no reciprocal edges, i.e. motifs 5 and 9. In [4] these four types are referred to as ‘in’, ‘out’, ‘middleman’ and ‘cycle’. Here we express these subtype clustering coefficients in terms of motif-role fingerprint vectors as

$$C_{2-sink} = 0.5\mathcal{F}_{R,(11,1:N)} \oslash \mathcal{F}_{R,(10,1:N)}, \tag{15}$$

$$C_{2-source} = 0.5\mathcal{F}_{R,(14,1:N)} \oslash \mathcal{F}_{R,(13,1:N)}, \tag{16}$$

$$C_{Relay} = \mathcal{F}_{R,(17,1:N)} \oslash \mathcal{F}_{R,(16,1:N)}, \tag{17}$$

$$C_{Cycle} = \mathcal{F}_{R,(18,1:N)} \oslash \mathcal{F}_{R,(16,1:N)}. \tag{18}$$

The factors of 0.5 arise from the two possible edges that can be added to motif-roles 10 and 13 to form closed directed feed-forward triangles.

We note that a comparison of the relative abundance of specific functional motif-role fingerprints for nodes of a given degree, with those in an in- or out-degree-preserving null-hypothesis network is equivalent to a comparison between elements of \mathbf{C} vectors in the two networks. This is because a degree-preserving null-hypothesis network ensures that counts of motif-roles 10, 13 and 16 do not change. On the other hand, the utility of per-node clustering coefficients is that normalisation enables comparisons between nodes with different degrees within the original network. The situation is different for structural motif-roles; a null-hypothesis network will not have the same counts of structural motif-roles 10, 13 and 16 as the original network, which suggests there is possible utility in defining directed *structural* clustering coefficients, as alternatives to those of [4].

This discussion also suggests that additional sub-type directed clustering coefficients could be of interest. For example, the *3-feedforward clustering*

coefficient:

$$\mathbf{C}_{3FF} = (\mathbf{V} - \mathcal{F}_{R,(18,1:N)}) \odot (\mathbf{T} - \mathcal{F}_{R,(16,1:N)}). \quad (19)$$

The global mean directed clustering coefficients are trivially obtained in the same way as the global mean directed clustering coefficient, i.e,

$$\bar{C}_{\text{subtype}} = \mathbf{C}_{\text{subtype}} \mathbf{1} / N_{\text{subtype},0}. \quad (20)$$

The different subtypes of clustering coefficient introduced by [4] suggest analogous forms of directed transitivity:

$$\begin{aligned} T_{2\text{-sink}} &= 0.5 \mathcal{F}_{R,(11,1:N)} \mathbf{1} / \mathcal{F}_{R,(10,1:N)} \mathbf{1} \\ &= 0.5 F_5 / F_4 \end{aligned} \quad (21)$$

$$\begin{aligned} T_{2\text{-source}} &= 0.5 \mathcal{F}_{R,(14,1:N)} \mathbf{1} / \mathcal{F}_{R,(13,1:N)} \mathbf{1} \\ &= 0.5 F_5 / F_1 \end{aligned} \quad (22)$$

$$\begin{aligned} T_{\text{Relay}} &= \mathcal{F}_{R,(17,1:N)} \mathbf{1} / \mathcal{F}_{R,(16,1:N)} \mathbf{1} \\ &= F_5 / F_2 \end{aligned} \quad (23)$$

$$\begin{aligned} T_{\text{Cycle}} &= \mathcal{F}_{R,(18,1:N)} \mathbf{1} / \mathcal{F}_{R,(16,1:N)} \mathbf{1} \\ &= 3 F_9 / F_2 \end{aligned} \quad (24)$$

$$\begin{aligned} T_{3FF} &= (\mathbf{V} - \mathcal{F}_{R,(18,1:N)}) \mathbf{1} / (\mathbf{T} - \mathcal{F}_{R,(16,1:N)}) \mathbf{1} \\ &= 3 F_5 / (2 F_1 + F_2 + 2 F_4). \end{aligned} \quad (25)$$

In the first of the two equations for *cycle transitivity*, we have been able to arbitrarily choose one of the three roles for motif 2 in the denominator, since when summed over all N , the results are identical for all three roles. The last expression, for *3-feedforward transitivity*, quantifies the total fraction of possible non-cyclic directed closed triangles that exist in a network.

Remarks on Undirected Networks

The transitivity of a directed network without regard to the direction of the edges could potentially be of interest. Given that S_i is the number of structural motif counts of type $m=i$, let $C_U = S_5 + S_6 + S_9 + S_{10} + S_{11} + S_{12} + S_{13}$. The undirected transitivity can be written as

$$T_U = \frac{C_U}{C_U + \frac{1}{3}(S_1 + S_2 + S_3 + S_4 + S_7 + S_8)}, \quad (26)$$

where C_U is the total number of closed triangles in a network written in terms of structural motifs counts. This result is equivalent to that of the standard definition of transitivity for an undirected network [7, 33], if the directed adjacency matrix was converted to undirected.

Examples: Application to analysis of the *C. elegans* neuronal connectome

As an example application, we calculated the structural and functional motif-role fingerprints for the *C. elegans* hermaphrodite and male neuronal networks. The results are shown in Table 1, which enumerates the motif role fingerprints for neuron AVAR in the hermaphrodite.

As mentioned, it is straightforward to derive the global subgraph ratio metrics (i.e., average directed clustering coefficients and directed transivities) from motif-role fingerprints, as indicated in the bottom right part of Fig. 3. As described above, consideration of motif-role fingerprints led us to define six directed transivities and six directed average clustering coefficients.

Fig. 5 compares each of these transivities and clustering coefficients for the two *C. elegans* neuronal networks, with those that result from in and out degree-preserving randomization of the *C. elegans* connectivity matrix. In each case, 20 randomized networks were created (we found that this was many more than were necessary to obtain consistent and significant changes in all metrics), and their transivities and average clustering coefficients are plotted. Our value of 0.22 for the directed clustering coefficient of the source role ($C_{2-source}$) is consistent with result published in [17], as is our value for the corresponding randomized network of 0.076) but none of the other directed clustering coefficients were mentioned in [17].

We observe that the *C. elegans* hermaphrodite chemical synapse network is between 1.2 and 3.3 times more transitive or clustered (depending on the specific metric) than degree-preserving randomizations of the network (ratios were calculated with respect to the mean of the statistics for all network randomizations). This result is consistent with previous evaluations of clustering coefficient for this network (e.g., [17]). It is also clear, however, that among all the metrics, cycles have the smallest ratio, for both directed transitivity and average directed clustering coefficient. This is also consistent with prior analysis, such as that obtained via standard directed motif analysis — see Figure 7 in [17]. We also found that the male has higher ratios than the hermaphrodite, ranging from 2.1 to 3.9 times more transitive or clustered than the corresponding null hypothesis networks.

It is potentially of interest (both for *C. elegans*, and any other neuronal network data) to consider whether functional significance can be inferred from this form of

Table 1. Example of structural and functional motif-role fingerprints: Neuron AVAR (node id 56) in the *C. elegans* neuronal network.

Role, <i>i</i>	$S_R(i,56)$	$\mathcal{F}_R(i,56)$
1	127	603
2	157	593
3	36	105
4	57	337
5	39	493
6	18	63
7	54	227
8	124	291
9	3	39
10	561	1176
11	62	254
12	7	31
13	615	1176
14	14	172
15	1	13
16	1258	2388
17	27	285
18	8	90
19	3	39
20	352	624
21	9	71
22	94	156
23	13	35
24	362	624
25	98	148
26	7	57
27	1	23
28	40	78
29	27	49
30	11	11

doi:10.1371/journal.pone.0114503.t001

analysis of directed clustering coefficient and transivities. We expect, however, that analysis of motif-role fingerprints will likely be more revealing.

Next, we aim to identify particular network nodes that participate in an overabundance of some specific role, compared to a randomized network.

A simple example that illustrates the utility of obtaining motif-role-fingerprints is as follows. For the *C. elegans* hermaphrodite, we obtained 20 randomized networks, and identified the individual neuron that participated in the greatest number of each of the 30 roles, above the mean obtained in the randomized network. For many of the roles, we observed that the highly ranked neuron according to this metric had a high in and/or out degree. So next, we scaled by the

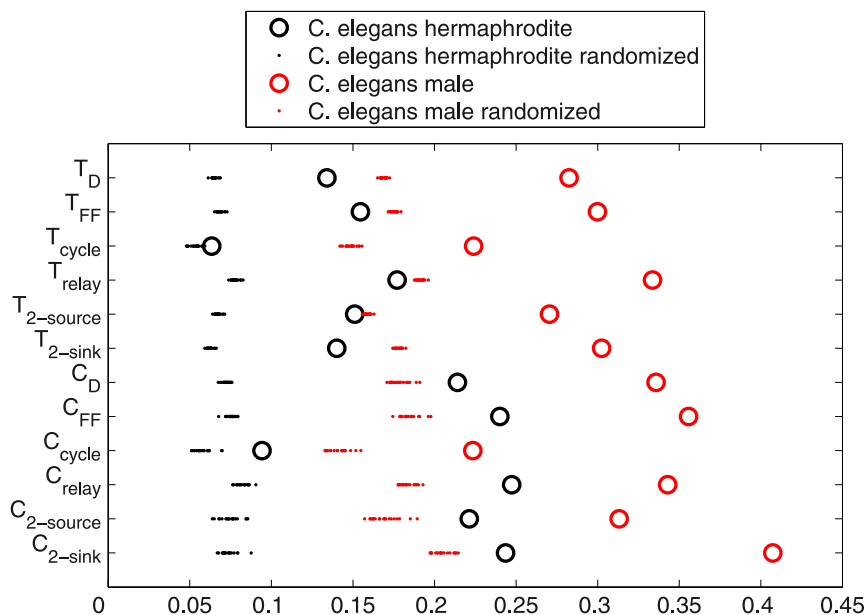


Fig. 5. Directed transitivities and average clustering coefficients for two directed *C. elegans* chemical synapse networks, and randomisations of those networks. Circles show each of the six directed transitivities and six directed clustering coefficient values for the *C. elegans* hermaphrodite and male networks. Dots show comparison points obtained from each of 20 degree-preserving randomizations of the two connectivity matrices. Clearly the male exhibits higher transitivity and clustering than the hermaphrodite, according to all 12 statistics, but both real networks are more transitive/clustered than corresponding null-hypothesis networks.

doi:10.1371/journal.pone.0114503.g005

total degree (i.e., in plus out degree) of each neuron, and examined the neurons with the highest ratios.

In this manner, we observed that neuron RIAL participates in 234 separate instances of functional motif-role 20, whereas in the corresponding randomized networks, RIAL on average participated in 53.3 instances of functional motif-role 20. This can be explained statistically, since RIAL participates in 9 reciprocal edge pairs to and from other neurons, and the our randomization algorithm does not preserve reciprocal degree, only in and out-degree.

A case of a neuron participating in an overabundance of a role that does not include reciprocal edges is that of neuron FLPR, and role 14. In the *C. elegans* network, FLPR participates in 80 instances of functional motif-role 14. The mean number of participations in the randomized networks, however, is only 14.75. Since motif-role 14 involves two outward edges from the reference node, and an edge between the two destination nodes, the motif-role analysis suggests that a role of neuron FLPR is to influence pairs of nodes that are themselves connected.

These few examples illustrate one of the potential applications for motif-role fingerprints: to identify interesting or anomalous nodes within a directed network so that further analysis or experimentation can be carried out on that node or its neighbors.

Future Extensions and Applications

In order to account for heterogeneity in network structure and node types, we have derived mathematical relationships that we expect to be useful when motif distributions need to be characterised, either structurally or functionally, on a node-participation basis, rather than relative to the entire network. We have demonstrated that a hierarchy of relevant metrics exist, with summary metrics such as transitivity derived from richer and more informative vector statistics. The dependencies between each metric discussed are summarized in [Fig. 3](#). We now discuss some anticipated applications and extensions of this work.

Analysis of Neuronal Connectome Data and Synaptic Polarities

Although the neuronal network of the nematode worm, *C. elegans*, is the only complete neuronal network obtained to date [17], network analysis will soon be required for the large neuronal network data sets that result from new experimental techniques currently under rapid development [28, 34]. Indeed, new methods have already resulted in a second partial neuronal network for the *C. elegans* male [19], and we used resulting network data in this paper.

In previous work on motifs applied to neuronal networks, it was observed that combining topological data with data on the functional role of neurons in *C. elegans* (sensory, motor or interneuron) allows a richer analysis of motif distributions with greater relevance to understanding than does describing structural motifs alone [18]. Both the work of [18], and the analysis of motifs in [16, 17, 22], however, characterized the hermaphrodite *C. elegans* neuronal network only in terms of overall abundance of each kind of motif, and did not study the number of motifs of each kind in which individual neurons participate. This is also the case for the analysis of the male posterior neuronal network reported by [19]. One possible direction is to use motif-roles to quantify the centrality of particular neurons within a network, such as by extending the work of [35] to take roles into account.

We anticipate that sophisticated analyses of directed complex neuronal network in future will make use of node-referenced role information, such as that provided by motif-role fingerprints discussed in this paper. Analysis of topological roles in neuronal connectome data could also be supplemented by physiological information, such as the polarity (excitatory or inhibitory) of synapses [30]. This could be modelled as signed edges, and motif-roles generalised to *Signed-motif-roles*.

Subgraphs with More Than Three Nodes

We note that the concept of motif-role fingerprints, either functional or structural, can be extended to arbitrary numbers of nodes per subgraph. For motifs with more than three nodes, however, the number of motif-role types becomes very large, which means that obtaining expressions for each element of \mathcal{F}_R is more difficult. For example, it is known that for four-node subgraphs, there are 199 different connected directed subgraphs. We have not counted how many unique roles there are within each of these, but obviously there are at most a total

of 4×199 motif-roles for 4-node subgraphs. Calculation of $\mathcal{M}_{S_R \rightarrow F_R}$ would also be tedious. Still, it need only be carried out once.

Although we leave this calculation for future work, we note that if this matrix was unknown, but alternative methods for finding both functional and structural motif-role fingerprint counts were available, then $\mathcal{M}_{S_R \rightarrow F_R}$ can readily be derived empirically using data from random directed networks. We have used this method to obtain the matrix $\mathcal{M}_{S \rightarrow F}$ (and its inverse) for the case of 4-node global motifs. This was achieved using the Matlab software package known as the *Brain Connectivity Toolbox*, made available in association with [8], which provides code for obtaining global functional and structural motif counts for 4-node connected subgraphs.

Extension to Weighted Network Edges

The definition of motifs (in the global sense) has previously been extended to incorporate information about edge weights [36]. The resulting metric was referred to as *subgraph intensity*. It is potentially useful to extend this idea to motif-roles, and perhaps it will be as simple as replacing the binary adjacency matrix \mathcal{A} with a weighted adjacency matrix \mathcal{W} in the equations shown in Fig. 2. However, we leave consideration of this possibility for future work.

Possible use in role detection and detection of similar nodes and similar networks

There has been recent interest in automatic discovery of network roles, and nodes that are structurally similar, and algorithms have been developed for achieving this [37]. The methods described in [37] are flexible in the sense that many different network statistics can be provided as inputs from which roles are identified. There is strong potential for including motif-role fingerprints as a subset of the network statistics used in such algorithms. If, in the future, many large connectome datasets become available, it may be potentially interesting to assess the resulting networks for overall similarity, or to search for similar nodes within or across networks.

Methods

Source code

Matlab code implementing the results of this paper is provided as [S1 Matlab File](#).

Notation for Functional Motif-Role Fingerprints in Fig. 2

For a network with N nodes, we denote the $N \times N$ binary directed adjacency matrix as \mathcal{A} ($\mathcal{A}_{i,j} \in \{0,1\} \forall i,j$). We assume that $\text{diag}(\mathcal{A}) = 0$, i.e there are no self-connections.

In the formulae listed in Fig. 2, we make use of the matrix $\mathcal{R} := \mathcal{A} \circ \mathcal{A}^T$, which is a binary matrix where each 1 indicates a reciprocal edge between two nodes.

Converting structural to functional motif fingerprints and *vice-versa*

The following matrix enables conversion from structural motif fingerprints, \mathcal{S} or structural motif counts, \mathbf{S} to functional motif fingerprints, \mathcal{F} or functional motif counts \mathbf{F} , as expressed in [Equations \(5\)](#) and (7) respectively.

$$\mathcal{M}_{\mathcal{S} \rightarrow \mathcal{F}} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 2 & 0 & 1 & 0 & 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 0 & 1 & 2 & 1 & 2 & 3 & 3 & 2 & 4 & 6 \\ 0 & 0 & 1 & 0 & 0 & 2 & 0 & 2 & 0 & 1 & 0 & 3 & 6 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 2 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (31)$$

The following matrix is the inverse of $\mathcal{M}_{\mathcal{S} \rightarrow \mathcal{F}}$, and can be used to convert from functional motif fingerprints, \mathcal{F} , or functional motif counts, \mathbf{F} to structural motif fingerprints, \mathcal{S} or structural motif counts, \mathbf{S} .

$$\mathcal{M}_{\mathcal{S} \rightarrow \mathcal{F}}^{-1} = \begin{pmatrix} 1 & 0 & -1 & 0 & -1 & 2 & 0 & 1 & 0 & 1 & 1 & -2 & 3 \\ 0 & 1 & -1 & 0 & -1 & 2 & -1 & 2 & -3 & 3 & 2 & -4 & 6 \\ 0 & 0 & 1 & 0 & 0 & -2 & 0 & -2 & 0 & -1 & 0 & 3 & -6 \\ 0 & 0 & 0 & 1 & -1 & 1 & -1 & 1 & 0 & 1 & 2 & -2 & 3 \\ 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 & -1 & -2 & 3 & -6 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & -1 & -2 & 3 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (32)$$

Deriving motif counts from motif-role fingerprints

Given that each motif is comprised from three motif-roles, deriving the motif counts from the motif-role counts, or *vice-versa* is trivial. To make this relationship explicit, we introduce the following 13×3 matrix composed from the elements of \mathbf{F}_R (denoted as $F_{R,k}$, $k=1, \dots, 30$) to explicitly denote which functional roles are associated with which functional motifs:

$$\mathcal{F}_{MR} = \begin{pmatrix} F_{R,1} & F_{R,2} & 0 \\ F_{R,3} & F_{R,4} & F_{R,5} \\ F_{R,6} & F_{R,7} & F_{R,8} \\ F_{R,9} & F_{R,10} & 0 \\ F_{R,11} & F_{R,12} & F_{R,13} \\ F_{R,14} & F_{R,15} & 0 \\ F_{R,16} & F_{R,17} & F_{R,18} \\ F_{R,19} & F_{R,20} & 0 \\ F_{R,21} & 0 & 0 \\ F_{R,22} & F_{R,23} & F_{R,24} \\ F_{R,25} & F_{R,26} & 0 \\ F_{R,27} & F_{R,28} & F_{R,29} \\ F_{R,30} & 0 & 0 \end{pmatrix} \quad (33)$$

The i -th row in \mathcal{F}_{MR} indicates motif $i, i=1, \dots, 13$. A zero appears for any motif in which more than one node plays the same role. Where \mathcal{F}_{MR} has three non-zero elements, they all have the same value, which is equal to the total number of functional motifs corresponding to that row. Where it has two elements, one element is twice the other, where the element multiplied by 2 is that indicated by $d_r = 2$ in Fig. 2. Similarly, where there is one element, it is multiplied by 3 as indicated by $d_r = 3$ in Fig. 2.

We also introduce \mathcal{S}_{MR} to denote the equivalent matrix for structural motifs. The total count of structural or functional motifs in a network can be trivially obtained from \mathcal{S}_{MR} and \mathcal{F}_{MR} respectively by

$$\mathbf{S}^\top = \frac{1}{3} \mathcal{S}_{MR} \mathbf{1}_3 \quad (34)$$

$$\mathbf{F}^\top = \frac{1}{3} \mathcal{F}_{MR} \mathbf{1}_3, \quad (35)$$

where $\mathbf{1}_3$ is a unit 3×1 column matrix.

Conversely, the vectors \mathbf{F}_R and \mathbf{S}_R can be trivially obtained from \mathbf{F} and \mathbf{S} respectively, since we also have

$$\mathcal{F}_{MR} = \begin{pmatrix} 2F_1 & F_1 & 0 \\ F_2 & F_2 & F_2 \\ F_3 & F_3 & F_3 \\ 2F_4 & F_4 & 0 \\ F_5 & F_5 & F_5 \\ F_6 & 2F_6 & 0 \\ F_7 & F_7 & F_7 \\ 2F_8 & F_8 & 0 \\ 3F_9 & 0 & 0 \\ F_{10} & F_{10} & F_{10} \\ F_{11} & 2F_{11} & 0 \\ F_{12} & F_{12} & F_{12} \\ 3F_{13} & 0 & 0 \end{pmatrix}. \quad (36)$$

Network data for *C. elegans* neuronal connectomes

For the hermaphrodite, we used network adjacency matrix data, based on chemical synapses, made publicly available in conjunction with [17]. For the male, we used network adjacency matrix data, based on chemical synapses, made publicly available in conjunction with [19].

Supporting Information

S1 Matlab File. Matlab code implementing the results of this paper.

[doi:10.1371/journal.pone.0114503.s001](https://doi.org/10.1371/journal.pone.0114503.s001) (M)

Author Contributions

Wrote the paper: MDM ONY BAS NI LMW. Derived the mathematical expressions: MDM BAS. Analysis and interpretation of mathematical results and applications: MDM ONY BAS NI LMW.

References

1. Bassett DS, Greenfield DL, Meyer-Lindenberg A, Weinberger DR, Moore SW, et al. (2010) Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Computational Biology* 6: e1000748.
2. McDonnell MD, Ward LM (2014) Small modifications to network topology can induce stochastic bistable spiking dynamics in a balanced cortical model. *PLoS One* 9: e88254.

3. **Cook S** (1971) The complexity of theorem-proving procedures. In: Proceedings of the Third annual ACM symposium on Theory of Computing. pp. 151–158.
4. **Fagiolo G** (2007) Clustering in complex directed networks. *Physical Review E* 76: 026107.
5. **Watts DJ, Strogatz SH** (1998) Collective dynamics of 'small-world' networks. *Nature* 393: 440–442.
6. **Ahnert SE, Fink TMA** (2008) Clustering signatures classify directed networks. *Physical Review E* 78: 036112.
7. **Newman MEJ** (2003) The structure and function of complex networks. *SIAM Review* 45: 167–256.
8. **Rubinov M, Sporns O** (2010) Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* 52: 1059–1069.
9. **Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al.** (2002) Network motifs: Simple building blocks of complex networks. *Science* 298: 824–827.
10. **Pržulj N, Corneil D, Jurisica I** (2004) Modeling interactome: Scale-free or geometric? *Bioinformatics* 20: 3508–3515.
11. **Itzkovitz S, Milo R, Kashtan N, Ziv G, Alon U** (2003) Subgraphs in random networks. *Physical Review E* 68: 026127.
12. **Itzkovitz S, Alon U** (2005) Subgraphs and network motifs in geometric networks. *Physical Review E* 71: 026117.
13. **Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU** (2006) Complex networks: Structure and dynamics. *Physics Reports* 424: 175–308.
14. **Itzkovitz S, Milo R, Kashtan N, Newman MEJ, Alon U** (2004) Reply to "Comment on 'Subgraphs in random networks' ". *Physical Review E* 68: 058102.
15. **Guimerà R, Sales-Pardo M, Amaral LAN** (2007) Classes of complex networks defined by role-to-role connectivity profiles. *Nature Physics* 3: 63–69.
16. **Reigl M, Alon U, Chklovskii DB** (2004) Search for computational modules in the *c. elegans* brain. *BMC Biology* 2: 25 (1–12).
17. **Varshney LR, Chen BL, Paniagua E, Hall DH, Chklovskii DB** (2011) Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Computational Biology* 7: e1001066.
18. **Qian J, Hintze A, Adami C** (2011) Colored motifs reveal computational building blocks in the *c. elegans* brain. *PLoS One* 6: e17013.
19. **Jarrell TA, Wang Y, Bloniarz AE, Brittin CA, Xu M, et al.** (2012) The connectome of a decision-making neural network. *Science* 337: 437–444.
20. **Towison EK, Vértes PE, Ahnert SE, Schafer WR, Bullmore ET** (2013) The rich club of the *C. elegans* neuronal connectome. *The Journal of Neuroscience* 33: 6380–6387.
21. **Yaveroğlu O, Malod-Dognin N, Davis D, Levnajic Z, Janjic V, et al.** (2014) Revealing the hidden language of complex networks. *Scientific Reports* 4: 4547.
22. **Sporns O, Kötter R** (2004) Motifs in brain networks. *PLoS Biology* 2: e369.
23. **Kashtan N, Itzkovitz S, Milo R, Alon U** (2004) Topological generalizations of network motifs. *Physical Review E* 70: 031909.
24. **Koschützki D, Schwöbbermeyer H, Schreiber F** (2007) Ranking of network elements based on functional substructures. *Journal of Theoretical Biology* 248: 471–479.
25. **Pržulj N** (2007) Biological network comparison using graphlet degree distribution. *Bioinformatics* 23: e177–e183.
26. **Guerrero C, Milenković T, Pržulj N, Kaiser P, Huang L** (2008) Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. *Proceedings of the National Academy of Sciences of the USA* 105: 13333–13338.
27. **Kuchaiev O, Pržulj N** (2011) Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics* 27: 1390–1396.
28. **Seung HS** (2009) Reading the book of memory: Sparse sampling versus dense mapping of connectomes. *Neuron* 62: 17–29.

29. **Prettejohn BJ, Berryman MJ, McDonnell MD** (2011) Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists. *Frontiers in Computational Neuroscience* 5: 11 (1–18).
30. **Dong CY, Cho KH** (2012) An optimally evolved connective ratio of neural networks that maximizes the occurrence of synchronized bursting behavior. *BMC Systems Biology* 6: 23.
31. **Mayo M, Abdelzaher AF, Perkins EJ, Ghosh P** (2012) Motif participation by genes in *e. coli* transcriptional networks. *Frontiers in Physiology* 3: 357.
32. **Wong E, Baur B, Quader S, Huang CH** (2012) Biological network motif detection: principles and practice. *Briefings in Bioinformatics* 13: 202–215.
33. **Barrat A, Weigt M** (2000) On the properties of small-world network models. *The European Physical Journal B* 13: 547–560.
34. **Marx V** (2012) Charting the brain's networks. *Nature* 490: 293–298.
35. **Wang P, Lu J, Yu X** (2014) Identification of important nodes in directed biological networks: A network motif approach. *PLoS One* 9: e106132.
36. **Onnela JP, Saramäki J, Kertész J, Kaski K** (2005) Intensity and coherence of motifs in weighted complex networks. *Physical Review E* 71: 065103(R).
37. **Henderson K, Gallagher B, Eliassi-Rad T, Tong H, Basu S, et al.** (2012) RoIX: Structural Role Extraction & Mining in Large Graphs. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 12–16, 2012, Beijing, China.
38. **Newman MEJ** (2010) *Networks: An Introduction*. Oxford University Press.