Unraveling the shape of networks with diffusion across scales

Alexis Arnaudon
EPFL

Introduction

This is a joint work with: Dr. Peach and Prof. Barahona at Imperial College London.

It has been published in three papers, some years ago:

[1] Peach, R. L., AA, & Barahona, M. (2020). Semi-supervised classification on graphs using explicit diffusion dynamics. *Foundations of Data Science, 2(1)*, 19.

[2] AA, Peach, R. L., & Barahona, M. (2020). Scale-dependent measure of network centrality from diffusion dynamics. *Physical Review Research, 2(3)*, 033104.

[3] Peach, AA, R. L., & Barahona, M. (2022). Relative, local and global dimension in complex networks. *Nature Communications, 13(1)*, 1-11

Today I will show you what we did in this “trilogy” of papers

Recap on diffusion

Continuous diffusion equation is:

$$\frac{\partial p(x,t)}{\partial t} = \sigma \nabla^2 p(x,t)$$

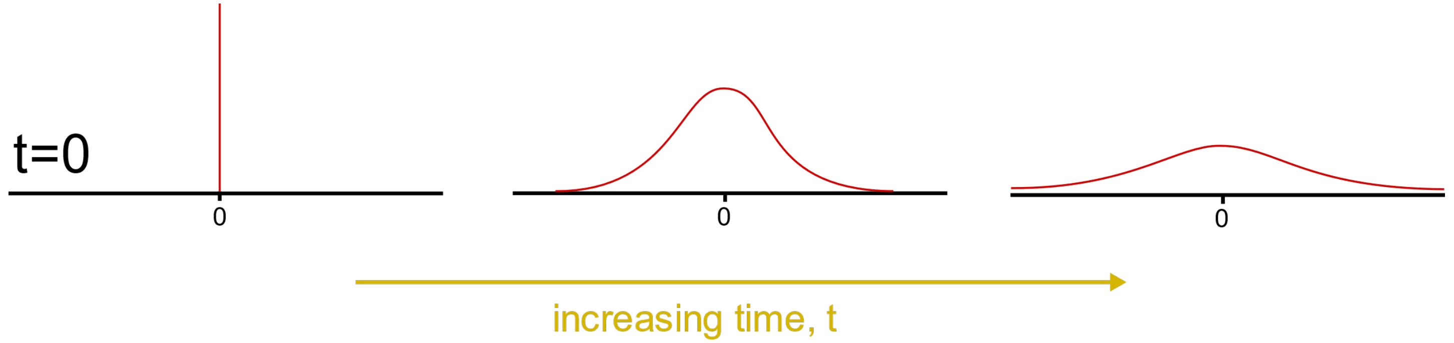
Solution is given by the Green's function, or heat kernel:

$$G_t(x) = \frac{1}{\sqrt{4\pi\sigma t}^d} exp(\frac{x^2}{4\sigma t})$$

Solution for any initial condition is sum of Green functions on non-compact spaces, but tricks can be used on compact spaces (method of mirror, etc…)

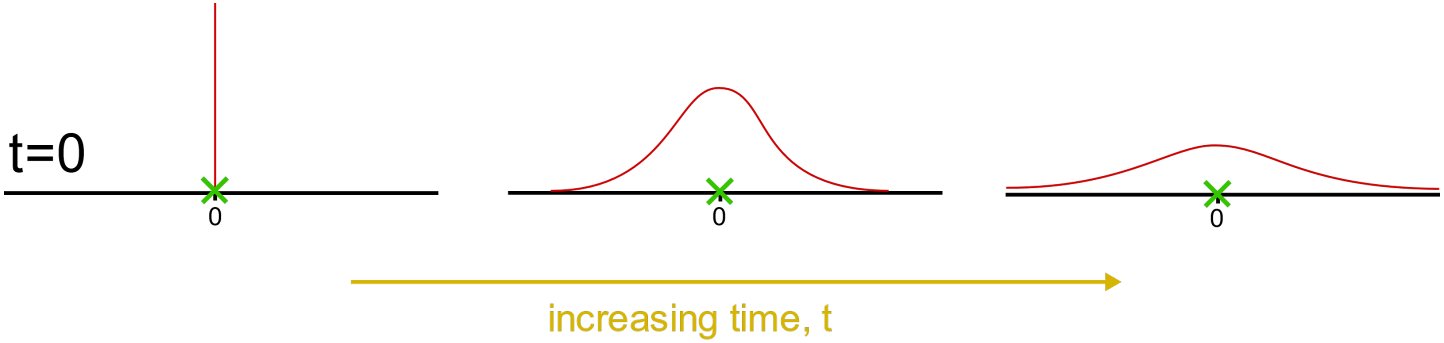
Recap on diffusion

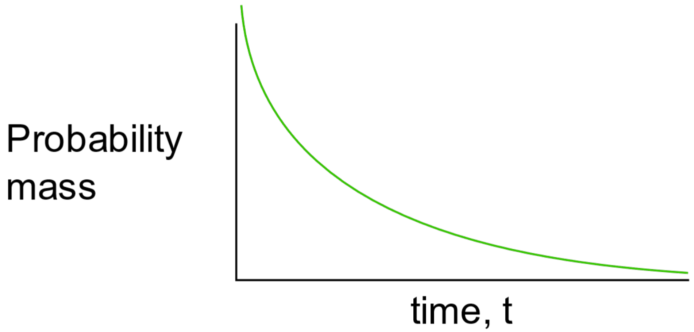
On infinite line, it is simple, starting from a point mass:



Recap on diffusion

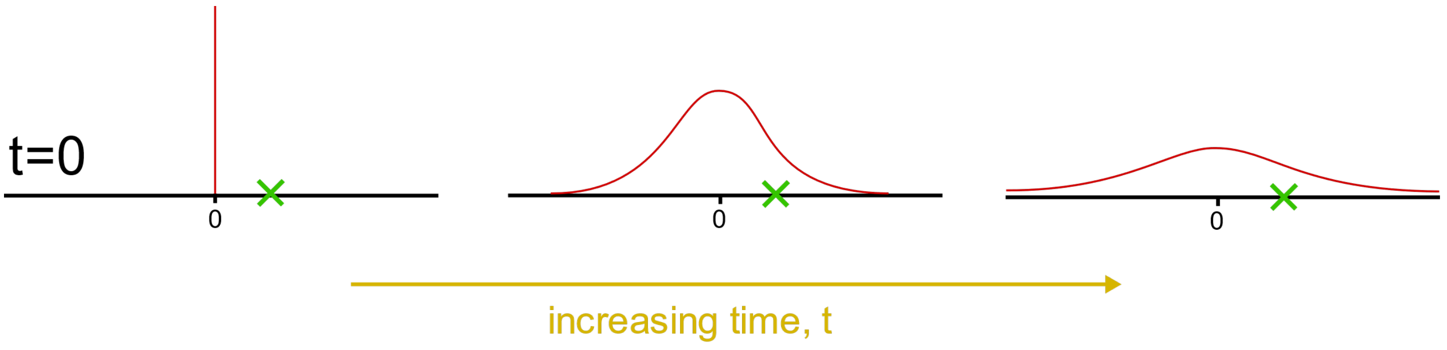
At the origin of the diffusion, the mass monotonically decays:

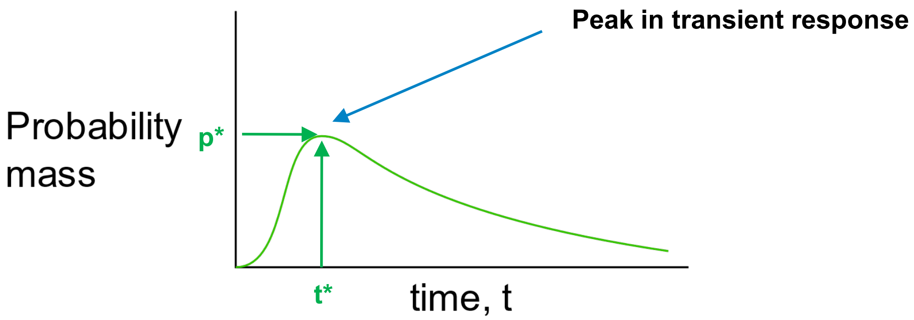




Recap on diffusion

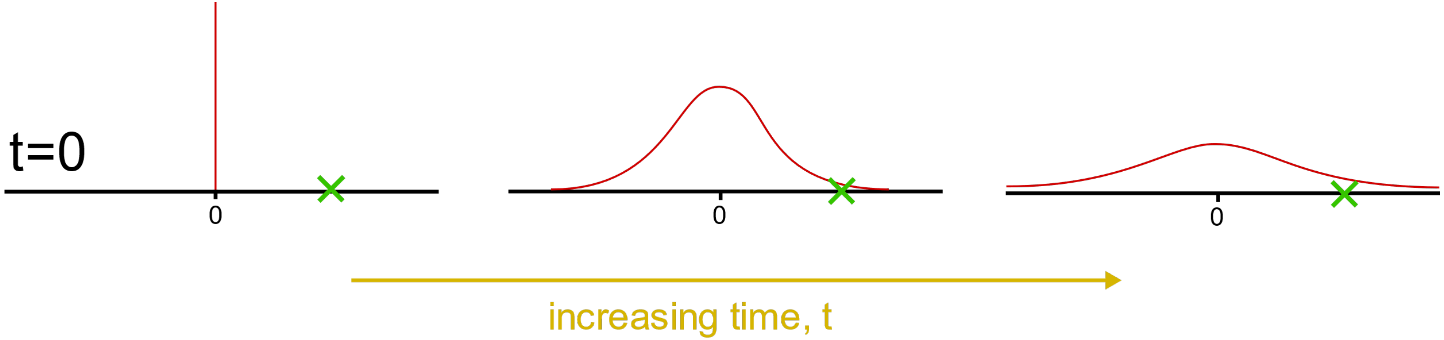
For any other position, it does not, and shows a maximum:

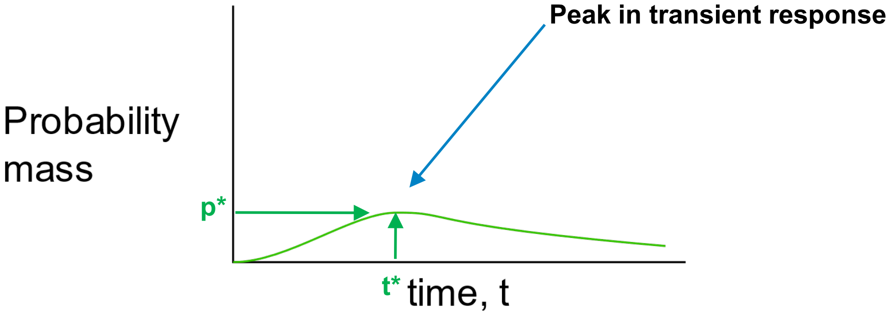




Recap on diffusion

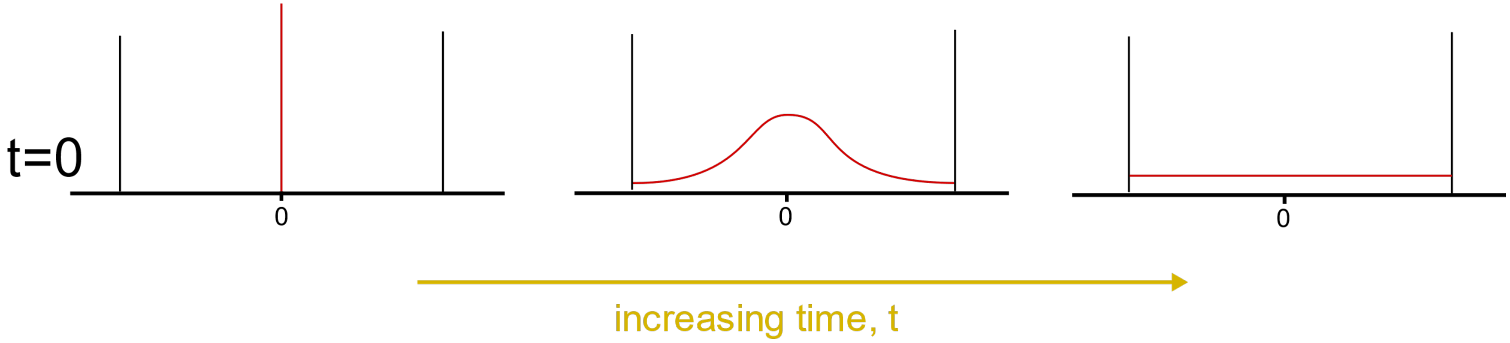
The amplitude and time of the peak depends on the distance to the source, the further away from the source, the later and smaller the peak:





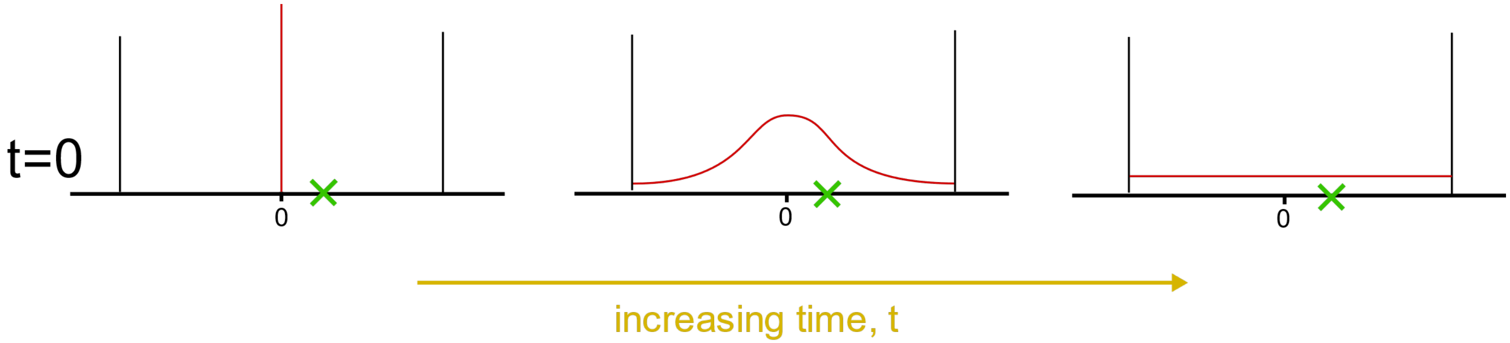
Recap on diffusion

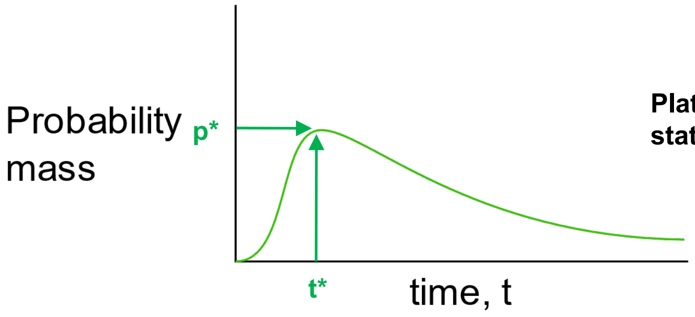
On compact spaces, like interval, diffusion will look different:



Recap on diffusion

Near the source, we have the same behaviour as in infinite line, as the walls are "far":

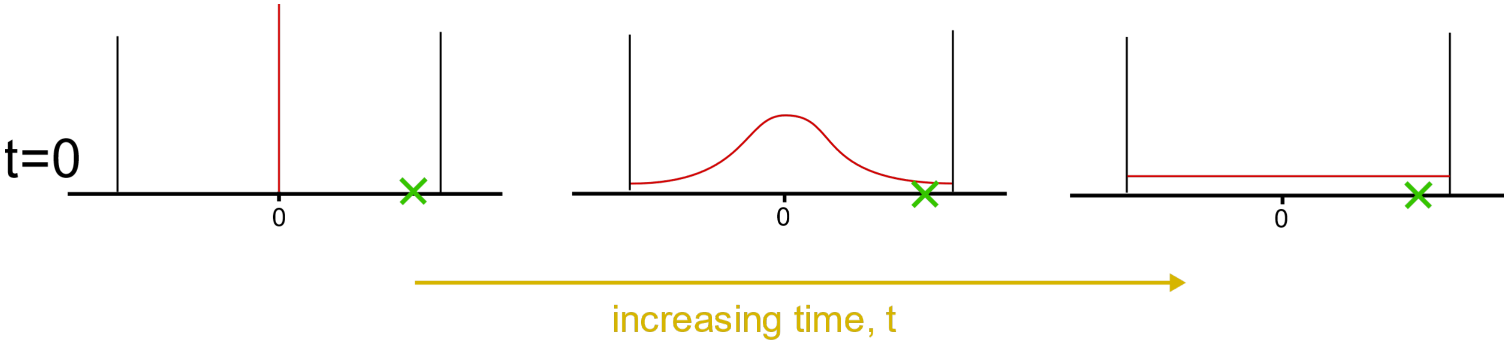


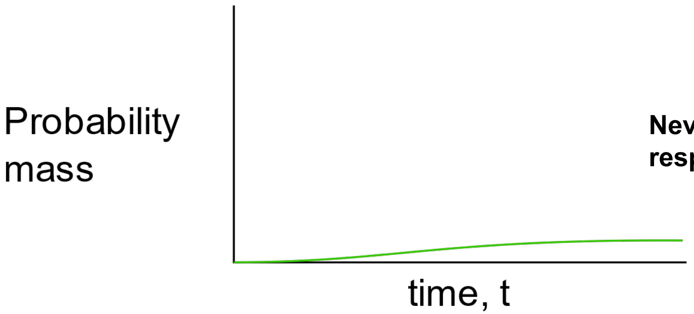


Plateaus at some stationary state as time tends to infinity

Recap on diffusion

But close to the boundary, we instead have monotonically increasing diffusion:





Never exhibits any in transient response

-> This change of behaviour of diffusion encodes information about the boundaries

Recap on graph theory and random walks

- Incidence matrix: $B_{i,(ij)} = 1$ if node I is head of edge (ij)

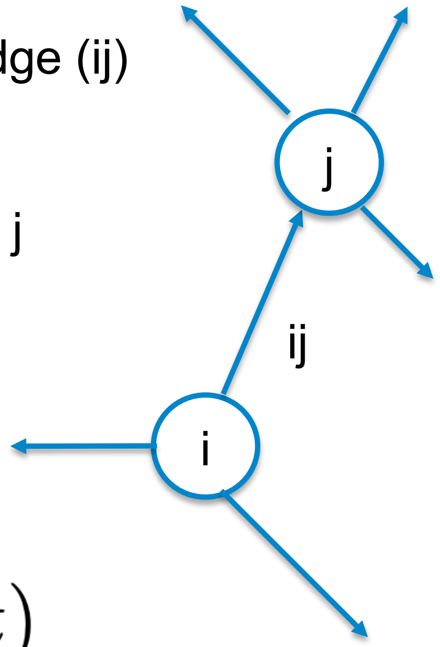
  $B_{j,(ij)} = -1$ if node j is head of edge (ij)

- Adjacency matrix $A_{ij} = 1$ if node i is connected to node j

  We can form discrete random walk: $p_{t+1} = Ap_t$

- Graph laplacian: $L = B^T B = D - A$

  We can form continuous random walk: $\frac{dp(t)}{dt} = Lp(t)$



From continuous diffusion to graph diffusion

We can go from continuous space: $$\frac{\partial p(x,t)}{\partial t} = \sigma \nabla^2 p(x,t)$$

to graphs: $$\frac{\partial \mathbf{p}(t)}{\partial t} = -L\mathbf{p}(t)$$ with graph Laplacian: $$L = D - A$$

For an initial condition at node i, the solution at the node j is the heat kernel:

$$p_j(t \mid i) = (e^{-tL})_{ij}$$

Discretising the previous example as a line graph will give the same results, but this diffusion equation works for any graph.

-> From these ‘diffusion trajectories’ on nodes, we can solve the heat equation and directly estimate peak time and amplitude, if it exists before a given “time horizon”.

How to use this observation to study graphs?

Given any initial condition (or node vector) on a graph and a diffusion-like process, we can compute for each node the time and amplitude of a ’transient peak’. With this information, we have three ‘options’:

1. Use amplitude: Graph Diffusion Reclassification [1]
2. Use time: Multiscale centrality [2]
3. Use time and amplitude: Network dimensionality [3]

[1] Peach, R. L., AA, & Barahona, M. (2020). Semi-supervised classification on graphs using explicit diffusion dynamics. *Foundations of Data Science, 2(1)*, 19.

[2] AA, Peach, R. L., & Barahona, M. (2020). Scale-dependent measure of network centrality from diffusion dynamics. *Physical Review Research, 2(3)*, 033104.

[3] Peach, AA, R. L., & Barahona, M. (2022). Relative, local and global dimension in complex networks. *Nature Communications, 13(1)*, 1-11

1. Using amplitude

Graph Convolutional Neural Network (GCN)

- takes a graph with node features and labels on subset of nodes
- Try to predict missing labels with two layers:

$$\begin{bmatrix} * \\ H_{\mathrm{GCN}} \end{bmatrix} = \mathrm{softmax}\left(\widehat{A}\ \mathrm{ReLU}\left(\widehat{A}\, \mathcal{X}\, W^0\right) W^1\right) ,$$

It works by propagating feature signal through the graph with adjacency matrix and learning weights matrices to predict the labels.

The signal propagation is a one-step random walk, that can be extended with diffusion. Instead, we use diffusion as a post processing step to “correct mistakes”.

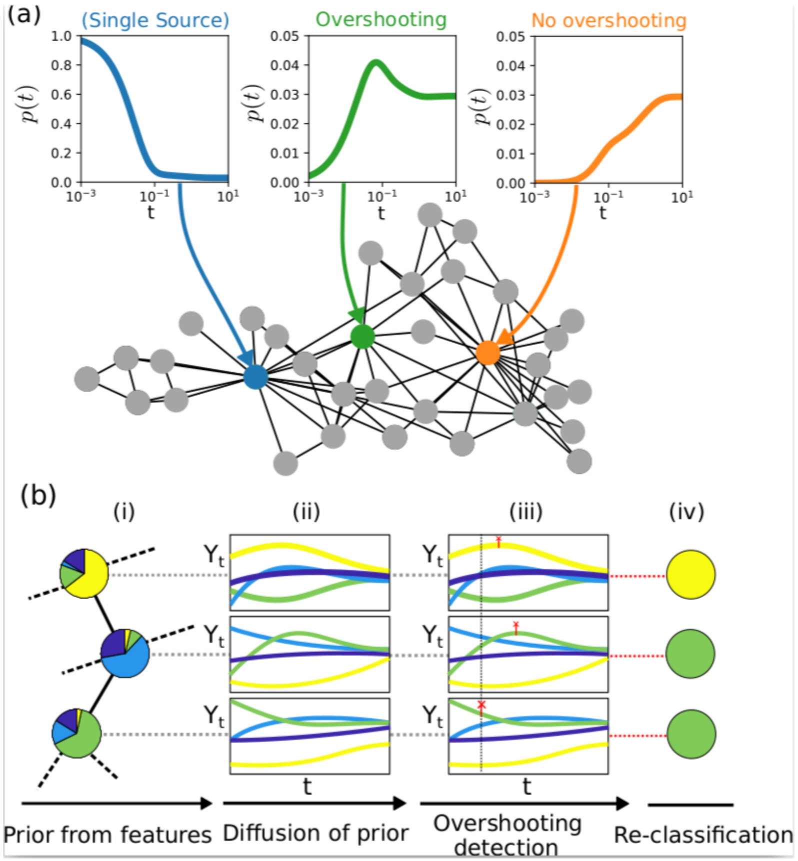
Graph Diffusion Reclassification

Node classifier such as GCN provide probability distribution on nodes to belong to each class.

For each class, we diffuse these distributions, and reclassify a node if a transient peak in another class is largest.

We can improve classification accuracies:

Method	Citeseer	Cora	Pubmed	Wikipedia
Planetoid	64.7	75.7	72.2	-
GCN	70.3	81.1	79.0	39.2
GDR (GCN)	**70.8** (+0.5)	**82.2** (+1.1)	**79.4** (+0.4)	39.5 (+0.3)

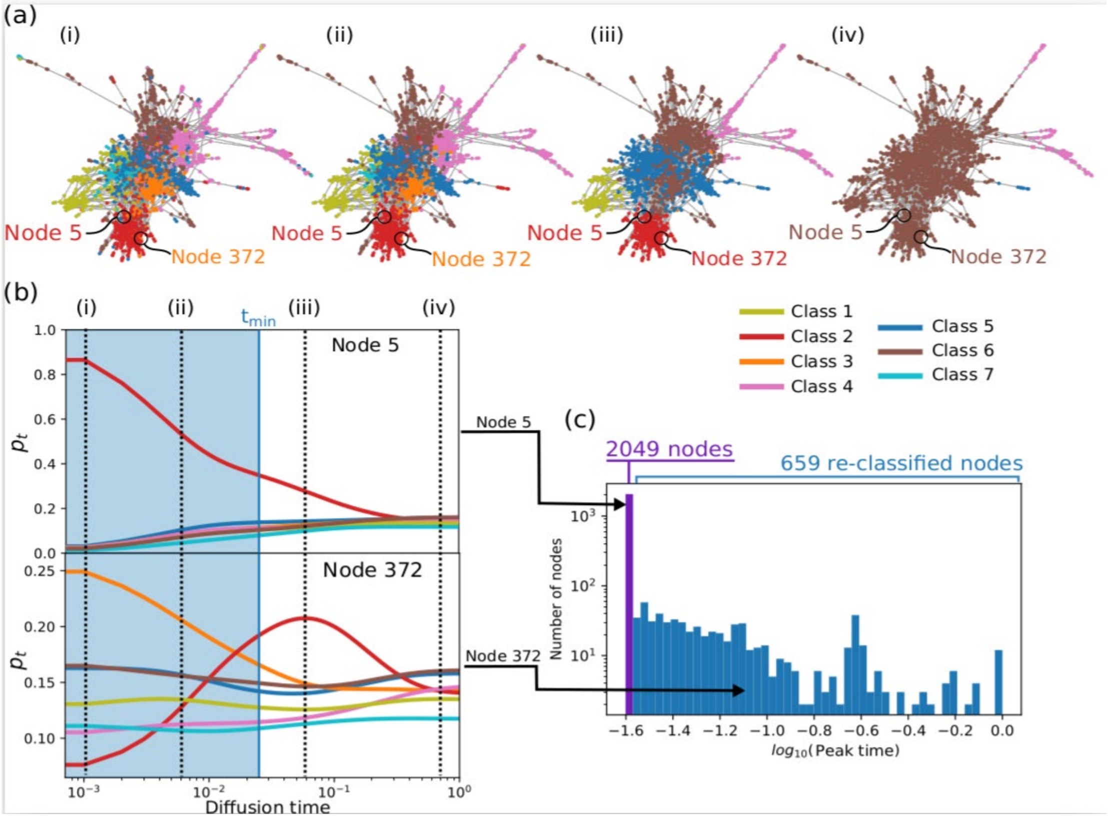


Reclassifying the Cora dataset

Diffusion of node classes:

Well classified node:

Reclassified node:



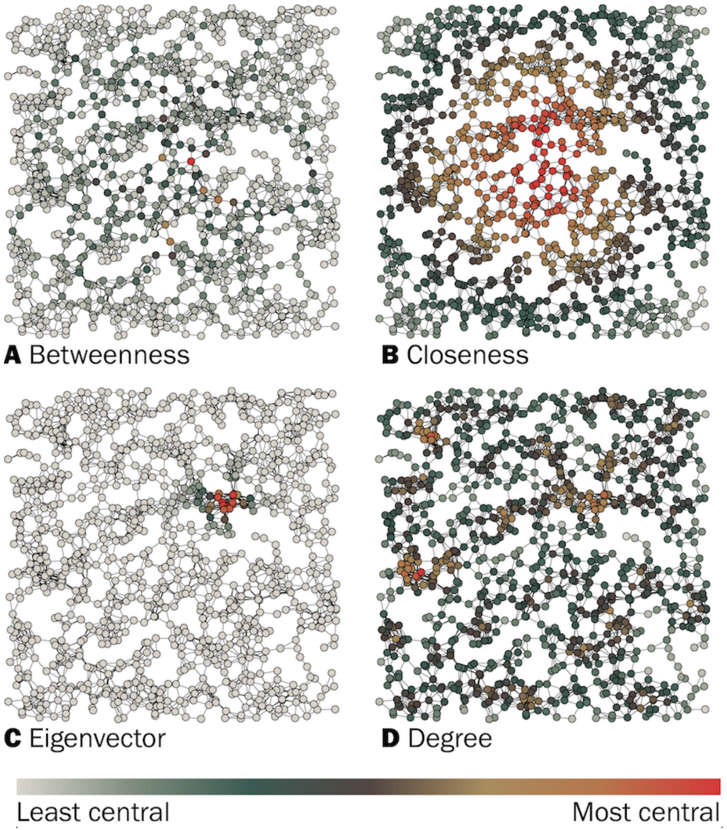
2. Using time

Graph centrality measures:

There exists many measures assigning a number to each node, the higher the more 'central'.

Here are a few simple ones:

- <u>Degree:</u> the node degree (first order measure, the more connected locally the more central)
- <u>Eigenvector:</u> largest eigenvector of adjacency matrix (the higher the centrality if my neighbours also have high values)
- <u>Betweenness:</u> number of shortest pass passing by the node (how often a node is between any two other)
- <u>Closeness:</u> average length of shortest path with all others



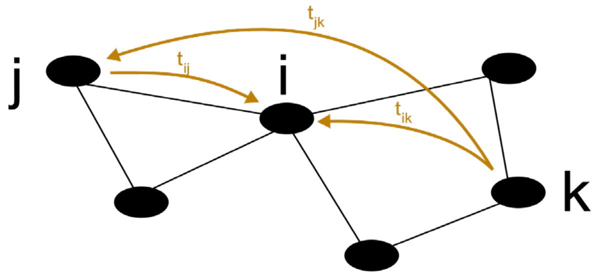
https://en.wikipedia.org/wiki/Centrality

Multiscale Centrality

We use the time of the transient response peak from a delta initial condition.

We could expect that this time would provide a distance measure between two nodes, but diffusion on graphs is ’non-trivial’, and often breaks the triangle inequality:

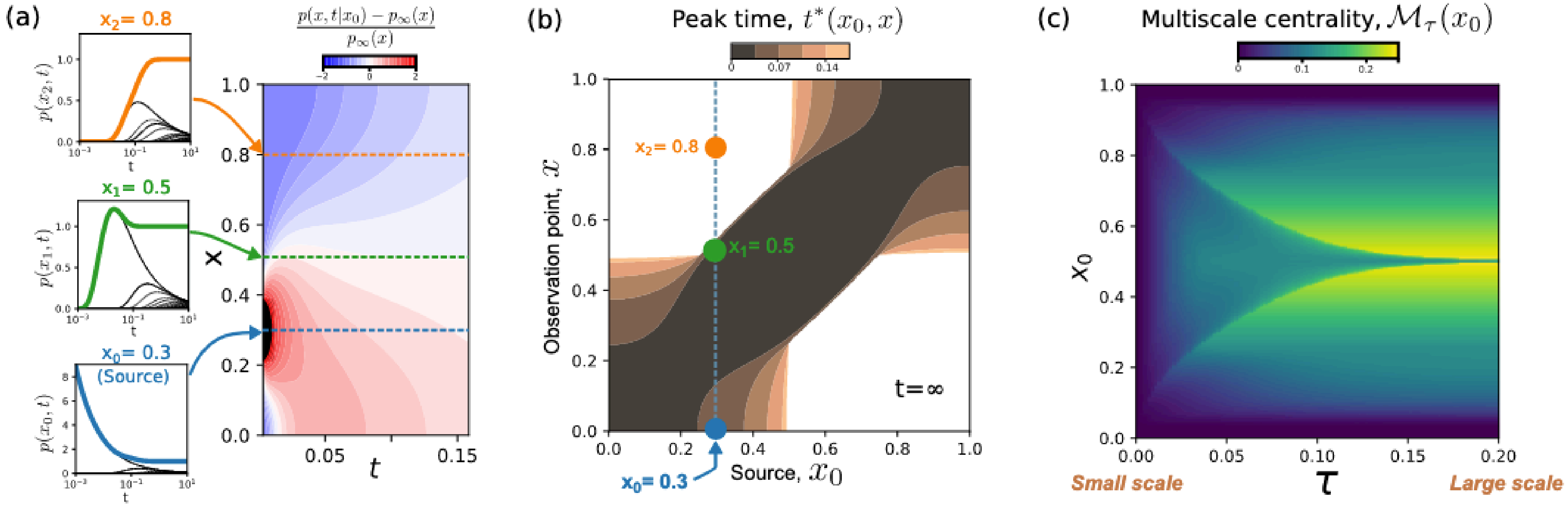
$$\Delta_{ij,\tau} := t^*_{ij,\tau} + t^*_{ik,\tau} - t^*_{jk,\tau} \leq 0$$



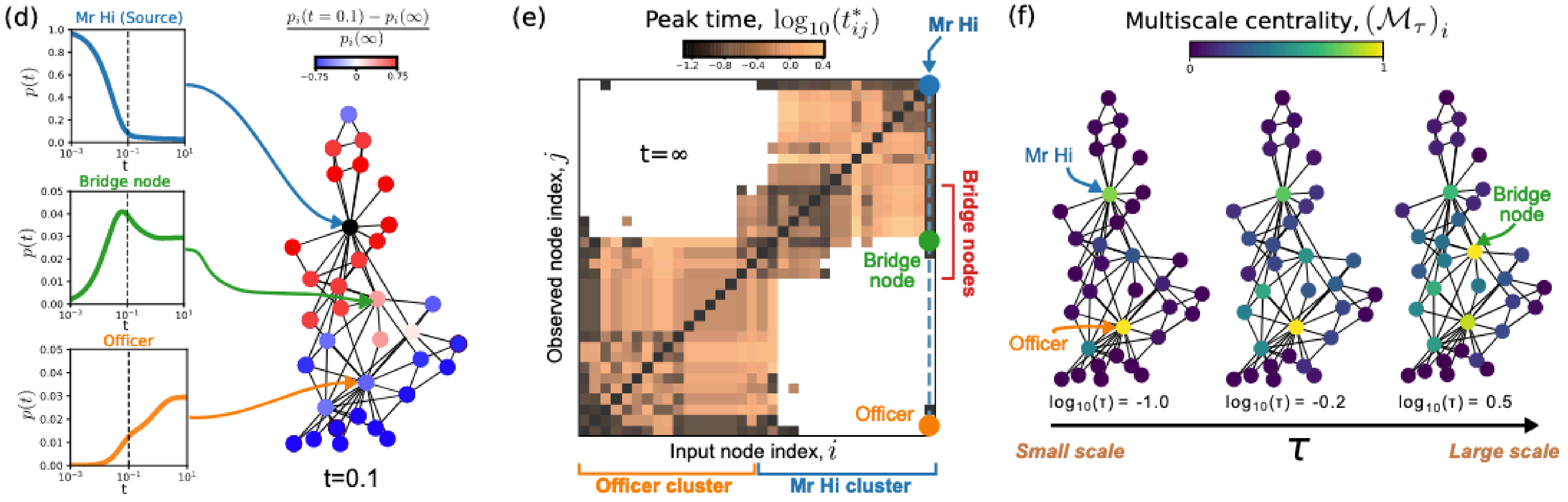
We then count the number of pairs of nodes that break the triangle inequality from source node, within a time horizon.

The fraction of such pairs yields our notion of multiscale centrality, parametrized by the time horizon.

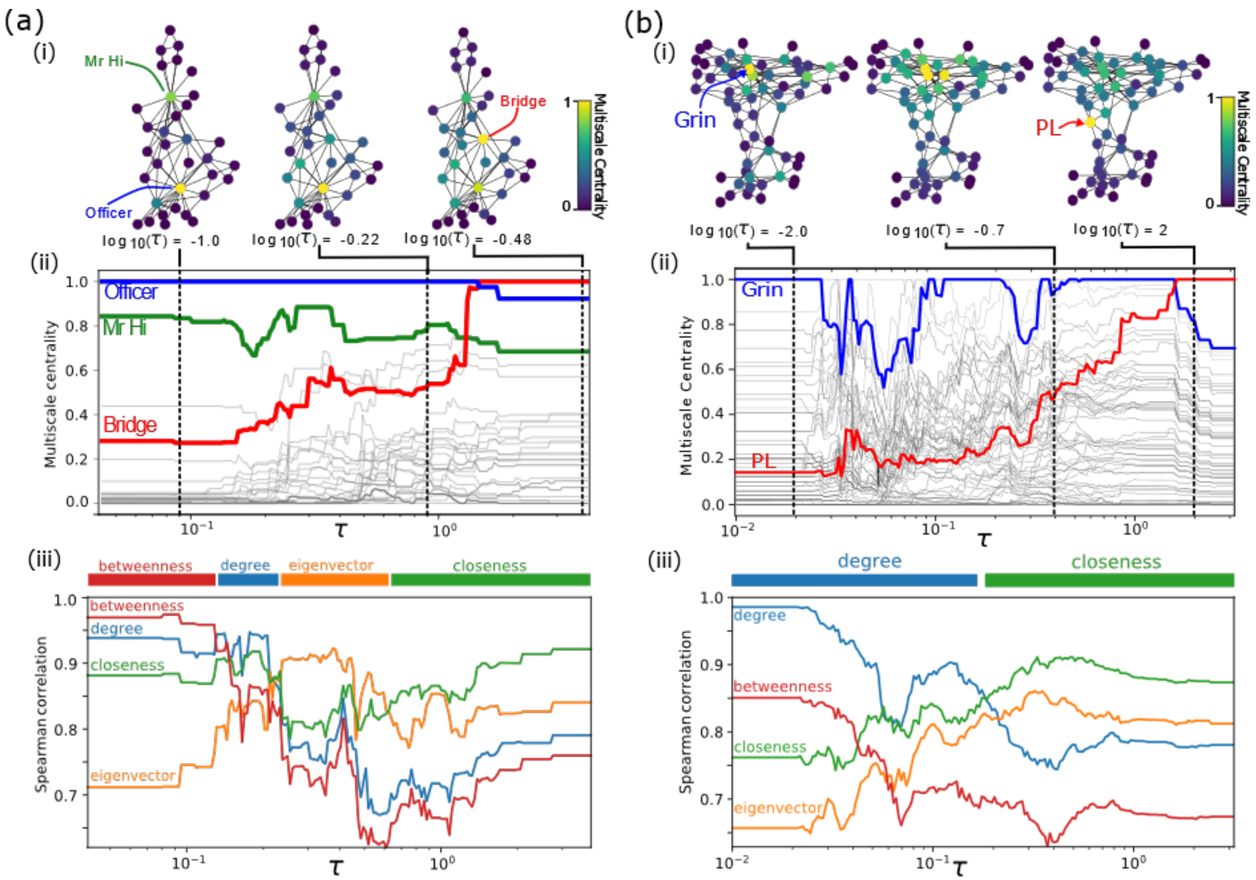
Multiscale centrality on the interval



Multiscale centrality on a graph



Comparison with other centrality measures

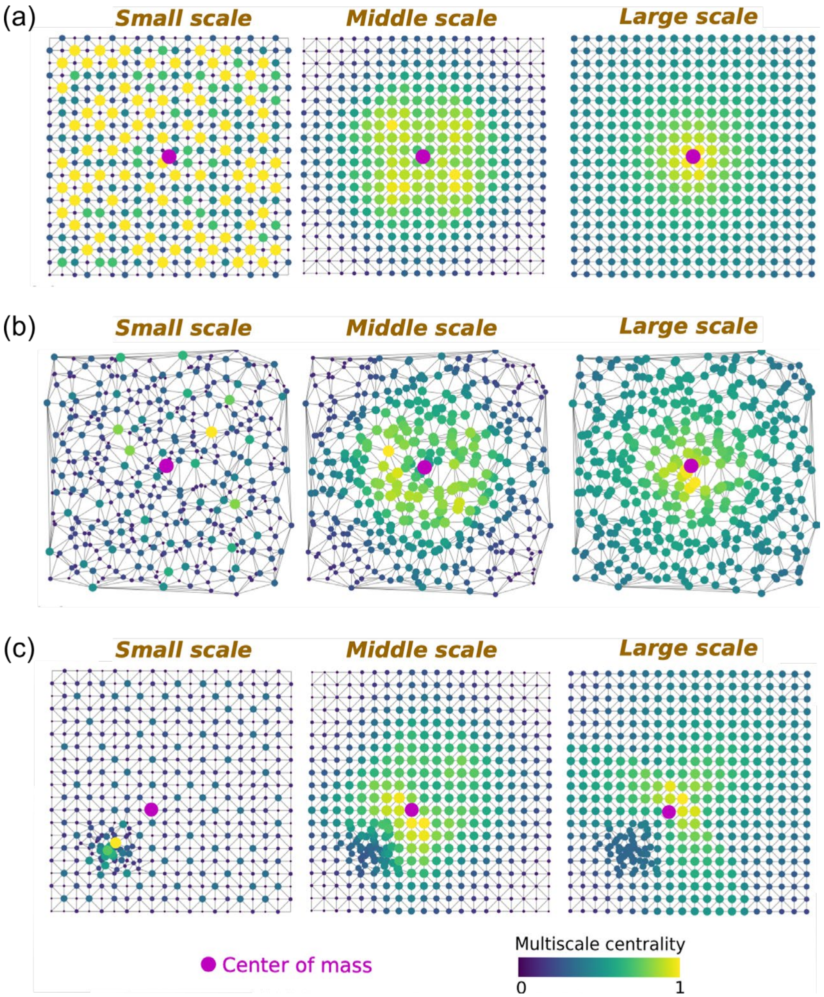


Simple examples

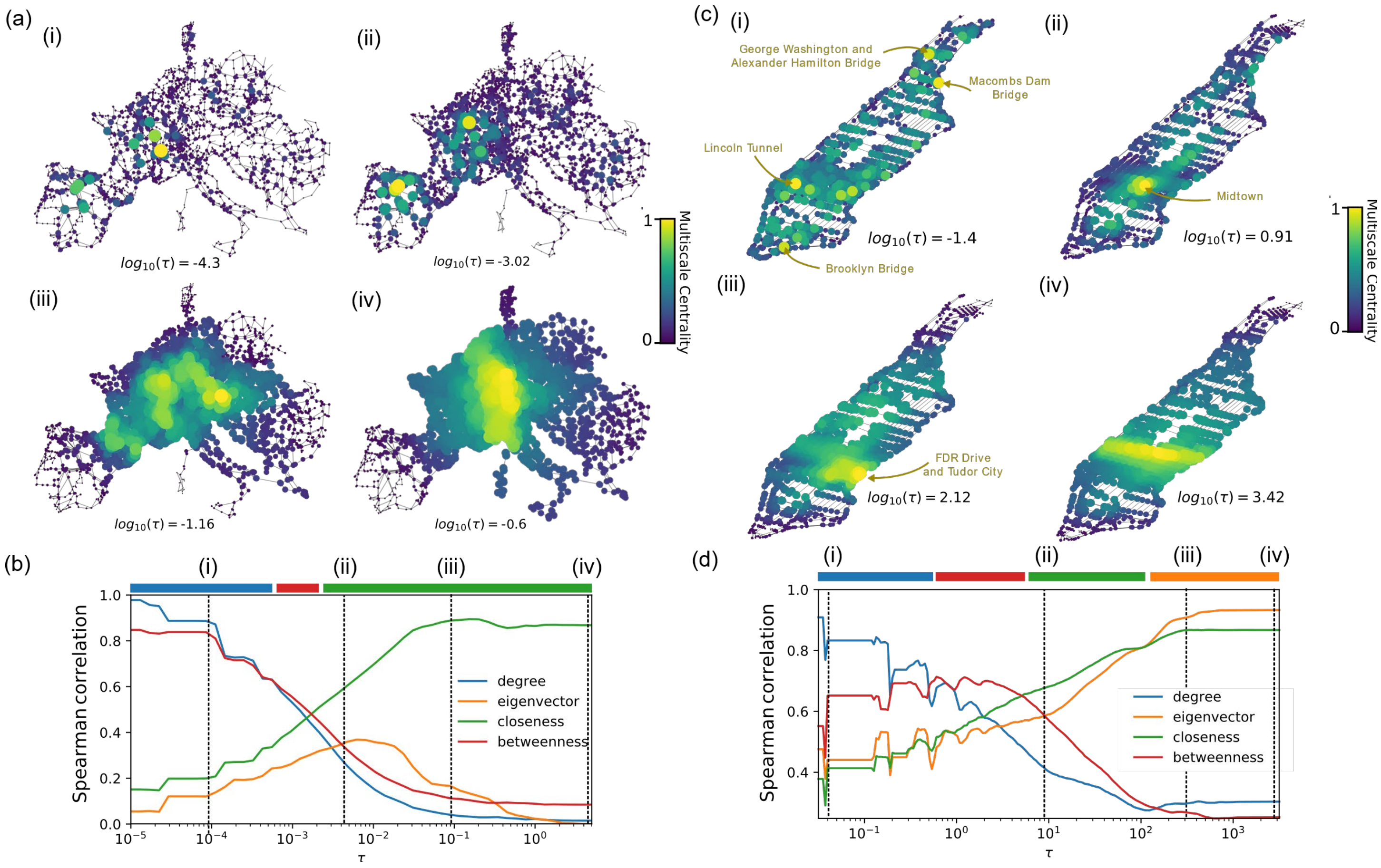
Uniform grid:

Delaunay mesh:

Uniform grid with mass:



More examples: European power grid and Manhattan road network



3. Using time and amplitude

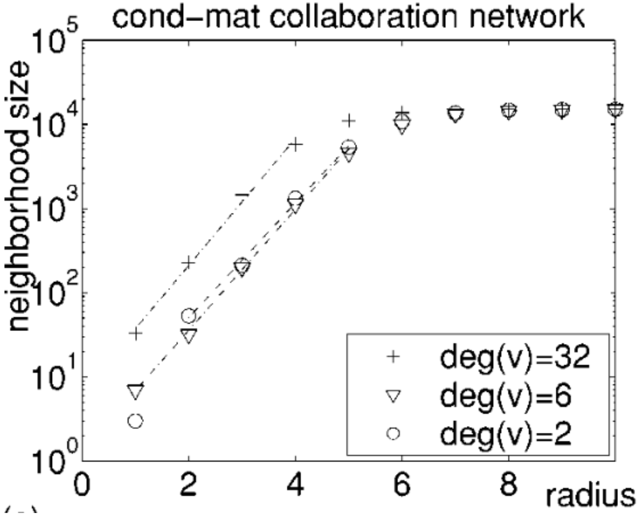
Short review on graph dimensions

- obvious for regular graphs (line 1d, grid 2d, etc…)

- In general, the 'static' definition is dimension of euclidean space to 'draw' the graph with edges of length = 1

- A physics-based definition how a number of nodes in a ball of radius r increases with r:

$$d = \lim_{r\to\infty} \frac{\log N_v(r)}{\log r} \qquad N_v(r) \sim r^d$$



Gábor and Szendrő (2004)

Multiscale dimensionalities on graphs

Finally, we use the time and amplitude of the transient peak to define notions of dimensions. Indeed, on Euclidean spaces, we can relate them with the dimension of the space.

From the Green function: $$G_t(x) = \frac{1}{\sqrt{4\pi\sigma t}^d} exp(\frac{x^2}{4\sigma t})$$

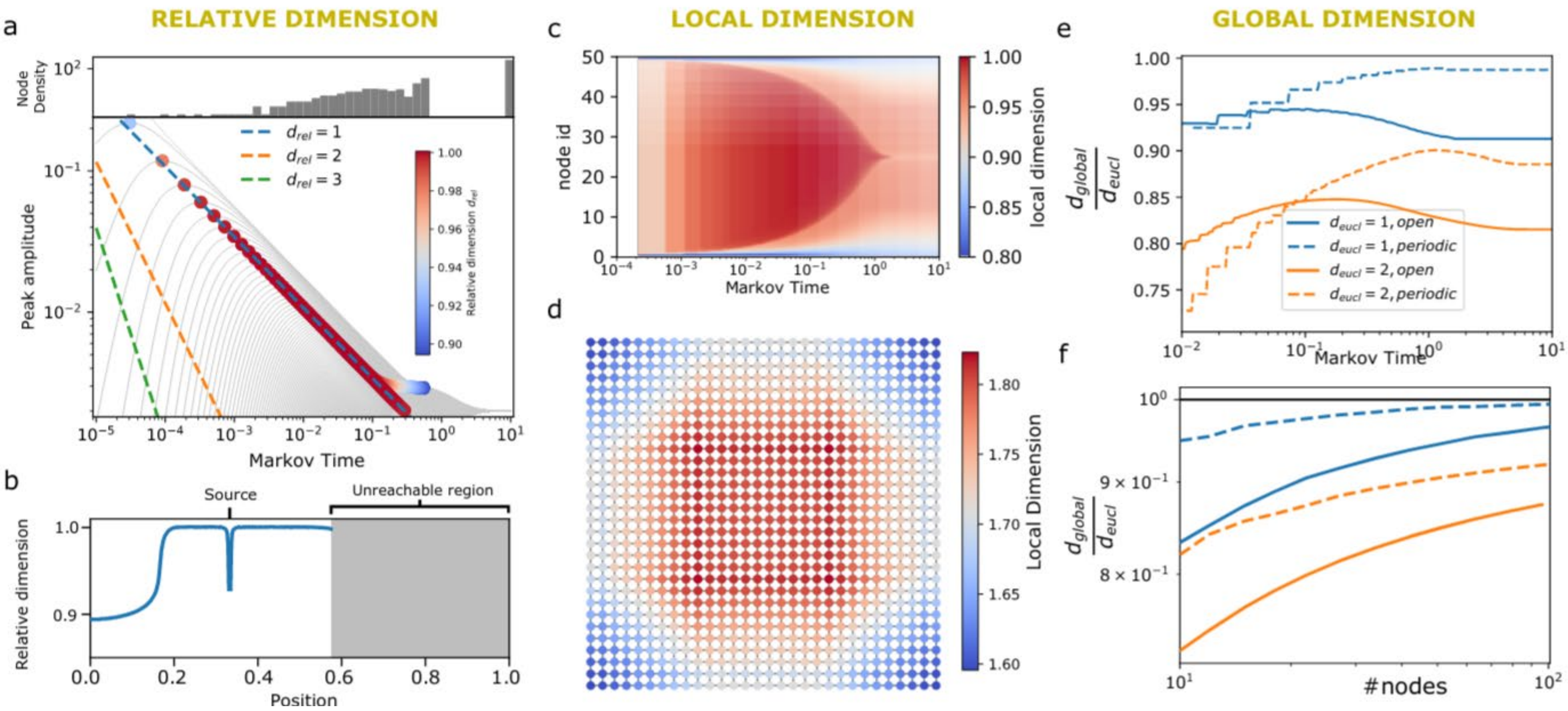
we have $$\hat{t}(\hat{\mathbf{x}}) = \frac{\|\hat{\mathbf{x}}\|^2}{2d\sigma}$$ and $$\hat{p}(\hat{t}) = (4e\pi\sigma\hat{t})^{-\frac{d}{2}}.$$

Which we can invert to get $$d(\mathbf{x}|\mathbf{x}_0) = \frac{-2\ln\hat{p}}{\ln\left(4e\pi\sigma\hat{t}\right)}$$ and on graphs: $$d_{ij} = \frac{-2\ln\hat{p}_{ij}}{\ln\left(4e\pi\sigma\hat{t}_{ij}\right)}.$$

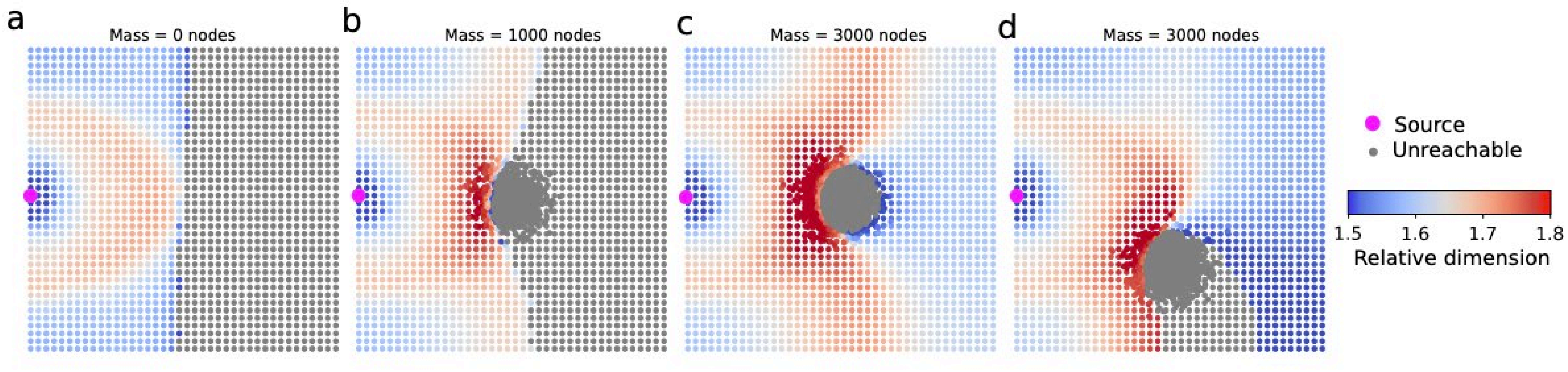
For each pairs of nodes, we have a notion **of relative dimension**. We also have a multiscale nature if we restrict times below a time horizon.

By averaging across nodes j, we have **local dimension**, and across all nodes I and j, a **global dimension**.

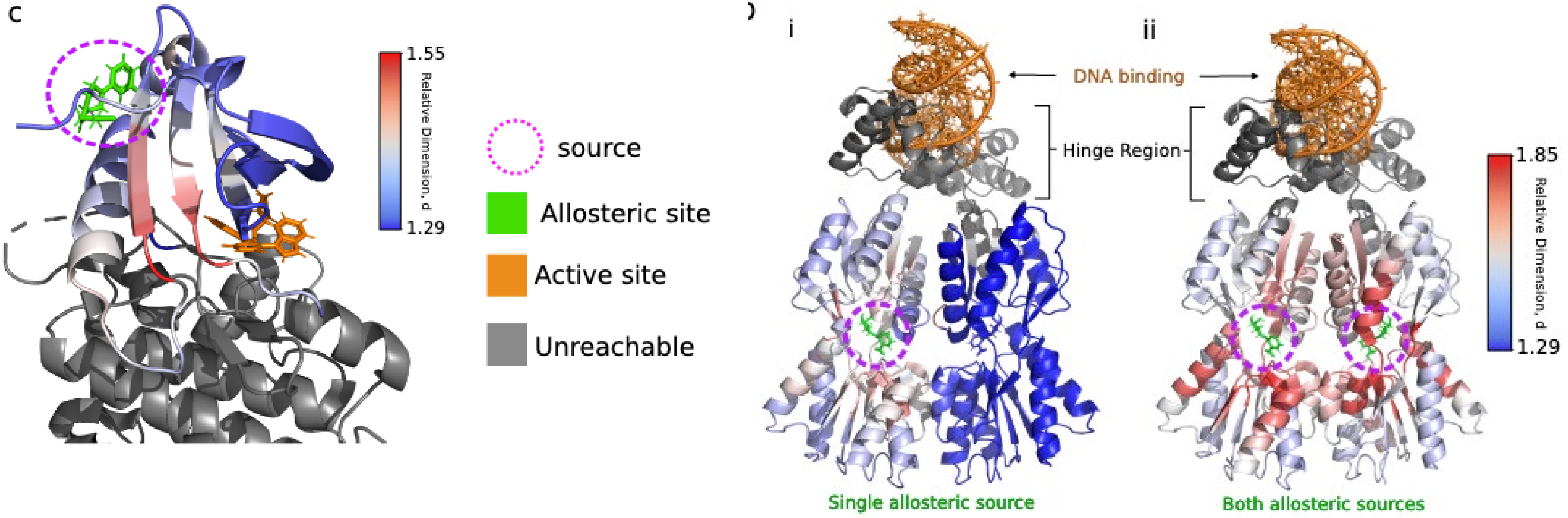
Line and grid example



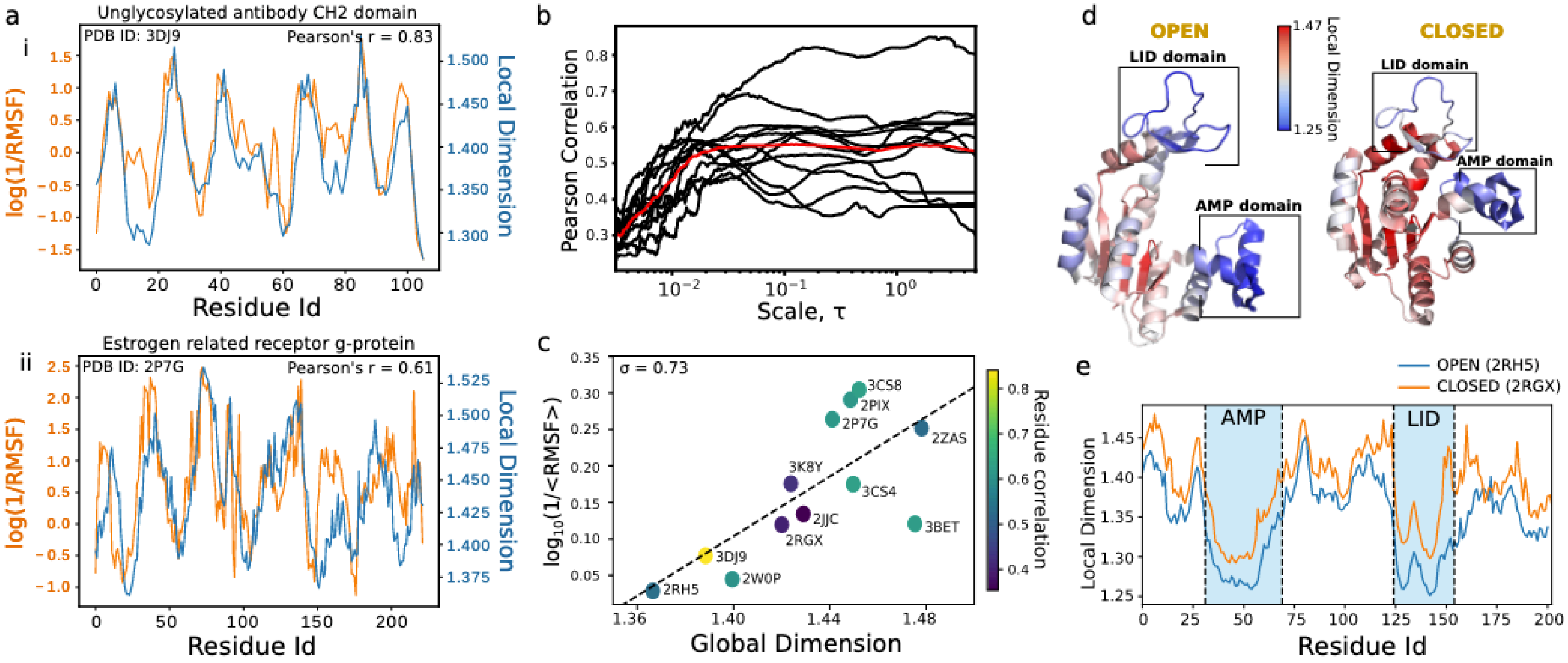
Some examples: ‘lensing’ effect



Some examples: protein structures

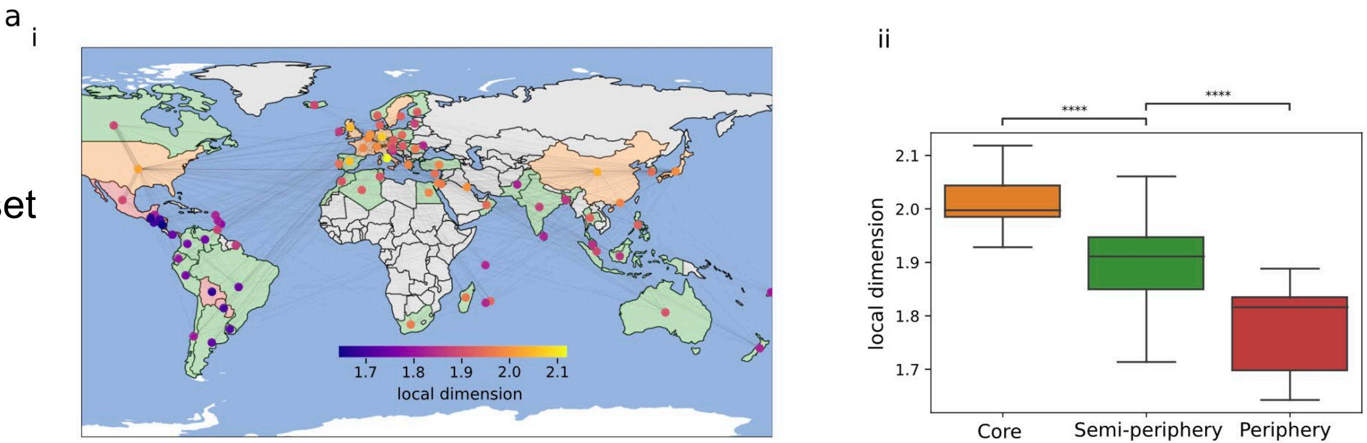


More protein structure example

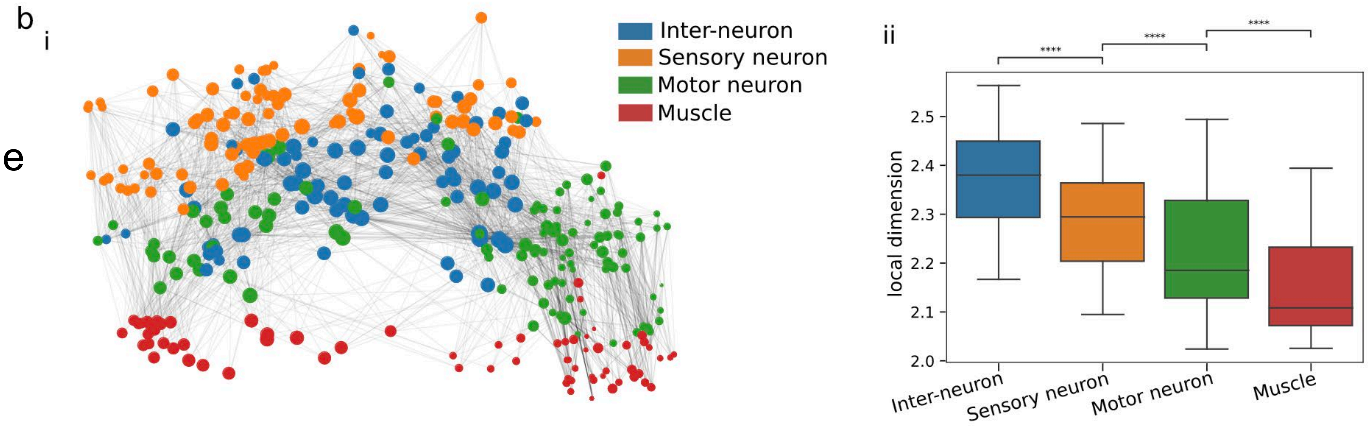


Some examples: trade data and connetome

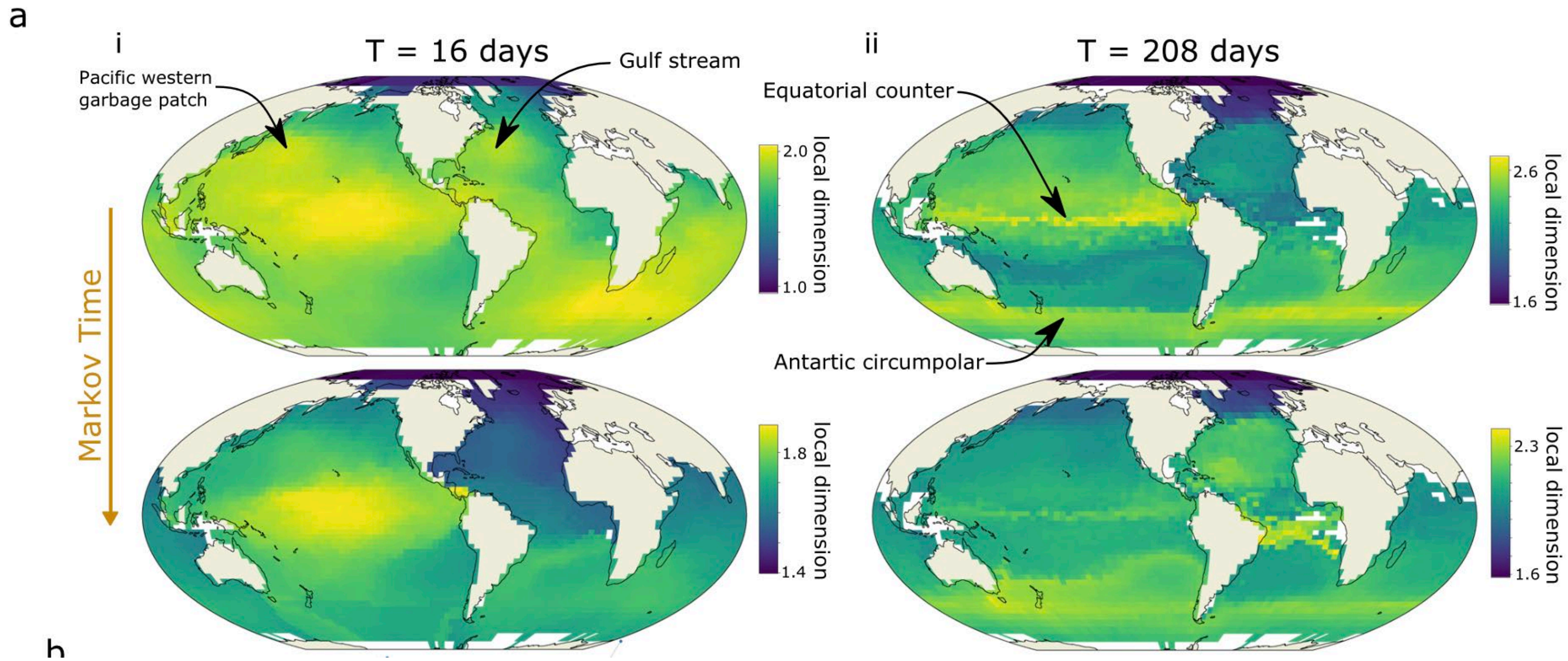
1994 world trade dataset



C. Elegans connectome



Some examples: ocean drifters



Summary

- Diffusion on networks is a powerful and versatile tool

- We have shown three ways to use it by only looking at the peak time/amplitude
    - Node classification
    - Multiscale centrality measures
    - Multiscale network dimensions (relative, local and global)

- These can be generalised, as long as a notion of diffusion is well-defined. These may include: directed graphs, multilayer graphs, higher order (hypergraphs, simplicity complexes, etc…), etc…

- Codes to compute them are available on github:
    - classification: https://github.com/barahona-research-group/GDR
    - centrality: https://github.com/barahona-research-group/MultiscaleCentrality
    - dimension: https://github.com/arnaudon/DynGDim