

Graph-based Methods for Brain Structural Connectivity Analysis

Nemo Fournier

Supervised by Pierre Maurel and Julie Coloigner

Abstract

I studied brain connectivity analysis, with an emphasis on graph-based methods. I adapted a recent measure to compare connectivity networks across subjects, based on graph signal processing and optimal transport ideas. I extended it by introducing a new way to leverage information from a whole cohort of subjects: barycenter computation. I also studied another tool, graph curvature, for which I proposed an adaptation to our context.

Contents

Introduction	1
1 Connectivity	1
1.1 Functional Connectivity and Structural Connectivity	2
1.2 Why Study Brain Connectivity?	3
1.3 State of the Art of Connectome Comparison	4
2 Graph Signal Processing and Optimal Transport	5
2.1 Graph Signals	5
2.1.1 Deriving Tools for the Analysis of Signals on Graphs . . .	6
2.1.2 Spectral Analysis of the Graph Laplacian	7
2.2 Optimal Transport of Smooth Graph Signals	9
2.3 Wasserstein Distance on our Connectivity Matrices	10
2.4 Barycenters of Gaussian distributions	11
2.5 Barycenters for Connectivity Analysis	13
3 Graph Curvature	14
3.1 From Riemannian manifolds to networks	15
3.1.1 Ricci curvature on manifolds	15
3.1.2 Ollivier-Ricci Curvature in Metric Space	17
3.2 Ollivier-Ricci Curvature on Connectivity Graphs	18
Conclusion	20
A Introduction on Optimal Transport	21
B Remerciements et Contexte Institutionnel et Social du Stage	24

Introduction

Analysing **brain connectivity** aims at studying how each region of the brain relates to each other. In the last decades, the improvement of non-invasive acquisition techniques has significantly broadened our capability to capture fine details about either physical or functional connections in the brain. Their diffusion among research centres has made connectivity analysis one of the spearheads of neuroscience studies. There is good hope that a better understanding of the brain's behaviour can come out of it.

I will present the different notions of connectivity in section 1, as well as their uses and the challenges arising from their analysis. The aim of my work is to develop a new metric to **compare** connectivity across subjects.

To this end, I will introduce the framework of **graph signal processing** in section 2. It has already found promising applications in the wider field of complex network analysis. We adapted a recently proposed metric to our setting, combining ideas from graph signal processing and the theory of **optimal transportation** (I introduce the tools of optimal transport as needed along the document, but a more comprehensive introduction is provided in appendix A for the interested reader). I proposed an extension of this metric, by exploiting a notion of barycenter which allows for comparisons of cohorts instead of individuals.

While adapting and devising those new methods, I apply them to data acquired for the study of depression. This data was collected in Rennes from patients suffering from mood depressive disorder to investigate the long term effects of the pathology on the brain as well as bio-markers of clinical aspects of the disease, such as drug-resistance. In this report, I will refer to **patients** for subjects suffering from depression, and **controls** for the healthy group.

I then studied another notion, **graph curvature**, that provides a new way to unveil the structural information of a graph. I describe its foundations in section 3, and I discuss how we could adapt it in our context. I present a new idea for its computation using a **diffusion distance**, as well as some preliminary results obtained with the aforementioned data.

Most of the implementations that I produced during this internship are available online ¹. I have also had the opportunity to contribute to the Python Optimal Transport library, that I used for some of my implementations, and I have been part of a paper submission to JMLR as a contributor of this library.

1 Connectivity

Studying human **brain connectivity** is all about understanding how the brain processes information. The physiology of the brain's microstructure, neurons interconnected via axons and communicating through synaptic action potentials, is well understood at the scale of individual cells. Yet the organ's sophistication (tens of billions of neurons, about 10^{14} synapses) prohibits its study at the scale of individual neurons, especially when it comes to explaining its complex behaviours.

We can nevertheless consider studying it as a whole system by relying on the hypothesis that the brain's organization at a macro-scale (where atomic elements are populations of neurons instead of individual cells) should allow for an overall

¹<https://github.com/little-nem/brain-connectivity-analysis>

efficient processing of information and that observing it at this level can still convey meaningful insights.

A landmark paper in this field is an article of Sporns et al [STK05], in which they coin the term **connectomic** for the study of the physical organization of the brain. It has been brought out that the macro-scale is indeed a relevant scale. Similarly to other complex-networks, such as social or biological networks, properties of information segregation (segmentation of the brain into highly specialised areas) and integration (with a global communication scheme between those specialised areas), as well as small-worldness or resilience have been described. This network is referred to as the **human connectome**.

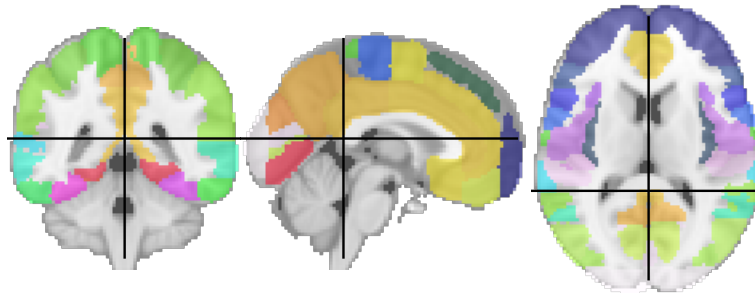


Figure 1: Harvard-Oxford Atlas, showing how the brain can be segmented into specialised areas.

Brain connectivity analysis can be divided into two main subdomains whose distinction lies in the nature of the studied network. On the one hand, there is the study of **functional connectivity**, which involves measuring the activity of the brain and understanding how different areas interact with each other. This can especially provide insights about how the brain behaves while performing some given cognitive task. On the other hand, **structural connectivity** can also be investigated. Here, we are heedful of the physical organization of the brain, which we infer from the observation of brain matter.

In both cases, we aim at defining a notion of connectivity in the brain, that accounts for how much each region of the brain interacts or is linked with each other. I will introduce in the next paragraphs the main ideas behind data acquisition in both contexts.

1.1 Functional Connectivity and Structural Connectivity

Functional connectivity. The main tool used to measure functional connectivity is Functional Magnetic Resonance Imaging (fMRI). This is an imaging technique introduced in the 1990s aiming at measuring brain activity. In its most widespread form, it exploits the fact that active neurons need oxygen, and will thus be supplied by a flow of oxygen-saturated blood. This exchange of oxygen induces perturbations in the magnetic properties of the tissues, which we can measure and localize both in space and time. An fMRI session thus results in a collection of time series, each associated with the absolute activity of a confined geographic region of the brain. One of the possible interpretations of the data comes from correlations between the measured activity of different parts of the brain. This allows defining the functional connectivity of the brain, describing a network in which parts of the brain with correlated activities are deemed connected.

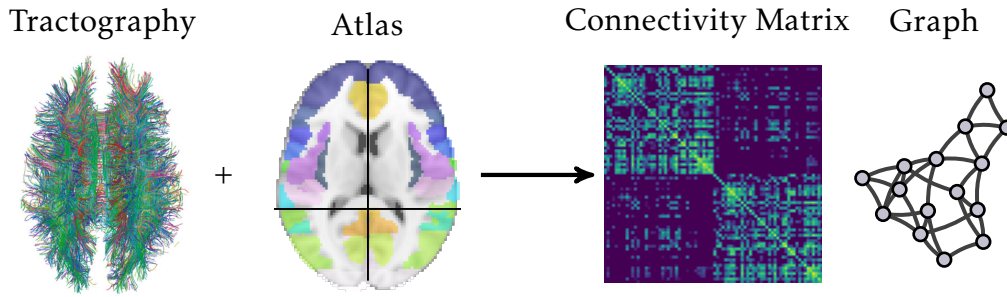


Figure 2: Classical structural connectivity pipeline. A tractography is acquired through diffusion MRI. We use an atlas (in our study the Freesurfer atlas, defining 80 regions) to compute a connectivity matrix: each entry is the number of fibers (obtained from the tractography) between two of the regions defined by the atlas. We can also compute the associated connectivity graph, which is such that its adjacency matrix is the connectivity matrix. (Tractography picture by Xavier Gigandet)

Structural connectivity. It is possible to unveil the physical structure of the brain through Diffusion Magnetic Resonance Imaging. In order to establish a map of the physical connections channels, this technique involves measuring diffusion patterns of water molecules in the brain. Those indirectly reveal the structure of brain fibers, since water molecules are more likely to diffuse following those fibers. This leads to the computation of a **tractogram** from which we can derive a network-like structure of the brain, where physical fiber connects the areas they go through. The pipeline to extract the networks we will be working with is visually represented in figure 2.

1.2 Why Study Brain Connectivity?

Brain connectivity analysis has already provided some promising results, showing that the instruments used to infer brain networks from actual subjects and to capture fine details about the brain’s organization. Here are some active research issues for which brain connectivity analysis has provided insights:

- Identify structures responsible for given cognitive functions and establish a “map” of the human connectome. See for example the Human Connectome Project (HCP) [VESB⁺13].
- Identify alterations in the connectome that could be markers of pathologies. For instance modifications of connectome topology have been observed in schizophrenia ([BBV⁺08]) or depression ([CBC⁺18]). Studies are now more focused on smaller bio-markers, *e.g.* drug-resistance, phenotype of the pathology...
- Record changes in the connectome along time, caused by ageing, a pathology, or a treatment. For instance, [DNC⁺10] uses support vector regressions to compute a “maturation index” from the strength of connections in the connectome, which can be used to predict the age of a subject. More recently, [WDC⁺19] used deep learning approaches to achieve a similar goal.
- Perform subject identification using connectivity maps as fingerprints, as is done in [VJP19] using nearest-neighbours predictors in a geodesic subspace of matrices.

- Map functional and structural connectivity, *i.e.* understand how both of them actually relate and if we can infer one notion of connectivity from the other. See for example [HK13] or [MTH⁺16].

A key element of those analyses is the ability to **compare** connectivity networks between subjects or across time (in the case of a longitudinal study), and explicit the meaningful differences between connectomes. This can help to account for individual variability among subjects, reveal differences between two cohorts, etc.

1.3 State of the Art of Connectome Comparison

The representation of connectomes through their connectivity matrix is often encountered in articles because it is especially efficient for visual comparison of connectomes: [ABR⁺13] shows that information is harder to extract visually from connectivity graph representations. Multiple studies have developed principles to compare connectivity matrices. Standard matricial norms have been used, such as the l_2 difference norm [PMPP⁺17] or the Pearson correlation between vectorized matrices [FSS⁺15]. More refined approaches exploit algebraic properties of connectivity maps to derive relevant notions of distance. For instance, in [VJP19], functional connectivity maps are proved to live in a submanifold of $\mathbf{R}^{n \times n}$ in which a geodesic distance can be defined and provides more reliable results than distances defined in the full matrix space. The ideas behind this approach are summed up in figure 3.

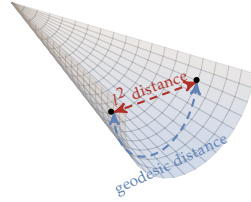


Figure 3: Functional connectivity matrices can be shown to live in conic submanifold of $\mathbf{R}^{n \times n}$. Distances can be computed by taking this structure into account, as in [VJP19]

Exploiting the duality between a connectivity matrix and the associated network, a trend in connectivity analysis is to embrace complex network inspired measures. This often involves computing several topological properties of the graph, ranging from individual node strengths to clustering coefficients as well as average shortest path lengths. An in-depth review of those markers can be found in [RS10], and I present some of them in figure 4. Most of those measures can find meaningful interpretations in terms of local network resilience, modularity, etc. They take part for example to [CBC⁺18] or [MWH19] where they are used to identify differences between cohorts.

Yet, those measures are not always fully satisfying. First of all, they often only depend on the local topology of the graph and fail to highlight more global properties. Some of those also put too much emphasis on the combinatorial structure of the network while disregarding the weights of the edges. It is shown in [XXG⁺19] that leveraging the information of the weights of edges could increase the reliability of complex network measures.

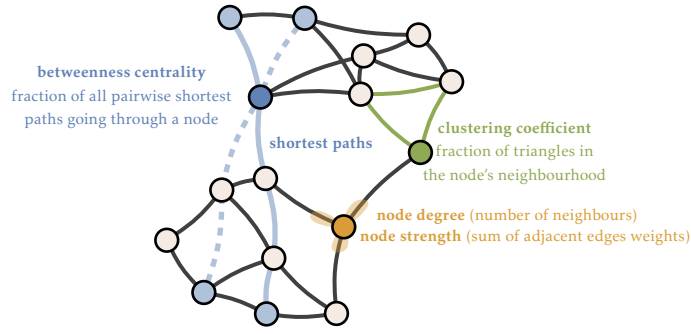


Figure 4: Some topologic markers used in complex network analysis.

This is the issue that we wanted to address in this work. We tried to devise new metrics and principles to meaningfully compare the global structure of connectivity graphs.

2 Graph Signal Processing and Optimal Transport

Signal processing is usually about the study of temporal signals, or time series: sequences of data points that are related through temporal dependencies. Each point follows the point measured just before it and is followed by the point measured right after. **Graph signal processing** is an extension of this framework, in which relations between data points are not limited to be temporal but are indeed more general. Interestingly enough, the tools introduced in this framework can provide deep insights on structural properties of graphs as well as new tools for data-processing in general. One can see for instance [RBT⁺19] for an introduction on the implication of this topic.

2.1 Graph Signals

Given a graph \mathcal{G} (a set V of nodes and a set E of — possibly weighted — edges between those nodes), a signal s on \mathcal{G} is simply a function that maps each node v to a value $s(v)$.

It is interesting to see that a time series is a particular graph signal, for which the support graph is a path where each node corresponds to a time sample. This is illustrated in figure 5.

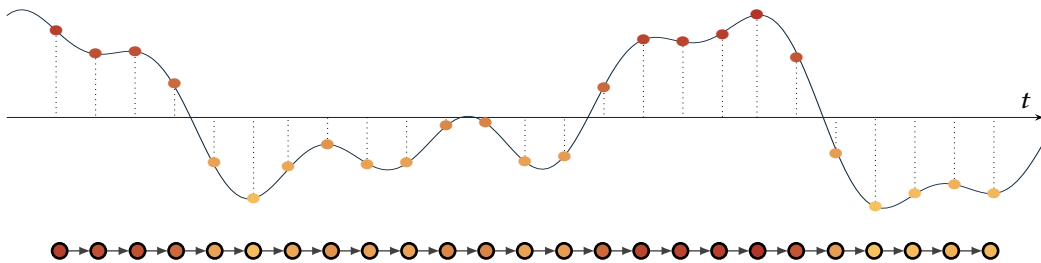


Figure 5: A time serie and its graph counterpart. The value of the signal on the graph is represented using colors on nodes, yellow being a low value, and red a high one

2.1.1 Deriving Tools for the Analysis of Signals on Graphs

It will be convenient to adopt a matricial perspective on the different objects that we manipulate. We can see a set of nodes $V = \{v_1, \dots, v_n\}$ as the standard basis of the space \mathbf{R}^n . A signal on a graph can thus be expressed as a vector $s = (s_1 \dots s_n)^T$ of \mathbf{R}^n , such that its coordinates are the value of the signal at each node. The adjacency matrix A of the graph is the matrix such that $A_{i,j} = 1$ if i and j are linked by an edge. The set of edges $E = \{e_1, \dots, e_p\}$ can also be seen as the standard basis of \mathbf{R}^p . Those objects are illustrated in figure 6.

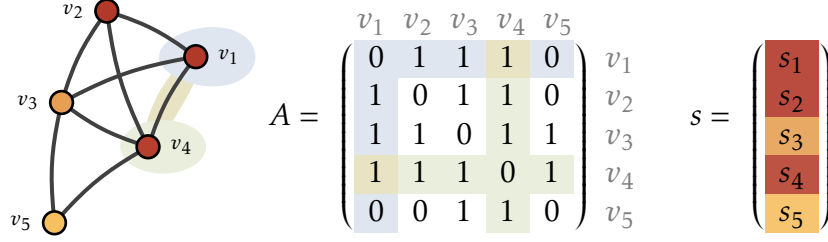


Figure 6: Graph with a signal defined on it (represented by the color on the nodes, red for a high value and yellow for a low one), its adjacency matrix A , and the signal vector s .

At this point, we will suppose that the graph is oriented, which means that each edge (i, j) leaves a vertex (i) to enter the other (j) . We will see later that the orientation does not matter for what we are interested in, but it eases the intuition behind the next objects we will introduce. To account for this orientation, we introduce the **incidence matrix** of the graph. It is the matrix of the application from \mathbf{R}^p to \mathbf{R}^n expressed in the bases E and V that maps e_t to $s_j - s_i$ when e_t is the edge (s_i, s_j) .

To follow the analogy with classical signal processing, we want to derive tools to describe and study the properties of a graph signal. A first notion that we can express is the notion of variation of a signal among its domain. The gradient of a signal over an edge $e = (i, j)$ is defined as the difference of the signal values at the extremities of the edge, following the orientation of the graph.

$$\text{grad}|_e(s) = s(j) - s(i) \quad (1)$$

We naturally extend this to the linear mapping from \mathbf{R}^n to \mathbf{R}^p that maps a signal to the vector of gradients along each edge. We denote ∇ its matrix expressed in the bases V and E , such that

$$\nabla \cdot s = \begin{bmatrix} \text{grad}|_{e_1}(s) \\ \text{grad}|_{e_2}(s) \\ \vdots \\ \text{grad}|_{e_p}(s) \end{bmatrix}$$

It is worth noting that this operator ∇ is nothing but the transposed incidence matrix of the graph.

Now that we have defined the gradient of the signal along its edges, we can express a notion of local variation at a node. The tool that we derive is the **graph Laplacian**. To pursue the analogy with classical vector calculus, the Laplacian of the signal at a node v is the divergence of the gradient through this point. Said

differently, in our setting, it is the “oriented sum” of the variations of the signal from this point toward each direction. Once again, the graph’s orientation is followed.

$$\text{laplacian}|_v(s) = \text{div}|_v(\nabla \cdot s) = \sum_{e \text{ enters } v} \text{grad}|_e(s) - \sum_{e \text{ exits } v} \text{grad}|_e(s) \quad (2)$$

This nodewise operation extends to a linear mapping from \mathbf{R}^n (the signal) to \mathbf{R}^n (the Laplacian of the signal at each node of the graph). From equation (2), we see that computing the divergence of the gradient over the edges is equivalent to applying the incidence matrix to the gradient field $\nabla \cdot s$. This incidence matrix equals ∇^\top . This is consistent with the classical calculus analogy where the divergence operator is the transpose of the gradient operator. We can thus express a Laplacian operator L of the form $L = \nabla^\top \nabla$ such that

$$L \cdot s = (\nabla^\top \nabla) \cdot s = \nabla^\top (\nabla \cdot s) = \begin{bmatrix} \text{laplacian}|_{v_1}(s) \\ \text{laplacian}|_{v_2}(s) \\ \vdots \\ \text{laplacian}|_{v_n}(s) \end{bmatrix}$$

Hence, the effect of the orientation of the graph cancels out because of the factorisation $\nabla^\top \nabla$ and plays no role in the definition of the Laplacian operator. L can thus also be defined for non-oriented graphs: injecting equation (1) into equation (2), it follows that L satisfies the following relations, independent of orientation:

$$(L \cdot s)|_{v_i} = \sum_{v_j \text{ connected to } v_i} (s(i) - s(j)) \quad L = D - A$$

Where D is the diagonal matrix whose entries are the degree of each node, and A is the aforementioned adjacency matrix. This result shows that the Laplacian matrix encodes structural properties of the graph. It also fully characterizes the graph, up to a reordering of its nodes (indeed those definitions initially depended upon the choice of the standard basis $V = \{v_1, \dots, v_n\}$ of \mathbf{R}^n).

We can provide a similar formulation for weighted graphs. We will just weight the gradient of the signal along an edge by the weight of this edge. Intuitively this expresses that two nodes linked together by a strong edge are heavily related, and thus a variation of the signal between those nodes should be more significant than the same variation along a weaker edge. The gradient operator is therefore expressed as $W \times \nabla$ and we can derive the following similar formula for the weighted Laplacian matrix, where D is the diagonal matrix of the weighted degrees.

$$L = \nabla^\top (W \nabla) = D - W \quad (3)$$

2.1.2 Spectral Analysis of the Graph Laplacian

We have derived an operator that solely depends on the structure of the graph, and quantifies the variation of a signal on this graph. Indeed, given a signal s , the Laplacian matrix can be used to express the **Dirichlet energy** $E(s)$ of the signal:

$$E(s) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{u,v} w_{u,v} (s(u) - s(v))^2 = s^\top L s$$

The Dirichlet energy quantifies how much the signal varies along each edge, taking into account the weight of each edge (*i.e.* how strongly the two nodes are related). A signal whose energy is low can be seen as a signal whose variations are **smooth** with respect to the structure given by the edges: a strong connection between two nodes leads to a similar value of the signal on both extremities. In the other hand, high energy, means that some strongly connected nodes have a significantly different value, hence some notion of abrupt variation.

Equation (3) also shows that the Laplacian matrix is real and symmetric, and thus admits a nice diagonal decomposition in a basis of orthogonal eigenvectors.

$$L = \chi \Lambda \chi^\top$$

Where $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix whose entries are the eigenvalues of L and χ is the matrix of the corresponding eigenvector basis (v_1, \dots, v_n) . We will assume that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Observe that since L can also be written as $L = \nabla^\top \nabla$, it is a semi-positive matrix, thus its eigenvalues are non-negative. Also note that since each row of L sums to 0 (see equation (3)), 0 is an eigenvalue of L , and thus we have $\lambda_1 = 0$.

Because the v_i 's are vectors of \mathbf{R}^n , they correspond to signals on the graph. We will call them the **(eigen-)modes** of the graph. The analogy with the notion of mode that we already knew in physics (those of a vibrating membrane for example) can be made because both are solutions of a Dirichlet problem $Lx = \lambda x$. Using the eigenvector property of v_i , we have:

$$E(v_i) = v_i^\top L v_i = v_i^\top (\lambda_i v_i) = \lambda_i \|v_i\|^2$$

Therefore, the energy of a normalized eigenvector only depends on the associated eigenvalue. This gives an intuitive idea behind the ordering of the eigenvalues: low eigenvalues correspond to modes with small energy (and thus smooth variations: the analogue of low frequency signals for time series), while high eigenvalues lead to mode with high energy (and thus high variations — high frequency). This phenomenon can be observed in figure 7.

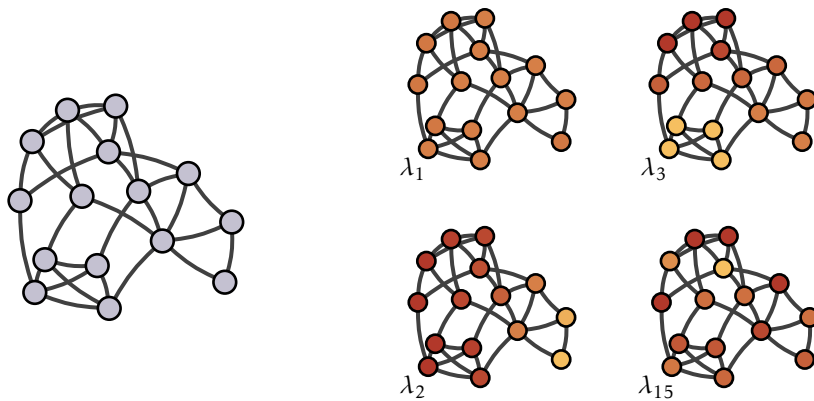


Figure 7: A graph, and some of the eigenvectors of its laplacian matrix. Note that the signal associated with λ_1 (the zero eigenvalue) is the constant signal across the graph. Also observe that higher eigenvalues are associated with signal presenting more abrupt variations.

Since we have a basis of signals, we can decompose any signal s into his representation h on the basis represented by χ , such that

$$s = \chi h = h_1 \begin{matrix} \text{graph} \\ \nu_1 \end{matrix} + h_2 \begin{matrix} \text{graph} \\ \nu_2 \end{matrix} + h_3 \begin{matrix} \text{graph} \\ \nu_3 \end{matrix} + \dots + h_{15} \begin{matrix} \text{graph} \\ \nu_{15} \end{matrix}$$

This is what is referred to as the **Fourier Transform** for graphs signals. We can thus decompose any graph signal into a sum of the modes of the graph, and the coefficients of this decomposition contain the spectral information of the signal. Because χ is orthogonal, this decomposition can be obtained through

$$h = \chi^\top s$$

2.2 Optimal Transport of Smooth Graph Signals

Optimal transportation theory provides tools formulating a notion of distance among **probability distributions**, such as the **Wasserstein distance** between probability measures (see appendix A for an in-depth introduction). Up to now, we have only been working with graphs, that are discrete objects but no distributions. An idea introduced in [DTFV16] is that a graph comes with a family of signals that are somehow natural on it: **smooth** signals. The core idea of the paper is that we can define a meaningful distribution of such smooth signals over the graph, and exploit the power of optimal transport tools on them. With the framework that we have introduced in the previous section, this corresponds to signals whose spectral mass is mainly concentrated around modes of low frequency.

The authors describe the following distribution of smooth signals over the graph. A random smooth signal should be such that the coefficient h_i associated to the eigenvector of ν_i in its spectral decomposition follows $\mathcal{N}\left(0, \frac{1}{\lambda_i}\right)$. Said differently: the higher the eigenvalue of a mode, the more its spectral mass will be squeezed toward 0. Such a distribution will thus favour the low frequency modes and naturally lead to a class of smooth signals over the graph. The vector h thus follows a multivariate distribution $\mathcal{N}\left(0, \Lambda^{-1}\right)$. Using the fact that a signal s and its decomposition h are linked through $s = \chi h$, this gives a distribution of signals that can be expressed as the following multivariate distribution:

$$\mathcal{N}\left(0, \underbrace{\chi \Lambda^{-1} \chi^\top}_{L^{-1}}\right)$$

For correctness sake, we should mention that the Laplacian matrix is not truly invertible (recall that 0 is one of their eigenvalues), and thus the inverses mentioned earlier should rather be written as L^\dagger , where \cdot^\dagger denotes the pseudo-inverse of the matrix.

The authors of [MEGCF19] use this link between a graph and a distribution of signals to use the framework of optimal transport. Since the objects manipulated have gone from the space of graphs to a space of distributions, they propose to use an optimal transport measure (the Wasserstein distance) to the problem of comparing two graphs. This idea is summed up in figure 8.

We thus need to compute an optimal transport distance between those measures. One of the drawbacks of today's optimal transport theory is its lack of tractable numerical schemes, especially for high dimensional distributions. Fortunately, the distributions considered here (multivariate Gaussians) enjoy quite a

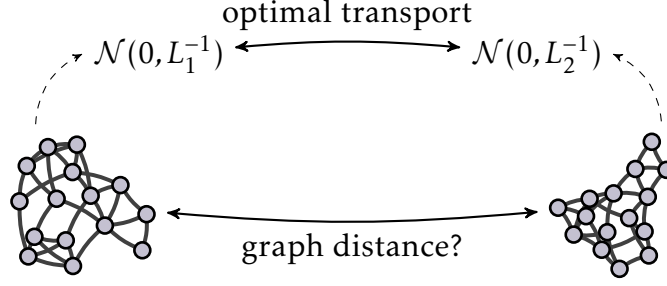


Figure 8: Using the optimal transport framework to compare two graphs: process used in [MEGCF19]

few good properties and the authors of [MEGCF19] use an explicit formulation for the 2-Wasserstein distance between two multivariate Gaussian distributions (also referred to as the Bures-Wasserstein distance [BJL19]): if $\nu_1 = \mathcal{N}(0, \Sigma_1)$ and $\nu_2 = \mathcal{N}(0, \Sigma_2)$.

$$W_2^2(\nu_1, \nu_2) = \text{tr } \Sigma_1 + \text{tr } \Sigma_2 - 2 \text{tr } \sqrt{\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2}}$$

The square roots are well defined in our case since the Laplacian matrix — and thus their (pseudo-)inverses — are positive semi-definite.

Note that for this approach to make sense, the graphs must have the same number of nodes. Moreover, they should be aligned, that is the ordering of the nodes of each graph, should be meaningfully comparable in both graphs: for instance to compare some networks associated to the economy of two countries, the capital of each country should correspond to the same node v_i in both graphs. The article [MEGCF19] dwells upon the issue of finding a good graph alignment, but we are not concerned interested by this since the networks that we will compare are issued from known brain atlases and thus we have enough prior knowledge of the networks to align them beforehand.

2.3 Wasserstein Distance on our Connectivity Matrices

At this point, we decided to try using the 2-Wasserstein distance on the dataset at our disposal. We wanted to observe the differences between controls and patients. A latent idea was that we could be able to classify a new subject from its connectome using the developed technique. This was backed up by studies having brought out significative connectivity difference between the control and the depressive cohorts. I thus implemented the distance introduced in [MEGCF19] that I discussed in the previous section. The hope was that the known structural differences ([CBC⁺18]) between the two cohorts would be captured by the transportation distance between graphs.

The approach I first chose is to compute a pairwise distance between each subject and each control. I aggregated those distance into an average distance to the control cohort for each subject. Our intuition was that, because of the connectivity change in specific networks related to the disease, the average distance between a control and the cohort of other controls would be smaller than the average distance between a patient and the cohort of control. In figure 9, I represented the distribution of average distances to the control group from either the control group or the

patient group. We expected to see a shift in the distributions, reflecting the greater homogeneity inside the control group.

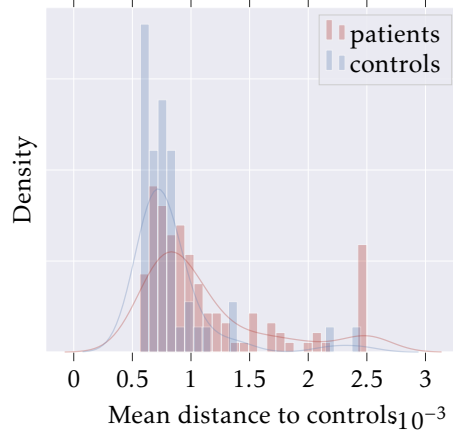


Figure 9: Pure Wasserstein distance to controls, the spike at the right corresponds to distances that were clipped to $2.5 \cdot 10^{-2}$, in order to keep a readable scale.

Although it captures the greater homogeneity among controls, the distance does not really discriminates between both groups. Even though we endorse the matrix space with a meaningful geometry, its high-dimensionality might still cause it hard to analyze, the same way k-Nearest-Neighbours classifier tend to struggle with high dimensionality data.

2.4 Barycenters of Gaussian distributions

An idea that I proposed to enhance results is to have a notion of barycenters of connectomes. This approach emerges naturally since we are willing to compare cohorts: being able to compute a mean connectome of a control group to which we compare individual subjects instead of averaging individual distances to subjects might provide better results than the previous approach. A geometric intuition of why this could enhance our results is provided in figure 10.

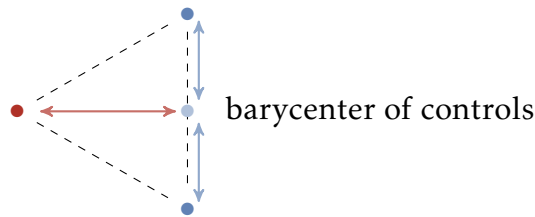


Figure 10: Geometric intuition of why to use barycenters in our study. Even though all subjects are roughly equidistant and thus hard to distinguish simply by computing the pairwise distances (in some arbitrary metric space), the controls (blue) are closer to their barycenter than the patient (red), and computing individual distances to the barycenter allows to classify more reliably.

The notion of barycenter (or centroid) is well defined in euclidean spaces. It finds a natural generalization in metric spaces with the formulation of the Fréchet mean. This gives, for instance, a natural way to define the geometric mean as a centroid: via the Fréchet mean for the hyperbolic distance.

As detailed earlier, we have embedded connectivity graphs into a probability space endowed with the Wasserstein metric. The formulation of the Fréchet mean $\bar{\mu}$ of a collection of measures (μ_1, \dots, μ_k) in this space (which we will also call a **Wasserstein barycenter**) is as follows:

$$\bar{\mu} = \arg \min_{\mu} \sum_{i=1}^k \lambda_i W_2^2(\mu, \mu_i) \quad \text{where } (\lambda_1, \dots, \lambda_k) \text{ are weights}$$

Computation of such barycenters in the theory of optimal transport often relies on samples of the distributions. See for instance [PC⁺19] for a review of numerical schemes behind this optimisation problem. Unfortunately, our distributions of interest are not described by samples but rather by their covariance matrix, which is convenient because sampling such high dimensionality distributions is expensive. Fortunately, multivariate Gaussians are easy to work with, and we will be able to use results of [ÁEdBCAM15] that gives an interesting and computationally fecund characterization of the Wasserstein barycenter, which happens to be the fixed point of a certain application:

$$\bar{\mu} = G(\bar{\mu})$$

G is defined as the function mapping a distribution μ to the distribution followed by $\sum_{i=1}^k \lambda_i T_i(X)$ where $X \sim \mu$ and T_i is the transport map sending μ to μ_i . One can get the intuition of this result as follows: the barycenter is the measure that is stabilized when geodesically moved toward each of the μ_i . Note that the subtlety here is that the distributions followed by each of the $T_i(X)$ are not independent of each other because the one and only random variable X is transported in each term. We will apply this result to our multivariate normal distributions.

A known result is the expression of T_i when we transport a Gaussian toward another Gaussian distribution, see e.g. [BJL19]. If $\mu = \mathcal{N}(0, \Sigma)$ and $\nu_i = \mathcal{N}(0, \Sigma_i)$,

$$T_i(x) = \Sigma^{-1/2} \left(\sqrt{\Sigma^{1/2} \Sigma_i \Sigma^{1/2}} \right) \Sigma^{-1/2} x \quad (4)$$

We denote A_i the matrix in front of x in the above expression. Applying the map G to the Gaussian measure $\mu = \mathcal{N}(0, \Sigma)$ when each ν_i is of the form $\mathcal{N}(0, \Sigma_i)$ yields that $G(\mu)$ is the distribution of a sum of affine transformations of Gaussians random variables, whose mean is 0, and whose covariance matrix can be expressed as:

$$\begin{aligned} (\lambda_1 A_1 \quad \dots \quad \lambda_k A_k) \begin{pmatrix} \Sigma & \dots & \Sigma \\ \vdots & \ddots & \vdots \\ \Sigma & \dots & \Sigma \end{pmatrix} \begin{pmatrix} \lambda_1 A_1 \\ \vdots \\ \lambda_k A_k \end{pmatrix} &= \sum_{i,j}^k (\lambda_i A_i) \Sigma (\lambda_j A_j) \\ &= \sum_{i,j}^k \lambda_i \lambda_j \left(\Sigma^{-1/2} \sqrt{\Sigma^{1/2} \Sigma_i \Sigma^{1/2}} \Sigma^{-1/2} \right) \Sigma \left(\Sigma^{-1/2} \sqrt{\Sigma^{1/2} \Sigma_j \Sigma^{1/2}} \Sigma^{-1/2} \right) \\ &= \Sigma^{-1/2} \left(\sum_{i=1}^k \lambda_i \sqrt{\Sigma^{1/2} \Sigma_i \Sigma^{1/2}} \right)^2 \Sigma^{-1/2} \end{aligned}$$

The authors of [ÁEdBCAM15] show that — under some mild hypothesis — this function G gives a natural iteration scheme $\mu_{n+1} = G(\mu_n)$ that converges to

the unique fixed point of G , which is the desired barycenter. If we denote $\mu^{(n)} = \mathcal{N}(0, \Sigma^{(n)})$, it provides the following iteration scheme that converges toward the covariance of the barycenter:

$$\Sigma^{(n+1)} = \Sigma^{(n)-1/2} \left(\sum_{i=1}^k \lambda_i \sqrt{\Sigma^{(n)1/2} \Sigma_i \Sigma^{(n)1/2}} \right)^2 \Sigma^{(n)-1/2}$$

Figure 11 shows the result of this iterative scheme to compute the barycenter of four two-dimensional Gaussians.

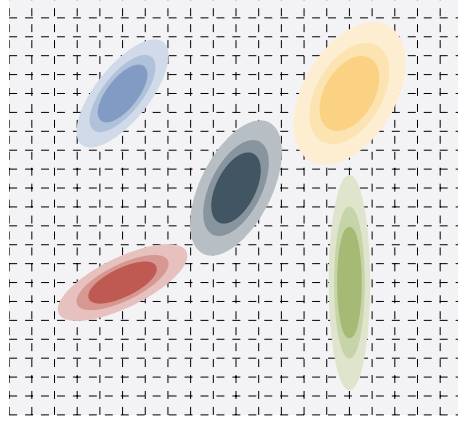


Figure 11: Barycenter of 2D Gaussians computed using the iterative scheme. The barycenter is the darkest distribution, in the middle. Observe that the barycenter have benefited from characteristics of all measures, while avoiding over-representation of individual variability, such as the squeezing of the green distribution.

2.5 Barycenters for Connectivity Analysis

The idea at this point was that we could continue to compare connectomes in the probability space introduced earlier, but instead of individually comparing subjects to each control, we can now start by computing a barycenter of the group of interest and then compare each subject to this barycenter. Not only this increases the speed of comparing a new subject to the control cohort since only one distance needs to be computed, but this also leverages that we are comparing a subject to the relevant features common across a whole cohort, thus avoiding too much emphasis put on individual variability. In figure 12 I show the barycenter computed from the full controls cohort.

To assess the utility of barycenter, I devised the experiment shown in figure 13. I repeatedly split the control cohort into two, used one part of the controls to compute a barycenter, and computed the distance between the other controls and the patients to the barycenter.

Once again, the results (see figure 14) were not significantly more satisfying than those obtained through the first method. The expected shift in the distributions, although existing because of the greater homogeneity in the control group, is not significant, and only a handful of patients are indeed discriminated by those distance-based methods.

I conducted further experiments, either by increasing the sparsity of the networks, or by splitting the connectome into hemispheric subnetworks to exploit

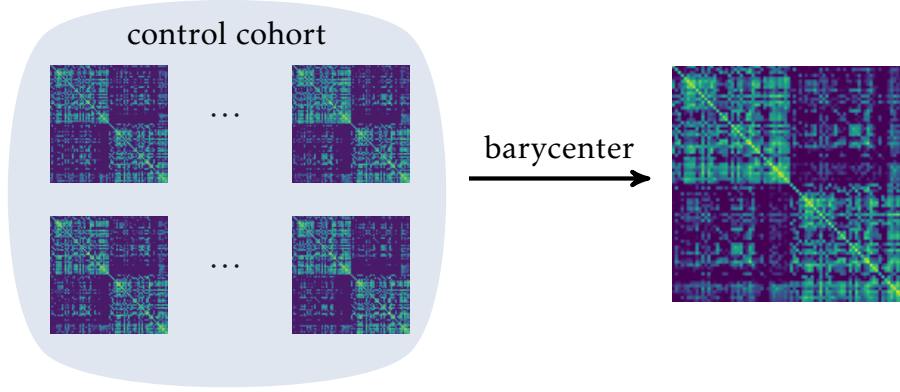


Figure 12: Barycenter computation. Note that we compute the barycenter in a probability space, and that the shown connectivity matrix only approximately corresponds to a true network.

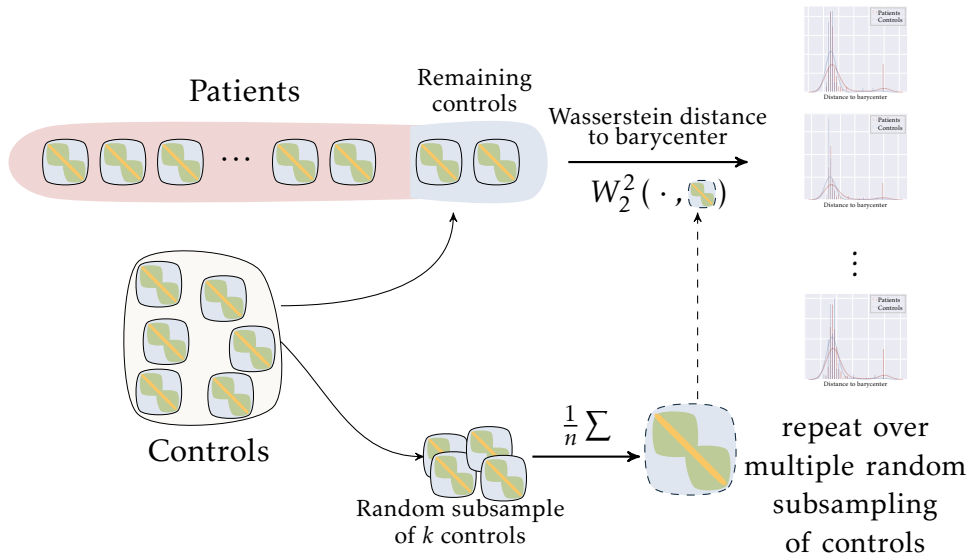


Figure 13: Barycenter folding tests. Note that the $\frac{1}{n} \sum$ arrow denotes the computation of the Wasserstein barycenter.

some properties of the distance described in [MEGCF19], but none of those yielded better results.

3 Graph Curvature

I investigated another marker that could be used to capture properties of the network. There is often a duality between **mesh processing** and graph analysis. Indeed, the tools introduced in section 2 in the context of graph signal processing, such as the graph Laplacian, are also defined and used in mesh processing. The intuition behind this is that shapes can often be represented through their combinatorial structure, and thus through an adjacency graph. Ollivier-Ricci curvature is another of those common tools, whose intuition first come from geometric objects, but can provide insight on the structure of the graph. This tool is starting to get used in complex network analysis, such as in [NLG⁺15] or [SJB19], and even in the context of neurosciences like in [FCG⁺19]. I will introduce this tool, the intuition

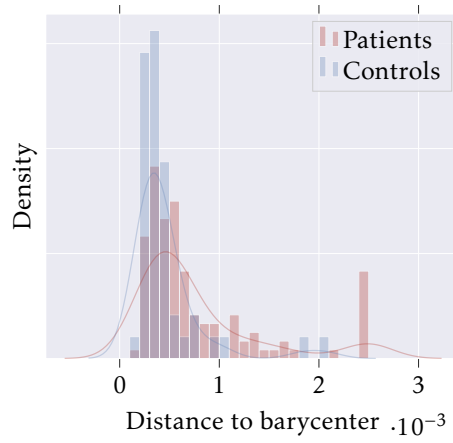


Figure 14: Distances to the control barycenter aggregated from the folds of the experiment presented in figure 13.

backing it as well as its theoretical foundations. I will then present ideas to adapt this tool in the context of our study.

3.1 From Riemannian manifolds to networks

Curvature is usually defined on geometric objects, as we will first see. I will then present how a notion of curvature can be computed on combinatorial objects (graphs in our context), and convey information on the structure of the object.

3.1.1 Ricci curvature on manifolds

A **manifold** can be roughly defined as a space that locally coincides with a Euclidean space. More precisely, a n -manifold it is a set $\mathcal{M} \subset \mathbf{R}^d$ such that for each point x of \mathcal{M} , there exist an affine subspace of \mathbf{R}^d of dimension n that coincides with x at first order. This space is called the tangent space of \mathcal{M} at x . An example of such set is a sphere embedded in \mathbf{R}^3 . Its local similarity with a Euclidean space is responsible for the phenomenon that makes the Earth hardly distinguishable from a plane at first order, at the human scale.

Thinking about **curvature** aims to quantify how much the manifold differs from a Euclidean space at second order. Several notions of curvature coexist. I will briefly introduce the **sectional curvature** and the **Ricci curvature**, both defined in the setting of **Riemannian manifolds**.

A Riemannian manifold is a manifold \mathcal{M} equipped with a **Riemannian metric**. That is for every point x in \mathcal{M} , there exist an inner product defined in the tangent space of \mathcal{M} at x , inducing a norm in this tangent space. A **canonical** such Riemannian metric is the restriction of the Euclidean inner product to the tangent space. This is illustrated in figure 15 for a sphere.

Given a smooth curve γ on \mathcal{M} ($t \mapsto \gamma(t) \in \mathcal{M}$), its derivative $\frac{d\gamma}{dt}(t)$ is a vector of the tangent space of \mathcal{M} at $\gamma(t)$, whose length can thus be defined thanks to the norm induced on the tangent space. By integration along the curve, we can therefore define the length of the curve. The **geodesic distance** between two points of \mathcal{M} is the infimum of the length of a smooth curve joining the two said points. This induces a metric d on the manifold, such that $d(x, y)$ is the geodesic distance between those two points.

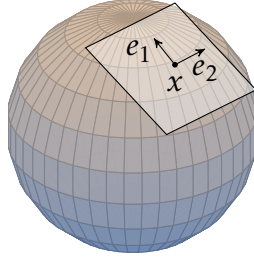


Figure 15: The sphere S embedded in \mathbf{R}^3 seen as a two dimensional Riemannian manifold. The tangent space of S at point x is the affine space represented, and an orthogonal basis (e_1, e_2) of this space can be used to define the canonical inner product inside this tangent space.

Using some geometrical intuition, we can easily define the **parallel transport** of a tangent vector along a curve. This is illustrated in figure 16 where we transport the tangent vector w_x from x to y along v , giving the vector w_y tangent to the manifold at y . This can be used to assess the local discrepancy between a manifold and a Euclidean space. Indeed, in a pure Euclidean setting, parallel vectors are actually rigorously parallel, in the sense that following both vectors does not result in either convergence or divergence. In a Riemannian manifold, parallel vectors are defined to be as parallel as the space allows it, which can result in convergence or divergence when following the geodesic they define.

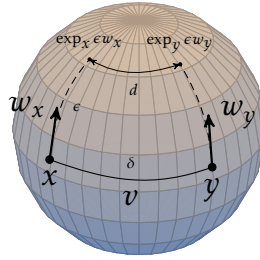


Figure 16: Parallel transport of w_x from x to y following v . The dotted lines represent the constant-speed geodesic starting from either x or y with initial speed given by the tangent vectors w_x and w_y , and running for a time ϵ . The sectional curvature at x in direction w along v (denoted $K(v, w_x)$) will be defined through the behaviour of d when $\epsilon \rightarrow 0$ and $\delta \rightarrow 0$.

More formally, using the notations from figure 16, the sectional curvature in direction w_x along v (denoted $K(v, w_x)$) is defined as a component of the asymptotic behaviour of $d(\exp_x \epsilon w_x, \exp_y \epsilon w_y)$:

$$d(\exp_x \epsilon w_x, \exp_y \epsilon w_y)_{\epsilon, \delta \rightarrow 0} = \delta \left(1 - \frac{\epsilon^2}{2} K(v, w_x) + \mathcal{O}((\delta + 1)\epsilon^2) \right)$$

This highlights that a positive sectional curvature leads to converging parallels (like on the above sphere), while a negative curvature leads to diverging parallels. This definition is dependant on both the direction v along which we transport and on the vector w_x that is transported. **Ricci curvature** is defined as the average of sectional curvature over all the directions w_x in the tangent space at x .

This can also be rephrased in term of asymptotic behaviour, in similar fashion to the definition of K . If we define \bar{d} to be the average of the distances d depicted in figure 16 for w_x describing the unit circle in the tangent space, we have

$$\bar{d}_{\epsilon, \delta \rightarrow 0} = \delta \left(1 - \frac{\epsilon^2}{2N} \text{Ric}(v) + \mathcal{O}((\delta + 1)\epsilon^2) \right)$$

3.1.2 Ollivier-Ricci Curvature in Metric Space

This definition is only valid for Riemannian manifolds. We would like to extend this, along with the intuitions behind it to general metric spaces, such as graphs. I will detail the approach formulated by Yann Ollivier in [Oll09]. His approach finds its roots in theorems, formulated by Lott and Villani, as well as Sturm, linking optimal transport distances, Boltzmann entropy (denoted H in what follows) and Ricci curvature on (still) Riemannian manifolds.

To understand the relationship between those notions, let us look at the following theorem, which has a very intuitive interpretation. If k is a lower bound for the Ricci curvature on the manifold \mathcal{M} , μ_0 and μ_1 are measures defined on \mathcal{M} , and $(\mu_t)_{t \in [0,1]}$ is the interpolation between those two measures (in the sense of the Wasserstein W_2 metric, see appendix A for more details), we have [vRS05]

$$H(\mu_{\frac{1}{2}}) \geq \frac{1}{2} (H(\mu_0) + H(\mu_1)) + \frac{k}{8} W_2^2(\mu_0, \mu_1) \quad (5)$$

If we remind, from information theory, that entropy basically accounts for the spread of the measure (entropy increases as the measure gets closer to being uniform), this theorem is actually saying that measures defined on a manifold tend to spread under positive curvature, while they are able to get shrunk while on negatively curved spaces. This can be observed in figure 17.



Figure 17: Dilatation (left) and contraction (right) property of the entropy under positive (left) and negative (right) curvature. Blue areas represent measures of finite support on the manifold. The red measure in the center represents (both in the left and right cases) the interpolation (in the Wasserstein sense) between the two others.

Another property, that will get us closer to the definition of Ollivier-Ricci curvature is the following. If we define $m_{r,x}$ to be the uniform measure on \mathcal{M} whose support is the circle of center $x \in \mathcal{M}$ and of radius r (for the geodesic distance on \mathcal{M}), one can show (cf. [vRS05]) that equation (5) under the condition $\text{Ric}(\mathcal{M}) \geq k$ is equivalent to

$$W_1(m_{r,x}, m_{r,y}) \leq \left(1 - \frac{k}{2(n+2)} r^2 + o(r^2) \right) d(x, y) \quad (6)$$

This last theorem is interesting, providing a link between a lower bound k of the curvature and the geodesic distance d on the manifold \mathcal{M} . It is the inspiration for the definition of a notion of curvature on metric spaces, removing the need for a Riemannian framework.

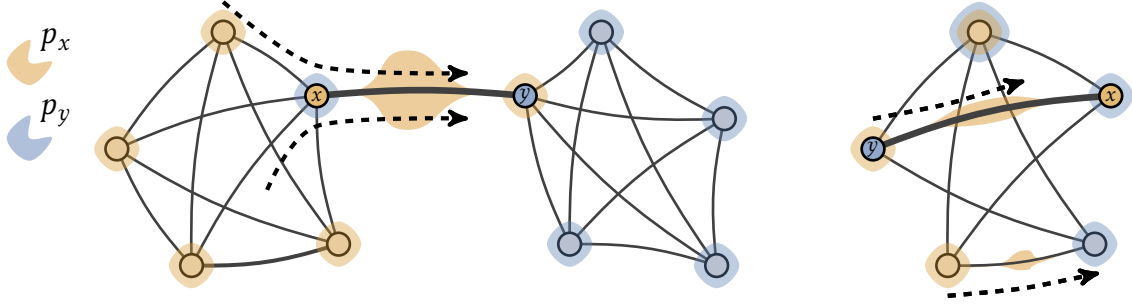


Figure 18: Intuition of Ollivier-Ricci curvature on graph. p_x and p_y are measures spread among the neighbourhoods of x and y . The curvature of the edge (x, y) depends on how p_x is transported to p_y . The bold (x, y) edge on the left has a negative curvature, because it connects two disjoint community and thus most of the mass of p_x has to be transported through this edge, causing the transportation distance to be at least $d(x, y)$. On the right, the edge has positive curvature, because it connects nodes from the same community and that share neighbours. There is less mass to move, and on shorter paths, so the total transportation distance is less than $d(x, y)$.

Hence, in a geodesic metric space (\mathcal{X}, d) equipped with a set of measures $\{p_x \mid x \in \mathcal{X}\}$, Ollivier defined in [Oll09] the **Ollivier-Ricci curvature** $\kappa(x, y)$ of the geodesic linking x and y using the following equation

$$W_1(p_x, p_y) = (1 - \kappa(x, y))d(x, y) \quad (7)$$

The p_x here essentially stands for a measure concentrated around x , which is similar to the $m_{r,x}$ introduced earlier. One can notice the direct analogy between the definition (7) and equation (6).

A direct interpretation of a positive Ollivier-Ricci curvature between x and y is the following: if one wants to transport a mass spread around x toward a mass spread around y , the total cost is smaller than the geodesic distance between x and y (from equation (7)). That might be because two points on the neighbourhood of x and of y are closer than x and y for the geodesic distance. Typically on a graph, nodes x and y might share neighbours, or well-connected neighbour's. Conversely, a negative curvature shows that the best way to transport the mass between x and y might be to actually follow the geodesic between x and y . In the context of a graph, positive curvature tends to happen inside communities, while negative curvature appears for inter-community edges. This is highlighted in figure 18

This provides a tool that can be used on graphs, as long as a geodesic distance can be defined on those. Indeed, 1-Wasserstein distance can easily be computed using dynamic programming. Some studies have provided good evidence that this can capture deep structural information about graphs (for example [NLLG19]). Those intuitions are backed up by some theoretical results: for instance [BJL11] shows a link with the spectrum of the Laplacian matrix of a graph, which is known to carry structural information as well.

3.2 Ollivier-Ricci Curvature on Connectivity Graphs

My idea here, was to follow an approach used in structural connectivity analysis with some of the topological indicators mentioned in section 1 (see figure 4). A way to use those local metrics is to compute them for each node of each subject, and

then, for each node, compare the distribution of metric behind the control group and the patient group for instance. Any significant difference would cause the node (and thus the corresponding brain area) to be considered of interest for the studied pathology. This is what is done in [CBC⁺18] for instance. I wanted to adopt a similar strategy, but using curvatures instead of the more “traditional” measures. It turned out that a recent paper ([FCG⁺19]) had already devised a similar method, with relative success.

I described in the previous theoretical section, the only requirement to compute an Ollivier-Ricci curvature, is to have a metric space. In our context, this means being able to compute a distance between two nodes on the graph. The choice of [FCG⁺19] was to consider shortest-paths in the connectivity graph, where the distance would simply be computed as the number of hops between two nodes. Even though accounting for the structure of the graph, choosing this notion of distance disregards the weights of connections, which is one of the drawbacks that we mentioned in section 1 and that we are trying to avoid.

I made the choice to introduce another notion of distance, based on the **heat kernel** of the graph. The heat kernel is basically the solution to the diffusion equation at the surface of the graph, involving the graph laplacian derived in section 2.

$$\left(\frac{\partial}{\partial t} + L\right)f = 0$$

There is a direct analogy with the heat equation from physics. This equation describes the diffusion of what could be the electrochemical energy of a signal in the brain. As so, it is relevant to study brain networks [CSK⁺16]. Links can also be formulated between this equation and random walks on connectivity graphs. Those random walks have been studied with interest in [Rob12] [RSPH14].

A solution of this equation is $f : t \mapsto e^{-tL}f_0$. The matrix $k_t = e^{-tL}$ is called the heat kernel of the graph, and each of its entries $k_t(x, y)$ represents the amount of heat initially in x having reached y at time t . As a kernel, it defines a distance between nodes of the graph

$$d_t(x, y) = k_t(x, x) + k_t(y, y) - 2k_t(x, y)$$

This distance, for well chosen t , takes into account the combinatorial structure of the graph, as well as the strengths of the connections, as shown in figure 19. This makes it more suited than the “shortest number of hops” distance. Now that a distance is defined between nodes, we can compute the curvatures of each connection.

Looking at connections having significative curvature differences between controls and patients yields more connections than what the methods using classic topological markers used to yield (see figure 20, where we performed a similar analysis than in [CBC⁺18]). This is both consistent with the claims of [FCG⁺19], and with the intuition that curvature, along with the heat kernel distance, is locally sensitive to the global topology of the graph, and thus a wider part of the networks finds its curvature affected by slight changes between the control cohort and the patients.

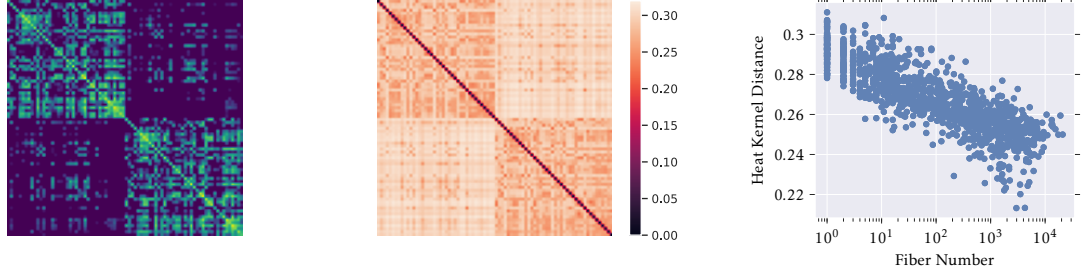


Figure 19: Heat Kernel distance computed on a connectivity matrix. Left: original connectivity matrix. Middle: heat kernel distance matrix. Right: each point represents a fiber connection (i.e. an entry of the original matrix), the x axis represents the strength of the anatomical connection, while the y axis is the heat kernel distance between each extremities of the connection.

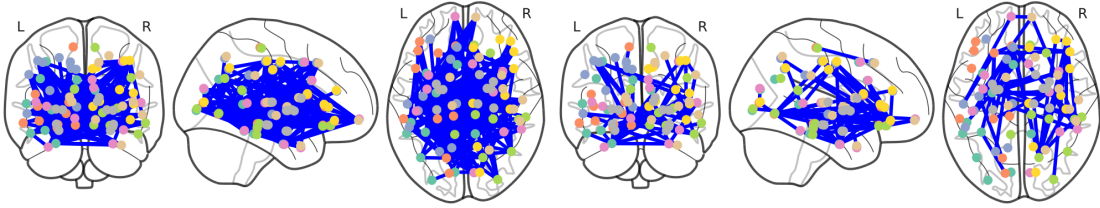


Figure 20: Left: connections with significant curvature difference between controls and patients. Right: connections with significant fiber number differences between controls and patients.

Conclusion

I have proposed to use a distance between graphs introduced recently in [MEGCF19]. I have also extended this distance by the use of barycenter computations. Having tried both approaches on the dataset at my disposal, it seemed that the distance and the barycenters computed with it were able to capture and reflect some of the structural information of the connectomes. Yet, the distance was not subtle enough to really discriminate the two groups, which mitigate its use for further studies in this context. I think that the notion of barycenter for connectomes that I introduced could still prove useful, thanks to its ability to leverage structural information from a whole cohort. It would be of interest to do further experiments with it to better understand what it could yield, maybe with toy graphs. Another point that I have not fully exploited is the knowledge of the transport map associated with the Wasserstein distance.

I also studied Ollivier-Ricci curvature and its potential application to connectome comparison. I proposed a framework to use it in the setting of structural connectivity data that is novel from how it had already been used. It should circumvent some of the drawbacks of topological markers that we discussed in section 1 and that are still present in the recent uses of Ollivier-Ricci curvature for connectivity analysis ([FCG⁺19]). I am still evaluating and comparing this approach to prior works. There are also techniques exploiting the Ollivier-Ricci curvature that are worth exploring, especially for connectome comparison. Among others lie the Ricci flow, that has been used with success for (non-biologic) complex networks, see for instance [WJS16] or [NLLG19].

A Introduction on Optimal Transport

The history of optimal transport formulations can be traced back to Gaspard Monge's *Mémoire sur la théorie des déblais et des remblais* (1781). In this document, he studied how to move masses of sand or soil from one place to another. The main motivation of this *Mémoire* being that a cost is naturally associated to the act of moving a mass particle from one place to another, and it is of particular interest to seek the optimal way of moving a global quantity of mass from one place to another, that should require the lowest amount of effort possible.

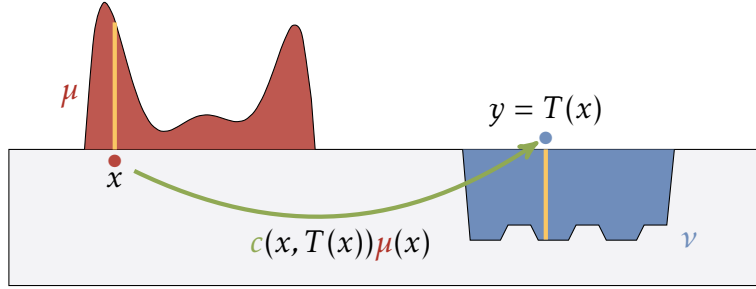


Figure 21: Earth mover problem. We want to fill the blue hole (the *déblai*, which is also called the **target distribution** ν) using the red matter (the *remblai*, which is also called the **source distribution** μ). T represents the transport map, describing a way to send particles from the source distribution to the target distribution. $c(x, T(x))$ is the cost of moving from x to $T(x)$

Monge's formalism. We work in a space Ω . We consider two probability measures μ and ν on Ω , respectively the **source** and the **target** distribution. We also consider a cost function $c : \Omega \times \Omega \rightarrow \mathbf{R}$. If Ω is some Euclidean space, c can, for instance, be the Euclidean distance induced in it. Monge's problem is formulated as follows:

$$\begin{aligned} \min_{T: \Omega \rightarrow \Omega} \int_{\Omega} c(x, T(x)) d\mu(x) \\ \text{s.t. } \forall A \in \mathcal{P}(\Omega), \nu(A) = \mu(T^{-1}(A)) \end{aligned}$$

The function T is called a transport map. It carries information about the displacement of each elementary unit of probability: the mass located in x is moved toward $T(x)$. The function c thus allows quantifying the cost of moving one elementary unit of mass. We want to minimize the average distance travelled by mass units during the optimal transport. The condition under which the optimization is performed ensures that all the mass of μ has been transported toward the distribution ν . Those concepts are introduced in figure 21.

The issue with such a formulation of an optimal transport problem is that it does not handle discrete probability measures. Indeed, if μ has a Dirac impulse bigger than every Dirac impulse of ν , then the transport map cannot be built: we are only performing displacements of mass and not any actual allocation of mass. For example in figure 21, we are not allowed to split any of the transported yellow units.

Kantorovich's formalism. Kantorovich came up with an alternative formulation for the optimal transport problem, that aims at handling discrete distributions.

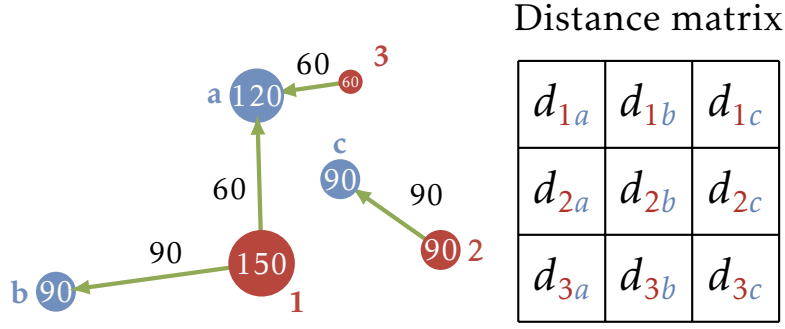


Figure 22: A discrete Kantorovich problem. The red distribution is still the source distribution and the blue the target one. Now that we work in a discrete setting, the cost function has been replaced by a distance matrix, expressing the pairwise costs of moving from one atom of the source distribution to an atom of the target distribution.

Kantorovich's idea was to change the function $T : \Omega \rightarrow \Omega$ into a probability P on the product space $\Omega \times \Omega$ called the **transport map**. The probability $P(x, y)$ describes how much of the mass in x is moved toward y . If we denote D the distance matrix encoding the costs of the problem, we obtain the following formulation in the discrete case:

$$\begin{aligned} \min_P \quad & \langle P, D \rangle = \sum_i \sum_j p_{ij} d_{ij} \\ \text{s.t.} \quad & \begin{cases} \forall i & \sum_j p_{ij} = \mu_i \\ \forall j & \sum_i p_{ij} = \nu_j \end{cases} \end{aligned}$$

The two conditions encode the fact that we aim to move all the mass of μ toward all the mass of ν . In the problem represented in figure 22, Monge's formulation forbids to split the mass of size 150 into two separate masses, and the problem is not feasible. Yet, Kantorovich's formulation allows this: as shown in figure 22, the mass of 150 is split into a mass of 90 and a mass of 60, both sent to different places.

This is clearly a **linear program**, for which efficient solvers are still under extensive study; [PC⁺19] provides a detailed overview of many numerical aspects of this problem.

Continuous Kantorovich's formulation. In the continuous setting, we introduce the following set of joint probabilities, which encodes the analogue of the two conditions in the discrete case, making sure that μ has been fully transported to ν .

$$\Pi(\mu, \nu) = \{P \in \mathcal{P}(\Omega \times \Omega) \mid \forall A, B \in \Omega, P(A \times \Omega) = \mu(A), P(\Omega \times B) = \nu(B)\}$$

We can rephrase the definition of $\Pi(\mu, \nu)$ as being the set of probability measures on $\Omega \times \Omega$ with marginals μ and ν . We consider once again a cost function c as in Monge's problem. Kantorovich's formulation now becomes the following optimization problem, where $X \sim \mu$ and $Y \sim \nu$:

$$\inf_{P \in \Pi(\mu, \nu)} \mathbb{E}_P[c(X, Y)] = \iint_{\Omega \times \Omega} c(x, y) P(x, y) dx dy$$

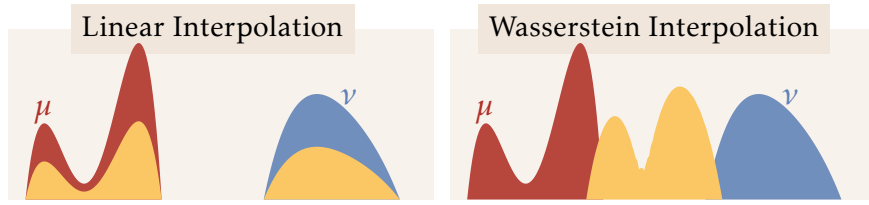


Figure 23: Two interpolations (in yellow) between μ and ν . Left: linear interpolation, defined as $(\mu + \nu)/2$ (or equivalently as $\arg\min_{\eta} l_2^2(\mu, \eta) + l_2^2(\nu, \eta)$, where l_2^2 is the squared difference between shapes) . Right: Wasserstein interpolation, defined as $\arg\min_{\eta} W_2^2(\mu, \eta) + W_2^2(\nu, \eta)$

Existence of solutions to Kantorovich’s problems. One can wonder, in light of what happened for Monge’s problems, whether solutions of Kantorovich’s problem as formulated above exist. A rich literature exists on this topic. For instance, in the case of measures on the \mathbf{R}^n space, with absolutely continuous measures μ and ν with respect to the Lebesgue measure, the existence (and uniqueness) of solutions will only depend on the properties of the cost function. If we consider the cost function $c_p(x, y) = \|x - y\|^p$, then we have the following discussion:

- $p > 1$: the strict convexity of c_p ensures that there is a unique solution to the Kantorovich problem
- $p = 1$: here we can obtain the existence of a minimizer for the Kantorovich, but we won’t have the unicity of the solution
- $p < 1$: in general, there will not exist solutions to the problem

More generally the existence of minimizers, as well as their uniqueness depends heavily on the structure of the space we consider, and the properties of both the measures and the cost function. A comprehensive study of those properties can be found in [Vil03].

Wasserstein distance. Having formulated this optimization problem allows us to define a distance between probability measures. Given a probability space Ω , a cost function $c : \Omega \times \Omega \rightarrow \mathbf{R}_+$ and two probability measures μ and ν in $\mathcal{P}(\Omega)$, the **p-Wasserstein-distance** between μ and ν is defined as

$$W_p(\mu, \nu) = \left(\inf_{P \in \Pi(\mu, \nu)} \iint_{\Omega \times \Omega} c(x, y)^p P(x, y) dx dy \right)^{1/p}$$

Wasserstein barycenters. This notion of distance allows to compute barycenters between probability distributions, and obtain results that are way more natural than barycenters obtained with respect to other norms, like the l_2 norm for instance. Figure 23 presents a barycenter obtained under Wasserstein-distance and a linear interpolation (the two averaged distributions being the blue and the red one, and the yellow shape being the barycenter). What emerges from those two interpolations is that the Wasserstein distance seems particularly well suited to manipulate distributions, as it appears to take into account the characteristics (like modes, shape, etc) of both the source and the target distribution, and the resulting barycenter looks like a distribution combining those into one distribution, while the l_2 barycenter does not merge those specificities into one distribution but rather re-normalize the source and target distributions.

B Remerciements et Contexte Institutionnel et Social du Stage

J’ai effectué mon stage dans l’équipe EMPENN (*cerveau* en breton) hébergée à l’IRISA à Rennes. L’équipe regroupe des personnels de l’INRIA, du CNRS, de l’INSERM, et de l’Université de Rennes 1. Je remercie de nouveau Pierre et Julie pour leur encadrement, et tous leurs conseils et éclairages, Armelle pour toute son aide et ses réponses à toutes mes questions, et puis tout le reste de l’équipe encore avec qui j’ai passé d’excellents mois dans la ville de Rennes qu’elle a su me faire découvrir sous son meilleur jour.

J’ai pu, pendant les premiers mois du stage, participer à la vie de l’équipe, tant à travers ses séminaires réguliers, des répétitions de thèses, et puis plus informellement tout au cours des repas, des pauses à la cafétéria, des quiz du lundi soir, des trajets jusqu’au RU. J’ai également eu l’opportunité de prévoir une participation à la *Semaine du Cerveau* à travers une intervention lors d’un *Bar en (neuro)sciences*, bien qu’elle n’ait finalement pas eu lieu pour les raisons sanitaires que l’on sait bien.

References

- [ABR⁺13] Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. Weighted graph comparison techniques for brain connectivity analysis. In **CHI ’13**, 2013.
- [ÁEdBCAM15] Pedro C. Álvarez-Esteban, Eustasio del Barrio, Juan Antonio Cuesta-Albertos, and Carlos Matrán. A fixed-point approach to barycenters in wasserstein space. 2015.
- [BBV⁺08] Danielle S. Bassett, Edward T. Bullmore, Beth A. Verchinski, Venkata S. Mattay, Daniel R. Weinberger, and Andreas Meyer-Lindenberg. Hierarchical organization of human cortical networks in health and schizophrenia. **The Journal of neuroscience : the official journal of the Society for Neuroscience**, 28 37:9239–48, 2008.
- [BJL11] Frank Bauer, Jurgen Jost, and Shiping Liu. Ollivier-ricci curvature and the spectrum of the normalized graph laplace operator. 2011.
- [BJL19] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the bures–wasserstein distance between positive definite matrices. **Expositiones Mathematicae**, 37(2):165–191, 2019.
- [CBC⁺18] Julie Coloigner, Jean-Marie Batail, Isabelle Corouge, Dominique Drapier, and Christian Barillot. White matter connectivity analysis in patients suffering from depression. 2018.
- [CSK⁺16] Ai Wern Chung, MD Schirmer, Michelle L Krishnan, Gareth Ball, Paul Aljabar, A David Edwards, and Giovanni Montana. Characterising brain network topologies: a dynamic analysis approach using heat kernels. **Neuroimage**, 141:490–501, 2016.
- [DNC⁺10] Nico U. F. Dosenbach, Binyam Nardos, Alexander Li Cohen, Damien A. Fair, Jonathan D. Power, Jessica A. Church, Steven M. Nelson, Gagan Wig, Alecia C. Vogel, Christina N Lessov-Schlaggar,

- Kelly Anne Barnes, Joseph W. Dubis, Eric Feczko, Rebecca S. Coalson, John R. Pruett, Deanna M. Barch, Steven E. Petersen, and Bradley L. Schlaggar. Prediction of individual brain maturity using fmri. **Science**, 329 5997:1358–61, 2010.
- [DTFV16] Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst. Learning laplacian matrix in smooth graph signal representations. **IEEE Transactions on Signal Processing**, 64:1–1, 12 2016.
- [FCG⁺19] Hamza Farooq, Yongxin Chen, Tryphon T. Georgiou, Allen Tannenbaum, and Christophe Lenglet. Network curvature as a hallmark of brain structural connectivity. In **Nature Communications**, 2019.
- [FSS⁺15] Emily S. Finn, Xilin Shen, Dustin Scheinost, Monica D. Rosenberg, Jessica S Huang, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable. Functional connectome fingerprinting: Identifying individuals based on patterns of brain connectivity. In **Nature Neuroscience**, 2015.
- [HK13] S. M. Hadi Hosseini and Shelli R. Kesler. Comparing connectivity pattern and small-world organization between structural correlation and resting-state networks in healthy adults. **NeuroImage**, 78:402–414, 2013.
- [MEGCF19] Hermina Petric Maretic, Mireille El Gheche, Giovanni Chierchia, and Pascal Frossard. Got: An optimal transport framework for graph comparison. In **Advances in Neural Information Processing Systems**, pages 13876–13887, 2019.
- [MTH⁺16] Jil Meier, Prejaas Tewarie, Arjan Hillebrand, Linda Douw, Bob W. van Dijk, Steven M. Stufflebeam, and Piet Van Mieghem. A mapping between structural and functional brain networks. **Brain connectivity**, 6 4:298–311, 2016.
- [MWH19] Ahmad Mheich, Fabrice Wendling, and Mahmoud Hassan. Brain network similarity: methods and applications. 2019.
- [NLG⁺15] Chien-Chun Ni, Yu-Yao Lin, Jie Gao, Xianfeng David Gu, and Emil Saucan. Ricci curvature of the internet topology. In **2015 IEEE Conference on Computer Communications (INFOCOM)**, pages 2758–2766. IEEE, 2015.
- [NLLG19] Chien-Chun Ni, Yu-Yao Lin, Feng Luo, and Jie Gao. Community detection on networks with ricci flow. **Scientific reports**, 9(1):1–12, 2019.
- [Oll09] Yann Ollivier. Ricci curvature of markov chains on metric spaces. **Journal of Functional Analysis**, 256(3):810–864, 2009.
- [PC⁺19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport. **Foundations and Trends® in Machine Learning**, 11(5-6):355–607, 2019.
- [PMPP⁺17] Vicente Ponsoda, Kenia Martínez, José Angel Pineda-Pardo, Francisco J. Abad, Julio Olea, Francisco J. Román, Aron K. Barbey, and Roberto Colom. Structural brain connectivity and cognitive ability differences: A multivariate distance matrix regression analysis. **Human Brain Mapping**, 38:803–816, 2017.

- [RBT⁺19] Benjamin Ricaud, Pierre Borgnat, Nicolas Tremblay, Paulo Gonçalves, and Pierre Vandergheynst. Fourier could be a data scientist: From graph fourier transform to signal processing on graphs. **Comptes Rendus Physique**, 20(5):474–488, 2019.
- [Rob12] Peter A. Robinson. Interrelating anatomical, effective, and functional brain connectivity using propagators and neural field theory. **Physical review. E, Statistical, nonlinear, and soft matter physics**, 85 1 Pt 1:011912, 2012.
- [RS10] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. **NeuroImage**, 52:1059–1069, 2010.
- [RSPH14] P. A. Robinson, S. Sarkar, Grishma Mehta Pandejee, and J. A. Henderson. Determination of effective brain connectivity from functional connectivity with application to resting state connectivities. **Phys. Rev. E**, 90:012707, Jul 2014.
- [SJB19] Jayson Sia, Edmond Jonckheere, and Paul Bogdan. Ollivier-ricci curvature-based method to community detection in complex networks. **Scientific reports**, 9(1):1–12, 2019.
- [STK05] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. **PLoS computational biology**, 1(4), 2005.
- [VESB⁺13] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. **Neuroimage**, 80:62–79, 2013.
- [Vil03] Cédric Villani. **Topics in optimal transportation**. Number 58. American Mathematical Soc., 2003.
- [VJP19] Manasij Venkatesh, Joseph Jaja, and Luiz Pessoa. Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. **NeuroImage**, 207:116398, 11 2019.
- [vRS05] Max-K. von Renesse and K. Sturm. Transport inequalities, gradient estimates, entropy and ricci curvature. 2005.
- [WDC⁺19] Xin Wen, Li Dong, Junjie Chen, Jie Xiang, Jung-Min Yang, Hechun Li, Xiaobo Liu, Cheng Luo, and Dezhong Yao. Detecting the information of functional connectivity networks in normal aging using deep learning from a big data perspective. In **Frontiers in Neuroscience**, 2019.
- [WJS16] Melanie Weber, Jürgen Jost, and Emil Saucan. Forman-ricci flow for change detection in large dynamic data sets. **Axioms**, 5(4):26, 2016.
- [XXG⁺19] Jie Xiang, Jiayue Xue, Hao Guo, Dandan Li, Xiaohong Cui, Yan Niu, Ting Yan, Rui Cao, Yao Ma, Yanli Yang, and Bin Wang. Graph-based network analysis of resting-state fmri: test-retest reliability of binarized and weighted networks. **Brain Imaging and Behavior**, pages 1–12, 2019.