Motifs and scale-free properties in brain connectivity networks across species

Brandon Barton and Lowell Kalman

Colorado School of Mines

Fall 2021

Abstract

Over the last 20 years modern technology has spawned the field of connectomics, allowing the creation of novel data sets mapping brains of various animals. Researchers have thoroughly analyzed these data sets, but many questions about the structure of the brain remain. Here we explore the structure of connectomes from different animals using tools from network science and graph theory. By examining motif occurrence in the brain networks we can learn about the structure and patterns of circuitry common to different animals. The degree distribution is another metric used to learn about the organization of the brain, and how it may share similar properties with other complex networks. Network motif spectra show that local structure in neuronal networks seems to follow simple connection patterns, and scale-free networks may provide a good model for the global trends in neuronal connectivity, giving insight into the structure and organization of the brain.

1 Background

1.1 Complex networks

Using networks as a model of real world systems captures relationships and patterns which are otherwise challenging to represent. However, what makes a system complex, and what scale of the network is useful to investigate in order to uncover what drives complex behaviour? To provide some background, one example of a complex system is a road transportation network. Road networks have many moving pieces (e.g. automobiles, traffic lights, construction zones) that all play a role in a larger behavior, such as traffic. Such emergent phenomena is a key attribute that makes a network complex, which network theory attempts to deconstruct [11]. Another aspect that makes complex network theory distinct from other methods is investigation of the system as a whole comprised of many subsystems [2]. The networks and subsystems of interest in this work are brain networks, and specifically connections between regions of the brain at different scales. This work uses a graph analytic modelling approach to provide insight into some of properties of brain networks at varying scale across multiple species.

1.2 Brain networks

To represent a complex network such as the brain, neuroscience turns to graph and network science. A brain network is represented by a graph $G = (N, E)$ where G is the graph consisting of nodes (N), and edges (E) that represent connections between nodes [1]. Further specification of node relationships can be made, such as as directed (one-way relationship) edges or signed (+/-) edges, which more accurately represent the type of relationship between nodes. While graphs can represent the true connections between individual neurons and their neighbors, these high resolution models are typically restricted to a small region in the brain due to the experimental challenges of generating such detailed data sets. While tremendous progress has been made on representing brain networks at single neuron resolution [10], network representations of larger networks such as human brains are still a significant challenge. However, lower resolution connectivity networks can provide insight into whole-brain connectivity by representing individual brain regions as nodes, and bundles of axonal projections as edges. In this work both single neuron networks and brain region connectivity networks are investigated.

In section 2, background is provided on prior work investigating network motifs and scale-free properties of networks. Following a summary of prior work, choices of brain connectivity networks at various resolutions are discussed in section 3. Graph analysis methods of network motif counting, and degree distribution analysis are described in section 3 as well. Following experimental methods used in this work, the results and original contributions of this work are presented in section 4. Finally, a discussion of the implications of our findings are discussed in section 5, followed by a discussion on future directions for research.

2 Related Work

2.1 Network Motifs

Given a host network $H = (N, E)$, a network motif is a subgraph $G \in H$. An example of a network motif query is provided in figure 1.

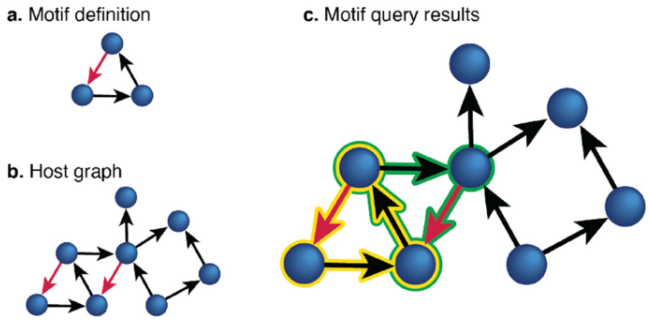

Figure 1: Example of a three node motif query on a host graph in DotMotif [7, 8].

Network motifs have been selected in previous works according to known recurrent connectivity patterns in neural circuity networks [9]. In previous work, counting the motifs in the network provides insight into statistically significant structural building blocks of brain networks, which may represent critical signalling or regulation networks [6]. Computing motif distributions in brain networks shows local structure on a neuron-to-neuron or region-to-region basis. Understanding patterns of connectivity in brain networks is critical to understanding how networks form, and the importance of structure in communication wiring pathways [6, 7]. While local structure of brain networks can be investigated using network motifs, analyzing the degree distribution of a network provides insight into patterns of global structure.

2.2 Power Law and Scale-free Networks

Formally, a scale-free network is one where the degree distribution follows a power-law,

$$p_k = Ck^{-\gamma}. \tag{1}$$

In 1 p_k is the probability of a node having k edges attached to it, C is determined by a normalization condition, and γ is the degree exponent of the network. In a log-log representation of p_k, a power-law distribution presents as a straight line. Scale-free networks often arise in quasi-organically evolving networks, such as the World Wide Web, academic paper citations, and protein interactions. One key feature of scale-free networks is the presence of hubs, which are nodes that have large numbers of edges. For particular values of γ, a scale-free network has a significantly shorter average distance

between nodes than an equivalent random network. Short path lengths gives rise to the small-world property for $3 < \gamma$, and the ultra-small-world property for $2 < \gamma < 3$ [1]. Figure S2 shows a comparison between a random and scale-free network, and how their degree distributions differ. Determining that a network is scale-free gives us insight into its structure, and how connections between nodes were created and organized over time.

3 Methods

3.1 Network Selection

3.1.1 Scale

Networks of varying scale were analyzed to investigate motif occurrence and scale-free properties in animals of varying anatomical size. The drosophila (fruit fly), mouse, cat, and human brain networks were chosen specifically to maximize the range of node-to-edge ratios, as well as representing a broad range of anatomical brain network sizes across species.

Network	Nodes (N)	Edges (E)	Density (ρ)	Avg Degree (k)	Source
drosophila	1780	17417	0.006	20	[3]
Mouse	213	21807	0.483	205	[3]
Cat	65	1139	0.274	35	[3]
Macaque	91	628	0.077	14	[3]

Table 1: Network attributes and summary statistics for drosophila, mouse, cat, and macaque networks. Network densities were calculated in NetworkX [5].

3.1.2 Resolution

Differing in scale, the drosophila, mouse, cat, and macaque brain networks were also mapped at varying resolutions. A summary of the network statistics are provided in table 1. The drosophila brain network has been mapped to the level of single neuron resolution, which captures the most accurate information content of neuron-to-neuron connections. The mouse, cat, and macaque brain networks were all abstracted networks representing brain region connectivity. Competing interests of network scale and network resolu-

tion present a challenging trade-of, as smaller scale networks are more easily investigated at higher resolution, while increasing the scale of brain networks compromises network resolution.

3.2 Network Motifs

3.2.1 Generating Motifs

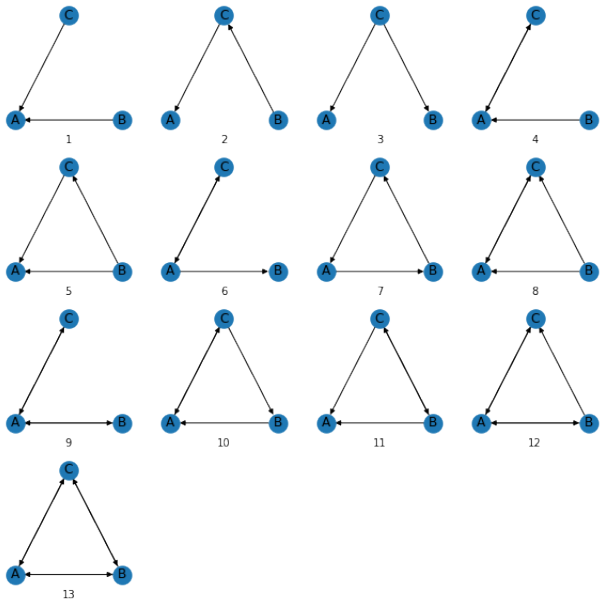

Figure 2: Directed (N = 3) motifs. Figure generated using NetworkX graph drawing capabilities.

For consistency in the directedness of the animal brain networks, three node ($N = 3$) directed, connected motifs were chosen for analysis. As shown in figure 2, these thirteen motifs were generated in NetworkX [5].

3.2.2 Counting Motifs

Motif concentrations across all were calculated using equation 2, where C_i is the concentration for motif i in the list of thirteen three node motifs, N_i is

the frequency of a single motif, and $T(G)$ is the sum of all motif frequencies.

$$C_i = \frac{N_i(G)}{T(G)}, T(G) = \sum_{i}^{13} N_i(G) \tag{2}$$

Motif frquencies were computed using DotMotif, an open source python library for network analysis specifically built for neuroscientists. DotMotif uses a variation of the VF2 subgraph isomorphism algorithm in S1 [4, 8]. While he algorithm implemented in DotMotif is a significant improvement from other existing implementations of graph isomorphism algorithms, it still carries a significant run time. Therefore, for the scope of this work, motif searches were limited to three node and computed across all networks.

3.3 Random Graph Generation

By comparing the animal brain connectivity networks to networks generated using various algorithmic building methods, the traits shared between random graph models provides valuable insight into the similarity of how network structure forms. The Barabási–Albert preferential attachment models were generated using NetworkX random graph generator, both in directed (SF) and undirected (BA) formats. After these graphs were generated, motif counts were performed in the same procedure as on the animal networks.

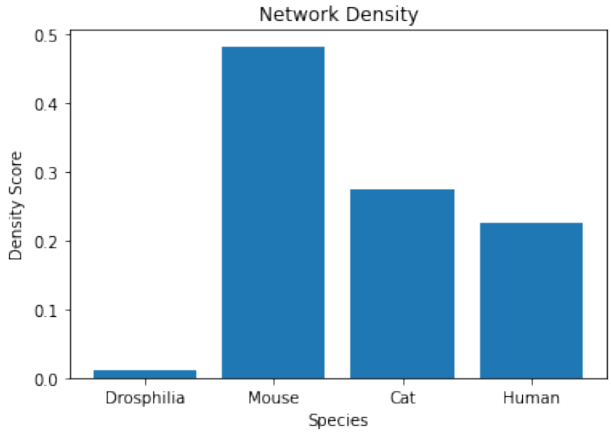

Figure 3: Network density distribution across species.

For a baseline comparison of the animal networks to the generated graphs, network densities were calculated for each network with equation 3, where n

is the number of nodes in the network, and m is the number of edges. As expected, the mouse network displays the highest network density due to the high number of edges relative to the number of nodes in the network, while the fly network has a relatively low network density. Network densities were compared across each network to the BA models.

$$d = \frac{2m}{n(n-1)} \tag{3}$$

4 Results

4.1 Motif Concentration

Motif concentration spectra of each animal brain network were compared to motif spectra of generated SF networks shown in figure 4. While the drosophila, cat, and mouse showed similar distributions of concentrations across the first three unidirectional motifs, the later motifs with additional edge complexity occurred relatively less frequently. Furthermore, while these three networks showed similar spectra to one another, shown in figure S3, each differed significantly from the SF motif spectra.

The macaque network, however, showed significant similarity to the SF motif concentration spectra as in figure 4d. In all SF generated networks, motif 1 showed the highest concentration, while motifs with bidirectional edges were much less prevelant. In the both the real drosophila network and the SF-drosophila network (generated with $N_{drosophila}$), there was disproportionate prevalence of motifs 1 through 3, while any motif with bidirectional edges was relatively insignificant.

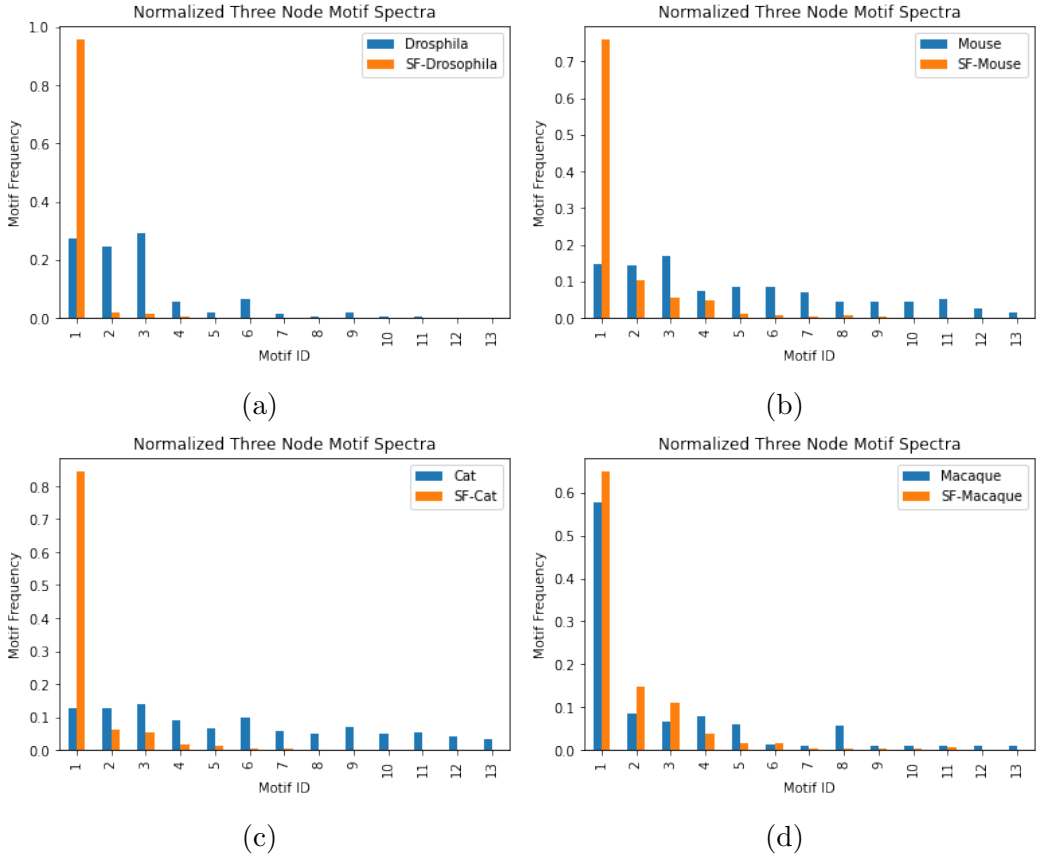

Figure 4: Motif concentrations across network species, compared to SF generated networks (a) drosophila network (b) Mouse network (c) Cat network (d) Macaque network

4.2 Scale-free attributes

The degree distribution for each network was generated using NetworkX. We also generated the degree distributions for each of the Barabasi-Albert generated graphs, in order to see if the structure was a good match for the real networks. For the mouse, macaque, and cat networks the degree distributions did not seem to follow any obvious pattern. The BA stochastic graphs we created also did not match well with the real networks, as seen in figure S4. However, the drosophila network not only showed a power-law degree distribution, but a strong correlation with the BA generated graph modelled

after the drosophila network. The BA model has an R^2 score of 0.73 when compared to the real drosophila network, indicating a strong correlation. Figure 4 shows the distributions for both networks both in a linear-linear plot and a log-log plot.

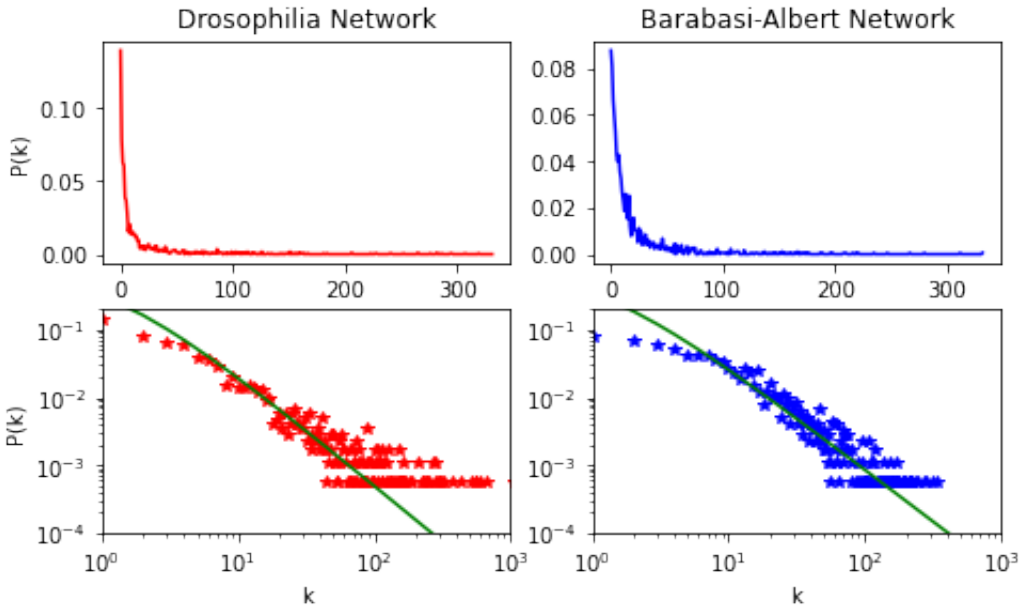

Figure 5: Degree distribution for Drosphilia Medulla. $n = 1781, k = 33641$. Data gathered with Electron Microscopy.

5 Discussion

5.1 Motif spectra comparison

Due to the significant variation among motif spectra among the neruonal networks across species considered compared to the SF generated networks, we can confidently conclude that the local strucutre of the drosophila, mouse, and cat networks all differ signicantly from a scale-free network generated accoriing to the rules of preferential attachment. However, the drosophila network's degree distribution did show strikingly similar results to the SF network. This leads us to the conclusion that the scale free properties of the drosophila network are not captured in the three node motifs considered in

this work, which would be a future avenue of work.

Furthermore, considering the overall trends in motif spectra across all species, high concentrations of unidirectional (1-3) motifs were most prevalent as expected. For single neuron resolution networks such as the Drosophila, bidirectional motif structures, while not completely absent, were significantly less prevalent than simpler motifs. This result leads us to the conclusion that at the smallest scale, neuron-to-neuron networks tend to prefer attachment protocols that are simple. In lower resolution networks like the mouse, cat, and macaque networks, brain region-to-region connection patterns display more complex interactions. Future work may consider known communication neuronal circuitry patterns for known behavior on a brain region-to-region basis, which may uncover the significance of neuronal structure in emergent behavior.

5.2 Scale-free properties across networks

We observe scale-free properties in the drosophila network, but not networks for other species. One explanation for this discrepancy is the differing resolutions of the networks. The drosophila data has a resolution of single neurons, while the other networks provide region-to-region connections. It is possible that higher resolution connectomes for the mouse, cat, and macaque would show scale-free properties as well. The scale-free properties of the drosophila network imply that there is a short average path length between nodes due to the presence of hubs. This means that an impulse travelling between neurons in a drosophila brain will travel through hub neurons that have synapses with many other neurons.

5.3 Future Research

There is certainly still more to be learned from analyzing the wealth of data in these connectomes. Improving algorithms for motif search and increasing computing power can facilitate the search for more complex motifs. Examining motif concentration based on region could expose the function of different circuits in the brain. Exploring more random graph models, such as sub-linear preferential attachment (limits of neuronal attachment), could shed more light on the organization of brain networks. There is also the gargantuan effort of gathering more data on more brains at a higher resolution. Mapping the drosophila brain took thousands of hours of manual

work by researches to identify connections from electron microscope images. It would have been impossible without help from machine learning models helping to identify possible connection sites. Developing even better machine learning models could allow for mapping brains at a level not possible before. Because network science is such a new and interdisciplinary field, as our understanding of complex networks as a whole increases, so does our knowledge of specific networks like the brain.

References

[1] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[2] Danielle S. Bassett and Edward T. Bullmore. Small-World Brain Networks Revisited. *Neuroscientist*, 23(5):499–516, 2017.

[3] Randal Burns, Eric Perlman, Alex Baden, William Gray Roncal, Ben Falk, Vikram Chandrashekhar, Forrest Collman, Sharmishtaa Seshamani, Jesse Patsolic, Kunal Lillaney, Michael Kazhdan, Robert Hider, Derek Pryor, Jordan Matelsky, Timothy Gion, Priya Manavalan, Brock Wester, Mark Chevillet, Eric T. Trautman, Khaled Khairy, Eric Bridgeford, Dean M. Kleissas, Daniel J. Tward, Ailey K. Crow, Matthew A. Wright, Michael I. Miller, Stephen J. Smith, R. Jacob Vogelstein, Karl Deisseroth, and Joshua T. Vogelstein. A community-developed open-source computational ecosystem for big neuro data. *arXiv*, 15(November):846–847, 2018.

[4] Luigi P. Cordella, Pasquale Foggia, Carlo Sansone, and Mario Vento. A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1367–1372, 2004.

[5] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[6] Marcus Märtens, Jil Meier, Arjan Hillebrand, Prejaas Tewarie, and Piet Van Mieghem. Brain network clustering with information flow motifs. *Applied Network Science*, 2(1), 2017.

[7] Jordan K. Matelsky, Elizabeth P. Reilly, Erik C. Johnson, Jennifer Stiso, Danielle S. Bassett, Brock A. Wester, and William Gray-Roncal. DotMotif: an open-source tool for connectome subgraph isomorphism search and graph queries. *Scientific Reports*, 11(1):1–14, 2021.

[8] Jordan K. Matelsky, Elizabeth P. Reilly, Erik C. Johnson, Brock A. Wester, and William Gray-Roncal. Connectome subgraph isomorphisms and graph queries with DotMotif. *bioRxiv*, 2020.

[9] Olaf Sporns and Rolf Kötter. Motifs in Brain Networks. *PLoS Biology*, 2(11), 2004.

[10] Martijn P. van den Heuvel, Edward T. Bullmore, and Olaf Sporns. Comparative Connectomics. *Trends in Cognitive Sciences*, 20(5):345–361, 2016.

[11] Ömer Nebil Yaveroäa Lu, Noël Malod-Dognin, Darren Davis, Zoran Levnajic, Vuk Janjic, Rasa Karapandza, Aleksandar Stojmirovic, and Nataša Pržulj. Revealing the hidden Language of complex networks. *Scientific Reports*, 4:1–9, 2014.

6 Supplementary Information

Algorithm 1: Our approach: Subgraph search

Input : A small motif graph M
Input : An arbitrarily large host graph H
Initialize an empty results set R;
Initialize an empty task queue Q;
Add the empty candidate mapping ($\{\}$) to Q;
while Q *is non-empty* **do**
    Pop a new candidate B from Q;
    Identify m_1 the most interesting node in motif M that does not yet have a mapping assigned in B;
    Identify all nodes that are valid mappings from the candidate to m_1 (based upon degree, attributes, etc.);
    if *multiple nodes are valid candidates* **then**
        Add each new candidate to Q;
    else
        (All nodes in M have a valid mapping in B to H);
        Add the mapping to the results set R;
    end
end
Result: The set R of resultant mappings from nodes in M to IDs in H

Figure S1: Algorithm implemented in DotMotif for subgraph isomorphism [8]. Developed for additional specifications to return subgraph monomorphism, or count-only simplification. This unnamed algorithm has shown significant speed-ups compared to NetworkX's implementation of the VF2 subgraph isomorphism algorithm.

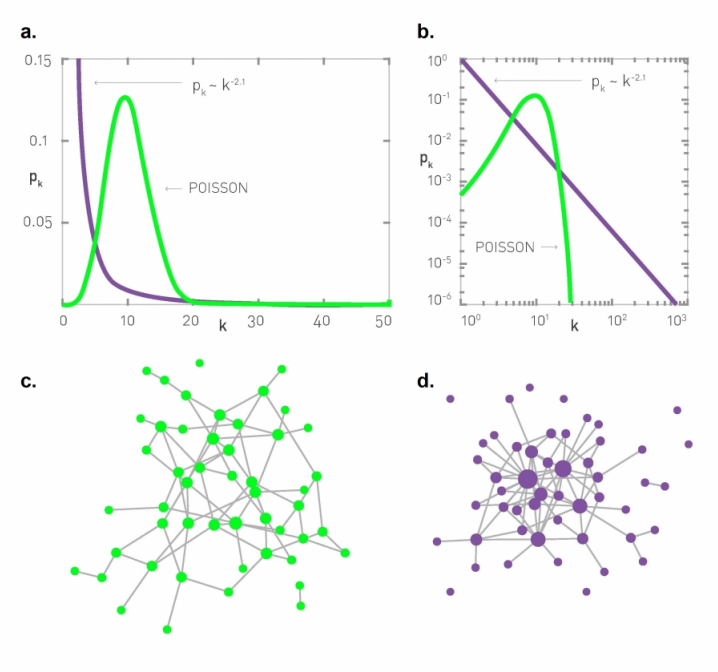

Figure S2: Degree Distribution Comparison of a Random vs. Scale-Free Network [1]

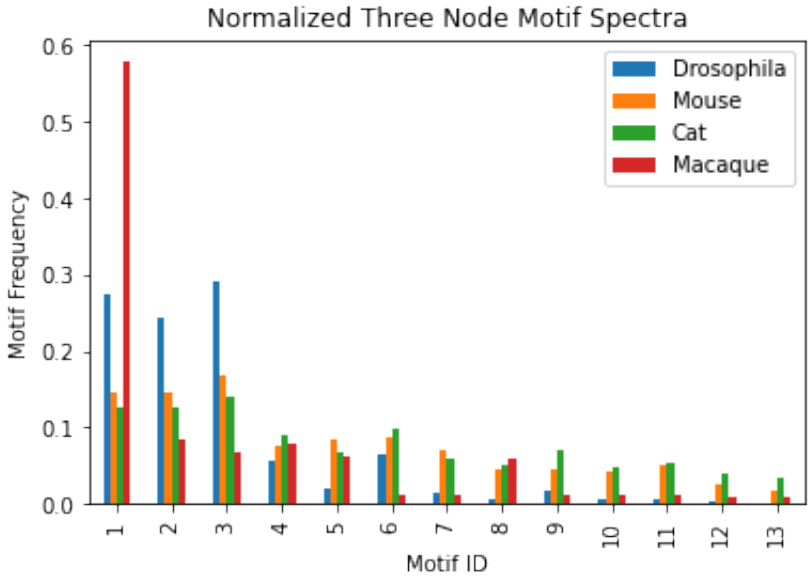

Figure S3: Network motif concentrations compared across species.

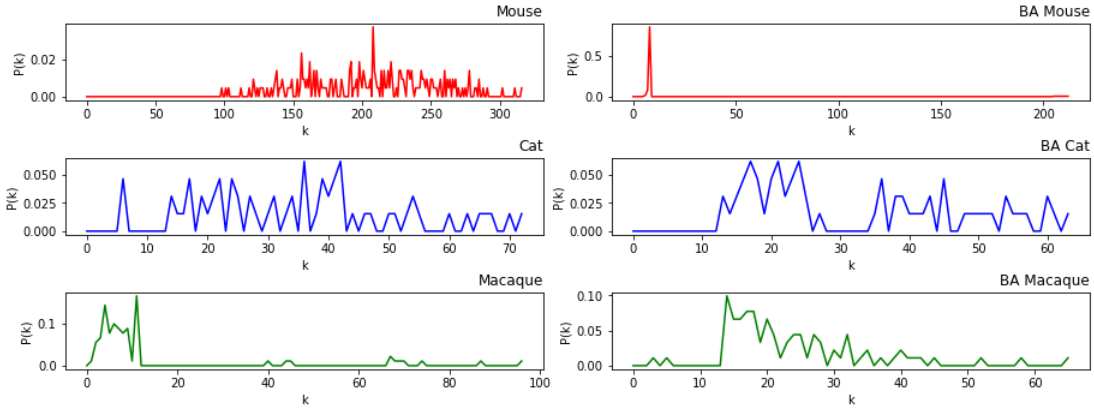

Figure S4: Degree distribution for mouse, cat, and human brain connectomes (Left) and BA stochastic graphs (Right).