



**Politecnico
di Torino**



Master's Degree Thesis

A multi-scale geometric approach to compare healthy and pathological brain networks

International Master of Science programme in
PHYSICS OF COMPLEX SYSTEMS

Internal Supervisor:

Prof. Alfredo BRAUNSTEIN

Candidate

Alice LONGHENNA

External Supervisor:

Fabrizio DE VICO FALLANI

Inria Research Scientist



Academic year 2020/2021

Summary

Complex networks have been increasingly used to model brain connectivity derived from experimentally obtained data [1]. Topological properties, such as node centrality, efficiency and modularity have been shown to reveal complementary information on the brain functioning in healthy and pathological conditions [2]. This thesis tries to go beyond a first descriptive approach in terms of single scale topological properties: considering the presence of multiple scales, and addressing the problem in the novel and promising framework of networks latent geometry.

Indeed, even if the brain is naturally embedded in Euclidean space, an effective hyperbolic geometry can better account for its most characteristic topological properties, such as community structure, heterogeneous degree distribution, efficient navigability. Random geometric (with distance dependent connection probability) graph models in hyperbolic space were developed [3] and merged with unsupervised learning techniques to convey algorithms able to map existing networks to hyperbolic space, assigning coordinates to nodes in order to maximally reproduce the observed topology.

Embedding networks to a latent geometric space can serve as a form of unsupervised learning. For instance one of these methods applied to the brain [4] has resulted in a separation of left and right hemisphere nodes, as well as a clear angular separation of groups of nodes corresponding to neuroanatomical regions. But it can also be the basis for a geometric renormalisation of the network.

Brain networks indeed possess a rich architecture organized over multiple scales linked to one another (from single nodes to groups or clusters). The characterization of multiscale properties in complex systems is commonly addressed using the renormalization technique of statistical physics. The first attempts to coarse-grain networks were carried out in topological space (contraction of shortest paths, degree thresholding). Nevertheless, the presence of correlations between different scales in some complex networks (linked to small world property), limits the applicability of renormalization procedures directly on the graph or the associated connectivity matrix, in which the metric structure is given by shortest paths between nodes. Following these observations a geometric renormalisation protocol in hyperbolic space was presented in [5].

My work was centered on applying those state of the art models and methods to the characterization of brain structural connectivity of a cohort composed by healthy subjects and Alzheimer diseased patients.

The datasets I used were in the form of weighted adjacency matrices, representing the strength of structural connections (number of axonal pathways), the information of which was derived from magnetic resonance (Diffusion Tensor) imaging by the neuroimaging core facility of CENIR.

In the first part I derived network representations in hyperbolic space, using *Mercator* [6]: an algorithm which combines dimensional reduction (Laplacian Eigenmaps) and likelihood maximization (based on a random graph model in hyperbolic space [3]) approaches to assign coordinates to each node.

The accuracy of each embedding was validated comparing local and global topological and properties of the original networks with that of an ensemble of random graphs, generated from the assigned coordinates, given the model.

I then applied the transformation of coordinates in [5] to unfold the multi-scale structure of the networks (starting from $O(10^3)$ nodes down to $O(10^1)$) and computed various global and local (node specific) topological and geometric measures for each connectivity matrix.

The results for the two groups (healthy and patients) were then statistically compared. For some of the properties analysed, the two groups are found to be indistinguishable at the original scale, while not indistinguishable on rescaled networks, suggesting the possibility for a theoretical description of how this disease is affecting brain functionality at multiple, interconnected scales.

A needed follow-up would be performing the same analysis across experimentally obtained networks at different resolutions, with a comparable number of nodes with respect to the model rescaled ones. The eventual correspondence of the results obtained would be in favour of the ability of this multi-scale modelisation to reproduce the observed network properties; the opposite case could lead to the conclusion that the proposed method conveys different or more information with respect to the observations, immediately linked to the necessity of understanding from where this new information comes from.

Acknowledgements

Table of Contents

List of Tables	VIII
List of Figures	IX
1 Introduction	1
1.1 What is a Network?	1
1.2 Measures on networks	2
1.3 Brain Networks	3
1.4 Aim of the thesis	4
2 From the brain to its structural brain network	6
2.1 Structural brain networks construction from MRI	6
2.1.1 Tractography	6
2.1.2 Parcellation	6
2.2 Filtering	7
2.2.1 Threshold proportional	8
2.2.2 MST and link adding	8
3 Network geometry	10
3.1 Random graph models in Hyperbolic space	11
3.1.1 \mathbb{S}^1 model	13
3.1.2 \mathbb{H}^2 model	13
3.1.3 Hyperbolic geometry and greedy navigation within the model.	14
3.2 Embedding networks in Hyperbolic space	15
3.2.1 A model based method: <i>Mercator</i> [6]	16
3.2.2 Validation of embedding accuracy	18
4 Geometric Renormalization	20
5 Results of the multi-scale analysis	24
5.1 Testing the embedding algorithm	24

5.2	Emergence of peculiar network properties	25
5.3	Multi-scale analysis of global properties	27
5.4	Multi-scale analysis of coordinates' distributions	32
6	Conclusions	34
7	Some validations of embedding plots	36
	Bibliography	42

List of Tables

- 5.1 Comparison between different networks of original clustering and the one of the synthetic ensemble, average distance. Shows that clustering is not well reproduced in networks with β close to 1. . . . 25

List of Figures

1.1	Illustration of the construction of structural and functional brain networks. Taken from [10]	4
2.1	Ellipsoids representing local fiber orientation and corresponding tractograms (adapted from [11]). Right: structural brain network illustration from [12]	7
3.1	Nodes on a latent hyperbolic space (surface of an hyperbola) and projected on the x-y plan, with their connections. Pink and teal shapes are circles in hyperbolic space. Image from [3]	12
3.2	Example network represented using model S^1 (left) and \mathbb{H} (right). Taken from [18]	14
4.1	Illustration of the geometric renormalization transformation (adapted from [5].	21
4.2	Illustration of 4 layers obtained with the GR transformation, in this analysis Louvain communities where computed at the original layer, at the downscaled layers the same partition was kept and the nodes coming from two different communities where counted (bottom right numbers are the fraction of that nodes) and colored in gray.	22
4.3	Proportionality between nodes degrees, hidden degrees coming from the transformations 4.1 and hidden degrees inferred from embedding at each layer (each divided by their average value). Up-left figure is the original most resoluted layer, bottom-right the less resoluted layer.	23
5.1	Global efficiency and modularity comparing an Alzheimer diseased patient with healthy control and a random graph across scales. Layer 0 is the network (not filtered) at the original scale, layers > 0 are rescaled using 4	26
5.2	connector hub node illustration	27

5.3	Clustering coefficient and fraction of supernodes mixing communities comparing an Alzheimer diseased patient with healthy control and a random graph across scales.	28
5.4	Global efficiency and average degree, statistical comparison of the 2 groups. Stars indicate a p-value < 0.05 is found at that scale. Network filtered to $k = 20$	29
5.5	Modularity flow on networks filtered with MST method to $k = 10$ (right) and $k = 5$ (left).	30
5.6	Gobal efficiency, clustering coefficient, average degree and assortativity flows on networks with $k = 3$ filtered with MST method.	31
5.7	First row: angular separation of nodes composing supernodes at layer 0, and at layer 4. Second row: hidden degrees distributions comparison at layer 0 and at layer 6. Again red colour indicates the pathological group while blue the non-pathological. Networks filtered down to $k = 20$. The distributions are visualised using <i>kernel density estimate (KDE)</i> from Seaborn python library.	33
7.1	Validation of embedding for non-pathological subject number 33. Filtered with MST backbone method. 2.2 up to average degree 20 . .	37
7.2	Validation of embedding for non-pathological subject number 33. Filtered with threshold method. 2.2 up to average degree 20	38
7.3	Validation of embedding for pathological subject number 27. Filtered with threshold method. 2.2 up to average degree 20	39
7.4	Validation of embedding for non-pathological subject number 33. Filtered with MST backbone method. 2.2 up to average degree 3 . .	40
7.5	Validation of embedding for pathological subject number 27. Filtered with MST backbone method. 2.2 up to average degree 3	41

Chapter 1

Introduction

1.1 What is a Network?

Many real-world systems are characterised by a large set of interacting elements. The well-founded mathematical branch of graph theory gave the idea and the basic formalism to model these systems as a set of nodes connected by edges. This approach was applied to biological, technological, social networks, and it soon became clear that the topological properties of these systems were peculiar with respect to that of well-known lattices or random graphs.

A research effort in the last two decades was targeted to widen the set of available models and measures for graphs, in order to be able to reproduce and characterise some peculiar features observed in graph representations of real-world interconnected systems: for instance dynamicity in time, presence of highly interconnected groups of nodes (communities), presence of correlations between node specific properties, relatively short number of steps between any two nodes (small-world property [7]). These irregular, dynamical graphs, usually with a very large number of nodes, are studied, under the name of *complex networks*, integrating graph theory formalism with fields such as statistical physics and machine learning.

The words *network* and *graph* are associated to the same mathematical object defined by the pair (V, E) , where V is a set of vertices/nodes of size $|V| = N$. $E \subseteq V \times V$ is the set of edges/links, that can be directed, undirected (non-relevant direction of the pair-wise connections), weighted or unweighted (if only the presence/absence of the connection is relevant).

A graph can thus be represented as a list/set of pairs of connected nodes (*edge list/set*), but a more effective representation is the *adjacency matrix*. Call it A , with element

$$(A)_{ij} = \begin{cases} \omega_{ij} & (i, j) \in E \\ 0 & (i, j) \notin E \end{cases} \quad (1.1)$$

$\omega_{ij} = 1$ if the graph is unweighted.

1.2 Measures on networks

In this section I explain some properties I will use to characterise networks in this work, from the basic ones to more complex organisational features.

basic measures The number of connections of a node is called *degree* and is computed as:

$$k_i = \sum_{j \in N} a_{ij} \quad (1.2)$$

where N is the number of nodes, a_{ij} is an element of the adjacency matrix which is 1 if an edge exists between nodes i and j , 0 otherwise. In the case of weighted networks, the same measure is called *strength* and takes the weight w_{ij} instead of a binary input.

The *shortest path-length* is the only metric defined on topological space, it's the count of the minimum number of edges connecting two nodes (the sum of their corresponding weights in the case of weighted networks).

measures of segregation These quantify the extent to which information processing is performed in densely interconnected, specialised groups.

The simplest measure of segregation is the *clustering coefficient* C , counting the fraction of nodes' neighbours which are also neighbours of each other (triangles):

$$C = \frac{1}{n} \sum_{i \in N} \frac{2t_i}{k_i(k_i - 1)} \quad t_i = \frac{1}{2} \sum_{j,h \in N} a_{ij}a_{ih}a_{jh} \quad (1.3)$$

where t_i is the number of triangles around node i .

The more informative¹ measure of segregation is the modular structure, a known organisational feature of the brain. It is typically derived from optimisation algorithms, extracting the partition which maximises *modularity* in the network, the latter defined as:

$$Q = \sum_{u \in M} \left[e_{uu} - \left(\sum_{v \in M} e_{uv} \right)^2 \right] \quad (1.4)$$

where M is the set of non overlapping subsets of network nodes (partition), e_{uv} is the proportion of edges linking nodes in module v with nodes in module u ; so that essentially quantifies the amount of intra-connections with respect to inter-connections between modules.

¹Giving also the composition of these specialised groups (optimal partition).

Measures of integration These quantify the ability to efficiently transport information in the network and rapidly integrate the outputs coming from different specialised areas.

The basic rule is that the shorter the distance between two points the faster the communication so that a first measure of integration is the *characteristic path-length* (obtained averaging over shortest paths between any pair of nodes in the network); nevertheless, if the network has disconnected components some of these shortest path-lengths d_{ij} are ∞ and that measure is not computable. *Global efficiency* has the same meaning but involves inverse path-lengths which are 0 for disconnected nodes, so that it is always computable:

$$E = \frac{1}{N} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^{-1}}{n - 1} \quad (1.5)$$

The overall efficiency of many real-world systems, and for instance the brain, is guaranteed by the trade-off between *integration* and *segregation*: a faster and more resilient communication system (more densely connected) comes with an increased wiring cost.

The optimal wiring in the brain is organised in well distinct specialised areas that at the same time are effectively connected, to converge in a complex though highly efficient communication system [8].

This is usually referred to as *small-world* organisation of complex systems, which is quantitatively reflected by efficiency comparable and clustering coefficient higher than a random graph². Equivalently this property is identified by an average path-lengths increasing as $\sim \log(N)$ with the number of nodes N.

1.3 Brain Networks

Graphs have been increasingly used to model brain connectivity, stimulated by the increasing availability of experimental tools to record dynamic patterns of interaction between neuronal elements [9, 1].

The more intuitive discrete, networked representation of the brain is derived associating nodes to anatomical regions of interest in the brain and edges to the physical/axonal connections between them. The weight of these connections is usually quantified with the number of fiber tracts connecting the two regions, or their length, or a combination of the two. The resulting object is called *structural*

²I mean a classical random graph, in which edges are just assigned with a fixed probability in topological space.

brain network, describing and modelling them we try to answer the question of whether and how the structure of neuronal wiring is related to brain functioning.

Another notable representation is that of *functional brain networks*, in which nodes represent regions of the brain in which we are measuring an activity (they can be for instance the electrodes in a EEG measurement), and edges represent the correlation between the recorded signals (usually while performing a certain task).

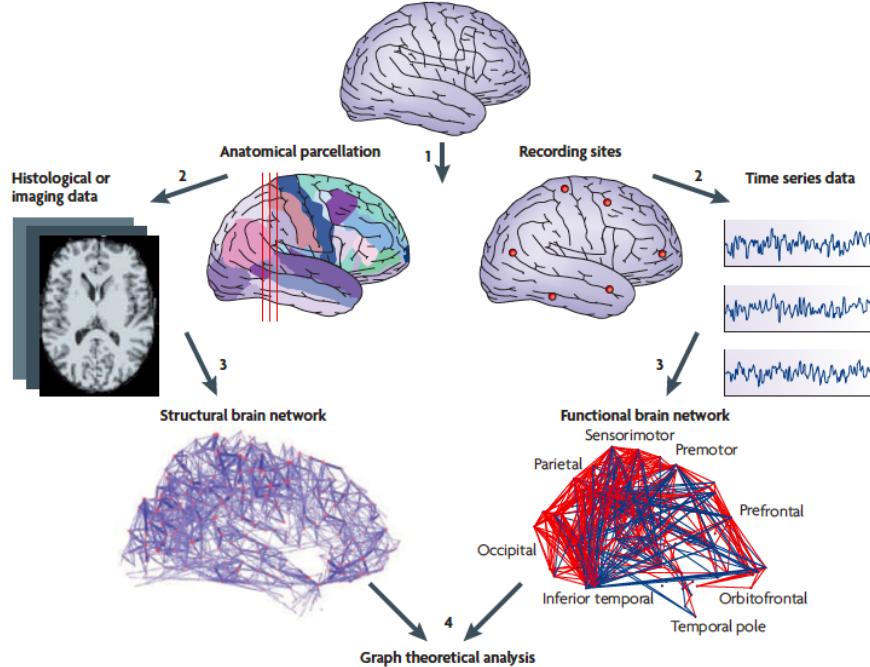


Figure 1.1: Illustration of the construction of structural and functional brain networks. Taken from [10]

In this work only structural brain networks will be treated, and their construction starting from Magnetic Resonance Imaging is described in ch.2.

1.4 Aim of the thesis

Topological properties (some illustrated in 1.2), such as the node centrality, efficiency (eq.1.5) and modularity (eq.1.4), have been shown to reveal complementary information on brain functioning in healthy and pathological conditions. However, the principles behind brain networks' organisation, and their relation to brain functioning, are still poorly understood.

This thesis tries to go beyond a first descriptive approach in terms of single scale

topological properties, considering the presence of multiple scales, and addressing the problem in the novel and promising framework of networks latent geometry (3).

Some recently developed models and methods are applied to human structural brain networks, coming from a group of healthy subjects and one of Alzheimer diseased patients. First, networks are embedded in an hyperbolic space, using a model based method mixing machine learning and likelihood maximisation, to obtain nodes coordinates (explained in ch. 3). Second, each network is coarse-grained and rescaled using a transformation of coordinates in the same hyperbolic space (explained in ch. 4).

Various global and local topological properties are computed on each down-scaled network and used to compare the two groups, with the goal of observing the difference between them at a certain scale, while it was not observable at the original one. The long term goals are to study brain organisation at multiple scales, also testing the relation with brain networks constructed at different resolutions; and determine whether and how these scales are interconnected to convey complex behaviour.

Considering the investigation of neurological disorders as a network problem has strong support from the fact that is well-known that many diseases manifest as the destruction of axonal pathways. Structural brain networks (their construction is explained in 2.1) represent these axonal connections, and enable a representation of their alteration in terms of edge weakening or removal, quantifiable through a wide range of global and local properties of the network.

Chapter 2

From the brain to its structural brain network

2.1 Structural brain networks construction from MRI

Magnetic Resonance Imaging (MRI) is a neuroimaging technique based on Nuclear Magnetic resonance: nuclei spin precession induced by strong magnetic fields causes the emission of an electromagnetic signal at radio frequencies which is recorded and localised in voxels (units of volume of the order of $1mm^3$) to construct a map of the system. These maps and the resulting images are different depending on the particular MRI technique adopted, the most common are T1 & T2 based, fMRI, DWI. The data I used in this project is coming from DWI (Diffusion Weighted Imaging), measuring diffusion trajectories of water molecules in the brain. The acquisition is performed for different gradient directions resulting in a tensor (represented as an ellipsoid).

2.1.1 Tractography

The information about the shape of white matter fiber tracts in the brain comes from tractography algorithms applied to DWI images, first estimating the distribution of the direction of propagation at each voxel, then building 3D trajectories from this local information.

2.1.2 Parcellation

Brain parcellation is the division of the brain into areas of interest that end up being the nodes of the network. There exist many techniques, anatomical-based

ones use anatomical landmarks, cortex curvature and architecture of neural cells, others are using clustering algorithms such as k-means, hierarchical clustering, applied to connectivity information derived for instance from tractography.

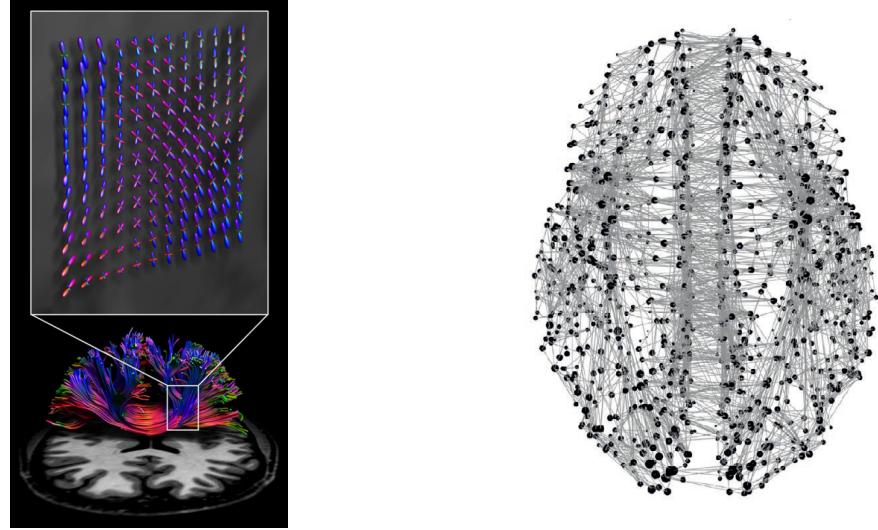


Figure 2.1: Ellipsoids representing local fiber orientation and corresponding tractograms (adapted from [11]). Right: structural brain network illustration from [12]

The networks I used were constructed from the anatomical parcellations built in [13], and composed of 23 brain networks coming from Alzheimer diseased patients, as well as 27 from non-pathological subjects, all with 1015 nodes and around 70000 links, weights represent the number of fiber tracts connecting the areas of interest.

2.2 Filtering

A connected network is obtained as a set of nodes and weighted edges, representable in the mathematical form of an adjacency matrix $A = \{w_{ij}\} \quad i, j = 1, \dots, N$. This representation of brain structure is usually thresholded, keeping only edges with weight above a certain value, or even binarised (edges that are kept are associated to an adjacency matrix element $a_{ij} = 1$). The reason for this operation is, other than simplifying the analysis, that weak links are more likely to be emerging from noise either in the acquisition or the image processing part and keeping them can prevent us to see some significant system's characteristics.

How to chose a significant threshold is a subtle issue, a range of values can be explored to see if at some point properties of the system that were before hidden emerge and remain observable.

I used two methods to filter the originally weighted networks and thus reduce the average degree:

2.2.1 Threshold proportional

From Brain Connectivity Toolbox for python¹, the function *threshold proportional* thresholds the network preserving a proportion $0 < p < 1$ of the strongest links, which means that we reduce the density of the network to p . Density of connections in a simple network² is related to the average number of connections (average degree) $\langle k \rangle$ by

$$p = \frac{\langle k \rangle}{N - 1} = \frac{2|E|}{N(N - 1)} \quad (2.1)$$

where E is the number of edges. Using this formula we can choose p in order to return a filtered network with a specific average degree.

Progressively pruning the links leads to the emergence of disconnected components and isolated nodes.

The embedding procedure (explained in ch.3), has problems when a disconnected network is given (in particular is returning the embedding of the largest connected component only).

After the thresholding, I kept only those networks that were still connected, to compare networks with the same number of nodes. In this way starting from the original networks with $\langle k \rangle \simeq 150$ and respectively 23 and 27 samples for the two categories (pathological and non-pathological) I ended up at $\langle k \rangle = 18$ having just 3 non-pathological subjects so that I stopped at $\langle k \rangle = 20$ to be able to perform meaningful statistical tests.

2.2.2 MST and link adding

In order to reach lower values of average degree while ensuring connectivity, I extracted a backbone of the original network computing the Maximum Spanning Tree, which is connected by definition. Then, having an initial average degree of $\simeq 1.5$, I added the missing edges in decreasing weight order. The number of edges to add $E - E_{mst} = E - (N - 1)$ in order to reach a target average degree k , is again given by the relation defining the graph density 2.1:

$$\frac{k}{N - 1} = \frac{2E}{N(N - 1)} \rightarrow E = \frac{kN}{2}$$

¹<https://github.com/aestrivex/bctpy>

²in which only one edge is defined between a pair of nodes

This way of reducing network density enables to obtain connected versions of the original network with an arbitrary reduced average degree. Nevertheless, the selection of a maximum spanning tree do not correspond to selecting the same number of edges with higher weight, and is thus not properly coherent with the aim of removing noise-related connections. The significance of networks filtered in this way has to be determined.

The embedding validation (explained in subsec.3.2.2) for the same network thresholded using the two different methods, down to the same average degree $k = 20$, is reported in 7. We can observe that the two representations are qualitatively indistinguishable, and both admit an accurate embedding.

Chapter 3

Network geometry

Some networks are spatially embedded in a geometric space, imagine traffic networks, social networks, the brain. Considering distance as a costly factor in links formation, seem a reasonable choice to model the behaviour of these interconnected spacial systems.

However, the actual geometry (usually Euclidean) is usually not satisfying in explaining the observed topological properties.

For instance, a random graph model in hyperbolic space [3] is currently the only one that can reproduce together: small-worldness and high clustering (eq.1.3), community structure, heterogeneous degree distribution, self-similarity [14], efficient navigability. Almost all the most characteristic properties of structural brain networks.

From these observation the study of *latent geometry* took the attention of the community lately. The fundamental question is whether it exists a geometric space from which the observed topology emerges naturally (for instance distributing nodes uniformly) as a reflection of the metric structure.

Beyond being a promising theoretical framework, network latent geometry is also paving the way for a bunch of practical applications to real-world problems as efficiently routing information in the Internet [15], community detection¹, link prediction [16] and many others.

¹Communities, as highly interconnected objects are expected to occupy delimited spacial regions according to a model in distance is inversely proportional to link formation.

3.1 Random graph models in Hyperbolic space

Random geometric graph models are characterised by a probability of connection between any pair of nodes depending on their distance in a metric space, with the general characteristic that the smaller the distance the higher the probability of connection between two nodes.

The rationale for the choice of a metric space in which distances are hyperbolic is that complex networks typically have a hidden hierarchical organization: indeed nodes are heterogeneous instances that can be usually classified into communities and sub-communities, and approximated by a tree-like structure, so that similarity distances in trees can be efficiently mapped to similarity distances in hyperbolic space, both exponentially expanding at a rate given respectively by the branching factor and the curvature. Moreover in [3] , Krioukov et al. formally show that generating points uniformly on the surface of a hyperboloid (negatively curved surface, two-dimensional hyperbolic space), and choosing a simple step probability of connection function

$$p(x) = \Theta(\mu - x) \quad (3.1)$$

is enough for heterogeneous topology to emerge. In particular, the resulting degree distribution is a power law. Intuitively, nodes that lay more near the center of the hyperboloid would have a higher degree, as a consequence of the structure of the metric space: hyperbolic distance x between points $(r, \theta), (r', \theta')$ is given by

$$\cosh(\zeta x) = \cosh(\zeta r) \cdot \cosh(\zeta r') - \sinh(\zeta r) \cdot \sinh(\zeta r') \cdot \cos\Delta\theta \quad (3.2)$$

where ζ is the curvature of the hyperbolic space, and $\Delta\theta = \pi - |\pi - |\theta - \theta'||$. This for sufficiently large $\zeta r, \zeta r'$ and $\Delta\theta$ can be approximated as:

$$x = r + r' + \frac{2}{\zeta} \ln\left(\sin\frac{\Delta\theta}{2}\right) \quad (3.3)$$

from which we can understand that central nodes (with smaller r) are likely to be at a small distance to many nodes, and thus have a higher degree.

The step connection probability function 3.1 is the simplest choice for a random geometric model and from a statistical physics' perspective can be seen as the zero temperature limit of a Fermi-Dirac connection probability:

$$p_{ij} = \frac{1}{e^{\beta(\epsilon_{ij}-\mu)} + 1} \quad (3.4)$$

which casts the ensemble of graphs generated to an entropy-maximising exponential family. Edges are modelled as non-interacting fermions with energy:

$$\epsilon_{ij} = f(x_{ij})$$

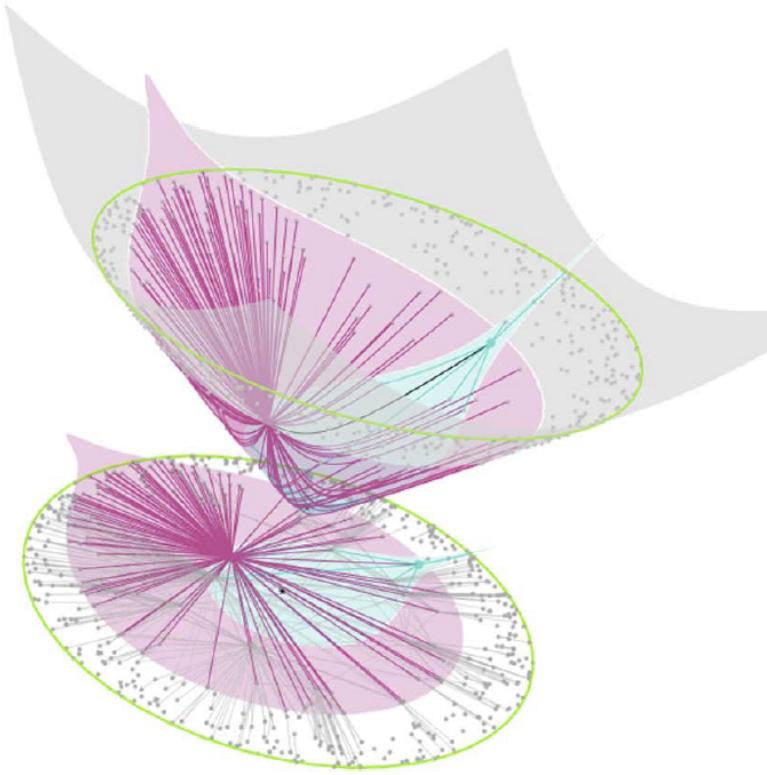


Figure 3.1: Nodes on a latent hyperbolic space (surface of an hyperbola) and projected on the x-y plan, with their connections. Pink and teal shapes are circles in hyperbolic space. Image from [3]

The ensemble is grand-canonical with chemical potential μ and inverse temperature β . Parameter μ fixes the average number of edges $|E|$ and thus controls the average degree in the network; parameter β controls clustering in the network, which is maximised at the ground state $T = 0 \leftrightarrow \beta \rightarrow \infty$ while tends to zero as $\beta \rightarrow 1$. As clustering decreases, the connectivity depends less and less on hyperbolic distance, as a consequence the metric structure is lost and so the congruency topology-geometry. Finally, at $T = 1$ the connection probability is not normalisable anymore and a phase transition point is encountered.

To reproduce small-worldness and high clustering the minimal requirement is $f(x) \propto \ln(x)$. To add heterogeneous degree distribution we have to include degree information in the probability, and in general any degree distribution can be reproduced adequately choosing the probability of connection.

In the following, I specify two well-known hyperbolic models in two dimensions which I will use in my analysis:

3.1.1 \mathbb{S}^1 model

It was introduced in [17], nodes are distributed along a circle of radius $R = N/2\pi$ (to set to 1 density along the circle) where N is the number of nodes. Each point is assigned a hidden degree $\kappa \in [\kappa_0, \infty)$, that is not equal to the actual degree but at least proportional (fig.4.3).

The energy is

$$\epsilon_{ij} = \ln \frac{x_{ij}}{\kappa_i \kappa_j} \quad x_{ij} = R \Delta \theta_{ij} = R(\pi - |\pi - |\theta_i - \theta_j||)$$

so that the corresponding probability of connection (eq.3.4) is decreasing with the arc-length (encoding similarity²) and increasing with the hidden degrees product (encoding popularity, the higher the degree the more "popular" is the corresponding node)³.

It can then be defined any distribution $\rho(\kappa, \theta)$ to obtain a desired degree distribution (for instance a Pareto distribution for the hidden degree κ leads to a power-law degree distribution).

3.1.2 \mathbb{H}^2 model

Defining radial coordinates from hidden degrees maps the hidden variable model \mathbb{S}^1 to a fully geometric one which is usually (as well as in my analysis) applied for visualisation purposes, but also makes the model invariant with respect to all isometries of Lorentz group. The two models are equivalent (statistically generate the same ensamble) choosing the change of variable:

$$r_i = R^{\mathbb{H}^2} - 2 \ln \frac{\kappa_i}{\kappa_0} \quad R^{\mathbb{H}^2} = 2 \ln \frac{N}{\mu \pi \kappa_0^2} \quad (3.5)$$

The probability of connection then becomes:

$$p_{ij} = \frac{1}{1 + e^{\beta/2(x_{ij} - R^{\mathbb{H}^2})}} = \frac{1}{1 + \chi_{ij}^\beta} = \tilde{p}(\chi_{ij}) \quad (3.6)$$

where χ is defined as a rescaled distance, x_{ij} is the approximate hyperbolic distance defined in 3.3.

Other dimensions D can be explored and a further parameter controlling curvature can be inserted. The higher the dimension the lower the clustering for the

²node similarity/dissimilarity can be defined in various ways, the simplest way is from presence/absence of edges between them, another one is the number of common neighbours.

³This *popularity-similarity* trade-off is a plausible principle for explaining connectivity of both static and growing networks.

same value of β so that in order to reproduce high clustering an upper bound on dimension is to be imposed, and $D=2$ is thus a reasonable choice.

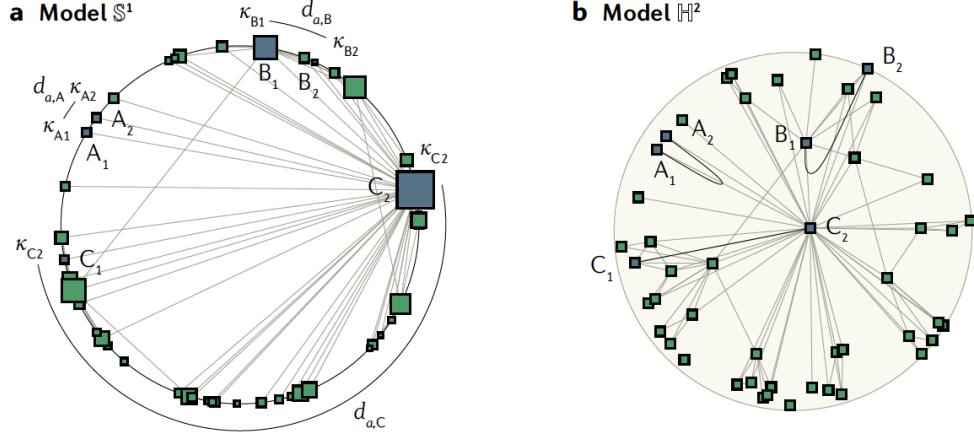


Figure 3.2: Example network represented using model S^1 (left) and H (right). Taken from [18]

3.1.3 Hyperbolic geometry and greedy navigation within the model.

Krioukov et al. in [3] studied information transport within the model, a phenomenon of interest for many systems to which the model can fit (brain, Internet, regulatory/metabolic networks). In a geometric space, nodes have an "address", and assuming they know the address of their neighbours and that the address of the target node is incorporated in the information packet to be sent, they can compute distances and perform a greedy navigation protocol, sending the message to the neighbour that is nearest to the target.

This process was simulated on random graphs generated from this hyperbolic model and efficiency of the process was tested computing: the percentage of successful paths (reaching their destination), the *hop stretch* (ratio between the number of edges in the greedy path and that in the shortest path, that ratio being equal to 1 for maximally efficient greedy navigation), the *hyperbolic stretch* (ratio between the distance travelled along the greedy path and the hyperbolic geodesic, again 1 for maximum efficiency). The result is that the process is maximally efficient, and thus network topology is maximally congruent with the hidden geometry, for strongest heterogeneity and clustering.

This because as I outlined before hyperbolic geodesics are maximally congruent with hierarchical paths, going from lower degree nodes to higher degree nodes

and back, for instance the percentage of shortest paths that are also hierarchical approaches 100% for networks with power law degree distributions of exponent $\gamma \rightarrow 2$. Another intuition is that high degree nodes cover a larger angular sector with their connections, making hierarchical paths the best way to forward information.

As for clustering it is pretty clear that the presence of triangles multiply the number of paths following approximately the same route and makes the network exceptionally robust to link removal/damages. Moreover, in the low clustering case (high temperature), the connectivity (manifested by connection probability behavior) depends less and less on hyperbolic distance, as a consequence the metric structure is lost (extreme case of classical random graphs) and so the congruency topology-geometry.

3.2 Embedding networks in Hyperbolic space

In a complementary way, we can think of mapping nodes to a geometric space, assigning coordinates to nodes in order to maximally reproduce the observed topology. This can work as a form of unsupervised learning of systems' organisational principles.

For instance one of these methods [19] applied to the brain [4] has resulted in a separation of left and right hemisphere nodes, as well as a clear angular separation of groups of nodes corresponding to neuroanatomical regions.

In general working in a geometric framework is a source of new measures and analyses that prospect to convey new information hard to detect at the adjacency matrix level.

digression on Euclidean space This problem was initially tackled as an unsupervised learning problem in Euclidean space, applying state of the art linear and nonlinear dimensional reduction techniques. It has soon been clear that linear mapping methods were not best suited to unveil the complex dependencies in the brain.

One example of this approach is [20], in which nonlinear dimensional reduction gave an intuitive interpretation of some network properties, for instance regions that were labelled as rich-clubs⁴ where found in central positions in the mapped representation, moreover rich-club nodes removal was leading to a ring-like shape as a reflection of a destruction of its centred architecture. Nevertheless, in original embedding realisations, many regions that were not labelled as rich-clubs were found to be central as well, revealing the need of a more coherent representation.

⁴Rich-club nodes are high degree nodes that are also connected with each other.

The following idea was that of considering non-Euclidean latent spaces, given for instance the success of hyperbolic random graph models in reproducing many complex networks properties.

For instance, even if the brain is naturally embedded in a three-dimensional Euclidean space, the study in [21] shows that a greedy targeted navigation process (the same explained in subsec.3.1.3), which is a plausible mechanism for information transfer in the brain, is much more efficient in brain networks *embedded in hyperbolic space* (sec.3.2), than considering the actual Euclidean distances.

The embedding algorithms that have been developed since now are either generative model based, or machine learning (non-linear dimensional reduction) based.

Model-based methods aim to extract model parameters that maximise the likelihood that the observed network topology emerges from the assigned coordinates. The result is more easily interpretable, but to be accurate requires the model used to be a good generative model for the system under study. Moreover, they are typically slower than any dimensional reduction algorithm. For instance, methods used in [22, 23] are of this kind, both based on the Popularity×Similarity Optimisation model (PSO), which essentially identifies node degree and similarity between nodes as the driving forces of connectivity. Their difference is the definition of *similarity* which in the first is given by presence/absence of an edge, while in the second by the number of common neighbours.

Machine learning based methods are instead fast, model independent and flexible enough to provide a non-trivial geometric representation of every network. They define a metric of pair-wise node similarity, and then seek the vectorised representation that best preserves the overall similarity of the observed graph. The main problem is that the majority of machine learning methods are not adapted to non-Euclidean space, and thus are only able to infer angular positions on S^1 circle. This is why these approaches are usually model-corrected in order to perform a two-dimensional hyperbolic embedding, ending up being model-dependent as well. Among them I cite coalescent embedding ($O(N^2)$) [19], employing manifold-based unsupervised machine learning techniques (like Isomap, Laplacian Eigenmaps) and minimum curvilinearity [24].

3.2.1 A model based method: *Mercator* [6]

Model based and data driven approaches can be efficiently combined to take advantage of both fastness and interpretability. Mercator, the algorithm I'm using in this work, does so by performing a first, fast embedding using model-corrected Laplacian Eigenmaps (LE). The latter can be further refined using itself as initial

condition for likelihood maximisation. The model used is the static PSO model \mathbb{S}^1 explained in sec.3.1.

Laplacian Eigenmaps LE was designed as a dimensional reduction technique, in that case a set of points in feature space $\{\bar{x}_i\}$, $\bar{x}_i \in \mathbb{R}^n$ is given and a random geometric graph is constructed according to some criteria (for instance connecting points at distance less than a certain threshold). The distances d_{ij} between points are then transformed into similarity scores to be used as weights for the network. The graph Laplacian is then constructed as:

$$L = D - W \quad w_{ij} = (W)_{ij} = a_{ij} \cdot e^{-d_{ij}^2/t}$$

where D is a diagonal matrix containing node degrees, W is the weighted adjacency matrix, $a_{ij} = (A)_{ij}$ the binary adjacency matrix entry, t a parameter and d_{ij} the distance. Then, spectral matrix decomposition is applied, obtaining the dimensionally reduced set of coordinates $\{\bar{y}_i\}$, $\bar{y}_i \in \mathbb{R}^m$, $m < n$ as $\bar{y}_i = (v_1^i, \dots, v_m^i)$ with v_j^i the i^{th} component of the j^{th} Laplacian eigenvector (ordered by eigenvalue magnitude).

Using this method in the context of network embedding results in a modified implementation: we essentially assume that the network to be embedded is the random geometric graph generated in the higher dimensional space. In this case we don't have access to the original coordinates, but we still need the distances to assign similarity weights and compute the graph Laplacian. The model is used for this purpose: distances are then estimated from the model as the expected chord length $d_{ij} = 2\sin(\langle\Delta\theta_{ij}\rangle/2)$, $\Delta\theta_{ij}$ expected angular separation in model \mathbb{S}^1 , given that the nodes are connected. At the end of this phase we learned angular positions of nodes $\{\theta_i\}$, $i = 1, \dots, N$.

ML estimation The likelihood of an embedding realisation is

$$\mathcal{L}(\{\kappa_i, \theta_i\} | \{a_{ij}\}, \mathbb{S}^1) = \frac{P(\{\kappa_i, \theta_i\}) \mathcal{L}(\{a_{ij}\} | \{\kappa_i, \theta_i\}, \mathbb{S}^1)}{\mathcal{L}(\{a_{ij}\} | \mathbb{S}^1)} \quad (3.7)$$

But since we don't know the prior distribution $P(\{\kappa_i, \theta_i\})$, we use an improper one, which makes it flexible enough to be applied to networks with arbitrary degree distributions. Assume $P(\{\kappa_i, \theta_i\}) = cost$ and maximise the log-likelihood:

$$\ln \mathcal{L}(\{a_{ij}\} | \{\kappa_i, \theta_i\}, \mathbb{S}^1) = \sum_{i < j} [a_{ij} \ln p_{ij} + (1 - a_{ij}) \ln (1 - p_{ij})] \quad (3.8)$$

The maximisation with respect to the hidden degrees, considering the form of connection probability p_{ij} in \mathbb{S}^1 leads to:

$$\frac{\partial}{\partial \kappa_l} \ln \mathcal{L}(\{a_{ij}\} | \{\kappa_i, \theta_i\}, \mathbb{S}^1) = \frac{\beta}{\kappa_l} \sum_{i \neq j} (a_{il} - p_{il}) = 0 \rightarrow k_l = \langle k_l \rangle \quad (3.9)$$

so that the hidden degrees are estimated optimising for each node the matching between the expected degree in \mathbb{S}^1 model and the observed degree k_i , starting from $\{\theta_i\}$ homogeneously distributed and $\kappa_i = k_i$.

Starting from this result, angular coordinates are refined visiting each node once, proposing new coordinates around it from a normal distribution, and choosing the one that maximally increases the likelihood. The result from model-corrected LE is thus used as initial condition for the optimisation procedure, strongly reducing the region of configuration space to explore.

The refinement of hidden degrees can then be repeated starting from the new angular positions inferred.

The parameter β is adjusted iteratively to match the average clustering predicted by the model with the observed one, the parameter μ (controlling average degree) is evaluated from β and expected degree.

3.2.2 Validation of embedding accuracy

Synthetic random networks can be generated from embedded coordinates given model connection probability. The accuracy of an embedding is evaluated from the comparison of various properties of the original network with that of an ensemble of synthetic networks generated from the inferred coordinates. In the following, I report a list of these properties computed and plotted for the scope:

1. Model connection probability as a function of rescaled distance $\frac{d_{ij}}{\mu \kappa_i \kappa_j}$, using parameter values estimated in the embedding procedure. To be compared with the empirical connection probability: computed as the fraction of connected nodes at a certain rescaled distance, divided by the total number of nodes at that distance.
2. The distribution of inferred angles.
3. The comparison of node by node degree, sum of degree of neighbours, average degree of neighbours, number of triangles, clustering, between the original network and the ensemble average on synthetic networks.

For each of these measures, the correspondence is quantified computing ζ (fraction of nodes whose original value is more than two standard deviations away from the ensemble average), Pearson correlation coefficient ρ and χ^2 .

4. The empirical on the original network degree distribution and that emerging from the ensemble.
5. The sum and average of degree of neighbours, number of triangles and clustering as a function of degree.

Some of these validation plots are reported in 7.

Chapter 4

Geometric Renormalization

Brain networks possess a rich architecture organised over multiple scales linked to one another (from single nodes to groups or clusters). The characterisation of this multi-scale property in complex systems is commonly addressed using the renormalization technique of statistical physics.

The presence of correlations between coexisting scales in complex networks limits, however, the application of renormalization procedures directly on the graph or the associated connectivity matrix (shortest-path-distance renormalization, degree-thresholding renormalization). Passing from the metric of shortest paths in topological space, to that of distances in latent space, opens the door to a proper geometric framework for studying the symmetries of the system and investigating its multi-scale organisation, with the possibility of borrowing concepts and methods from renormalisation group in statistical physics.

If the network is efficiently embedded, the geometric renormalisation protocol presented in [5], based on the hyperbolic static network model \mathcal{S}^∞ [3], offers the possibility to unfold its multi-scale structure.

Once nodes are mapped to the hyperbolic disk, the transformation is implemented as follows:

1. In layer l , group nodes in blocks of r consecutive elements along the angular direction¹, which are merged into a single node (*supernode*) at layer $l+1$.
2. Recompute links for layer $l+1$: If at least one node of a block is connected to at least one node of another block at layer l , the correspondent supernodes are connected at layer $l+1$.

¹The starting point is not statistically relevant

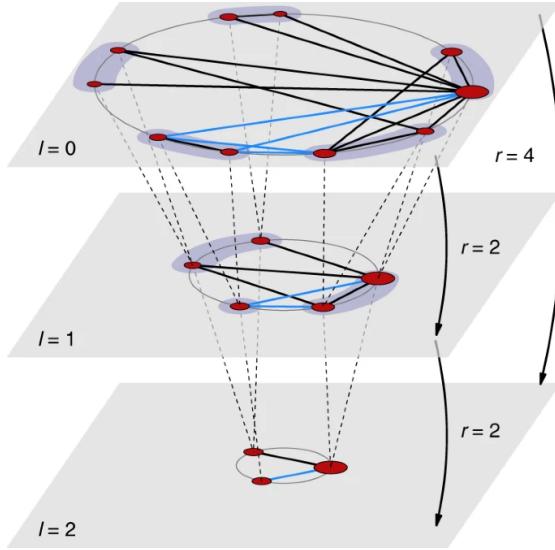


Figure 4.1: Illustration of the geometric renormalization transformation (adapted from [5].

3. Recompute coordinates for layer $l+1$: the transformations

$$\kappa_i^{(l+1)} = \left(\sum_{j=1}^r (\kappa_j^{(l)})^\beta \right)^{1/\beta} \quad \theta_i^{(l+1)} = \left(\frac{\sum_{j=1}^r (\theta_j^{(l)} \kappa_j^{(l)})^\beta}{\sum_{j=1}^r (\kappa_j^{(l)})^\beta} \right)^{1/\beta} \quad (4.1)$$

with the rescaling of parameters:

$$\beta^{(l+1)} = \beta^{(l)} \quad \mu^{(l+1)} = \mu^{(l)}/r \quad R^{(l+1)} = R^{(l)}/r \quad (4.2)$$

are chosen since they keep the probability of connection maximally congruent to \mathbb{S}^1 model and preserve semi-group structure $(\kappa_i'')_r = (\kappa_i')_{r^2}$, $(\theta_i'')_r = (\theta_i')_{r^2}$.

If the network admits an accurate embedding, and is thus compatible with \mathbb{S}^1 model, that transformation maintains the probability of connection in its original form. This produces, at any scale (apart from finite size effects), networks belonging to the same ensemble, but with a different average degree, that is to be rescaled to obtain self-similar network replicas.

I developed an algorithm that gets coordinates and parameters coming out from the embedding, rescales them following equations 4.1 and 4.2, and recomputes links. The coordinates' transformation is based on \mathbb{S}^1 model, while I used \mathbb{H}^2 model for visualisation purpose. I only considered grouping nodes in blocks of 2 consecutive elements ($r=2$) in my analysis.

As I already said I included the computation of various global and local (node-specific) topological properties that I will use to compare the pathological and non-pathological groups at multiple scales.

I also included at each layer the evaluation of communities using Louvain community detection algorithm [25], and computed the fraction of nodes belonging to different communities that are merged into the same supernode at the following layer (that I will call "fraction of nodes mixing communities"). This essentially to see the resilience of communities after successive transformations, and thus in some way the ability of the embedding procedure to separate communities into distinct angular sectors.

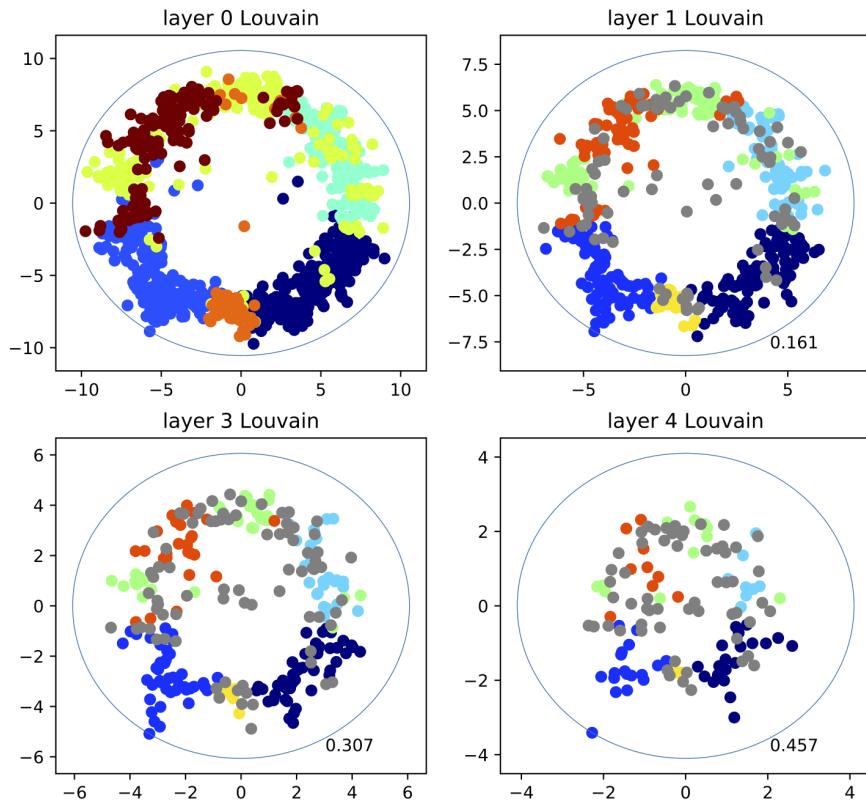


Figure 4.2: Illustration of 4 layers obtained with the GR transformation, in this analysis Louvain communities where computed at the original layer, at the downscaled layers the same partition was kept and the nodes coming from two different communities were counted (bottom right numbers are the fraction of that nodes) and colored in gray.

Limit in \mathbb{H}^2 representation degree magnitudes $\{\kappa_i\}$ are increasing in the flow, while maintaining their proportionality with observed degrees $\{k_i\}$ at each layer (fig. 4.3), as the networks remain congruent with \mathbb{S}^1 model. Nevertheless this effect leads the argument of the logarithm defining $R^{\mathbb{H}}$ 3.5 to become < 1 at a certain layer (since $N^{(l+1)} = N^{(l)}/2$ and $\mu^{(l+1)} = \mu^{(l)}/2$ and $R^{\mathbb{H},(l)} \sim \ln(1/\kappa_0^2)$), so that the radius collapse to 0 and network is not representable in \mathbb{H} model anymore.

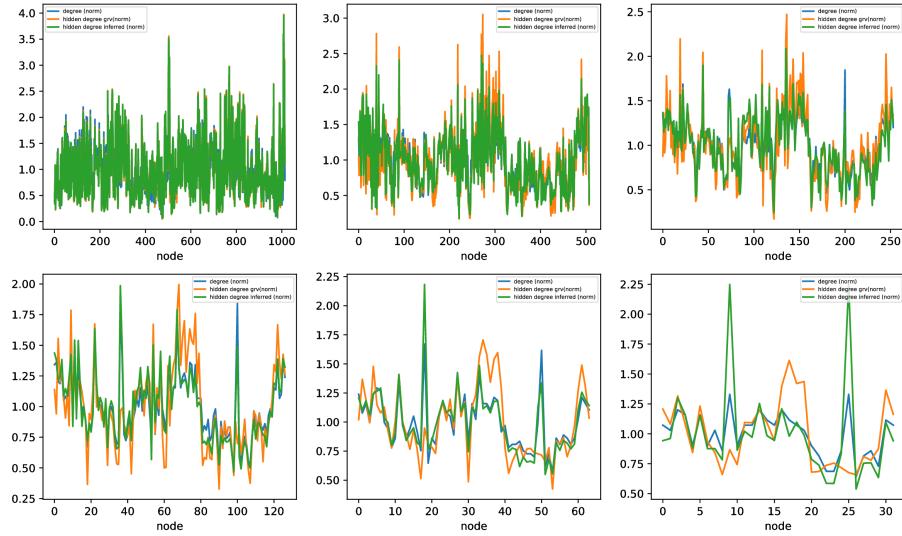


Figure 4.3: Proportionality between nodes degrees, hidden degrees coming from the transformations 4.1 and hidden degrees inferred from embedding at each layer (each divided by their average value). Up-left figure is the original most resolute layer, bottom-right the less resolute layer.

Chapter 5

Results of the multi-scale analysis

5.1 Testing the embedding algorithm

The first analysis that I conducted was testing the accuracy of the embedding applied to synthetic networks possessing different peculiar properties, which I could control. What I wanted to investigate was the eventual increased or decreased accuracy of the embedding in correspondence with certain identifiable characteristics.

The embedding of network data, apart from being an alternative data representation from which we can learn about network organisational features, can be seen as the mapping of topological instances back to the geometric latent space from which they arise. If this latent space assumption is true and geometry actually plays a role in defining topology, clustering can be explained in the network as a manifestation of the triangle inequality in the metric space. At the same time, if the geometric hyperbolic model involved in the embedding is suitable to describe the observed structure of the network, we expect the embedded nodes to generate synthetic networks accurately reproducing original network properties.

Resuming, clustering suggests the existence of hidden geometrical relations, and an accurate embedding proves that the particular geometrical relations included in the model are a plausible explanation for the observed topology.

I generated some synthetic networks (using *NetworkX* python library) from known models such as Watts-Strogatz, Barabasi-Albert and one controlling community structure¹. All of them are generated with the same number of nodes.

¹Dividing nodes into a desired number of groups and assigning a higher probability of connection intra-groups than inter-groups.

For each network, I looked at the original clustering coefficient and the one averaged over the ensemble of synthetic networks.

I also computed the average $\1 distance between connected nodes $R\Delta\theta_{ij}/|E|$, with the hypothesis that it could be an indicator of the goodness of the embedding, given that nodes that are connected in the original network should be embedded in similar angular regions.

	β	clustering original	clustering ensemble	avg. $\1 distance
random communities	4.26	0.626	0.629	0.39
Barabasi	1.01	0.142	0.233	1.29
Watts-Strogatz $p = 0.1$	2.65	0.471	0.477	0.54
Watts-Strogatz $p = 0.9$	1.01	0.094	0.178	2.15

Table 5.1: Comparison between different networks of original clustering and the one of the synthetic ensemble, average distance. Shows that clustering is not well reproduced in networks with β close to 1.

What I noticed is that the method can more efficiently reproduce networks with higher clustering (corresponding to a higher beta value), while networks with lower clustering (limiting case: $\beta \sim 1$) are embedded to networks that reproduce clustering worse (in particular it is increased).

When original clustering is higher, the average distance between connected nodes is found to be smaller, indicating that the method is recognising and efficiently mapping topological triangles into geometrical structures verifying triangular inequality.

β in brain networks included in my dataset is around 2, clustering coefficient around 0.5 and is accurately reproduced in the synthetic ensemble generated for the validation. The overall embedding results as accurate for all the subjects involved in the study, proving the fitness of the model and the applicability of the multi-scale analysis. In section 7 I report the result of validation for some networks.

5.2 Emergence of peculiar network properties

I investigated the behaviour of some global measures on pathological and non-pathological brain networks across scales and compared it to a random network with approximately the same number of nodes and edges (generated using `gnm_random_graph(n,m)` function from NetworkX), to observe the peculiarities of these systems and compare them to the literature.

I selected just one subject from each group with respectively 72664 and 72164 edges and I generated a random graph with 72164 edges and the same number of nodes.

A higher modularity (eq.1.4) in both brain networks with respect to the random and I found it higher in the brain affected by Alzheimer with respect to the healthy one.

The higher modularity in pathological conditions reflects a lack of interconnections between modules, and consequently, a deficit in the integration of information, further confirmed by a lower efficiency (eq. 1.5). This consideration is valid at all scales and notably, while efficiency difference is of order 10^{-4} at layer 0, it becomes of order 10^{-1} between networks rescaled 3 times (~ 150 nodes). The result for modularity and efficiency is shown in fig.5.1.

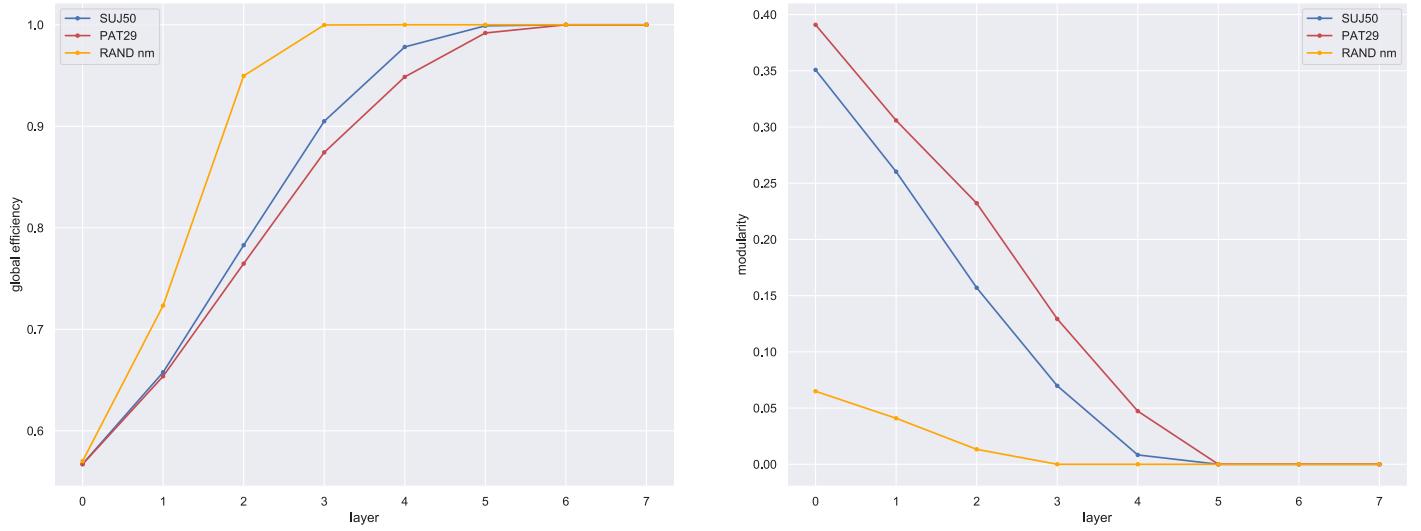
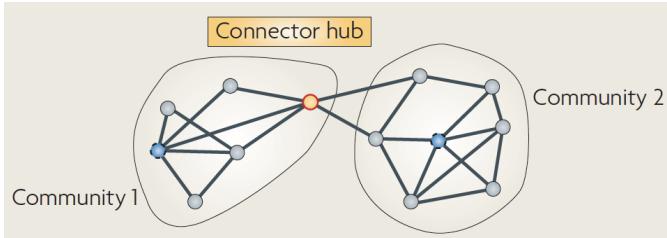


Figure 5.1: Global efficiency and modularity comparing an Alzheimer diseased patient with healthy control and a random graph across scales. Layer 0 is the network (not filtered) at the original scale, layers > 0 are rescaled using 4

The renormalisation procedure causes a general increase in network efficiency and a decrease in modularity. Which means that while many nodes belonging to the same community (placed in similar angular positions) are grouped, decreasing the dimension of communities, the connections between different communities are not cut off overall, effectively preserving the balance between integration and segregation.

This is expected considering how nodes are grouped (explained in 4) and the fact that nodes involved in linking communities (connector hubs (fig.??)) are likely connected to many nodes of both communities, maintaining the interconnection for many layers.

Figure 5.2: connector hub node illustration



I report (fig.5.3) two more results for the clustering coefficient (eq.1.3) and the fraction of supernodes mixing communities (communities computed with *Louvain* algorithm at layer 0), the interpretation of which is more subtle. Clustering is higher in the pathological subject in line with the increased segregation also reflected by the higher modularity, but the tendency is inverted at layer 2. Also clustering is steeply increasing for the random graph overcoming the two complex networks, as its lack of structure makes it rapidly tend to a completely connected graph.

The flow of the fraction of supernodes mixing communities shows that pathological networks communities are more resilient through scales, again expected from the higher segregation.

This measure show a difference, even if just visually, between the pathological and non-pathological and is related to the application of the geometric renormalisation procedure, thus not obtainable from alternative multi-scale analyses involving experimental data at different resolutions. The idea is that quantities like this can be employed in new diagnosis tests if proven to be actually conveying additional information with respect to imaging or graph theoretical analysis.

At the original scale, global efficiency in brain networks is comparable with the random, while clustering is higher, thus reflecting the expected small-world organisation of brain networks; but this feature is weakened at less resolute scales.

5.3 Multi-scale analysis of global properties

I analysed the flow of some global properties of networks across layers, to see if I could find a statistically significant difference between the two groups of data (27 healthy and 23 diseased structural brain networks), at a certain reduced scale that was not observable at the original one.

At each scale I computed **global efficiency**, **clustering**, **average degree**,

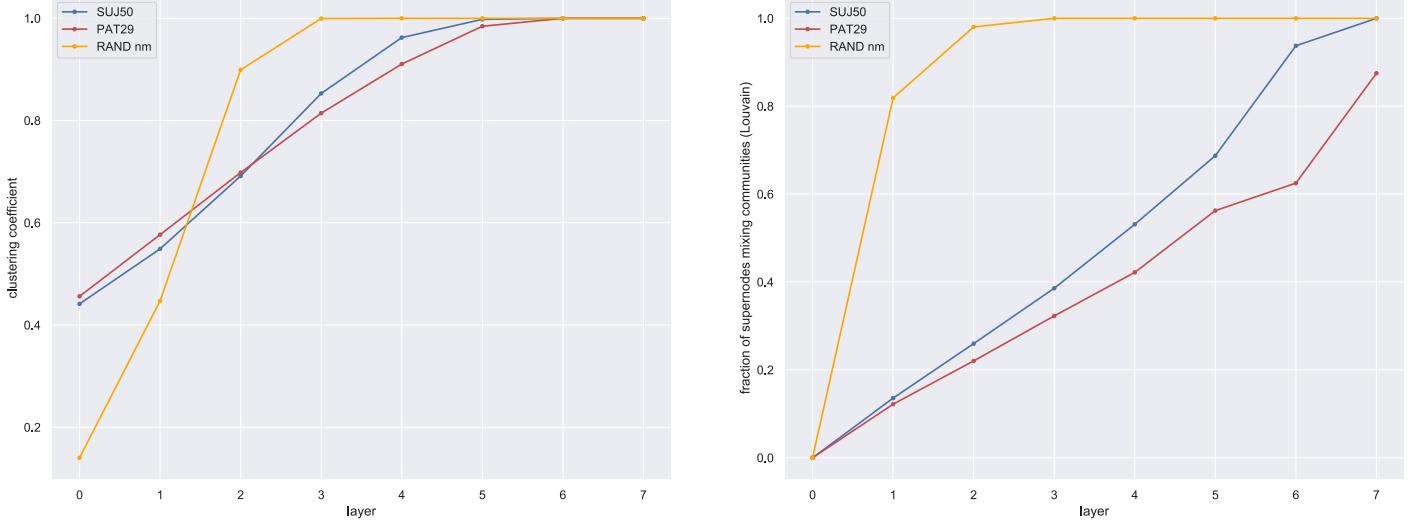


Figure 5.3: Clustering coefficient and fraction of supernodes mixing communities comparing an Alzheimer diseased patient with healthy control and a random graph across scales.

minimum degree, modularity, number of communities, fraction of supernodes mixing communities, assortativity² for each subject in the two groups, then I performed a statistical Wilcoxon rank-sums test³ between the samples obtained. This was repeated varying the strength of filtering (done as explained in subsection 2.2).

No difference is observed on non-filtered networks at any layer ($k \simeq 150$), as well as for networks filtered with *threshold proportional* (subsec.2.2.1) down to average degree $k = 60$ and $k = 30$. At $k = 20$ instead, I start to observe a distinction between the two groups, on the average degree flow and global efficiency (fig. 5.4).

Notably, the same results are not found on the network filtered to $k = 20$ using the MST backbone method (subsec.2.2.2). Instead, significant difference in modularity is found in the network using both methods, but the difference is arising already at layer 0, which is a less relevant result for the scope of testing the

²Assortativity is a measure of correlation between degrees of neighbours: the more high degree nodes are connected to other high degree nodes the more is positive.

³Non-parametric test against the alternative hypothesis of one group values being larger than the others. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ranksums.html>

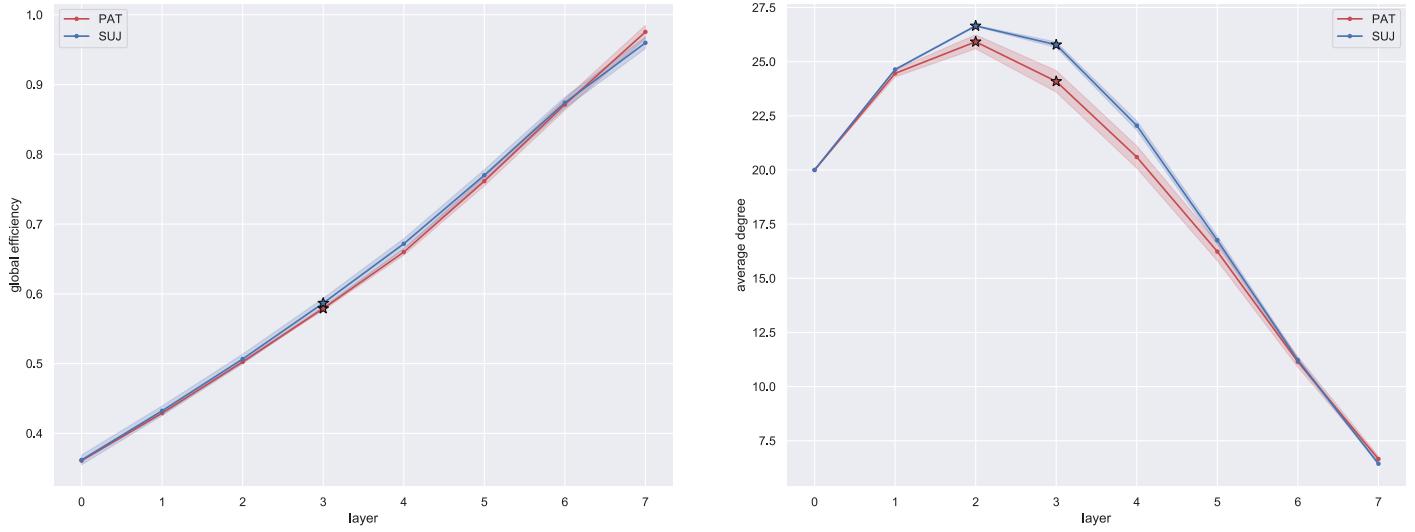


Figure 5.4: Global efficiency and average degree, statistical comparison of the 2 groups. Stars indicate a p -value < 0.05 is found at that scale. Network filtered to $k = 20$.

power of the method in unveiling differences undetectable at the original scale. A difference at layers > 0 is found for $k = 10$ and $k = 5$ (fig. 5.5), while strangely enough no difference in modularity is observed for $k = 3$ at any layer.

Apart from modularity many properties become informative for the networks filtered down to $k = 3$ (fig. 5.6).

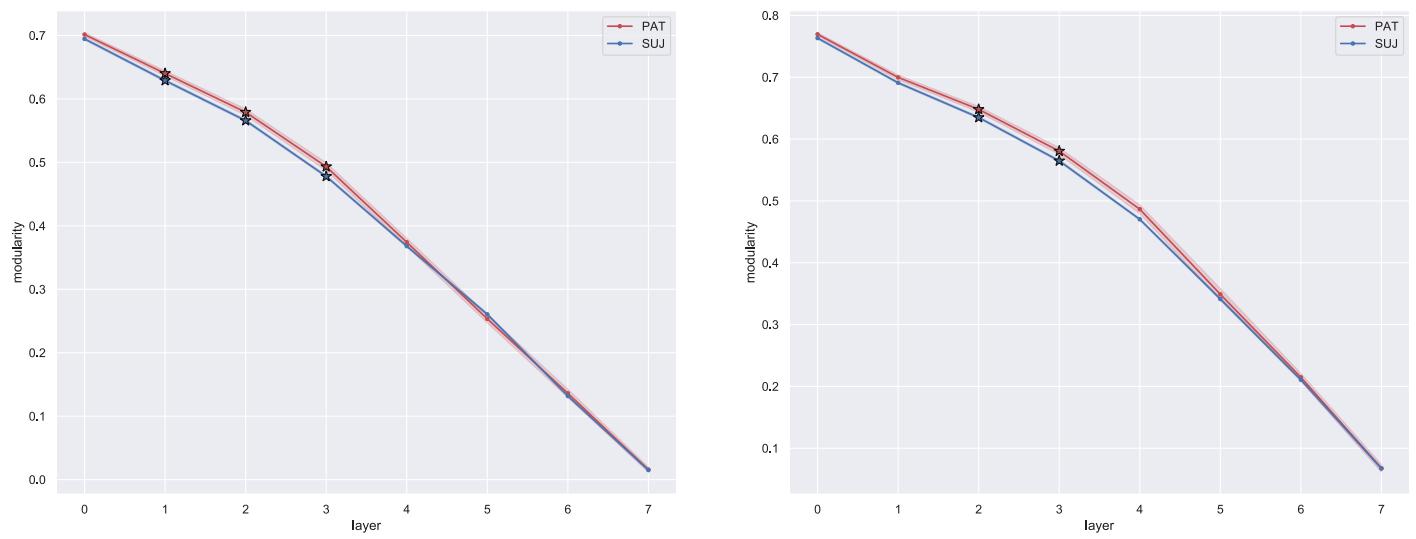


Figure 5.5: Modularity flow on networks filtered with MST method to $k = 10$ (right) and $k = 5$ (left).

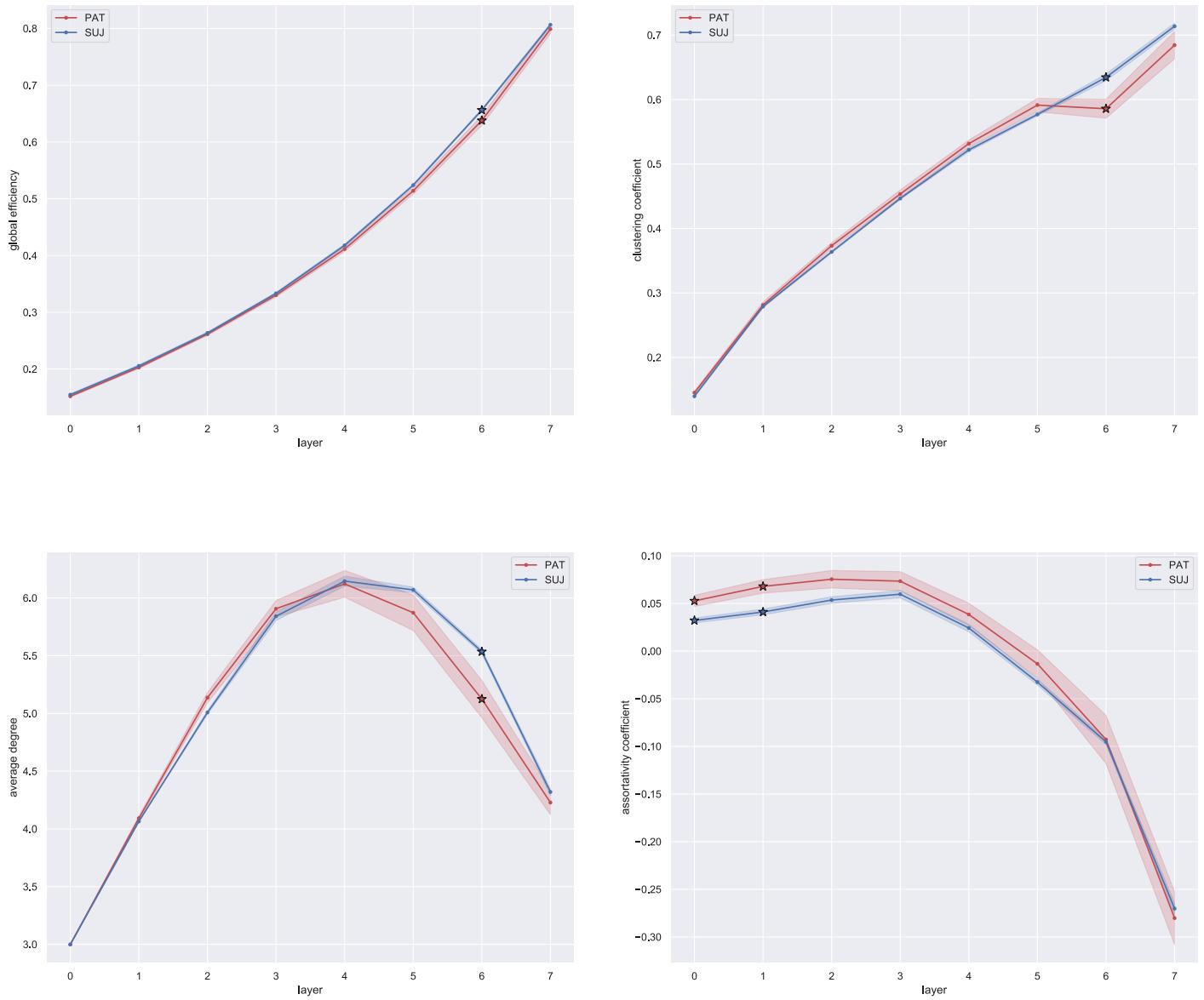


Figure 5.6: Global efficiency, clustering coefficient, average degree and assortativity flows on networks with $k = 3$ filtered with MST method.

5.4 Multi-scale analysis of coordinates' distributions

From transformed coordinates saved for each layer, I extracted **angular distance between nodes merged into supernodes** and **hidden degrees** κ_i .

At the original scale level, it is possible to uniquely associate nodes in one subject's network to others, cause they still represent the same areas of interest in the brain. But as they are grouped into supernodes following the angular ordering, this information is lost, so I decided to compare the empirical distributions of these local quantities, which is independent on nodes' labels. The analysis is again performed for various filtering strengths.

I performed a two sample Kolmogorov-Smirnov test⁴ and computed the Wasserstein distance⁵. The first tests the hypothesis that the two samples represent the same distribution, the second is a measure of the amount of work needed to move a distribution to the other.

The networks filtered with threshold proportional are not showing a difference in distributions. While the networks filtered starting from their MST exhibit a progressive distancing of distributions across layers.

The distributions of angular distance are found not equal with a p-value < 0.001 from layer 3 (~ 150 nodes) on. As for the distributions of hidden degrees the same result is found from layer 4 (~ 70 nodes). The Wasserstein distance for the angular distance distributions increases slightly, from 0.00 to 0.18 at layer 7, while the same measure for the distribution of hidden degrees increases from 0.29 to 99.74.

The same behaviour is found for stronger filtering (smaller average k), meaning k = 20 is already an informative density in this case.

I had computed also distributions of **local clustering** and **local efficiency** using functions from *bctpy*, finding no significant results. This is not in disagreement with the fact that some significance was found for the correspondent global quantities, which are just the average of the local ones; the non-parametric t-test used in that case was in fact comparing medians of the two groups and not means.

⁴https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ks_2samp.html

⁵https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wasserstein_distance.html

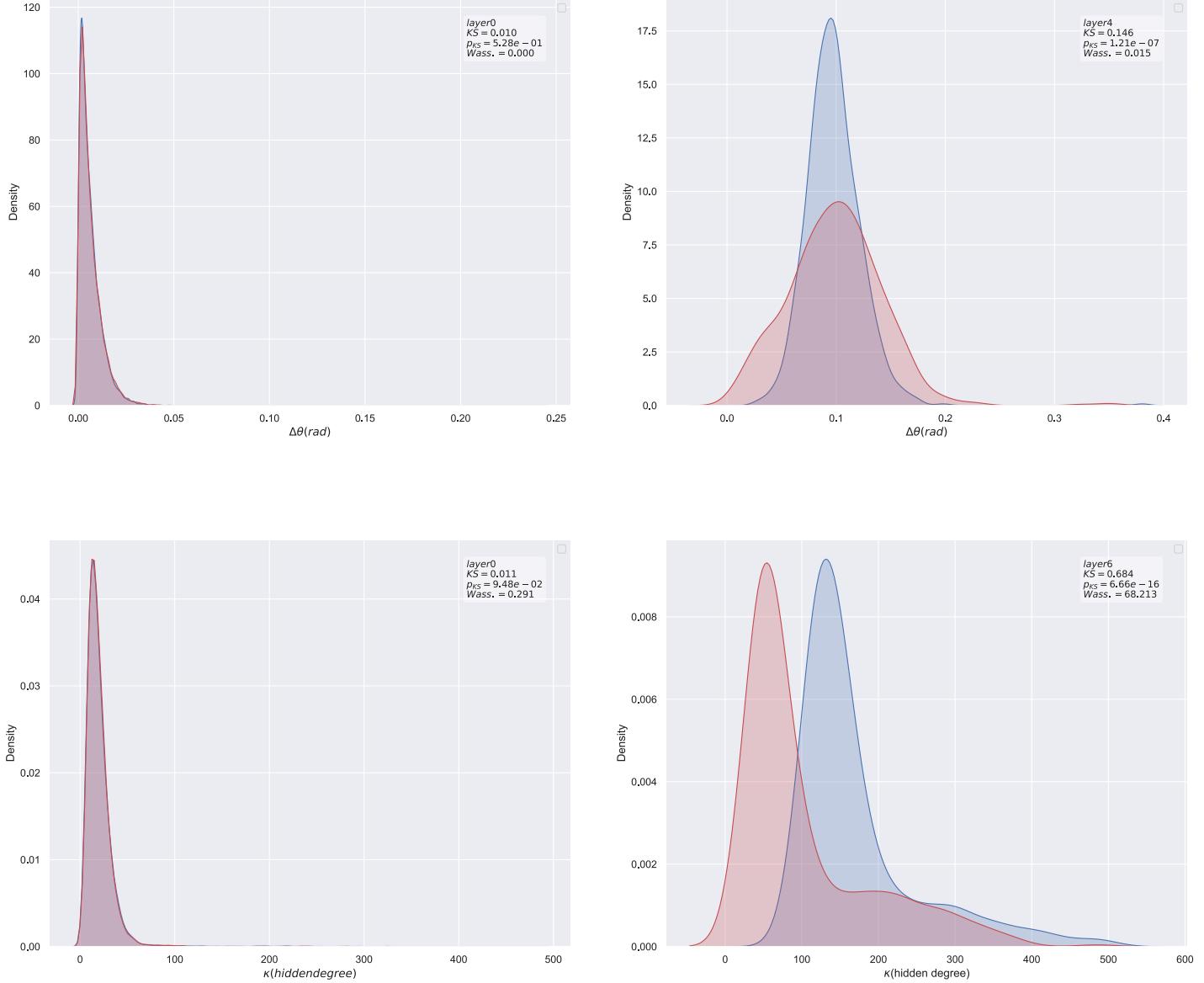


Figure 5.7: First row: angular separation of nodes composing supernodes at layer 0, and at layer 4. Second row: hidden degrees distributions comparison at layer 0 and at layer 6. Again red colour indicates the pathological group while blue the non-pathological. Networks filtered down to $k = 20$. The distributions are visualised using *kernel density estimate (KDE)* from Seaborn python library.

Chapter 6

Conclusions

Brain networks can represent the system at multiple scales, resolution can go from the neuronal level to that of neural ensembles or cortex parcels. However, different scales are usually analysed in a distinct way and a theoretical framework able to reproduce the multi-scale organisation, as well as explain how the different scales are inter-related, is still missing.

The relevance of this geometric approach is the possibility to model network scaling properties in a unified manner, supported by the evidence in the literature that it is indeed able to reproduce some important features and symmetries of the system.

I applied the method to the classification of Alzheimer diseased patients and found that, in some cases, a statistically significant difference can be found at underlying scales while not at the original scale, suggesting the possibility for a theoretical description of how this disease is affecting brain functionality at multiple, interconnected scales. Modularity and efficiency have been proven to be good measures to distinguish the two groups (as expected from literature), but also the predictive power of information encoded in hyperbolic representations has emerged from measures involving angular coordinates and hidden degrees.

Some results were obtained after network filtering but without a clear understanding of the role of filtering in the emergence of significant properties, which should be investigated further.

The filtering with the MST method has revealed a similar embedding but conveyed different results, in general, with respect to the other one, and notably was the only one able to unveil differences in coordinates' distributions; leaving the need to better understand the difference between thresholding and backbone method for filtering.

A needed follow-up would be performing the same analysis across experimentally obtained networks at different resolutions, with a comparable number of nodes

with respect to the model rescaled ones. The eventual correspondence of the results obtained would be in favour of the ability of this multi-scale modelisation to reproduce observed network properties; the opposite case could lead to the conclusion that the proposed method conveys different or more information with respect to the observations, immediately linked to the necessity of understanding from where this new information comes from.

Some perspectives would be to apply this method to functional brain networks¹, or include dynamical features.

Indeed brains are intrinsically dynamical systems: functional brain networks are rapidly changing in response to environmental stimuli; but also structural brain networks, even though stable at short time scales (seconds/minutes), are indeed altered over longer time scales by plasticity effects (synapses strength change, remapping of cognitive functions), ageing, injury and pathology.

¹Network in which edges represent correlations between signals measured in the areas of interest (becoming nodes), in resting state or performing a certain task.

Chapter 7

Some validations of embedding plots

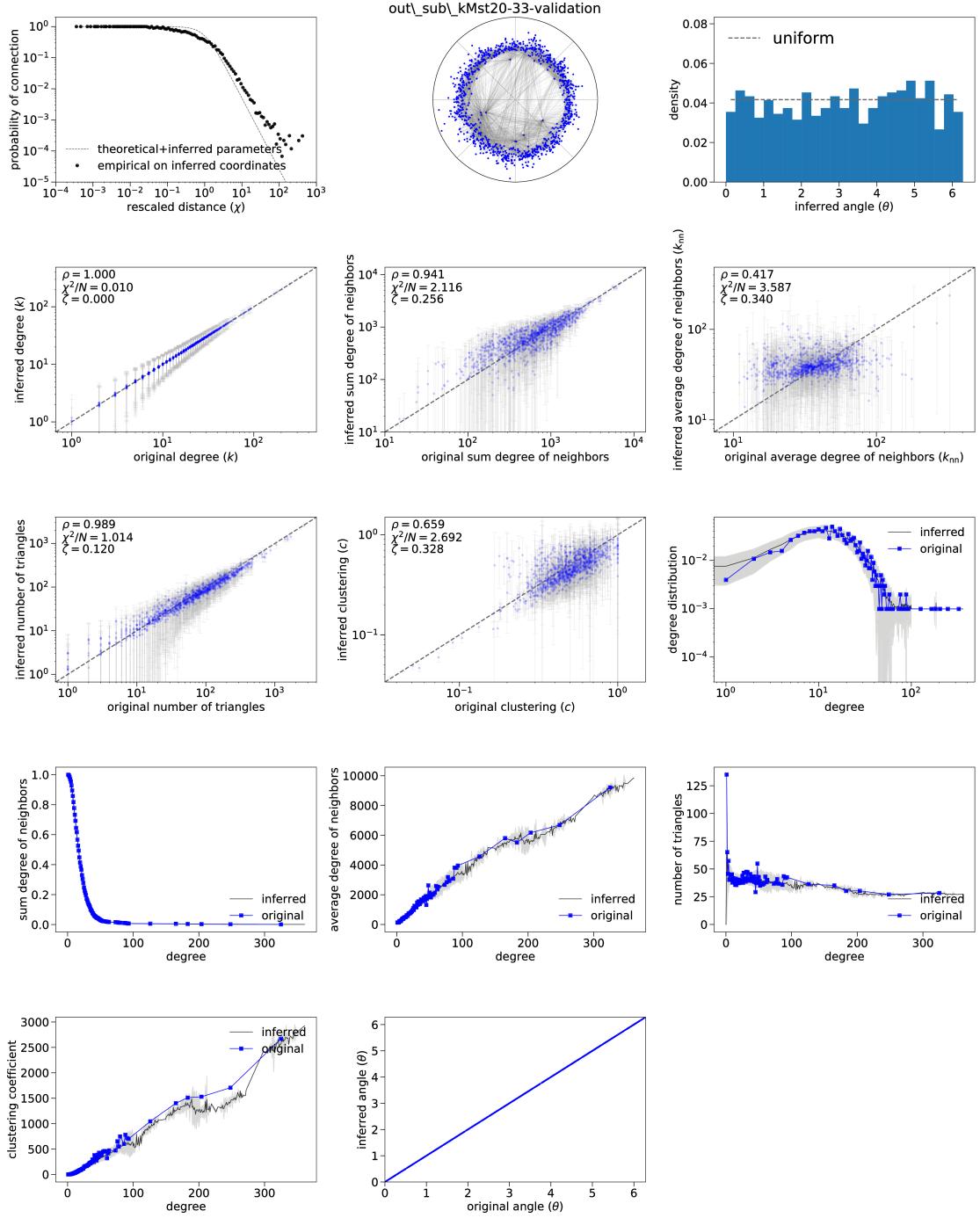


Figure 7.1: Validation of embedding for non-pathological subject number 33. Filtered with MST backbone method. 2.2 up to average degree 20

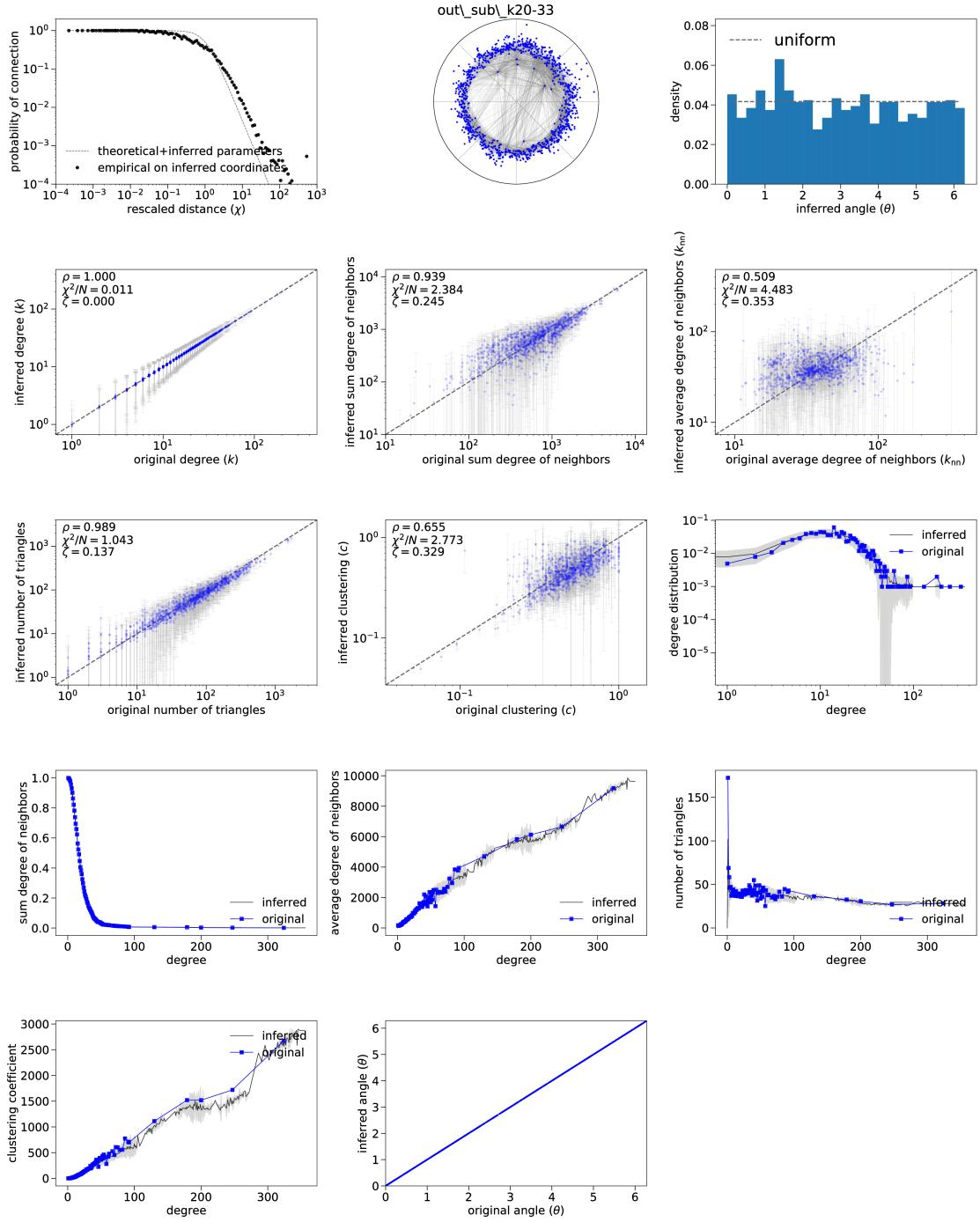


Figure 7.2: Validation of embedding for non-pathological subject number 33. Filtered with threshold method. 2.2 up to average degree 20

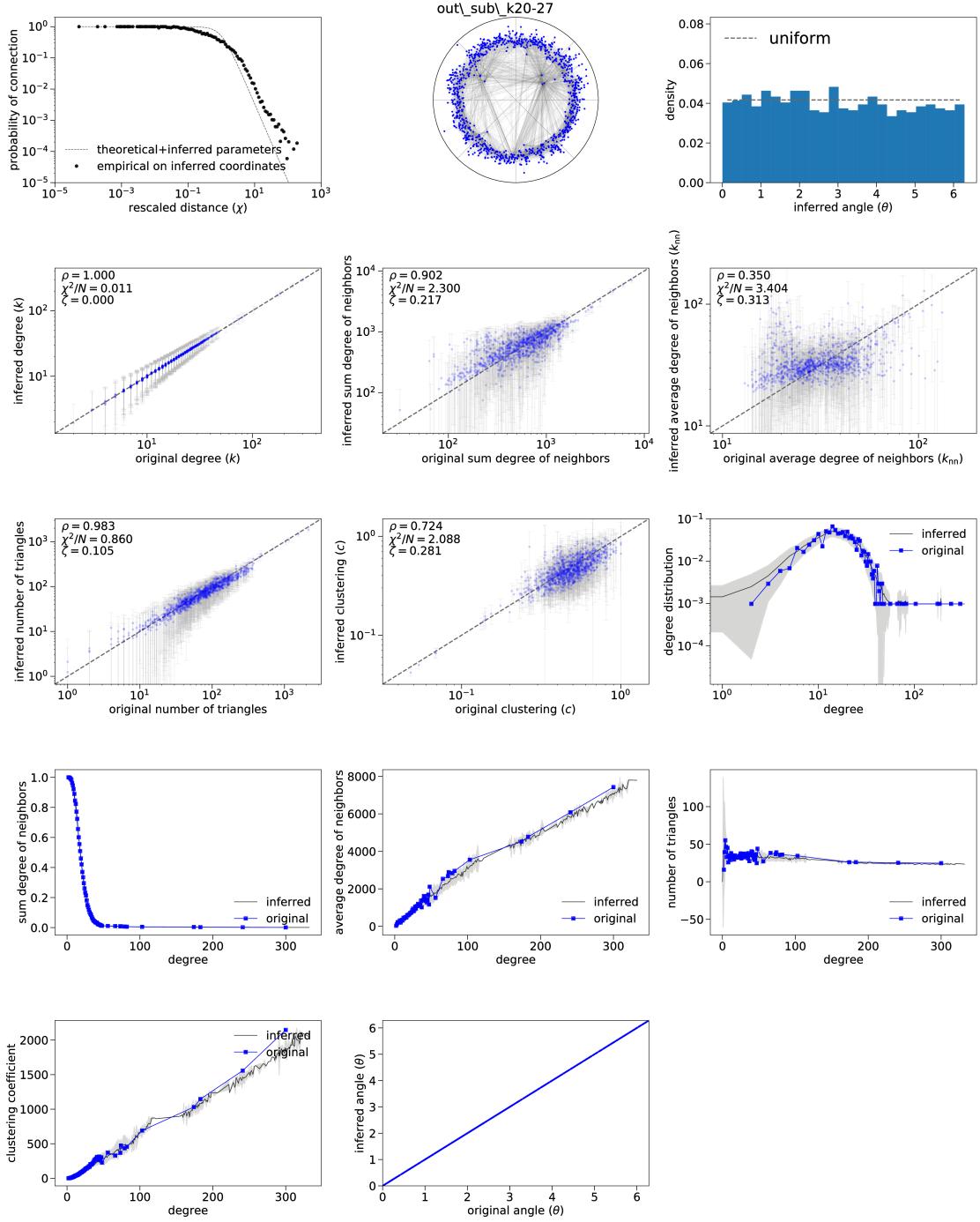


Figure 7.3: Validation of embedding for pathological subject number 27. Filtered with threshold method. 2.2 up to average degree 20

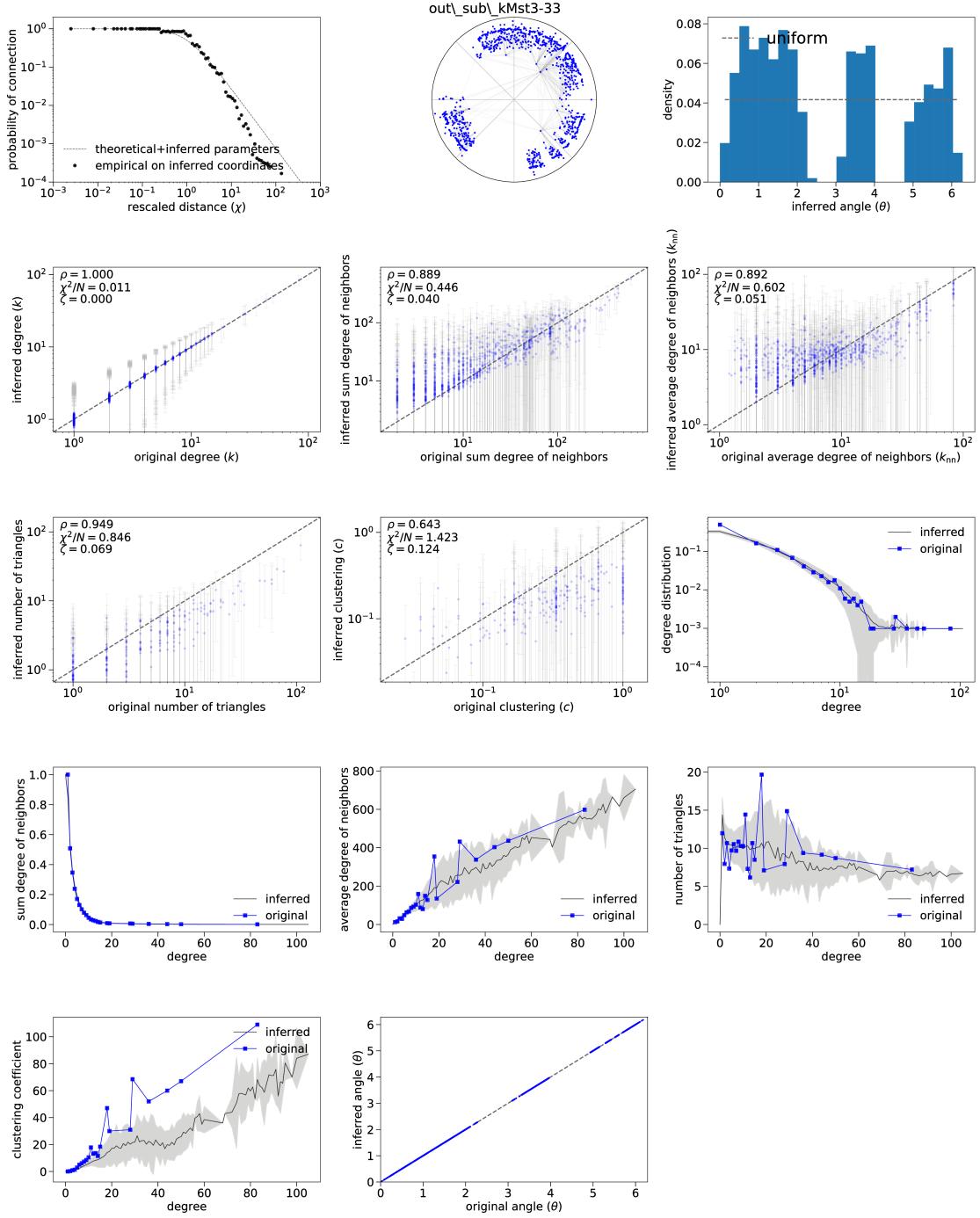


Figure 7.4: Validation of embedding for non-pathological subject number 33. Filtered with MST backbone method. 2.2 up to average degree 3

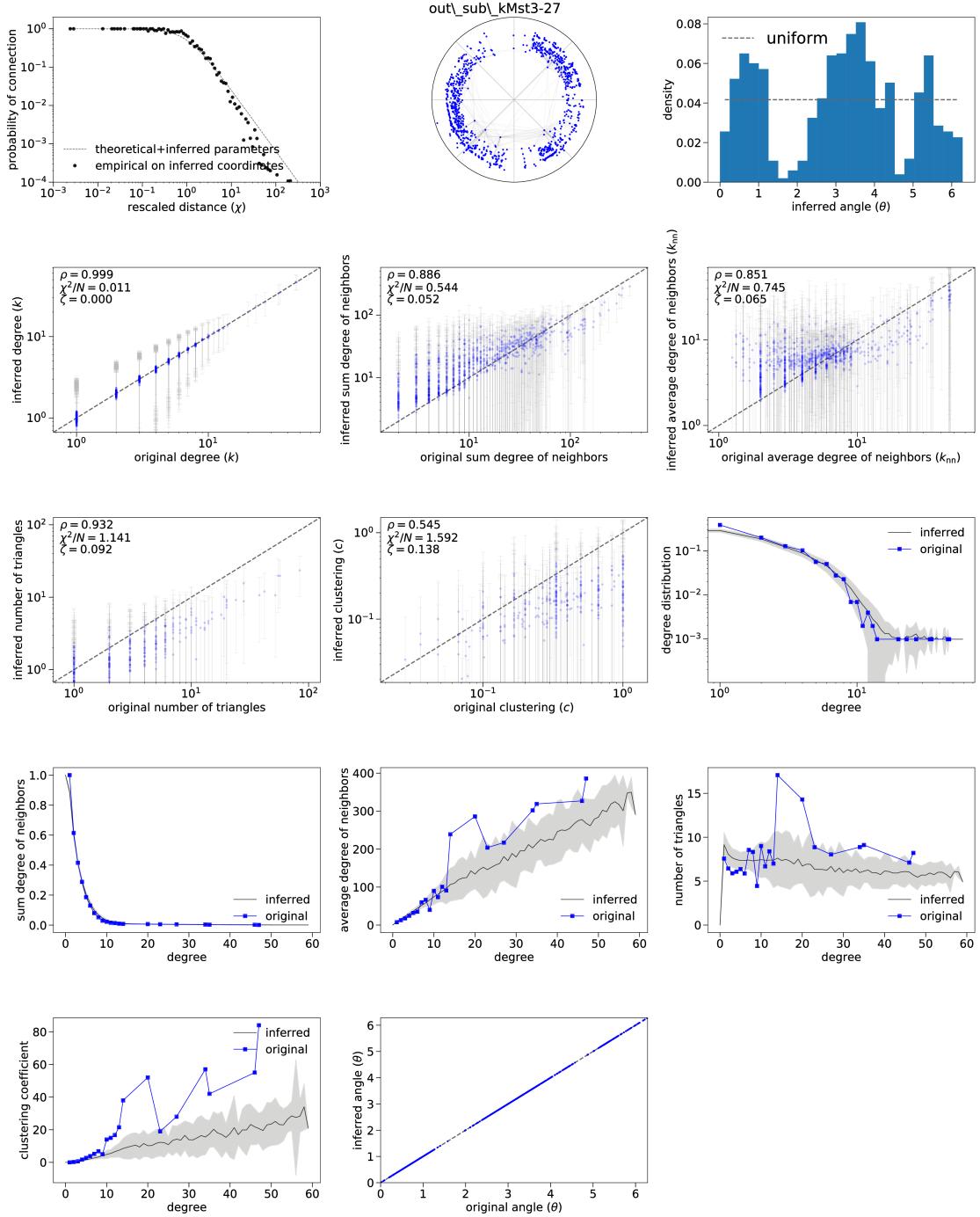


Figure 7.5: Validation of embedding for pathological subject number 27. Filtered with MST backbone method. 2.2 up to average degree 3

Bibliography

- [1] Bassett D.S. Lynn C.W. «The physics of brain network structure, function and control.» In: *Nat Rev Phys* 1, 318–332 (2019) (cit. on pp. ii, 3).
- [2] Lin Q et Al. Dai Z. «Disrupted structural and functional brain networks in Alzheimer’s disease.» In: *Neurobiol Aging* 75:71-82. (2019) (cit. on p. ii).
- [3] Kitsak M et al. Krioukov D Papadopoulos F. «Hyperbolic geometry of complex networks.» In: *Phys Rev E.* 2010; 82:036106. (2010) (cit. on pp. ii, iii, 10–12, 14, 20).
- [4] Narula V. et al. Cacciola A. Muscoloni A. «Coalescent embedding in the hyperbolic space unsupervisedly discloses the hidden geometry of the brain.» In: *arXiv preprint arXiv:1705.04192.* (2017) (cit. on pp. ii, 15).
- [5] Boguñá M. & Serrano M.Á. García-Pérez G. «Multiscale unfolding of real networks by geometric renormalization.» In: *Nature Phys* 14, 583–589 (2018) (cit. on pp. ii, iii, 20, 21).
- [6] Guillermo García-Pérez et al. «Mercator: uncovering faithful hyperbolic embeddings of complex networks.» In: *New J. Phys.* 21 123033 (2019). URL: <https://github.com/networkgeometry/mercator/blob/master/README.md#python-module> (cit. on pp. iii, 16).
- [7] Watts D.J. and Strogatz S.H. «Collective Dynamics of “Small-World” Networks.» In: *Nature,* 393, 440-442 (1998) (cit. on p. 1).
- [8] Sporns O. Bullmore E. «The economy of brain network organization.» In: *Nat Rev Neurosci* 13, 336–349 (2012) (cit. on p. 3).
- [9] Sporns O. Bassett D.S. «Network neuroscience.» In: *Nat Neurosci* 20, 353–364 (2017) (cit. on p. 3).
- [10] Sporns O. Bullmore E. «Complex brain networks: graph theoretical analysis of structural and functional systems.» In: *Nat Rev Neurosci* 10, 186–198 (2009) (cit. on p. 4).
- [11] Descoteaux M et al. Jeurissen B. «Diffusion MRI fiber tractography of the brain.» In: *NMR Biomed.* 2019 Apr;32(4):e3785. (2017) (cit. on p. 7).

- [12] Misic B & Sporns O Avena-Koenigsberger A. «Communication dynamics in complex brain networks.» In: *Nat Rev Neurosci* 19, 17–33 (2018) (cit. on p. 7).
- [13] Cammoun L et al. Hagmann P. «Mapping the Structural Core of Human Cerebral Cortex.» In: *PLOS Biology* 6(7): e159. (2008) (cit. on p. 7).
- [14] Muhua Zheng et al. «Geometric renormalization unravels self-similarity of the multiscale human connectome.» In: *PNAS*, 117 (33) 20244-20253 (2020) (cit. on p. 10).
- [15] Papadopoulos F. & Krioukov D. Boguñá M. «Sustaining the Internet with hyperbolic mapping.» In: *Nat Commun* 1, 62 (2010) (cit. on p. 10).
- [16] Maksim Kitsak, Ivan Voitalov, and Dmitri Krioukov. «Link prediction with hyperbolic geometry». In: *Phys. Rev. Research* 2 (4 Oct. 2020), p. 043113 (cit. on p. 10).
- [17] Krioukov Dmitri Serrano M.Angeles and Boguñá Marián. «Self-Similarity of Complex Networks and Hidden Metric Spaces». In: *Phys. Rev. Lett.* 100 (7 2008) (cit. on p. 13).
- [18] De Domenico M. et al. Boguñá M. Bonamassa I. «Network geometry.» In: *Nat Rev Phys* 3, 114–135 (2021) (cit. on p. 14).
- [19] Ciucci S. et al. Muscoloni A. Thomas J.M. «Machine learning meets complex networks via coalescent embedding in the hyperbolic space.» In: *Nat Commun* 8, 1615 (2017) (cit. on pp. 15, 16).
- [20] Allen Q Ye et Al. «The intrinsic geometry of the human brain connectome.» In: *Brain Inform.* 2(4):197-210. (2015) (cit. on p. 15).
- [21] Serrano MÁ Allard A. «Navigable maps of structural brain networks across species.» In: *PLoS Comput Biol* 16(2): e1007584 (2020) (cit. on p. 16).
- [22] Serrano M. et al. Papadopoulos F. Kitsak M. «Popularity versus similarity in growing networks.» In: *Nature* 489, 537–540 (2012) (cit. on p. 16).
- [23] Fragkiskos Papadopoulos, Rodrigo Aldecoa, and Dmitri Krioukov. «Network geometry inference using common neighbors». In: *Phys. Rev. E* 92 (2 2015) (cit. on p. 16).
- [24] Carlo Vittorio Cannistraci et Al. «Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding.» In: *Bioinformatics, Volume 29, Issue 13* (2013) (cit. on p. 16).
- [25] Thomas Aynaud. *python-louvain x.y: Louvain algorithm for community detection*. 2020. URL: <https://github.com/taynaud/python-louvain> (cit. on p. 22).