Chaos, Solitons and Fractals 164 (2022) 112707

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Impact of finite size effect on applicability of generalized fractal and spectral dimensions to biological networks

Adam Craig [a,*], Mesut Yücel [b], Lev Muchnik [c,d], Uri Hershberg [e,**]

[a] School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, PA, USA
[b] GlaxoSmithKline, London, UK
[c] School of Business Administration, The Hebrew University of Jerusalem, Jerusalem, Israel
[d] Microsoft Research Israel, Alan Turing 3, Hertzliya, Israel
[e] Department of Human Biology, University of Haifa, Haifa, Israel

ARTICLE INFO

Keywords:
Biological networks
Fractal dimension
Spectral dimension
Multifractal analysis
Multi-spectral analysis
Memory-biased random walk

ABSTRACT

In recent years, researchers have explored fractal dimension, spectral dimension, and multifractal analysis as ways of describing the emergent hierarchical structure of complex networks. However, fractality implies an infinite recursion that is impossible for finite networks describing real-world biological systems. We show that there is a substantial finite size effect on two widely used empirical methods (box-covering and sandbox) of estimating generalized fractal dimensions. As a partial solution to this issue we introduce here a generalized method for calculating network spectral dimension using a memory-biased random walk (MBRW). To observe the impact of network size, we start with an ensemble of networks representing a variety of biological systems, identify their community structures using Infomap, and use a modified stochastic block model to generate networks with similar community structure but varying size. We find that, compared to shortest-path-based generalized fractal dimension methods, the MBRW generalized spectral dimension (D_q) shows a clearer and more consistent ordering of networks by community structure for all orders (q) considered. We also find that, among the measures of multifractality and multispectrality, only MBRW multi-spectrality (range of D_q values) changes in a consistent direction under randomization of each level of community structure. Our results show that network size is an important consideration when comparing the fractal or spectral dimensions of real-world networks and that observing the interaction between network structure and an agent acting in time with memory provides insights into network structure not available through calculations based on purely topological features.

1. Introduction

1.1. Time and memory in networks

Using networks to represent interactions and relationships in real-world systems is a widely accepted and useful methodology for understanding and analysis [3]. Many common approaches to looking at a network rely primarily on static topological features of the network, such as degree distribution and shortest paths [3]. However, it is also possible to analyze dynamics of the processes occurring on networks, which is particularly likely to provide new insights when the network represents a set of dynamic underlying processes, only some subset of which are happening at any one time [4]. Clusters of entities whose interactions occur close together in time often constitute functional units or other communities, such as ganglia in the nervous system [28] or complexes of proteins [13]. Furthermore, many such systems feature memory in the sense that components that recently interacted are more likely to interact again in the immediate future. As examples, neuronal circuits often contain feedback loops such that activity in a certain part of the network is self-sustaining over a short period of time, even without additional stimuli [16], and post-translational modifications of proteins provide a form of memory that modifies which proteins in a protein-protein interaction network can interact [7]. While accurately modeling the dynamics of such networks is often a complex and application-specific task, random walks provide a simple and broadly applicable way of observing how network topology influences sequences

* Correspondence to: A. Craig, School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, PA, USA.
** Correspondence to: U. Hershberg, Department of Human Biology, University of Haifa, Haifa, Israel
E-mail addresses: agcraig@hkbu.edu.hk (A. Craig), uri@sci.haifa.ac.il (U. Hershberg).

https://doi.org/10.1016/j.chaos.2022.112707
Received 14 April 2022; Received in revised form 29 August 2022; Accepted 14 September 2022
Available online 28 September 2022

of interactions. In particular, researchers have previously made use of the ability of communities to slow the progression of a random walk agent through the network in order to develop random-walk-based community detection algorithms, including the widely used Infomap algorithm [23,27,29]. However, Yücel and Hershberg showed that a simple random walk agent is often more heavily influenced by local-scale features, particularly hubs, compared to a memory-biased random walk (MBRW) agent, which is primarily influenced by community structure [36].

1.2. Fractal and dimension and network size

In precent years, researchers have made a variety of claims about network fractal dimension as a measure of macro-scale properties of the network, such as robustness or susceptability to attack [33,34]. This concept of network fractality arises from an analogy with the fractality of shapes embedded in a Euclidean coordinate space, such as an outline in a plane, which intuitively corresponds to roughness in the sense of nested levels of detail: the higher the fractal dimension, the faster the perimeter of the shape grows as the resolution with which we measure that outline increases [22]. Starting with [31], researchers have tried to extend this notion of levels of nested detail to networks of nodes and edges not embedded in a coordinate space through an analogy between measurement of the perimeter of a rough shape with a straight edge and renormalization of a graph by collapsing all nodes within a certain shortest-path distance of some central node into a single box node [31]. Prior work observing that the box-counting fractal dimension of a scale-free network decreases as the network grows interpreted this as indicating decreasing robustness of the network [33]. However, this raises the question of at what size the concept of fractality even becomes applicable to a finite real-world network and whether such observations could reflect an artificially inflated estimate of fractality in smaller networks, where the reality is further from the ideal of infinitely nested structures implied by the concept of fractality. Furthermore, whereas attention has previously focused on fractality as a measure of the vulnerability or robustness of entire networks, such as the fragility of communication networks or the ease with which a virus can travel through a contact network, and on the importance of individual nodes, relatively little prior work has studied the relationship between network fractality and community structure [33]. More fundamentally, the concept of fractality in metric spaces implies an infinite nesting of structure within structure, while networks describing real-world systems consist of finite numbers of discrete nodes and edges, often in the lower hundreds or thousands, small enough that the assumption of infinite size seems implausible, raising the question of how applicable the concept of fractal dimension is to small real-world networks and how comparable results are for networks of different sizes. For example, prior research has found that scale-free networks with the same degree distribution have monotonically decreasing fractal dimension as the size increases, which was then interpreted as showing that larger networks were less robust [33]. However, it is possible that the difference is an artifact of a finite-size effect. In this work, we set out to explore the relationship between fractal dimension, network size, and community structure.

1.3. Generalized MBRW spectral dimension and community structure

The common methods of finding the fractal dimension of a network place networks into a box or cluster based on shortest path distance from some seed node [33]. In many networks, however, agents and information do not always travel by the shortest path. This is particularly true for networks that reflect sequences of interactions, or shaped by diffusion of some entity or a dynamic process that leverages existing relations, like the networks commonly used to represent biological systems. The idea of a dimension associated with fractality that governs diffusion originated with [2] and has since evolved into the concept of the spectral dimension [6]. However, prior works defined spectral dimension in terms of a simple random walk. We here introduce a new version of spectral dimension based on the rate at which an MBRW agent explores the network and hypothesize that it will bear a clearer relationship to community structure than does either simple random walk spectral dimension or any shortest-path-based fractal dimension.

Additionally, real-world networks are heterogeneous, with different regions having different levels of interconnectedness. Several works have used generalized fractal dimensions, ones that have a tunable parameter, generally an exponent, that one can adjust to give more emphasis to denser or more sparse parts of the network and used the range of this generalized dimension over a given domain of the parameter as a measure of network heterogeneity [10,20,21,32,33]. In order to achieve this same ability to study network heterogeneity with spectral dimension, we propose a generalized MBRW spectral dimension. We start with the formula $S_N \sim N^{\mathrm{d}/2}$ from [25], where d is the spectral dimension, N is the length of a random walk on the network, and S_N is the expected number of distinct nodes visited in the walk. We replace the simple random walk with an MBRW and replace the simple mean with the generalized mean (See [5]). The exponent of the power mean serves as the tunable parameter giving more weight to the larger or smaller values [5], here the walk segments covering more or fewer nodes, so that the range of generalized spectral dimension values over a sufficiently wide domain of positive and negative exponents (the MBRW multi-spectrality) can serve as a measure of heterogeneity of community structure.

To test how strongly MBRW spectral dimension and multi-spectrality relate to the kinds of community structures found in real-world biological networks, we tested them with synapse, protein-protein interaction (PPI), and metabolic networks. To test the relative impacts of network size and structure on the generalized MBRW spectral dimension and two shortest-path-based approaches to calculating generalized fractal dimensions, we compared values across a set of artificial networks with community structures modeled on those of the real-world networks but rescaled to different sizes. We found that all the dimensions showed a pronounced finite size effect that only became small relative to differences in community structure at sizes much larger than those of the real-world networks. We also compared the impacts of different scales of community structure on MBRW multi-spectrality and two shortest-path-based multifractality measures by comparing values of the real-world networks to those of partially and fully randomized counterparts with the same size and mean degree. We found that only MBRW multi-spectrality changed in a consistent way under each randomization, increasing as the randomness of the network increased and showed relatively low variability within most sets of networks generated by applying the same randomization method to the same original network. These results show that memory keeps the generalized MBRW spectral dimension strongly tied to community structure across all values of its power parameter, making it more effective as a way of investigating the heterogeneity of network structure.

2. Methods

2.1. Memory-biased random walk

Whereas a conventional random walk agent on an unweighted, undirected network has an equal chance of moving from its current location to any adjacent node, the MBRW agent computes new weights for all possible transitions at each time step [37]. It has two parameters, a bias, α, and memory length, s (Ibid.). An additional rule forbids backtracking along the edge by which the agent arrived at its current location to prevent memory from causing it to pass back and forth repeatedly between the same two nodes, a behavior that would not tell us anything about community structure (Ibid.). The probability of any other transition it has taken within the past s steps, that is, a remembered transition, is α times higher than the probability of any other transition(i.e., those

not taken recently). For example, if the current node has 10 edges, arrived by one edge, and has left via two of those edges on previous visits within memory, then the transition back to the previous node has a probability of 0, each of the 2 remembered transitions has a probability of $\alpha/(2\alpha+7)$, and each of the 7 unremembered transitions has a probability of $1/(2\alpha+7)$. Previously, Yücel et al. found that the community-seeking effect of MBRW holds over a wide range of biases and memory lengths (Ibid.). In selecting parameters, we started with the values $\alpha =$ 1000 and $s = 5$, shown to work well for community detection in artificial networks in [38], and sampled the effect of parameter choice on the MBRW spectral dimension for α of 1000, 5000, 10,000, 50,000, and 100,000 and s of 3 (the minimum for which memory can have any effect), 5, 10, 50, 100, 500, 1000, and 5000 by comparing the values on original and fully randomized networks. Using a larger value of α increases the difference between the spectral dimension (D_1) values of real and randomized networks but at the cost of needing to run the simulation longer in order to adequately cover the network. We selected $\alpha =$ 10,000, as it provided an adequate balance between these two concerns. As reported in (Ibid.), even with longer memory, the MBRW agent spends most of its time in short loops, rarely longer than 5 steps. To check that our results were not sensitive to memory length, we repeated all simulations with both $s = 5$ and $s = 500$. The results were qualitatively similar but showed a slightly stronger finite-size effect with shorter memory. Since the goal of our tests with rescaled networks is to observe the finite-size effect, we show results for $s = 5$ (See Fig. 3 in Results). However, when examining the effect of randomizing different scales of community structure on multi-spectrality, we considered finite size effect a noise factor and so display results for $s = 500$ (See Fig. 6 in Results). For an example of a side-by-side comparison of results with short and long memory, see the plots of effect of randomization on spectral dimension in Figs. S3–S26 vs S27–S49 and S118 vs S119 in Supplementary Materials.

2.2. Spectral dimension and multi-spectral analysis

Spectral dimension, d, is a measure of the relationship between scale and rate of diffusion through an infinite network, which Rammal and Toulouse showed related N, the length of a random walk on the network, to S_N, the expected number of distinct nodes visited in the walk, by the power-law $S_N \sim N^{d/2}$ when $d < 2$ and linearly ($S_N \sim N$) when $d > 2$ [25]. We build on this approach by using this to empirically compute an approximate spectral dimension, D_1, for a finite network using either simple random walk or MBRW. We then create a generalized spectral dimension D_q by using the power mean for a power q (q-mean, defined for a set of values $X = \{x_1, \ldots, x_N\}$ as $\left(\frac{1}{N}\sum_{i=1}^{N} x_i^q\right)^{1/q}$) as the expected number [5]. The use of the power q allows us to tune the sensitivity of the generalized spectral dimension D_q to different parts of the network structure. Larger positive values give more weight to the walk segments that visit the most unique nodes and thus to less modular portions of the network, where the MBRW agent can explore more freely, while larger negative values give more weight to the walk segments that visit the fewest unique nodes and thus to the most modular portions of the network, which most strongly trap the MBRW agent. Our method is as follows:

1. Run a random walk simulation on the network for a set number of steps, T. In initial tests, we found that results were not sensitive to the total number of steps used so long as it was sufficient to visit every node on the network multiple times. We chose $T = 2^{27}$ based on practical limitations of available computing hardware, and, in all runs, we found that either length was sufficient for the walk agent to cover even the largest networks used in this work multiple times (as shown by how the y-coordinates of the circles plateau in Fig. 1). When we did try a larger value, $T = 2^{30}$, the results were visually identical. For example, out of all rescaled networks, the largest

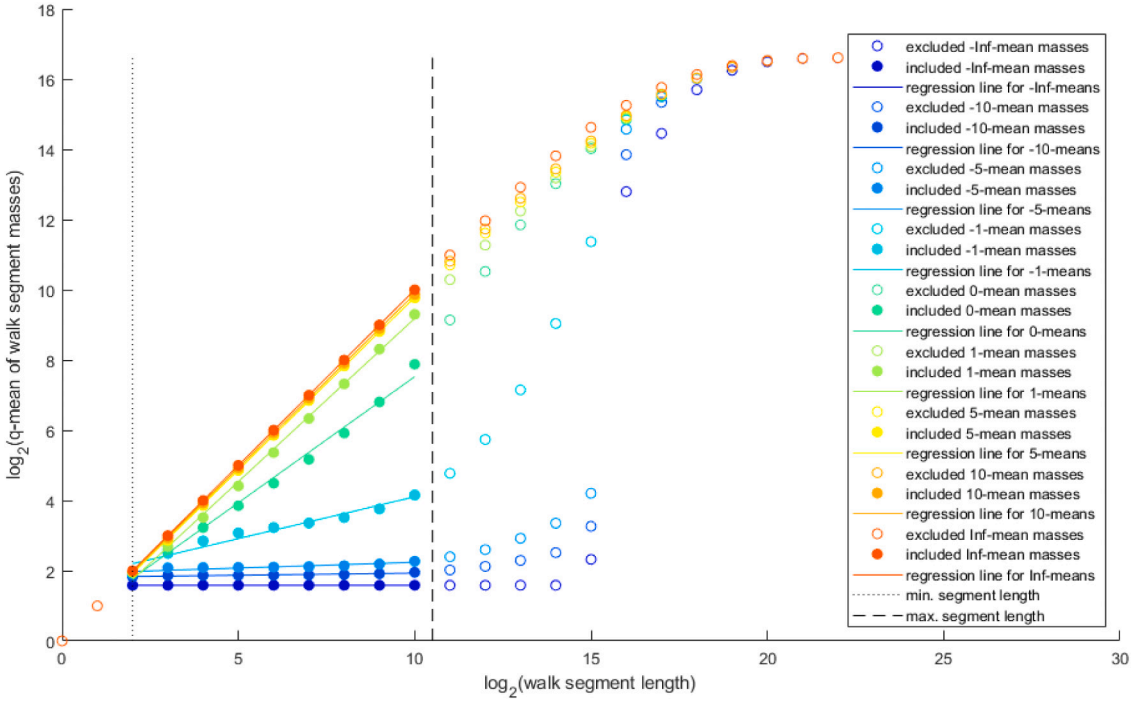

Fig. 1. log-log plot of q-mean segment mass vs segment length for MBRW (bias = 10,000, memory = 5) segments on a version of the *D. melanogaster* PPI network rescaled to 100,000 nodes and selected q values of -infinity, −10, −5, −1, 0, 1, 5, 10, and infinity. The vertical lines indicate constraints placed on the segment length, in this case between 4 and 1453, the number of nodes in the original *D. melanogaster* PPI network. Filled circles represent (length, mass) pairs within the desired range while empty circles represent pairs outside of it. Each line represents a linear regression on the in-range points of the same color. Twice the slope of the regression line is our estimate of the MBRW generalized spectral dimension, D_q.

absolute difference between two corresponding D_q values (bias = 10,000, memory = 5, one calculated with $T = 2^{27}$, the other calculated with $T = 2^{30}$) was 10.6 %, and the mean absolute difference was 0.42 %. This stage produces a long sequence of nodes visited by the memory-biased random walk.

2. For each random walk segment length N that is a power of 2 greater than or equal to some minimum value N_{min} (shown by the black dotted line in Fig. 1) and less than some maximum value N_{max} (shown by the black dashed line in Fig. 1), split the walk into segments of length N, and compute $S_N(q)$, the q-mean number of unique nodes visited in a segment (the filled circles in Fig. 1). We chose N_{min} to be the shortest walk length for which the rules of the walk do not pre-determine the number of distinct nodes in a segment. For a simple random walk this is $N_{min} = 2$, whereas, for MBRW, due to the no-backtracking rule $N_{min} = 4$. In general, we found that using a single value of N_{max} for all networks being compared facilitates comparison. It should be a value well before that at which the mean segment mass plateaus due to the finite size of the network. For each comparison of a real-world network to a set of rescaled or randomized versions, we set N_{max} to the number of nodes in the original network.
3. Use linear regression to find D_q and b such that $log_2(S_N(q)) \approx b + (D_1/2)log_2(N)$ (illustrated by the lines in Fig. 1). Take D_q as our estimate of the generalized simple random walk or MBRW spectral dimension.

To measure the multi-spectrality of the network, we computed$\Delta D_q := \max_{q\in Q} D_q - \min_{q\in Q} D_q$, that is, the range of D_q over all q in the domain $Q = \{-10, -9, \ldots, 9, 10\}$. We chose these bounds of q in keeping with the convention adopted in [10,20,21] for computing the box-covering and sandbox multifractalities. In most cases, we found that the slope of the log-scale 10-mean of the segment mass as a function of length is often close to that of the Inf-mean (maximum value), while that of the −10-mean was close to that of the -Inf-mean, suggesting that this range encompases most of the variability of D_q. See Fig. 1 for an illustrative example. For further discussion of this choice, see section S5 in Supplementary Materials.

2.3. Box-covering and sandbox multifractal methods

We compare MBRW spectral dimention to two methods frequently used for calculating the generalized fractal dimension of a network: the box-covering method [20] and sandbox method [21]. The key difference between them is that boxes in the box-covering method are disjoint, whereas they are overlapping in the sandbox method. For each, we considered both this individual D_q value and the multifractality as measured by ΔD_q, the range of D_q over integer values of q from −10 to 10.

To compute the generalized box-covering fractal dimensions of a network, we start by randomly generating 1000 coverings of the network with boxes of radius r and choosing the one that covers the network with the fewest boxes for each r from 1 to some maximum [20]. Li et al. do not specify a method for selecting the cutoff radius (Ibid.). When comparing the original network to its randomizations, all networks are the same size, so we use one less than the network diameter as the cutoff. For the rescaled networks, we use one less than the smallest diameter (4) for all networks (See section 2.9). For $q = 0$, the box-covering generalized fractal dimension reduces to the box-counting fractal dimension (Ibid.).

Unlike the box-covering method, the sandbox method uses overlapping boxes, eliminating the computational complexity of computing a covering of disjoint boxes [21]. In our implementation, we use every node in the network as the center of a sandbox instead of just a random sampling of them. As with the box-covering method, we use all radii from 1 to one less than the diameter of the network when comparing the original network and its same-size randomizations and 1 to 3 for the rescaled networks and, for the rescaled networks, we use radii from 1 to one less than the smallest network diameter.

2.4. Real-world biological networks

To test our methods' applicability, we applied them to 12 biological networks of three types: synapse wiring diagrams, protein-protein interaction networks (PPIs), and metabolic networks. We analyzed 3 different attempts to map the neuronal network of *C. elegans* [1,18,28,35], the proteome-scale PPIs of *D. melanogaster* [14] and *H. sapiens* [26], the embryogenesis gene-focused PPI of *C. elegans* [13] and interferon signalling gene-focused PPI of *H. sapiens* [19], and five alternate reconstructions of the metabolic network of *A. thaliana* with different criteria for inclusion of edges [24]. See Table S1 in Supplementary Materials.

2.5. Preprocessing

In order to use MBRW on these networks, we must first pre-process them in order to obtain versions that the walk agent can completely traverse without reaching a dead-end. Additionally, although some of the selected networks, such as Human IIS, are directed, we treat all networks as undirected so that we can better compare results across different networks. Because of the MBRW agent's no-backtracking rule, it will have no valid moves whenever it reaches a node with degree 1 (a dead-end). To prevent this, we take the 2-core of each network [30], iteratively removing each dead-end until none remain. Note that this may require multiple passes over all nodes, since removing a node of degree one decreases the degree of the node adjacent to it, possibly dropping it to 1. After taking the 2-core, we remove any components not connected to the largest connected component so that the agent can cover the entire pre-processed network. After preprocessing, the network sizes ranged from 116 to 2880 (See Table S2 in Supplementary Materials.). We used Infomap [8] to detect the community structure of each network and found that these networks displayed diverse community structure, with some, such as the *C. elegans* synapse networks, dominated by a small number of large communities and others, such as the interactome scale Human and *D. melanogaster* PPI networks, consisting of a large number of small communities, though all showed substantial community structure with modularities ranging from 0.35 (*C. elegans* neural 1) to 0.74 (*D. melanogaster* PPI) (See Table S3 in Supplementary Materials.).

2.6. Rescaled networks

As a way of testing the effect of network size on our measures, we created ensembles of rescaled networks, each modeled on one of the real-world networks. The goal was to create networks of varying size while retaining the same mean degree and proportion of edges that were intra-community edges (intra-edges) within each community and of inter-community edges (inter-edges) that connected each pair of communities. We based our rescaled network model on the stochastic block model [17] but modified it in order to ensure that the resulting networks are MBRW-traversable. Our modified stochastic block model algorithm is as follows:

1. Let N_1 be the size of the original network, G_1, and N_2 be the target size of the rescaled network, G_2.
2. Let M_1 be the number of edges in G_1, and set the target number of edges in G_2 to $M_2 = \text{round}(M_1N_2/N_1)$.
3. Use a community detection algorithm (Infomap [8] in this case) to partition G_1 into discrete communities with sizes $n_{1,1}$, $n_{1,2}$, …, $n_{1,r}$.
4. For each community i in G_1, create a community in G_2 of size $n_{2,i} = \max(3, \text{round}(n_{1,i}N_2/N_1))$.

5. For each community i in G_1, let $m_{1,i,i}$ be the number of intra-community edges in i, and set the target number of intra-community edges in the corresponding community in G_2 to $m_{2,i,i}$ = ceiling(max(min($n_{2,i}$($n_{2,i}$-1)/2, $m_{1,i,i}M_2/M_1$), $n_{2,i}$)).
6. For each community in G_2, connect all the nodes in a cycle. For instance, if community i has nodes numbered a through $a + n_{2,i}$-1, then add edges ($a + 1$,$a + 2$), ($a + 2$,$a + 3$), …, ($a + n_{2,i}$-2,$a + n_{2,i}$-1), and ($a + 1$,$a + n_{2,i}$-1).
7. Randomly choose whether to connect each pair of nodes not connected in the cycle with probability ($m_{2,i,i}$-$n_{2,i}$)/($n_{2,i}$($n_{2,i}$-1)/2-$n_{2,i}$).
8. For each pair of communities s and t in G_1, let $m_{1,s,t}$ be the number of inter-community edges between s and t, and set the target number of inter-community edges between the corresponding communities in G_2 to $m_{2,s,t}$ = ceiling(max(min($n_{2,s}n_{2,t}$, $m_{1,s,t}M_2$/M_1), $n_{2,i}$)).
9. For each pair of communities s and t in G_2, if $m_{2,s,t} > 0$, then connect the respective first (lowest index) nodes of the two communities.
10. For each pair of communities s and t in G_2, if $m_{2,s,t} > 0$, then randomly choose whether to connect each pair of nodes $i \in s$ and $j \in t$ such that i and j are not both the first nodes of their respective communities with probability ($m_{2,s,t}$-1)/($n_{2,s}n_{2,t}$-1).

For a visual representation of the overall logic of the algorithm, see Fig. S2 in Supplementary Materials.

For each pair of one of the 12 real-world networks and one of the 100 target network sizes 1000, 2000, …, 100,000, we generated 10 instances. Over 99 % of networks had numbers of nodes and edges that both differed from the target numbers by <5 %. We then computed the MBRW (bias = 10,000, memory = 5, 10 runs per network) generalized spectral dimensions and box covering (10 runs, each with 1000 iterations) and sandbox (using every node as a sandbox center as described above) generalized fractal dimensions of these networks.

2.7. Same-size fully and partially randomized networks

Whereas comparing rescaled networks provided a way to test the effect of network size while keeping community structure consistent, we subsequently used randomization to test the ability of each measure to distinguish networks with community structure from random networks of equivalent size and density. Specifically, we posited that a measure that reflects the community structure of a network should change in a consistent direction when randomization removes some or all of that structure and that any such change should be larger than variation due to small-scale random differences within a set of networks that are similar in size and community structure. Our partial randomization methods take a partitioning of the network into disjoint communities and preserve the total number of inter-edges and the number of intra-edges within each community. We used Infomap using the flag for a 2-level partition of the network and all others settings left at their default values [8], as this provided communities that were generally large enough to have meaningful intra-community structure as well as structure in the connections among communities.

For each real-world network (Original), we generate 4 types of randomized network (Fig. 2):

1. Inter-Edges: Preserve edges between communities (inter-edges) while replacing each edge between nodes of the same community (intra-edge) with a randomly generated intra-edge within the same community.
2. Intra-Edges: Preserve intra-edges while replacing inter-edges with an equal number of randomly generated inter-edges.
3. Communities: Replace each intra-edge with a randomly generated intra-edge in the same community, and replace each inter-edge with a randomly generated inter-edge. This method preserves communities, while changing connections within them and alters connections between communities.
4. Mean Degree: Replace each edge with a randomly generated edge without regard for community membership. This method preserve the number of the network edges thus keeping the average network degree unchanged, but destroying the network structure.

Using Erdős and Rényi's method or Gilbert's method for randomly generating Erdős-Rényi networks could have generated dead-ends or unconnected components [9,12], which we would then need to prune, changing the size and mean degree. Instead, we developed an algorithm that reliably preserves the numbers of nodes and edges, minimum degree of 2, and connectedness, Randomize-to-Equilibrium:

1. Save a copy of E in E_0.
2. Set $\Delta_{max} = 0$.
3. Iterate over all edges in the graph $G = (V, E)$.
4. Let the current edge be $\{A, B\} \in E$.
5. Randomly select a second edge $\{C, D\} \in E$.
6. If $\{A, D\} \notin E$, $\{C, B\} \notin E$, $A \neq D$, and $C \neq B$, then replace $\{A, B\}$ and $\{C, D\}$ with $\{A, D\}$ and $\{C, B\}$ in E.
7. Use breadth-first search to determine whether the graph is still connected.
8. If the graph is not connected, remove $\{A, D\}$ and $\{C, B\}$, and put back $\{A, B\}$ and $\{C, D\}$.
9. Randomly select nodes $P, Q \in V$.
10. If $\{P, Q\} \in V$, $P \neq Q$, ($k_A > 2$ *OR* $A = P$), and ($k_B > 2$ *OR* $B = Q$), then replace $\{A, B\}$ with $\{P, Q\}$ in E where k_A.and k_B are the degrees of nodes A and B, respectively. If one of the nodes from which we removed an edge had a degree of 2, it would become a dead-end.
11. Use breadth-first search to determine whether the graph is still connected.
12. If the graph is not connected, remove $\{P, Q\}$, and put back $\{A, B\}$.
13. Move on to the next edge in E.
14. Count the edges in Enot present in E_0, $\Delta = |E| - |E \cap E_0|$.
15. If $\Delta > \Delta_{max}$, then set Δ_{max}equal to the new maximum Δ.
16. Repeat steps 3–15 until we have performed 100 such iterations without updating the value of Δ_{max}.

The algorithm attempts to perform both endpoint swaps between two edges and replacements of single edges with completely new edges, because swapping allows us to change edges that are currently between two nodes with degree 2, whereas replacement does not. Applying

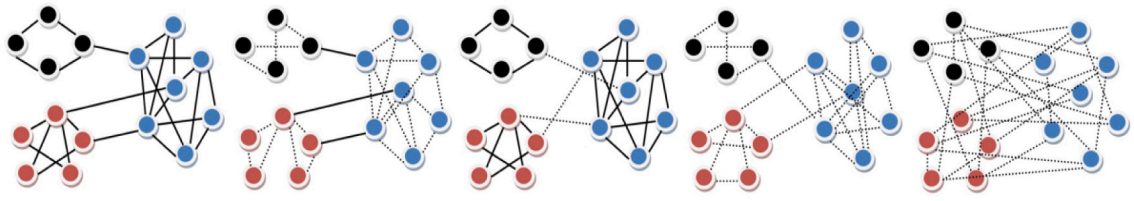

Fig. 2. Different randomizations of a model network. From left to right: original, inter-edges, intra-edges, communities, mean degree.

Randomize-to-Equilibrium as described above to the entire graph produces an instance of a Mean Degree network. To produce an instance of Inter-Edges, we ran Randomize-to-Equilibrium separately on each subgraph defined by a community. To generate an instance of Intra-Edges, we ran the following modified version of Randomize-to-Equilibrium:

1. Copy all inter-edges to set $F = \{\{A,B\} \in E \mid c(A) \neq c(B)\}$, where $c(X)$ is the community to which node X belongs.
2. Save a copy of F in F_0
3. Set $\Delta_{max} = 0$.
4. Iterate over all edges F.
5. Let the current edge be $\{A,B\} \in F$.
6. Randomly select a second edge $\{C,D\} \in F$.
7. If $\{A,D\} \notin F$, $\{C,B\} \notin F$, $A \neq D$, $C \neq B$, $c(A) \neq c(D)$, and $c(C) \neq c(B)$, then replace $\{A,B\}$ and $\{C,D\}$ with $\{A,D\}$ and $\{C,B\}$ in E,F.
8. Use breadth-first search to determine whether the graph is still connected.
9. If the graph is not connected, remove $\{A,D\}$ and $\{C,B\}$, and put back $\{A,B\}$ and $\{C,D\}$.
10. Randomly select nodes $P,Q \in V$.
11. If $\{P,Q\} \in V$, $P \neq Q$, $c(P) \neq c(Q)$, $(k_A > 2\ OR\ A = P)$, and $(k_B > 2\ OR\ B = Q)$, then replace $\{A,B\}$ with $\{P,Q\}$ in E,F where k_A.and k_B are the degrees of nodes A and B, respectively. The requirement that they have degree above 2 is to avoid creating deadends.
12. Use breadth-first search to determine whether the graph is still connected.
13. If the graph is not connected, remove $\{P,Q\}$, and put back $\{A,B\}$.
14. Move on to the next edge in F.
15. Count the edges in Fnot present in F_0, $\Delta = |F| - |F \cap F_0|$.
16. If $\Delta > \Delta_{max}$, then set Δ_{max}equal to the new maximum Δ.
17. Repeat steps 3–15 until we have performed 100 such iterations without updating the value of Δ_{max}.

To make an instance of Communities, we first generated an instance of Intra-Edges, then applied Randomize-to-Equilibrium to each of its communities separately, as done when creating an instance of Inter-Edges. For each pairing of original network and randomization type, we generated 100 randomized networks. For each such network, we used 100 separate runs of MBRW and simple random walk to generate generalized spectral dimension estimates. For box-covering generalized fractal dimension, we performed 100 runs on each network, each with 1000 iterations. Since our version of sandbox generalized fractal dimension is deterministic, we only ran it once on each randomized network. For MBRW and simple random walk spectral dimensions, since the networks we want to compare to each other are the same size, we selected the number of nodes in the network as N_{max}, as this provided a cutoff well below where the walk agent fully covered the network in all cases. For the box-covering and sandbox methods, we used the diameter of the network as the maximum box/sandbox radius.

3. Results

3.1. Effect of rescaling on fractal and spectral dimensions

We tested how sensitive each fractal or spectral dimension is to community structure, network size, and small-scale random variation by creating ensembles of networks with community structures modeled on those of real-world networks but rescaled to different sizes and including partially randomly generated intra-edges and inter-edges. For each real-world network, we generated 10,000 networks with similar community structure of size ranging from 1000 to 100,000 and computed their MBRW spectral dimension, box-covering and sandbox fractal dimensions.

Generalized MBRW spectral dimension shows certain behaviors that are consistent across all q values tested (-10 to 10) (Fig. 3): the variation among networks of the same size, based on the same original is small relative to differences due to size and community structure, showing that it is insensitive to small-scale random variation in structure. The networks based on the same original network form a distinct band, in most cases not overlapping with the bands of other networks based on other originals, showing that the measure is sensitive to community structure. Each band shows a clear upward trend approaching a different asymptotic value as the network size grows large. This upward trend is steep compared to the distances between bands in the size range of the real-world networks (116 to 2880 nodes), indicating that the finite size effect would substantially skew any direct comparison of real-world networks of different sizes within this range. Despite this, the ordering of the bands stabilizes well before the finite size effect begins to level off around 10,000 nodes and remains stable as the size increases. As q increases, the value of each D_q increases at different rates for different networks. The differences between bands are largest for the middle values of q (Fig. 3a), suggesting these provide the overall best sensitivity to differences in community structure. The ordering of bands remains mostly stable except at extreme negative q values where *C. elegans* neural 1 and *C. elegans* neural male switch places and the networks other than the *C. elegans* neuronal networks collapse together (Fig. 3b). Lower D_q values generally indicate stronger community structure, so the higher values of *C. elegans* neuronal networks at negative q compared to the others reflects that they are dominated by a small number of large communities with little of the networks occupied by the small, dense communities to which negative-q D_q dimensions are most sensitive.

The two generalized fractal dimensions also have regimes of q for which they show low variability among networks of the same size based on the same original, separation into bands by community structure, and finite size effects that are pronounced in the range of the real world network sizes but level off in the tens of thousands of nodes. In both cases, we can see these properties clearly for q values near the middle of the range (Fig. 4). However, in most other respects, they behave very differently from each other. Whereas the trend for sandbox fractal dimension, as for MBRW spectral dimension, is to increase with increasing size (Fig. 4a), the box-covering fractal dimension decreases with increasing size (Fig. 4b). Furthermore, whereas the sandbox fractal dimension and MBRW spectral dimension both have distinct bands for all original networks, the box-covering fractal dimension only has three bands, one for the neuronal networks, one for *D. melanogaster* PPI, and one for all of the other networks. Furthermore, whereas the box-covering fractal dimension is most sensitive to small-scale randomness when q is positive (Fig. 5a), the sandbox fractal dimension is most sensitive when q is positive (Fig. 5b). At the opposite ends of the spectrum, at large negative q, box-covering fractal dimension shows little separation into distinct bands according to network structure (Fig. 5c), whereas, for large positive q, sandbox fractal dimension shows distinct bands (Fig. 5d). These differences suggest that sandbox fractal dimension and box-covering fractal dimension are fundamentally different measures and that box-covering fractal dimension is far less sensitive to community structure.

The behavior of sandbox fractal dimension mainly differs from that of MBRW spectral dimension in that its ordering does not stabilize until the networks reach a size of over 40,000 nodes. Additionally, the ordering of bands changes substantially as q changes. For negative q, the sandbox fractal dimension shows high sensitivity to randomness in the neuronal networks but not the others, because the small number of large communities leaves more room for variability in intra-community structure (Fig. 5d). These differences suggest that the sandbox fractal dimension depends on community structure but is less strongly tied to it and more influenced by other factors than is the MBRW spectral dimension. For additional figures showing how the behavior of each generalized fractal or spectral dimension measure changes with q, see Section S3 in Supplementary Materials.

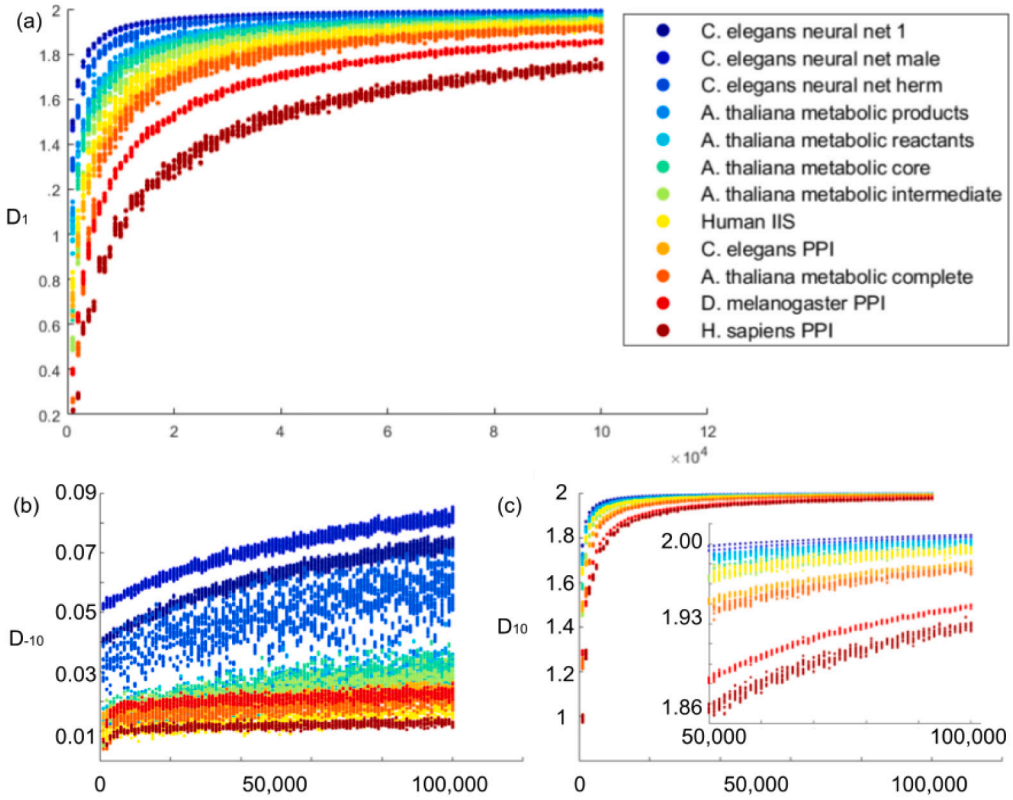

Fig. 3. MBRW (bias = 10,000, memory = 5) generalized spectral dimensions D_1 (a), D_{-10} (b), and D_{10} (c) of rescaled artificial networks modeled on real-world networks. The inset in the plot of D_{10} shows that the networks remain stratified by community structure at larger sizes.

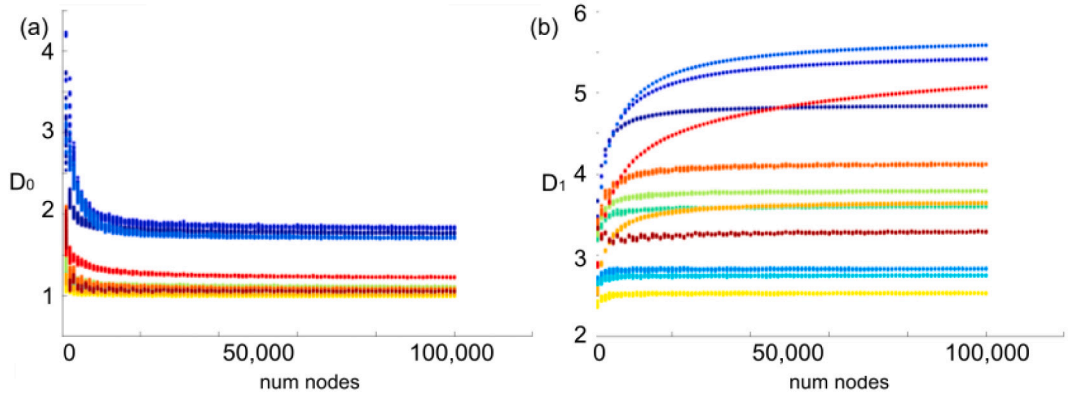

Fig. 4. Box-covering (1000 iterations) fractal dimension D_0 (a) and sandbox fractal dimension D_1 (b) of rescaled artificial networks modeled on real-world networks. The color key is the same as that for Fig. 4.

3.2. *Effect of randomization on fractality and multifractality*

Whereas our tests with the rescaled networks varied the size of the networks while retaining community structure, our network randomization tests kept network size constant while removing community structure. A consistent direction of change from original to partially randomized to fully randomized networks shows that a measure detects community structure. That is, if the direction of change of the median values from Original to Intra-Edges (by randomization of inter-edges), from Original to Inter-Edges (by randomization of intra-edges), from Intra-Edges to Communities (by randomization of intra-edges), from Inter-Edges to Communities (by randomization of inter-edges), and from Communities to Mean Degree (by full randomization of the network) should all occur in the same direction for all real-world networks. Ideally, these changes would also be larger than the variability within a set of networks produced by applying the same randomization to the same original network, showing that the influence of community structure is larger than the influence of small-scale random variation. A smaller range of values on fully randomized networks is also desirable, as it indicates a smaller finite size effect.

For ΔD_q, the range of D_q taken over the selected domain of q for each measure, as a measure of multifractality or multi-spectrality, that is, of the heterogeneity of the network, only the MBRW multi-spectrality behaves completely consistently, always increasing from less randomized to more randomized networks based on the same original network (Fig. 6a). For a discussion of the choice of range of q, see Methods 2.2 and Supplementary Materials S5. In most cases, it also has the smallest variability among networks with the same original and randomization

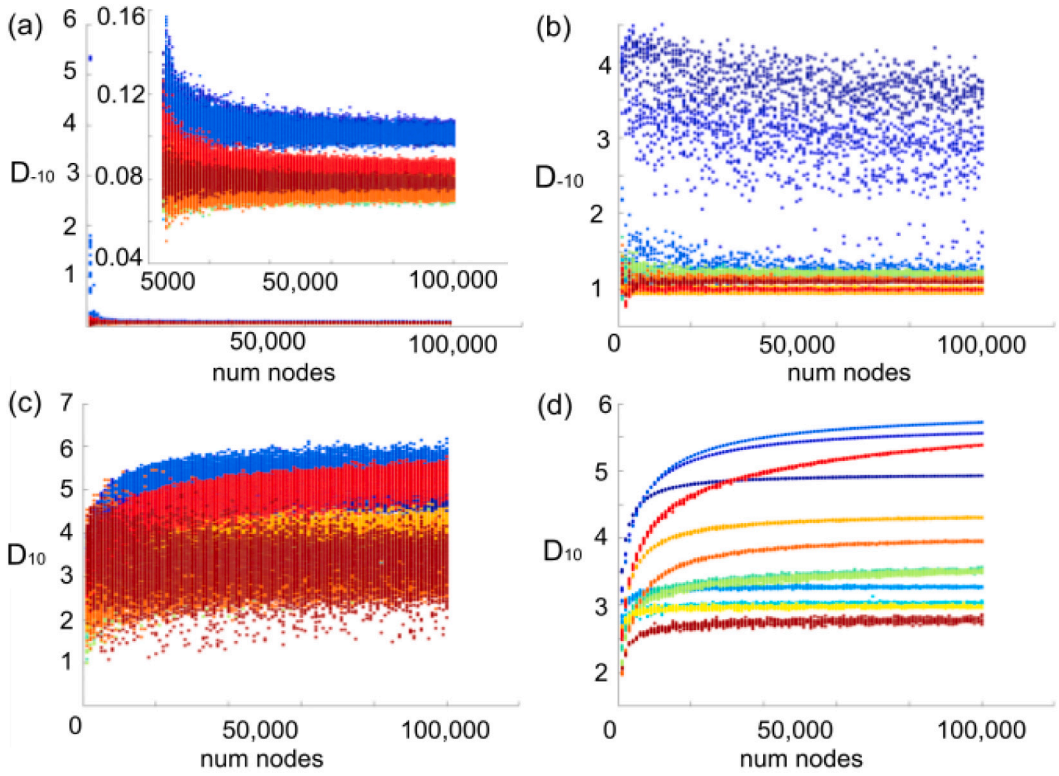

Fig. 5. (a) box-covering (1000 iterations) D_{-10}, (b) sandbox D_{-10}, (c) box-covering D_{10} and (d) sandbox D_{10} generalized fractal dimensions of rescaled artificial networks modeled on real-world networks. The inset for box-covering D_{-10} shows that networks separate into two non-overlapping bands at larger sizes. The color key is the same as that for Fig. 4.

type, though the range of values for fully randomized Mean Degree networks remains substantial due to finite size effect (around 0.15 compared to a range of around 0.30 for the Original networks). The simple random walk multi-spectrality (Fig. 6b) and box-covering multifractality (Fig. 6c) show no clear trends in direction and a high degree of variability within each same-original-and-randomization group. The sandbox multifractality (Fig. 6d) shows a clear downward trend with increasing randomness and small range of fully randomized values but shows inconsistent direction of change between Original and Inter-Edges and between Communities and Mean Degree (Fig. 6e).

We observed that individual shortest-path fractal dimensions (box-covering D_0 and sandbox D_1) changed direction in a highly inconsistent manner, with the box-covering method showing high variability for the same randomization applied to the same original network (See Figs. S115 and S116 in Supplementary Materials). While the simple random walk spectral dimension D_1 consistently increased with increasing randomization, MBRW spectral dimension only increased in most cases with exceptions occurring for some networks upon randomization of intra-community edges (See Figs. S117, S118, and S119 in Supplementary Materials).

4. Discussion and conclusion

In this work, we introduce a measure of network fractality and a measure of multifractality based on the dynamics of a random walk agent with memory running on the network as opposed to relying on static shortest path distances: MBRW spectral dimension and MBRW multi-spectrality, the range values of the generalized MBRW spectral dimensions on a given network. We demonstrate that the memory of the MBRW agent ties these measures strongly to network community structure, making them robust against small-scale random perturbations of the network across a variety of real-world biological networks, while the generalized fractal dimensions and measures based on shortest-path distances had regimes of their power parameters where they were hypersensitive to such small-scale variability and showed less clear stratification by community structure. This suggests that memory helps to stabilize the lower end of the distribution of segment masses for each segment length by tying walk agent behavior more closely to community structure, even when the sizes of the communities themselves are larger or smaller than the number steps the walk agent can keep in memory.

We find that, compared to other methods, the MBRW generalized spectral dimension (D_q) is more sensitive to differences in community structure (differences between networks derived from different real-world networks) than to random variation in small-scale structure (differences among networks derived from the same real-world network) across all orders (q) considered (-10 to $+10$). Every measure, for a q value where it is stable enough against small random variations, exhibits a finite size effect characterized by progression (decrease for box-covering, increase for all other measures) of values toward a distinct asymptotic value for each real-world network with increasing network size. In all cases, the values only approach their asymptotic limits at sizes larger than 30,000 nodes, much larger than the sizes of the real-world network used, 116 to 2880 nodes. To assess how strongly each measure related to community structure without interference from the finite size effect, we compare how values differ between the original networks and fully and partially randomized versions.

All measures of fractality showed a finite size effect that was large relative to the differences due to community structure in the size range of the real-world biological networks considered, 116 to 2880 nodes. This shows that the finite size of real-world networks limits the applicability of fractal and spectral dimensions and that any attempt to compare networks based on such measures must take into account differences in network size. Using a rescaling method similar to that employed in this work may be an acceptable substitute for working with the original networks in some cases, but it will be important to ensure

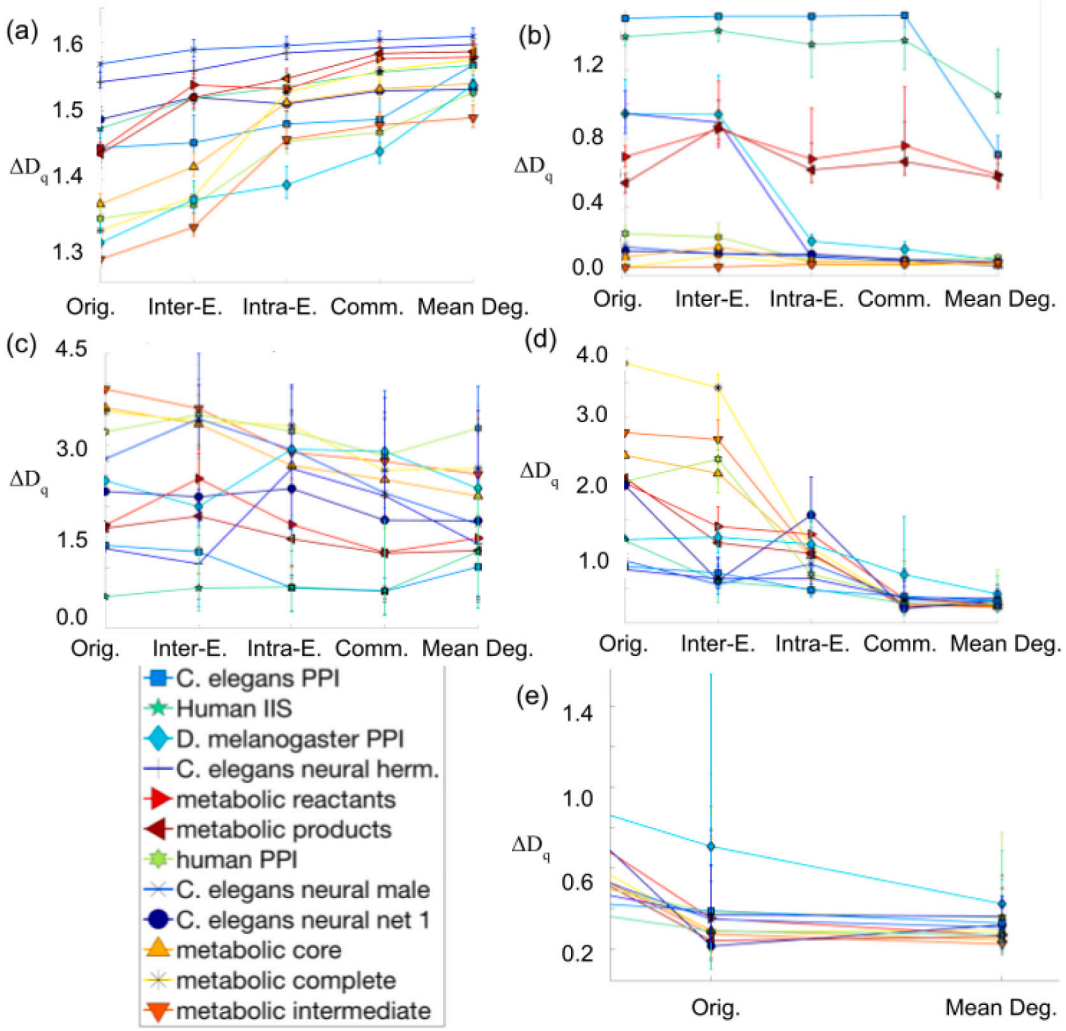

Fig. 6. ΔD_q (range of D_q for q in domain [−10,10]) versus randomization type for MBRW spectral dimension (a), simple random walk spectral dimension (b), box-covering fractal dimension (c), and sandbox fractal dimension (d). (e) shows sandbox ΔD_q for only Communities and Mean Degree, showing that the direction of change is inconsistent.

that the rescaling approach used preserves the characteristics of interest in the original networks.

Another observation that suggests that static topological measures are not sufficient to guage the fractality or multifractality of a network is that only a minority of networks were disassortative, showing a negative correlation between the degrees of edge endpoints, and, of those, only one, Human IIS, has a Pearson correlation of degrees less than −10 %, showing that it was more than weakly disassortative (See Table S2 in Supplementary Materials.). Even though some works, such as [11] cite disassortativity as a defining feature of fractal networks, many real-world biological networks do not exhibit simple disassortativity and instead show dichotomous degree correlation [15]. This implies that the topological approach to measuring fractal dimension may not be applicable to these networks.

That box-covering and sandbox generalized fractal dimensions behaved very differently from one another suggests that they are measuring fundamentally different features of the network and cannot simply be lumped together as alternate methods of estimating an underlying "fractal dimension" that is an intrinsic property of the network. Furthermore, neither one showed a consistent relationship with community structure, suggesting that they are less useful than spectral dimension in general and MBRW spectral dimension in particular for studying meso-scale features and may only be appropriate for their already well-documented [33] applications as macro-scale estimates of network robustness and vulnerability on large networks. We thus show that taking into account the motion of information on a network is key for understanding its structure.

CRediT authorship contribution statement

Adam Craig: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization. **Mesut Yücel:** Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft. **Lev Muchnik:** Conceptualization, Investigation, Writing – review & editing. **Uri Hershberg:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We used data already published elsewhere and cited the sources. Our code is available on GitHub. Please see the end of the supplementary

materials.

Acknowledgements

Computations presented in this work were performed on the Hive computer cluster at the University of Haifa, which is partly funded by Israel Science Foundation grant 2155/15. Lev Muchnik is supported by Israel Science Foundation grant no. 2566/21 and the David Goldman Data-Driven Innovation Research Center. Funding through the Boren Fellowship from the Institute of International Education and through the Louis and Bessie Stein Family Fellowship conferred by the Drexel University Office of Global Engagement enabled coauthor Adam Craig to live in Israel for a year, facilitating collaboration on this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chaos.2022.112707.

References

[1] Achacoso TB, Yamamoto WS. AY's neuroanatomy of C. elegans for computation. CRC Press; 1991.
[2] Alexander S, Orbach R. Density of states on fractals: «fractons». J Phys Lett 1982;43(17):625–31.
[3] Barabási AL. Network science. Cambridge, England: Cambridge University Press; 2016.
[4] Barrat A, Barthelemy M, Vespignani A. Dynamical processes on complex networks. Cambridge, England: Cambridge University Press; 2008.
[5] Bullen PS. Handbook of means and their inequalitiesVol. 560. Springer Science & Business Media; 2013.
[6] Burioni R, Cassi D. Universal properties of spectral dimension. Phys Rev Lett 1996;76(7):1091.
[7] Creixell P, Linding R. Cells, shared memory and breaking the PTM code. Mol Syst Biol 2012;8(1):598.
[8] Edler D, Rosvall M. The MapEquation software package. available online at, http://www.mapequation.org; 2013.
[9] Erdös P, Rényi A. On random graphs I. Publ Math 1959;6(26):290–7.
[10] Furuya S, Yakubo K. Multifractality of complex networks. Phys Rev E 2011;84(3):036118.
[11] Gallos LK, Song C, Makse HA. A review of fractality and self-similarity in complex networks. Physica A 2007;386(2):686–91.
[12] Gilbert EN. Random graphs. Ann Math Stat 1959;30(4):1141–4.
[13] Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JDJ, Hao T, Li N. Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. Nature 2005;436(7052):861.
[14] Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, McKillip E. A protein complex network of Drosophila melanogaster. Cell 2011;147(3):690–703.
[15] Hao D, Li C. The dichotomy in degree correlation of biological networks. PloS One 2011;6(12):e28322.
[16] Hasselmo ME, McClelland JL. Neural models of memory. Curr Opin Neurobiol 1999;9(2):184–8.
[17] Holland PW, Laskey KB, Leinhardt S. Stochastic blockmodels: first steps. Soc Networks 1983;5(2):109–37.
[18] Jarrell TA, Wang Y, Bloniarz AE, Brittin CA, Xu M, Thomson JN, Emmons SW. The connectome of a decision-making neural network. Science 2012;337(6093):437–44.
[19] Li S, Wang L, Berman M, Kong YY, Dorf ME. Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. Immunity 2011;35(3):426–40.
[20] Li BG, Yu ZG, Zhou Y. Fractal and multifractal properties of a family of fractal networks. J Stat Mech: Theory Exp 2014;2014(2):P02020.
[21] Liu JL, Yu ZG, Anh V. Determination of multifractal dimensions of complex networks by means of the sandbox algorithm. Chaos 2015;25(2):023103.
[22] Mandelbrot B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 1967;156(3775):636–8.
[23] Pons P, Latapy M. Computing communities in large networks using random walks. In: International symposium on computer and information sciences. Berlin, Heidelberg: Springer; 2005, October. p. 284–93.
[24] Radrich K, Tsuruoka Y, Dobson P, Gevorgyan A, Swainston N, Baart G, Schwartz JM. Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. BMC Syst Biol 2010;4(1):114.
[25] Rammal R, Toulouse G. Random walks on fractal structures and percolation clusters. J Phys Lett 1983;44(1):13–22.
[26] Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Kamburov A. A proteome-scale map of the human interactome network. Cell 2014;159(5):1212–26.
[27] Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 2008;105(4):1118–23.
[28] Sammut M, Cook SJ, Nguyen KC, Felton T, Hall DH, Emmons SW, Barrios A. Glia-derived neurons are required for sex-specific learning in C. Elegans. Nature 2015;526(7573):385.
[29] Sarantopoulos I, Papatheodorou D, Vogiatzis D, Tzortzis G, Paliouras G. TimeRank: a random walk approach for community discovery in dynamic networks. In: International Conference on Complex Networks and their Applications. Cham: Springer; 2018, December. p. 338–50.
[30] Seidman SB. Network structure and minimum degree. Soc Networks 1983;5(3):269–87.
[31] Song C, Havlin S, Makse HA. Self-similarity of complex networks. Nature 2005;433(7024):392.
[32] Wang D. Multifractal characterisation and analysis of complex networks. Queensland University of Technology; 2011. Doctoral dissertation.
[33] Wang DL, Yu ZG, Anh V. Multifractal analysis of complex networks. Chin Phys B 2012;21(8):080504.
[34] Wen T, Cheong KH. Invited review: the fractal dimension of complex networks: a review. Information fusion. 2021.
[35] Wu Y, Chen Z, Yao K, Zhao X, Chen Y. On the correlation between fractal dimension and robustness of complex networks. Fractals 2019;27(04):1950067.
[36] Xu M, Jarrell TA, Wang Y, Cook SJ, Hall DH, Emmons SW. Computer assisted assembly of connectomes from electron micrographs: application to Caenorhabditis elegans. PloS One 2013;8(1):e54050.
[37] Yucel M, Hershberg U. Memory as an organizer of dynamic modules in a network of potential interactions. SCW at AAMAS; 2014.
[38] Yucel M, Muchnik L, Hershberg U. Detection of network communities with memory-biased random walk algorithms. J Complex Networks 2016;5(1):48–69.