

Unfolding the multiscale structure of networks with dynamical Ollivier-Ricci curvature

Adam Gosztolai (✉ adam.gosztolai@epfl.ch)

École Polytechnique Fédérale de Lausanne <https://orcid.org/0000-0002-0699-5825>

Alexis Arnaudon

Imperial College London

Article

Keywords: network structure, learning algorithms, network processes, Ollivier-Ricci curvature

Posted Date: February 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-222407/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on July 27th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-24884-1>.

Unfolding the multiscale structure of networks with dynamical Ollivier-Ricci curvature

Adam Gosztolai*¹ and Alexis Arnaudon^{†2}

¹*Neuroengineering Laboratory, Brain Mind Institute & Interfaculty Institute of Bioengineering, EPFL, Lausanne, Switzerland*

²*Department of Mathematics, Imperial College London, London, United Kingdom*

Abstract

Defining the geometry of networks is typically associated with embedding in low-dimensional spaces such as manifolds. This approach has helped design efficient learning algorithms, unveil network symmetries and study dynamical network processes. However, the choice of embedding space is network-specific, and incompatible spaces can result in information loss. Here, we define a dynamic edge curvature for the study of arbitrary networks measuring the similarity between pairs of dynamical network processes seeded at nearby nodes. We show that the evolution of the curvature distribution exhibits gaps at characteristic timescales indicating bottleneck-edges that limit information spreading. Importantly, curvature gaps robustly encode communities until the phase transition of detectability, where spectral clustering methods fail. We use this insight to derive geometric modularity optimisation and demonstrate it on the European power grid and the *C. elegans* homeobox gene regulatory network finding previously unidentified communities on multiple scales. Our work suggests using network geometry for studying and controlling the structure of and information spreading on networks.

Real-world networks are rarely embedded in physical or Euclidean spaces, which complicates their analysis. However, to correctly represent node similarities, it is typical to assume that the network's nodes lie in a low-dimensional subspace, such as a manifold or linear subspace¹. Having this geometric backbone permits the efficient functioning of standard clustering methods, including ones based on Euclidean geometric features such as k-means or expectation maximisation². A related means of geometrising networks is possible by embedding nodes into a continuous space. For example, the hyperbolic space of constant negative curvature provides a natural parametrisation of complex networks to unveil their self-similar clusters across scales^{3,4}. Likewise, embedding networks into a geometric space based on a suitable distance metric between dynamical network processes has helped reveal their functional organisation^{5,6}. However, in general, there is no guarantee that a network is compatible with a given metric space without suffering significant distortion⁷. At the same time, a network may have several, not necessarily self-similar, geometric representations arising, for example, from clusters at multiple resolutions⁸. Thus, there is a need for a geometric notion that does not require embedding, yet allows studying the multiscale structure of a general class of networks.

A promising candidate is the Ollivier-Ricci (OR) curvature⁹, which measures the change of local connectivity from one node to another, given by the cost of transporting a unit of mass between their respective neighbourhoods. Instead of imposing a geometry on the network through embedding, the OR curvature induces an effective geometry that has precise interpretation in limiting cases. In fact, it is the only one among a number of discrete curvature notions^{10,11} known to converge rigorously to the traditional Ricci curvature of a Riemannian manifold¹². The OR curvature is also related to graph theoretical objects, including the local clustering coefficient and bounds on the spectrum of the graph Laplacian^{13,14}, and has

*corresponding author: adam.gosztolai@epfl.ch

[†]Current address: Blue Brain Project, École polytechnique fédérale de Lausanne (EPFL), Campus Biotech, 1202 Geneva, Switzerland

41 lead to advances in applications such as studying the robustness of economic networks¹⁵, characterising
 42 the human brain structural connectivity¹⁶ and designing clustering heuristics^{17,18}.

43 However, several properties of the OR curvature hinder its widespread applicability to study network
 44 clusters. Since it depends on structural neighbourhoods, it lacks a resolution parameter to define a geometry
 45 on different resolutions, that is necessary to study the multiscale structure in real-world networks. Further,
 46 the OR curvature of an edge is a local quantity, in a sense that it is controlled by the degree of its
 47 endpoints¹³. Thus, it may provide a suboptimal geometric representation of sparse networks - including
 48 many real-world networks where each node connects only to a few others - in which node degrees vary
 49 widely. This lack of robustness of the OR curvature for sparse networks also precludes its use for studying
 50 from a geometric perspective the phase transition occurring as the community structure gets weaker and
 51 become abruptly undetectable^{19–21}. In other words, there is a need for a geometric notion that does not rely
 52 on embeddings, robustly captures multiscale clusters in real networks, and captures the phase transition
 53 at the limit of cluster detection.

54 Results

55 Dynamical Ollivier-Ricci curvature from graph diffusion

56 We address this need by combining two distinct frameworks – network-driven dynamical processes and
 57 geometry with OR curvature. The spreading of network-driven dynamical processes is shaped by the
 58 heterogeneity of the network. In turn, one may infer the network structure by observing properties of their
 59 evolution. We focus on Markov diffusion processes^{8,22–24}, a class of linear dynamical systems which is rich
 60 enough to capture several properties of nonlinear processes on networks^{25,26}. On a connected network G
 61 weighted by pairwise distances w_{ij} , the continuous time diffusion is constructed by the standard procedure²⁷
 62 of defining the normalised graph Laplacian matrix $\mathbf{L} := \mathbf{K}^{-1}(\mathbf{K} - \mathbf{A})$, where \mathbf{K} is the diagonal matrix of
 63 node degrees with $K_{ii} = \sum_j A_{ij}$ and \mathbf{A} is the weighted adjacency matrix encoding similarities between
 64 nodes. For example, one may simply take $A_{ij} = \max_{ij} w_{ij} - w_{ij}$, or $A_{ij} = e^{-w_{ij}}$. Then, the probability
 65 measure of the diffusion started from the unit mass δ_i on node i (Fig. 1a, b) evolves according to

$$\mathbf{p}_i(\tau) = \delta_i e^{-\tau \mathbf{L}}. \quad (1)$$

66 In analogy to the Ricci curvature on a manifold, the classical OR curvature^{9,28} measures the change of
 67 one-step neighbourhoods between nodes (see Methods and Supplementary Fig. 1a for background). Here,
 68 instead of structural neighbourhoods we consider distributions generated by diffusion processes across
 69 scales τ . Specifically, we start a diffusion process at each node $i = 1, \dots, n$ to obtain a set of measures
 70 $\mathbf{p}_i(\tau)$. We then define the *dynamic* Ollivier-Ricci curvature of an edge as the distance of adjacent pairs of
 71 measures relative to the distance of their starting points

$$\kappa_{ij}(\tau) := 1 - \frac{\mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau))}{w_{ij}}, \quad (2)$$

72 whenever ij is an edge and 0 otherwise. Intuitively, Eq. (2) measures how much 'closer' diffusions get over
 73 time when started w_{ij} distance apart, measured by \mathcal{W}_1 , the optimal transport distance²⁹. It is obtained as
 74 a solution to a minimisation problem (Eq. (13) in Methods) and encodes the least cost of transporting the
 75 measure $\mathbf{p}_i(\tau)$ to $\mathbf{p}_j(\tau)$ via the edges on the graph. The minimiser of this problem is the optimal transport
 76 plan represented as a matrix $\zeta(\tau)$. The entries of this matrix shown on Fig. 1c, d quantify how much mass
 77 is moved between each pair of nodes u and v along their connecting geodesic of length d_{uv} (Fig. 1e).

78 As expected, our definition recovers the classical OR curvature⁹ as a first-order approximation for
 79 small times $\tau \ll 1$. Indeed, $\mathbf{p}_i(\tau) \simeq \delta_i \mathbf{K}^{-1} \mathbf{A}$ is the one-step measure encoding the local connectivity.
 80 Further, the dynamical OR curvature inherits the geometric intuition of the classical definition. Notably,
 81 on canonical trees-like and cliques-like networks the $\kappa_{ij}(\tau)$ is negative and positive, respectively, for all
 82 finite scales τ analogously to the Ricci curvature on hyperboloids and spheres (Supplementary Fig. 1b, c).

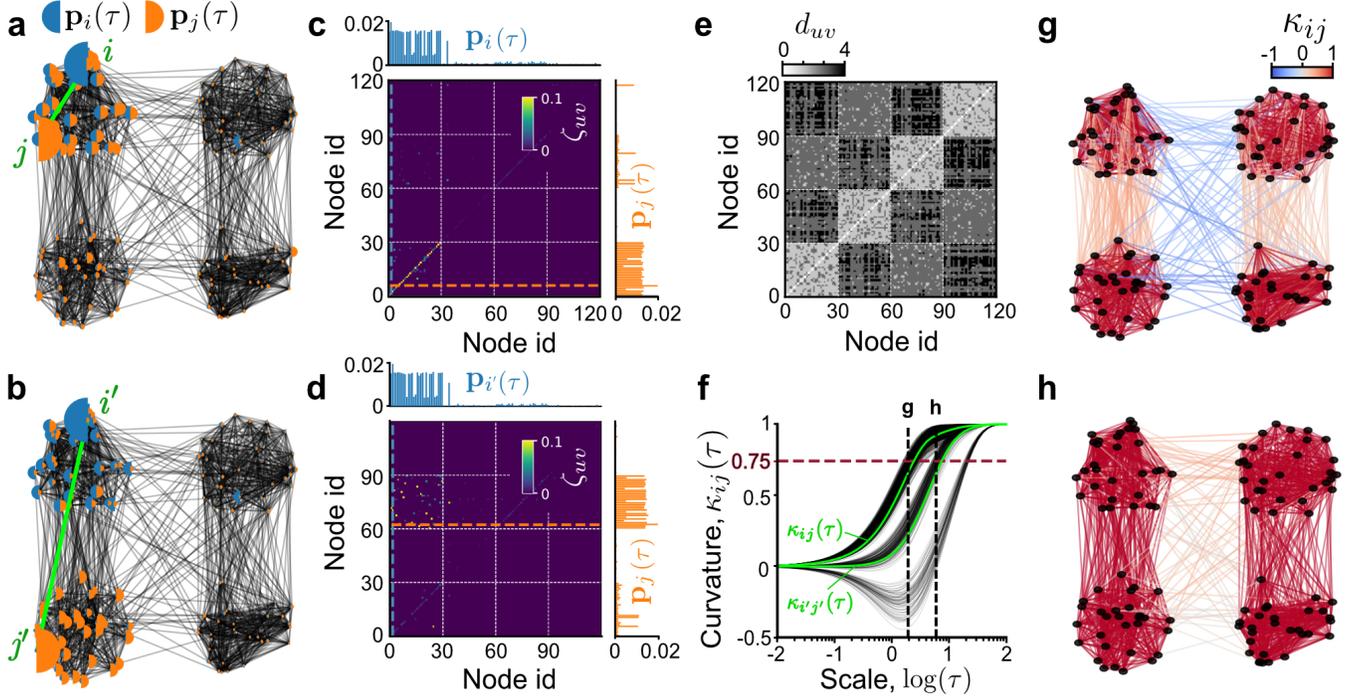


Figure 1: Dynamical Ollivier-Ricci curvature capturing the spreading of diffusion processes. **a** Snapshot at time $\log \tau = 0.15$ of a pair of diffusion measures $\mathbf{p}_i(\tau)$ and $\mathbf{p}_j(\tau)$ started at nodes i, j of a stochastic block model network ($n = 120$ with four equal clusters with edge probabilities 0.8 within clusters and 0.1 or 0.02 between clusters). When i and j are in the same cluster, the measures overlap significantly. The size of half-circles is proportional to the amount of mass on the respective nodes. **b** For i', j' in different clusters the measures remain largely disjoint. **c** Optimal transport plan $\zeta(\tau)$ superimposed with $\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)$. When i, j (colored dashed lines) lie in the same cluster only diagonal elements ζ_{uu} are positive, meaning only geodesics within a cluster transport significant mass. The white dashed lines correspond to the four clusters. **d** Same as **c**, but with diffusions started at nodes i' and j' in different clusters. Only entries ζ_{uv} with u and v in different clusters have significant nonzero weight. **e** Geodesic distance matrix showing the block structure of the network. **f** The evolution of the edge curvatures (Eq. (2)) against time, with the highlighted lines corresponding to edges in **a, b**. Here $\kappa_{ij}(\tau) \simeq 0.75$ indicates scales when local mixing occurs between diffusion pairs. The dashed vertical lines show two such scales ($\log \tau = 0.15, 0.43$). **g, h** Graph edges coloured by the curvature reveals the clusters at the two scales.

83 In the following we are interested in studying the curvature distribution across edges when the network
 84 structure deviates from these canonical topologies.

85 Edge curvature gap differences in rate of information spreading

86 Most real-world networks exhibit organisation on several scales. As an illustration, the unweighted stochastic
 87 block model (SBM) network³⁰ of four equal-size clusters contains two nontrivial scales if the edges are
 88 drawn independently with probability 0.8 within clusters and 0.1 or 0.02 between clusters (Fig. 1a, b). We
 89 show that multiscale structure can be revealed by scanning through a finite range of scales τ and studying
 90 snapshots of curvature distribution across edges.

91 The characteristic scales of a network are related to the overlap between pairs of diffusion measures
 92 $\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)$. This overlap depends on the starting points i, j and on network clusters which can confine
 93 diffusions on well-connected regions for long times before reaching the stationary state π ^{8,22,23,31}, with
 94 $\pi_i = \mathbf{K}_{ii} / \sum_i \mathbf{K}_{ii}$. This transient phenomenon is reflected by the structure of the optimal transport matrix
 95 $\zeta(\tau)$. If i, j lie within the same subnetwork, the measures quickly overlap (Fig. 1a) and only diagonal
 96 entries of $\zeta(\tau)$ are positive (Fig. 1c), weighing only short, within-cluster geodesics. By contrast, started
 97 at different subnetworks, the measures remain almost disjoint (Fig. 1b) and $\zeta(\tau)$ is forced to select longer

98 geodesics (Fig. 1d, e), reflected by the large entries in the off-diagonal block.

99 The evolution of the edge curvature $\kappa_{ij}(\tau)$ (Fig. 1f) aggregates the information in $\zeta(\tau)$ into a single
100 number that is related to the rate of mass exchange between subnetworks at a given scale. We see in
101 Fig. 1f that, initially, when all nodes support disjoint point masses and the diffusions have not yet mixed,
102 $\lim_{\tau \rightarrow 0} \kappa_{ij}(\tau) \rightarrow 1 - \mathcal{W}_1(\delta_i, \delta_j)/d_{ij} = 0$. At the other extreme, as the diffusions reach stationary state,
103 $\lim_{\tau \rightarrow \infty} \kappa_{ij}(\tau) \rightarrow 1 - \mathcal{W}_1(\boldsymbol{\pi}, \boldsymbol{\pi})/d_{ij} = 1$. At intermediate scales, the curvature can take values between
104 1 and some finite negative number depending on the graph. We find that, as the curvature of an edge
105 evolves, the scale at which it approaches unity indicates how easy it is to propagate information between
106 the subnetworks. More precisely, in the Methods, we prove that this scale gives an upper bound on the
107 mixing time τ_{ij}^{mix} of the diffusion pair, namely,

$$\begin{aligned} \tau_{ij}^{\text{mix}} &:= \frac{1}{2} \sum_{uv} |\zeta_{uv}(\tau) - \zeta_{uv}(\infty)| \\ &\leq \min\{\tau : \kappa_{ij}(\tau) \geq 0.75\}, \end{aligned} \quad (3)$$

108 where $\zeta(\tau)$ is the optimal transport plan with marginals $\mathbf{p}_i(\tau)$ and $\mathbf{p}_j(\tau)$. Note that $\kappa_{ij}(\tau) \geq 0.75$ does
109 not imply that the corresponding diffusion processes have approached stationary state independently, but
110 only that they exchange negligible mass at that or larger scales.

111 Importantly, a gap in the distribution of curvatures appears when the curvature exceeds 0.75 for some
112 edges while remaining less than 0.75 for others indicating a network bottleneck that limits mass flow. To
113 illustrate this, Fig. 1f shows three groups of edges, those with most positive curvature are found within
114 clusters, while the other two groups of edges are found between pairs of clusters. Figs. 1g, h correspond
115 to two scales on Fig. 1f ($\log \tau = 0.15, 0.43$) where the curvature has exceeded 0.75 for some groups of
116 edges, indicating the diffusions are well mixed within these groups, but not across other edges for which
117 the curvature is less than 0.75. The latter mark bottleneck edges which lie between the expected partitions
118 with 4 and 2 clusters, respectively. This simple example shows the importance of the scale parameter τ
119 in our curvature definition to capture the network structure at multiple scales. Before applying this to
120 real networks, we take a closer look at the curvature gap in the theoretical context of the stochastic block
121 model.

122 Curvature gap is a robust indicator of clusters in stochastic block models

123 Since in our example any pair of diffusions are supported by one (Fig. 1a) or two (Fig. 1b) clusters, we
124 focus on studying the subgraph G induced by two clusters (Fig. 2a). The subgraph G is a realisation of
125 $\mathcal{G} = \text{SSBM}(n/2, p_{\text{in}}, p_{\text{out}})$, the symmetric SBM composed of two planted partitions of equal size. Edges are
126 generated independently with probability p_{in} within-clusters and probability p_{out} between-clusters. We will
127 denote the ground truth with $C_i \in \{1, -1\}$ for each node i and define $\bar{k} = n(p_{\text{in}} + p_{\text{out}})/2$ as the average
128 degree.

129 Classical spectral clustering methods²⁷ perform well for dense graphs (Fig. 2a), where \bar{k} is an increasing
130 function of n . This suppresses fluctuations for large n causing a spectral gap to appear when the eigenvalue
131 λ_c of the Laplacian matrix \mathbf{L} of G separates from bulk eigenvalues arising from randomness²⁷ (Fig. 2c).
132 In this dense regime, λ_c is well approximated by $\langle \lambda_c \rangle_{\mathcal{G}} = 2p_{\text{out}}/(p_{\text{in}} + p_{\text{out}})$, the second eigenvalue of the
133 ensemble averaged Laplacian $\langle \mathbf{L} \rangle_{\mathcal{G}}$ (see Supplementary Note 1). Since λ_c can be identified due to the
134 spectral gap, clustering involves simply labelling nodes by the sign of the entries of the corresponding
135 eigenvector, which also approximates the ensemble average eigenvector $\phi_c(u) = 1/\sqrt{n}$ when $C_u = 1$ and
136 $-1/\sqrt{n}$ when $C_u = -1$. However, for sparse graphs (Fig. 2b), where \bar{k} is constant (independent of n), the
137 spectral gap ceases to exist³² (Fig. 2d). Thus, spectral algorithms relying on identifying λ_c perform no
138 better than chance. To perform clustering in this regime, one needs to go beyond spectral clustering using,
139 for example, the belief propagation method in statistical physics or the related non-backtracking operator
140 whose spectrum is better behaved^{19,21}.

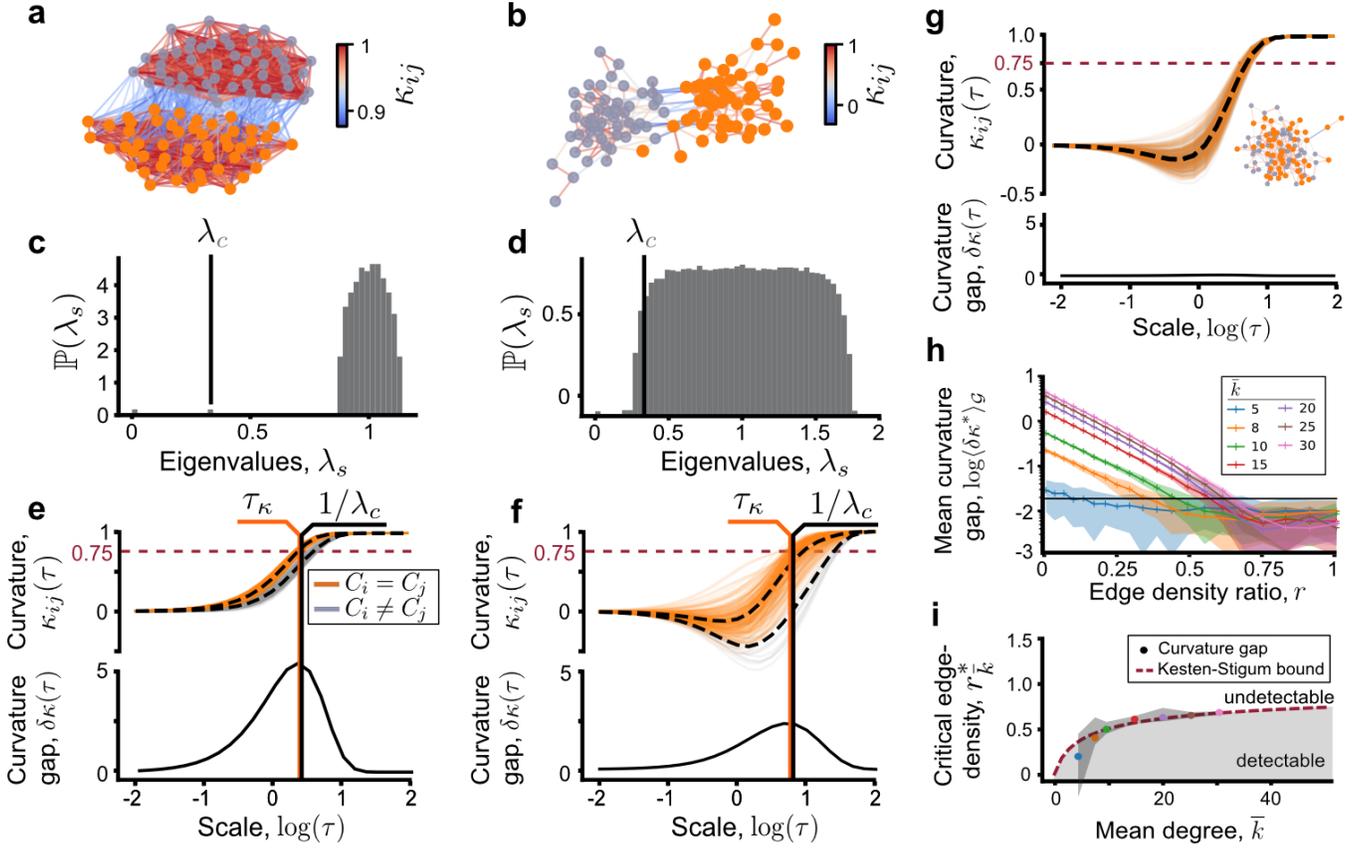


Figure 2: Edge curvature gap indicates the presence of clusters where spectral clustering fails. Two-partition symmetric SBM graph in the **a** dense regime ($p_{in} = 0.5$, $p_{out} = 0.1$) and **b** sparse regime ($p_{in} = 8/n$, $p_{out} = 0.5/n$). Edges are coloured by the curvature ($n = 100$, $\log \tau = 0.83$). **c**, **d** The histogram of eigenvalues obtained from five SBM realisations in dense and sparse regime, respectively. In the dense regime, the eigenvalue λ_c corresponding to the community structure is well separated from the bulk eigenvalues, but overlaps in the sparse regime. **e**, **f** The evolution of edge curvatures driven against diffusion time. A gap between the curvatures of within-edges and between-edges is associated with the presence of clusters. When $\kappa_{ij} > 0.75$ (horizontal dashed line) the diffusions are well mixed across the respective edges. The curvature gap is maximal at $\tau_\kappa \approx \lambda_c^{-1}$ (orange and black vertical lines). **g** There is no curvature gap in the limiting ER graph (inset, $p_{in} = p_{out} = (8 + 0.5)/(2n)$). **h** Maximal curvature gap averaged over 20 SBM realisations for each fixed \bar{k} with 10^4 nodes, against edge density ratio. The horizontal line marks the estimated background noise level. The intersection of this line with the mean curvature gap defines $r_{\bar{k}}^*$, the largest possible edge density ratio to detect clusters. **i** Phase diagram of critical edge density ratio against average degree. The numerically obtained critical edge density ratios computed from the curvature gap are superimposed with the theoretical Kesten-Stigum detection limit (dashed line) and show excellent agreement. Gray shaded area denotes the regime where detection is possible.

141 To see how robustly the dynamical OR curvature indicates the presence of clusters in the symmetric
 142 SBM, let us construct a measure on the curvature evolution. To this end, we define the curvature gap as
 143 the difference between the mean curvatures of within- and between-edges

$$\delta\kappa(\tau) := \frac{1}{\sigma} \left| \langle \kappa_{ij}(\tau) \rangle_{C_i=C_j} - \langle \kappa_{ij}(\tau) \rangle_{C_i \neq C_j} \right| \quad (4)$$

144 where the averages are on within and between-edges, normalised by $\sigma = \sqrt{\frac{1}{2}(\sigma_{\text{within}}^2 + \sigma_{\text{between}}^2)}$ in terms
 145 of the standard deviations of both sets of curvatures. This measure is adapted from the sensitivity index in
 146 signal detection theory, known to be, asymptotically, the most powerful statistical test for discriminating
 147 two distributions³³. Large curvature gap $\delta\kappa(\tau)$ indicates that the within and between edges have curvatures
 148 different enough for the clusters to be recovered (Fig. 2e, f). Correspondingly, in the limits $\tau \rightarrow 0, \infty$

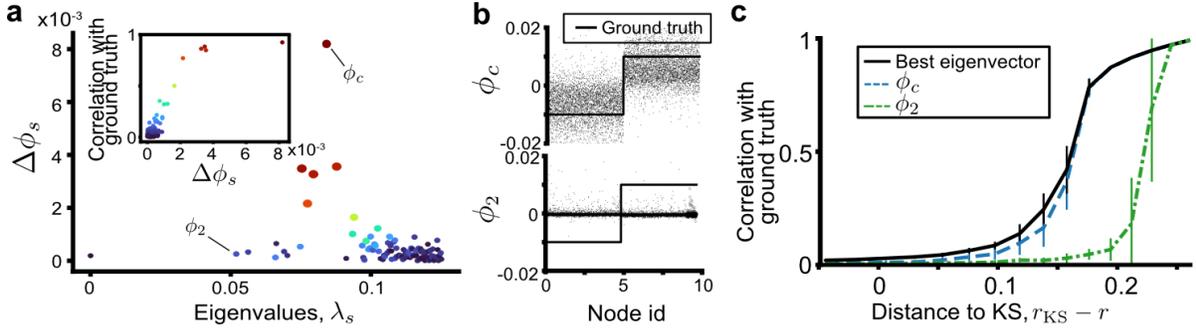


Figure 3: Detecting communities using pairs of diffusions near the weak recovery limit. **a** Difference in eigenvectors $\Delta\phi_s$ (Eq. (8)) between diffusion processes started at adjacent nodes for a single sparse SBM network ($p_{\text{in}} = 3/n$, $p_{\text{out}} = 0.5/n$, $n = 10^4$). Each dot marks $(\lambda_s, \Delta\phi_s)$ for the 50 smallest eigenvectors, colored by the correlation of the corresponding eigenvector with the ground truth, shown in the inset. **b** The eigenvector with the highest $\Delta\phi_s$ encodes the cluster structure (solid line), whereas the second eigenvector ϕ_2 , used by spectral clustering methods, are driven by high random fluctuations. **c** Correlation of eigenvectors with ground truth against distance to KS limit ($n = 10^5$, $\bar{k} = 3$). The eigenvector identified by the highest $\Delta\phi_s$ approaches the correlation with the ground truth of the best eigenvector in the spectrum. All eigenvectors become uncorrelated with the ground truth at the KS limit.

149 where the curvatures are uniform across the graph $\delta\kappa(\tau)$ vanishes and, likewise, in the absence of structure
 150 ($p_{\text{in}} \approx p_{\text{out}}$ in the Erdős-Rényi (ER) limit) we have $\delta\kappa(\tau) = 0$ for all τ (Fig. 2g). At intermediate scales, we
 151 find that the scale of maximal curvature gap occurs at τ_κ at which point the curvatures of within-edges is
 152 $\kappa_{ij}(\tau_\kappa) \approx 0.75$. In agreement with Eq. (3), this indicates well-mixed diffusions across these edges relative
 153 to low-curvature bottleneck edges between clusters, which indicate incomplete mixing. We also find that
 154 $\tau_\kappa \approx \lambda_c^{-1}$ (Fig. 2e, f). These results show that positive curvature gap is associated with the presence of
 155 clusters.

156 What is the minimum curvature gap needed to detect clusters? Previous works on the limits of cluster
 157 detection has shown that if the clusters are too weak (high $r := p_{\text{out}}/p_{\text{in}}$) or the graph too sparse (low
 158 \bar{k}), no algorithm can assign the vertices to communities better than chance, or distinguish G from an
 159 Erdős-Rényi graph ($r = 1$). This is known as the limit of weak-recovery or detection and is characterised
 160 by the Kesten-Stigum (KS) threshold $r = r_{\text{KS}} = (\bar{k} - \sqrt{\bar{k}})/(\bar{k} + \sqrt{\bar{k}})$ ^{19,20,34}.

161 To study this limit, we sampled 20 networks from \mathcal{G} for a range of \bar{k} and r . For each sample, we computed
 162 the maximal curvature gap $\delta\kappa^* := \max_\tau \delta\kappa(\tau)$ and formed the ensemble average quantity $\langle \delta\kappa^* \rangle_{\mathcal{G}}$. As r
 163 increases for a given \bar{k} we observe that $\langle \delta\kappa^* \rangle_{\mathcal{G}}$ decreases exponentially until a certain noise level (Fig.
 164 2h). The critical edge density ratio r_k^* can be estimated as the smallest r where $\langle \delta\kappa^* \rangle_{\mathcal{G}}$ dropped below a
 165 threshold background noise level, estimated here at 0.035 (black horizontal line). This choice of threshold
 166 is not absolute, as it is affected by the finite-size effect of the SBM graphs. An analytical derivation of this
 167 threshold is out of scope of this work, but our numerical experiment clearly shows that the curvature gap
 168 detects a signal from the planted partitions up to the KS limit (Fig. 2i).

169 Geometric cluster detection in the sparse regime

170 Given that the curvature gap (Eq. (4)) indicates the presence of clusters until the fundamental KS limit we
 171 asked if this information could be used to recover the ground truth partition. The definition of curvature gap
 172 (Eq. (4)) suggests looking for equilibrium configurations of the unit-temperature Boltzmann distribution
 173 over the cluster assignments C ,

$$\mathbb{P}(C|\boldsymbol{\kappa}) \propto e^{\sum_{ij} \kappa_{ij}(\tau)\delta(C_i, C_j)}, \quad (5)$$

174 where $\boldsymbol{\kappa}$ is a matrix with entries κ_{ij} and the sum is over all edges ij . The distribution involves only
 175 within-edges because finding those is equivalent to finding between-edges, up to a normalisation factor.

176 The distribution $\mathbb{P}(C|\kappa)$ is important because all of its equilibrium states are equivalent and correlate
 177 with the ground truth partition of the symmetric SBM \mathcal{G} . To see this, we connect $\mathbb{P}(C|\kappa)$ to the posterior
 178 distribution $\mathbb{P}(C|G)$ of the cluster assignments obtained given the graph drawn from \mathcal{G} . In the sparse
 179 regime, the likelihood of observing G with a given cluster assignment C is

$$\mathbb{P}(G|C) \propto \prod_{ij} \left(\frac{p_{\text{in}}}{p_{\text{out}}} \right)^{\delta(C_i, C_j)} \propto \mathbb{P}(C|G) \quad (6)$$

180 (see Eq. (17) in Methods). The second part of Eq. (6) results from Bayes' theorem using a uniform prior
 181 on C , since a priori all configurations are equally likely. It has been previously shown¹⁹ that $P(C|G)$ is
 182 equivalent to the Boltzmann distribution of an Ising model with constant interaction strength

$$\mathbb{P}(C|G) \propto e^{\beta \sum_{ij} \delta(C_i, C_j)} \quad (7)$$

183 with inverse temperature $\beta = \log(p_{\text{in}}/p_{\text{out}}) \approx p_{\text{in}} - p_{\text{out}}$. Note that one of the equilibrium states is trivial
 184 assigning all nodes to one cluster. However, asymptotically ($n \rightarrow \infty$) the probability of this state vanishes
 185 and the Boltzmann distribution is uniform over all other configurations with group sizes $n/2$ and $p_{\text{out}}n/2$
 186 between-edges¹⁹. The fact that one of these states is the ground truth partition, and all equilibrium states
 187 of Eq. (7) are equivalent up to a permutation of nodes within clusters means they are indistinguishable
 188 from the ground truth partition.

189 Due to the equivalence between Eq. (6) and (7), to prove the equivalence between Eq. (5) and (6) we
 190 show that Eq. (5) can also be reduced to Eq. (7). The main insight is that the dynamical OR curvature
 191 (Eq. (2)) is constructed using pairs of diffusions, as opposed to single diffusions. Thus, eigenmodes
 192 arising from random fluctuations are reflected equally in the spectrum of both diffusions and cancel out
 193 upon taking differences over all adjacent node pairs. This allows recovering the community eigenvector
 194 ϕ_c even in the sparse regime where when there is no spectral gap and λ_c is no longer identifiable from
 195 the spectrum (Fig. 2d). Specifically, using pairs of diffusions, we use the spectral expansion to write
 196 $\sum_{ij} (p_i^u(\tau) - p_j^u(\tau)) = \sum_s e^{\lambda_s \tau} \phi_s \Delta \phi_s$ where

$$\Delta \phi_s := \sum_{ij} (\phi_s(i) - \phi_s(j)) . \quad (8)$$

197 We find that, on a single SBM realisation, $\Delta \phi_s$ is large for only a few eigenvectors ϕ_s and diminishing for
 198 others Fig. 3a). Importantly only those eigenvectors with large $\Delta \phi_s$ correlate strongly with the ground
 199 truth (Fig. 3a inset). As seen in Fig. 3b, the best eigenvector is not ϕ_2 , i.e., the one whose eigenvalue
 200 is second in the spectrum and is used by spectral clustering methods, but the one whose eigenvalue is
 201 inside the bulk in Fig. 2d and thus cannot be identified by looking at the spectrum alone. The correlation
 202 with the ground truth for ϕ_c with the highest $\Delta \phi_s$ averaged over 50 SBM realisations remains close to the
 203 highest achievable among all eigenvectors as the KS bound is approached. Meanwhile, ϕ_2 is suboptimal
 204 (Fig. 3c). We also found that, close to the KS bound, often a few other eigenvectors with similarly high
 205 $\Delta \phi_s$ appear, suggesting an improved clustering method combining several top eigenvectors, but this is out
 206 of scope here.

207 To express the curvature in the exponent of Eq. (5) we use the dual formulation of the optimal
 208 transport distance (Eq. (14) in Methods). The fact that $\Delta \phi_c$ dominates the contribution from other
 209 eigenvectors, allows us to approximate $\sum_{ij} (p_i^u(\tau) - p_j^u(\tau)) = e^{\lambda_c \tau} \phi_c \Delta \phi_c + \epsilon_\phi \propto e^{\lambda_c \tau} \phi_c + \epsilon_\phi$, where ϵ_ϕ is an
 210 asymptotically small term. We use this expression, together with the duality formula (Eq. (14)) to express
 211 Eq. (5). Finally, in the sparse regime, we may make a tree-like approximation of the neighbourhoods of i
 212 and j to find that Eq. (5) reduces to

$$\mathbb{P}(C|\kappa) \propto e^{|\kappa| \sum_{ij} \delta(C_i, C_j)} . \quad (9)$$

213 We refer the reader to the Methods for details. Eq. (9) is the same as Eq. (7) when the communities are
 214 assortative ($p_{\text{in}} > p_{\text{out}}$). We then conclude that the curvatures encode the communities of the symmetric
 215 SBM and allow it to be recovered until close to the Kesten-Stigum bound.

216 In the next section, we present a clustering algorithm based on this insight that can find multiscale
 217 clusters in real-world networks.

218 Geometric modularity for the multiscale clustering of networks

219 To exploit the property of the dynamical OR curvature to give multiple geometric representations, we
 220 develop a multiscale graph clustering algorithm for real-world networks. Using Eq. (5), we introduce the
 221 geometric modularity function

$$Q_{\kappa}(C, \tau) = \frac{1}{2m_{\kappa}} \sum_{ij} (\kappa_{ij}(\tau) - \kappa_0) \delta(C_i, C_j), \quad (10)$$

222 where $2m_{\kappa} = \sum_{ij} |\kappa_{ij}|$ is a normalisation factor and $\kappa_0 = \max_{ij} \kappa_{ij}(\tau_{\text{min}})$ is a constant ensuring that all
 223 edges have small non-positive curvature at the smallest computed scale τ_{min} . Hence optimising Eq. (10) at
 224 small times yields separate communities for each node whereas at large times, when $\kappa_{ij}(\tau) \rightarrow 1$ for all ij ,
 225 all nodes are merged to a single community. At intermediate scales, the curvatures will have negative and
 226 positive values on different edges, making the detection of non-trivial clusters possible without a statistical
 227 null-model. This is in contrast to classical modularity³⁵, which minimises the expected number of edges
 228 between clusters, and requires a statistical null-model (typically the configuration model), which can hinder
 229 identifying functional communities based on dynamics⁵.

230 To detect robust partitions at several scales, we sample the cluster landscape $Q_{\kappa}(C, \tau)$ at a sequence
 231 of scales τ spanning the entire dynamical range of the curvature and, at each τ , optimising Eq. (10) using
 232 the Louvain algorithm^{36,37} with 200 random initialisations. At a given τ , we take the cluster with the
 233 highest geometric modularity and deem it robust if it has a low variation of information VI_{τ} against 50
 234 other randomly chosen clusters at this scale, as well as low variation of information $\text{VI}_{\tau\tau'}$ against the best
 235 cluster assignments at nearby scales τ' . As an example, we show in Fig. 4a the result of this computation
 236 on our four-partition SBM graph with two hard-coded scales. We clearly see two large plateaus with low
 237 VI_{τ} and $\text{VI}_{\tau\tau'}$, corresponding to robust clusters, shown in Fig. 4b,c. At the smallest scales we find no
 238 robust communities shown by the sharp increase in the number of communities and the large VI_{τ} .

239 Due to the link between high edge curvature and well-mixed state (Eq. (3)), we expected that at robust
 240 scales the clusters will correspond to those regions which have a high amount of redundant information, and
 241 thus can be disconnected without affecting the dynamics within them. To see this, we applied this clustering
 242 algorithm to the European power grid graph in Fig. 4d,e,f, an unweighted network of major electrical lines,
 243 which has been previously analysed for robustness³⁸, multiscale communities³⁹ and centrality²⁴. The
 244 multiscale community structure can be clearly seen with the many minima of the VI_{τ} function in Fig. 4d.
 245 We displayed two scales in Fig. 4e,f which unfold parts of the power grid which have been historically
 246 independently developed. The smaller scale (at around $\log \tau = -0.95$) marks countries or economical
 247 and historical alliances (Skandinavia, Benelux, Czechoslovakia, Balkans, etc.). Likewise, the larger scale
 248 (at around $\log \tau = -0.5$) divides historical Eastern-Western Europe. Interestingly with the boundary
 249 in Germany runs along the iron curtain, which also demarcates the regions between major electricity
 250 companies.

251 Finally, we analysed a recent dataset of homeobox gene expression in single neurons of *C. elegans* in
 252 Fig. 4g, h, i and j⁴⁰. This work found based on a multivariate linear regression that the homeobox gene
 253 expression profile in a given anatomical neuron class can explain on average 74% of the expression level
 254 of the remaining genes in that neuron class. We therefore asked whether the homeobox gene expression
 255 profile has sufficient information to cluster neurons into their known anatomical classes.

256 The data contains a binary feature vector for each of the 301 neurons, indicating the presence of a
 257 protein expressed by any of the 105 homeobox genes in the given neuron. To convert this data into a

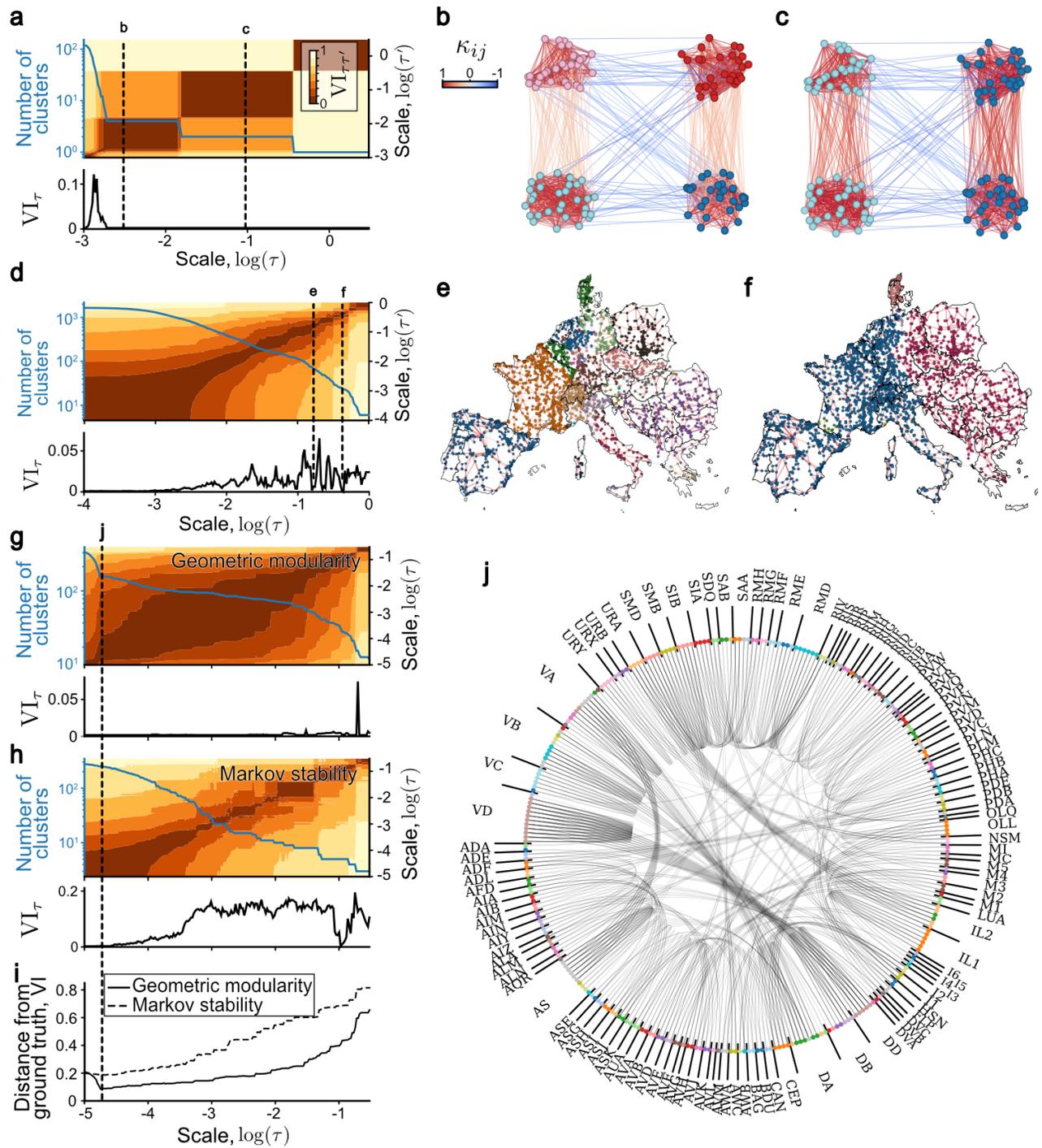


Figure 4: Clustering networks based on multiscale geometric modularity a Clustering statistics computed based on 10^2 Louvain realisations for the multiscale stochastic block model graph. Vertical dashed lines show scales at which stable clusters are detected based on low variation of information at a given scale and persistent low variation of information between Louvain realisations across scales. The communities obtained at these scales are shown on b for $\log \tau = -2.3$ and c for $\log \tau = -1$. Edges are coloured by the curvatures at the respective scales. d Clustering statistics for the European power grid. Two representative stable scales are shown e for $\log \tau = -0.95$ and f for $\log \tau = -0.5$. g Clustering statistics for and the network of *C. elegans* single-neuron homeobox gene expressions show a plateau of stable scales with very similar partitions. h Clustering statistics obtained with Markov stability shows stable scales only at small times with single-node communities, indicating overfitting, and many non-robust partitions at larger scales with high variation of information. i Distance from ground truth based on structural neuronal types or the predicted clusters. Geometric modularity obtained significantly better performance than Markov stability. j Clustering of the *C. elegans* homeobox gene expression data obtained from geometric modularity optimisation superimposed with the ground truth.

258 graph with nodes being neurons, we first eliminated all homeobox genes co-expressed in none or more
259 than 90% of the neurons to retain 67 homeobox genes. We then constructed an all-to-all graph adjacency
260 matrix weighted by the Jaccard similarity index between expression profiles of neurons. To increase the
261 number of edges with negative curvature, thus improve the detection at the smallest scales, we sparsified
262 this network using a geometric sparsification method⁴¹ with parameter $\gamma = 0.01$. This method retains at
263 most a fraction γ edges of the original graph as minimum spanning tree augmented by edges relevant for
264 preserving local or global geometry of the graph.

265 The results of our clustering algorithm on this graph is shown on Fig. 4g and compared with the
266 result of Markov stability⁸ (Fig. 4h), a multiscale method based on persistence of diffusions. Geometric
267 modularity obtains a large range of robust scales with highly similar clusters - as shown by the low VI_τ and
268 $VI_{\tau\tau'}$. These scales correlate closely with the known ground truth of 117 anatomical neuron classes (Fig.
269 4i). In contrast, for Markov stability⁸, the scales with low VI_τ overfit the graph finding too many clusters
270 (Fig. 4h) which correlate less with the ground truth (Fig. 4i). Likewise, hierarchical clustering fails to
271 identify the ground truth communities⁴⁰. On Fig. 4j we superimpose the best clustering from geometric
272 modularity against the ground-truth. We observe little differences, apart from VA and AS nodes as well
273 as VD and DD often clustered together. Careful look reveals close biological relationship between these
274 classes; all four classes correspond to motor neurons, with pairs expressing the same neurotransmitters
275 - VA, AS expressing acetylcholine and VD, DD expressing gamma-aminobutyric acid (GABA). These
276 novel results give direct quantitative support to the claim that homeobox gene expression patterns encode
277 structural neuron types. We also observe other stable partitions at larger scale, but they did not correlate
278 the ground-truth.

279 Overall, these results give a strong demonstration that our method is able to find stable clusters in
280 sparse graphs, and provide meaningful insights into distinct types of real-world networks.

281 Discussion

282 We introduced the concept of dynamical Ollivier-Ricci (OR) curvature which defines an effective geometry
283 from pairs of diffusion processes on the network. Instead of imposing the requirement of a manifold
284 approximation or embedding, used by previous geometric approaches^{4-6,23}, our approach constructs a
285 geometric object - the weighted and signed edge curvature matrix - capturing progressively coarser features
286 as the diffusion processes evolve.

287 Real-world networks often exhibit community structure on multiple scales, based on difference between
288 the rates of information propagation in regions the network on various timescales. We showed that the edge
289 curvature matrix carries a precise meaning in this context and bounds the rate of information flow across
290 edges. Consequentially, curvature gaps, differences between edge curvatures within and between regions,
291 indicate network bottlenecks. This result does not rely on the dynamics being linear diffusions, making
292 it suitable to study the interaction of arbitrary dynamical processes. We expect that, in the future, this
293 approach can be used to tune the geometry of the graph to control the flux or interaction of network-driven
294 dynamical processes, for example, leading to better insights to synchronisation problems or metapopulation
295 models⁴².

296 Although diffusion processes constructed from the graph Laplacian have been explored for network
297 clustering^{8,23}, our work differs in the use of diffusion pairs to construct the curvature. Two diffusions pick
298 up random variations in the graph independently, which can be exploited to average out non-informative
299 fluctuations. On stochastic block models, this feature allows the curvature gap to robustly indicate clusters
300 in the sparse regime down to the fundamental limit, where clustering methods relying on the spectral gap in
301 the Laplacian fail³². We also found a new measure of eigenvalue quality, able to select the best eigenvector
302 to be used in spectral methods. Interestingly, the edge curvatures are defined on the set of shortest paths
303 which cannot contain the same edge twice, a subset of the set of non-backtracking walks. Our results
304 are therefore consistent with previous works on the limits of cluster detection using statistical physics
305 objects including the spectrum of non-backtracking operator²¹ or related message passing approaches¹⁹.
306 We expect this insight to provide a new avenue to study the fundamental limits of efficient clustering from

307 a geometric perspective.

308 Finally, using geometric modularity we built an easy-to-use algorithm to study the multiscale commu-
309 nity structure of real networks. We demonstrated on the European power grid network and a recent dataset
310 of *C. elegans* single-neuron homeobox gene expressions that our method can find robust and interpretable
311 communities on multiple scales on diverse datasets without the tendency of overfitting.

312 Overall, we expect our insights connecting dynamical processes, geometry and network clustering to
313 open new avenues to studying and controlling the structural and dynamical properties of networks.

314 Acknowledgements

315 AG acknowledges support from an HFSP Cross-disciplinary Postdoctoral Fellowship (LT000669/2020-C).
316 We thank Mauricio Barahona for insightful discussions on this topic, Jonas Braun and István Tomon for
317 their helpful comments on the manuscript and Daniel Morales for inspiring us to analyse the *C. elegans*
318 dataset.

319 Author contributions

320 A.G. and A.A. contributed equally to this work.

321 Code availability

322 The code to reproduce the results in our paper and to perform geometric modularity optimisation is
323 available at https://github.com/agosztolai/geometric_clustering.

324 Data availability

325 The raw data supporting the results is available from the authors upon request.

326 Methods

327 Classical Ollivier-Ricci edge curvature

328 To contrast the dynamical Ollivier Ricci curvature in Eq. (2), we recap the definition of the classical formulation⁹, which is
329 a generalisation of the Ricci curvature of manifolds in differential geometry. Briefly, consider two close points x and y on a
330 manifold as well as a vector \mathbf{v} on the tangent plane at x and another tangent vector \mathbf{v}' in the tangent plane at y that is parallel
331 to \mathbf{v} , i.e., obtained by parallel transport along the geodesic connecting x, y (Supplementary Figure 1a). These vectors shift x
332 and y to nearby points x' and y' , which will be at a distance $d_{x'y'} \approx d_{xy}(1 - \|\mathbf{v}\|^2 K_w/2)$, where K_w is the sectional curvature.
333 The Ricci curvature Ric_{xy} between points x, y is then defined as the average sectional curvature and is proportional to

$$Ric_{xy} \propto 1 - \frac{\langle d_{x'y'} \rangle}{d_{xy}}, \quad (11)$$

334 where $\langle \cdot \rangle$ denotes the average over all vectors \mathbf{w}, \mathbf{w}' running over the unit sphere in the tangent planes at x and y . In other
335 words, it measures how much geodesics expand or contract on average around points x, y . On flat planes the geodesics stay
336 equally separated hence $Ric_{xy} = 0$, on spheres the geodesics contract hence $Ric_{xy} > 0$, whereas in hyperbolic spaces they
337 expand, hence $Ric_{xy} < 0$ (Supplementary Figure 1a).

338 The classical Ollivier-Ricci curvature⁹ is defined by direct analogy to this. Consider two adjacent nodes i and j and place
339 weights on their immediate neighbours in proportion to the edge weights, namely, $\mathbf{p}_i = \delta_i \mathbf{K}^{-1} \mathbf{A}$. Then the Ollivier Ricci
340 curvature becomes

$$\kappa_{ij} = 1 - \frac{\mathcal{W}_1(\mathbf{p}_i, \mathbf{p}_j)}{d_{ij}}. \quad (12)$$

341 Contrast this expression to the dynamical Ollivier Ricci curvature in Eq. (2), which considers diffusion measures which weight
342 progressively larger neighbourhoods as τ increases.

343 We remark that Eq. (12) and the dynamical OR curvature in (2) are valid for any two nodes i, j if the denominator is
344 replaced by the weighted geodesic distance between i and j , but for this work, it suffices to consider adjacent nodes. Indeed,
345 for any non-adjacent nodes uv , $\kappa_{uv} \geq \kappa_{u'v'}$, where $u'v'$ is an adjacent pair lying on the geodesic connecting u, v (Proposition
346 19 in Ref.⁹), meaning that local curvatures control global curvatures.

347 Wasserstein distance

348 To measure the distance between a pair of measures $\mathbf{p}_i(\tau)$ and $\mathbf{p}_j(\tau)$ we use the optimal transport distance²⁹ (also known as
349 1-Wasserstein or earth-mover distance), defined as

$$\begin{aligned} \mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)) &= \min_{\zeta} \sum_{uv} d_{uv} \zeta_{uv}, \\ \text{subject to } \sum_v \zeta_{uv} &= p_i^u(\tau), \quad \sum_u \zeta_{uv} = p_j^v(\tau). \end{aligned} \quad (13)$$

350 The constraints in Eq. (13) ensure that the optimal transport plan $\zeta(\tau) \in \mathbb{R}^{n \times n}$ is a coupling of the measures $\mathbf{p}_i(\tau)$, $\mathbf{p}_j(\tau)$,
351 i.e., $\zeta(\tau)$ is a joint distribution that admits $\mathbf{p}_i(\tau)$ and $\mathbf{p}_j(\tau)$ as marginals.

352 An equivalent formulation of this distance can be constructed from the Kantorovich-Rubinstein duality²⁹, given by

$$\mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)) = \sup_f \sum_u f(u) [p_i^u(\tau) - p_j^u(\tau)] \quad (14)$$

353 where the supremum is taken over all 1-Lipschitz functions f on the graph, that is,

$$|f(u) - f(v)| \leq d_{uv} \quad (15)$$

354 for any node pair u, v .

355 Computational complexity

356 The computational complexity of our clustering method is determined by three components: the computation of the diffusion
357 measures (Eq. (1)), the computation of the optimal transport distance (Eq. (13)) and the computation of the clustering.

358 We compute each diffusion measure in Eq. (1) by the scaling and squaring algorithm of Ref.⁴³. To our knowledge, the
359 complexity of this algorithm is not known, but we found it to be better than computing the matrix exponential which runs
360 in time $\mathcal{O}(n^3)$ and then multiplying by the initial condition. The exact computation of Eq. (13) is performed by interior
361 point methods which have a complexity $\mathcal{O}(n^3 \log n)$. However, note that in our work the explicit computation of $\zeta(\tau)$ is not
362 required and, moreover, there is typically a significant overlap between the measures $\mathbf{p}_i, \mathbf{p}_j$ whenever i, j lie in the same
363 highly connected region. Based on these observations we use recent approximate algorithms to compute the transport cost
364 permitting near $\mathcal{O}(n)$ -time computation of the optimal transport distance^{44,45}. This yields a computational complexity of
365 $\mathcal{O}(nm)$ for sparse graphs with worst case $\mathcal{O}(n^3)$. These methods also allow GPU parallelisation, which we recommend using
366 for large ($n \gg 10^3$) and dense graphs.

367 The third component is the complexity of the Louvain algorithm which is linear $\mathcal{O}(n)$ for sparse networks³⁶.

368 Upper bound on the mixing time in terms of curvature

369 Here we prove inequality (3), which gives an upper bound on the mixing time of the coupled diffusions with measures $\mathbf{p}_i(\tau)$,
370 $\mathbf{p}_j(\tau)$ in terms the dynamical OR curvature. The ϵ -mixing time is defined as the smallest τ where the law of the coupled
371 process, the optimal transport plan $\zeta(\tau)$, is within an ϵ radius of the stationary distribution

$$\tau_{ij}(\epsilon) := \min\{\tau : \|\zeta(\tau) - \zeta(\infty)\|_{\text{TV}} \leq \epsilon\}, \quad (16)$$

372 where the notion of "close to stationarity" is quantified by the total variation distance $\|\zeta(\tau) - \zeta(\infty)\|_{\text{TV}} := \frac{1}{2} \sum_{uv} |\zeta_{uv}(\tau) -$
373 $\zeta_{uv}(\infty)|$. Since $\mathbf{p}_i(\tau)$ and $\mathbf{p}_j(\tau)$ are marginals of $\zeta(\tau)$ we have that

$$\begin{aligned} \tau_{ij}(\epsilon) &= \min\{\tau : \|\mathbf{p}_i(\tau) - \boldsymbol{\pi}\|_{\text{TV}} + \|\mathbf{p}_j(\tau) - \boldsymbol{\pi}\|_{\text{TV}} \leq \epsilon\} \\ &= \min\{\tau : \|\mathbf{p}_i(\tau) - \mathbf{p}_j(\tau)\|_{\text{TV}} \leq \epsilon\}, \end{aligned}$$

374 where we used the independence of the diffusion processes. From here, we may follow Ref.⁴⁶ and use the Csiszár-Kullback-
375 Pinsker inequality for the optimal transport distance

$$\|\mathbf{p}_i(\tau) - \mathbf{p}_j(\tau)\|_{\text{TV}} \leq (1/d_0) \mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)),$$

376 where $d_0 = \min_{ij} d_{ij}$ is a global graph constant, which can therefore be absorbed into ϵ . This gives an upper bound

$$\begin{aligned} \tau_{ij}(\epsilon') &\leq \min\{\tau : \mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)) \leq \epsilon'\} \\ &= \min\{\tau : \kappa_{ij}(\tau) \geq 1 - \epsilon'\}, \end{aligned}$$

377 with $\epsilon' = d_0 \epsilon$ which is what we set out to show. Note that choosing any $\epsilon' \in (0, 1/2)$ ensures exponential convergence rate to
378 the stationary measure⁴⁷ and by convention, we take the middle of this range and define $\tau_{ij}^{\text{mix}} := \tau_{ij}^{\text{mix}}(1/4)$ to obtain Eq. (3).

379 Connection between geometric modularity and the symmetric stochastic block model

380 In this section, we prove that the Boltzmann distribution of cluster assignments given the edge curvatures $\mathbb{P}(C|\kappa)$ (Eq. (5))
 381 has equilibrium states which are indistinguishable from the ground truth partition of the SBM. We show this by reducing
 382 $\mathbb{P}(C|\kappa)$ as well as the posterior distribution $\mathbb{P}(C|G)$, to the same constant interaction Ising model (Eq. (7)). In the remainder
 383 of this section we work in the sparse regime, where $p_{\text{in}}, p_{\text{out}} = O(1/n)$.

384 First, we recap the well-known equivalence of the SBM and the Ising model¹⁹. Let E denote the set of edges. The
 385 probability distribution of the symmetric SBM for two clusters can be written as³⁰

$$\begin{aligned} \mathbb{P}(G|C) &= p_{\text{out}}^e (1 - p_{\text{out}})^{\binom{n}{2} - e} \times \\ &\times \prod_{ij \in E} \left(\frac{p_{\text{in}}}{p_{\text{out}}} \right)^{\delta(C_i, C_j)} \prod_{ij \notin E} \left(\frac{1 - p_{\text{in}}}{1 - p_{\text{out}}} \right)^{\delta(C_i, C_j)} \\ &\propto \prod_{ij \in E} \left(\frac{p_{\text{in}}}{p_{\text{out}}} \right)^{\delta(C_i, C_j)} \end{aligned} \quad (17)$$

386 where e is the total number of edges and in the last line we used that the effect of non-edges is weak in the sparse regime.
 387 Therefore, by Bayes' theorem with uniform prior one obtains the posterior distribution $\mathbb{P}(C|G) \propto \mathbb{P}(G|C)$. As a result,
 388 the probability of clusters generated by the SBM is equivalent to the Ising model with uniform interaction with Boltzmann
 389 distribution given by Eq. (7)¹⁹.

390 Second, we reduce the Boltzmann distribution of clusters given the edge curvature to same Ising model in Eq. (7). From
 391 Eq. (5) we have

$$\begin{aligned} \mathbb{P}(C|\kappa) &\propto e^{\sum_{ij} \kappa_{ij}(\tau) \delta(C_i, C_j)} \\ &\propto e^{\sum_{ij} [1 - \mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau))] \delta(C_i, C_j)}, \end{aligned} \quad (18)$$

392 where in the last line we used the definition of the curvature in Eq. (2). Comparing Eq. (18) with Eq. (7) note that
 393 $1 - \mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau))$ is non-constant and has a non-linear dependence on the scale τ . However, it is possible to express it in
 394 terms of $p_{\text{in}}, p_{\text{out}}$ to make the connection to the Ising model. Let us write the diffusion measures in Eq. (1) in terms of the
 395 spectral decomposition of \mathbf{L} as

$$p_i^k(\tau) = \delta_i \sum_{s=1}^n e^{-\lambda_s \tau} \phi_s(k) \phi_s(i) = \sum_{s=1}^n e^{-\lambda_s \tau} \phi_s(k) \phi_s(i). \quad (19)$$

396 At this point let us remark that in the dense regime where $p_{\text{in}}, p_{\text{out}} = O(1)$, the first two eigenmodes (λ_1, ϕ_1) and (λ_c, ϕ_c)
 397 dominate and the second eigenmode contains the anti-symmetric eigenvector $\phi_c(u) = 1/\sqrt{n}$ when $C_u = 1$ and $-1/\sqrt{n}$ when
 398 $C_u = 2$ that is associated with the community structure (Fig. 2c). Thus, one can follow spectral clustering methods²⁷ to
 399 find the sparsest cut between clusters using ϕ_c . In contrast, in the sparse regime, the dominant eigenmodes will be driven by
 400 random fluctuations in the node degrees across the graph⁴⁸, thus spectral clustering algorithms based on \mathbf{L} are suboptimal
 401 (Fig. 2d).

402 However, the coupled diffusion pair allows for cancelling out random fluctuations in their spectrum. To see this, consider
 403 for a between-edge ij the difference

$$\begin{aligned} \sum_{ij \in E} p_i^k(\tau) - p_j^k(\tau) &= \sum_{ij \in E} \sum_{s=1}^n e^{-\lambda_s \tau} \phi_s(k) [\phi_s(i) - \phi_s(j)] \\ &= \sum_{s=1}^n e^{-\lambda_s \tau} \phi_s(k) \Delta \phi_s, \end{aligned} \quad (20)$$

404 where $\Delta \phi_s$ is defined in Eq. (8). The first term involves the constant eigenvector ϕ_1 corresponding to the stationary state.
 405 Therefore, $\phi_1(i) = \phi_1(j)$ for all ij and thus its contributions cancels out when taking differences. Further, for eigenvectors ϕ_s
 406 with $s \neq 1, c$ we have asymptotically ($n \rightarrow \infty$) that

$$\Delta \phi_s \rightarrow 0$$

407 (Fig. 3). As a result, the only contribution we are left with is coming from the anti-symmetric eigenmode (λ_c, ϕ_c) . Thus we
 408 have that

$$\sum_{ij \in E} (p_i^u(\tau) - p_j^u(\tau)) = \begin{cases} \epsilon_\phi, & \text{if } C_i = C_j, \\ e^{-\lambda_c \tau} \phi_c \Delta \phi_c + \epsilon_\phi, & \text{if } C_i \neq C_j, \end{cases} \quad (21)$$

409 where ϵ_ϕ represents the contribution from the random eigenvectors which is negligible in the limit $n \rightarrow \infty$.

410 To compute \mathcal{W}_1 in the exponent of Eq. (18), we use Kantorovich-Rubinstein duality (Eq. (14)). Using Eq. (21) in Eq.
 411 (14) and ignoring asymptotically small terms, we consider the quantity

$$\sum_{ij \in E} \sum_u f(u) [p_i^u(\tau) - p_j^u(\tau)]$$

$$\begin{aligned}
&= e^{-\lambda_c \tau} \sum_u f(u) \phi_c(u) \\
&= \frac{e^{-\lambda_c \tau}}{n} \left[\sum_{u: C_u=1} f(u) - \sum_{u: C_u=2} f(u) \right] \\
&= \frac{e^{-\lambda_c \tau}}{n} \left[\sum_{u: C_u=1} (f(u) - f(i)) - \sum_{u: C_u=2} (f(u) - f(j)) \right. \\
&\quad \left. + \sum_{u: C_u=1} f(i) - \sum_{u: C_u=2} f(j) \right]. \tag{22}
\end{aligned}$$

412 In the sparse regime, we may make a tree-like approximation in the neighbourhood of i . This means that the number of
413 neighbours of i at distance q inside the cluster is $p_{\text{in}}^q (n/2)^q$, ignoring terms of order $O(1/n)$ and beyond. Considering only
414 nodes at unit distance ($q = 1$), we approximate Eq. (22) as

$$\begin{aligned}
&\frac{e^{-\lambda_c \tau}}{n} \left[\sum_{\substack{u: C_u=1 \\ u \sim i}} (f(u) - f(i)) - \sum_{\substack{u: C_u=2 \\ u \sim i}} (f(u) - f(i)) \right. \\
&\quad + \sum_{\substack{u: C_u=1 \\ u \sim j}} (f(u) - f(j)) - \sum_{\substack{u: C_u=2 \\ u \sim j}} (f(u) - f(j)) \\
&\quad \left. + \sum_{\substack{u: C_u=1 \\ u \sim i}} f(i) - \sum_{\substack{u: C_u=2 \\ u \sim i}} f(i) + \sum_{\substack{u: C_u=1 \\ u \sim j}} f(j) - \sum_{\substack{u: C_u=2 \\ u \sim j}} f(j) \right] \\
&= \frac{e^{-\lambda_c \tau}}{n} \left[\sum_{\substack{u: C_u=1 \\ u \sim i}} (f(u) - f(i)) - \sum_{\substack{u: C_u=2 \\ u \sim i}} (f(u) - f(i)) \right. \\
&\quad + \sum_{\substack{u: C_u=1 \\ u \sim j}} (f(u) - f(j)) - \sum_{\substack{u: C_u=2 \\ u \sim j}} (f(u) - f(j)) \\
&\quad \left. + \frac{n}{2} p_{\text{in}} (f(i) - f(j)) - \frac{n}{2} p_{\text{out}} (f(i) - f(j)) \right].
\end{aligned}$$

415 Then, taking the supremum over all 1-Lipschitz functions f , we obtain

$$\begin{aligned}
&\sum_{ij \in E} \mathcal{W}_1(\mathbf{p}_i(\tau), \mathbf{p}_j(\tau)) (1 - \delta(C_i, C_j)) \\
&\quad \approx e^{-\lambda_c \tau} (p_{\text{in}} + p_{\text{out}}) \left(1 + \frac{|p_{\text{in}} - p_{\text{out}}|}{2(p_{\text{in}} + p_{\text{out}})} \right) \tag{23}
\end{aligned}$$

416 Substituting this into Eq. (18) and noting that $p_{\text{in}} + p_{\text{out}}$ is constant we obtain at a fixed τ

$$\mathbb{P}(C|\boldsymbol{\kappa}) \propto \exp \left[\left(\frac{|p_{\text{in}} - p_{\text{out}}|}{2(p_{\text{in}} + p_{\text{out}})} \right) \sum_{ij \in E} \delta(C_i, C_j) \right],$$

417 which up to a constant of proportionality equals the expression in Eq. (9).

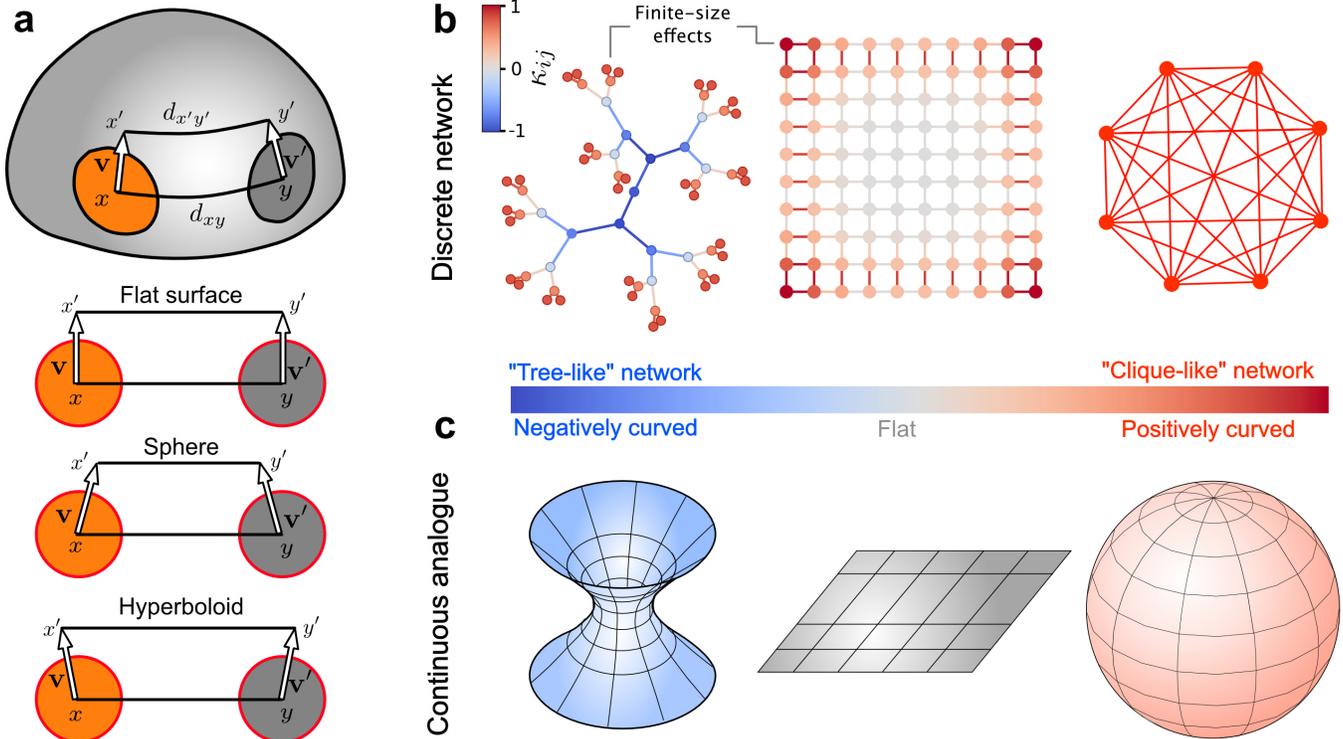
418 References

- 419 [1] Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction.
420 *Science* **290**, 2319–2323 (2000).
- 421 [2] Ding, C., Xiaofeng He, Hongyuan Zha & Simon, H. D. Adaptive dimension reduction for clustering high dimensional
422 data. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 147–154 (2002).
- 423 [3] Serrano, M. A., Krioukov, D. & Boguñá, M. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev.*
424 *Lett.* **100**, 078701 (2008).
- 425 [4] García-Pérez, G., Boguñá, M. & Serrano, M. Á. Multiscale unfolding of real networks by geometric renormalization. *Nat.*
426 *Phys.* 1–9 (2018).
- 427 [5] De Domenico, M. Diffusion geometry unravels the emergence of functional clusters in collective phenomena. *Phys. Rev.*
428 *Lett.* **118**, 168301 (2017).

- 429 [6] Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**,
430 1337–1342 (2013).
- 431 [7] Matousek, J. *Lectures on discrete geometry*, vol. 212 (Springer New York, 2013).
- 432 [8] Delvenne, J.-C., Yaliraki, S. N. & Barahona, M. Stability of graph communities across time scales. *Proc. Natl. Acad.*
433 *Sci. U.S.A.* **107**, 12755–12760 (2010).
- 434 [9] Ollivier, Y. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256**, 810–864 (2009).
- 435 [10] Sturm, K.-T. On the geometry of metric measure spaces. *Acta Math.* **196**, 65–131 (2006).
- 436 [11] Lott, J. & Villani, C. Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.* **169**, 903–991 (2009).
- 437 [12] van der Hoorn, P., Cunningham, W. J., Lippner, G., Trugenberger, C. & Krioukov, D. Ollivier-Ricci curvature convergence
438 in random geometric graphs. *Preprint at <https://arxiv.org/abs/2008.01209>* (2020).
- 439 [13] Jost, J. & Liu, S. Ollivier’s Ricci curvature, local clustering and curvature-dimension inequalities on graphs. *Discrete &*
440 *Comput. Geom.* **51**, 300–322 (2014).
- 441 [14] Bauer, F., Jost, J. & Liu, S. Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator. *Math.*
442 *Res. Lett.* **19**, 1185–1205 (2012).
- 443 [15] Sandhu, R. S., Georgiou, T. T. & Tannenbaum, A. R. Ricci curvature: An economic indicator for market fragility and
444 systemic risk. *Sci. Adv.* **2** (2016).
- 445 [16] Farooq, H., Chen, Y., Georgiou, T. T., Tannenbaum, A. & Lenglet, C. Network curvature as a hallmark of brain structural
446 connectivity. *Nat. Commun.* **10**, 4937 (2019).
- 447 [17] Sia, J., Jonckheere, E. & Bogdan, P. Ollivier-Ricci curvature-based method to community detection in complex networks.
448 *Sci. Rep.* **9**, 9800 (2019).
- 449 [18] Ni, C.-C., Lin, Y.-Y., Luo, F. & Gao, J. Community detection on networks with Ricci flow. *Sci. Rep.* **9**, 9984 (2019).
- 450 [19] Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Asymptotic analysis of the stochastic block model for modular
451 networks and its algorithmic applications. *Phys. Rev. E* **84**, 066106 (2011).
- 452 [20] Abbé, E. & Sandon, C. Community detection in general stochastic block models: Fundamental limits and efficient
453 algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, 670–688 (2015).
- 454 [21] Massoulié, L. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the Forty-Sixth*
455 *Annual ACM Symposium on Theory of Computing*, STOC ’14, 694–703 (Association for Computing Machinery, New
456 York, NY, USA, 2014).
- 457 [22] Gfeller, D. & De Los Rios, P. Spectral coarse graining of complex networks. *Phys. Rev. Lett.* **99**, 038701 (2007).
- 458 [23] Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion
459 maps. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7426–7431 (2005).
- 460 [24] Arnaudon, A., Peach, R. L. & Barahona, M. Scale-dependent measure of network centrality from diffusion dynamics.
461 *Phys. Rev. Res.* **2**, 033104 (2020).
- 462 [25] Schaub, M. T., Billeh, Y. N., Anastassiou, C. A., Koch, C. & Barahona, M. Emergence of slow-switching assemblies in
463 structured neuronal networks. *PLoS Comp Biol* **11**, e1004196–28 (2015).
- 464 [26] Young, H. P. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *Am.*
465 *Econ. Rev.* **99**, 1899–1924 (2009).
- 466 [27] Chung, F. R. K. *Spectral Graph Theory*, vol. 92 (American Mathematical Soc., 1997).
- 467 [28] Veysseire, L. Coarse Ricci curvature for continuous-time Markov processes. *Preprint at <https://arxiv.org/abs/1202.0420>*
468 (2012).
- 469 [29] Villani, C. *Optimal transport: old and new* (Springer, Berlin Heidelberg, 2009).
- 470 [30] Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109–137 (1983).

- 471 [31] Gosztolai, A., Carrillo, J. A. & Barahona, M. Collective search with finite perception: Transient dynamics and search
472 efficiency. *Front. Phys.* **6**, 153 (2019).
- 473 [32] Kawamoto, T. & Kabashima, Y. Limitations in the spectral method for graph partitioning: Detectability threshold and
474 localization of eigenvectors. *Phys. Rev. E* **91**, 062803 (2015).
- 475 [33] Kay, S. M. *Fundamentals of statistical signal processing*. Prentice Hall signal processing series (Prentice Hall PTR, Upper
476 Saddle River, NJ, 1993).
- 477 [34] Mossel, E., Neeman, J. & Sly, A. A proof of the block model threshold conjecture. *Combinatorica* **38**, 665–708 (2018).
- 478 [35] Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577–8582
479 (2006).
- 480 [36] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat.*
481 *Mech. Theory Exp.* **2008**, P10008 (2008).
- 482 [37] PyGenStability: unsupervised clustering with generalised Louvain and Markov Stability. *In preparation* (2021).
- 483 [38] Rosas-Casals, M., Valverde, S. & Solé, R. V. Topological vulnerability of the European power grid under errors and
484 attacks. *Int. J. Bifurcat. Chaos* **17**, 2465–2475 (2007).
- 485 [39] Schaub, M. T., Delvenne, J.-C., Yaliraki, S. N. & Barahona, M. Markov dynamics as a zooming lens for multiscale
486 community detection: Non clique-like communities and the field-of-view limit. *PLOS One* **7**, 1–11 (2012).
- 487 [40] Reilly, M. B., Cros, C., Varol, E., Yemini, E. & Hobert, O. Unique homeobox codes delineate all the neuron classes of *C.*
488 *elegans*. *Nature* **584**, 595–601 (2020).
- 489 [41] Beguerisse-Diaz, M., Vangelov, B. & Barahona, M. Finding role communities in directed networks using role-based
490 similarity, markov stability and the relaxed minimum spanning tree. *2013 IEEE Global Conference on Signal and*
491 *Information Processing* (2013).
- 492 [42] Davis, J. T., Perra, N., Zhang, Q., Moreno, Y. & Vespignani, A. Phase transitions in information spreading on structured
493 populations. *Nat. Phys.* 1–8 (2020).
- 494 [43] Al-Mohy, A. H. & Higham, N. J. A new scaling and squaring algorithm for the matrix exponential. *SIAM J. Matrix*
495 *Anal. Appl.* **31**, 970–989 (2010).
- 496 [44] Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. In Burges, C. J. C., Bottou, L., Welling,
497 M., Ghahramani, Z. & Weinberger, K. Q. (eds.) *Adv. Neural. Inf.. Process. Syst. 26*, 2292–2300 (Curran Associates, Inc.,
498 2013).
- 499 [45] Atasu, K. & Mittelholzer, T. Linear-complexity data-parallel earth mover’s distance approximations. In Chaudhuri, K.
500 & Salakhutdinov, R. (eds.) *Proceedings of the 36th International Conference on Machine Learning*, vol. 97 of *Proceedings*
501 *of Machine Learning Research*, 364–373 (PMLR, Long Beach, California, USA, 2019).
- 502 [46] Paulin, D. Mixing and concentration by Ricci curvature. *J. Func. Anal.* **270**, 1623–1662 (2016).
- 503 [47] Levin, D. A., Peres, Y. & Wilmer, E. L. *Markov chains and mixing times* (American Mathematical Society, 2006).
- 504 [48] Krivelevich, M. & Sudakov, B. The largest eigenvalue of sparse random graphs. *Preprint at*
505 <https://arxiv.org/math/0106066v1> (2001).

Supplementary Figures



Supplementary Figure 1: Ricci and dynamical Ollivier Ricci curvature on canonical surfaces and graph structures. **a** Ricci curvature on a manifold. The geodesic distance of close points x and y on average changes when translated by parallel vectors \mathbf{v} and \mathbf{v}' on the unit circle in the tangent planes at x and y . On planes the points remain equidistant, on spheres the points contract and on hyperbolic surface they expand. **b** Canonical graphs with edges coloured by the dynamical OR curvature (Eq. (2)) for $\tau = 1$ show that positively and negatively curved graphs have qualitatively different topologies. Away from the boundaries, tree-like topologies are negatively curved, grid-like topologies are flat (zero curvature), whereas clique-like topologies attain positive curvature. Nodes are coloured by the average edge curvature across the neighbours. **c** Analogously, the differential geometric notion of Ricci curvature is negative on hyperbolic surfaces, zero on planes and positive on spherical surfaces.

Supplementary Notes

508 Supplementary Note 1

509 In this section, we compute the spectrum of the expected normalised Laplacian matrix of the symmetric
 510 SBM $\mathcal{G}(n, k_{\text{in}}/n, k_{\text{out}}/n)$. Here k_{in} and k_{out} are constants representing the expected number of edges within
 511 and across clusters, respectively. The expected adjacency matrix of the symmetric stochastic block model
 512 is:

$$\langle \mathbf{A} \rangle_{\mathcal{G}} = \begin{pmatrix} \frac{k_{\text{in}}}{n} \mathbf{1}_{n/2 \times n/2} & \frac{k_{\text{out}}}{n} \mathbf{1}_{n/2 \times n/2} \\ \frac{k_{\text{out}}}{n} \mathbf{1}_{n/2 \times n/2} & \frac{k_{\text{in}}}{n} \mathbf{1}_{n/2 \times n/2} \end{pmatrix}. \quad (1)$$

513 Then, the expected normalised Laplacian matrix is given by

$$\langle \mathbf{L} \rangle_{\mathcal{G}} = \mathbf{I} - \begin{pmatrix} \frac{2k_{\text{in}}}{n(k_{\text{in}}+k_{\text{out}})} \mathbf{1}_{n/2 \times n/2} & \frac{2k_{\text{out}}}{n(k_{\text{in}}+k_{\text{out}})} \mathbf{1}_{n/2 \times n/2} \\ \frac{2k_{\text{out}}}{n(k_{\text{in}}+k_{\text{out}})} \mathbf{1}_{n/2 \times n/2} & \frac{2k_{\text{in}}}{n(k_{\text{in}}+k_{\text{out}})} \mathbf{1}_{n/2 \times n/2} \end{pmatrix} \quad (2)$$

514 The first eigenvector is $\phi_1 = \mathbf{1}_n/\sqrt{n} \in \mathbb{R}^n$, with the corresponding eigenvalue being $\lambda_1 = 0$. The second
 515 eigenvector has two values, one on each cluster of the SBM. Taking $\phi_c(u) = 1/\sqrt{n}$ for $1 \leq u \leq n/2$ and
 516 $-1/\sqrt{n}$ for $n/2 < u \leq n$ one has asymptotically

$$\langle \mathbf{L} \rangle_{\mathcal{G}} \phi_c = \left[1 - \frac{k_{\text{in}}}{k_{\text{in}} + k_{\text{out}}} \frac{2}{n} - \frac{k_{\text{in}}}{k_{\text{in}} + k_{\text{out}}} \frac{2}{n} \left(\frac{n}{2} - 1 \right) + \frac{k_{\text{out}}}{k_{\text{in}} + k_{\text{out}}} \frac{2}{n} \frac{n}{2} \right] \phi_c \xrightarrow{n \rightarrow \infty} \frac{2k_{\text{out}}}{k_{\text{in}} + k_{\text{out}}} \phi_c. \quad (3)$$

Figures

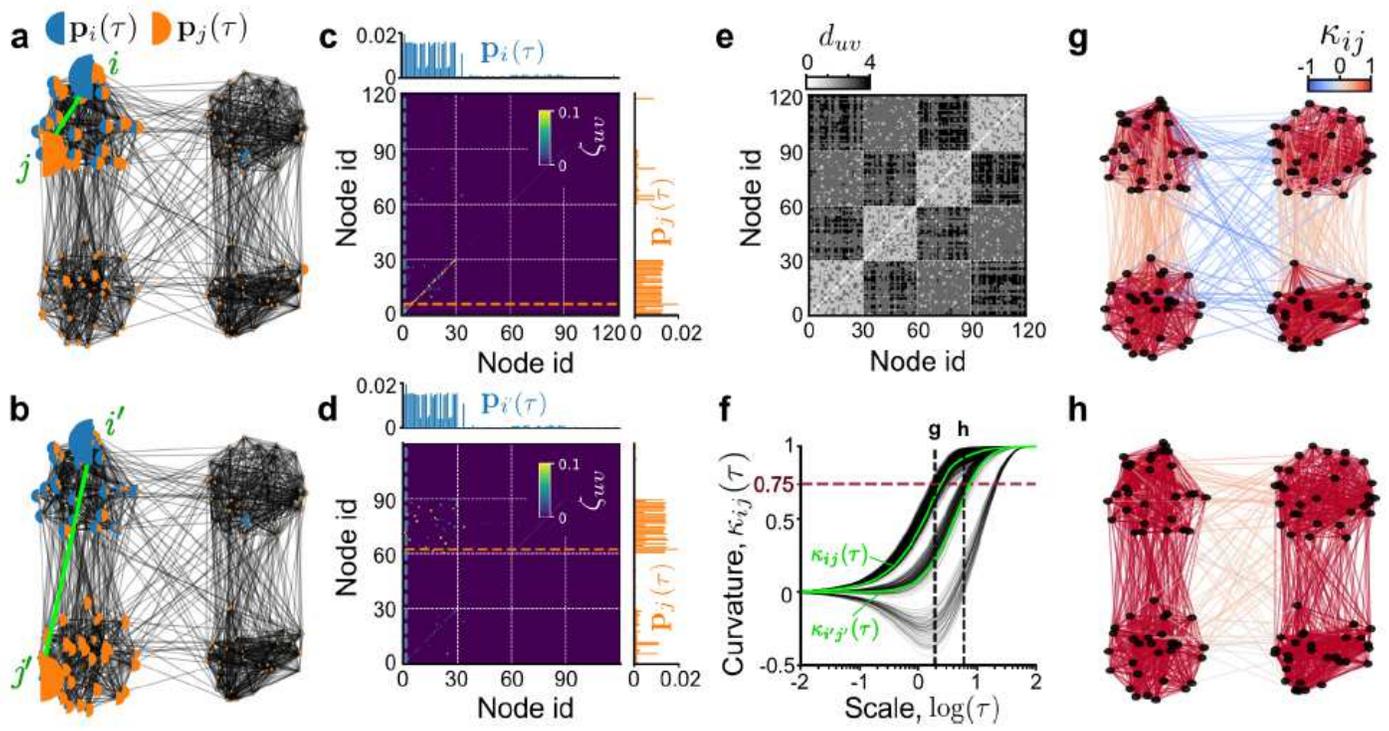


Figure 1

Please see the manuscript file to view figure caption.

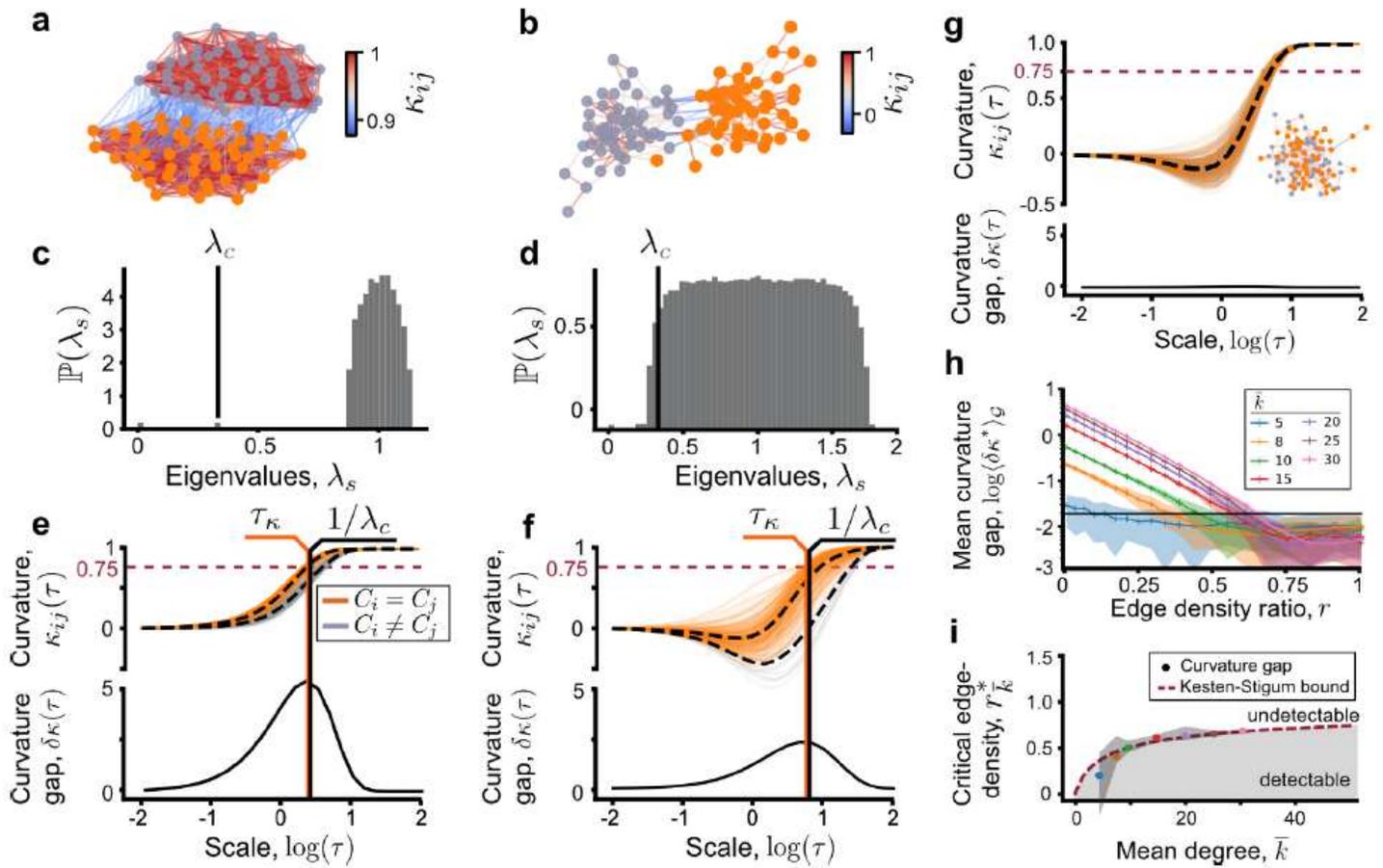


Figure 2

Please see the manuscript file to view figure caption.

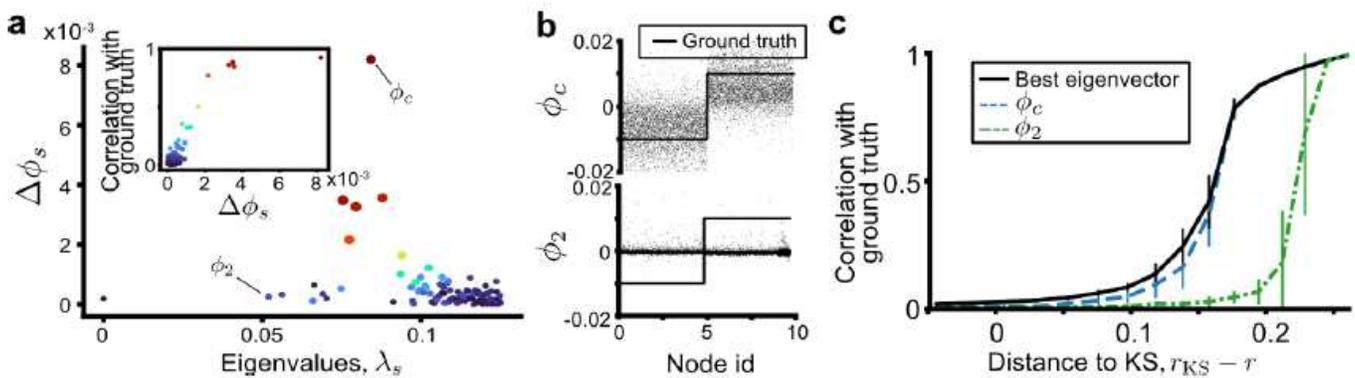


Figure 3

Please see the manuscript file to view figure caption.

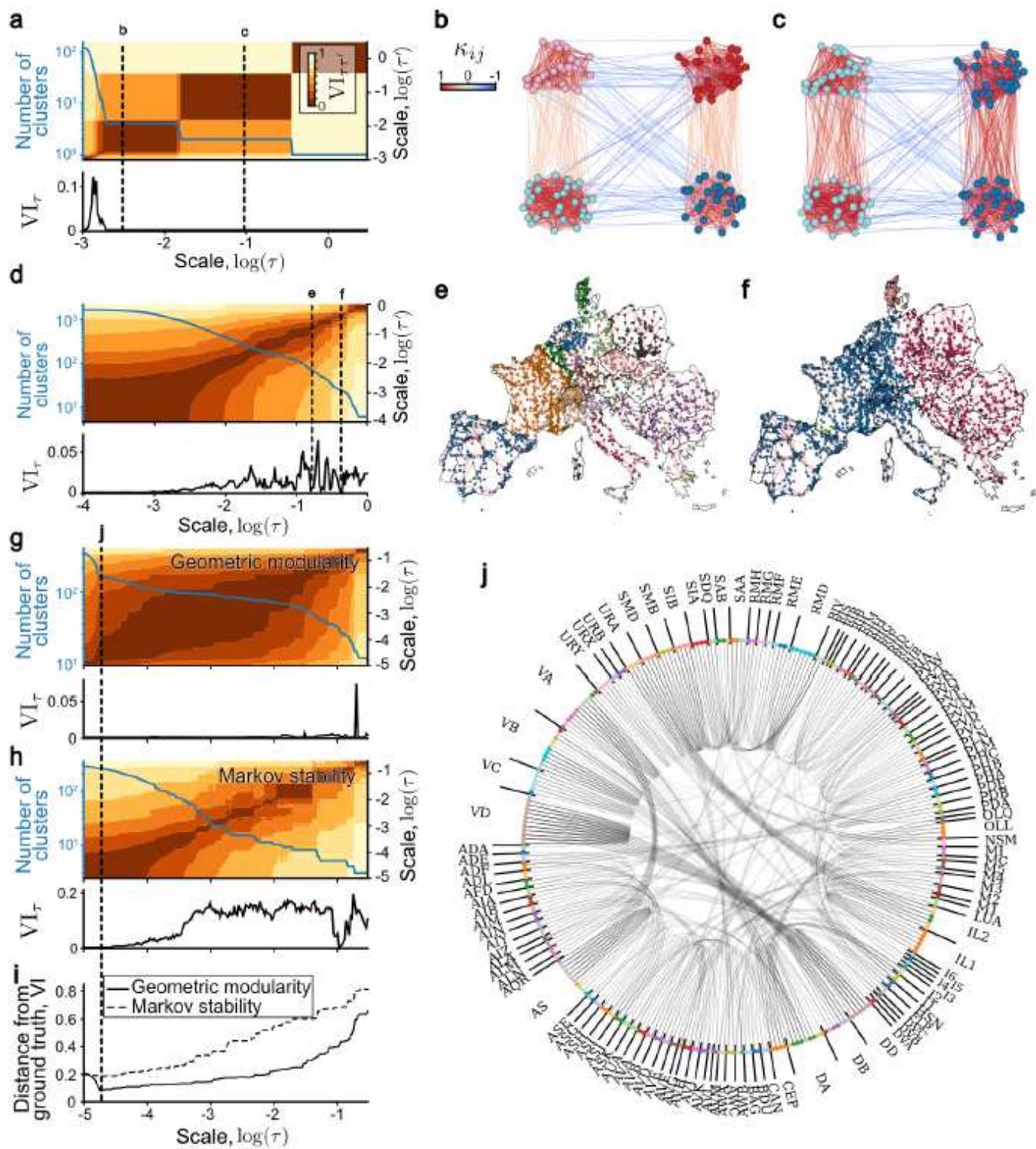


Figure 4

Please see the manuscript file to view figure caption.