Relative, local and global dimension in complex networks

Robert Peach
University Hospital of Wuerzburg https://orcid.org/0000-0002-8738-5825
Alexis Arnaudon
Imperial College London
Mauricio Barahona (m.barahona@imperial.ac.uk)
Imperial College London https://orcid.org/0000-0002-1089-5675

Article

Keywords: Ideal Notions of Dimension, Boundary Constraints, Inhomogeneities Distortions, Discrete Systems, Diffusive Process, Allosteric Communication, Scale-dependent Definition

Posted Date: July 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-637874/v1

Version of Record: A version of this preprint was published at Nature Communications on June 2nd, 2022. See the published version at https://doi.org/10.1038/s41467-022-30705-w.

Relative, local and global dimension in complex networks

Robert Peach,[1,2,*] Alexis Arnaudon,[1,3,*] and Mauricio Barahona[1]

[1] *Department of Mathematics, Imperial College, London SW7 2AZ, UK*
[2] *Department of Neurology, University Hospital Würzburg, Würzburg, Germany*
[3] *Blue Brain Project, École Polytechnique Fédérale de Lausanne (EPFL), CampusBiotech, 1202 Geneva, Switzerland*

(Dated: June 9, 2021)

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.

One of the first forays into graph dimensionality originated with Erdös, when he explored the embedding of graphs into a minimum finite dimensional Euclidean space [1]. This line of study helped realise the algorithmic importance of geometric interpretations of graphs [2] but was unfortunately no more than a by-product of the graph embedding process, yielding little actionable information [3]. Later, by characterising the fractal properties of complex networks, a measure of network dimension was defined in terms of the scaling property of a networks 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 percolation [7], 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 [8] and have been used to define a node-centric dimension that can identify influential/central nodes [9, 10] or vital spreaders in infection models [11].

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 heterogeneities [5]. Similarly, in classic papers such as [12] where the dimension of a node is defined using the decay rate of diffusion, the same assumptions of homogeneity are required. Take for example 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 our *relative dimension*.

We start with the Green's function of the diffusion equation

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

which, together with an initial condition as a delta function at some position $\mathbf{x_0}$, provides a solution of diffusion equation as $p(\mathbf{x}, t) = G_t(\mathbf{x} - \mathbf{x}_0)$. As already considered in [13], these solutions have a maxima in there transient response at any other location $\mathbf{x}$, at time $\widehat{t}$ and amplitude $\widehat{p}$ given as

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

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

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

Notice that 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 this Equation (3) that allows us to generalise 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 is intrinsic as it does not rely on any Euclidean embedding, but only on the existence of diffusion dynamic on

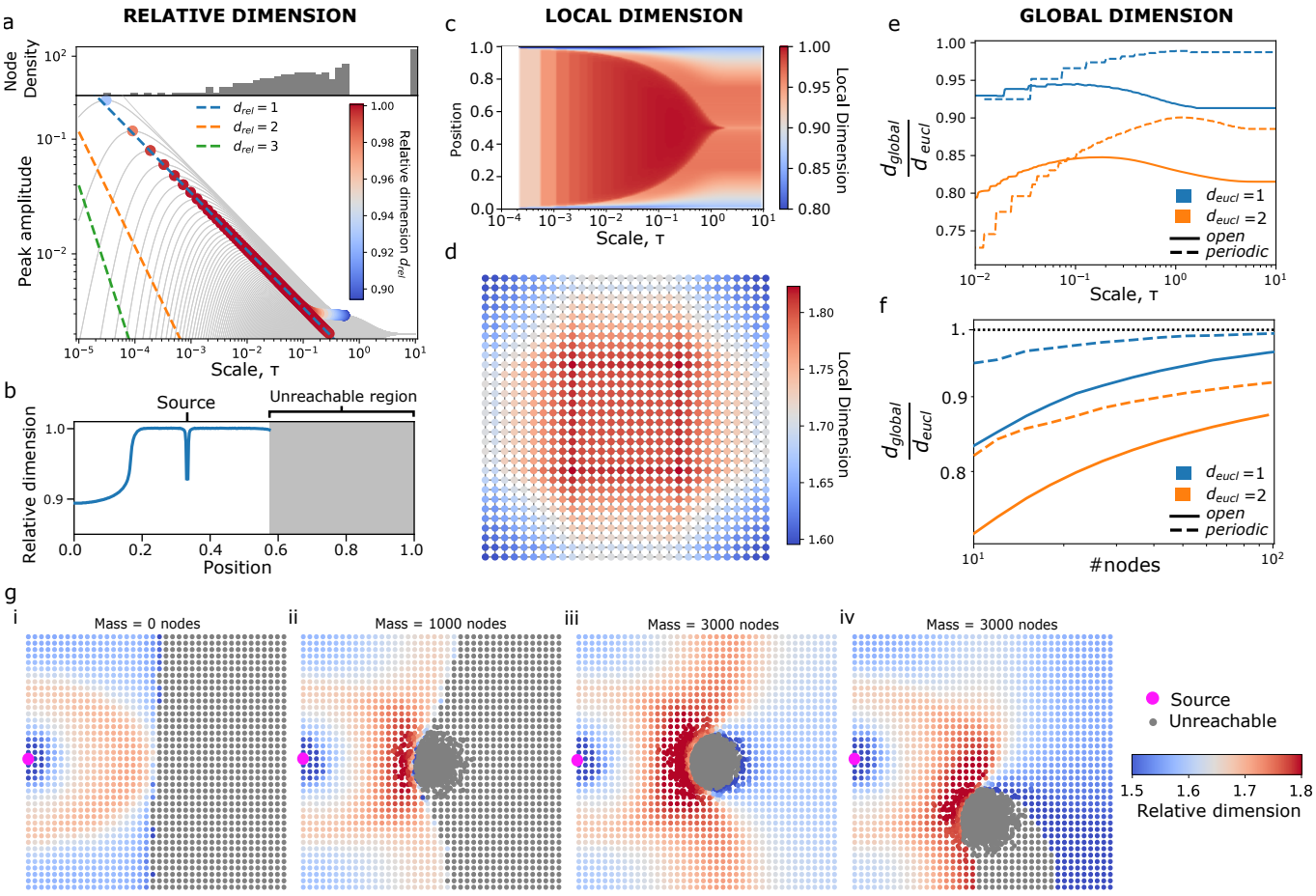

FIG. 1. **a-c** Line graph example with $n = 500$ nodes representing the interval $[0, 1]$. **a** The relative dimension of nodes given a source located at $x = 1/3$. The grey lines are the transient responses of the (non-source) nodes and the position of the peaks in the transient responses are highlighted by dots, coloured by their relative dimension. Dashed lines are isolines of integer relative dimensions. Top inset, a histogram of transient response peaks where the far right bin corresponds to nodes where no peak was observed. **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. Below, we also observe the increasing effect of the boundaries. **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 as a function of scale for the same line and grid graph as well as their periodic equivalent graphs, illustrating differing behaviors 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 to the underlying Euclidean dimension d_{eucl} of $d_{global} = \max_\tau \mathcal{D}(\tau)$, showing a faster convergence for lower dimensional spaces and periodic grids. **g** The relative dimension from a point source (pink dot) to other nodes in a Delaunay grid graph (edges not shown). Grey nodes indicate unreachable nodes for which no transient response peak was detected. The regular grid is shown in **i**, and an additional mass is added in **ii-iv**, with varying mass and position, showing a effect similar to gravitational lensing.

the original space. In particular, on graphs we can use the standard diffusion process

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

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 limit [14]). With delta function at node i with mass m_i as our initial condition, the j-th coordinate of the solution of Equation (4) is given by the heat kernel

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

Following the previous derivation, we measure the time $\widehat{t}_{ij}$ and amplitude $\widehat{p}_{ij}$ of transient response peaks for a node j relative to the node i by numerically solving Equation (5), and compute the $N \times N$ matrix or relative dimensions

$$d_{ij} = \frac{-2\ln \widehat{p}_{ij}}{\ln\left(4e\pi\sigma\widehat{t}_{ij}\right)}\,. \tag{6}$$

To illustrate our definition of relative dimension, we used a line graph (Figure 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 $d = 1$. It is then natural to define the *local dimension* of the source node by averaging the relative dimension of the nodes displaying a peak before a given time τ as

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

where $\mathbb{1}_{\widehat{t}_{ij}<\tau}$ is the indicator function for nodes that display a transient response peak before time τ. Whilst the local dimension can be likened to a measure of centrality, it is directly capturing 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 center 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. The '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 at each scale to obtain the global dimension

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

still dependent on τ. In Fig. 1e we display the global dimension (as a ratio to the expected Euclidean dimension) for the line and grid graphs and their periodic equivalents. The non-periodic graphs display a maximum in global dimension 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 this curve can be seen as an approximation of the Euclidean dimension, whereas the global dimension at largest scale characterise 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, and can be considered together. In addition, 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 grids is only affected by the topology, and thus converges faster.

To develop some intuition for our measure of relative dimension, we consider a simple constructive example using Delaunay meshes (Figure 1g). Given a source-node located at the left boundary or an homogeneous delaunay mesh, relative dimension displays an inhomogeneous distribution radially from the source until nodes are unreachable (Figure 1g(i)). 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. 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 if the mass is sufficiently large. Small masses are reminiscent of weak lensing (Figure 1g(ii)), whereas larger masses are closer to strong lensing (Figure 1g(iii)) [15]. 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 (Figure 1g(iv)) shows a coupling between the lensing effect and the presence of the boundary. All three possible effects, boundaries, topology and imhomogeneities 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 intrinsic graph theoretical measure.

To further illustrate the benefits of relative dimension we examine 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 Figure 2 (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. 2a(i)), whilst in reverse the allosteric site is unreachable to the active site (Fig. 2a(ii)). Whilst an exact statement of allosteric mechanism is not our purpose here, its interesting to note that a low relative dimension suggests a more 'direct' or 'funneled' communication from the allosteric site to the active site. Moreover, the asymmetry of the communication suggests that each half of the protein is purposed for diffusion in opposing directions. 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 region [16]. Given that the allosteric mechanism is cooperative, we don't expect 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. 2b(i)) vs. both allosteric sites as sources

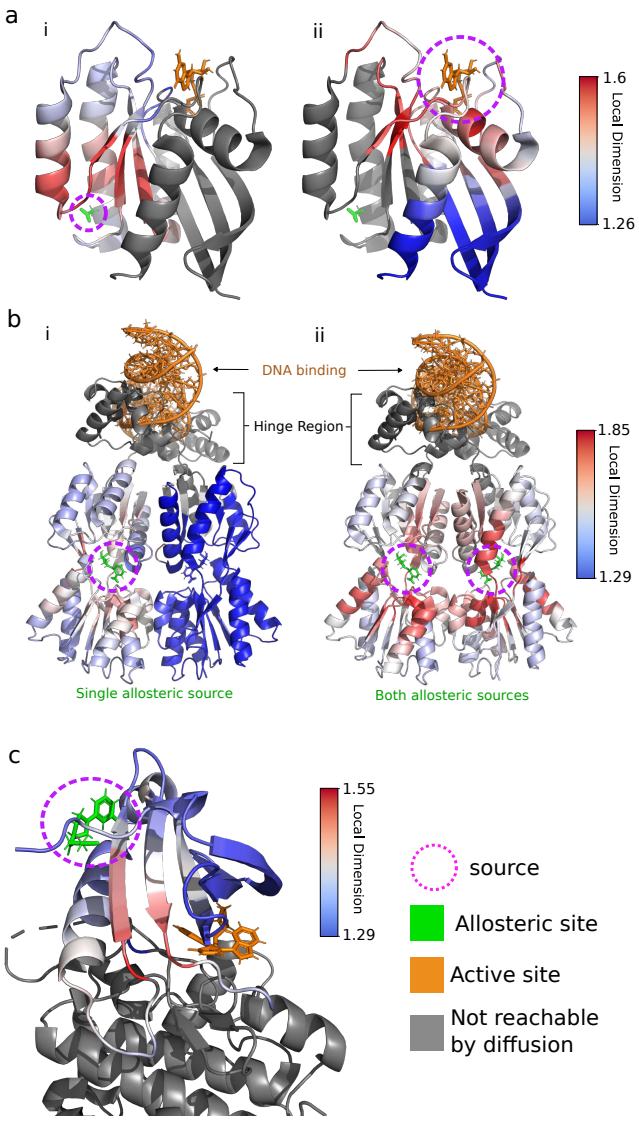

FIG. 2. Relative dimensionality in allosteric proteins. **a** The relative dimensionality of all atoms in HRas GTPase (PDB ID: 3K8Y) given (i) the allosteric site and (ii) the active site as the source of diffusion. **b** Relative dimension in the multi-allosteric site Lac Repressor protein (PDB ID: 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).

simultaneously (Fig. 2b(ii)). We find that when binding NFP to just one monomer the relative dimension across the entire protein is lowerwhen 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 site [17], and thus we would expect direct communication between the active and allosteric site. Using the allosteric site as the source of our diffusion, we find that the activation loop is the only protein region that is reachable and returns a measure of relative dimension (Figure 2c). Similarly to HRas, we observe a lower relative dimension at the active site suggesting that communication is more focused from the allosteric to active site.

Whilst the relative dimension provides insights into allostery, we can look to the local and global dimension to examine protein dynamics. In Figure 3(a), we show a strong correlation between the local dimension and $\log_{10}(1/\text{RMSF})$ of residues for Figure 3a(i) an unglycosylated antibody CH2 domain and Figure 3a(ii) an Estrogen Related Receptor g protein, and thus a larger local dimension reduces the flexibility. To examine this further, we plotted the Pearson correlation between local dimension and $\log_{10}(1/\text{RMSF})$ for 12 randomly chosen proteins in Figure 3(b). We see that at middling to long timescales of diffusion the correlation plateaus 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 Figure 3(c), we find a strong correlation (Pearson $\sigma = 0.73$) between global dimension and the $\log_{10}(1/\langle\text{RMSF}\rangle)$ 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 dimensionality is generally < 2 and decreases with an increase in flexibility [12, 18].

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 (Figure 3d). The AMP and lid domains are known to open and close around substrate, agreeably we find both have a lower local dimension relative to the core domain (Figure 3e). Furthermore, we find the AMP domain to have a lower average local dimension than the lid domain in both conformations, a result that we validated using experimental fluorescence correlation spectroscopy that shows the AMP domain to open and close at a faster rate ($16.2\mu s$) than the lid domain ($46.6\mu s$) [19, 20].

Until now we have considered systems with spatially embedded networks and the relationships between dimension and structure. But what about dynamical processes on networks? Using an SIR model on Watts Strogatz small-world networks [21] and by scanning the infection probability β, we show that the local dimension of a node strongly predicts its infectiousness (Fig. 4a). Below the critical regime of large infectiousness, we find that β 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 we observe a behavior similar to a phase transition, whereby the best time scale diverges towards values near unity, corresponding to the largest scale of the local

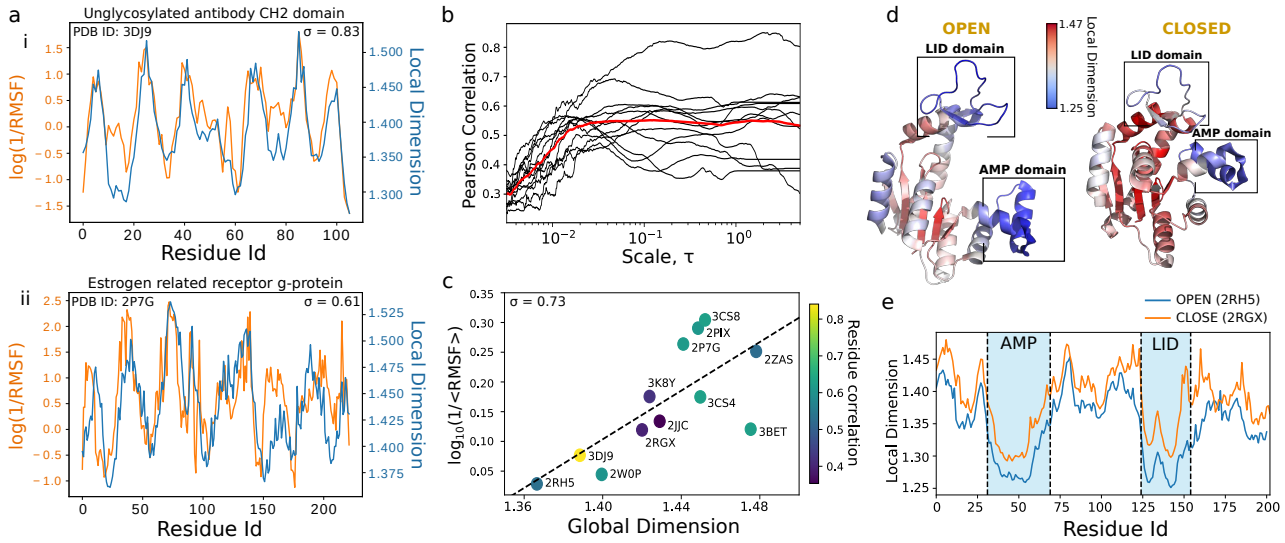

FIG. 3. The relationship between root-mean-square fluctuations (RMSF) of protein residues and their local and global dimension. **a** The log-inverse RMSF vs local dimensionality for each residue in **i** unglycosylated antibody CH2 domain, **i** Estrogen Related Receptor g. **b** The positive correlation (0.539) between local dimension and log-inverse RMSF at the residue level across 12 different proteins as a function of Markov time. **c** A strong positive correlation between global dimensionality of 12 proteins against log-inverse RMSF. **d** The local dimension of each residue (i) mapped onto Aquifex Adenylate Kinase in the open (PDB ID: 2RH5) and closed (PDB ID: 2RGX) conformations and (ii) plotted by residue id.

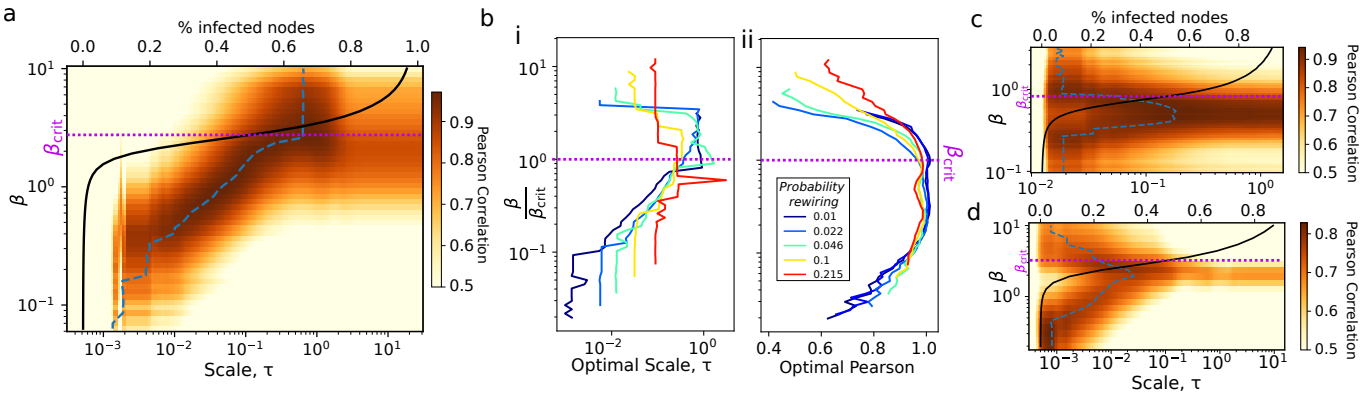

FIG. 4. **a** Heatmap of Pearson correlation between local dimension and node infectiousness for small-world graph ($n = 100$, average degree $k = 10$, probability of rewiring $p = 0.015$). The black line is the average proportion of infected nodes given a single-seed node for a given infection probability β. The transition from low to high proportion of infected nodes indicates the critical point. 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 of the small-world graph for **c** a Delaunay grid graph ($n = 400$) and **d** the European powergrid network, and show similar linear relationships between scale and infection probability β prior to criticality.

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 random graphs. We find that the relationship between the optimal scale to determine local dimension and infectiousness of a node disappears with the randomness of the network (Fig. 4). At low β, node infectiousness is determined by the distance from hubs in a small-world graph and as β increases, the spreading dynamics are faster and nodes further away can be infected. A local dimension with larger time horizons is therefore necessary to obtain a better prediction on node infectiousness. However, in random networks all nodes are on average at equal distance from hubs and no meaningful scale exists. We find similar linear relationships between β and scale in a Delaunay grid graph (Fig. 4c) and the European power grid (Fig. 4d). The decrease in scale for the local dimension to be a good predictor beyond β 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.

In this paper we have introduced a new framework to define intrinsic notions of dimensions not only on graphs, but on any spaces where a diffusion process can be defined, or even more generally, where a dynamical process from which we can infer the Euclidean dimension can be defined. We have shown the relevance of this approach to examine real world systems such as protein dynamics or epidemic spreading by exploiting the underlying graph structure, but many other applications are within the scope of these three measures, through detailed studies with the relative dimension, probing local dimensions at various scales, or characterising entire graphs with the global dimension.

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 and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 424778381-TRR 295.

* These authors contributed equally.

[1] P. Erdös, F. Harary, and W. T. Tutte, Mathematika **12**, 118 (1965).
[2] L. Lovász, *Graphs and geometry*, Vol. 65 (American Mathematical Soc., 2019).
[3] N. Linial, E. London, and Y. Rabinovich, Combinatorica **15**, 215 (1995).
[4] G. Csányi and B. Szendrői, Physical Review E **70**, 016122 (2004).
[5] M. T. Gastner and M. E. Newman, The European Physical Journal B-Condensed Matter and Complex Systems **49**, 247 (2006).
[6] O. Shanker, Modern Physics Letters B **21**, 321 (2007).
[7] L. Daqing, K. Kosmidis, A. Bunde, and S. Havlin, Nature Physics **7**, 481 (2011).
[8] F. N. Silva and L. d. F. Costa, arXiv preprint arXiv:1209.2476 (2012).
[9] J. Pu, X. Chen, D. Wei, Q. Liu, and Y. Deng, EPL (Europhysics Letters) **107**, 10010 (2014).
[10] T. Bian and Y. Deng, Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 043109 (2018).
[11] T. Wen, D. Pelusi, and Y. Deng, Knowledge-Based Systems , 105717 (2020).
[12] S. Reuveni, R. Granek, and J. Klafter, Proceedings of the National Academy of Sciences **107**, 13696 (2010).
[13] A. Arnaudon, R. L. Peach, and M. Barahona, Physical Review Research **2**, 033104 (2020).
[14] A. Singer, Applied and Computational Harmonic Analysis **21**, 128 (2006).
[15] C. W. Misner, K. S. Thorne, J. A. Wheeler, *et al.*, *Gravitation* (Macmillan, 1973).
[16] B. Müller-Hill and S. Oehler, *The lac operon* (Walter de Gruyter New York:, 1996).
[17] R. M. Biondi, A. Kieloch, R. A. Currie, M. Deak, and D. R. Alessi, The EMBO journal **20**, 4380 (2001).
[18] S. Reuveni, R. Granek, and J. Klafter, Physical review letters **100**, 208101 (2008).
[19] R. Peach, *Exploring protein dynamics using graph theory and single-molecule spectroscopy*, Ph.D. thesis, Imperial College London (2017).
[20] R. L. Peach, D. Saman, S. N. Yaliraki, D. R. Klug, L. Ying, K. R. Willison, and M. Barahona, bioRxiv , 847426 (2019).
[21] D. J. Watts and S. H. Strogatz, nature **393**, 440 (1998).
[22] N. Masuda, M. A. Porter, and R. Lambiotte, Physics reports **716**, 1 (2017).
[23] F. Song, S. N. Yaliraki, and M. Barahona, (2021), https://doi.org/10.6084/m9.figshare.14039723.v1.
[24] B. R. Amor, M. T. Schaub, S. N. Yaliraki, and M. Barahona, Nature communications **7**, 12477 (2016).
[25] S. Mersmann, L. Strömich, F. J. Song, N. Wu, F. Vianello, M. Barahona, and S. Yaliraki, Nucleic Acids Research (2021), 10.1093/nar/gkab350.
[26] A. Kuriata, A. M. Gierut, T. Oleniecki, M. P. Ciemny, A. Kolinski, M. Kurcinski, and S. Kmiecik, Nucleic acids research **46**, W338 (2018).
[27] F. McCormick, Molecular reproduction and development **42**, 500 (1995).
[28] G. Buhrman, G. Holzapfel, S. Fetics, and C. Mattos, Proceedings of the National Academy of Sciences **107**, 4931 (2010).
[29] N. A. Becker, A. M. Greiner, J. P. Peters, and L. J. Maher III, Nucleic acids research **42**, 5495 (2014).
[30] C. Wilson, H. Zhan, L. Swint-Kruse, and K. Matthews, Cellular and molecular life sciences **64**, 3 (2007).
[31] J. D. Sadowsky, M. A. Burlingame, D. W. Wolan, C. L. McClendon, M. P. Jacobson, and J. A. Wells, Proceedings of the National Academy of Sciences **108**, 6056 (2011).
[32] K. a. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. a. Wilson, G. a. Petsko, M. Karplus, C. G. Hübner, and D. Kern, Nature **450**, 838 (2007).
[33] I. Z. Kiss, J. C. Miller, P. L. Simon, *et al.*, Cham: Springer **598** (2017).

METHOD

Graph diffusion. A network (or a graph) G is a tuple $G = (\mathcal{V}, \mathcal{E})$, consisting of the set of nodes $N = |\mathcal{V}|$ vertices and $M = |\mathcal{E}|$ 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 forms 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 matrix [14]. Using the definition of the Laplacian, we can state the diffusion equation for a $N \times 1$ time-dependent node vector $\mathbf{p}(t)$ as (4), which

is also known as consensus dynamics [22]. For an initial condition with a delta function of mass m at node i, the j-th coordinate of the solution of (4) is given by (5).

For comparability across different graphs, we normalise the Markov time of diffusion by the second smallest eigenvalue of the graph Laplacian, λ_2 (the spectral gap), thus $\tau = 1$ is the timescale 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 $m_i = \overline{k}/(nk_i)$ where $\overline{k}$ 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/DynGDim, 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 using [23], an extension of [24]. 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 [23, 24], see [25] 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 model [26].

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 regulation [27]. 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 Q61 [28]. Here, 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 Figure 2b, 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 [29, 30]. 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 Figure 2b). On each monomer there is an allosteric site for the binding of NPF molecules that activate the lac repressor.

PDK1. Our final allosteric protein is a well-known protein Kinase called PDK1 (PDB ID: 3ORX) that is implicated in the progression of Melanoma's [31]. The allosteric site of PDK1 is a sequence of amino acids, called the PDK1 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 site [17]. The crystallographic structure (PDB ID: 3ORX) used for our analysis has the molecule BI4 bound at the active site [31] via three hydrogen bounds 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, 5 0mM NaCl). TCEP and protease inhibitor were added throughout the purification process. Alexa 488-labelled ADK was prepared overnight using 20 μM 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) [32]. Samples were diluted to 200 pM in pH 7.5 FRET buffer (20 mM TRIS, 50mM 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,

$$G(\tau) = G(0)\left(\frac{1}{1+\frac{\tau}{\tau_D}}\right)\left(1-F+Fe^{\frac{\tau}{\tau_m}}\right)\left(1-F_2+F_2e^{\frac{\tau}{\tau_{conf}}}\right),$$

where τ_c, τ_m and τ_D are the dynamical timescales 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.

SIR. For the example with SIR dynamics, we simulated the standard SIR model on networks, using the fast approximation of [33], 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.