



Relative, local and global dimension in complex networks

Robert Peach ^{1,2,4}, Alexis Arnaudon^{1,3,4} & Mauricio Barahona ¹✉

Dimension is a fundamental property of objects and the space in which they are embedded. Yet ideal notions of dimension, as in Euclidean spaces, do not always translate to physical spaces, which can be constrained by boundaries and distorted by inhomogeneities, or to intrinsically discrete systems such as networks. To take into account locality, finiteness and discreteness, dynamical processes can be used to probe the space geometry and define its dimension. Here we show that each point in space can be assigned a relative dimension with respect to the source of a diffusive process, a concept that provides a scale-dependent definition for local and global dimension also applicable to networks. To showcase its application to physical systems, we demonstrate that the local dimension of structural protein graphs correlates with structural flexibility, and the relative dimension with respect to the active site uncovers regions involved in allosteric communication. In simple models of epidemics on networks, the relative dimension is predictive of the spreading capability of nodes, and identifies scales at which the graph structure is predictive of infectivity. We further apply our dimension measures to neuronal networks, economic trade, social networks, ocean flows, and to the comparison of random graphs.

¹Department of Mathematics, Imperial College, London SW7 2AZ, UK. ²Department of Neurology, University Hospital Würzburg, Würzburg, Germany. ³Blue Brain Project, École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech 1202 Geneva, Switzerland. ⁴These authors contributed equally: Robert Peach, Alexis Arnaudon. ✉email: m.barahona@imperial.ac.uk

One of the first forays into graph dimensionality originated with Erdős, when he explored the embedding of graphs into a minimum finite-dimensional Euclidean space¹. This line of study helped realise the algorithmic importance of geometric interpretations of graphs² but was unfortunately no more than a by-product of the graph embedding process, yielding little actionable information³. Later, by characterising the fractal properties of complex networks, a measure of network dimension was defined in terms of the scaling property of a network topological volume^{4–6}. Whilst the fractal approach showed that dimension plays an important role in characterising network topology and governing dynamical processes such as percolation⁷, it was initially limited to global descriptions of network dimension. Extensions that considered the local scaling properties of the volume at different topological distances from a node were introduced in⁸ and have been used to define a node-centric dimension that can identify influential nodes^{9,10} or vital spreaders in infection models¹¹.

However, methodologies based on fractal approaches assume that the topological volume follows a power-law distribution, a strong assumption, not necessarily accurate in real-world networks exhibiting heterogeneities⁵. Similarly, in classic papers such as¹², where the dimension of a node is defined using the decay rate of diffusion, or in¹³, where a random walk is used to create node embeddings, the same assumptions of homogeneity are required and an intermediate scale of dynamics must be chosen. As an example, with a diffusive source located at the joining of a 1-d and a 2-d space, by measuring the decay rate we immediately ignore the heterogeneity of the space and simply find a dimension somewhere between 1 and 2. In this paper, we posit that the dimension at a node can, and should be, defined as *relative* to another node. Using the solution of diffusion at other nodes relative to the source we are able to define a *relative dimension*.

Results

Graph dimension from diffusion dynamics. We start with the Green's function of the diffusion equation in d dimensions

$$G_t(\mathbf{x}) = (4\pi\sigma t)^{-d/2} \exp\left(-\frac{\|\mathbf{x}\|^2}{4\sigma t}\right), \quad (1)$$

which, together with an initial condition as a delta function at some position \mathbf{x}_0 , provides a solution of diffusion equation as $p(\mathbf{x}, t) = G_t(\mathbf{x} - \mathbf{x}_0)$. From hereon, we refer to the time evolution of $p(\mathbf{x}, t)$ as the transient response. As already considered in our previous works^{14,15}, these solutions have a maxima in their transient response at any other location \mathbf{x} , at time \hat{t} and amplitude \hat{p} given as

$$\hat{t}(\mathbf{x}) = \frac{\|\mathbf{x}\|^2}{2d\sigma}, \quad \hat{p}(\hat{t}) = (4e\pi\sigma\hat{t})^{-\frac{d}{2}}, \quad (2)$$

where, without loss of generality, $\mathbf{x}_0 = 0$. Then, the dimension at any point \mathbf{x} relative to \mathbf{x}_0 can be evaluated to yield the definition of the relative dimension

$$d(\mathbf{x}|\mathbf{x}_0) = \frac{-2 \ln \hat{p}}{\ln(4e\pi\sigma\hat{t})}. \quad (3)$$

Clearly, on the Euclidean space \mathbb{R}^d , the relative dimension is always equal to d , independently of \mathbf{x} and \mathbf{x}_0 . However, if we instead consider a compact subspace $\Omega \subset \mathbb{R}^d$, the diffusion dynamics will deviate from those prescribed in Equation (1) due to the presence of boundaries relative to \mathbf{x} and \mathbf{x}_0 .

The key property of Equation (3) that allows us to generalise it to graphs is that the positions \mathbf{x}_0 and \mathbf{x} are not explicit in the right-hand side but only used as labels to initialise the diffusion dynamics and measure the transient response. Consequently, the

relative dimension can be seen as intrinsic as it does not rely on any Euclidean embedding, but only on the existence of a diffusion dynamics on the original space. In particular, on graphs we can use the standard diffusion process

$$\partial_t \mathbf{p}(t) = -L\mathbf{p}, \quad (4)$$

for a time-dependent node vector $\mathbf{p}(t)$ with L the normalised graph Laplacian $L = K^{-1}(K - A)$ (corresponding to Euclidean diffusion in the continuous limit¹⁶), where K is the diagonal matrix of node degrees. Using a delta function at node i with mass m_i , $\mathbf{p}(0) = (0, 0, \dots, m_i, \dots, 0)$, as our initial condition, the j -th coordinate of the solution of Equation (4) (the so-called transient response of j) is given by the heat kernel

$$p_j(t|i) = m_i(e^{-tL})_{ij}. \quad (5)$$

By numerically solving (5), we can measure the time \hat{t}_{ij} and amplitude \hat{p}_{ij} at which a maximum appears in the transient response peak (time evolution) of node j given a delta function initial condition at node i . In analogy to Equation (3), we can then compute the full $N \times N$ matrix of relative dimensions with elements

$$d_{ij} = \frac{-2 \ln \hat{p}_{ij}}{\ln(4e\pi\sigma\hat{t}_{ij})}. \quad (6)$$

To illustrate the notion of relative dimension, we used a line graph (Fig. 1a, b) as a discrete representation of the continuous 1-D interval. We observe that due to the boundaries, a large fraction of nodes do not have a peak in transient response, however for nodes near the source, where the boundary has no influence, the relative dimension is close to the expected $d = 1$. We emphasise that the dimension is not derived from a fit to the data, as is common in measures of fractal dimensions^{4–6}, but instead is directly observed at the transient response relative to a source node.

It is then natural to define the *local dimension* of a node i by averaging the relative dimension of the nodes displaying a peak in their transient responses relative to i before a given time τ as

$$\mathcal{D}_i(\tau) = \frac{\sum_{j=1, j \neq i}^n d_{ij}(\tau) \mathbb{1}_{\hat{t}_{ij} < \tau}}{\sum_{j=1, j \neq i}^n \mathbb{1}_{\hat{t}_{ij} < \tau}}, \quad (7)$$

where $\mathbb{1}_{\hat{t}_{ij} < \tau}$ is the indicator function. Whilst the local dimension can be likened to a measure of centrality, it also directly captures the dimension of the local embedding space. In Fig. 1c we observe the increasing effect of the boundaries on local dimension as we increase the scale. Near the centre of the line, and when considering nearby nodes (at short scales), one can expect to estimate a dimension near 1, or equivalently 2 for the grid shown in Fig. 1d. We observe in Fig. 1c a central region with $\mathcal{D}_i \sim 1$ that becomes increasingly smaller as scale τ increases; at short scales, the central region is insensitive to the boundaries since the diffusion has not yet reached them. This 'boundary insensitive central region' collapses at $\tau = 1$ (corresponding to the spectral gap of the graph) when all nodes have aggregated information about the boundaries of the line graph.

Finally, we can define a graph measure of dimension by averaging the local dimensions across multiple scales to obtain the global dimension

$$\mathfrak{D}(\tau) = \frac{1}{n} \sum_{i=1}^n \mathcal{D}_i(\tau), \quad (8)$$

still dependent on τ . In Fig. 1e we display the global dimension (as a ratio to the expected Euclidean dimension) for the line and

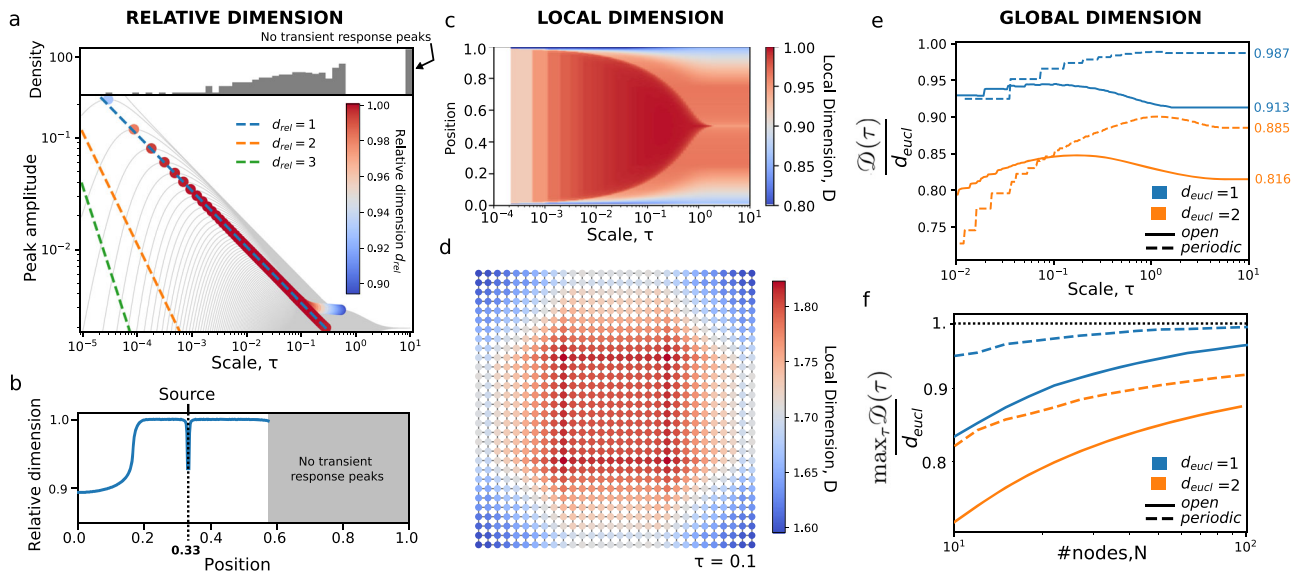


Fig. 1 The relative, local and global dimension. **a-c** Line graph example with $n = 500$ nodes representing the interval $[0, 1]$. **a** The relative dimension of nodes given a source located at $x = 0.33$. The grey lines are the transient responses of the (non-source) nodes and the position of the peaks in the transient responses are highlighted by dots, coloured by their relative dimension. Top inset, a histogram of transient response peaks where the far right bin corresponds to nodes where no peak in the transient response was observed, and thus no relative dimension could be calculated. **b** The relative dimension as a function of position in $[0, 1]$ shows a plateau near $d_{rel} = 1$ for nodes near the source. The grey region indicates the set of nodes for which no peak was observed. **c** The local dimension of each node as a function of scale, where above $\tau = 1$, the stationary state is attained and the local dimension is stable. **d** The local dimension of the grid graph ($n = 500$) at scale $\tau = 0.1$, showing inhomogeneities due to the boundaries similar to the line graph. **e** The evolution of the global dimension $\mathcal{D}(\tau)$ (normalised by the expected Euclidean dimension) as a function of scale for the same line and grid graph as well as their periodic equivalent graphs, illustrating differing behaviours emerging from the influence of the boundaries or the topology. **f** For the same graphs as in **e**, we increase the number of nodes in each dimension to measure the convergence rate of $\max_{\tau} \mathcal{D}(\tau)$ to the underlying Euclidean dimension d_{eucl} , showing a faster convergence for lower dimensional spaces and periodic grids.

grid graphs and their periodic equivalents (the circle and sphere graphs respectively).

Whilst the periodic equivalents do not contain boundaries, they are still constrained to a compact space that will introduce topological effects, e.g., on a periodic graph the diffusion will interact with itself at the opposite side to the initial condition. We first notice that the non-periodic graphs display a maximum in global dimension, likely when the effect of the boundaries is lowest. In contrast, the periodic graphs do not exhibit a peak of the same magnitude suggesting that the topological effect of a compact space has less impact on the global dimension than the presence of a boundary.

In the context of graphs as discrete Euclidean spaces, the maximum of the global dimension curve (Fig. 1e) can be seen as an approximation of the Euclidean dimension, whereas the global dimension at largest scale characterises the effect of the boundary or topology of the graph. It should be noted that for a non grid-like graph, what is a boundary or a topological effect is not clear. By increasing the graph size, and thus reducing the effects of the boundaries, the global dimension converges towards the expected Euclidean dimension (Fig. 1f). For the grid, the surface of the boundary increases with respect to the volume of the space and results in a slower convergence, whereas the global dimension of the periodic grid is only affected by the topology, and thus converges faster.

Delaunay meshes and inhomogeneities. To develop more intuition for our measure of relative dimension, we consider a simple constructive example using Delaunay meshes in Fig. 2. Given a source-node located at the left boundary of a homogeneous delaunay mesh, relative dimension displays an inhomogeneous distribution radially from the source until nodes do not have a transient response peak (Fig. 2(a)). Adding nodes near

the centre of the Delaunay grid graph creates local inhomogeneities modifying the underlying space, with a clear analogy to the theory of gravitation and gravitational lensing¹⁷. In particular, the added mass acts as a gravitational lens for the diffusion process, whereby nodes directly behind the point mass that were previously 'unreachable' can be 'reached by the diffusion' if the mass is sufficiently large. Small masses are reminiscent of weak lensing (Fig. 2(b)), whereas larger masses are closer to strong lensing (Fig. 2(c))¹⁸. The behaviour of relative dimension in the presence of inhomogeneities suggests that diffusion effectively occurs on a curved geometry induced by the presence of the mass. Moving the mass towards one boundary (Fig. 2(d)) shows some coupling between the lensing effect and the presence of the boundary. All three possible effects, boundaries, topology and inhomogeneities, are thus important in the notion of dimensions, but may not be distinguishable in more complex networks. Nevertheless, our notion of relative dimension is able to capture them all in one graph-theoretical measure.

Dimensions in protein structure: rigidity and allostery. We then apply the relative dimensions on a real-world example with allostery in proteins, a phenomena whereby a subset of a protein (active site) can be modulated (activated or inactivated) through binding of a ligand at another subset of the protein (allosteric site). We examine three well-studied allosteric proteins: HRas GTPas, Lac repressor and PDK1 in Fig. 3 (for more details on these proteins, see Methods). In HRas, we find a low relative dimension at the active site given the allosteric site as the source (Fig. 3a(i)), but in reverse the allosteric site does not see a transient peak from the diffusion started in the active site (Fig. 3a(ii)). Even if an exact statement of allosteric mechanism is not our purpose here, it is interesting to note that a low relative dimension suggests a more 'direct' or 'funneled' communication from

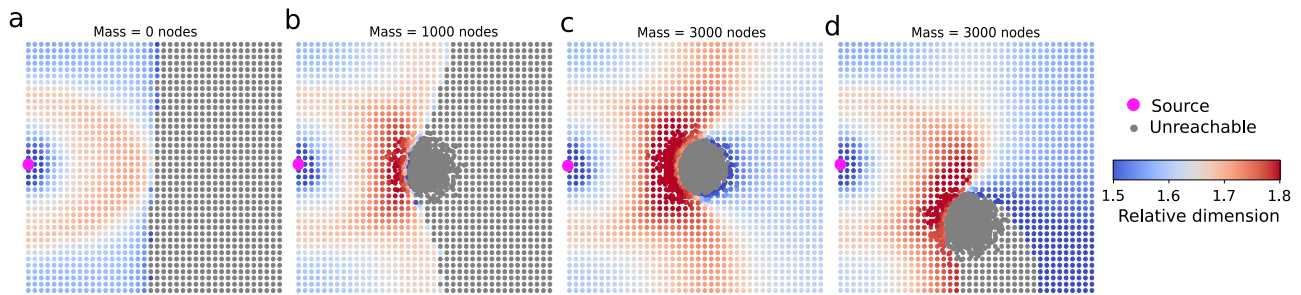


Fig. 2 Inhomogeneities and lensing. The relative dimension from a point source (pink dot) to other nodes in a Delaunay grid graph (edges not shown). Grey nodes indicate nodes for which no transient response peak was detected. The regular grid is shown in **a**, and an additional mass is added in **b–d**, with varying mass and position, showing a effect similar to gravitational lensing.

the allosteric site to the active site. Moreover, the asymmetry of this communication may relate to different functions for each half of the protein.

The lac repressor protein is constructed from two separate monomers and it is generally understood that binding of both NPF molecules (one on each monomer) is required to activate the lac repressor via a cooperative allosteric effect acting on the hinge region¹⁹. Given that the allosteric mechanism is cooperative, we do not expect a direct communication to the active site from the allosteric site, and instead we examined the change in relative dimension upon using a single allosteric site as a source (Fig. 3b(i)) vs. both allosteric sites as sources simultaneously (Fig. 3b(ii)). We find that when binding NFP to just one monomer the relative dimension across the entire protein is lower when compared to using both allosteric sites as sources of diffusion.

Finally, binding at the PDK1 interacting fragment (PIF) on PDK1 triggers a signal to start the phosphorylation of the activation loop of the substrates at the ATP pocket, or active site²⁰, and thus we would expect direct communication between the active and allosteric sites. Using the allosteric site as the source of our diffusion (Fig. 3c), we find that a large region of PDK1 does not return a relative dimension (grey region in Fig. 3c). We remind the reader that to calculate relative dimension we must observe a peak in the transient response. Of those residues for which relative dimension was computed, the activation loop displays the lowest relative dimension to the allosteric site. We hypothesise that a lower dimension pathway from the allosteric to active site will improve the efficiency of communication transfer since it becomes more direct.

Whilst the relative dimension provides insights into allostery, we can leverage the local and global dimension to examine protein dynamics. In Fig. 4(a), we show a strong correlation between the local dimension and $\log_{10}(1/\text{RMSF})$ of residues for Fig. 4a(i) an unglycosylated antibody CH2 domain and Fig. 4a(ii) an Oestrogen Related Receptor g protein. The results here suggest that a residue with a larger local dimension is associated with a lower flexibility and thus lower degrees of freedom.

To examine this further, we plotted the Pearson correlation between local dimension and $\log_{10}(1/\text{RMSF})$ for 12 randomly chosen proteins in Fig. 4(b). We see that at middling to long time scales of diffusion the correlation plateau with an average at about $\sigma = 0.55$ suggesting that the relationship between local dimension and protein flexibility is robust. Calculating the global dimension for the same set of proteins in Fig. 4(c), we find a correlation (Pearson $\sigma = 0.73$) between global dimension and the $\log_{10}(1/\text{RMSF})$ of a protein. The global values of dimension sit between 1.36 and 1.5 for the 12 proteins. These results agree with studies that show spectral dimension is generally < 2 and decreases with an increase in flexibility^{12,21}.

We now take a deeper look at *Aquifex* Adenylate Kinase (ADK), a dynamical protein with three subdomains: the lid, AMP and core domains. We find that the closed conformation displays a higher local dimension due to the presence of stabilising interactions, not present in the open conformation, creating a more compact structure (Fig. 4d). The AMP and lid domains are known to open and close around substrate. We find that both have a lower local dimension relative to the core domain (Fig. 4e) and that the AMP domain to have a lower average local dimension than the lid domain in both conformations. The latter we validated using experimental fluorescence correlation spectroscopy that shows that the AMP domain to open and close at a faster rate (16.2 μs) than the lid domain (46.6 μs)^{22,23}.

Local dimension as a means to differentiate node roles. To further explore our measure of dimension in the context of identifying roles of nodes within the network, we present two examples of real-world complex networks in Fig. 5 where nodes have pre-assigned roles. The first example explores the world trade network (consisting of 80 nodes) of metal manufacturing in 1994²⁴, where nodes correspond to countries and directed incoming edges represent the amount of weighted imports from another country. A well established concept in economic theory partitions countries based on their positioning (1. core, 2. semi-peripheral, 3. peripheral) within the world economy²⁵. For the largest scale, we find significant differences between distributions of the local dimension for each of the world partitions (Fig 5b). There is almost no overlap in local dimension between the two extreme partitions, core and periphery, but the distribution of local dimension for semi-peripheral nodes is wider, suggesting that this class of countries is more diverse.

Our second example is the undirected connectome ($N = 377$) of the nematode *Caenorhabditis elegans* (Fig. 5b(i)) with the inclusion of muscles, important for examining control²⁶ (<https://www.wormatlas.org/neuronalwiring.html>), and where scales have previously been shown as important²⁷. We compare the dimension of the three different neuronal types (inter neurons, sensory neurons, motor neurons) and muscles, at long scales in Fig. 5b(ii), and find significant differences in their local dimensions. Inter-neurons are central nodes of neural circuits that enable communication between sensory and motor neurons, thus we would expect them to sit in a higher-dimensional space, where muscles are peripheral as they display the lowest local dimension, likely aiding with the direct propagation of signals. In addition, we find the highest dimensional nodes are the important control motor neurons AVA/AVB neurons (both left and right), resulting in uncontrolled motion if ablated²⁶ (see Supplementary Table 1 for top 40 local dimension neurons).

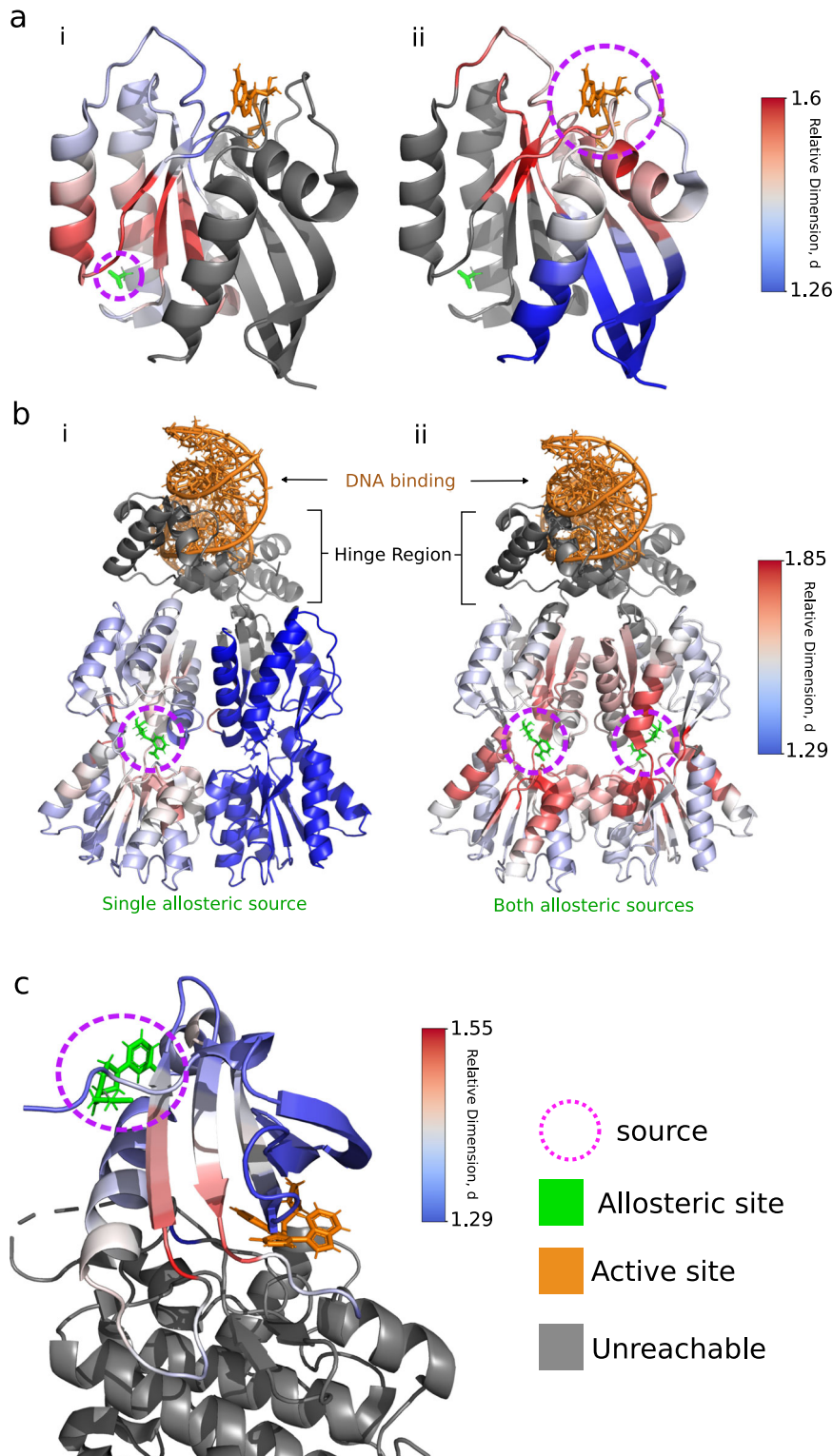


Fig. 3 Relative dimensions in allosteric proteins. Protein residues (amino acids) are coloured according to their relative dimension to the source region, where grey indicates that no peak in the transient response was identified and thus relative dimension could not be calculated. **a** The relative dimensions of all atoms in HRas GTPase (PDB ID: 3K8Y) given (i) the allosteric site and (ii) the active site as the source of diffusion. **b** Relative dimension in the multi-allosteric site Lac Repressor protein (PDB ID: 1EFA) given a (i) single allosteric site source and (ii) for both allosteric sites simultaneously. **c** The relative dimension give the allosteric site as the source in PDK1 (PDB ID: 3ORX).

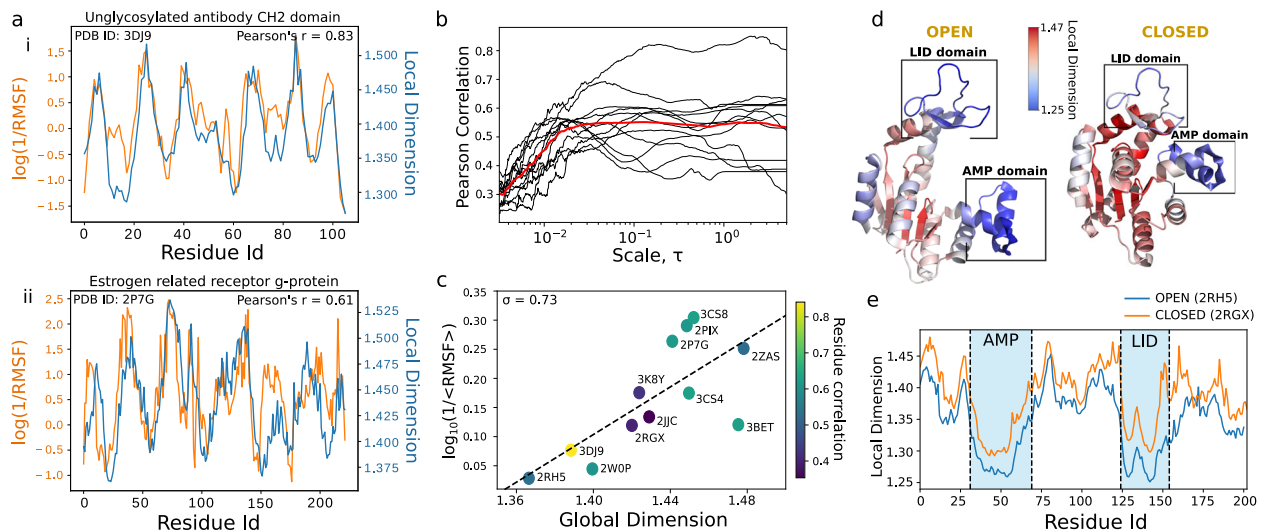


Fig. 4 The relationship between root-mean-square fluctuations (RMSF) of protein residues and their local and global dimension. **a** The log-inverse RMSF vs local dimensionality for each residue in **i** unglycosylated antibody CH2 domain, **ii** Oestrogen Related Receptor g. **b** The positive correlation (0.539) between local dimension and log-inverse RMSF at the residue level across 12 different proteins as a function of scale. **c** A strong positive correlation between global dimensionality of 12 proteins against log-inverse RMSF. **d** The local dimension of each residue mapped onto Aquifex Adenylate Kinase in the open (PDB ID: 2RH5) and closed (PDB ID: 2RGX) conformations and **e** plotted by residue id.

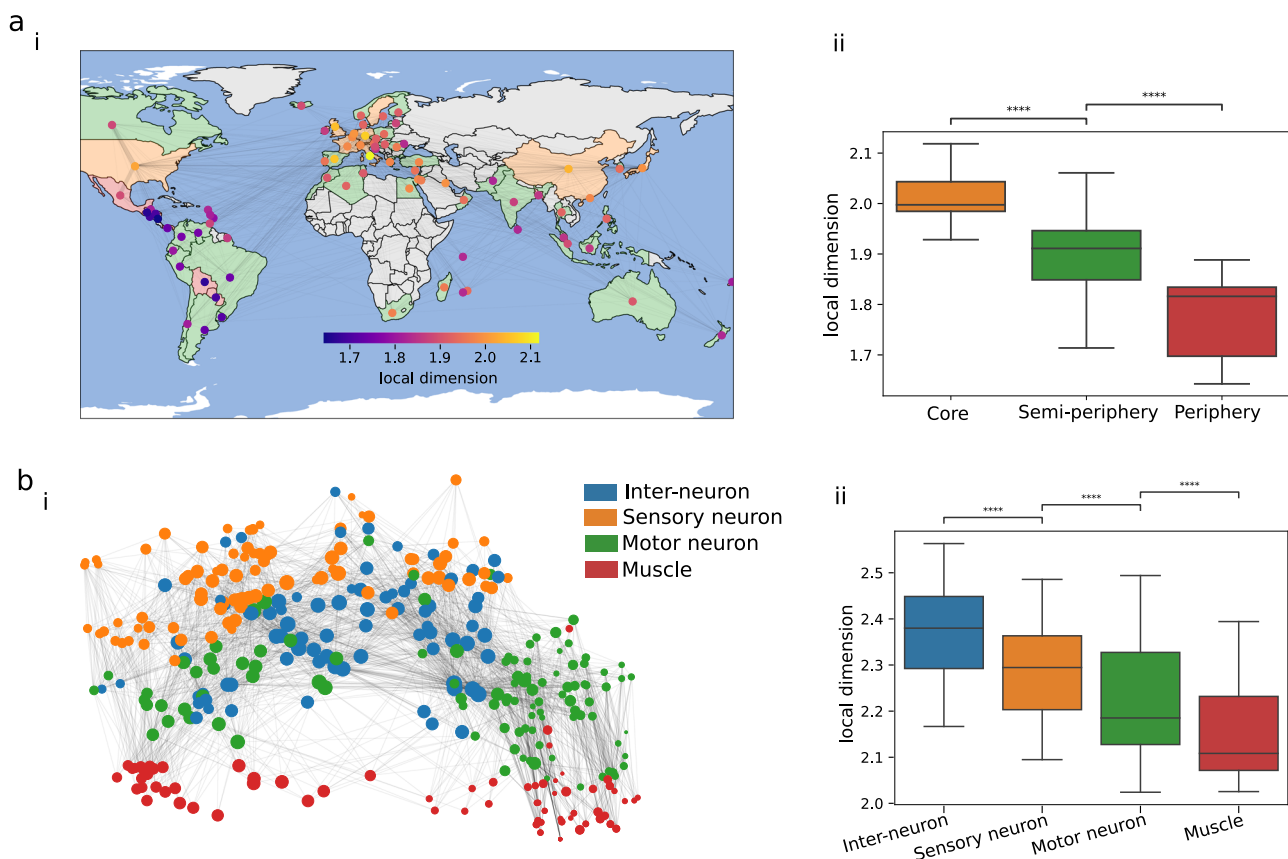


Fig. 5 Spatially embedded networks with long range interactions. **a** **i** The local dimension (coloured nodes) of the 1994 world trade dataset²⁴ ($\tau = 0.57$). Countries are coloured according to their position in the world partition (grey countries not in dataset). **ii** Comparison of local dimension by world partition shows a decreasing local dimension as the countries become more peripheral. **b** **i** The *C. elegans* connectome consisting of 279 neurons and 98 muscles coloured by neuron type/muscle are shown in **i**. In **ii**, the boxplots of local dimensions at large scale show that the neuron types statistically differ. All statistical tests were Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction, **** $p \leq 0.0001$.

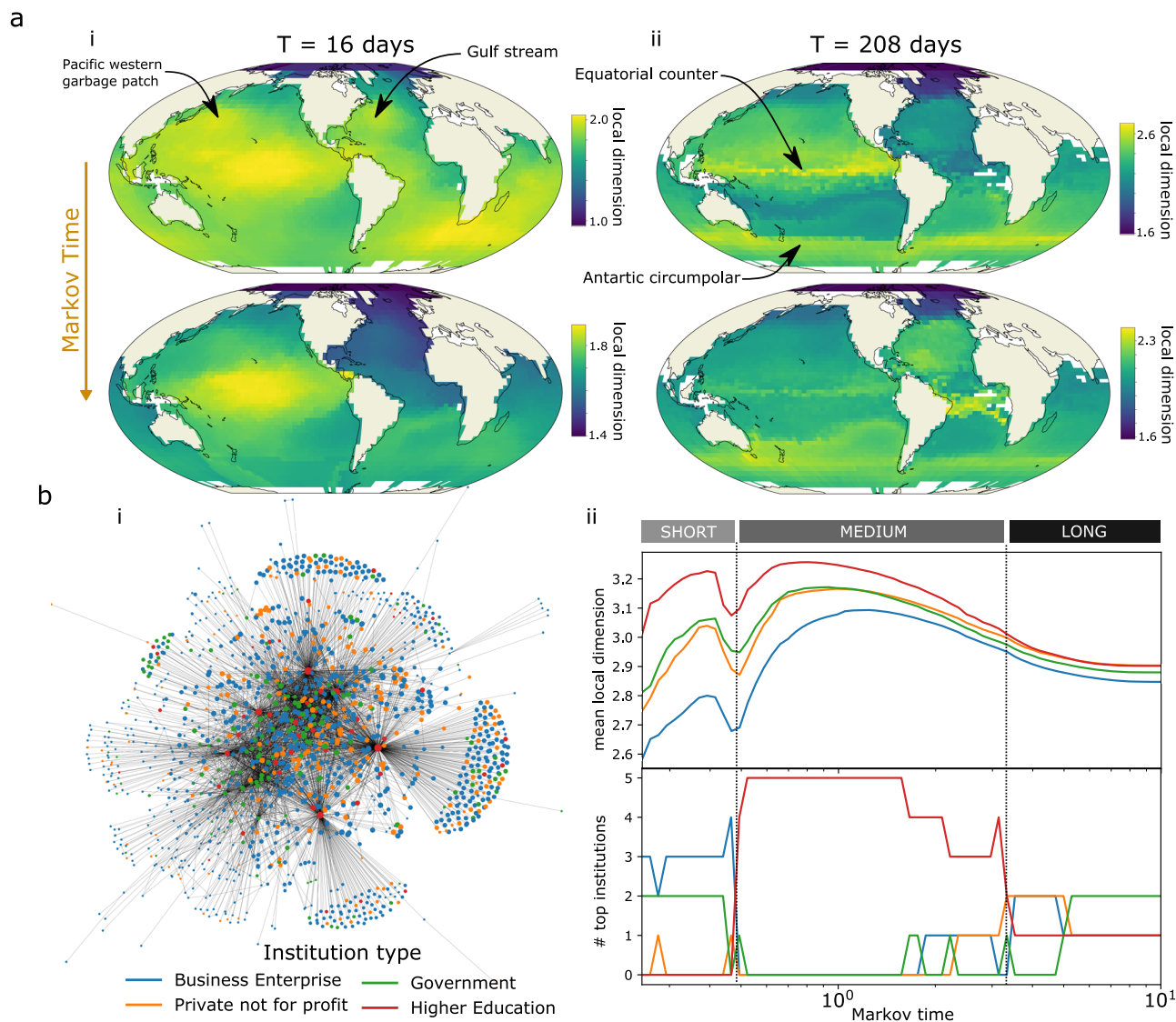


Fig. 6 Illustration of local dimension at several scales in two dataset. In **a**, we considered two graphs extracted from the ocean drifter by ref. ²⁹ where edges are the number of drifters crossing two regions of the ocean within $\mathbf{i} T = 16$ and $\mathbf{ii} 208$ days. For each, we selected a small and a large scale of local dimension, each representing various known features of the ocean dynamics, mostly located between the main gyres such as between the north and south equatorial in the pacific, or along the antarctic circumpolar current. In **b**, we explored a **i** social network of scientific collaborations between New Zealand institutions **ii** across scales, to find that the top 5 institutions are businesses at small scales, universities at longer scales, and a mix of institutions at stationarity. From middle scales, the University of Auckland remains the top-ranked node until stationarity.

Local dimension as scale-dependent measure of centrality.

Measures of centrality are some of the most fundamental tools in network theory. Here, we show that the local dimension can also be utilised as a scale-dependent centrality measure, such as those derived in ^{15,28}. To illustrate the use of the local dimension as a centrality measure for complex networks we analysed two datasets where the importance of nodes changes substantially with scale.

First we look at the global network of ocean surface currents derived from the Global Drifters programme (<http://www.aoml.noaa.gov/phod/gdp/index.php>) constructed by ²⁹ (https://github.com/maurofaccin/ocean_surface_dataset). Each node is associated with a small region of the ocean, and an edge between two nodes counts the number of drifters passing from one to another region in a given time interval T . For short times, such as $T = 16$ days, the graph connectivity remains local with respect to the spatial embedding of the nodes on the earth surface, but with larger times ($T = 208$ days) the connectivity becomes long range

and complex (see also the degree distribution in Supplementary Fig. 1). We can examine both time intervals at short and long scales of our local dimension (Fig. 6a); the small or large scale local dimension provide different perspectives on regions of high dimensions, related to regions where the ocean flow has a more complex dynamics. At small time intervals and short scales (Fig. 6a(i) top), we identify locally high dimension regions such as the Gulf stream or the Pacific garbage patch where drifters remain trapped and circulate quickly. If we look at long time intervals (Fig. 6a(ii)), we notice bands of high dimension which represent the boundaries between main gyres, such as that along the equator. At short scales, the drifters have lower dimensional dynamics while they follow these currents. However, at longer scales the drifters can drift north or south of the equator and be further transported to widely different regions throughout the world, and thus the dimension of the boundaries between major ocean currents is larger. We also note a visual similarity between the small time interval and long scale (Fig. 6a(i) bottom) and long

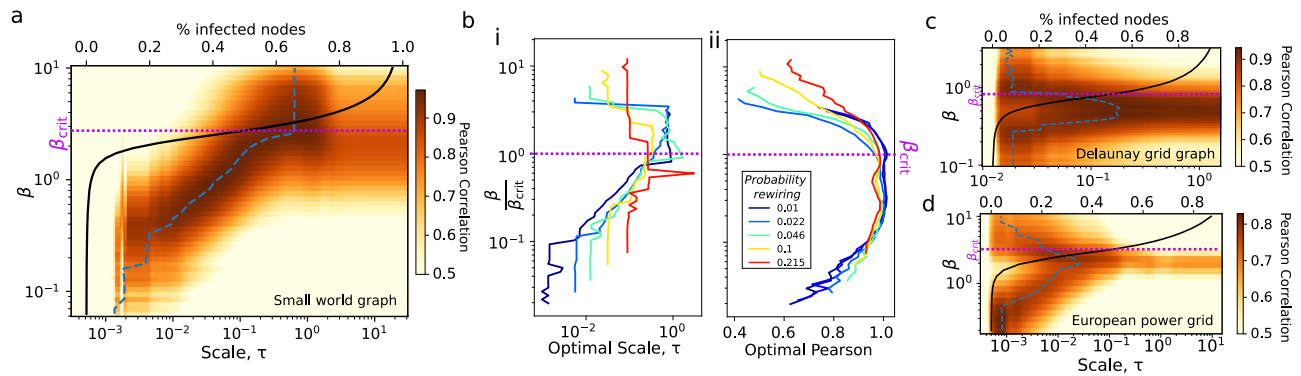


Fig. 7 Dimension and epidemic spreading. **a** Heatmap of Pearson correlation between local dimension and node infectiousness for small-world graph ($n = 100$, average degree $k = 10$, probability of rewiring $p = 0.015$). The black line is the average proportion of infected nodes given a single-seed node for a given infection probability β . The transition from low to high proportion of infected nodes indicates the critical point while the dashed line is the maximum correlation for each β . **b** We vary the probability of rewiring edges p of small world graphs and display **i** the diffusion time that maximises the correlation between local dimension and infectiousness for varying β , and **ii** the associated correlation coefficient. The correlation is near one close to criticality and above 0.8 for a large range of β . We repeat the analysis in **a** for **c** a Delaunay grid graph ($n = 400$) and **d** the European powergrid network to observe a similar linear relationships between scale and infection probability β prior to criticality.

time interval and short scale (Fig. 6a(ii) top), whereby the drifter movements are generally split between north and south. Our results provide further evidence that a notion of scale in the analysis of ocean flow is crucial to exploit and interpret the dynamics²⁹.

Finally, we examine a complex social network of scientific collaborations between New Zealand institutions (Fig. 6b). Each node represents an institution which falls into the following categories: higher education, Government, Private not for profit, or Business Enterprise. Edges are weighted by the number of collaborations between two institutions in the time period 2010–2015, measured by co-authored publications on Scopus³⁰. We compute the local dimension as a function of scale on this network and identify three main scales (short, medium and long; Fig. 6b(ii)). On average and across scales, the higher education institutions displayed the highest local dimension and business enterprises were lowest. However, if we instead look only at the 5 nodes with the highest local dimension, we find that at short scales, businesses and government institutions comprised the top 5 local dimension nodes, highlighting their high dimension to a small neighbourhood. For a wide range of medium time scales, we find that the universities display the largest local dimension, reflecting their hub-like role in the network (Fig. 6b(i)). At long time scales (in the limit close to stationarity) we find a mixture of nodes from all institutions appear in the top 5 nodes. A previous study used betweenness and eigenvector centrality to show that most central institutions were not solely universities, but was also comprised of other institution types³⁰. Here, we show that the precise role of each node depends on the choice of scale, as already discussed in ref. 15.

Dimension in epidemic spreading. What about dynamical processes on networks? In Fig. 7a, we use an SIR model on Watts-Strogatz small-world networks³¹ and by scanning the infection probability β , we show that the local dimension of a node strongly predicts its infectiousness. Below the critical regime of large infectiousness, we find that infection probability is positively correlated with the scale, i.e. the size of the local neighbourhood that should be considered grows with the infection probability. However, near criticality β_{crit} (a threshold infection probability), we observe a behaviour similar to a phase transition, whereby the time scale that local dimension correlates best with node infectiousness diverges towards values near unity, corresponding to the largest scale of the local dimension.

We further computed the local dimension and SIR dynamics for small-world graphs whilst varying the probability of rewiring p parameter, to interpolate between near regular graphs to Erdős-Rényi random graphs. In Fig. 7b we observe that the relationship between the optimal scale to determine local dimension and infectiousness of a node disappears with the randomness of the network. At low β , node infectiousness is dominated by the distance from high degree nodes in a small-world graph and, as β increases, the spreading dynamics accelerates and nodes further away can be infected. A local dimension at longer time scales τ is therefore necessary to obtain a better prediction on node infectiousness. However, in Erdős-Rényi random networks all nodes are on average at equal distance from high degree nodes and no meaningful scale exists.

We find similar linear relationships between β and scale in a Delaunay grid graph (Fig. 7c) and the European powergrid (Fig. 7d). The decrease in scale for the local dimension to be a good predictor beyond β_{crit} for both graphs echoed the results of high probability re-wiring in small-world graphs, suggesting that global graph structure becomes less important if the infection probability is sufficiently high.

Graph classification from distributions of local dimensions.

Random graphs, such as the Watts-Strogatz graph used above, sit at the intersection of graph theory and probability theory, and are often used to investigate the properties of ‘typical’ graphs. Various models of random graphs exist to cover the diversity of complex networks encountered in the real-world, but the most commonly discussed are Erdős-Rényi, Watts-Strogatz, and Barabasi-Albert graphs. To understand whether the distribution of local dimension differed across these three types random graphs, we generated a large dataset with various choices of parameters to generate each type of random graphs of similar sizes (see “Methods”). We then computed the local dimension of each node of each graph and extracted three features from the distribution of local dimension (mean, standard deviation and skewness) and used a Random Forest model to classify between the random graph types. The classification model achieved 0.95 ± 0.014 accuracy with a stratified 10-fold split, suggesting that different random graphs types display inherently different dimensional properties. A Shap feature importance analysis revealed that the skewness and standard deviation of the distributions were most informative in differentiating the random graph types (Fig. 8(a)). The skewness and standard deviation of Barabasi-Albert graphs were larger

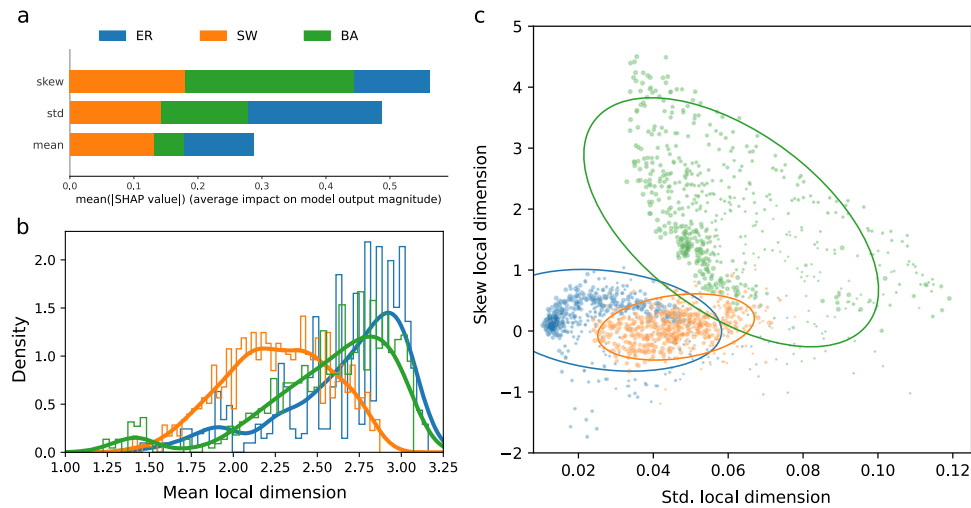


Fig. 8 Comparison of random networks (Erdos Rényi, Watts-Strogatz, Barabasi-Albert) by features derived from the distribution of local dimension (mean, standard deviation, skewness). For each graph type, 600 graphs with varying choices of parameters was chosen (see Methods). **a** The shap value for each feature reveals the importance of each feature in distinguishing random graph types in the trained random forest model. **b** A density histogram revealing the distribution of the mean local dimension of each random graph type. **c** Each graph is plotted by their skewness and standard deviation of local dimension, coloured by the random graph type and marker size is proportional to the number of nodes. The ellipses indicate a 2 standard deviation confidence interval using a Pearson correlation for each graph type.

reflecting their extremely broad and non-homogenous degree distribution. As expected, an overlap in the distribution of Erdős-Rényi and Watts-Strogatz graphs is observed (Fig. 8(c)) owing to the fact that Watts-Strogatz graphs were designed specifically to interpolate between lattices and fully disordered states (similar to, but not exactly Erdős-Rényi³²) via a rewiring of edges. Despite their overlap, Erdős-Rényi graphs display a smaller standard deviation, likely resulting from a more homogeneous degree distribution.

Discussion

In this paper we have introduced a new framework to define notions of dimensions not only on graphs, but on any space where a dynamical process (from which the Euclidean dimension can be inferred) can be defined. Our measure of dimension is defined using consensus dynamics on graphs, which is most similar to Euclidean diffusion, and naturally links with the dimension in the d-dimension diffusion equation. In this sense, our measure is intrinsically defined through the diffusive process taking place on a discrete system and recovers the intuitive definition of dimension as the system loses its discreteness. In doing so, we are also able to give a geometric meaning (through the notion of dimension) to the effect of boundaries and density inhomogeneities. We have shown the relevance of this approach to examine real-world systems such as protein dynamics, neuronal or social networks, ocean currents or epidemic spreading by examining the underlying graph structure.

Through various detailed studies with the relative dimension, probing local dimensions at various scales, or characterising entire graphs with the global dimension, we have provided evidence for the wide applicability of our dimension measures to both non-complex and complex networks (see SI for characterisation of degree distributions of graphs used in this paper). There are a variety of practical applications where probing network geometry is of great utility³³ and are within the scope of these dimension measures. For example, spatially modulated neurons (such as place cells or grid cells), whose network architecture plays a fundamental role in the representation of space and spatial memory, could be studied with our measures to understand the local and global lattice arrangement of firing fields³⁴.

Alternatively, our measures could be used to provide insights into the manifestation of material properties. For example, the angle at which two stacked layers of graphene are oriented relative to each other dictates the presence of superconductivity and fragile topology³⁵. Further analysis of graph classification problems using the distribution of dimension measures (relative or local) are also promising in view of our preliminary results using random generative networks.

Methods

Graph diffusion. A network (or a graph) G is a tuple $G = (V, E)$, consisting of the set of nodes $N = |V|$ vertices and $M = |E|$ edges connecting them. The network can be described by its $N \times N$ adjacency matrix which indicates the existence and the weight of a connection (edge) between each pair of nodes. On a graph, there are several non-equivalent definitions of diffusion, which are defined by different forms of the graph Laplacian. However, only one form corresponds to the Euclidean diffusion, described by the normalised Laplacian $L = K^{-1}(K - A)$ where K is the diagonal matrix of weighted degrees and A the weighted adjacency matrix¹⁶. Using the definition of the Laplacian, we can state the diffusion equation for a $N \times 1$ time-dependent node vector $p(t)$ as in Equation (4), which is also known as consensus dynamics³⁶. For an initial condition with a delta function of mass m at node i , the j th coordinate of the solution of Equation (4) is given by Eq. (5). For comparability across different graphs, we normalise the times of diffusion by the second smallest eigenvalue of the graph Laplacian, λ_2 (the spectral gap), thus $\tau = 1$ is the time scale for the diffusion to reach stationarity.

From our choice of Laplacian, the relative dimension matrix d (that we introduce in the next section) is symmetric if the initial masses m are chosen inversely proportional to the weighted node degrees.

In addition, to ensure that the stationary state of the diffusion sums to unity, we take $m_i = \bar{k}/(nk_i)$ where \bar{k} is the mean weighted degree and n is the number of nodes in the source. This is used in the protein example, where the initial mass are distributed on all the atoms of the allosteric or active site.

Comparison with fractal dimension. Looking more closely at our definition of relative dimension of Equation (6), it is proportional to the ratio of natural logarithms of peak amplitude and time, which displays similarities to the fractal based approaches where an approximate dimension can be derived from the ratio of natural logarithms of mass at a radius r ,

$$d \sim \frac{\log(M)}{\log(r)}, \tag{9}$$

where the mass M is simply the number of nodes within some link distance r ⁷.

Computational aspects. Python code to compute the relative, local and global dimensions is available at <https://github.com/barahona-research-group/DynGDdim>, based on the package NetworkX and numpy/scipy standard libraries.

Delaunay mesh with mass. We apply Delaunay triangulation to a 40 by 40 grid to return a weighted planar graph for which no point is inside the circumcircle of any triangle. The size of the grid is one unit of the code distance units. We define the weights of each edge as the inverse Euclidean lengths between points and thus obtain a discretisation of the plane. To simulate the gravitational lensing effect, we added additional nodes sampled from a Gaussian distribution with parameters with variance 0.05 in the unit square with various positions and number of nodes.

Protein graph construction. The graph representation of the proteins used in this work are computed using³⁷, an extension of³⁸. In short, from a pdb file, each atom is represented by a node, and bonds between atoms by an edge weighted by the energy of the bond. The choice of bonds is key to create a meaningful graph representation, and is explained in^{37,38}, see³⁹ to access the code.

Root-mean-square fluctuation calculations. Enzymatic proteins are inherently flexible and known to exhibit motions across a wide range of temporal and spatial scales. Using simulations, each atom can be assigned a root-mean-square fluctuation (RMSF). We calculate the RMSF using the CABS-flex 2.0 webserver which simulates protein dynamics using a coarse-grained protein model⁴⁰.

Protein dataset. We present here more details on the main set of proteins we used in this work.

HRas. HRas plays an important role in signal transduction during cell-cycle regulation⁴¹. Previous studies have shown that calcium acetate acts as an allosteric activator and its mechanism of allostery is mediated by a network of hydrogen bonds, involving structural water molecules, that link the allosteric site to the catalytic residue Q61⁴². We treat the allosteric and active sites, that are located at opposite ends of the protein (PDB ID: 3K8Y), as the source or target nodes in our relative dimension (since multiple atoms compose the allosteric and active sites, we use all nodes as the source of the diffusive process with a uniform distribution on them).

Lactose repressor (lac). As a second example, we examine the well-studied lactose repressor (lac) (PDB ID: 1EFA) in Fig. 3b, present in *E. coli* and which binds to the lac operon, a section of DNA, to inhibit the expression of proteins for the metabolism of lactose when no lactose is present^{43,44}. In its complete form, it consists of 4 monomers, with two binding sites to a single DNA strand, inhibiting the genes located between them. The combination of two monomers co-operate to form one of the two binding sites (orange region in Fig. 3b). On each monomer there is an allosteric site for the binding of NPF molecules that activate the lac repressor.

PKD1. PKD1 is a well-known protein Kinase (PDB ID: 3ORX) that is implicated in the progression of Melanoma's⁴⁵. The allosteric site of PKD1 is a sequence of amino acids, called the PKD1 interacting fragment (PIF), that binds to a phosphate on the catalytic domain. This binding triggers a signal to start the phosphorylation of the activation loop of the substrates at the ATP pocket, or active site²⁰. The crystallographic structure (PDB ID: 3ORX) used for our analysis has the molecule B14 bound at the active site⁴⁵ via three hydrogen bonds to a region of high relative dimension, and interacts through hydrophobic forces on a region of low relative dimension.

Fluorescence correlation microscopy experiments. Protein plasmids of *Aquifex* Adenylate Kinase (ID:18092 Plasmid:peT3a-AqAdk/MVGDH) were purchased from AddGene as deposited by 'Dorothee Kern Lab Plasmids'. The plasmids were already encoded with two cysteine mutations for maleimide conjugation. ADK was expressed in a 1 litre culture BL21 (DE3) cells via inoculation with 1 mM IPTG. BugBuster was used for cell lysis and TCEP and protease inhibitor was added to the lysate. ADK was purified via HIS-tag with a gravi-trap (GE-healthcare), and a PD-10 column was used to remove imidazole and exchange into protein buffer (20 mM TRIS, 50 mM NaCl). TCEP and protease inhibitor were added throughout the purification process. Alexa 488-labelled ADK was prepared overnight using 20 μ M protein with molar ratio 1:10 of protein:Alexa 488. Excess dye was removed using HIS-tag purification and a PD-10 column. A Typhoon was used to examine the gel of the purified-labelled ADK product and showed no excess fluorophore. The label sites for the FRET experiment were Tyr 52 (AMP_{bd} domain) changed to Cys and Val 145 changed to Cys (lid domain)⁴⁶. Samples were diluted to 200 pM in pH 7.5 FRET buffer (20 mM TRIS, 50 mM NaCl) with 0.3 mg/ml BSA to prevent surface adsorption. Measurements were taken at thermal equilibrium such that all processes under analysis are statistical fluctuations around the equilibrium. Freely diffusing single molecules were detected using a home-built dual-channel confocal fluorescence microscope. A tunable wavelength argon ion laser (model 35LAP321-230, Melles Griot, Carlsbad, CA) was set to 514.5 nm to excite Alexa 488. The beam was focused into the sample solution to a diffraction-limited spot with a high numerical aperture oil-immersion objective (Nikon Plan Apo TIRF 60x, NA 1.45). The closer refractive indexes of oil and glass relative to water and glass make oil immersion preferable due to reduced light reflection. Type FF immersion oil (Cargille, USA) was used due to its negligible fluorescent properties. The obtained

fluctuations of fluorescence intensity are autocorrelated. We fit the autocorrelation curves with a global model that includes components for triplet excitation, conformational dynamics and diffusion, with the assumption that they differed by a factor of 1.6 to distinguish the components,

$$G(\tau) = G(0) \left(\frac{1}{1 + \frac{\tau}{T_D}} \right) \left(1 - F + F e^{-\frac{\tau}{\tau_m}} \right) \left(1 - F_2 + F_2 e^{-\frac{\tau}{\tau_c}} \right),$$

where τ_c , τ_m and T_D are the dynamical time scales of the protein conformational dynamics, mean triplet relaxation and the protein diffusion respectively. F_1 is the fraction of molecules entering the triplet state and F_2 is the fraction of molecules conformationally fluctuating.

Root-mean-square fluctuation analysis. We use the cabs flex 2 server that generated fast simulations of near-native dynamics. The dynamics uses Monte Carlo dynamics and an asymmetric metropolis scheme. CABS is a well established coarse-grained (i.e. atoms are combined into larger units) protein modelling tool. CABS uses a forcefield derived from statistical regularities seen in known protein structures, and it includes side-chain-side-chain mean field potentials, coarse-grained models of main chain hydrogen bonds, and local peptide-chain geometric preferences. The solvent effect is accounted for in an implicit fashion through protein structure statistics used in the derivation of the CABS force field. The dynamics of CABS-based coarse-grained proteins is simulated by a random series of local conformational transitions (controlled by a Monte Carlo method). The results show strong similarities with fully atomistic MD simulations. (Description here <http://biocomp.chem.uw.edu.pl/sites/default/files/publications/ct300854w.pdf>) The resulting trajectory from the MD simulation is analysed and clustered to a representative ensemble of protein models that reflect the flexibility of the input structure. In short, the simulation (like other MD simulations) examines the dynamic evolution of interacting units (atoms or coarse-grained units). The trajectories are determined by solving Newtons equations of motion, where the forces between units are determined by the proposed forcefield. Therefore, inherently one can study the thermodynamic properties of a system via a MD simulation.

SIR model. For the example with SIR dynamics, we simulated the standard SIR model on networks, using the fast approximation of⁴⁷, with open sourced code available at <https://github.com/springer-math/Mathematics-of-Epidemics-on-Networks> and estimated the infectiousness of each node as the averaged number of removed nodes when the spread started from this node over 500 realisation of the dynamics. To estimate the critical value for the infectiousness β , we computed the average infectability across all nodes for each β and estimated β_{crit} as the value for which half of the nodes are infected.

Graph classification dataset. We generated 600 graphs of each of the three classes, Erdos-Renyi, Barabasi-Albert and Small Worlds. We sampled the number of nodes with 10 bins from 100 to 1000, and repeated that 3 times with different random seed. For in each case, we created 20 networks of each types with the following range of parameters: ER from with probabilities from 0.03 to 0.1. BA with number of edges per nodes from 1 to 20 and SW with probability from 0.1 to 0.7 and number of neighbours from 5 to 10. Improvements to the random graph classification results can be made using other graph theoretic features⁴⁸.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplemental Information files.

Code availability

The code is shared under the GNU General Public License v3.0. It can be found at <https://github.com/barahona-research-group/DynGDim> and 10.5281/zenodo.6496778⁴⁹.

Received: 18 June 2021; Accepted: 13 May 2022;

Published online: 02 June 2022

References

1. Erdős, P., Harary, F. & Tutte, W. T. On the dimension of a graph. *Mathematika* **12**, 118–122 (1965).
2. Lovász, L. *Graphs and Geometry*, vol. 65 (American Mathematical Soc., 2019).
3. Linial, N., London, E. & Rabinovich, Y. The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15**, 215–245 (1995).

4. Csányi, G. & Szendrői, B. Fractal–small-world dichotomy in real-world networks. *Phys. Rev. E* **70**, 016122 (2004).
5. Gastner, M. T. & Newman, M. E. The spatial structure of networks. *Eur. Phys. J. B Condens. Matter Complex Syst.* **49**, 247–252 (2006).
6. Shanker, O. Defining dimension of a complex network. *Mod. Phys. Lett. B* **21**, 321–326 (2007).
7. Daqing, L., Kosmidis, K., Bunde, A. & Havlin, S. Dimension of spatially embedded networks. *Nat. Phys.* **7**, 481–484 (2011).
8. Silva, F. N. & Costa, L. d. F. Local dimension of complex networks. Preprint at <https://arxiv.org/abs/1209.2476> (2012).
9. Pu, J., Chen, X., Wei, D., Liu, Q. & Deng, Y. Identifying influential nodes based on local dimension. *EPL (Europhys. Lett.)* **107**, 10010 (2014).
10. Bian, T. & Deng, Y. Identifying influential nodes in complex networks: a node information dimension approach. *Chaos* **28**, 043109 (2018).
11. Wen, T., Pelusi, D. & Deng, Y. Vital spreaders identification in complex networks with multi-local dimension. *Knowl. Based Syst.* **195**, 105717 (2020).
12. Reuveni, S., Granek, R. & Klafter, J. Anomalies in the vibrational dynamics of proteins are a consequence of fractal-like structure. *Proc. Natl Acad. Sci. USA* **107**, 13696–13700 (2010).
13. Lacasa, L. & Gómez-Gardenes, J. Correlation dimension of complex networks. *Phys. Rev. Lett.* **110**, 168703 (2013).
14. Peach, R. L., Arnaudon, A. & Barahona, M. Semi-supervised classification on graphs using explicit diffusion dynamics. *Found. Data Sci.* **2**, 19 (2020).
15. Arnaudon, A., Peach, R. L. & Barahona, M. Scale-dependent measure of network centrality from diffusion dynamics. *Phys. Rev. Res.* **2**, 033104 (2020).
16. Singer, A. From graph to manifold laplacian: the convergence rate. *Appl. Comput. Harmonic Anal.* **21**, 128–134 (2006).
17. Einstein, A. Lens-like action of a star by the deviation of light in the gravitational field. *Science* **84**, 506–507 (1936).
18. Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* (Macmillan, 1973).
19. Müller-Hill, B. & Oehler, S. *The Lac Operon* (Walter de Gruyter New York, 1996).
20. Biondi, R. M., Kieloch, A., Currie, R. A., Deak, M. & Alessi, D. R. The pif-binding pocket in pdk1 is essential for activation of s6k and sgk, but not pkb. *EMBO J.* **20**, 4380–4390 (2001).
21. Reuveni, S., Granek, R. & Klafter, J. Proteins: coexistence of stability and flexibility. *Phys. Rev. Lett.* **100**, 208101 (2008).
22. Peach, R. Exploring Protein Dynamics Using Graph Theory and Single-molecule Spectroscopy. *Imperial College London*, Ph.D. thesis (2017).
23. Peach, R. L. et al. Unsupervised graph-based learning predicts mutations that alter protein dynamics. *bioRxiv* Preprint at <https://doi.org/10.1101/847426> (2019).
24. De Nooy, W., Mrvar, A. & Batagelj, V. *Exploratory Social Network Analysis with Pajek: Revised and Expanded Edition for Updated Software*, Vol. 46 (Cambridge University Press, 2018).
25. Smith, D. A. & White, D. R. Structure and dynamics of the global economy: network analysis of international trade 1965–1980. *Soc Forces* **70**, 857–893 (1992).
26. Yan, G. et al. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* **550**, 519–523 (2017).
27. Bacik, K. A., Schaub, M. T., Beguerisse-Diaz, M., Billeh, Y. N. & Barahona, M. Flow-based network analysis of the *Caenorhabditis elegans* connectome. *PLoS Comp. Biol.* **12**, 1511.00673 (2016).
28. Estrada, E. & Hatano, N. Communicability in complex networks. *Phys. Rev. E* **77**, 036111 (2008).
29. Faccin, M., Schaub, M. T. & Delvenne, J.-C. State aggregations in Markov chains and block models of networks. *Phys. Rev. Lett.* **127**, 078301 (2021).
30. Aref, S., Friggens, D. & Hendy, S. Analysing scientific collaborations of New Zealand institutions using scopus bibliometric data. In *Proc of the Australasian Computer Science Week Multiconference, Association for Computing Machinery*, 1–10 (2018).
31. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
32. Maier, B. F. Generalization of the small-world effect on a model approaching the Erdős–Rényi random graph. *Sci. Rep.* **9**, 1–9 (2019).
33. Boguna, M. et al. Network geometry. *Nat. Rev. Phys.* **3**, 114–135 (2021).
34. Ginosar, G. et al. Locally ordered representation of 3d space in the entorhinal cortex. *Nature* **596**, 1–6 (2021).
35. Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
36. Masuda, N., Porter, M. A. & Lambiotte, R. Random walks and diffusion on networks. *Phys. Rep.* **716**, 1–58 (2017).
37. Song, F., Yaliraki, S. N. & Barahona, M. Bagpype: A python package for the construction of atomistic, energy-weighted graphs from biomolecular structures. figshare preprint figshare:10.6084 (2021).
38. Amor, B. R., Schaub, M. T., Yaliraki, S. N. & Barahona, M. Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nat. Commun.* **7**, 12477 (2016).
39. Mersmann, S. et al. ProteinLens: a web-based application for the analysis of allosteric signalling on atomistic graphs of biomolecules. *Nucleic Acids Res.* <https://doi.org/10.5281/zenodo.6496778> (2021).
40. Kuriata, A. et al. Cabs-flex 2.0: a web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res.* **46**, W338–W343 (2018).
41. McCormick, F. Ras-related proteins in signal transduction and growth control. *Mol. Reprod. Dev.* **42**, 500–506 (1995).
42. Buhman, G., Holzapfel, G., Fetis, S. & Mattos, C. Allosteric modulation of ras positions q61 for a direct role in catalysis. *Proc. Natl Acad. Sci. USA* **107**, 4931–4936 (2010).
43. Becker, N. A., Greiner, A. M., Peters, J. P. & Maher III, L. J. Bacterial promoter repression by dna looping without protein–protein binding competition. *Nucleic Acids Res.* **42**, 5495–5504 (2014).
44. Wilson, C., Zhan, H., Swint-Kruse, L. & Matthews, K. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell. Mol. Life Sci.* **64**, 3–16 (2007).
45. Sadowsky, J. D. et al. Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc. Natl Acad. Sci. USA* **108**, 6056–6061 (2011).
46. Henzler-Wildman, K. A. et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).
47. Kiss, I. Z., Miller, J. C., Simon, P. L. et al. *Mathematics of Epidemics on Networks* 598 (Springer, 2017).
48. Peach, R. L. et al. hcga: Highly comparative graph analysis for network phenotyping. *Patterns* **2**, 100227 (2021).
49. Peach, R., Arnaudon, A. & Barahona, M. Relative, local and global dimension in complex networks: code. <https://doi.org/10.5281/zenodo.6496779> (2022).

Acknowledgements

We thank David Infield, Thomas Higginson, Francesca Vianello, Florian Song, Paul Expert, Asher Mullokandov and Sophia Yaliraki for valuable discussions. We acknowledge funding through EPSRC award EP/N014529/1 supporting the EPSRC Centre for Mathematics of Precision Healthcare at Imperial. R.P. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 424778381-TRR 295. A.A. was supported by funding to the Blue Brain Project, a research centre of the École polytechnique fédérale de Lausanne (EPFL), from the Swiss government’s ETH Board of the Swiss Federal Institutes of Technology.

Author contributions

R.P., A.A., M.B. contributed to conceiving the original idea. RP and AA contributed to writing the code. R.P. and A.A. contributed to the main analyses. R.P., A.A., M.B. contributed to writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30705-w>.

Correspondence and requests for materials should be addressed to Mauricio Barahona.

Peer review information *Nature Communications* thanks Filippo Radicchi, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2022