

Evidence of Rentian Scaling of Functional Modules in Diverse Biological Networks

Javier J. How

jhow@ucsd.edu

Saket Navlakha

navlakha@salk.edu

*Salk Institute for Biological Studies, Integrative Biology Laboratory,
La Jolla, CA 92037, U.S.A.*

Biological networks have long been known to be modular, containing sets of nodes that are highly connected internally. Less emphasis, however, has been placed on understanding how intermodule connections are distributed within a network. Here, we borrow ideas from engineered circuit design and study Rentian scaling, which states that the number of external connections between nodes in different modules is related to the number of nodes inside the modules by a power-law relationship. We tested this property in a broad class of molecular networks, including protein interaction networks for six species and gene regulatory networks for 41 human and 25 mouse cell types. Using evolutionarily defined modules corresponding to known biological processes in the cell, we found that all networks displayed Rentian scaling with a broad range of exponents. We also found evidence for Rentian scaling in functional modules in the *Caenorhabditis elegans* neural network, but, interestingly, not in three different social networks, suggesting that this property does not inevitably emerge. To understand how such scaling may have arisen evolutionarily, we derived a new graph model that can generate Rentian networks given a target Rent exponent and a module decomposition as inputs. Overall, our work uncovers a new principle shared by engineered circuits and biological networks.

1 Introduction ---

One approach to understand how biological networks are organized is to study their scaling properties. Developmental constraints often force evolution to conserve certain properties of networks as they change in size. By relating these properties to function, we can uncover new structure-function relationships. Here, we describe a new scaling relationship present in functional modules in diverse biological networks.

Molecular interaction networks are highly modular, containing subsets of nodes (genes or proteins) that interact more strongly with each other

compared to the rest of the network (Spirin & Mirny, 2003; Sharan, Ulitsky, & Shamir, 2007; Navlakha, Schatz, & Kingsford, 2009; Jiang & Singh, 2010; Davis, Yaveroglu, Malod-Dognin, Stojmirovic, & Przulj, 2015). These modules represent core biological processes that occur in the cell, such as RNA metabolism, protein methylation, or autophagy (Ashburner et al., 2000). Although less studied, there are also many interactions between proteins in different modules (Pinkert, Schultz, & Reichardt, 2010), such as pleiotropic genes that have multiple functions and belong to multiple modules. These together suggest there is substantial cross-talk between biological modules, including across modules that are seemingly unrelated (Stearns, 2010). Here, our goal is to uncover structure in this intermodule communication using theories developed from engineered circuit design.

E. F. Rent, a scientist at IBM in the 1960s, studied the structure of computer circuits and found that when he plotted the number of gates (nodes) in a logical block (module) of the circuit versus the number of connections to or from that block, the two scaled according to a power law (Landman & Russo, 1971; Lanzerotti, Fiorenza, & Rand, 2005). This relationship, now known as Rent's rule, was preserved across many spatial scales, from very small to very large modules, indicating that interconnections can be described by a self-similar function (Stevens, 2009). The rule has also been shown to hold over time as circuits evolved to become larger (more on-board components), more energy efficient, and higher performing (Bakoglu, 1990; Lanzerotti et al., 2005; Bassett et al., 2010).

The exponent of the power law relationship is called Rent's exponent. Higher exponents denote random, complex wiring, though with greater logical capacity; smaller values denote less capacity but a more economical design that prefers short-versus long-range connections (Christie & Stroobandt, 2000). Thus, the exponent encodes a trade-off between the economy of design and performance complexity. In digital circuits, the exponent is important because it can be used to predict the total amount of wiring needed to realize a specific design topology (Christie & Stroobandt, 2000), and thus how efficiently the network is embedded in space (Bullmore & Sporns, 2012). It can also be used to estimate various chip layout parameters (Christie & Stroobandt, 2000). Proteins in molecular networks do not have fixed positions in space, and thus there is no exact analog of "wiring length"; however, proteins are largely partitioned into physical cellular compartments, and thus Rent's exponent may provide insight into the efficiency of cross-compartment communication. Remarkably, circuit design engineers at IBM did not explicitly try to optimize for Rent's rule; it emerged naturally from the human design process. In molecular networks, a certain amount of independence of modules is clearly desired, yet the diameters of most molecular networks and signaling pathways are rather small, suggesting that intermodule connections are significant and useful for information integration (Mason & Verwoerd, 2007; Gitter,

Klein-Seetharaman, Gupta, & Bar-Joseph, 2011; Deeds, Krivine, Feret, Danos, & Fontana, 2012; Navlakha, Gitter, & Bar-Joseph, 2012). This trade-off between economy and complexity could also signify how disruption of one biological process affects another (closely or distantly related) biological process (Dillin, Gottschling, & Nystrom, 2014) and how efficiently these effects propagate throughout the network (Navlakha, He, Faloutsos, & Bar-Joseph, 2014).

In this work, we find evidence of Rentian scaling of functional modules in a diverse set of molecular and cellular networks, including physical protein interaction networks, cell-type specific gene regulatory networks, and neural circuits. In contrast to prior work testing Rentian scaling, which decomposed networks into topological modules using graph partitioning algorithms (Bassett et al., 2010; Klimm, Bassett, Carlson, & Mucha, 2014; Bullmore & Sporns, 2012), we decomposed networks into biological modules based on known functional annotations of nodes. This allowed us to test whether natural modules designed by evolution exhibited Rentian scaling in the same way that the human design process naturally led to Rentian scaling of engineered circuits. We find strong evidence in support of Rentian scaling, even when using different definitions of *biological module* and different protein interaction data sources for defining molecular networks. In gene regulatory networks, we also found that Rent's exponents in embryonic cells are significantly smaller than the exponents in differentiated cells, indicating that Rentian scaling may provide a useful network feature for understanding cellular maturation throughout development. We also find Rentian scaling in functional modules of the *C. elegans* neural network, but, interestingly, not in three social networks, implying that information processing may play a role in producing Rentian structure. Finally, to help explain how Rentian scaling may have emerged biologically, we derived a new evolutionary graph model to generate networks with a specified Rent's exponent. Overall, our work describes a new principle of how intermodule edges are distributed within diverse biological networks.

2 Calculating the Rent's Exponent of a Network

As input, we are provided an undirected network $G = (V, E)$ and a decomposition of the nodes V into k modules, $\mathcal{M} = (M_1, M_2, \dots, M_k)$, where each $M_i \subset V$. An individual node can belong to multiple modules, and hence $M_i \cap M_j \neq \emptyset$, necessarily.

To compute the Rent's exponent p for each network (see Figure 1), we calculated two numbers for each module M_i : the number of nodes inside the module (n_i), and the total number of external edges (e_i) from nodes in M_i to nodes in other modules. This is straightforward to calculate, except in the case where modules overlap. For example, if node $x \in M_i, M_j$ and node $y \in M_j$, and if edge $(x, y) \in E$, then for M_i , the edge is counted as external,

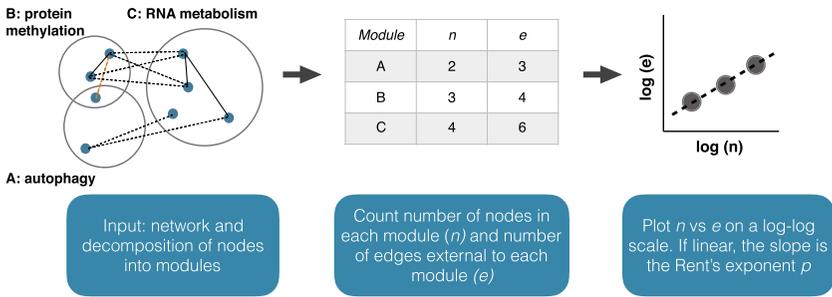


Figure 1: Computing the Rent's exponent of a network. (Left) The input is a network and a decomposition of the nodes into modules. Solid and dotted lines denote internal and external edges, respectively. The orange dotted line is considered an external edge for module A and an internal edge for module B. (Middle) Count the number of nodes in each module (n) and the number of external edges from nodes in that module to nodes in other modules (e). (Right) Plot n versus e on a log-log scale. If linear, the slope of the line is the Rent's exponent.

but for M_j , it is internal since both x and y belong to M_j . We then plot n_i versus e_i for all i and test whether the following relationship between the two variables holds:

$$e = cn^p, \quad (2.1)$$

where c is an integration constant equal to the average number of external connections per node.

The significance of this relationship was tested by plotting n_i versus e_i for all i on a log-log scale, computing a regression line for the data using least squares, and then computing the coefficient of determination (R^2) for the line. It is well known that topological and physical Rentian scaling does not hold for large modules—called “Region II” by the Rent community—and thus we removed large modules from our analysis (see section A.1 in the appendix).

If the points fall on a straight line on a log-log plot, the network exhibits Rentian scaling, with an exponent $0 < p \leq 1$. Thus, $p = 1$ corresponds to a random arrangement of external connections, with no placement optimization (i.e., no preference between inter- or intramodule connections). For digital circuits with a homogeneous layout, the smaller the value of p , the more efficiently the network is embedded (Christie & Stroobandt, 2000; Bassett et al., 2010), indicating a preference of economy over complexity. In functional modules in molecular networks, the embedding can correspond to a diffusion distance for proteins to find their correct

interaction partners, with smaller values of p indicating a preference for local interactions.

Testing Rentian scaling requires a network to be decomposed into modules (\mathcal{M}). Prior work has defined these modules using the topology of the network (Bassett et al., 2010), for example, by recursively bipartitioning the network using graph cuts (Karypis, Aggarwal, Kumar, & Shekhar, 1999) and computing Rent's exponent using modules at all levels in the hierarchy. However, one downside of this approach is that a graph clustering method needs to be selected, and this method might implicitly bias toward modules that display Rentian scaling. An alternative approach for networks that are physically embedded in space is to randomly place randomly sized "boxes" onto the network; each box corresponds to a module and includes all the nodes it covers (Yang, Bozorgzadeh, & Sarrafzadeh, 2001; Bassett et al., 2010). This is also undesirable for us because most proteins diffuse and do not have fixed positions in space. Here, we depart from both of these approaches and instead define modules biologically (instead of topologically), based on the known cellular function of individual nodes. Specifically, a biological module contains a set of genes or proteins that takes part in the same biological process within the cell, such as RNA metabolism, protein methylation, or autophagy. Assigning genes/proteins to modules is largely done independently from network topology, often based on a battery of experimental assays, including gene knockout experiments and expression data on transcript levels in different conditions. These natural modules also range across many sizes, which is important to test any scaling rule. Critically, this approach allows us to test whether evolution-defined modules exhibit Rentian scaling similar to how the human design process naturally led to Rentian scaling of digital circuits.

2.1 Network Randomization to Test Rentian Scaling. To determine whether a network's Rent's exponent was not due to chance alone, for each network we compared its empirical Rent's exponent with three random controls. We reason that any given network that follows Rent's rule should, if randomized, have a larger Rent's exponent, as found by Bassett et al. (2010). We repeated each type of randomization 100 times and then compared each network's empirical and random Rent's exponents using a one-sample two-tailed t -test. The controls were generated as follows:

1. *Random modules.* We selected two random nodes u and v and for each listed the modules they do not share in common. Assuming these two lists are not empty, we then randomly selected one module in each list (call them m_u and m_v) with probability proportional to the size of the module, so that larger modules were more likely to be randomized. We then moved u into m_v and v into m_u . This procedure ensured that module affiliations are randomized, but the distribution of module sizes remains the same as in the empirical

network. We swapped nodes $\sum n_i$ times (i.e., the sum of the number of nodes in every module) and calculated Rent's exponent with the new modules. We repeated this procedure 100 times and report the average Rent's exponent. The edges remain unchanged for this control.

2. *Random edges.* We performed a standard edge-swapping procedure that preserves the degree distribution: we selected two edges (u, v) and (x, y) and swapped their end points to create (u, y) and (v, x) , assuming neither of the new edges already exists. We swapped pairs of edges $|E|$ times, where $|E|$ is the number of edges in the network, and then computed Rent's exponent. We repeated this procedure 100 times and report the average Rent's exponent. The modules remain unchanged for this control.
3. *Random both.* We first randomized the modules and then randomized the edges.

3 An Evolutionary Algorithm to Generate Rentian Networks _____

Here, our goal is to derive a simple, biologically feasible graph model that could generate networks that display Rentian scaling. Formally, the problem is as follows:

Given: A set of nodes V , a decomposition of V into modules \mathcal{M} , and a target Rent's exponent p

Find: A graph $G = (V, E)$ with Rent's exponent $\approx p$, computed using the modules \mathcal{M}

Our goal is to develop a generative model to define the edges connecting the given nodes in V . Contrary to most generative graph models (Leskovec, Kleinberg, & Faloutsos, 2005; Barabási & Albert, 1999; Watts & Strogatz, 1998; Vázquez, Flammini, Maritan, & Vespignani, 2003) and prior work in Rentian scaling (Klimm et al., 2014), our problem formulation includes a module decomposition as input used to compute Rent's exponent. The modules are used as input so that we can use the same decomposition and target Rent's exponent of any empirical network, which ensures that a solution exists and allows for a direct comparison between the synthetic and real networks.

Duplication-divergence represents one common biological mechanism used to evolve molecular networks. Below, we describe a standard model for this process (called DMC) followed by our extension of this model (called DSC) that uses the same duplication principle but better captures Rentian properties of biological networks. We generated an ensemble of 100 DMC and 32 DSC networks. We generated only 32 DSC networks because of the time required to generate networks with an extra parameter (namely, q_{fav}). We then generated 100 DSC networks for the best parameter

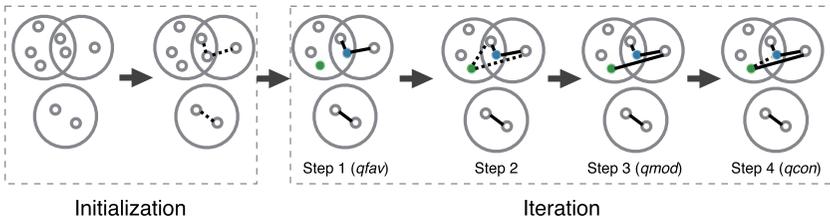


Figure 2: Duplication-specialization with complementarity (DSC) model. (Left) To initialize, we connect two random nodes in each module; some nodes may be chosen more than once because modules overlap. (Right) Step 1: We select a “new” node u (green) and an “anchor” node v (blue) from which u duplicates. With probability q_{fav} , the anchor node is selected from a module shared with u . Step 2: Neighbors of v are copied onto u . Step 3: For each common neighbor x , with probability $1 - q_{mod}$, both edges (u, x) and (v, x) are retained; with probability q_{mod} , (u, x) is deleted if u and x share fewer modules than v and x , or vice versa. Step 4: The two duplicated nodes are connected with probability q_{con} . This procedure is repeated until all nodes join the network.

combination for each species in order to better compare the DMC and DSC networks. The results we describe are averages over the ensemble.

3.1 The Duplication-Mutation with Complementarity Model (DMC).

In this model (Vázquez et al., 2003), an existing gene (node u) duplicates to initially form a topologically equivalent node (v). Then their set of common interaction partners diverges, indicating subspecialization of the two genes. These two processes, duplication and divergence, are controlled by two parameters: q_{mod} and q_{con} . Specifically, after each duplication step, each common neighbor x of u and v is retained independently by both u and v with probability $1 - q_{mod}$; with probability q_{mod} , either (u, x) or (v, x) is removed, with equal probability. Then, with probability q_{con} , an edge (u, v) is added between the duplicates. This model has been extensively validated in terms of its ability to reproduce many known topological features of protein interaction networks (Middendorf, Ziv, & Wiggins, 2005; Navlakha & Kingsford, 2011; Navlakha, Faloutsos, & Bar-Joseph, 2015), yet it does not use or incorporate any module decomposition in its growth procedure.

In each step of the model, we choose one random unconnected node (u , the “new” node) and a random node that already joined the network (v , the “anchor” node), and copy v to u by applying the DMC procedure described above. We iterated this process until all nodes joined the network.

3.2 The Duplication-Specialization with Complementarity Model (DSC). To directly incorporate the module decomposition into the growth procedure, we created the DSC model (see Figure 2). This model differs

from DMC in two ways:

1. We introduced a module-aware parameter, q_{fav} , used in the duplication step. With probability q_{fav} , the anchor node v is selected randomly from a module in which the new node u lies (instead of randomly from the entire network in DMC). This is motivated by the biological observation that duplicated nodes retain some functional association with their ancestral node (Wagner, 2001; Pereira-Leal & Teichmann, 2005); q_{fav} allows us to model this variability. Due to module overlap, even if $q_{\text{fav}} = 1$, there may still be external edges between modules.
2. We modified the divergence procedure. If an edge to neighbor x is chosen to diverge, then instead of randomly selecting which edge (u, x) or (v, x) to delete, we deterministically delete the edge depending on which node (u or v) has fewer modules in common with x . In other words, if u and x have fewer modules in common than v and x , then edge (u, x) is deleted. This represents the specialization of duplicated proteins observed by Nasvall, Sun, Roth, and Andersson (2012) and agrees with the observation by Zinman, Zhong, and Bar-Joseph (2011) that interactions between proteins in the same module are more likely to be conserved than interactions between proteins in different modules. Thus, when choosing which edge will diverge, the DSC model takes into account the module overlap with the neighbor in question, whereas DMC merely flips a coin.

An illustration of the DSC model is shown in Figure 2. Pseudocode of the model is shown in algorithm 1.

4 Results

First, we asked if three classes of biological networks (molecular, gene regulatory, and neural) and two classes of nonbiological networks (information and social) exhibit Rentian scaling. Second, we evaluated the ability of two generative graph models (DMC and DSC) to recapitulate Rentian properties for the biological networks studied.

4.1 Molecular Networks Demonstrate Rentian Scaling. We collected protein-protein interaction (PPI) networks for six species from BioGRID (Stark et al., 2006; see the appendix): *Drosophila melanogaster* (fly), *Homo sapiens* (human), *Mus musculus* (mouse), *Arabidopsis thaliana* (plant), *Schizosaccharomyces pombe* (fission yeast), and *Saccharomyces cerevisiae* (baker's yeast). We assigned each protein in the network to modules based on the protein's known annotations under the Biological Process ontology in the Gene Ontology database (GO; Ashburner et al., 2000). Each module represents a biological process; all proteins annotated to that biological process are assigned to the module. These annotations are largely determined using biological

Algorithm 1: DSC ($M, q_{con}, q_{mod}, q_{fav}$).

```

1: G = Graph() # Initialize empty graph.
2: for all m ∈ M do
3:   u = m.random_node()
4:   v = m.random_node()
5:   G.add_edge(u,v) # Start with dumbbell in each module.
6: end for
7:
8: VC = set of connected nodes
9: VU = set of unconnected nodes
10:
11: while |VU| > 0 do
12:   u = VU.pop_random_node()
13:   Mu = set of modules that u is in
14:   Mrest = set of modules that u is not in
15:
16:   #1.Module-aware duplication.
17:   if rand() < qfav then
18:     m = Mu.random_module()
19:   else
20:     m = Mrest.random_module()
21:   end if
22:   mC = m ∩ VC # Set of connected nodes in m
23:   v = mC.random_node()
24:
25:   G.copy_node(v,u) # Duplicate v to u.
26:   for all x ∈ G.neighbors(v) do
27:     #2.Neighbor divergence.
28:     Overlapu = number of modules shared by x and u
29:     Overlapv = number of modules shared by x and v
30:     if rand() < qmod then
31:       if Overlapu < Overlapv then
32:         G.delete_edge(u,x)
33:       else
34:         G.delete_edge(v,x)
35:       end if
36:     end if
37:   end for
38:
39:   #3.Duplicate interaction.
40:   if rand() < qcon then
41:     G.add_edge(v,u)
42:   end if
43: end while
44: return G

```

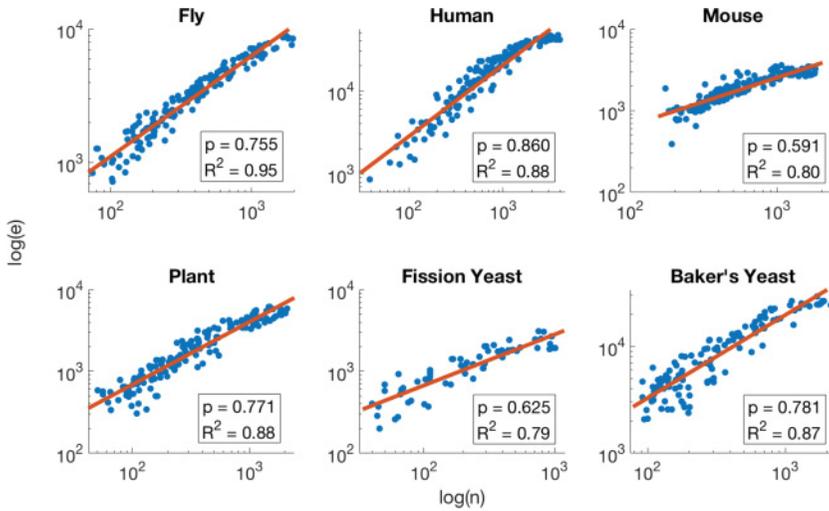


Figure 3: Molecular networks exhibit Rentian scaling. Rent's exponents for protein interaction networks of six species with REVIGO cutoff of 0.7. The x -axis is the log of the number of nodes (n), and the y -axis is the log of the number of external edges (e) from that module to other modules. All networks display Rentian scaling, with the Rent's exponent p shown in the legend.

assays, such as gene knockout experiments, that did not include network topology as a criteria.

To test Rent's rule for modules over a range of sizes, we selected nonredundant GO annotations from all levels of the GO hierarchy using REVIGO (Supek, Bosnjak, Skunca, & Smuc, 2011). REVIGO uses a clustering algorithm to find one GO annotation (module) that represents many semantically similar GO annotations. Semantic similarity accounts for how close a pair of GO annotations is to its lowest common ancestor in the GO hierarchy (Schlicker, Domingues, Rahnenfuhrer, & Lengauer, 2006). A user-defined cutoff parameter can be used to vary the number and semantic overlap of annotations. The resulting sizes of modules varied by an order of magnitude, for example, from 174 to 1803 nodes per module for *M. musculus* at a REVIGO cutoff of 0.7. There was also significant module overlap (pleiotropy), with nodes belonging to an average of 12.5 modules (*S. pombe*) to 35.8 modules (*M. musculus*), with the other species lying in between.

We found that all six PPI networks displayed Rentian scaling—each least-squares regression line achieved an $R^2 > 0.75$ —with Rent's exponents ranging from 0.591 for mouse to 0.860 for human (see Figure 3). This means that as modules contain more nodes, they have more external connections in accordance with a power law, indicating that a single scaling rule can

capture the distribution of intermodule connections for modules of many sizes.

We also found a strong positive correlation between the number of edges in the network and the Rent's exponent ($R = 0.80$, over all species), suggesting that larger networks tend to increase cross-module communication, as opposed to increasing intramodularity. Moreover, all randomized controls produced Rent's values that were significantly higher than the empirical Rent's exponent observed for the real network (see Table 1). Randomizing both edges and modules consistently produced networks with larger Rent's exponents than randomizing either modules or edges alone (see Table 1). There was, however, no clear pattern when comparing Rent's exponents when we randomize either edges or modules; in half the species, the former produced lower Rent's exponents, whereas the opposite was true in the other species. While some of these differences may appear small, they are differences in log-space, and they mimic the range of differences observed in prior Rentian analysis of brain networks (Bassett et al., 2010).

4.1.1 Robustness to Different Module Decompositions. We tested the robustness of this observation in two ways. First, we varied the cutoff parameter in REVIGO to generate module decompositions with fewer (cutoff = 0.5) and greater (cutoff = 0.9) numbers of modules. In both cases, we observed similar Rent's exponents for each species, which continued to be significantly lower than randomized controls (see Table 1). Rent's exponents generally increased as the number of modules increased for all species.

Second, we assigned proteins to all modules in the entire GO Biological Process hierarchy (we kept semantically redundant modules) using the GoTermFinder tool (Boyle et al., 2004). This increased the range of module sizes to over four orders of magnitude, and it produced even more highly overlapping modules; for example, each node belonged to an average of 41.3 modules in *S. pombe* and 106.7 modules in *M. musculus*. This extensive amount of module overlap increased the number of cross-module connections; thus, all Rent's exponents increased, but we still observed Rentian scaling over separate hierarchical levels of every PPI network (see Table 4 and Figure 7 in the appendix).

These tests together demonstrate that Rentian scaling in PPIs is robust to different definitions of a module and extent of module overlap.

4.1.2 Robustness to Different Data Sources. Protein interaction networks are notoriously noisy and incomplete (Bader, Chaudhuri, Rothberg, & Chant, 2004; Huang & Bader, 2009), affecting the conclusions that can be drawn from their analyses. To generate an alternative set of interactions, we used the STRING database, which collects interactions over a broader range of data sources (Szkarczyk et al., 2015; see also the appendix). We considered physical "binding interactions" only with a confidence score greater than 700 to minimize noise. We again used REVIGO at three different

Table 1: Rent's Exponents for PPI Networks and the *C. elegans* Neural Network.

Species	Number of Nodes	Number of Modules	Number of Edges	REVIGO	Rent's Exponent	Random Modules	Random Edges	Random Both
Fly	4544	95	18,791	0.5	0.742 (0.94)	0.751 (0.97)	0.751 (0.96)	0.754 (0.97)
	4585	171	18,952	0.7	0.755 (0.95)	0.764 (0.97)	0.765 (0.96)	0.767 (0.97)
Human	4589	280	18,981	0.9	0.762 (0.95)	0.770 (0.97)	0.772 (0.96)	0.774 (0.97)
	9615	87	102,292	0.5	0.845 (0.86)	0.869 (0.92)	0.857 (0.88)	0.873 (0.92)
	9656	158	102,884	0.7	0.860 (0.88)	0.881 (0.93)	0.873 (0.90)	0.886 (0.92)
Mouse	9657	305	102,895	0.9	0.868 (0.88)	0.886 (0.93)	0.880 (0.90)	0.891 (0.93)
	3796	96	7047	0.5	0.565 (0.76)	0.582 (0.83)	0.582 (0.86)	0.588 (0.86)
	3817	185	7068	0.7	0.591 (0.80)	0.607 (0.83)	0.607 (0.88)	0.613 (0.87)
Plant	3820	292	7072	0.9	0.604 (0.80)	0.620 (0.81)	0.620 (0.87)	0.626 (0.86)
	5390	79	15,556	0.5	0.762 (0.88)	0.781 (0.97)	0.784 (0.93)	0.788 (0.97)
	5418	158	15,606	0.7	0.771 (0.88)	0.789 (0.97)	0.792 (0.93)	0.796 (0.97)
Fission yeast	5418	280	15,606	0.9	0.770 (0.90)	0.786 (0.97)	0.791 (0.94)	0.793 (0.97)
	2030	53	6943	0.5	0.561 (0.81)	0.605 (0.94)	0.612 (0.92)	0.614 (0.95)
	2030	73	6943	0.7	0.625 (0.79)	0.669 (0.93)	0.674 (0.90)	0.680 (0.93)
Baker's yeast	2030	109	6943	0.9	0.674 (0.77)	0.719 (0.91)	0.727 (0.87)	0.733 (0.91)
	4743	92	65,775	0.5	0.770 (0.86)	0.802 (0.94)	0.791 (0.90)	0.806 (0.94)
	4747	124	65,810	0.7	0.781 (0.87)	0.810 (0.95)	0.802 (0.91)	0.816 (0.95)
Worm	4759	202	66,056	0.9	0.825 (0.87)	0.853 (0.94)	0.846 (0.90)	0.859 (0.93)
	239	8	1612	—	0.859 (0.97)	0.950 (0.95)	0.920 (0.99)	0.960 (0.95)

Notes: All Rent's exponents are significantly smaller than each of the corresponding randomized controls ($p < 0.01$, one-sample two-tailed t -test). Values in parentheses are R^2 .

cutoffs to define modules and found that these networks continued to display Rentian scaling (see Table 5 in the appendix), further suggesting that Rentian scaling is a robust and conserved property of PPI networks. Of note, when the STRING database is used, the human PPI network has a much lower R^2 value compared to other species and compared to the full human PPI network (see Table 1). This discrepancy may be due to the substantially fewer nodes in the high-confidence STRING human PPI network (~ 3300) versus in the full human PPI network (~ 9600). This reduction suggests that there are many human PPIs yet to be mapped at a high-confidence level, leaving the network more incomplete compared to other species. It could also suggest that human PPI networks simply do not exhibit Rentian scaling.

4.1.3 Comparison of Rent's Exponents Using Functional versus Topological Modules. Conventional analyses have used topological modules instead of functional modules to derive Rent's exponents. Here, we compared the exponents generated by the two approaches. To derive topological modules, we used hMetis (Bassett et al., 2010; Karypis et al., 1999), which recursively bipartitions the graph to minimize the number of edges between partitions. This procedure resulted in 11 to 13 hierarchical levels of partitions across the different PPI networks. The average number of nodes and external degrees in a partition at a given hierarchical level were plotted on a log-log scale, and the slope of the line was defined as the topological Rent's exponent. The Rent's exponents ranged from 0.74 to 1.04 for mouse and human, respectively, and the R^2 values ranged from 0.25 to 0.85 for human and fission yeast, respectively (see Table 3 in the appendix). The Rent's exponents from the topological modules were always larger than those for the functional modules at all Revigo cutoff values and for every species, which implies more random wiring in the topological modules. Further, in five out of six networks, the fits were poorer. Thus, Rentian scaling in PPIs seems to be better defined using functional, not topological, modules.

The differences in Rent's exponents observed between topological modules and functional modules could be attributed to several factors. First, there are numerous graph-theoretic algorithms to partition networks, and it is not clear which definition of *topological module* makes the most biological sense. While hMetis is commonly used in the Rent community, recent work has shown that the Markov clustering algorithm (Enright, Van Dongen, & Ouzounis, 2002; Brohee & van Helden, 2006) and graph-summarization-based approaches (Navlakha et al., 2009) produce the highest correlation between topological and functional modules in PPI networks useful for protein function prediction. These methods, however, do not return hierarchical modules, which is important in order to test Rentian scaling across scales. Second, PPI networks are notoriously noisy, with many spurious interactions and false negatives that obfuscate the true modules. Thus, we focused our study here on more ground-truth modules, based on known

functional annotations. There are also some limitations in using functional modules (see section A.2).

4.2 Analysis of Cell-Type Specific Regulatory Networks. We collected cell-type-specific regulatory interactions for 41 human and 25 mouse cell lines (Neph et al., 2012; Stergachis et al., 2014) to test if Rentian scaling also appears in transcriptional networks. Nodes in these networks correspond to genes, and an edge exists between two genes if one regulates the transcription of the other. Each of the 66 networks contained interactions that occur within only one specific cell type (e.g., fetal brain, hepatoblastoma, and embryonic stem cells). Using REVIGO-defined modules, we found that all cell-type-specific networks displayed Rentian scaling, with exponents ranging from 0.795 to 0.937 (mouse; see Table 6 in the appendix) and 0.843 to 0.921 (human; see Table 7 in the appendix). The largest of the Rent's exponents were attributed to immune and cancer cells.

We also found that embryonic stem cells (ESCs) had significantly lower Rent's exponents than differentiated cells in both species (see Figure 4). Unlike the PPI networks, where more interactions correlated with significantly higher Rent's exponents, the opposite was true here: ESCs contained more interactions than other cell types ($17,883 \pm 3518$ for ESCs versus $14,810 \pm 3522$ for differentiated cells in mouse) yet had lower Rent's exponents (see Figure 4). This suggests that as a cell develops and defines its functional identity, its Rent's exponent increases by eliminating more intramodule edges than intermodule edges. Thus, one signature of cellular differentiation may be changes in the distribution of intermodule connections, from less to more intermodule cross-talk as the cell matures, an observation consistent with the balanced lineage specifier hypothesis (Loh & Lim, 2011).

4.3 Analysis of Neural, Social, and Information Networks. Next, we tested the generality of Rent's rule when applied to topology-independent module decompositions of other biological (neural) and nonbiological (social, information) networks.

Prior work has analyzed the Rentian properties of the *C. elegans* neural network (Bassett et al., 2010); however, here we applied a Rentian analysis based on a module decomposition defined by known functional annotations of neurons, as opposed to defining modules using a graph partitioning algorithm, as was previously done. Specifically, we obtained the neural network for *C. elegans* from WormAtlas (Varshney, Chen, Paniagua, Hall, & Chklovskii, 2011) and decomposed this network into eight natural modules, where each module contained neurons with the same function (Varshney et al., 2011). These functions were chemosensation, mechanosensation, nociception, thermosensation, proprioception, sensation of oxygen, olfaction, and motion. We found that the network displayed Rentian scaling ($R^2 = 0.97$), with a Rent's exponent of 0.859 (see Figure 5a), which was

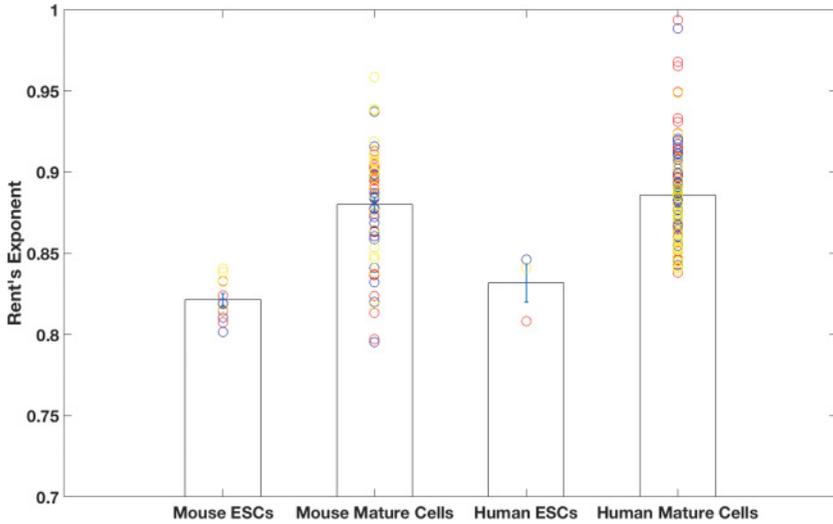


Figure 4: Embryonic stem cells (ESCs) exhibit a lower Rent's exponent than differentiated cells. The average Rent's exponent for ESCs was lower than that of mature cells in both mice and humans. Bars are standard error of the mean, and circles are the Rent's exponents for individual cell lines with modules defined by REVIGO at three cutoffs. The three different colors for the circles denote the three Revigo cutoffs (red = 0.5, blue = 0.7, yellow = 0.9). Differences in Rent's exponents between ESCs and non-ESCs in a given species are statistically significant across all Revigo values ($p < 0.01$, two-sample two-tailed t -test) and within individual cutoffs ($p < 0.01$, two-sample Kolmogorov-Smirnov test for mouse at Revigo values of 0.5, one-sample two-tailed t -tests for mouse at Revigo values of 0.7 and 0.9, and one-sample two-tailed t -tests for human because the latter had just one ESC line).

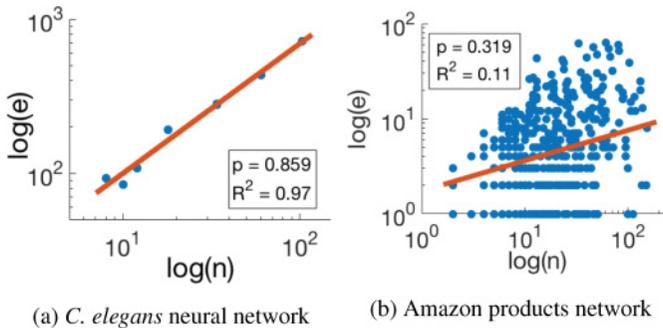


Figure 5: The *C. elegans* neural network displays Rentian scaling, but the Amazon product network does not. The R^2 for the worm neural network was 0.97; for the Amazon network, it was 0.11.

smaller than randomized controls (see Table 1). This exponent suggests two things. First, neural circuits have some preference for short-range over long-range connections, as supported by the wiring economy principle (Ramón y Cajal, 1899; Chklovskii, Schikorski, & Stevens, 2002; Rivera-Alba, Peng, de Polavieja, & Chklovskii, 2014), which states that wire is a commodity in space- and resource-constrained neural circuits (as in digital circuits). Second, due to the relatively high exponent, there is some deviation from this principle, as may be expected since intermodule connections are needed to integrate data from multiple neural types to determine appropriate behavioral responses (Perez-Escudero & de Polavieja, 2007). Prior work found a lower Rent's exponent (0.74) when computed using random partitions in a physical placement of the neurons (Bassett et al., 2010). Thus, Rentian scaling appears robust, but the exponent is sensitive to the choice of module decomposition.

We then tested Rentian scaling on three social and information networks with ground-truth modules (Yang and Leskovec, 2015; Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007). In the Amazon network, products (nodes) were linked to other products that were purchased together (edges). The modules represented groups of at least three products that share a common function, ascribed using predefined product categories (Yang & Leskovec, 2015). The Amazon network did not display Rentian scaling, as the log-log regression was not a straight line ($R^2 = 0.11$; see Figure 5b). We also tested the DBLP collaborations network and YouTube social groups network (see the appendix), and found that neither exhibited Rentian scaling.

Together, these results suggest that Rentian scaling is not an inevitable consequence of any network developmental process. Further, many classes of random networks also do not exhibit Rentian behavior (Stroobandt, 2007; Klimm et al., 2014). One hypothesis, then, is that Rentian scaling may be unique to information processing networks (e.g., digital circuits on a computer chip or biological circuits) and is derived implicitly by some general growth process, which we investigate next.

4.4 A Random Graph Model to Recapitulate Rentian Properties of Networks. The fact that Rentian scaling is found in molecular and cellular networks of several species and cell types indicates that this property may be a consequence of a common evolutionary process. We first tested whether a popular, existing generative graph model (DMC) could form networks that exhibit Rentian scaling while agnostic to the module decomposition. Overall, we found that DMC mostly failed in generating the empirical Rent's exponents of PPI networks, whereas our extended model (DSC) performed better, as described below. The DMC and DSC models are described in section 3.

To test each model's ability to generate Rentian networks, we used as input to the model the same set of nodes and their module decomposition

from each PPI network, separately. The challenge was to recapitulate the Rentian properties for each PPI network by defining the edges. The synthetic, model-generated network “succeeded” in this regard if it: (A) it had a Rent’s exponent within 0.04 of the empirical value; (B) contained within 10% of the number of edges as the empirical network; and (C) had a similar distribution of external edges per node as the empirical network. Attempting to lower the error rate of measure B resulted in an inability to match Rent’s exponents for some networks; thus, more model parameters may be needed for a tighter fit. We tested all combinations of the three model parameters (q_{mod} , q_{con} , q_{fav} ; see section 3) with their values ranging from 0.1 to 0.9 in intervals of 0.1. We report the results for the best parameter settings for each model in terms of measures A and B.

In the analysis of the ensemble, we found that the average DSC Rent’s exponents were closer to the empirical Rent exponents than the DMC Rent’s exponents across all species (see Figure 6A). In this figure, we selected the parameters that generated networks with the closest Rent’s exponent to the empirical while abiding by the constraint on number of edges. Figure 6B shows example Rentian plots with the number of external edges averaged over the entire ensemble (the sizes of the modules remain the same across the ensemble). In all six species, the DSC model produced networks with more similar Rent’s exponent and/or more similar number of edges as the empirical network, compared to DMC-grown networks. For example, when using the same set of nodes and module decomposition as the fly PPI network (see Table 2), there were 18,952 edges in the empirical network, 17,658 edges in the DSC network, and 19,289 edges in the DMC network. Furthermore, the DMC network had a Rent’s exponent of 0.87, which was further away from the DSC and empirical networks’ Rent’s exponents of 0.78 and 0.76, respectively.

Interestingly, the parameter settings for DSC that led to successful networks may highlight the relative roles of duplication and divergence during evolution. For all parameter values that generated successful networks (as defined above), $q_{\text{mod}} = 0.64 \pm 0.09$, $q_{\text{con}} = 0.63 \pm 0.23$, $q_{\text{fav}} = 0.53 \pm 0.26$. Prior studies have determined that $q_{\text{mod}} = 0.4$ and $q_{\text{con}} = 0.7$ produce networks that closely match other topological properties of the yeast PPI network (Navlakha & Kingsford, 2011; Navlakha et al., 2015); these parameters lie roughly within the range of best parameters we found, which further validates their use here.

We also found that the DSC model better reproduced the degree distribution of the number of external edges per node compared to DMC (see Table 2). This is a common metric used to evaluate the similarity between a real network and a synthetic network generated by a model. This measure (dubbed “homogeneity” by the Rentian community) is a common feature of digital circuits and suggests that the interconnect complexity of nodes is similar across modules of different sizes (Christie & Stroobandt, 2000). To quantify the similarity in degree distributions, we calculated the

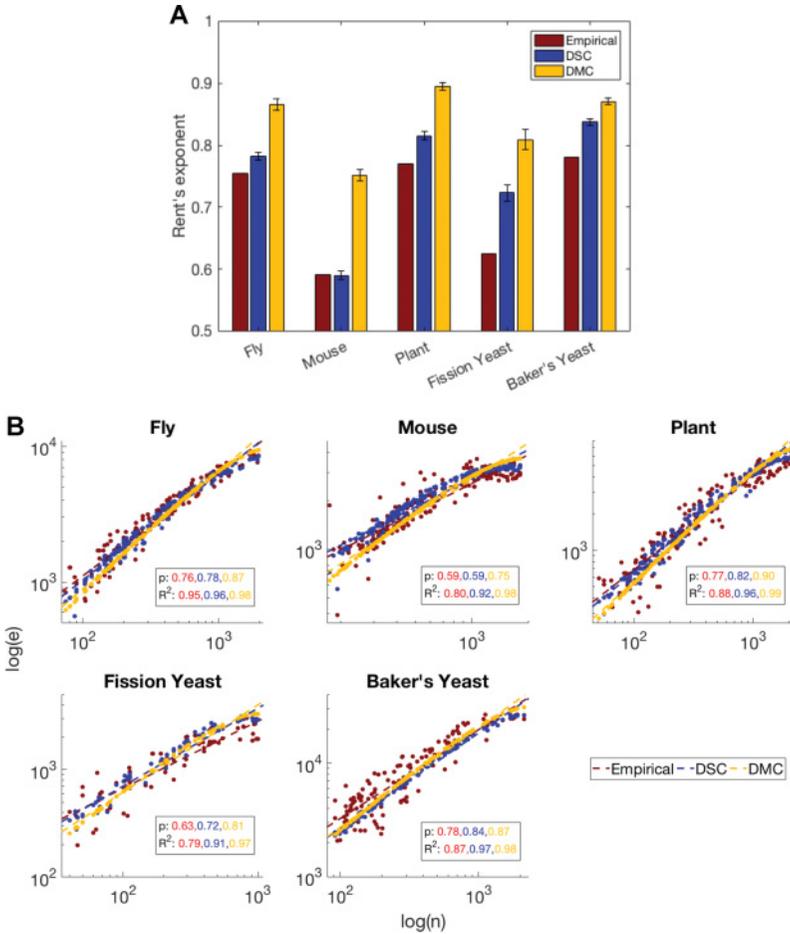


Figure 6: The DSC model better recapitulates the Rentian properties of real PPI networks compared to DMC. (A) Compared to DMC, the DSC model generates networks with Rent's exponents closer to the empirical Rent's exponents of five PPI networks. Ensembles of networks were generated for each possible parameter combination for each model. Bars indicate mean Rent's exponent, and error bars represent standard deviations over the ensemble. We used the same nodes and module decomposition of each real PPI network (REVIGO cutoff of 0.7). (B) Rent plots of networks generated by DMC and DSC (REVIGO cutoff of 0.7). Empirical data are shown in red, the DSC model is shown in blue, and the DMC model is shown in yellow. Each dot corresponds to one module. Ideal performance would be an exact overlap with the red data, indicating the same number of edges and the same Rent's exponent as the PPI network. Performance is shown for the best parameters for each model. Overall, the DSC line is closer than DMC to the empirical line across all species. The Rent's exponents (p) and fits (R^2) are shown in boxes.

Table 2: The DSC Generative Model Better Recapitulates Key Characteristics of the Empirical Networks than the DMC Model.

Species	Rent's Exponent	Number of Edges	DMC Rent's Exponent	DMC Number of Edges	DMC KL Distance	DSC Rent's Exponent	DSC Number of Edges	DSC KL Distance
Fly	0.755	18,952	0.866 ± 0.009	19,289 ± 1178	0.025 ± 0.003	0.783 ± 0.006	17,658 ± 446	0.018 ± 0.002
Mouse	0.591	7068	0.751 ± 0.010	7226 ± 234	0.033 ± 0.002	0.590 ± 0.006	6756 ± 132	0.026 ± 0.001
Plant	0.771	15,606	0.896 ± 0.006	14,724 ± 511	0.055 ± 0.004	0.815 ± 0.007	15,126 ± 267	0.041 ± 0.003
Fission yeast	0.625	6943	0.809 ± 0.016	6645 ± 464	0.089 ± 0.009	0.723 ± 0.014	7014 ± 239	0.087 ± 0.008
Baker's yeast	0.781	65,810	0.871 ± 0.006	63,787 ± 3887	0.060 ± 0.003	0.837 ± 0.005	60,010 ± 2359	0.053 ± 0.005

Notes: All values are means ± standard deviations from 100 network instantiations for the DMC and DSC models. KL Distance refers to the average Kullback-Leibler divergence of the number of external edges per node in a module from the PPI networks to 100 instantiations of DMC or DSC networks, plus or minus the standard deviation. Results here used a REVIGO cutoff of 0.7.

Kullback-Leibler (KL) divergence to compute the distance between homogeneity histograms, from empirical to DSC and from empirical to DMC. The KL distance from empirical to DSC was smaller than the distance from empirical to DMC in all species (e.g., 0.018 ± 0.002 versus 0.025 ± 0.003 , respectively, for *D. melanogaster*), except for *S. pombe*, where the two distances were nearly identical. This further validates the ability of DSC to reproduce Rentian properties of real networks compared to DMC. There are numerous other measures used to establish the similarity between two networks (e.g., motifs, graphlets, random walk distributions, graph spectra). We do not claim that DSC is the definitive model in this regard; rather, we show how a simple model based on established biological principles can capture a new feature (Rentian scaling) that future, more complex models should attempt to satisfy.

5 Discussion

We provided evidence that the structure of several molecular and cellular networks studied here is Rentian. While modularity has long been an appreciated facet of molecular networks, we find that these networks are “more than mere modules” (Pinkert et al., 2010), with significant structure in their intermodule connectivity. Inspired by theories developed from engineered digital circuits, we showed that Rentian scaling is a conserved property of several protein interaction networks and gene regulatory networks across multiple species and cell types; it also applies to functional modules in the *C. elegans* neural network but not to three social or information networks. Important to our work was testing this theory using module decompositions that were derived largely independently from network topology and instead by the underlying biology itself. This allowed us to test whether evolution converged onto biological modules with Rentian properties.

This power law relationship in module interconnects may aid in other applications, such as for function prediction tasks (Pinkert et al., 2010) or for ranking false or missing interactions in the network. Rentian scaling may also be used as a feature to discriminate between networks, especially across time. For instance, Sperry, Telesford, Klimm, and Bassett (2017) found that the present-day London Underground railroad displays Rentian scaling, but it did not in the year 1900. We found that regulatory networks for embryonic stem cells displayed Rentian scaling but that Rent’s exponents significantly increased with cellular development. Hence, the emergence of Rentian scaling, or the increase in Rent’s exponent, may provide another feature of how networks develop over time (Leskovec et al., 2005). Rentian scaling, however, cannot be trivially explained by changes in network density. For the PPIs, we found a positive correlation between Rent’s exponent and network density; however, this relationship did not hold when PPIs were taken from the STRING database ($R = 0.80$ and -0.05 for the full PPIs and those from STRING, respectively). Furthermore, in the regulatory

networks, there was an inverse correlation between Rent's exponent and network density ($R = -0.64$ and -0.26 for the mouse and human cell lines, respectively).

While this scaling law was persistent, exact interpretation of the Rent's exponent for an individual species may be muddled by the fact that protein interaction networks are still very noisy and incomplete. Although we attempted to account for this by studying networks derived from different data sources, it is difficult to determine whether the "true" Rent's exponents for these networks are higher or lower than the values derived here. Erroneous edges would likely count more as intermodule edges than intramodule edges; thus, removing noise may lower Rent's exponents. On the other hand, there are likely many missing edges between proteins in different modules, which might increase Rent's exponent when mapped. Our results here can serve as a benchmark for future studies as these networks continue to be mapped.

We also provided a new graph model (DSC) to generate graphs with a desired Rent's exponent, given a module decomposition. This model may be useful when designing synthetic biocircuits (Nielsen et al., 2016) or in other engineered network design applications. If a module decomposition is not available, one can use common network partitioning algorithms (Bassett et al., 2010), though there is no guarantee that these modules will be relevant to actual biological modules. Thus, an important contribution of this study is that Rentian scaling is found in networks decomposed into functional, not topological, modules and that the DSC model can build networks that exhibit Rentian scaling using these modules. While our algorithm requires a module decomposition as input, a more sophisticated generator would also assign proteins to modules during the growth process. How to evolve such modularity is an important problem (Kashtan & Alon, 2005; Clune et al., 2013), but one that we do not consider here.

There may also be multiple mechanisms that can generate networks with Rentian scaling. The DSC graph model under a duplication model can produce Rentian networks. On the other hand, neither the London Underground railroad nor digital chips likely evolved using a duplication model, yet both demonstrate Rentian scaling (Sperry et al., 2017). Thus, while we posit a biological mechanism, we recognize that other mechanisms can produce Rentian scaling. Yet not every network growth mechanism generates Rentian networks. For example, the three social and information networks we studied likely did not evolve using a duplication model, and neither exhibited Rentian scaling. Thus, future work needs to better understand the theoretical basis underlying which mechanisms can produce Rentian networks and which cannot.

It is possible that mechanisms that produce Rentian networks are those constrained by a cost-performance trade-off. Proteins, for instance, must interact with specific partners with cognate structural domains. Proteins can also be sequestered into distinct cellular compartments (e.g., the nucleus),

which limits the number of proteins with which they may interact. Neuronal circuits, similarly, are neither completely randomly wired nor minimally wired (Bassett et al., 2010). A biological network, then, must determine its topology by balancing cost and performance. Rentian scaling might be a by-product of these various constraints (i.e., a spandrel—Rubinov, 2016), and not an evolutionary adaptation. This would be consistent with our finding that Rentian scaling is absent in three social and information networks, where the cost of making edges across any distance within the graph is likely to be much lower than in biological networks. Future work needs to provide a better theoretical basis for this observation.

Finally, how do the Rent's exponents calculated here compare with those observed in digital (e.g., VLSI) circuits? Mid- and large-sized circuits, taken from circuit repositories from the Microelectronics Center of North Carolina and IBM-PLACE (Alpert, 1998), display Rentian scaling with exponents ranging from 0.449 to 0.648 (Yang et al., 2001; Karypis et al., 1999). Another benchmark VLSI circuit, the ISCAS89 s953, had a higher Rent's exponent of 0.730 (Bassett et al., 2010). Overall, these computer circuits have Rent's exponents that are typically smaller than that of the PPI networks studied here. It has been observed that more recent complex circuits have higher Rent values compared to older designs (Bassett et al., 2010). We also observed that the Rent's exponent for the most complex organism studied, *H. sapiens*, was larger than that of the other species (see Table 1), though this may be coincidental.

Overall, the fact that both evolution and the human design process produced Rentian circuits suggests another close correspondence between biological and engineered networks (Navlakha & Bar-Joseph, 2011; Del Vecchio et al., 2016).

Appendix: Supplemental Methods and Results

A.1 Additional Technical Details. Following Bassett et al. (2010), we removed modules that contained more than $N/2$ nodes, where N equals the number of nodes in the network, in order to avoid the Region II boundary effects of Rentian plots, where a single module contains most of the nodes in a network. We also excluded nodes that appeared solely in these larger modules. This Region II has been observed in physical and topological partitions for a network (Christie & Stroobandt, 2000), and in VLSI chips, it refers to the fact that there is a limited number of input-output terminals at the boundary of the chip (Landman & Russo, 1971; Bassett et al., 2010). Nonetheless, it is clear from Figures 3, 7, and 8 that a lower threshold may be needed, as there still appears to be a plateau or turning down of the scaling relationship for modules with a large number of nodes.

In all randomization procedures, we fixed c (the average number of external connections per node) to the value found in the corresponding

empirical network; otherwise, the change in slope was masked by a change in this value.

A.2 Methodological Limitations. One novelty of this study is that we use functional, not topological, modules to partition the networks we studied. This approach, however, comes with some limitations. First, graph partitioning approaches used previously (e.g., hMetis used by Bassett et al., 2010) can be used to study modules over different scales of the topological hierarchy that allow one to study fractal and other forms of scaling (Song, Havlin, & Makse, 2005). We studied a functional hierarchy; in the Gene Ontology, we find modules that vary over one to two orders of magnitude, with larger modules corresponding to broader biological processes (e.g., cellular physiological process) and smaller sized modules corresponding to more specific processes (e.g., pyrimidine metabolic process). The Gene Ontology, however, is not structured as a tree but rather as a directed acyclic graph, which does not provide clear distinctions between levels. Furthermore, there has been extensive prior work in developing graph partitioning methods to extract topological modules from protein interaction networks that can be used to predict protein function (Brohee & van Helden, 2006; Navlakha et al., 2009; Song & Singh, 2009; Sharan et al., 2007); however, success largely depends on the partitioning algorithm used. Thus, it may be difficult to compare insights from topological and functional analyses. Second, nodes (e.g., proteins) in our networks do not have fixed positions in physical space, but rather diffuse more broadly in cellular components. These compartments do provide a coarse physical embedding of nodes, but the embedding is within a limited physical region as opposed to a static position. Thus, further work is needed to relate Rent's exponents to typical notions of wiring economy and embedding when proteins can diffuse within limited physical regions. Third, Rentian scaling was originally developed to study nonoverlapping logic blocks (modules); in using functional modules, a node can be assigned to many modules, which may modify the interpretation of Rent's exponent. Fourth, the networks studied here are binary, whereas many biological networks also have weights indicating the strength or the confidence of interactions.

A.3 Code and Data Availability. All code and data, in both raw and processed form, are available online at <https://github.com/javierhow/Rentian-Scaling-NECO-2018>.

A.4 Data Sets

A.4.1 Protein interaction networks. We downloaded protein-protein interactions (PPIs) from BioGRID version 3.4.141 for six species: *M. musculus*,

S. pombe, *H. sapiens*, *A. thaliana*, *D. melanogaster*, and *S. cerevisiae* (Stark et al., 2006). We removed all interactions that were genetic, between genes in different organisms, or found using AffinityCapture RNA or Protein-RNA assays. We annotated each protein to all significantly enriched Biological Process GO terms using GOTermFinder (Boyle et al., 2004). We excluded annotations that were inferred electronically (IEA annotations). We set a p -value cutoff of 1 to find all associated GO terms for each protein (Boyle et al., 2004). The items on the list of GO terms and their associated p -values were then supplied to REVIGO to group GO terms into semantic clusters, which allowed us to exclude overly redundant modules (Supek et al., 2011). We used three values for the semantic cutoff parameter in REVIGO—0.5, 0.7, and 0.9—in order to test the robustness of our analysis against modules of varying sizes and degrees of overlap. When we did not use REVIGO, for computing the Rent plots, we averaged the number of external edges for every module of the same size, as done before (Stroobandt, 1998).

We also analyzed Rentian scaling in PPIs from the STRING database v10, which collects interactions from several sources and assigns each one a confidence score (Szkarczyk et al., 2015). We kept only interactions that were annotated as “binding” (physical interactions) and had a score greater than or equal to 700 (high confidence). We used FlyBase version FB2016_05 to convert the protein IDs for *D. melanogaster* from the format used by STRING to one usable by GOTermFinder (Attrill et al., 2016). We used bioDBnet’s db2db tool to convert the proteins IDs for *H. sapiens* and *M. musculus* from their Ensembl Protein ID format to Gene Symbol, which could be used in GOTermFinder (Mudunuri, Che, Yi, & Stephens, 2009). We also removed proteins that began with “LOC” and picked the first Gene Symbol when several were provided for one Ensembl Protein ID. We then used GOTermFinder to annotate the nodes, and REVIGO at three different cutoffs to reduce semantic redundancy.

A.4.2 Social and Information Networks. The Amazon network was described in the main text. The DBLP network represented scientific collaborations between authors (nodes) who have coauthored a paper (edges). Here, publication venues (conferences) serve as the modules (communities). These modules are highly overlapping (Yang & Leskovec, 2015), much like the PPI networks, and yet they fail to display Rentian scaling. The YouTube social network was taken from January 2007, where users (nodes) were connected to other users they were friends with (edges). Modules in this network were user-defined groups. For all three networks, only the top 5000 highest-quality modules were used, where each module was assigned a score that corresponded to an average of four goodness metrics that mathematically describe how module-like it was (Yang & Leskovec, 2015).

Table 3: PPIs for Six Species Using Topological Modules Display Poorer Rentian Scaling.

Species	REVIGO	Rent Exponent	R^2
Fly	0.5	0.907	0.78
	0.7	0.882	0.77
	0.9	0.869	0.91
Human	0.5	0.826	0.34
	0.7	1.039	0.25
	0.9	0.943	0.24
Mouse	0.5	0.679	0.98
	0.7	0.744	0.77
	0.9	0.915	0.87
Plant	0.5	0.745	0.80
	0.7	0.753	0.83
	0.9	0.816	0.72
Fission Yeast	0.5	0.667	0.85
	0.7	0.844	0.85
	0.9	0.857	0.84
Baker's Yeast	0.5	0.726	0.67
	0.7	0.964	0.61
	0.9	0.906	0.78

Note: All Rent's exponents are larger and R^2 fits are poorer from topological modules derived from hMetis (v.1.5.3) compared to those derived using functional modules (except for Fission Yeast, where the fit is higher with topological modules).

A.5 Rentian Scaling Using Topological Modules. Table 3 shows the Rent's exponents for modules derived from topological partitions using hMetis.

A.6 Rentian Scaling Using GoTermFinder. Table 4 and Figure 7 show Rentian scaling for all six species using modules defined with GoTermFinder (Boyle et al., 2004).

A.7 Rentian Scaling Using STRING Protein Interactions. Table 5 shows the Rent analysis for PPIs of all six species using high-confidence protein-protein interactions from the STRINGv10 database interactions (Szklarczyk et al., 2015). Figure 8 shows Rent plots for a REVIGO cutoff of 0.7. "High confidence" was defined as having a STRING score ≥ 700 .

A.8 Analysis of Cell-Type-Specific Regulatory Networks. Tables 6 and 7 show the Rentian analysis for all the mouse and human cell-type specific regulatory networks.

Table 4: Rentian Scaling Using All Modules Assigned from GOTermFinder.

Species	Number of Nodes	Number of Modules	Number of Edges	Rent's Exponent	Random Modules	Random Edges	Random Both
Fly	7174	4539	31,948	0.880 (0.93)	0.882 (0.94)	0.884 (0.93)	0.886 (0.94)
Human	13,400	9744	134,638	0.887 (0.84)	0.896 (0.87)	0.894 (0.85)	0.901 (0.87)
Mouse	4159	8971	7659	0.809 (0.88)	0.821 (0.88)	0.820 (0.89)	0.826 (0.89)
Plant	7486	3520	25,715	0.917 (0.90)	0.927 (0.95)	0.932 (0.92)	0.935 (0.95)
Fission yeast	2477	2942	8248	0.846 (0.80)	0.886 (0.89)	0.884 (0.84)	0.898 (0.89)
Baker's yeast	4866	3861	66,749	0.907 (0.83)	0.934 (0.89)	0.930 (0.86)	0.940 (0.88)

Notes: All Rent's exponents are smaller than that of the randomized controls ($p < 0.01$, one-sample two-tailed t -test). Values in parentheses are R^2 .

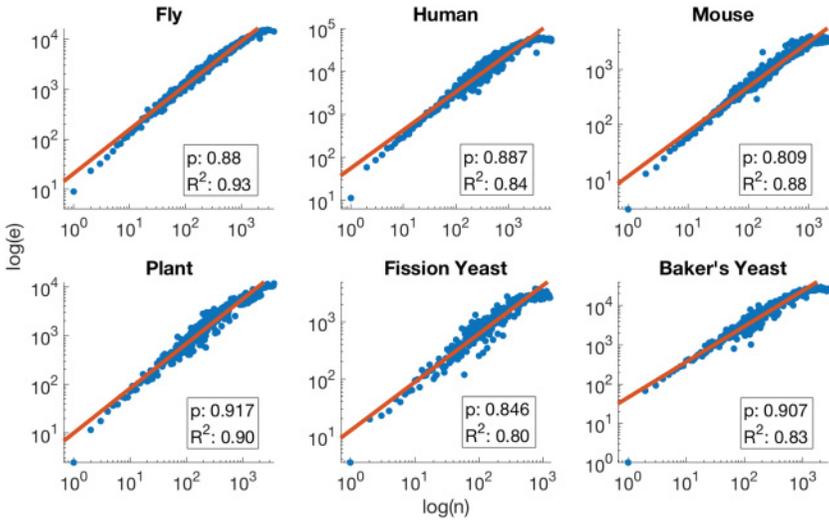


Figure 7: Rentian scaling using GOTermFinder. PPIs from six species display Rentian scaling, even when using overlapping modules defined by several thousand GO terms.

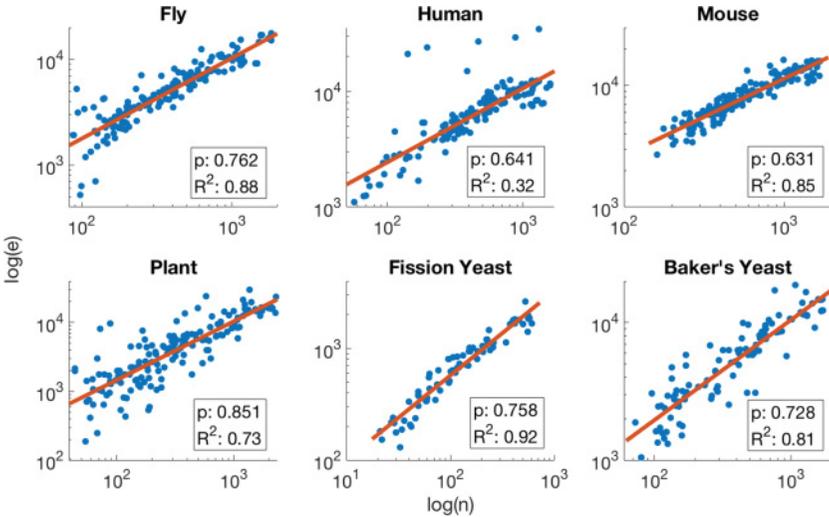


Figure 8: Rentian scaling using STRING. PPIs from six species display Rentian scaling, even when using the high-confidence networks with a score greater than or equal to 700 from the STRING database and a REVIGO cutoff of 0.7. Note the low R^2 for the Human network, which may be due to many missing proteins and interactions when compared to the BioGRID network.

Table 5: PPIs for Six Different Species from the STRING Database Also Display Rentian Scaling.

Species	Number of Nodes	Number of Modules	Number of Edges	REVIGO	Rent Exponent	Random Modules	Random Edges	Random Both
Fly	4010	100	41,052	0.5	0.812 (0.85)	0.864 (0.97)	0.858 (0.91)	0.872 (0.96)
	4055	176	41,186	0.7	0.762 (0.87)	0.808 (0.97)	0.805 (0.92)	0.817 (0.97)
Human	4055	281	41,186	0.9	0.771 (0.87)	0.815 (0.97)	0.813 (0.91)	0.824 (0.97)
	3344	79	81,481	0.5	0.615 (0.22)	0.707 (0.97)	0.710 (0.50)	0.731 (0.98)
	3347	144	81,503	0.7	0.641 (0.32)	0.723 (0.98)	0.735 (0.65)	0.752 (0.98)
Mouse	3347	277	81,503	0.9	0.636 (0.33)	0.710 (0.98)	0.730 (0.62)	0.742 (0.98)
	3530	86	40,871	0.5	0.614 (0.84)	0.645 (0.98)	0.653 (0.95)	0.659 (0.99)
	3551	175	41,097	0.7	0.631 (0.85)	0.661 (0.98)	0.667 (0.95)	0.674 (0.98)
Plant	3552	294	41,100	0.9	0.644 (0.85)	0.670 (0.98)	0.678 (0.95)	0.683 (0.98)
	5404	95	73,304	0.5	0.853 (0.76)	0.941 (0.96)	0.905 (0.82)	0.951 (0.96)
	5404	173	73,304	0.7	0.851 (0.73)	0.939 (0.96)	0.901 (0.83)	0.949 (0.96)
Fission Yeast	5404	283	73,304	0.9	0.839 (0.69)	0.913 (0.96)	0.885 (0.80)	0.924 (0.96)
	1329	60	9333	0.5	0.762 (0.92)	0.903 (0.96)	0.864 (0.94)	0.918 (0.96)
	1329	83	9333	0.7	0.758 (0.92)	0.881 (0.97)	0.852 (0.95)	0.899 (0.97)
Baker's Yeast	1329	121	9333	0.9	0.781 (0.9)	0.895 (0.97)	0.880 (0.95)	0.918 (0.97)
	4034	76	47,175	0.5	0.751 (0.78)	0.831 (0.95)	0.812 (0.88)	0.841 (0.95)
	4043	108	47,355	0.7	0.728 (0.81)	0.806 (0.95)	0.787 (0.90)	0.816 (0.95)
4048	175	47,405	0.9	0.757 (0.81)	0.826 (0.96)	0.818 (0.90)	0.839 (0.95)	

Notes: All Rent's exponents are smaller than that of the randomized controls ($p < 0.01$, one-sample two-tailed t -test). Values in parentheses are R^2 .

Table 6: Mouse Regulatory Networks Demonstrate Rentian Scaling.

Cell line	Number of Nodes	Number of Modules	Number of Edges	Rent Exponent	Random Modules	Random Edges	Random Both
mffBrain-DS14536	503	121	18524	0.795 (0.92)	0.798 (0.94)	0.795 (0.92)	0.798 (0.94)
ZhBtc4-DS15236 ^a	504	119	22248	0.802 (0.94)	0.803 (0.95)	0.801 (0.94)	0.803 (0.95)
ZhBtc4-DS17562 ^a	503	119	18631	0.810 (0.93)	0.813 (0.94)	0.810 (0.93)	0.813 (0.94)
ZhBtc4-DS17616 ^a	503	119	16822	0.819 (0.92)	0.823 (0.94)	0.819 (0.92)	0.822 (0.94)
mCJ7-DS13320 ^a	500	119	13832	0.819 (0.93)	0.821 (0.95)	0.819 (0.93)	0.822 (0.95)
mRetina-DS20004	504	119	15963	0.820 (0.91)	0.824 (0.94)	0.82 (0.91)	0.824 (0.94)
mHeart-DS18138	503	121	18867	0.832 (0.88)	0.840 (0.93)	0.832 (0.88)	0.841 (0.93)
mGenitalFatPad-DS18182	504	119	24360	0.841 (0.90)	0.847 (0.94)	0.841 (0.90)	0.848 (0.94)
mB-DS16168	496	117	15081	0.859 (0.86)	0.867 (0.93)	0.859 (0.87)	0.868 (0.93)
m3134-DS8497 ^b	468	120	13763	0.861 (0.89)	0.870 (0.93)	0.861 (0.89)	0.870 (0.93)
mLung-DS14479	500	120	18258	0.864 (0.88)	0.874 (0.94)	0.864 (0.88)	0.874 (0.94)
mLiver-DS14605	490	120	10586	0.869 (0.85)	0.881 (0.92)	0.868 (0.85)	0.881 (0.92)
mATn-DS17070 ^c	500	120	18768	0.872 (0.87)	0.882 (0.94)	0.872 (0.87)	0.882 (0.94)
mNIH_3T3-DS16900 ^b	491	121	14025	0.877 (0.88)	0.886 (0.93)	0.877 (0.88)	0.886 (0.93)
mKidney-DS13948	498	120	14593	0.878 (0.90)	0.885 (0.94)	0.878 (0.90)	0.886 (0.94)
m416B-DS14099	490	118	12091	0.884 (0.85)	0.895 (0.93)	0.884 (0.85)	0.896 (0.93)
mTN-DS16171 ^c	497	119	15013	0.884 (0.86)	0.896 (0.93)	0.884 (0.86)	0.896 (0.93)
mThymus-DS18819 ^c	463	121	11336	0.885 (0.86)	0.896 (0.93)	0.885 (0.86)	0.896 (0.93)
mTN-DS17080 ^c	492	122	15747	0.887 (0.87)	0.898 (0.93)	0.887 (0.87)	0.898 (0.93)
mB-DS17866	451	117	9279	0.898 (0.84)	0.909 (0.92)	0.899 (0.84)	0.910 (0.92)
MEL-DS13036 ^b	491	120	13449	0.898 (0.86)	0.910 (0.93)	0.899 (0.86)	0.910 (0.93)
mATr-DS20149 ^c	496	118	14184	0.907 (0.85)	0.919 (0.93)	0.907 (0.85)	0.920 (0.93)
mBrain-DS12727	483	120	10182	0.907 (0.87)	0.915 (0.93)	0.906 (0.87)	0.915 (0.93)
mA20-DS16695 ^b	492	122	13382	0.916 (0.87)	0.927 (0.93)	0.916 (0.88)	0.927 (0.94)
mTR-DS17864 ^c	488	119	13577	0.937 (0.84)	0.950 (0.92)	0.937 (0.84)	0.950 (0.92)

Notes: Only results for a REVIGO cutoff of 0.7 are shown. Similar results hold for other REVIGO cutoffs of 0.5 and 0.9. Values in parentheses are R^2 .

^aRows correspond to ESCs.

^bRows correspond to cancer cell lines.

^cRows correspond to immune cells.

Table 7: Human Regulatory Networks Demonstrate Rentian Scaling.

Cell line	Number of Nodes	Number of Modules	Number of Edges	Rent Exponent	Random Modules	Random Edges	Random Both
fLung-DS14724	322	87	6478	0.843 (0.91)	0.856 (0.94)	0.845 (0.91)	0.856 (0.94)
hESC-T0-DS11909 ^a	325	86	6244	0.846 (0.93)	0.854 (0.94)	0.848 (0.93)	0.855 (0.94)
fHeart-DS12531	312	84	5115	0.855 (0.90)	0.866 (0.93)	0.856 (0.90)	0.867 (0.93)
NHDF_Neo-DS11923	314	85	5415	0.859 (0.91)	0.866 (0.94)	0.861 (0.91)	0.867 (0.94)
HFP-DS13390	321	86	5033	0.860 (0.90)	0.875 (0.94)	0.863 (0.90)	0.876 (0.94)
HepG2-DS7764 ^b	298	83	4600	0.865 (0.90)	0.874 (0.93)	0.866 (0.90)	0.874 (0.93)
HEpIC-DS12763	317	85	5101	0.867 (0.89)	0.881 (0.93)	0.870 (0.89)	0.881 (0.93)
NHLF-DS12829	318	86	5145	0.868 (0.90)	0.884 (0.94)	0.870 (0.90)	0.884 (0.94)
HVMF-DS13981	318	86	5356	0.868 (0.91)	0.880 (0.93)	0.870 (0.90)	0.881 (0.93)
HMVEC_LLy-DS13185	314	85	5445	0.869 (0.89)	0.887 (0.93)	0.873 (0.88)	0.886 (0.93)
HFF-DS15115	309	88	4249	0.871 (0.90)	0.883 (0.93)	0.873 (0.90)	0.883 (0.93)
HMVEC_dBlAd-DS13337	312	85	4699	0.872 (0.89)	0.882 (0.93)	0.875 (0.89)	0.883 (0.93)
HMF-DS13368	320	86	4632	0.873 (0.90)	0.885 (0.94)	0.876 (0.89)	0.886 (0.94)
HCF-DS12501	315	86	5090	0.876 (0.89)	0.895 (0.93)	0.878 (0.89)	0.897 (0.93)
HMVEC_dBINeo-DS13242	317	86	6031	0.881 (0.89)	0.895 (0.93)	0.883 (0.89)	0.896 (0.93)
HSMN-DS14426	315	86	4761	0.881 (0.90)	0.895 (0.93)	0.884 (0.90)	0.896 (0.94)
IMR90-DS13219	314	89	3828	0.881 (0.90)	0.895 (0.94)	0.882 (0.90)	0.896 (0.93)
fBrain-DS11872	313	88	4254	0.881 (0.90)	0.894 (0.93)	0.882 (0.90)	0.893 (0.93)
SkMC-DS11949	323	86	6025	0.881 (0.91)	0.895 (0.94)	0.882 (0.91)	0.897 (0.94)
HAEpIC-DS12663	316	87	4620	0.882 (0.89)	0.896 (0.92)	0.884 (0.89)	0.897 (0.92)
HPAF-DS13411	321	83	4715	0.883 (0.88)	0.903 (0.92)	0.884 (0.88)	0.905 (0.92)
HAh-DS15192	322	84	5928	0.883 (0.89)	0.901 (0.92)	0.884 (0.89)	0.902 (0.92)
SAEC-DS10518	310	86	3299	0.883 (0.89)	0.891 (0.93)	0.884 (0.89)	0.892 (0.93)
NHA-DS12800	310	88	3070	0.883 (0.90)	0.897 (0.93)	0.884 (0.90)	0.897 (0.93)
AoAF-DS13513	320	86	4898	0.887 (0.90)	0.902 (0.93)	0.889 (0.89)	0.903 (0.93)
AG10803-DS12374	310	87	4140	0.890 (0.89)	0.903 (0.93)	0.893 (0.89)	0.905 (0.93)

Table 7: Continued.

Cell line	Number of Nodes	Number of Modules	Number of Edges	Rent Exponent	Random Modules	Random Edges	Random Both
HCM-DS12599	317	85	5273	0.893 (0.89)	0.908 (0.93)	0.895 (0.89)	0.908 (0.93)
HMVEC_dlyNeo-DS13150	316	86	5467	0.893 (0.89)	0.912 (0.93)	0.897 (0.89)	0.914 (0.93)
NHDF_Ad-DS12863	318	86	4471	0.893 (0.91)	0.906 (0.94)	0.895 (0.91)	0.907 (0.94)
HIPEpiC-DS12684	316	84	4291	0.899 (0.90)	0.909 (0.93)	0.900 (0.89)	0.909 (0.93)
SKNSH-DS8482 ^b	298	86	4599	0.902 (0.88)	0.919 (0.92)	0.904 (0.88)	0.920 (0.92)
CD34-DS12274	318	82	5894	0.905 (0.89)	0.927 (0.92)	0.908 (0.88)	0.929 (0.93)
CD20-DS18208 ^c	315	85	5920	0.908 (0.89)	0.931 (0.92)	0.912 (0.88)	0.930 (0.93)
HCPepiC-DS12447	319	85	5001	0.910 (0.89)	0.927 (0.92)	0.912 (0.89)	0.926 (0.92)
Gm06990-DS7748 ^c	300	83	4545	0.911 (0.85)	0.937 (0.91)	0.914 (0.85)	0.937 (0.91)
NB4-DS12543	319	86	6632	0.912 (0.90)	0.930 (0.93)	0.915 (0.90)	0.929 (0.93)
HRCe-DS10666	309	87	3083	0.917 (0.89)	0.933 (0.92)	0.921 (0.89)	0.936 (0.92)
HPdLF-DS13573	315	85	4362	0.917 (0.90)	0.927 (0.93)	0.918 (0.90)	0.928 (0.93)
K562-DS9767	296	88	3100	0.920 (0.89)	0.935 (0.93)	0.923 (0.89)	0.936 (0.93)
Gm12865-DS12436 ^c	315	81	5223	0.921 (0.86)	0.949 (0.92)	0.923 (0.86)	0.951 (0.92)

Notes: Only results for a REVIGO cutoff of 0.7 are shown. Similar results hold for other REVIGO cutoffs of 0.5 and 0.9. Values in parentheses are R^2 .

^aRow corresponds to ESCs.

^bRows correspond to cancer cell lines.

^cRows correspond to immune cells.

Acknowledgments

J.J.H. thanks the National Science Foundation Graduate Research Fellowship Program (grant DGE-1650112). S.N. thanks the Army Research Office (grant DOD W911NF-17-1-0045) for funding support.

References

- Alpert, Charles J. (1998). The ISPD98 Circuit Benchmark Suite. In *Proceedings of the 1998 International Symposium on Physical Design* (pp. 80–85). New York: ACM.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 25(1), 25–29.
- Attrill, H., Falls, K., Goodman, J. L., Millburn, G. H., Antonazzo, G., Rey, A. J., . . . Baker, P. (2016). FlyBase: Establishing a gene group resource for *Drosophila melanogaster*. *Nucleic Acids Res.*, 44(D1), D786–792.
- Bader, J. S., Chaudhuri, A., Rothberg, J. M., & Chant, J. (2004). Gaining confidence in high-throughput protein interaction networks. *Nat. Biotechnol.*, 22(1), 78–85.
- Bakoglu, H. B. (1990). *Circuits, interconnections, and packaging for VLSI*. Reading, MA: Addison-Wesley.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.
- Bassett, D. S., Greenfield, D. L., Meyer-Lindenberg, A., Weinberger, D. R., Moore, S. W., & Bullmore, E. T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput. Biol.*, 6(4), e1000748.
- Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., & Sherlock, G. (2004). GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20(18), 3710–3715.
- Brohee, S., & van Helden, J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*, 7, 488.
- Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.*, 13(5), 336–349.
- Chklovskii, D. B., Schikorski, T., & Stevens, C. F. (2002). Wiring optimization in cortical circuits. *Neuron*, 34(3), 341–347.
- Christie, P., & Stroobandt, D. (2000). The interpretation and application of Rent's rule. *IEEE Transactions on Very Large Scale Integration Systems*, 8(6), 639–648.
- Clune, J., Mouret, J.-B., & Lipson, H. (2013). The evolutionary origins of modularity. *Proc. R. Soc. B*, 1755, 20122863.
- Davis, D., Yaveroglu, O. N., Malod-Dognin, N., Stojmirovic, A., & Przulj, N. (2015). Topology-function conservation in protein-protein interaction networks. *Bioinformatics*, 31(10), 1632–1639.
- Deeds, E. J., Krivine, J., Feret, J., Danos, V., & Fontana, W. (2012). Combinatorial complexity and compositional drift in protein interaction networks. *PLoS ONE*, 7(3), 1–14.

- Del Vecchio, D., Densmore, D., El-Samad, H., Ingber, D., Khalil, A., Kosuri, S., . . . Tang, C. (2016). What have the principles of engineering taught us about biological systems? *Cell Syst*, 2(1), 5–7.
- Dillin, A., Gottschling, D. E., & Nystrom, T. (2014). The good and the bad of being connected: The integrons of aging. *Curr. Opin. Cell Biol.*, 26, 107–112.
- Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, 30(7), 1575–1584.
- Gitter, A., Klein-Seetharaman, J., Gupta, A., & Bar-Joseph, Z. (2011). Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Res.*, 39(4), e22.
- Huang, H., & Bader, J. S. (2009). Precision and recall estimates for two-hybrid screens. *Bioinformatics*, 25(3), 372–378.
- Jiang, P., & Singh, M. (2010). SPiCi: A fast clustering algorithm for large biological networks. *Bioinformatics*, 26(8), 1105–1111.
- Karypis, G., Aggarwal, R., Kumar, V., & Shekhar, S. (1999). Multilevel hypergraph partitioning: Applications in VLSI domain. *IEEE Transactions on Very Large Scale Integration Systems*, 7(1), 69–79.
- Kashtan, N., & Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 13773–13778.
- Klimm, F., Bassett, D. S., Carlson, J. M., & Mucha, P. J. (2014). Resolving structural variability in network models and the brain. *PLoS Comput. Biol.*, 10(3), e1003491.
- Landman, B. S., & Russo, R. L. (1971). On a pin versus block relationship for partitions of logic graphs. *IEEE Trans. Comput.*, 20(12), 1469–1479.
- Lanzerotti, M. Y., Fiorenza, G., & Rand, R. A. (2005). Microminiature packaging and integrated circuitry: The work of E. F. Rent, with an application to on-chip interconnection requirements. *IBM J. Res. Dev.*, 49(4/5), 777–803.
- Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proc. 11th Intl. ACM Conf. on Knowledge Discovery and Data Mining* (pp. 177–187). New York: ACM.
- Loh, K. M., & Lim, B. (2011). A precarious balance: Pluripotency factors as lineage specifiers. *Cell Stem Cell*, 8(4), 363–369.
- Mason, O., & Verwoerd, M. (2007). Graph theory and networks in biology. *IET Systems Biology*, 1(2), 89–119.
- Middendorf, M., Ziv, E., & Wiggins, C. H. (2005). Inferring network mechanisms: The *Drosophila melanogaster* protein interaction network. *Proc. Natl. Acad. Sci. U.S.A.*, 102(9), 3192–3197.
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proc. of the 7th ACM SIGCOMM Conf. on Internet Measurement* (pp. 29–42). New York: ACM.
- Mudunuri, U., Che, A., Yi, M., & Stephens, R. M. (2009). bioDBnet: The biological database network. *Bioinformatics*, 25(4), 555–556.
- Nasvall, J., Sun, L., Roth, J. R., & Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. *Science*, 338(6105), 384–387.
- Navlakha, S., & Bar-Joseph, Z. (2011). Algorithms in nature: The convergence of systems biology and computational thinking. *Mol. Syst. Biol.*, 7, 546.

- Navlakha, S., Faloutsos, C., & Bar-Joseph, Z. (2015). Massexodus: Modeling evolving networks in harsh environments. *Data Mining and Knowledge Discovery*, 29(5), 1211–1232.
- Navlakha, S., Gitter, A., & Bar-Joseph, Z. (2012). A network-based approach for predicting missing pathway interactions. *PLoS Comput. Biol.*, 8(8), e1002640.
- Navlakha, S., He, X., Faloutsos, C., & Bar-Joseph, Z. (2014). Topological properties of robust biological and computational networks. *J. R. Soc Interface*, 11(96), 20140283.
- Navlakha, S., & Kingsford, C. (2011). Network archaeology: Uncovering ancient networks from present-day interactions. *PLoS Comput. Biol.*, 7(4), e1001119.
- Navlakha, S., Schatz, M. C., & Kingsford, C. (2009). Revealing biological modules via graph summarization. *J. Comput. Biol.*, 16(2), 253–264.
- Neph, S., Stergachis, A. B., Reynolds, A., Sandstrom, R., Borenstein, E., & Stamatoyannopoulos, J. A. (2012). Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, 150(6), 1274–1286.
- Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, . . . Voigt, C. A., (2016). Genetic circuit design automation. *Science*, 352(6281), aac7341.
- Pereira-Leal, J. B., & Teichmann, S. A. (2005). Novel specificities emerge by stepwise duplication of functional modules. *Genome Res.*, 15(4), 552–559.
- Perez-Escudero, A., & de Polavieja, G. G. (2007). Optimally wired subnetwork determines neuroanatomy of *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.*, 104(43), 17180–17185.
- Pinkert, S., Schultz, J., & Reichardt, J. (2010). Protein interaction networks—more than mere modules. *PLoS Comput. Biol.*, 6(1), e1000659.
- Ramón y Cajal, S. (1899). *Textura del Sistema Nervioso del Hombre y Vertebrados*, vol. 2. Nicolas Moya.
- Rivera-Alba, M., Peng, H., de Polavieja, G. G., & Chklovskii, D. B. (2014). Wiring economy can account for cell body placement across species and brain areas. *Curr. Biol.*, 24(3), R109–R110.
- Rubinov, M. (2016). Constraints and spandrels of interareal connectomes. *Nature Communications*, 7, 13812.
- Schlicker, A., Domingues, F. S., Rahnenfuhrer, J., & Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, 7, 302.
- Sharan, R., Ulitsky, I., & Shamir, R. (2007). Network-based prediction of protein function. *Mol. Syst. Biol.*, 3, 88.
- Song, C., Havlin, S., & Makse, H. A. (2005). Self-similarity of complex networks. *Nature*, 433(7024), 392–395.
- Song, J., & Singh, M. (2009). How and when should interactome-derived clusters be used to predict functional modules and protein function? *Bioinformatics*, 25(23), 3143–3150.
- Sperry, M. M., Telesford, Q. K., Klimm, F., & Bassett, D. S. (2017). Rentian scaling for the measurement of optimal embedding of complex networks into physical space. *Journal of Complex Networks*, 5, 199–218.
- Spirin, V., & Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U.S.A.*, 100(21), 12123–12128.

- Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.*, 34(Database issue), D535–D539.
- Stearns, F. W. (2010). One hundred years of pleiotropy: A retrospective. *Genetics*, 186(3), 767–773.
- Stergachis, A. B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A. P., Zhang, M., . . . Stamatoyannopoulos, J. A. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature*, 515(7527), 365–370.
- Stevens, C. F. (2009). Darwin and Huxley revisited: The origin of allometry. *J. Biol.*, 8(2), 14.
- Stroobandt, D. (1998). *Analytical methods for a priori wire length estimates in computer systems*. 1998. Ph.D. dissertation, University of Ghent.
- Stroobandt, D. (2007). Rent's rule coincidence or the result of the design process? In *Proc. of the IEEE System Level Interconnect Prediction Workshop*. Piscataway, NJ: IEEE.
- Supek, F., Bosnjak, M., Skunca, N., & Smuc, T. (2011). REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE*, 6(7), e21800.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, 43(database issue), D447–D452.
- Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., & Chklovskii, D. B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.*, 7(2), e1001066.
- Vázquez, A., Flammini, A., Maritan, A., & Vespignani, A. (2003). Modeling of protein interaction networks. *ComplexUs*, 1, 38–44.
- Wagner, A. (2001). The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Molecular Biology and Evolution*, 18(7), 1283–1292.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442.
- Yang, J., & Leskovec, J. (2015). Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1), 181–213.
- Yang, X., Bozorgzadeh, E., & Sarrafzadeh, M. (2001). Wirelength estimation based on rent exponents of partitioning and placement. *Proc. of the Intl. Workshop on System-Level Interconnect Prediction* (pp. 25–31). New York: ACM.
- Zinman, G. E., Zhong, S., & Bar-Joseph, Z. (2011). Biological interaction networks are conserved at the module level. *BMC Syst. Biol.*, 5, 134.