Universitat Oberta de Catalunya uoc.edu

Multilayer approach to diagnose and classify Multiple Sclerosis phenotypes using graph theory measures

Joan Ginard Illescas

Master in Science in Data Science
Machine Learning in Medicine

Project supervisor
Eloy Martínez de las Heras
Coordinating professor
Ferran Prados Carrasco

Date of submission
25-06-2023

Universitat Oberta de Catalunya

SUMMARY OF THE FINAL PROJECT

Title of the project:	Multilayer approach to diagnose and classify Multiple Sclerosis phenotypes using graph theory measures
Author name:	Joan Ginard Illescas
Project supervisor:	Eloy Martínez de las Heras
Coordinating professor:	Ferran Prados Carrasco
Date of submission (MM/YYYY):	06/2023
Name of the degree:	Master of Science in Data Science
Topic of the final project:	Machine Learning in Medicine
Language:	English
Keywords:	Multiple Sclerosis, network analysis, classification

Abstract

Multiple sclerosis (MS) is a chronic disease that affects the central nervous system and is a leading cause of disability in young adults. Magnetic resonance imaging (MRI) is a key tool for disease diagnosis, but lesions seen on an MRI do not always correlate with disease progression, known as the "clinical-radiological paradox."

Network science has proved to be a powerful tool for characterizing brain connectivity patterns, and in this work, we propose using network connectivity measures to classify MS patients. We will construct a three-layer network per subject based on MRI data and obtain a set of measures to enable the application of machine learning algorithms to differentiate between healthy subjects and MS patients and distinguish patients with worse clinical outcomes.

During this process we will find most suitable connectivity measures and determine the usefulness of considering the network layers separately or integrating them into a multilayer network.

Index

List of Figures

List of Tables

1. Introduction

Multiple sclerosis is a chronic disease of the central nervous system and is the first non-traumatic cause of disability in young adults (D. T. Chard et al. 2021) It is characterized by inflammation, demyelination and progressive neurodegeneration (Haider et al. 2016).

Patients can evolve from a clinically isolated syndrome (CIS) into a primary progressive course (PPMS) or a relapsing-remitting course (RRMS), which will evolve into a secondary progressive course (SPMS) within in a period that may vary between 10 and 20 years (Kocevar et al. 2016).

It is of paramount importance to determine which patients will follow each disease course as early treatment can delay disease progression (Comi et al. 2009) and improve living standards of these patients.

1.1. Context and motivation

Magnetic resonance imaging (MRI) is established as a key tool for the disease diagnosis, although lesions shown by this test seem to have no direct correlation with the evolution of the disease (Chard and Trip 2017). Those who have more lesions on an MRI do not necessarily present more symptoms, which is known as the "clinical-radiological paradox." (Barkhof 2002).

Currently, there are no other tools that can measure the progression of the disease. Therefore, it is of great interest to find a system that can classify patients based on the available diagnostic tests, primarily MRI, according to the progression of the disease.

On the other hand, increasing size and complexity of neurobiological data is met with theoretical and computational advances in data analysis (Bassett and Sporns 2017). Network science has proved to be a powerful tool to cope with this challenge and to characterize brain connectivity patterns (Fornito, et al. 2016)

In this work, we propose using network connectivity measures to classify MS patients. For this purpose, we will use data obtained from MRI resulting in the construction of a 3 layer network per subject, with 76 nodes each layer. Based on the connectivity properties of the network, we will obtain a set of measures that will enable us to apply different machine learning algorithms, which we hope will help us distinguish patients with worse clinical outcomes, or at least to differentiate between healthy subjects and MS patients.

Previous studies have explored the feasibility of several machine learning models (Kocevar et al. 2016) or (Zhao et al. 2020), but few, to the best of our knowledge, have focused on using connectivity measures (Solana et al. 2019)

1.2. Personal motivation

One of the subjects that I have enjoyed the most during this Master Degree has been graph theory. That is why I decided I wanted to do my Master Thesis on something related to it. However, I did not want to explore typical examples in graph theory like as social networks or transportation.

While reviewing available thesis topics, my wife, a pediatrician, pointed out this one as the most original one.

I delved into the specific topic of the project and I found very attractive to be able to relate the functioning of the brain or model some aspect of it with a mathematical model such as a graph. As I learned more about this area, my interest grew, and I believe we are at an important moment in advancing our knowledge of how the brain works.

1.3. Goals

In this work we aim to investigate the potential of using network analysis and machine learning algorithms applied to MRI data to classify MS patients.

Our main goal is to find algorithms or an ensemble of algorithms that can classify individuals into healthy subjects and MS patients and discriminate patients in different stages of the disease.

We have additional secondary objectives that are either desirable or serve as preliminary milestones. Here the most prominent ones:

- Find out most suitable network connectivity measures for the task. This is not only relevant in improving the algorithm's subsequent performance but could also assist in the research of the disease itself. It should be noted that previous studies have already explored this selection process (Solana et al. 2018) or (Casas-Roma et al. 2022)
- Determine if it is more useful to consider the layers separately, or to try to integrate them into a multilayer network.
- Obtain results with algorithms that allow for interpretation and avoid “black box” algorithms

1.4. Sustainability, diversity and ethical/social challenges

The Ethical and Global Engagement Competence (EGEC) is defined at the Master’s level as follows: *“Act in an honest, ethical, sustainable, socially responsible and respectful manner with respect to human rights and diversity, both in academic practice and in the professional, and design solutions to improve these practices.”* It addresses three main dimensions: Sustainability, Ethical behavior and social responsibility and Diversity and Human Rights.

Our goal could potentially lead to improved living standards for those affected by MS so we can state this work main impact is on Ethical behavior and Social responsibility.

Diversity is also addressed in the sense that exists no differentiation or discrimination of patients by skin color, religion, sexual orientation or any possible source of discrimination.

However, patients sex and age certainly has to be taken into account as it is well documented that are gender and age differences in WM (Hsu et al. 2008). This does not mean we pursue results which apply to only one sex and age, in fact quite the opposite.

1.5. Approach and Methodology

A project like this has a previous step which is reviewing available literature. A thorough literature review is a valuable tool for ensuring the accuracy of the theoretical framework adopted, refining project goals, and applying relevant findings to the present research.

Data science projects typically follow a set of common stages, such as exploratory data analysis and data processing, feature selection, model creation and model assessment. In addition, there may be additional stages that are specific to the particular topic or domain of the project. It is important to note that the data science process is not strictly linear, but rather an iterative one. Additionally, the boundaries between the different stages can be blurry at times, as there is often overlap and feedback loops between them.

Accordingly project has been divided in four major steps: Data processing, Network connectivity measures, feature selection, model creation and comparison.

Data processing

Project's tutor provides the data. It is composed of a cohort of 147 patients and 18 healthy volunteers. As it has been noted before, data comprises a multi-layer network for each subject, encoded as 3 data matrix per subject. Those 3 matrices represent: structural white matter (WM) network, structural gray matter (GM) network and a resting-state functional network. For each subject some clinical information is available, including age, sex, disease duration, EDSS score, binary classification informing whether the subject is a patient or a healthy subject and the MS type they suffer.

Although data has already been processed in order to obtain the matrices, there are still some decisions to make. For instance, in GM and WM matrices there could be some connections that are not really present or in functional matrix there are negative correlations that would imply the existence of negative weights in our network.

Network connectivity measures

In this step, different connectivity measures for each network and patient will be calculated. Literature will serve as a guide to select the measurements to use.

Also in this step we will check whether it is convenient to work with a 3 layer network, combine it into one single and/or discard one or more layers if they are not meaningful.

Feature Selection

In order to optimize the performance of our models, we need to select the most relevant features. This can involve conducting statistical test on our data to determine the most informative variables, or using dimensionality reduction techniques to reduce the complexity of the dataset.

Model creation and assessment

This stage involves training and testing different models with the same set of train and test data. To assess model performance we will use metrics like accuracy, recall and F-scores among others.

To carry out the project, R and Python, via Jupyter notebook, will be used. In R we will use specific libraries to analyze and perform network measurements, like *igraph* and *muxViz* (De Domenico, Porter, and Arenas 2015), which is a library specially focused on multilayer networks. Besides those libraries we will also use *tidyverse* libraries.

Regarding to Python, data science most relevant libraries will be used: *Pandas, Numpy, Sckit-Learn, Matplotlib* and *Seaborn* and *Scipy*. In addition two specific libraries will also be used: *NetworkX (network measurments)* and *ComBat* to correct biases in our array due to the use of different scanners (Behdenna et al. 2021).

1.6. Schedule

The work plan is organized around a series of milestones, which will be completed in each Continuous Assessment Test. Each milestone is then further divided into smaller steps. As shown in the table below and in the Gantt Diagram (Fig. 1), the main phase of the project (phase 3) is based on the stages outlined in the previous section.

STAGE	START DATE	END DATE
1. Work planning	01/03/2023	12/03/2023
2. State of the art – Bibliographic review	08/03/2023	21/03/2023
2.1. Literature review	08/03/2023	17/03/2023
2.2. Draft	18/03/2023	25/03/2023
3. Work implementation	26/03/2023	27/05/2023
3.1. Data preprocessing	26/03/2023	12/04/2023
3.2. Network connectivity	13/04/2023	27/04/2023
3.3. Feature selection	28/04/2023	12/05/2023
3.4. Creation of Models	19/05/2023	27/05/2023
4. Writing Report	29/05/2023	25/06/2023
4.1. Draft	29/05/2023	11/06/2025
4.2. Final version	12/06/2025	25/06/2025
5. Project Defense	26/06/2023	02/07/2023
5.1. Slides and Video	26/06/2023	02/07/2023
5.2. Public Defense	Date to be determined	

Table 1. Project Schedule

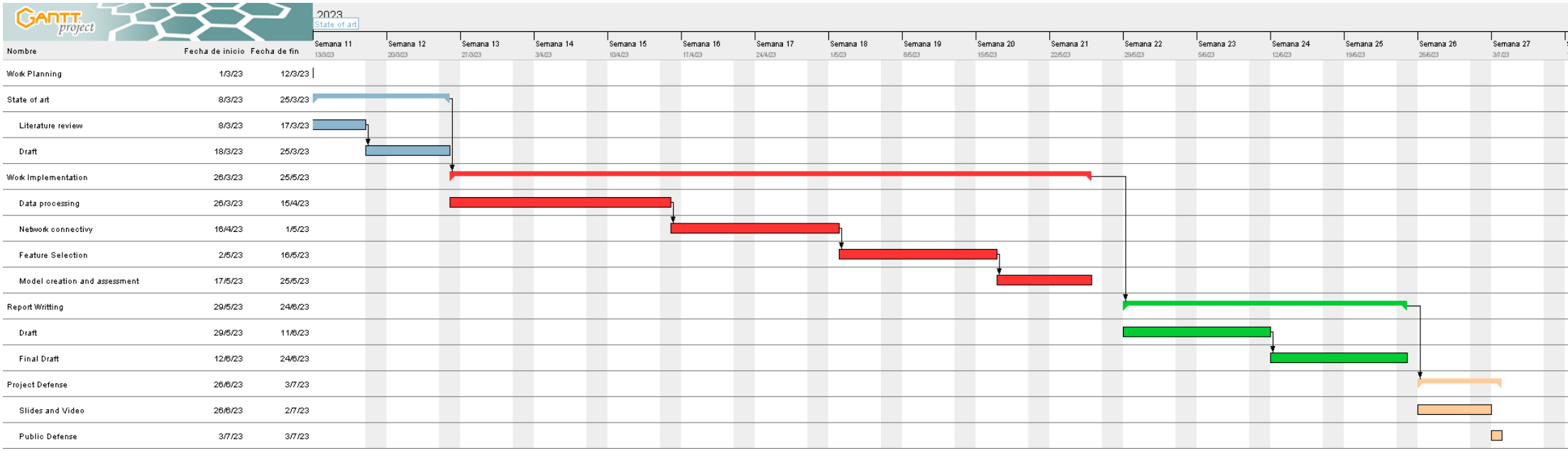

Fig 1. Project Gantt diagram

There are several reasons why planned schedules may be disrupted, including over-optimistic planning, illness, technological setbacks, and excessive workload at the workplace. That is precisely why we have allocated more time in Phase 3, as it allows for the review of the process thus far and adjustments to be made if necessary.

1.7. Summary of the outputs of the project

The primary outputs of the project include the following: the current document, a virtual presentation discussing the key findings, and a GitHub repository. The repository contains the code files needed to reproduce the different results with the corresponding tables and images available.

In particular in the code files (and described in repository README file), stages like data processing, network measures calculation and machine learning model development and evaluation can be found.

1.8. Brief description of the remaining chapters of the report

In Chapter 2,"State of the art", a comprehensive review of existing literature in the field is provided. This examination serves to delineate the current state of research

Chapter 3, "Graphs and input data", offers an introduction to graphs and their relationship to our dataset. Some basic concepts of both single and multilayer graphs are presented.

Chapter 4, "Methods and resources", outlines the methodology employed in the project, commenting the data sources, processes and some significant decisions made during the study.

In chapter 5, "Results" obtained are presented and discussed in detail. Following that, chapter 6 "Conclusions and future work", draws together the main conclusions and proposes potential directions for future research.

Final chapters, are devoted to "List of Abbreviations", "Bibliography" and "Appendices" that contains supplementary information, such as mathematical definitions for graph metrics and versions of Python and R libraries.

2. State of the art

As previously mentioned, Multiple Sclerosis (MS) is a neurodegenerative disease that poses a challenge when it comes to relating MRI-detected lesions to physical disability and cognitive impairment in patients. (Fleischer et al. 2016) and (Schoonheim, Meijer, and Geurts 2015) have suggested that during the early stages of the disease, a compensatory mechanism may exist that allows for the reorganization of brain functional networks to cope with disease progression. However, this mechanism is only possible when structural damage is not yet severe (Schoonheim, Broeders, and Geurts

2022). While the proposal of such a mechanism remains controversial (Schoonheim, Broeders, and Geurts 2022), it nevertheless suggests that there is something happening with brain networks in MS, which has led to the classification of MS as a network disease (Schoonheim, Broeders, and Geurts 2022; D. T. Chard et al. 2021). Hence, utilizing network analysis and graph-based measures to study MS, as this work proposes, is a valid and appropriate approach.

2.1. Brain networks

Although the focus of this work is on the characteristics of brain networks in individuals with MS, it is worth noting that a healthy brain is characterized by a combination of segregated and integrated processing. In a comprehensive review of the literature on brain structure and function, (Schoonheim, Broeders, and Geurts 2022) describe how a healthy brain is represented by high local clustering and short average path lengths between distant regions. The measures of network integration in this kind of network are characteristic path length and global efficiency, while segregation is quantified by modularity and clustering of local efficiency. This organization of networks is referred to as rich club organization (Heuvel and Sporns 2011; Fornito et al. 2016). In rich club networks, high-degree nodes (or network hubs) are more densely connected to each other than to lower-degree nodes.

As Fornito, et al. 2016 point out, there are 2 main networks studied in brain connectivity: Structural and Functional networks. Structural networks are based on the anatomical connections between different regions of the brain, while functional networks are based on the patterns of synchronized activity between those regions. This means that while the structural network provides information about the anatomical pathways that connect different brain regions, the functional network provides information about the strength and efficiency of the communication between those regions.

2.2. Single layer networks applied to MS

In their review, Fleischer et al. 2019 enumerated all network measures found in the literature up to that point (Table 2) to distinguish between healthy subjects and MS patients or between patients in different clinical stages of the disease. Some studies have focused on studying structural network disruption, such as (Kocevar et al. 2016; Shu et al. 2016; Llufriu et al. 2016), while others have focused more on functional networks, such as (Welton et al. 2020)

	Measures	**Interpretation**
Measures of centrality	Degree Centrality	The higher the value the higher the influence of the region
	Eigenvector Centrality	Higher values correspond to regions which are connected to regions that are central in the network
	Nodal Efficiency	A higher value indicates a higher

		ability of the region to propagate information with the other nodes
Measures of segregation	Clusteriing coefficient	Fraction of a node's neighbor that also neighbors. So, it will indicate an organization principle which is cost-efficient
	Transitivity	Variant of clustering coefficient
	Local efficiency	It shows the capacity of the network to transfer information between neighboring regions
	Modularity	Modules are densely connected nodes that are sparsely connected to the rest of the network. Increased values represents an optimized network in response to changing environments
Measures of integration	Global efficiency	Information transfer across the whole brain is more efficient
	Path length	An increase will show a lower ability to transfer information in parallel
Measures of network resilience	Assortativity	Increase describes brain ability to continue functioning as response to continuous damage.

Table 2. Graph Based Measures in literature.

Adapted from Fleischer et al. 2019

Among the first group, (Llufriu et al. 2016) observed an increase in Path Length and a decrease in Global Efficiency, which could indicate a disruption in network integration. (Shu et al. 2016) found a decrease in local and global efficiency. On the other hand, (Fleischer et al. 2016) found that, at least in the early stages of the disease, there is an increase in network clustering and modularity, which could be indicative of the compensatory mechanism mentioned previously.

According to (Schoonheim, Broeders, and Geurts 2022) it could be concluded that patients tend to show more segregated and less integrated structural networks overall, particularly in patients with cognitive impairment. In the same review, they pointed out that existing studies on functional networks are more complex and that hypothetical connections between network efficiency and cognition are less clear.

Some authors, such as (Pontillo et al. 2022), have concluded that to this date, there is no "hallmark of multiple sclerosis" in the sense that conflicting results still arise when studying the brain and multiple sclerosis as a single layer network.

2.3. Brain and multilayer networks

Multilayer networks are a relatively new approach in network analysis (Bianconi 2022), and their application to the human brain is even more recent. (Schoonheim, Broeders, and Geurts 2022) note that considering the brain as a multilayer network leads to

emergent properties that cannot be fully captured by analyzing individual layers separately. (Sporns 2018) predicts that the use of a multilayer framework is likely to become more widespread.

Regarding the brain, some studies explore the application or extension of single-layer measures to a multilayer setting, such as (Vaiana and Muldoon 2020; Mandke et al. 2018). Others have proposed models, such as the core-periphery organization from a multiplex point of view (Battiston et al. 2018). With respect to disease, it is worth noting that the disruption of the core-periphery structure has been studied in Alzheimer's disease (Guillon et al. 2019).

2.4. Multilayer networks applied to MS

Given what have been discussed about single layer networks, it is not surprising that (Pontillo et al. 2022; Schoonheim, Broeders, and Geurts 2022) suggest that multilayer networks may provide better insights into the organization of the brain and multiple sclerosis. This approach is so new that to best of our knowledge there are only four papers applying multilayer networks to multiple sclerosis. (Kennedy et al. 2023) used five biological layers, which are quite different from the data available to the present work, and (Martí-Juan et al. 2023) studied the relationship between functional and structural networks using a tool called The Virtual Brain. Therefore, the primary focus will be on the other two papers.

(Casas-Roma et al. 2022). examined a three-layer network with the same layers as in this study, including a GM morphological network, a structural brain network, and a functional network. In their approach, all nodes are the same across the layers, but each layer represents a different type of relationship between nodes. One of the main innovations in their study is the use of the GM structural network to represent interlayer connections between the other two layers. They employed global and local measures to describe the properties of the multilayer network, including Strength, Degree, Betweenness centrality, Closeness centrality, and local efficiency. The authors found that all MS patients had lower local efficiency, and most of them had lower closeness centrality and node degree.

In (Pontillo et al. 2022), the researchers also used a three-layer network similar to the one in the (Casas-Roma et al. 2022). study. However, they used a different approach by constructing a multiplex network, which is a type of multilayer network where nodes have a one-to-one correspondence between layers. This allows for the integration of different layers into a single layer. They measured coreness using the definition proposed by (Battiston et al. 2018) and also introduced a Coreness disruption index, which represents a global measure of core-periphery reorganization. They found that the weakening of the multiplex core-periphery structure depends on the disease phase and is associated with physical disability and cognition. They also noted that the modeling of different layers together is still a topic of debate, and new solutions may emerge in the future.

2.5. Machine Learning and MS

It is interesting to note that various machine learning models have been employed to predict or classify MS patients using different types of MRI data. This can be seen in the reviews of (Nabizadeh et al. 2022; Seccia et al. 2021). However, it is surprising that there are no ML algorithms applied to graph metrics, and as far as we know, only (Kocevar et al. 2016; Solana et al. 2019) have applied SVM to connectivity measures. To our knowledge, no one has applied ML algorithms to a multilayer network in the context of MS.

2.6. Conclusions from the state-of-art

After reviewing the literature, it is evident that analyzing MS as a network disorder and employing graph-based measures is a valid approach. Although, it appears that there is an increasing agreement that a multilayer network approach is necessary to fully capture the complexity of MS. However, there is still no consensus on the best way to model the different layers of the network.

This presents a challenge for the present work, as the main goal was to explore the use of network analysis and graph-based metrics to study MS. Given the current state of the field, it may be necessary to focus on developing and comparing different approaches to modeling the multilayer network.

3. Graphs and input data

3.1. Single-layer graphs and input data form

In mathematics a graph is a structure that captures relationships between objects. It is made up of vertices and edges. Each edge connects a pair of vertices, with the vertices serving as the objects and the edges representing relations between them.

There are different types of graphs. They can be directed or undirected. In an undirected graph each edge connects two vertices without specific direction, while in a directed graph, edges point from one vertex to another in a specified direction.

Graphs can also be weighted or unweighted. In a weighted graph, each edge is assigned a numerical value, known as weight. Depending on the context of the graph, weights can represent different things, such as distances or cost.

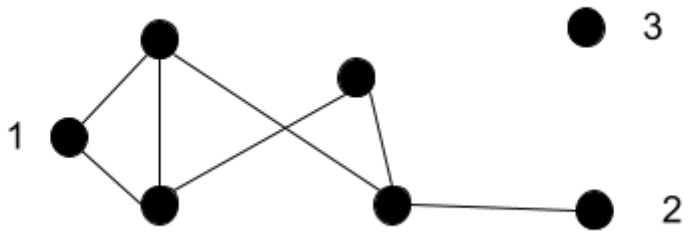

Fig 2. Example of a simple graph

In real-world applications, these mathematical graphs are often referred to as networks, with the vertices and edges renamed as nodes and links, respectively. However, these names are used interchangeably, and there is no strict rule dictating their usage.

Graphs can be represented by matrices. One of those matrices is the adjacency matrix. They are squared matrices representing connections between nodes, i.e. element in position i,j represents whether a relationship exists between node i and node j. Given a graph, G, its adjacency matrix $A = (a_{ij})$ is:

$$a_{ij} = \begin{cases} 1 \; if \; \exists \; an \; edge \; between \; node \; i \; and \; node \; j \\ 0, otherwise \end{cases} \quad (1)$$

Obviously in undirected graphs adjacency matrices are symmetric $a_{ij} = a_{ji}$.

If graph is weighted adjacency matrix is

$$a_{ij} = \begin{cases} \omega_{ij} \; if \; node \; i \; is \; linked \; to \; node \; j \; with \; weight \; \omega_{ij} \\ 0, otherwise \end{cases} \quad (2)$$

Where ω_{ij} is the edge of the weight connecting nodes i and j.

The input data takes the form, or could be interpreted as, weighted adjacency matrices, and thus interpreted as **undirected and weighted** networks. Each participant has one matrix per type of brain network (see section 4.2), 3 matrices in total.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16
1	0.0000000	0.3955671	0.0000000	0.0000000	0.0000000	0.3824387	0.3921771	0.4909023	0.0000000	0.3562779	0.0000000	0.3394741	0.3955272	0.0000000	0.3285685	0.3388446
2	0.3955671	0.0000000	0.3908107	0.0000000	0.3993008	0.3931075	0.4079483	0.4299556	0.3875951	0.3203500	0.0000000	0.3146355	0.4117320	0.0000000	0.3887990	0.3477685
3	0.0000000	0.3908107	0.0000000	0.3982200	0.3377938	0.3318887	0.3715550	0.4057192	0.2791992	0.3821787	0.2768503	0.3629678	0.3776060	0.3925704	0.0000000	0.0000000
4	0.0000000	0.0000000	0.3982200	0.0000000	0.2466624	0.3815739	0.2739551	0.3405749	0.3686030	0.0000000	0.3327310	0.0000000	0.2723174	0.2476177	0.0000000	0.0000000
5	0.0000000	0.3993008	0.3377938	0.2466624	0.0000000	0.3700973	0.3365139	0.4089708	0.3262968	0.3853814	0.3109272	0.3814012	0.3586648	0.2479760	0.0000000	0.3922712
6	0.3824387	0.3931075	0.3318887	0.3815739	0.3700973	0.0000000	0.3818969	0.3978764	0.3234102	0.4092844	0.3413446	0.3951709	0.3430973	0.4195099	0.3894821	0.3801424
7	0.3921771	0.4079483	0.3715550	0.2739551	0.3365139	0.3818969	0.0000000	0.4265760	0.3590941	0.3851834	0.3733197	0.3751427	0.3189608	0.3038593	0.0000000	0.4043695
8	0.4909023	0.4299556	0.4057192	0.3405749	0.4089708	0.3978764	0.4265760	0.0000000	0.4238879	0.0000000	0.3127143	0.0000000	0.4412924	0.3382747	0.3244210	0.4150146
9	0.0000000	0.3875951	0.2791992	0.3686030	0.3262968	0.3234102	0.3590941	0.4238879	0.0000000	0.3919756	0.2929028	0.3800123	0.3291073	0.4088283	0.0000000	0.3735961
10	0.3562779	0.3203500	0.3821787	0.0000000	0.3853814	0.4092844	0.3851834	0.0000000	0.3919756	0.0000000	0.3870152	0.2893733	0.3972814	0.3273961	0.0000000	0.3385863
11	0.0000000	0.0000000	0.2768503	0.3327310	0.3109272	0.3413446	0.3733197	0.3127143	0.2929028	0.3870152	0.0000000	0.3806801	0.3670595	0.2481466	0.0000000	0.0000000
12	0.3394741	0.3146355	0.3629678	0.0000000	0.3814012	0.3951709	0.3751427	0.0000000	0.3800123	0.2893733	0.3806801	0.0000000	0.3986426	0.0000000	0.0000000	0.3449569
13	0.3955272	0.4117320	0.3776060	0.2723174	0.3586648	0.3430973	0.3189608	0.4412924	0.3291073	0.3972814	0.3670595	0.3986426	0.0000000	0.4012147	0.0000000	0.4162551
14	0.0000000	0.0000000	0.3925704	0.2476177	0.2479760	0.4195099	0.3038593	0.3382747	0.4088283	0.3273961	0.2481466	0.0000000	0.4012147	0.0000000	0.0000000	0.0000000
15	0.3285685	0.3887990	0.0000000	0.0000000	0.0000000	0.3894821	0.0000000	0.3244210	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.3773880
16	0.3388446	0.3477685	0.0000000	0.0000000	0.3922712	0.3801424	0.4043695	0.4150146	0.3735961	0.3385863	0.0000000	0.3449569	0.4162551	0.0000000	0.3773880	0.0000000
17	0.3155368	0.3273883	0.0000000	0.0000000	0.0000000	0.3494350	0.3865658	0.0000000	0.0000000	0.3260983	0.3980971	0.3538678	0.3956479	0.0000000	0.0000000	0.3265707
18	0.3110019	0.3291985	0.0000000	0.0000000	0.3723179	0.3588899	0.3783199	0.4185516	0.0000000	0.3292225	0.0000000	0.3242423	0.3903443	0.0000000	0.3570988	0.2809069
19	0.0000000	0.0000000	0.2687262	0.3361815	0.3275459	0.3278713	0.3638053	0.4111633	0.2796271	0.3825269	0.2261392	0.3791775	0.3554826	0.3668764	0.0000000	0.0000000
20	0.4238250	0.3985462	0.0000000	0.0000000	0.3872252	0.3679960	0.3979624	0.3888070	0.3699909	0.0000000	0.0000000	0.0000000	0.3968617	0.0000000	0.2677327	0.3700924
21	0.4412755	0.4024135	0.0000000	0.0000000	0.0000000	0.4050306	0.4217114	0.3859337	0.0000000	0.0000000	0.0000000	0.0000000	0.4347333	0.0000000	0.3250864	0.3779773
22	0.3959998	0.3757150	0.3862442	0.0000000	0.4083014	0.3888624	0.4207095	0.4119613	0.3828929	0.3636515	0.0000000	0.0000000	0.4290655	0.0000000	0.3194129	0.3401566
23	0.4518925	0.4018281	0.3184170	0.3876748	0.3684271	0.3766337	0.3907858	0.3302396	0.3671429	0.4264825	0.1895665	0.4199782	0.4060235	0.3654822	0.3127781	0.3967154
24	0.3639365	0.3214786	0.0000000	0.0000000	0.0000000	0.0000000	0.4017102	0.4753933	0.3809726	0.2943138	0.3646769	0.2737196	0.3943428	0.0000000	0.3867349	0.3086390
25	0.3246084	0.3245639	0.3580042	0.0000000	0.3756715	0.3633965	0.3835710	0.4172639	0.3629144	0.3244038	0.3978492	0.3340053	0.3866901	0.0000000	0.3610654	0.3246968

Fig 3. Sample of input data.

Image shows a sample of input data: one matrix from one of the subjects. Each participant has 3 matrices of 76 rows and 76 columns (not all shown in this image), where numbers could be interpreted edges weights. For instance, number in position (4, V5) is the connection between nodes 4 and 5. ROIs corresponding to the nodes are encoded in another file.

In an undirected network, the degree of a node i, denoted as k_i, is the number of edges incident to it. As an example, in Fig 2, node 1 has degree 2, while node 2 has degree 1. For weighted graphs there is an analogous concept: the node strength. Rather than counting the number of links, it sums up all weight edges connected to the node.

In network analysis, a set of metrics is employed to assess and quantify certain aspects or properties of graphs. This assessment or quantification is necessary to compare different networks in numerical terms. Depending on the aspect or property being evaluated, metrics can be divided into those associated with the entire network, referred as global measures in this work, metrics associated with the nodes of the network (local measures in this document), metrics associated with a substructure or grouping of network nodes, and metrics associated with the edges or arcs of the network.

No single metric is universally applicable in the sense that multiple metrics are typically required for a comprehensive understanding of a graph's behavior. Metrics used in this text are defined in appendix 9.1. Table 2 provides a useful summary of these measures and their meaning.

3.2. Multilayer networks

Multilayer networks is a relatively recent development in the field (Bianconi 2022). This approach focuses on the interactions between different interconnected networks. For instance, in social interactions, each network can represent a different type of social

interaction (work, friends, family, etc.), see Fig 4. Alternatively, in transportation networks, each network might represent a mean of transportation (train, bus, metro, etc.). In these examples it is obvious that there can be “jumps” from one network to another, a person can receive a message from a colleague and pass it on to a friend, moreover it is also possible that a colleague is also a friend. Each network that composes this “superior network” is called layer, hence the name multilayer network.

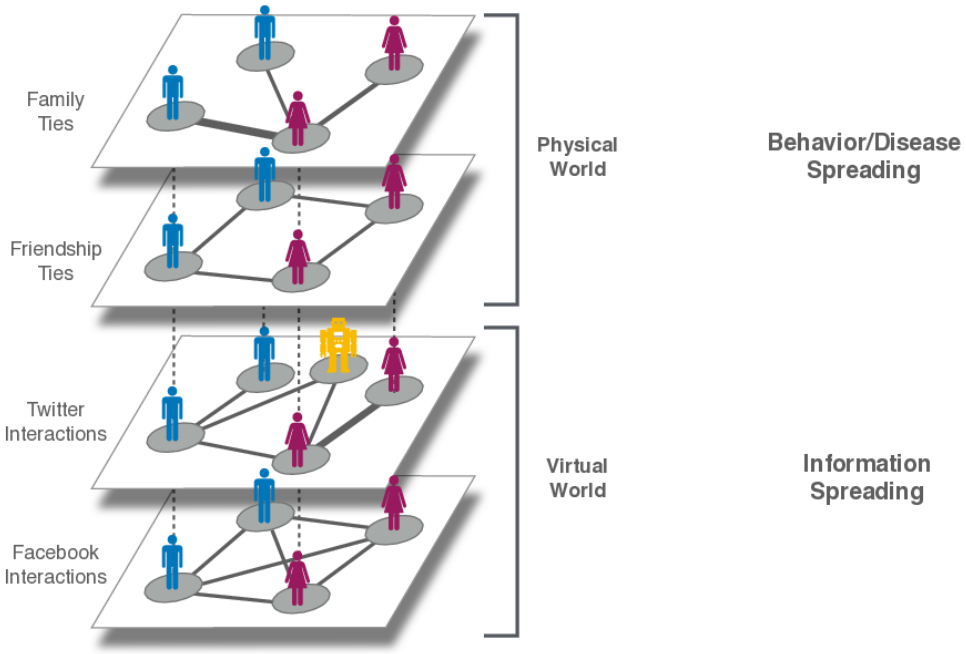

Fig 4. Multilayer social network.

Sample of a multilayer network. Each layer represent a different context, within a wider topic (Social interactions). A subject can interact with other people through different means (represented by the different layers). Alternatively, it may be possible that a subject can engage with certain people in some layers but not in others. Figure from (De Domenico 2020) under Creative Commons Attribution-ShareAlike 4.0 International License

An increasing number of studies have underscored the relevance of investigating the relation between structural and functional brain networks (Bullmore and Sporns 2009). As discussed in section 2.3., this can be addressed as a genuine multilayer problem where different brain networks constitute the layers of a multilayer network.

One kind of multilayer network, especially relevant to this case, is **multiplex network.** In this type of multilayer network, each layer has the same set of nodes, and these nodes are interconnected across different layers, see **¡Error! No se encuentra el origen de la referencia.**. Corresponding nodes belonging to different layer are called replica nodes.

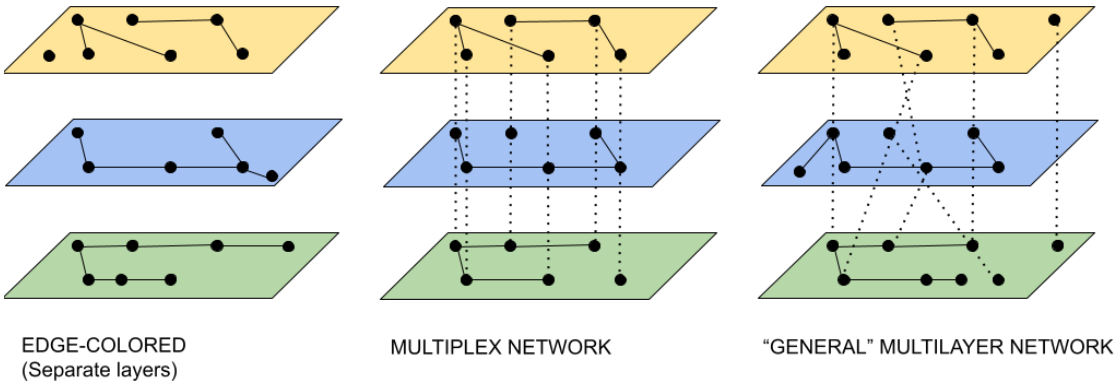

Fig 5. Types of multilayer networks

The image illustrates three examples of multilayer networks, first and second from the left are the ones of interest this work. The leftmost could be regarded not as a multilayer network but rather as three networks considered simultaneously, although contrary to depiction in this image, all layers in this work contain the same nodes. The central illustration shows a multiplex network where each layer has the same nodes, and connection between layers solely exists between these identical nodes, called replica nodes

Multilayer networks can be addresses in two main ways: using tensors (De Domenico 2022) or a **supra-adjacency matrix** (Bianconi 2022). In this work the latter approach has been employed.

The full information about a multiplex network composed of M layers is encoded in M distinct adjacency matrices, where the matrix for layer α is represented as $a^{[\alpha]}$. In undirected and weighted networks, these matrices take the form:

$$a_{ij}^{[\alpha]} = \begin{cases} \omega_{ij}^{[\alpha]} \; if \; node \; i \; is \; linked \; to \; node \; j \; in \; layer \; \alpha \; with \; weight \; \omega_{ij} \\ 0, otherwise \end{cases} \quad (3)$$

As nodes are only connected to their replica nodes, supra-adjacency matrices is given by:

$$\mathcal{A}_{i\alpha,j\beta} = \begin{cases} a_{ij}^{[\alpha]} \;\; if \; \alpha = \beta \\ \delta(i,j) \;\; if \; \alpha \neq \beta \end{cases} \quad (4)$$

where $\mathcal{A}_{i\alpha,j\beta}$ indicates whether node i in layer α is connected to node j in layer β. And δ is the Kronecker delta.

Fig 6 provides a clear visualization of how a supra-adjacency matrix is constructed in a multiplex network.

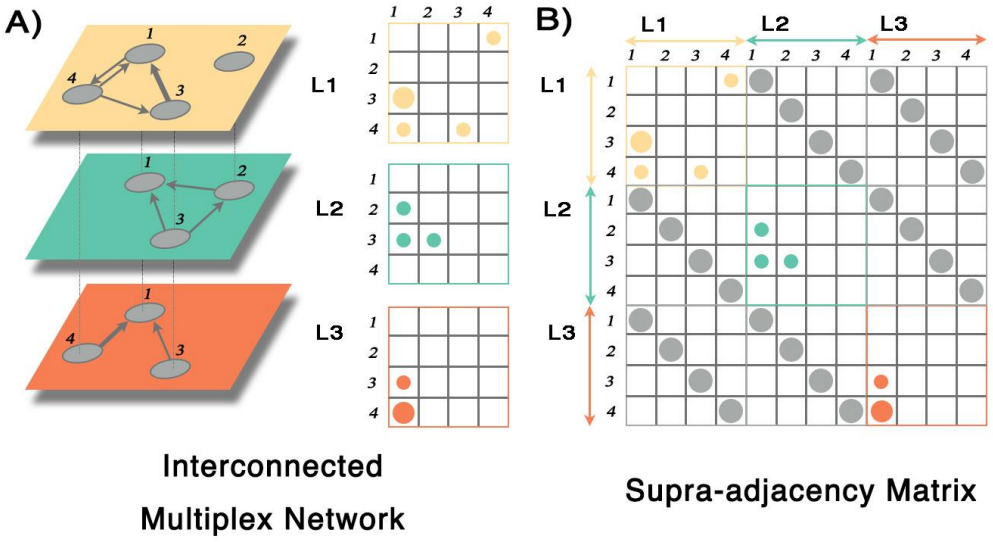

Fig 6. Multiplex network and Supra-adjacency matrix

Part A) depicts a situation with three independent networks, where L1, L2 and L3 are their respective adjacency matrices. B) graphically illustrates how these three matrices integrate to form a supra-adjacency matrix, with grey points representing the connection between layers. Figure from (De Domenico 2020) under Creative Commons Attribution-ShareAlike 4.0 International License

3.3. Edge-colored multigraphs and multiplex networks

In this project, as previously introduced, the main objective is to utilize machine learning models and graph measures to distinguish between HS and PwMS, and if feasible, further differentiate between various MS types. To accomplish this, different scenarios will be considered:

1. Edge-colored multigraph (*Separate Layers)*: participants' matrices form a multigraph without any connection between layers (see Edge-Colored Multigraph in Fig 5).
2. Multiplex network with 3 layers (*3-multilayer*): matrices are integrated into a multiplex network, creating a single, interconnected network.
3. Multiplex network with 2 layers (*2-multilayer*): same case as before but considering only two layers. Preliminary results showed no real contribution to relevant measures from GM network (see section 5.1), so it is decided to compare the 3 layer approach with 2 layer case.

Names enclosed in parentheses serve as the terminology used throughout this document to reference each scenario

With this cases, it will be possible to compare the results from the different approaches and to obtain an insight into how various graph measures vary between three separate layers and a unified network (see also Fig 9)

4. Methods and resources

4.1. Participants

The dataset encompasses data from 165 subjects, with ages spanning from 22 to 72 years. Among these participants, 18 where healthy subjects (HS), who volunteered for the study, and the remaining 147 were patients with Multiple Sclerosis (PwMS). This group can be further subdivided based on the type of MS: 6 with PPMS, 16 with SPMS and the remaining 125 with RRMS. Physical disability is assessed using the EDSS.

Below is a table summarizing data from all participants:

	PwMS (n = 147)	HS (n = 18)
Female, n	104	15
Age, years	47.3 ± 10.1	36.6 ± 9.6
MS type RRMS, n	125 (90 female)	----
MS type SPMS, n	16 (10 female)	----
MS type PPMS, n	6 (4 female)	----
EDSS, median (range)	2 (0 – 7.5)	----
Disease duration, year	14.1 ± 10.1	----

Table 3. Participants clinical, demographic and cognitive characteristics

Continuous variables are given as the mean ± standard deviation

This dataset is significantly imbalanced, which not only undermines the results from statistical tests but also poses challenges for the utilization of machine learning methods (see section 4.8). Additionally, this imbalance complicates the generalization of results.

Data distribution is shown Fig 7.

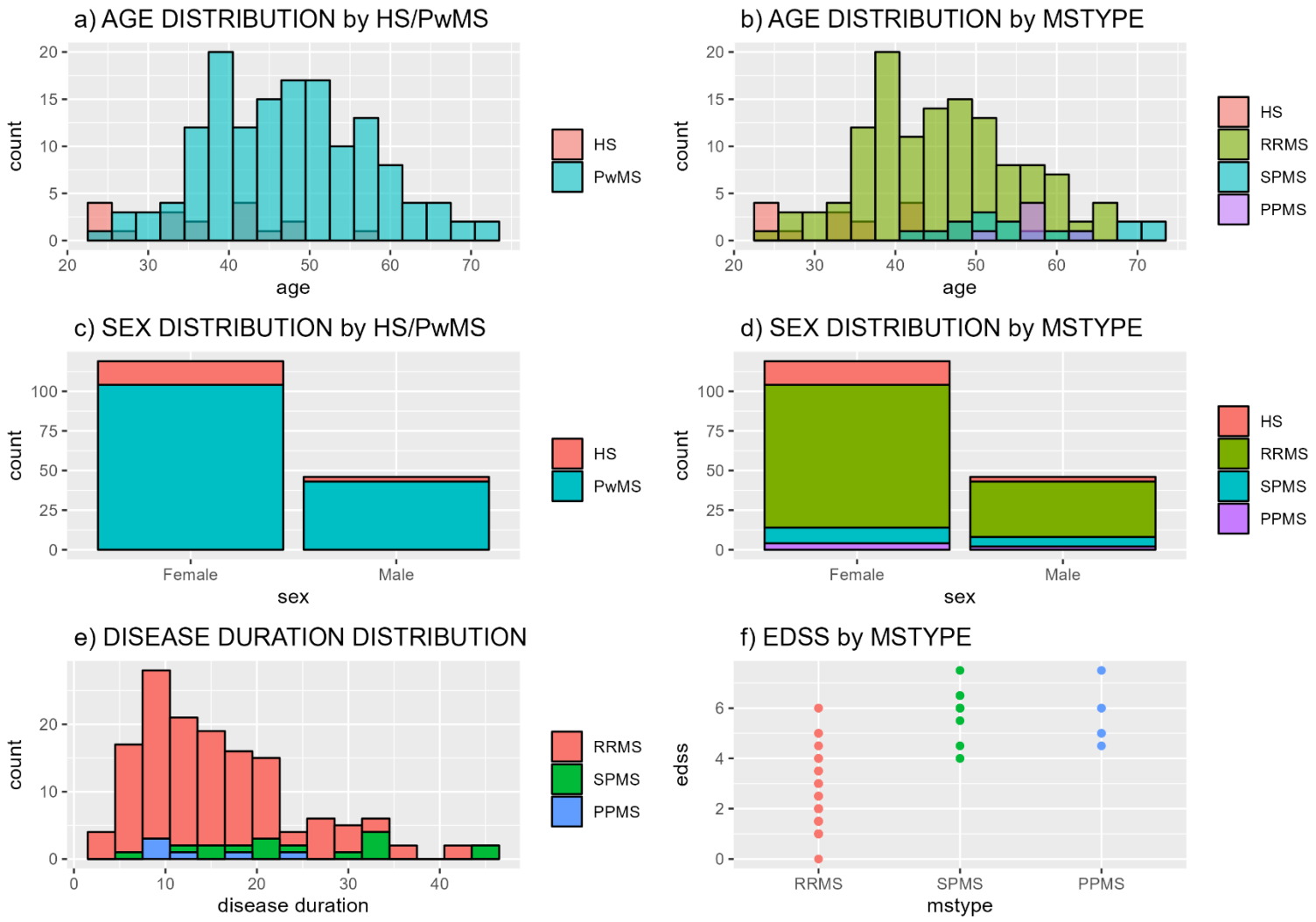

Fig 7. Participants data distribution.

The meaning of each color is found in the legend next to each plot. There is no particular meaning associated with color besides indicating MS type or whether a subject is HS or PwMS. From graphs a) and b), it can be inferred that HS group tends to be younger, while patients with more advanced types of MS lean towards the higher age spectrum (b). Graphs c) and d) clearly show a significantly higher representation of females. In graph e), the disease duration appears to cluster around the 10 to 20-year mark. Finally, graph f) illustrates how disability tends to exacerbate during the more advanced stages of the illness.

4.2. Brain networks and processing steps

As it has been discussed before there are 3 matrices per subject. Scanner data acquisition and preprocessing steps needed to obtain these matrices are beyond the scope of this work, but it is important to note that these matrices still require some additional processing and verification.

During data acquisition, the brain was segmented into 76 regions of interest (ROIs), or nodes, leading to 76 x 76 matrices (see Fig 8).

Preprocessing was conducted to ensure all matrices are symmetrical. Consequently, when interpreting these matrices as graphs, they will be recognized as undirected

graphs, given that the connection from node i to node j is identical to that from j to i. The main diagonal in matrices represents self-connections and will be removed if present.

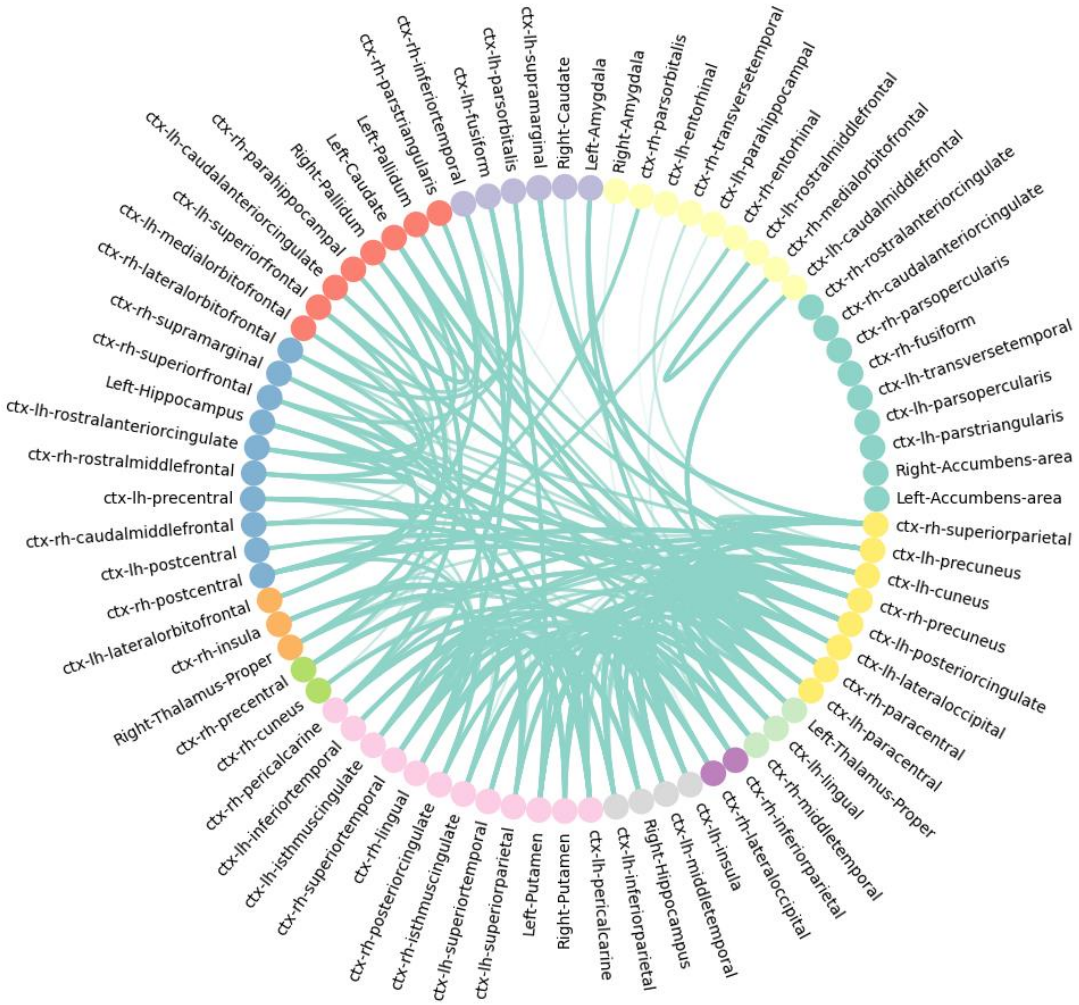

Fig 8. Parcellation scheme in circus plot

Image depicts ROIs and relevant connections (statistically significant differences between groups). Node colors correspond different groups. These groups are formed by ROIs with similar number of significant connections.

4.2.1. Structural brain network (FA network)

As described in section 2.1., this network maps the anatomical pathways connecting different regions of the brain. To generate matrices that embodies this network, the diffusion of water molecules is studied (hence the name diffusion MRI). Diffusion of water molecules in the brain is anisotropic, indicating that diffusion does not occur freely (and isotropically) but rather following pathways, running in parallel to the barriers imposed by brain structure. Fractional anisotropy (FA) quantifies how water diffusion is constraint in a given direction within a voxel.

Despite the preprocessing steps effectively minimizing or eliminating factors that could introduce noise in our data, a threshold is set to further ensure the elimination of all non-connections. This threshold is applied to data matrices to remove these spurious connections (Fornito, et al. 2016). However, since this could inadvertently remove true, albeit weak, connections, an additional criterion is used. If a connection is present in at least 60% of the healthy subjects (HS) - that is, 11 out of 18 - the connection is retained.

These conditions can be encapsulated as follows:

1. Eliminate all connections that fall below a threshold, which is set at **0.1**.
2. However, if these connections are present in at least **11 out of 18** HS, they should be retained.

The preprocessing steps proved to be precise, as no changes were observed after applying this condition.

Fig 10 shows a distribution of weights across all matrices, with a significant number of outliers at zero. This situation is partially rectified with the age and sex correction (see section 4.3). It is decided *not to impute* any value to remove these outliers, under the consideration that this values are likely to represent weak structural connections, thus any imputation will not greatly affect the overall data. Furthermore, brain graphs are sparse and this zeros are compatible with this kind of graph. Given that graph measures are being applied to the matrices/graphs, these values should not pose any numerical problems.

4.2.2. Structural gray matter brain network (GM network)

For each participant, there is another structural network. This network is derived from the similarities in gray matter (GM) morphological patterns according to the defined parcellation scheme. As mentioned before acquisition of this data, as well as parcellation scheme fall beyond the scope of this work.

The same threshold and conditions applied to the FA networks are also implemented for these matrices, and it is observed that this processing does not induce any changes.

4.2.3. Resting-state functional network (fMRI network)

Functional MRI is used to measure participants' brain activity during resting-state. After preprocessing, signal correlations between Regions of Interest (ROIs) form the matrices. As described in section 2.1. this network provides information about the communication between nodes.

As matrices are formed by correlation coefficients, its values range from approximately -1 to 1, with diagonal values representing self-correlations. When interpreting these matrices as the graph adjacency matrices, their values will serve as network edge weights. This requires to remove negative values, as they would imply negative weights. However, discarding negative correlations could lead to a loss of significant information (Fleischer et al. 2019).

Given that negative values also suggest a relation between ROIs, absolute value is taken (like in Battiston et al. 2017) to retain this information while setting the diagonal values to zero.

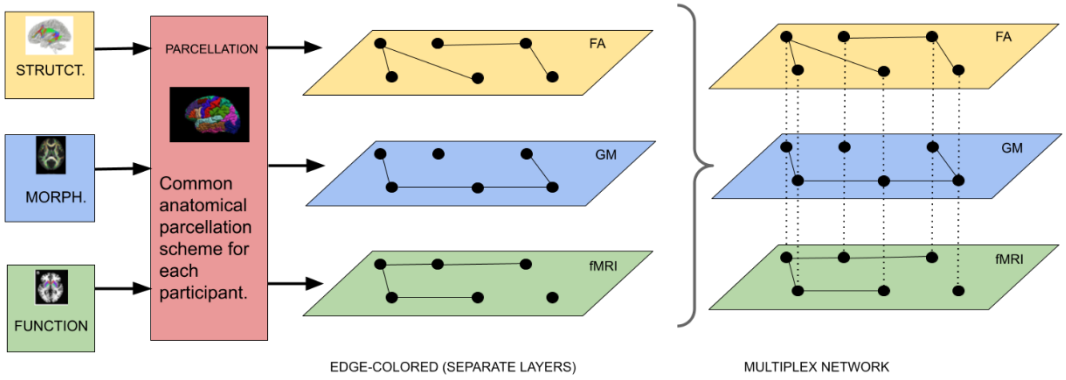

Fig 9. From MRI to networks

The figure illustrates the process of transitioning from original MRI data to networks. Initially, there are three sets of MRI data per subject, represented by the three squares on the left. Following this, a common anatomical parcellation process converts these data into the matrices that serve as the starting point of input data for this work. Each of these three matrices can be interpreted as a network, as shown in the central part of the illustration. These three networks can then be converted into a multiplex network, as depicted in the rightmost part of the figure

4.3. Age and sex correction

The brain, similar to many other biological systems or structures, undergoes changes with aging and exhibits gender-related differences (Hsu et al. 2008).

To control for age and gender, and allow comparison across participants, the matrix values are adjusted using a linear regression with age and gender as regressors. For each specific i,j position in all matrices, all 165 values are collected, one per participant, and linear regression is applied. Since the matrices are symmetric and the diagonal is zero, only one linear regression for each matrix position in the upper triangle has to be considered. Linear regression formula is:

$$\hat{y} = \propto * age + \beta * sex \qquad (5)$$

Where $\hat{y}$ are predicted values, α and β are the regression coefficients. The difference between actual value (y) and predicted value, is known as residuals.

$$\hat{y} - y = residuals \qquad (6)$$

These residuals account for information not explained by our regressors, i.e., sex and age. Therefore and theoretically, any changes caused by the disease will be captured in these residuals. The final value of i,j position in our matrix is our residuals plus and "standard" value for this position. This standard value should represent a normal brain, corrected for age and sex. However as this value is not uniform across all brains, the

mean of the healthy volunteers is considered as the standard value. Consequently, for patient m, the final value in position *i,j* is given by:

$$y_{i,j}^{final\ value} = residual_{i,j} + mean(HS_{i,j}) \quad (7)$$

Please note that the linear correction may result in values marginally below zero or slightly above one. Since negative values cannot represent graph weights, they are set to zero. However, values above one are retained, as they are infrequent and only marginally exceed one.

In the following figure, the differences in values distribution before and after accounting for age and sex are clearly shown:

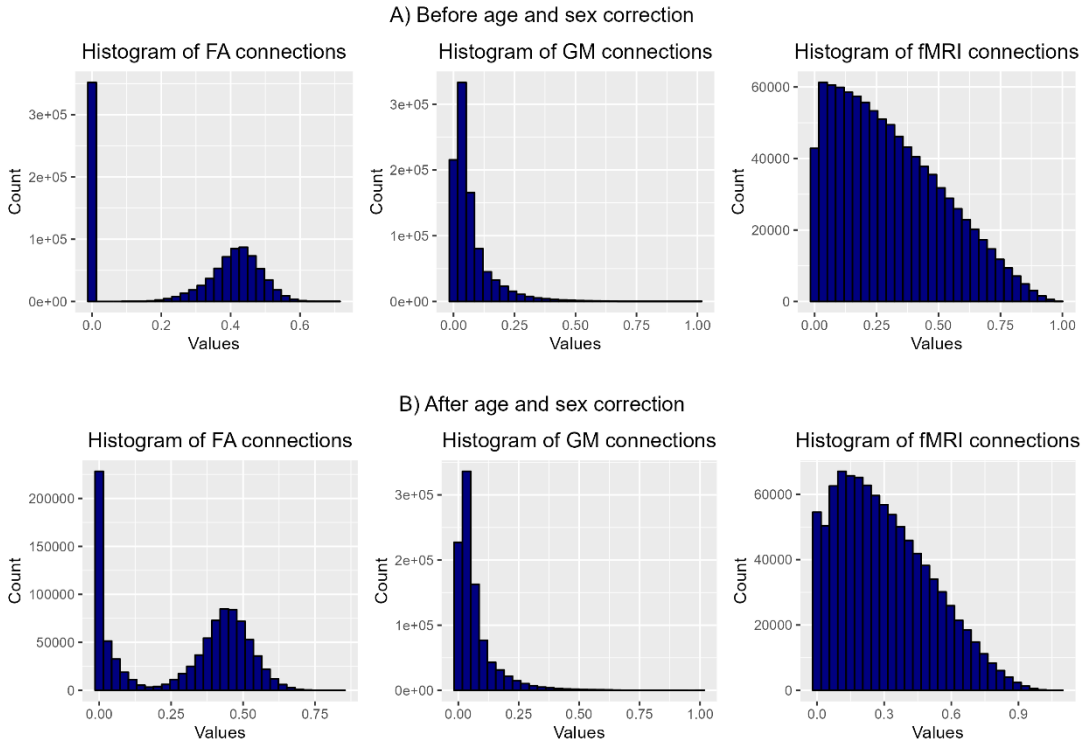

Fig 10. Weights distribution, before and after sex and age correction.

Differences are especially noticeable in FA connections, as depicted in the leftmost histograms of (A) before, and (B) after correction. In these FA histograms, the y-axis range clearly indicates that a significant quantity of zero values have been adjusted. Visible changes, though not as pronounced as in FA, are also present in fMRI connections, which are shown in the rightmost histograms. However, no substantial differences are apparent in GM connections, as seen in central plots.

4.4. Data harmonization using ComBat

Participants data have been collected using two different scanners. According to existing literature (Fortin et al. 2017), diversity in data acquisition methods may lead to an increase in the variance of matrix values.

To verify if there is an increase in variance, principal component analysis, PCA, is conducted on data (see Fig 11). The analysis indicate effect of data acquisition is only evident in FA data.

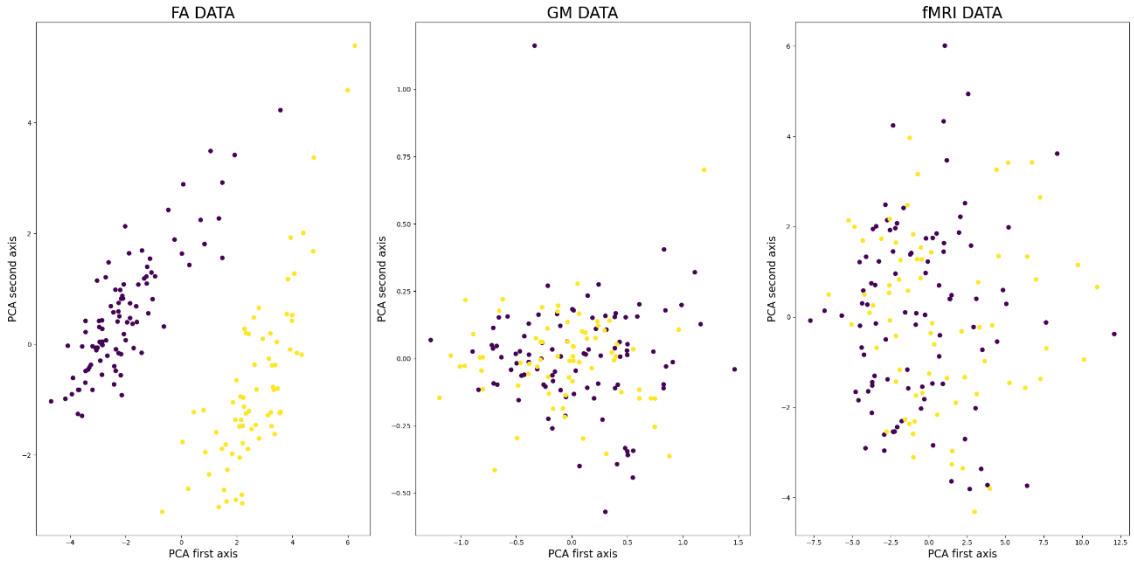

Fig 11. PCA before harmonization

Different colors in the plots denote whether data come from one scanner or the other. In central and right plot (GM and fMRI) points are mixed, thereby showing no effect of data acquisition. Conversely, the plot from the FA layer (on the left), clearly exhibits this data acquisition effect, as evidenced by the two discernible groups differentiated by color (and thus by scanner).

To overcome this problem neuroCombat library (Fortin et al. 2018) is applied to FA data. The following figure displays data changes in PCA for FA data.

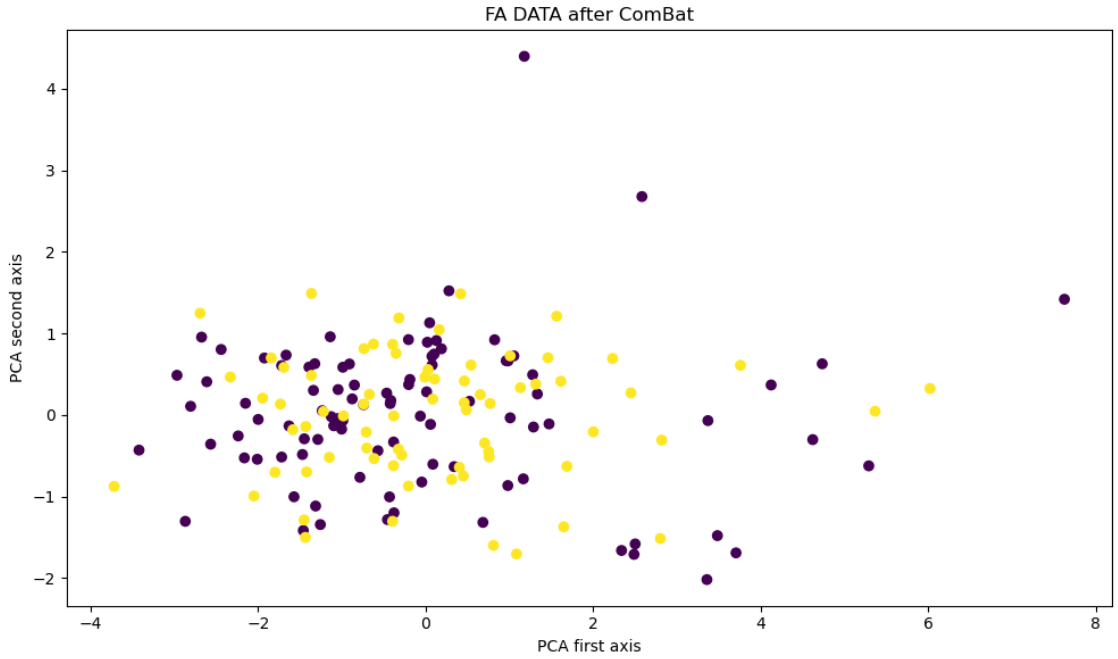

Fig 12. PCA with FA data after ComBat

Contrary to the FA plot in the preceding figure, this one- generated after the harmonization process-, exhibits no signs of data acquisition, with points appearing mixed similarly to the GM and fMRI data in the previous figure.

4.5. Graphs measures

Measures mathematical definitions can be found in appendices (see section 9.1)

4.5.1. SVD normalization

In order to integrate different layers in one single multi-layer network, like the multiplex network intended to use, it is necessary to correct differences in link weight across layers. According to (Mandke et al. 2018) uncorrected weights could introduce biases, and thus, it is advisable to apply a form of correction before conducting multilayer analysis

Singular Value Decomposition (SVD) technique for weight adjustment is used, as recommended by (Mandke et al. 2018) and also utilized by (Pontillo et al. 2022). If the adjacency matrix is represented as A, SVD can be applied in the following manner:

$$A = U\Lambda V^T \quad (8)$$

In this formula, U and V hold the left and right singular vectors respectively, and Λ represents the singular values of A. For link weight correction, rescaling using the largest singular value λ_1 is applied. Hence, the rescaled matrix is given by

$$\tilde{A} = U\left(\frac{10}{\lambda_1}\Lambda\right)V^T \quad (9)$$

A factor of 10 is incorporated to ensure the range of values is not too small and approximately varies between 0 and 1.

Fig 13 shows the changes in matrices weights before and after applying this SVD “normalization”.

SVD normalization results in two sets of measures for single layers (see next section). The first set corresponds to the separate layers case where normalization is not applied because, these layers essentially function as distinct networks. However, all measures from these separate networks will be collectively considered. The second set is relevant when the layers are integrated into a multiplex, a situation that necessitates this normalization.

Although it could be argued that normalization is also required in the separate layers case, it is decided to proceed without SVD correction in order to compare the two situations. In section 5.1 differences in results obtained in graph measures with and without SVD are discussed.

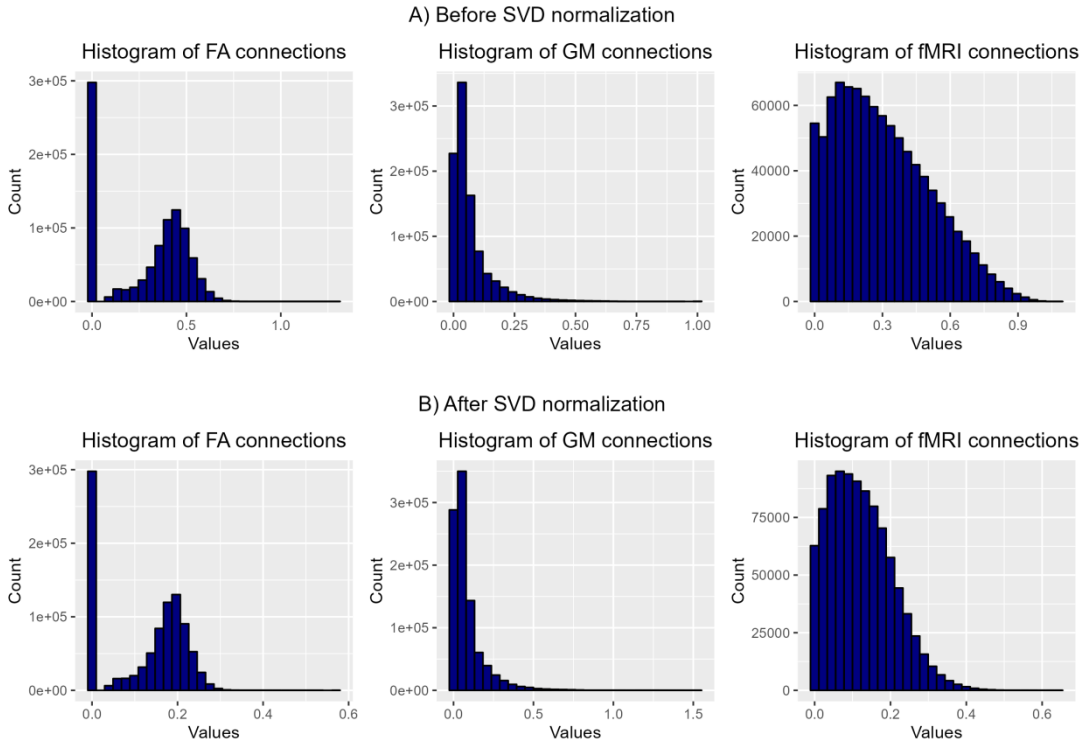

Fig 13. Distribution of weights before and after SVD normalization

Figure shows that values in FA (left) and fMRI (right) have been compressed into a more confined range after normalization (B), although the overall shape remains roughly the same. Conversely, in the GM connection, central plots, most noticeable feature is that values can escalate up to 1.5 after normalization, while there is also some compression in the lower values, it does not appear as important as in FA and fMRI layers.

4.5.2. Measures on single layers

Various graph measures are carried out, both global and local, on the separate layers case. These measures will also be applied to each layer within the multilayer network. As discussed in the previous section, since matrices in the first scenario have not undergone SVD correction, this approach will show how such correction impacts the results of graph measurements.

Programming libraries are employed to directly perform all measurements. This approach removes the need to predetermine which measures to include. These libraries offer comprehensive functions and methods that allow for a wide range of network measures to be calculated. This ease of use enables exploration of numerous measures without the need for complex coding.

However, it has been decided to include at least those measures that previous authors have identified as relevant (see section 2.2). It was also decided to incorporate density, even though it is less likely to yield distinctive results, given that the brain is a sparse network, making changes due to illness challenging to detect using this metric.

For global measures following measures have been considered:

1. Global Efficiency
2. Mean Path Length
3. Transitivity, global clustering
4. Diameter
5. Modularity
6. Global Strength
7. Density

In the case of local measures the following have been taking into account:

1. Efficiency
2. Closeness Centrality
3. Strength

Many measures are somehow related. Global efficiency is approximately the inverse of the average shortest path, and local efficiency and closeness centrality are clearly related (see mathematical definitions in the appendices). As a result it is likely that there will be some highly correlated graph measures, which will need to be removed prior to applying machine learning models. However, before that step, statistical analysis will be conducted to identify which measures exhibit statistical differences between the groups (see sections 4.6, 5.1 and 5.2)

4.5.3. Measures on multiplex graph

The measures that consider the entire graph and not just individual layers as in the previous section, are referred generally as measures on multiplex or multilayer graph

In this case, preliminary results showed that global measures were not statistically significant. Thus, to reduce code execution time, it was decided to focus solely on local measures.

Three measures of centrality have been applied

1. MultiPageRank centrality
2. Sum of total multi-Strength (a multilayer “version” of strength)
3. Closeness Versatility (closeness centrality for multilayers on some plots or tables “MultiPath” due to the name programming library muxViz gives to the functions that performs the measure)

These three measures have been chosen, while disregarding others, for two main reasons: to my knowledge they have not been used in previous studies, and they showed variability as inter-layer weights changed (see following section).

4.5.4. Inter-layer weights

In a multiplex network, all nodes in one layer are connected to their replica nodes in other layers. However, what should be the weight assigned to these inter-layer links (or interlinks)?

Literature does not provide standard methods for choosing these interlink weights: most of the techniques use some sort of *a posteriori* method, choosing weights (or a

parameter related to the inter-layer edges or the layer) that maximize an output of interest (Battiston et al. 2018, Vaiana and Muldoon 2020, Guillon et al. 2019). Only in Mandke et al. 2018, where the inter-layer weight is chosen as the average of weights in layers, an explicit a priori method can be found.

Given this, both approaches are tested (results are shown in 5.3):

1. Interlink weight as the average of weights
2. A comprehensive search of links that maximize statistically significant measures.

4.6. Statistical test and correlations

Statistical tests on measures are conducted to keep only those measures that show statistically significant differences between groups. This step is not strictly required but assists in reducing the number of features fed to the machine learning models. Furthermore, it helps to identify measures and nodes that are more important to differentiate between groups.

The statistical testing process is as follows:

1. The Shapiro-Wilk test is applied to ascertain if the data follows a normal distribution.
2. If data does not conform a normal distribution the Mann-Whitney-Wilcoxon test is applied.
3. If data does follow a normal distribution, Bartlett test is uses to check for homoscedasticity before proceeding to the t-test
4. The previous test were conducted to distinguish between HS and PwMS. To further refine the feature selection, Bonferroni correction is employed to identify measures that show statistical differences between the MS types.

The initial 3 steps form what, throughout the document, is referred to as the 'first test' or 'test 1'. Meanwhile, step 4 will be referred to as ‘second test’, ‘test 2’ or simply as ‘Bonferroni’.

One note of caution is that our dataset has very reduced number in some groups and this undermines results from statistical tests

It is important to note that our dataset contains a very small number of samples in certain groups (Table 3), which potentially undermines the reliability of results derived from statistical tests

Results are shown in section 5.1, 5.2 and 5.3.

4.7. Network models for machine learning algorithms.

This section is devoted to explain how measures to be included in the different models are chosen.

For the scenario where each layer is considered separately (Separate Layers), data from measurements without SVD correction was utilized. As layers are considered individually, this should be the case, take the measures as they are with no correction.

In the case of multilayer configurations (either 2 or 3 layers), both single layer measurements and multilayer measurements are contemplated. However, for this instance, single layer measurements with SVD correction are employed. It is important to note that single layer measures can be made on any network, as they only contemplate one layer at a time.

Measurements to include in each case will depend on the statistical tests (see Section 5.4). As commented in the previous section, the purpose of conducting these tests is to select most suitable features for the machine learning models. Depending on the results, it will be decided whether to include only those that pass test 1 or those that also pass test 2.

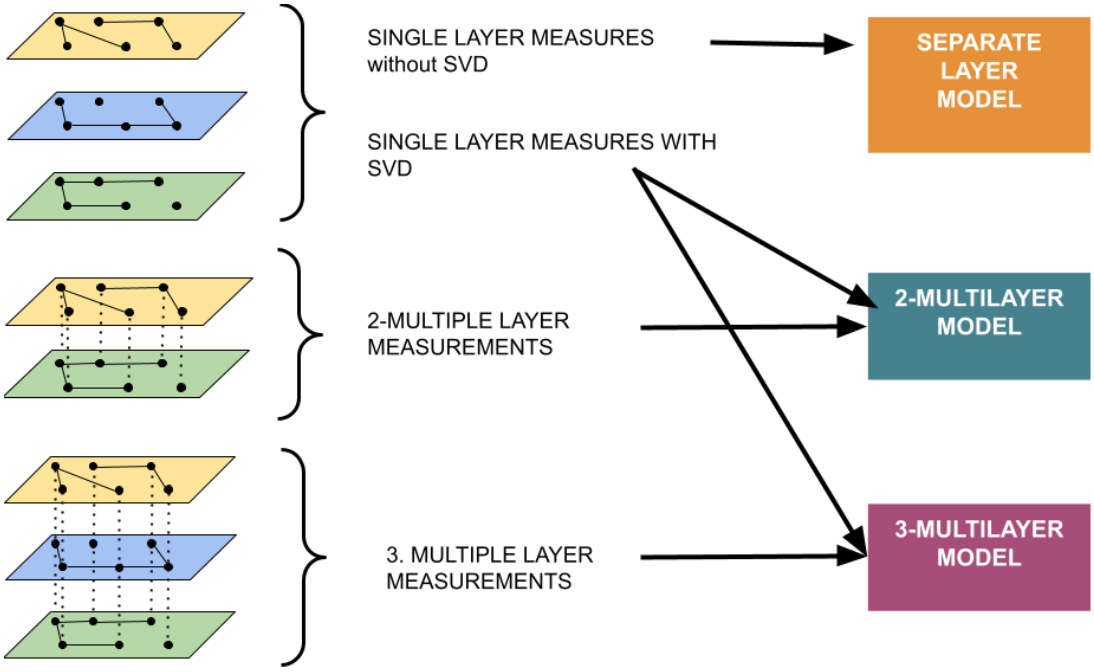

Fig 14. Measures to create the models.

The figure illustrates the process of constructing the models for the machine learning. On the left, the three networks are depicted. Single layer measures are extracted from the networks with separate layers, and multilayer measures are derived from the corresponding networks. Single layer measures are also included in the multilayer models, however, they are not necessarily derived from the corresponding multilayer network, as they only apply to individual layers.

Some authors (Battiston et al. 2018, Vaiana and Muldoon 2020) suggest that each layer can have a different weight in the overall multilayer network, though in this work this path has not been explored, and all layers are considered to contribute equally to the final network.

4.8. Data augmentation

One well known problem in classification task is the issue of imbalanced datasets, where there is a majority class significantly more represented that the other(s). In this project, the majority class is PwMS when differentiating control subjects from participants with MS, and RRMS when differentiating MS types. This imbalance can lead to scenarios where a model primarily predicts the majority class, resulting in artificially high accuracy despite poor predictive power for the minority class. This is because the model may struggle to identify patterns in the minority class due to insufficient sample size.

In order to solve this, augmentation techniques that overrepresents the minority class are used. One such method is Synthetic Minority Oversampling Technique (SMOTE) proposed by (Chawla et al. 2002). Briefly, SMOTE algorithm works by randomly selecting an example from the minority class and then one of its nearest neighbors. It then draws a line in the feature space between these two samples and creates a new instance in a point along the line (Brownlee 2021).

One drawback of this method is that since it generates new samples based on existing ones, it must be applied after partitioning the dataset into training and testing sets. This is to avoid having very similar samples in both sets, which could affect model performance. However, applying SMOTE after this split further limits the availability of minority class instances for the algorithm. This can potentially compromise the diversity of the minority instances, and subsequently the robustness of the model.

Another problem is that SMOTE works better when used in conjunction with under-sampling of majority class (Brownlee 2021). This approach might be suitable for large datasets, but it seems impractical for our case with just 147 instances of the majority class. Consequently, it was considered that the application of under-sampling will further reduce the already limited data.

4.9. Machine learning models

One of the major goals in this work is the interpretability of the machine learning models outcomes. As such, this study consciously avoids “black box” algorithms that do not provide insights into the significance attributed to different features.

The task in this study is classification: therefore, unsupervised machine learning models and regression models are not considered. There’s a plethora of models for classification to choose, like decision trees, logistic regression, probabilistic methods (like Naïve Bayes), KNN, SVM and combination of classifiers. As discussed in the work of Gironés et al. 2017., the state of the art in data mining is either focused on the development of new algorithms (which seldom occurs) or on improving existing ones and creating new models by combining simpler models. Consequently, it seems reasonable to commence by selecting from these "improved" methods.

Among this “combination” models there are ensemble algorithms like Random Forest and boosting algorithms like Gradient Boosting, both of which are based on the fact that a combination of decision trees delivers a superior performance. This combination of decision trees are achieved in different ways in ensemble and boosting.

It is decided to include two additional models to those mentioned, with the aim of increasing the diversity of algorithms considered. This will help to determine whether models with differing approaches can deliver enhanced performance.

The first additional model selected is the widely utilized *Support Vector Machines* (SVM or SVC for the case of a classifier) (Gironés et al. 2017). SVM has also been the model of choice in different studies on MS (Zurita et al. 2018, Kocevar et al. 2016, Solana et al. 2019), and it is chosen as a way to compare a well-established method in the field with the ensemble and boosting models.

And as a second additional method, it was decided to apply a relatively simple algorithm, K-Nearest Neighbors (KNN), in part because it is not found in literature, so it could be taken as a baseline to compare with the theoretically superior methods.

To recap, 4 models have been utilized:

1. Random Forest Classifier
2. Support Vector Classifier
3. KNN Classifier
4. Extreme Gradient Boosting Classifier (XGBoost)

4.9.1. Metrics

To evaluate machine learning models performance metrics that are not only the most sensitive for our case but also the most commonly used in literature (Gironés Roig et al. 2017) have been employed:

Accuracy	$ACC = \frac{TP + TN}{FP + FN + TP + TN}$	*(10)*
Precision	$PRE = \frac{TP}{FP + TP}$	*(11)*
Recall	$REC = \frac{TP}{FN + TP}$	*(12)*
F1-Score	$F1 = 2 \times \frac{PRE \times REC}{PRE + REC}$	*(13)*

Where TP, TN, FP, FN stand for True Positive, True Negative, False Positive and False Negative, respectively.

All these metrics are derived from the confusion matrix which is a contingency table representing true values and predicted ones, as illustrated in the following image.

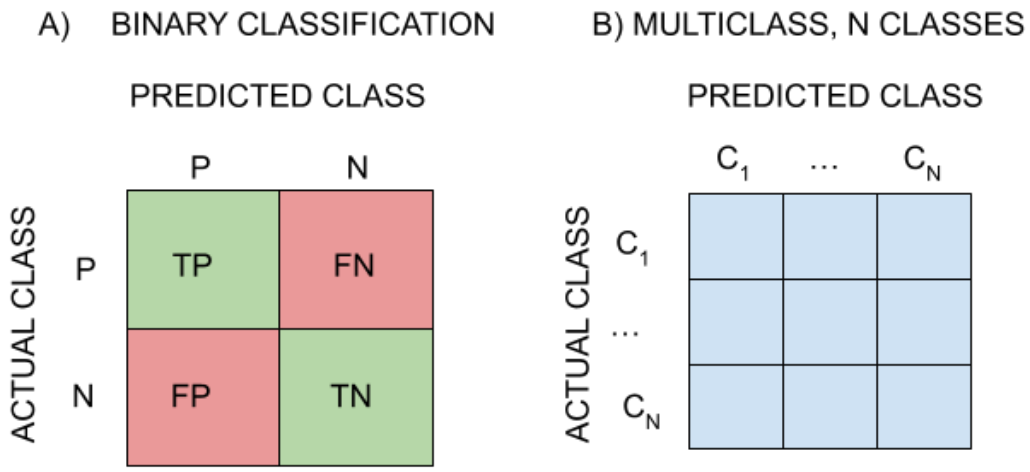

Fig 15. Confusion matrix definition

Figure A) represents a confusion matrix for binary classification, in green color, cells corresponding to correctly classified values, TP and TN, and in red, cells corresponding to errors, FN and FP. B) illustrates a confusion matrix for a multiclass classification scenario, where, as in binary case, in the matrix diagonal true values are located, while off-diagonal correspond to errors.

4.9.2. Grid Search and K fold

The identification of the best hyperparameters for each model was facilitated through a grid search. This method evaluates the model by testing every combination of possible parameters provided and returns the combination that yields the best value for the chosen metric - in this case, the F1 score, see definition in the Metrics section.

The model's training utilizes a K-fold approach. In this procedure training set is randomly divided into k number of observations of approximately equal size and the model is trained on k-1 folds and validated using the remaining one, and this process is repeated k times; each time a different fold is treated as a validation set. (James et al. 2013)

4.10. Feature importance

Constructing a model that makes accurate predictions is important, but identifying which measures and nodes play a key role in the model is equally significant. Therefore, the final part of model construction is dedicated to explore feature importance, which pinpoint the nodes and measures most essential for prediction.

To identify the importance of the features a technique implemented in Scikit-Learn, called *Permutation Feature Importance* will be used. Initially, the model is trained using the training set, and a score with set data is obtained. Then one feature within the model is shuffled and the test data is passed to get a new score. If the feature is important, the model performance will deteriorate and the score will drop. Otherwise the score will remain fairly the same.

Owing to time constraints, a comprehensive analysis of all models and scenarios is not feasible. Consequently, attention will be concentrated on ensembles and boosting, specifically Random Forest and XgBoost models.

4.11. Software and libraries

The initial part of the project which encompasses preprocessing, age and sex correction and graph measures was carried out using R, version 4.2.3. The igraph package was employed for single layer measures, whereas muxViz was used for multilayer measures. However there was one notable exception in this part, to conduct data harmonization, Neurocombat library for Python was used.

The final part of the project, involving machine learning models, was conducted using Python, version 3.10.10, and Scikit-Learn library. Additionally, customary libraries like Pandas, Numpy, Matplotlib and Seaborn were also utilized. Some other specific libraries used in this part were imblearn for data augmentation and XGBoost for this classifier

In order to perform other specific tasks, like plotting results other libraries were needed. Complete list of libraries and versions can be found in appendix 9.2

5. Results

5.1. Global single layer graph measures, with and without SVD correction

This section examines the result of applying the network measures to single layers, while comparing how the result depends on whether SVD normalization is applied or not. The noticeable change in the range of values of the weights when SVD correction is applied, has already been observed (see Fig 13). The focus now is to verify if this translates into a difference in the measurements of the graphs.

Fig 16, Fig 17 and Fig 18 display boxplots with the local graph measures and a comparison between case a) without normalization and case b) with SVD correction. Diameter and average path length are related and approximately inversely proportional to efficiency so this measures trend will be approximately correlated (see section 3.1 and appendix 9.1).

For the **FA layer,** the images for the case without SVD, show a decrease in average path length and diameter and an increase of efficiency for PwMS, There is also a clear decrease in global strength, modularity shows an slightly increase, while transitivity and density show no differences between groups. Comparing these sets of boxplots with the case where SVD is applied, it can be observed that the differences smooth out.

In the case of **GM layer** (Fig 17) it is apparent that there are no clear differences between groups in either case. This observation anticipates what will be a trend found throughout this project, suggesting that GM layer may not provide discriminatory information.

The plots corresponding to the functional layer (Fig 18) exhibit a behavior opposite to that of the FA layer when SVD is applied. In this case, the correction process amplifies the differences between groups, making them more evident. Another really interesting differences between FA layer and fMRI layer is the reversal of trends observed in almost all measures, where an increase for PwMS is noted in FA layer, a decrease is observed in the fMRI layer.

FA LAYER

a) WITHOUT SVD NORMALIZATION

b) SVD NORMALIZATION

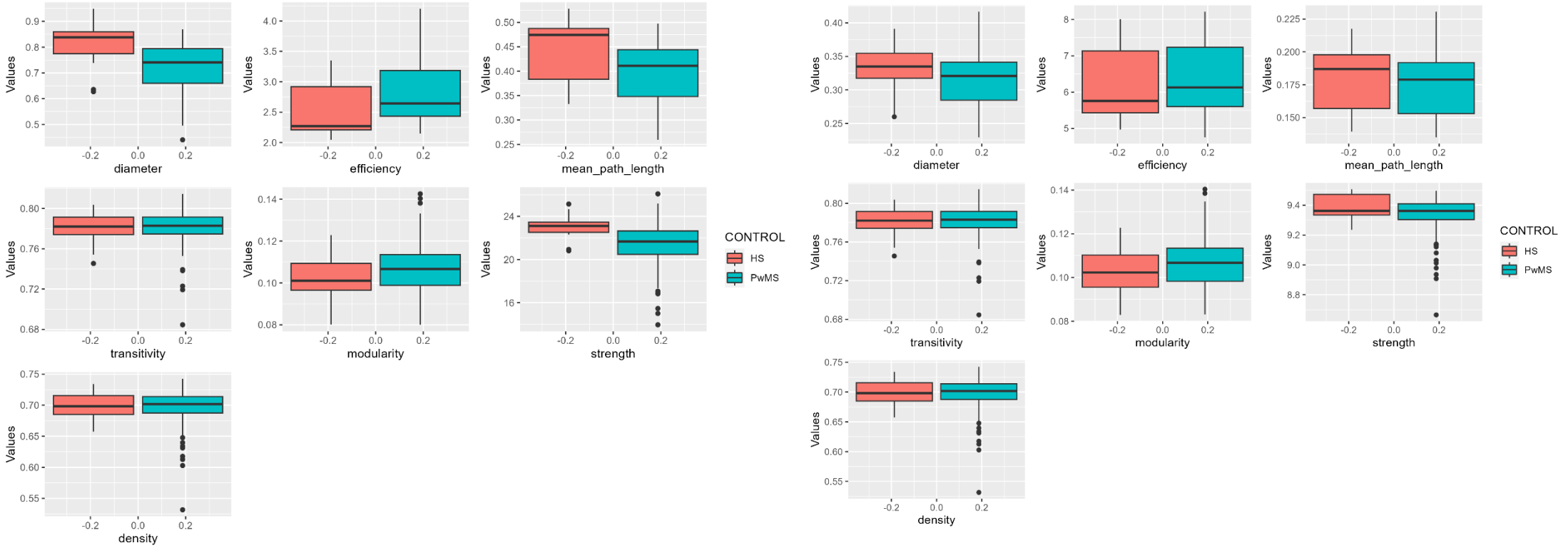

Fig 16. Boxplot graph global measures. FA Layer. With and without SVD.

Figure a) demonstrates that the measures of transitivity and density do not exhibit significant differences between groups, HS in red and PwMS in green, with only minimal disparities noted in modularity. Other measures do appear to present differences. However, in Figure b), these differences seem to be less distinct or even vanish, particularly in the case of efficiency and mean path length. Notably, while the box plots for strength visually indicate differences, the means are essentially equivalent.

GM LAYER

a) WITHOUT SVD NORMALIZATION

b) SVD NORMALIZATION

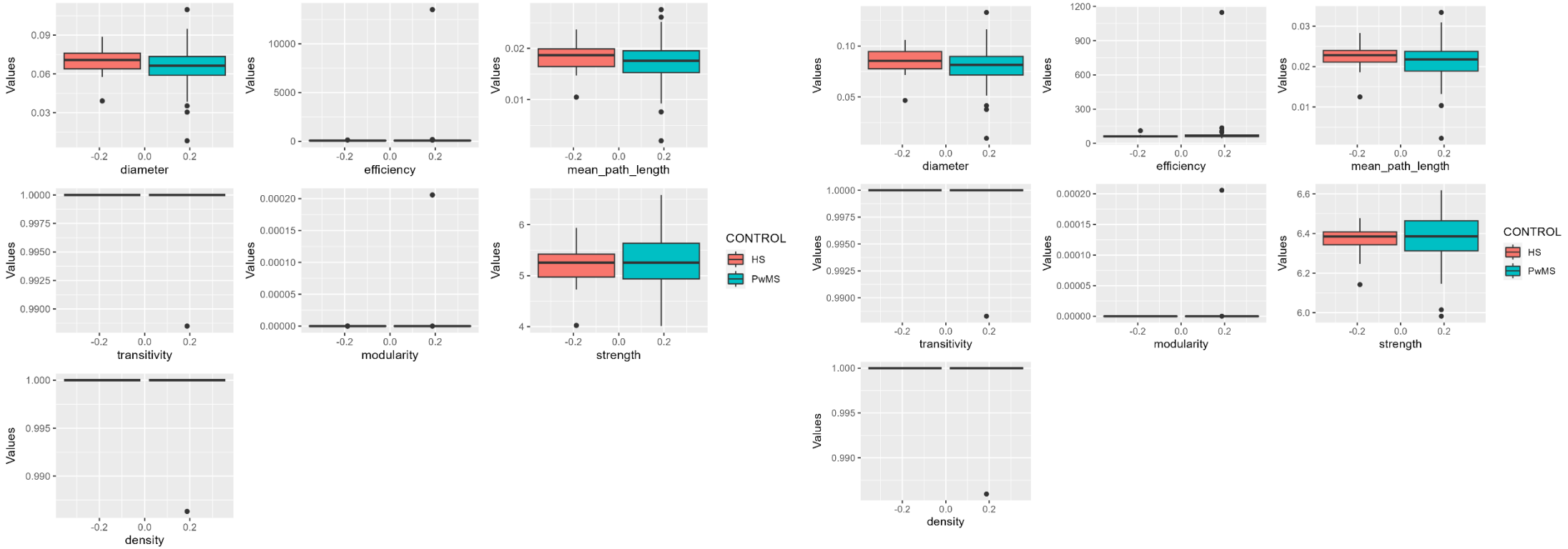

Fig 17. Boxplot graph global measures. GM Layer. With and without SVD.

Both figures, without SVD (a) and with SVD applied (b) show no real differences between HS (in red) and PwMS (in green). In some cases values are so concentrated that the boxplot is not different from a line.

fMRI LAYER

a) WITHOUT SVD NORMALIZATION

b) SVD NORMALIZATION

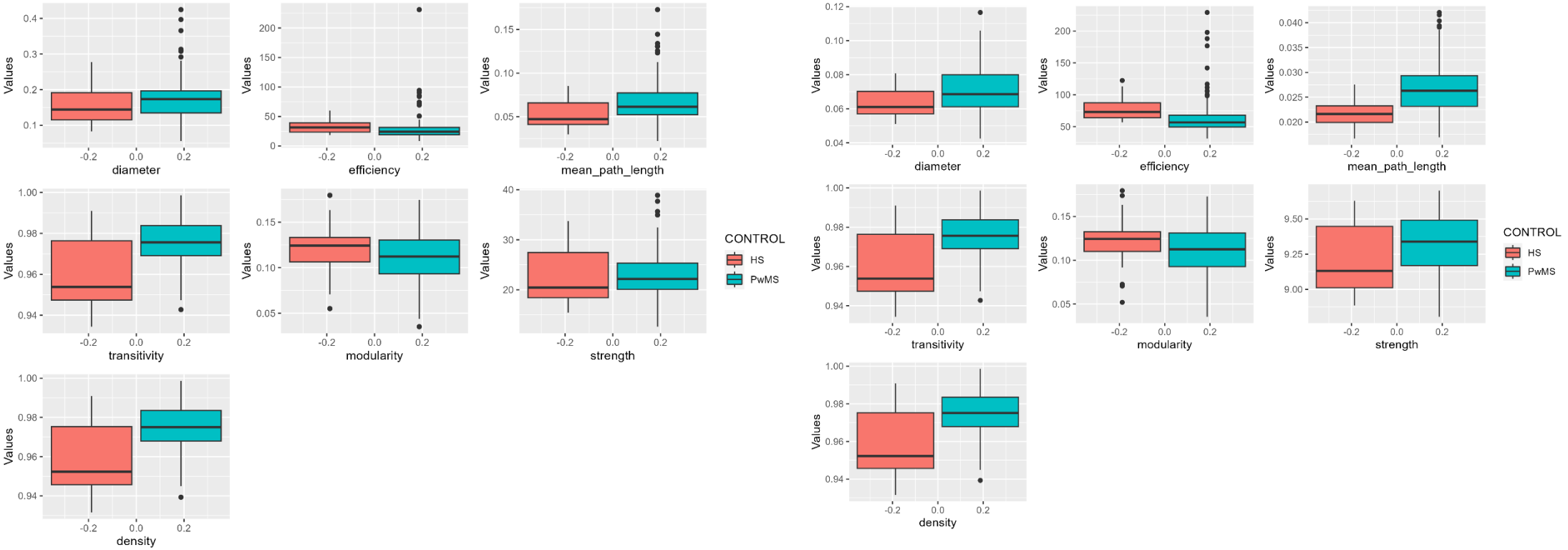

Fig 18. Boxplot graph global measures. fMRI Layer. With and without SVD.

Figure illustrates differences between groups in fMRI layer, HS in red and PwMS in green, for different measures in case without SVD normalization a) and with SVD b). In this case, contrary to what happens in the FA layer (Fig 16). the differences between groups are greater in b), especially in the case of efficiency, mean path length and strength.

Statistical differences between global measures

Results obtained after conducting statistical test 1 are shown in Table 4 and Table 5. For the sake of brevity, only measures with a p-value less than 0.05 are displayed.

WITHOUT SVD NOMALIZATION

LAYERS	MEASURE	HS	PwMS	p value
FA	diameter	0.82 ± 0.09	0.72 ± 0.1	0.000
	efficiency	2.56 ± 0.46	2.81 ± 0.46	0.009
	mean path length	0.44 ± 0.07	0.4 ± 0.06	0.007
	strength	23.06 ± 1.13	21.42 ± 1.9	0.000
fMRI	efficiency	34.12 ± 13.85	28.64 ± 22.36	0.018
	mean path length	0.05 ± 0.02	0.07 ± 0.02	0.011
	transitivity	0.96 ± 0.02	0.98 ± 0.01	0.001
	density	0.96 ± 0.02	0.98 ± 0.01	0.001

Table 4. Single layers results of statistical test. No SVD normalization

Continuous variables are given as the mean ± standard deviation. Color highlights measures that are present in both cases, Table 4 and 5, with and without SVD.

SVD NORMALIZATION

LAYERS	MEASURE	HS	PwMS	p value
FA	diameter	0.33 ± 0.03	0.31 ± 0.04	0.041
fMRI	diameter	0.06 ± 0.01	0.07 ± 0.01	0.015
	efficiency	77.67 ± 19.56	63.61 ± 28.36	0.000
	mean path length	0.02 ± 0.00	0.03 ± 0.00	0.000
	transitivity	0.96 ± 0.02	0.98 ± 0.01	0.001
	density	0.96 ± 0.02	0.98 ± 0.01	0.001

Table 5 Single layers results of statistical test. SVD normalization

Continuous variables are given as the mean ± standard deviation. Color highlights measures that are present in both cases, Table 4 and 5, with and without SVD.

It is noteworthy that, in the case without normalization, there are four statistically significant measures half from the white matter layer (FA) and half from the functional layer (fMRI).

However, upon applying normalization, while the functional layer (fMRI) yields one additional significant measure, the FA layer is left with just one. Interestingly, in both scenarios, no significant measures from the grey matter (GM) layer is observed.

It is also important to point out that density yield distinctive results in the case of the functional layer.

It can also be interesting to examine the correlations between graph measures and compare the difference between both cases. In the case involving SVD, as depicted in Fig 19, the diameter, efficiency, and mean path length in the WM layer display a strong correlation, either positive or negative, as predicted. In contrast, efficiency and mean path length in the functional layer do not exhibit significant correlation. Simultaneously, for both scenarios—SVD and non-SVD—the functional layer's transitivity and density display an extremely high correlation, so much so that they appear as 1.00 in the figure

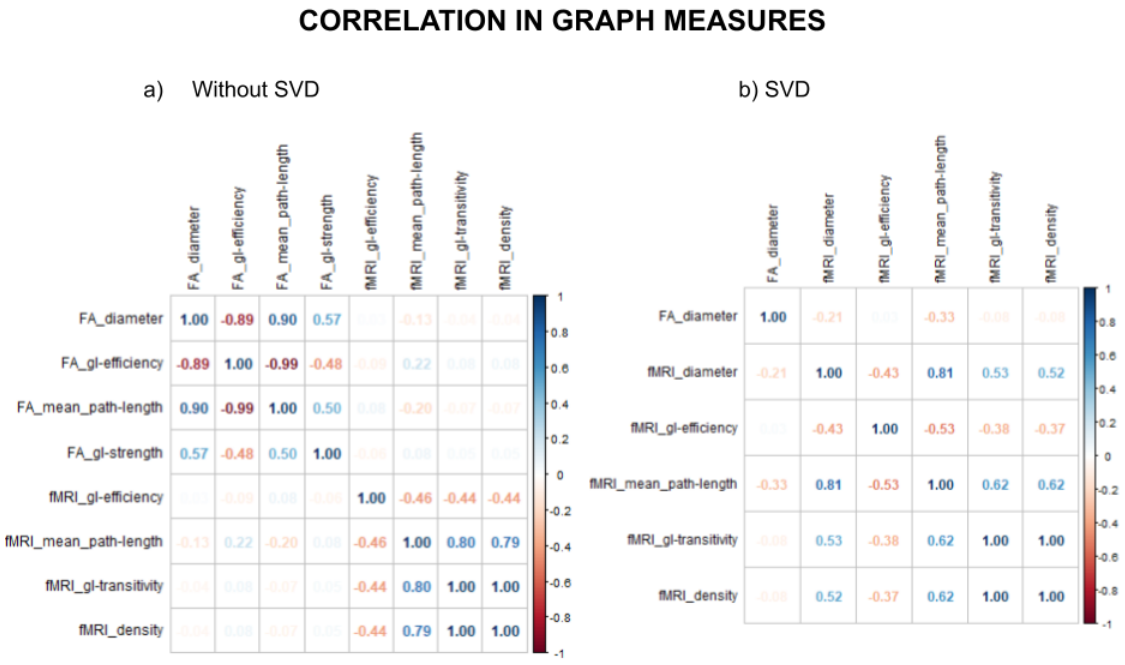

Fig 19.Correllations between global measures. Separate Layers. With and without SVD correction.

Labels represent a specific layer (FA or fMRI) and measure separated by an underscore. 'gl' if present denotes global. It is clear from both cases (with SVD b) and without SVD a)) that only measures from the same layer show some correlation. More measures show a significant correlation in a) than in b), as can be observed in the figure.

After conducting statistical tests with Bonferroni correction, the number of measures that pass the test shrinks to just 2 and 1, for cases with and without SVD, respectively. The limited number of available PPMS and SPMS participants must be taken into consideration and could explain why there are so few measures.

LAYERS	Without SVD	With SVD
FA	Diameter	Diameter
	Efficiency	-----

Table 6. Global single layer measures that pass test 2. Color highlight measures present in both cases.

For the machine learning models in this study, the decision was made to incorporate all measures that successfully passed the first statistical test.

5.2. Local single layer graph measures, with and without SVD correction

As it is explained in section 4.5.2, 3 local measures will be made. This implies a total of 684 measures (calculates as 3 measures x 76 nodes x 3 layers). It complicates the analysis and summarization of results compared to the previous case, making it impossible to present the results in the same way.

Statistical differences between local measures

The following results are obtained for local measures

684 total measures	**Without SVD**	**With SVD**
	# measures	# measures
HS vs PwMS	326	191
Between MS types (Bonferroni correction)	89	9

Table 7. Local single layer measures that pass statistical tests.

The most notable aspect of these results lies in the considerable disparity between the outcomes with and without the application of SVD. As discussed in Section 4.5.1, applying this correction tends to even out differences in weight distribution. Therefore, it seems a logical consequence that fewer measures will be statistically different as range of weights is reduced.

Results for each case after Bonferroni are examined more closely in the following tables. Main measures and layer for the case without SVD are shown in Table 8.

WITHOUT SVD		
LAYERS	MEASURE	TOTAL NODES
FA	Closeness centrality	28
	strength	56
GM	Strength	1
fMRI	Closeness centrality	4

Table 8. Local single layer measures and number of nodes that pass statistical test without SVD.

The subsequent table shows the nodes that have statistically significant measures in more than one layer. Only the right hemisphere's paracentral is statistically significant in all layers.

ROI	# of layers
ctx-rh-paracentral	3
ctx-lh-caudalmiddlefrontal	2
ctx-lh-entorhinal	2
ctx-lh-fusiform	2
ctx-lh-inferiorparietal	2
ctx-lh-isthmuscingulate	2
ctx-lh-lingual	2
ctx-lh-middletemporal	2
ctx-lh-paracentral	2
ctx-lh-postcentral	2
ctx-lh-precentral	2
ctx-lh-rostralmiddlefrontal	2
ctx-rh-caudalanteriorcingulate	2
ctx-rh-cuneus	2
ctx-rh-entorhinal	2
ctx-rh-fusiform	2
ctx-rh-lateraloccipital	2
ctx-rh-parstriangularis	2
ctx-rh-precentral	2
ctx-rh-superiorfrontal	2
ctx-rh-supramarginal	2
Left-Pallidum	2
Right-Amygdala	2

Table 9. ROIs that pass both statistical tests in more than one layer.

Next table shows results for the case where SVD is applied, as there are only 9 measures it is easier to summarize and review the results.

WITH SVD

LAYERS	MEASURE	NODE
FA	Closeness centrality	lh-isthmuscingulate
		Left-Hippocampus
		ctx-rh-caudalanteriorcingulate
		ctx-rh-lateralorbitofrontal
		ctx-rh-precentral
	strength	Right-Caudate
fMRI	Closeness centrality	ctx-rh-entorhinal
		ctx-rh-parahippocampal
		ctx-rh-paracentral

Table 10. Local single layer measures and number of nodes that pass statistical test witht SVD

Again it should be noted that only one measurement from GM layer shows a statistically significant differences.

5.3. Results for local multilayer measurements

As it has been already commented, in light of the aforementioned results, where the GM layer does not yield significant results, two cases were considered: One multilayer network with all three layers, another multilayer network with only two layers (FA and fMRI)

The first decision to be made concerns the weight to be assigned to inter-layer links. As previously mentioned (Mandke) uses the average of all weights. A close examination of these averages show that they are quite similar, moving within a tight range, as can be seen in Fig 20. The range is approximately 0.0125 and 0.025 for 3 and 2 layers respectively, moreover range does not seem different for HS and PwMS. Although the range is larger for the latter group as it includes more subjects

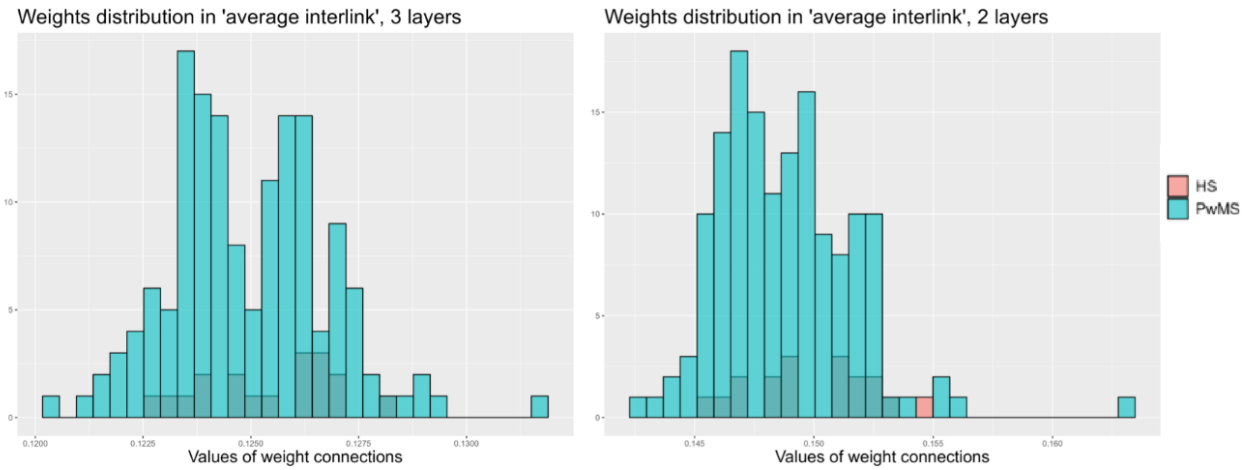

Fig 20. Distribution of interlayer weights in case of averaging.

Left plot illustrates the average inter-layer link weight histogram in the case of three layers, with groups differentiated by color (pink for HS and blue for PwMS). While right plot depicts the case of 2 layers. It is clear from both images that values are concentrated in a confined range of values and that no clear difference between groups is distinguishable. Range is approximately 0.0125 for the case of three-layers and 0.025 for 2 layers.

Following these findings, another option was considered, selecting the interlayer weight that maximizes the number of measures that passes the statistical tests. A range of weights from 0.1 to 1 was evaluated, as presented in the next table.

	3 LAYERS		2 LAYERS	
interlayer weights	Test 1	Bonferroni	Test 1	Bonferroni
0.1	28	3	52	2
0.2	29	4	52	2
0.3	29	4	52	2
0.4	29	4	52	2
0.5	29	4	52	2
0.6	31	4	52	2
0.7	31	4	52	2
0.8	31	4	52	2
0.9	31	4	52	2
1.0	31	4	52	2

Table 11. # of measures that pass statistical test when varying interlayer weights

In the scenario with two layers, the variation in interlayer weights makes no difference, and somewhat surprisingly, although more measures pass Test 1 compared to the three-layer case, fewer measures pass Test 2. When considering three layers, minor differences emerge between weights, indicating that an increase in weight results in an slightly increase in the number of measures that pass Test 1. Ultimately, an interlayer weight of 1 was selected.

The following table provides brief summary of those measures that passed the first statistical test in either case:

	2 layers	**3 layers**
	# measures	# measures
MultiPageRank	7	15
Sum Multi-Strength	8	9
Closeness centrality	37	7

Table 12. Summary of multi-layer measures that passed statistical test 1

There is an important difference in closeness centrality between both cases.

After test with Bonferroni correction the following results are obtained:

	2 layers	3 layers
	# Nodes	# Nodes
MultiPageRank		ctx-lh-precentral Left Accumbens area Ctx-rh-lateraloccipital
Sum Multi-Strength		Ctx-rh-superiorfrontal
Closeness centrality	Ctx-lh-medialorbitofrontal Ctx-lh-parsorbitalis	

Table 13. Summary of multi-layer measures that passed statistical test 2

Following the second statistical test, it is worth mentioning that statistical significant measures and nodes differ between the two scenarios.

Upon comparing all measures from single layer (with and without SVD) and multilayer (with 2 and 3 layers), the node 'ctx-lh-precentral' stands out as the most frequently appearing.

While the GM layer might not seem to contribute significantly, its presence still influences the results of multilayer measurements as results in both cases are clearly different.

5.4. Summary of data for machine learning models

To recap, three cases were considered for the machine learning models: separate layers, multilayer with three layers (3-multilayer), and multilayer with two layers (2-multilayer). Section 4.7 outlines the data chosen for each model, indicating that the final selection would depend on the results of the statistical tests that were just presented in the preceding sections. Based on these results, the final data was selected in the following manner:

- Separate layers: For *local measures* only those that passed the 2nd test were considered, while *global* ones were selected from those that passed the first test.
- Multilayer: From the single layer measurements, only data with SVD correction was considered as explained in section 4.7. Measures are added following the same criterion as for Separate Layers, global measures that pass test 1, but only local that passes both tests. All multilayer measures that passed the first statistical test were selected, due to the limited number of measures that passed the second test.

This leaves us with the following number of features

1. Separate Layers: 97 features
2. 3-multilayer: 53 features
3. 2-multilayer: 74 features

5.4.1. Removal of correlated features

Machine learning models perform better when features are not correlated. Given the substantial number of features in each case, there is a strong likelihood that some may indeed be correlated.

A threshold of ±0.8 is established and measures correlating above (or below) this threshold are removed. The number of measures taken out are shown in the table below

CASE	Features removed
Separate layers	86
3-multilayer	9
2-multilayer	10

Table 14. Features removed due to high correlation

The influence of the SVD correction on the results is evident in the count of correlated features that are eliminated. Barely any measure is removed in scenarios where the correction has been applied compared to scenario where it has not been applied.

Though this is a necessary step for model construction, it is important to note that correlations may well vary with another dataset and potentially lead to alternate features. Moreover, this approach can complicate model interpretation. When a feature is identified as important in the final model, it raises a question: Is it always this feature that holds importance, or it could be one of its highly correlated counterparts?

5.5. Machine learning model results HS vs PwMS

Subsequent table shows results for separate layers, 3-multilayer and 2-multilayer, green background color highlights best model for that metric and scenario.

Separate layers

					# Errors out of 33	
	accuracy	precision	recall	F1 score	FP	FN
RF	0.901	0.943	0.949	0.945	0	3
SVC	0.871	0.878	0.991	0.931	2	0
KNN	0.818	0.953	0.836	0.888	0	4
XGBoost	0.894	0.943	0.940	0.940	0	2

Table 15. Results from machine learning models HS vs PwMS. Separate layers.

Color in this table and in Table 16 Table 17 highlight model(s) that perform better for the corresponding metric.

3- multilayer

	accuracy	precision	recall	F1 score	# Errors out of 33	
					FP	FN
RF	0.894	0.926	0.957	0.940	1	1
SVC	0.872	0.878	0.992	0.931	2	0
KNN	0.749	0.953	0.749	0.839	1	8
XGBoost	0.901	0.949	0.939	0.944	0	2

Table 16. Results from machine learning models HS vs PwMS. 3-multilayer.

2-multilayer

	accuracy	precision	recall	F1 score	# Errors out of 33	
					FP	FN
RF	0.856	0.894	0.949	0.920	2	3
SVC	0.879	0.879	1.000	0.936	2	0
KNN	0.787	0.924	0.827	0.871	2	8
XGBoost	0.841	0.907	0.914	0.910	1	2

Table 17. Results from machine learning models HS vs PwMS. 2-multilayer.

With some exceptions, results exhibit a pattern that is, a model that performs better in one metric, tends to perform better in that metric in all scenarios, this could mean that one must choose one model over another depending on the target. For instance, if goal is to minimize false negatives, clearly SVC is the best choice and KNN is the worst

BEST METRIC RESULT FOR EACH MODEL

	Accuracy	Precision	Recall	F1 score	FP	FN
RF	2-multi	Sep-layers	3-multi	Sep-layers	Sep-layers	3-multi
SVC	2-multi	2-multi	2-multi	2-multi	ALL	ALL
KNN	Sep-layers	Sep/3-multi	3-multi	3-multi	Sep-layers	Sep-layers
XGBoost	3-multi	3-multi	Sep-layers	3-multi	Sep/3-multi	ALL

Table 18. Summary of best network for each metric. HS vs PwMS

On each cell is displayed the network that performs better on that metric; 2-multi for 2-multilayer, 3-multi for 3-multilayer and Sep-layers for separate layers. Sep/3-multilayers indicate that both, separate-layer and 3-multilayer performance is equally good. ALL indicates that 3 layers performs equally well.

Previous table shows which scenario performs better on each model. Differences between metrics is tiny, 0.001 in some case. It seems no scenario performs better than any other across all models. However it must be noted that SVC has all its better metrics in 2-multilayer (with the caveat that differences are minimal)

5.6. Machine learning model results for MS types classification

For classification between MS types (RRMS, SPMS, PPMS), only PwMS participants were selected, as in most diagnostic situations, general diagnosis (MS) comes first, and then diving into details is possible.

As has already been discussed, limitation in the number of patients with PPMS (6), potentially limits results of this work and generalization.

Separate layers

	accuracy	precision	recall	F1 score	# Errors out of 30
RF	**0.855**	**0.804**	**0.855**	**0.823**	**6**
SVC	**0.855**	**0.744**	**0.855**	**0.795**	**6**
KNN	**0.727**	**0.818**	**0.727**	**0.758**	**9**
XGBoost	**0.846**	**0.803**	**0.846**	**0.819**	**5**

Table 19. Results from machine learning model. MS types. Separate layers

Color in this table and in Table 20Table 21 highlight model(s) that perform better for the corresponding metric.

3- multilayer

	accuracy	precision	recall	F1 score	# Errors out of 30
RF	**0.863**	**0.824**	**0.863**	**0.838**	**5**
SVC	**0.863**	**0.745**	**0.863**	**0.800**	**6**
KNN	**0.693**	**0.829**	**0.693**	**0.739**	**8**
XGBoost	**0.803**	**0.764**	**0.803**	**0.783**	**3**

Table 20. Results from machine learning model. MS types. 3-multilayer

2-multilayer

	accuracy	precision	recall	F1 score	# Errors out of 30
RF	**0.855**	**0.804**	**0.855**	**0.823**	**6**
SVC	**0.854**	**0.744**	**0.855**	**0.796**	**6**
KNN	**0.727**	**0.818**	**0.727**	**0.759**	**9**
XGBoost	**0.846**	**0.803**	**0.846**	**0.819**	**5**

Table 21. Results from machine learning model. MS types. 2-multilayer

A surprising finding is that XGBoost does not perform better in any metric, but is the model that commit less mistakes in all scenarios. As in the previous case of binary classification, models that perform better in one metric in one case, does so in all scenarios.

BEST METRIC RESULT FOR EACH MODEL

	Accuracy	Precision	Recall	F1 score	FP
RF	**3-multi**	**3-multi**	**3-multi**	**3-multi**	**3-multi**
SVC	**3-multi**	**3-multi**	**3-multi**	**3-multi**	**ALL**
KNN	**Sep/2-multi**	**3-multi**	**Sep/2-multi**	**2-multi**	**3-multi**
XGBoost	**Sep/2-multi**	**Sep/2-multi**	**Sep/2-multi**	**Sep/2-multi**	**3-multi**

Table 22. Summary of best network for each metric. MS types

On each cell is displayed the network that performs better on that metric; 2-multi for 2-multilayer, 3-multi for 3-multilayer and Sep-layers for separate layers. Sep/2-multil indicate that both, separate-layer and 2-multilayer performance is equally good. ALL indicates that 3 layers performs equally well.

In multiclass classification, though differences are still minimal, 3-multilayer performs better in the metrics of 2 out of 4 models, and makes fewer errors in all models. This could suggest that although GM layer does not yield to relevant measures in single layer, it plays a role in the overall multilayer network.

5.7. Feature importance in HS vs PwMS

Following figures show feature importance for Random Forest and XGBoost for separate layers and 2-multilayer and 3-multilayer in the binary classification (HS vs PwMS)

Keys to read the figures.

To interpret the figures, the following conventions should be understood as they are consistent across all figures:

Features that show an average importance of zero have been removed

Global measures of single layers, follow this convention “Layer_measure”. For instance, “FA_diameter” refers to the diameter of FA layer or “fMRI_gl_efficiency” for global efficiency of fMRI layer

Local measures follow this convention: “Measure_ROI”, for instance “MultiPRCent_ctx-lh_inferiorparietal”, where abbreviations indicate the following:

Abbreviation	Measure
MultiPRCent	MultiPageRank Centrality
MultiPath	Closeness centrality for multilayers
MultiSTSum	Strength sum in multilayers
loc-clos-cent	Local closeness centrality (single layers)
loc-strength	Strength

Abbreviation in ROI	meaning
lh	Left hemisphere
rh	Right hemisphere
ctx	cortex

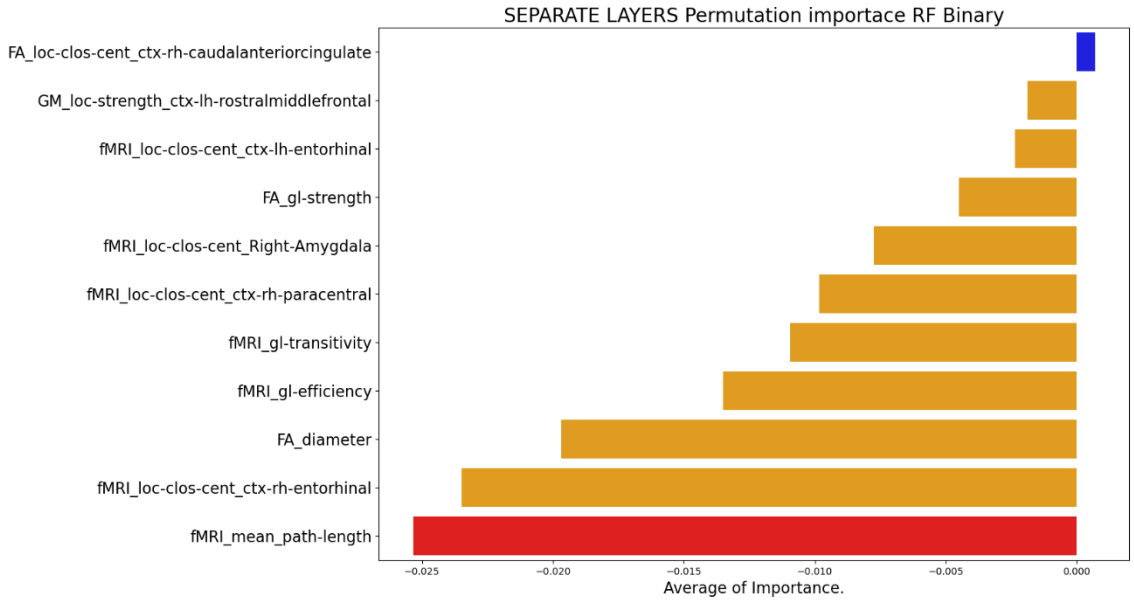

Fig 21. Permutation Importance for Random Forest. Separate Layers

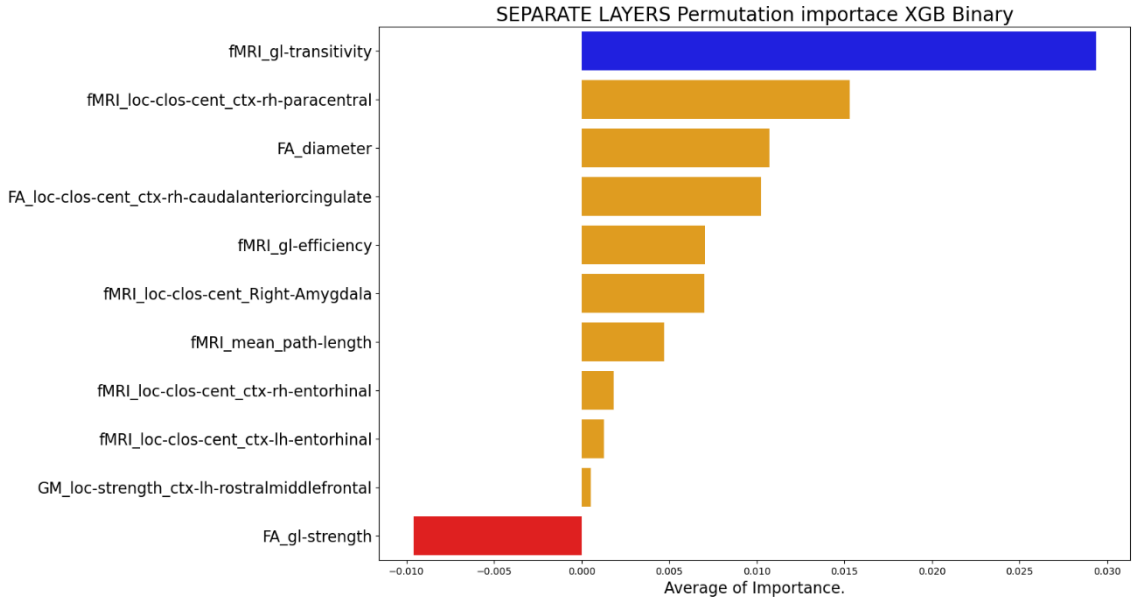

Fig 22. Permutation Importance for XGBoost. Separate Layers

Figures display a curious pattern. In RF all features except one have a negative impact on the model (upper image), whereas all features except one have a positive impact on XGBoost model (bottom image). Positive feature in RF is closeness centrality for ctx-rh-cuadalanteriorcingulate which is also positive for XGBoost, and most negative feature in XGBoost is global strength for FA layer and is also negative in RF. Blue color indicates most positive impact and red color the most negative.

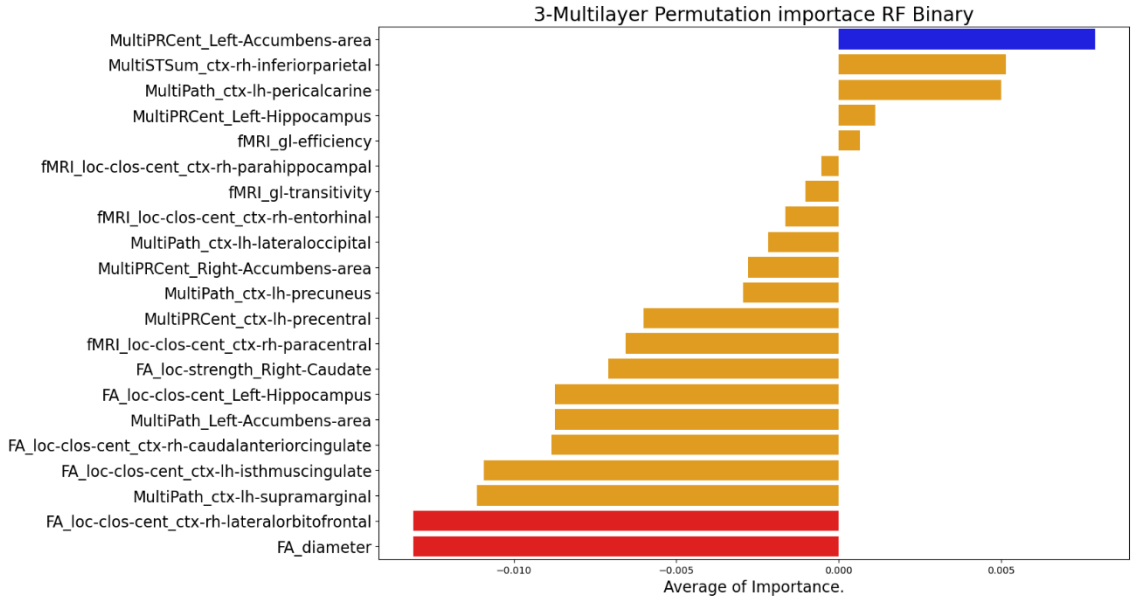

Fig 23. Permutation Importance for Random Forest. 3-multilayer

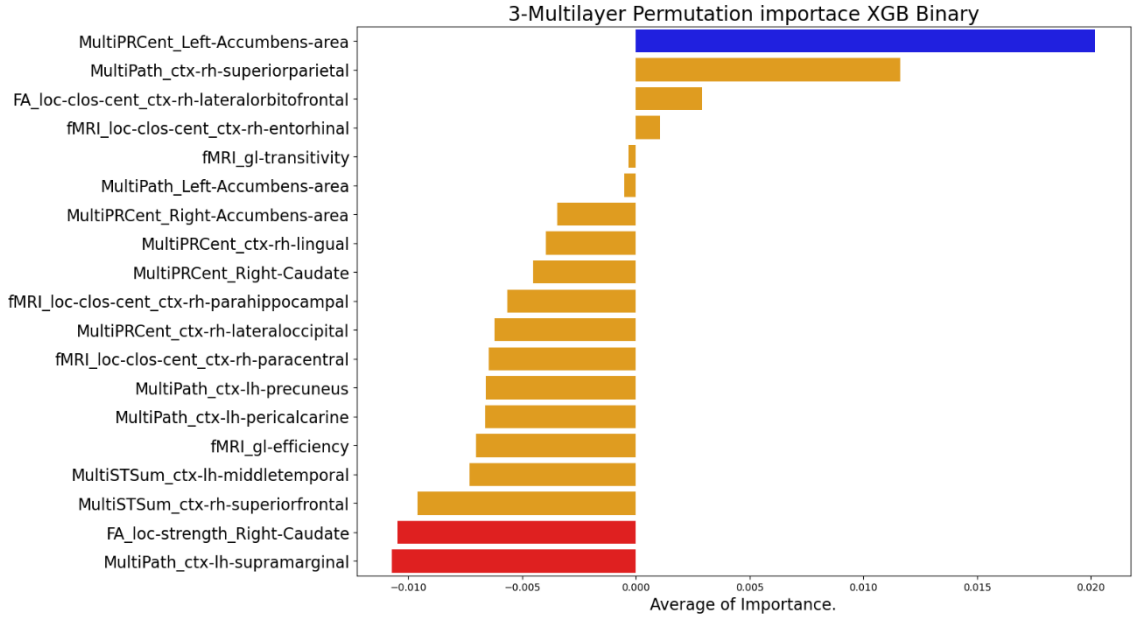

Fig 24. Permutation Importance for XGBoost. 3-multilayer

In this case, the figures show how RF (upper image) and XGBoost (bottom image) display more or less the same number of features over and below zero. MultiPageRank closeness centrality for Left Accumbens is the feature with higher impact in both models, whereas among the most negative both have multilayer closeness centrality for lh-supramarginal ROI. Blue color in plot indicate most positive feature and red one, the two most negative ones.

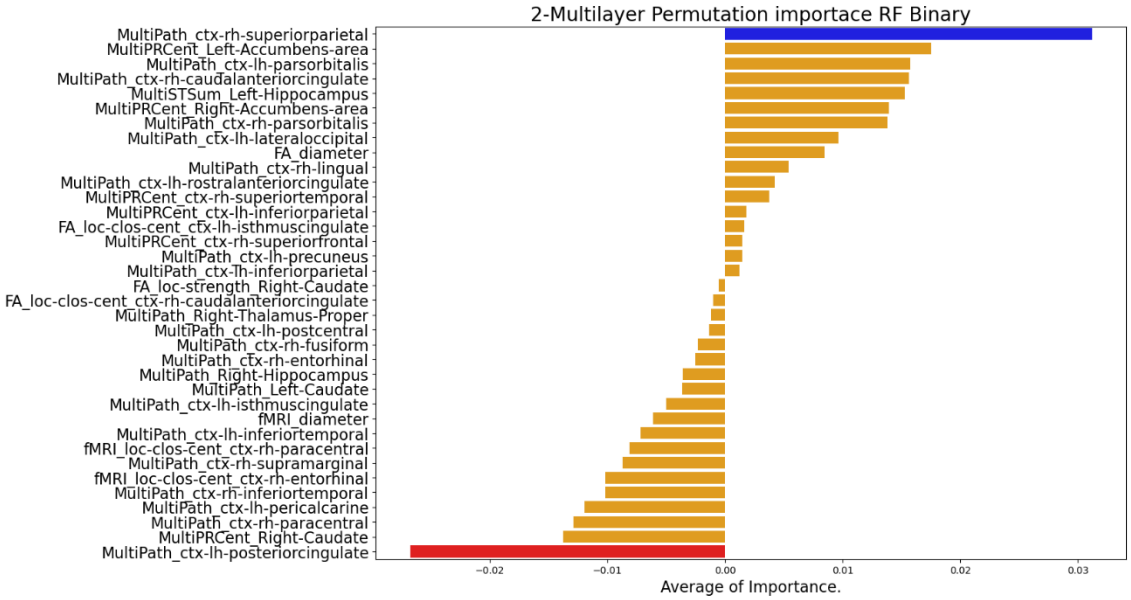

Fig 25. Permutation Importance for Random Forest. 2-multilayer.

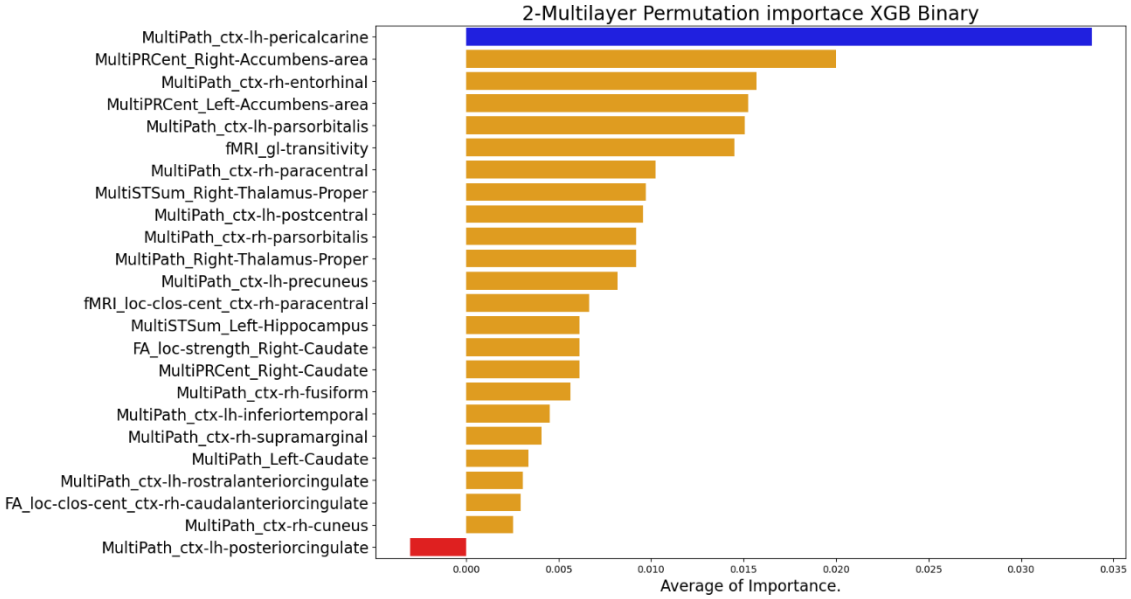

Fig 26. Permutation Importance for XGBoost. 2-multilayer

For RF (upper image) we have a lot of features that either show positive or negative impact, while XGBoost, (bottom image) shows only one feature below zero (closeness centrality for multilayers in lh-posteriorcingulate) which is also the most negative in RF. On the other hand 3 most positive ones in one model are not among the top three in the other model. Although if more features (5 or 6) are considered, Left Accumbens MultiPageRank centrality and lh-parsobitalis for multilayer closeness centrality are found in both. Plots show in blue color most positive feature and in red most negative.

While some common features above or below zero with similar importance do appear between models within the same scenario, no consistent pattern emerges across different scenarios or even within the same case. Especially because some models show only one positive or negative feature reducing and, in fact, eliminating the possibility of identifying a common node and measure. Nonetheless, a summary across all scenarios and models reveals that certain ROIs and measures do appear more frequently than others. The subsequent tables provide a summary of these frequently occurring ROIs and measures in both cases - values above zero and below zero:

POSITIVE IMPORTANCE (ABOVE ZERO)

ROI	# of times present	SCENARIOs	MODELS	MEASURES
ctx-rh-caudalanteriorcingulate	4	Separate layers	RF, XGB	Closeness centrality FA layer (3)
		2-multilayer	RF, XGB	Multilayer closeness centrality (1)
Left Accumbens Area	4	2-multilayer	RF, XGB	MultiPageRank centrality (4)
		3-multilayer	RF, XGB	
ctx-rh-entorhinal	4	Separate layers	XGB	Closeness centrality fMRI layer (3)
		2-multilayer	XGB	Multilayer closeness centrality(1)
		3-multilayer	RF, XGB	
ctx-rh-paracentral	3	Separate layers	XGB	Closeness centrality fMRI layer (2)
		2-multilayer	XGB	Multilayer closeness centrality (1)
ctx-rh-superiorparietal	3	2-multilayer	RF	Multilayer closeness centrality(3)
		3.multilayer	RF, XGB	

Table 23. More frequent ROIs with positive importance.

NEGATIVE IMPORTANCE (BELOW ZERO)

ROI	# of times present	SCENARIOs	MODELS	MEASURES
Right Caudate	6	2-multilayer	RF	Local strength Fa layer (3)
		3-multilayer	RF, XGB	MultiPageRank centrality (3)
ctx-rh-paracentral	5	Separate layers	RF	Closeness centrality fMRI layer (4)
		2-multilayer	RF	Multilayer closeness centrality (1)
		3-multilayer	RF, XGB	
ctx-lh-pericalcarine	3	2-multilayer	RF	Multilayer closeness centrality(3)
		3-multilayer	RF, XGB	
ctx-rh-entorhinal	3	Separate layers	RF	Closeness centrality fMRI layer (2)
		2-multilayer	RF	Multilayer closeness centrality (1)

Table 24. More frequent ROIs with negative importance

It is of no surprise that 2-multilayer appears more frequently than the rest, given that it shows more measures above and below zero, see Fig 25 and Fig 26 (it also has more available features Table 12). Although, as it has been discussed, there are no nodes that consistently appear in all models, scenarios and measures, it is indeed the case that some can be regarded as more important to differentiate between HS and PwMS. Table 23 show how cortex, right hemisphere caudalanteriorcingulate, left Accumbens area show importance above zero, while right caudate and cortex, left hemisphere paracentral zone show importance below zero. It is noteworthy how cortex right hemisphere entorhinal zone seems to play opposite roles depending on the case (see both Table 23 and Table 24)

6. Conclusions and future work

6.1. Conclusions related to graph construction

SVD correction and single-layer results

This study presented the opportunity to compare two approaches to single layer measures, one applying an SVD correction and another without. The SVD correction reduces the range of values of the graph weights, which appears to dilute differences among graphs (from the perspective of the measures applied). This is evidenced by the fact fewer *local* measures passes both statistical tests, 9 with SVD vs 89 without. Conversely, the measures with SVD normalization that pass the tests show no significant correlation, while 86 out of 97 single layer measures, both global and local, show high correlation in the case without SVD.

This observation that SVD reduces the range of weights and statistically significant measures could potentially be interpreted as a reduction of noise or bias as suggested by Mandke et al. 2018. Whether this approach should be used when considering three separate layers, or even just a single layer, remains a question for future research.

Inter-layer weights

Another important conclusion is the lack of clear criteria for selecting interlayer weights. Indeed, setting these values presents a challenge due to the distinct biological attributes each layer holds, making it difficult to propose a theoretical approach.

Considering the primary objective of this project, which is to differentiate between participant classes, it make sense to select values that maximize features passing statistical tests, thereby providing more features available for machine learning models. However, it has been observed that changes in weights (ranging from 0.1 to 1) barely affect the number of features passing one or both statistical tests. It may be beneficial to extend the range of values beyond 1, or developing a theoretical framework to establish this value.

GM layer

The finding that only one measurement out of 235 from the GM layer passed the statistical tests raises the question of whether this layer provides information. This realization emerged during the project's development, led to the consideration of a multilayer with only FA and fMRI layers (2-multilayer).

A retrospective analysis reveals 2 conflicting tendencies. On one hand 3-multilayer models seem to outperform 2-multilayer models in distinguishing between MS types suggesting that this layer indeed plays a role in the global multiplex network. On the other hand, in the analysis of feature importance, 2-multilayer show more features with average importance above and below zero, possibly as a consequence of having more available features after the statistical tests. The former is more important for model construction, while the latter helps in identifying potentially clinical important ROIs.

Multilayer results

The difference between the 2-multilayer and 3-multilayer scenarios is not as significant as it is SVD in the single-layer measurements. In the 2-multilayer case, 52 measures pass the first statistical test and 2 measures pass the second, while in the 3-multilayer scenario, 31 and 4 measures pass the first and second statistical tests respectively (see Table 12 and Table 13). When examining correlations, a similar number of features (10 and 9) are removed due to high correlation in both scenarios.

From the point of view of the graph construction, it does not appear to be an objective reason to prefer one network over the other.

6.2. Conclusions from machine learning models

Shortage of data

One of the major challenges encountered during this project is the limited size of the dataset, with only 165 participants and particularly low representation of certain classes, such as PPMS (6). Although data augmentation technique like SMOTE have been used, it cannot really solve the problem only mitigates it.

Data augmentation techniques "artificially" inflates minority classes by creating new samples from the existing ones. This could potentially introduce inaccurate new instances. Furthermore, the division of the dataset in into train and test sets reduces the available samples for SMOTE to operate. It also downsize to the minimum the number of samples from the minority class in the test dataset, thus making it difficult to evaluate the model and reducing its validity.

The scarcity of data presents a challenge not only in the construction of machine learning models but also during feature selection, as statistical tests would undoubtedly benefit from a larger dataset.

HS vs PwMS

The application of machine learning models to distinguish between HS and PwMS yield mixed results across algorithms and network types. SVM shows a slight advantage as

it shows the highest recall and fewest number of false negatives across all networks. SVM also the highest accuracy and F1 score in the 2-multilayer case. However differences in scores are minimal as well as differences in the number of errors between SVM, RF and XGBoost (see Table 15, Table 16 and Table 17)

RRMS vs SPMS vs PPMS

When it comes to differentiating between types of MS, SVM and Random Forest models generally outperform other models in most cases. However, somewhat surprisingly, XGBoost, despite having slightly lower scores, makes the fewest errors. This counterintuitive result could be attributed to the minimal differences in the scores between these models (Table 19, Table 20 and Table 21)

Upon analyzing the performance of the network across models, 3-multilayer outperforms the other two networks in two models (RF and SVM) while separate layers and 2-multilayer are better in XGBoost. KNN model presents mixed results, with no clear superior network (Table 22)

Feature importance between HS and PwMS

Probably as a consequence of all mixed results there is no single ROI or measure that completely stands out when comparing their importance in the different models. However, upon closer examination, certain ROIs like the right hemisphere's caudalanteriorcingulate and entorhinal and left Accumbens area seem to slightly exceed others in positive importance. Conversely right caudate and paracentral of the right hemisphere surface more often as negatively important.

Regarding to the measures, the situation is even more mixed, and no single clearly distinguishing itself from the rest.

It is important to recall that a significant number of measures were eliminated from the separate layers network due to high correlation. This makes difficult to infer which nodes could be important as their measures are all related. This may be a clear sign that multilayer networks should be preferred.

6.3. Future work

To enhance the outcomes of the current project it is of paramount importance to increase the number of subjects. Some efforts has been done in this direction with the creation of SUMMIT (Bove et al. 2018) or the use of Virtual Brain for modeling (Martí-Juan et al. 2023).

In relation to the creation and analysis of graphs, additional measures that have not been considered in this study such as k-coreness (Pontillo et al. 2022) could be incorporated. Additionally, to further optimize the creation of machine learning models, the exploration of different thresholds (Zhao et al. 2020), could prove useful.

However one important path not explored in this work involves assigning different weights to different layers, thereby altering the contribution of each individual layer. Considering the obtained results, where GM does not contribute to statistically relevant individual layer features, but 2-multilayer and 3-multilayer show different and mixed

results, it suggests that neither completely removing the layer nor fully including it is the optimal solution. In this sense it is worth remembering that Casas-Roma et al. 2022 used GM layer to model interlayer connections between the other two layers, which is a potential use of this layer which is also not explored in this work.

7. List of Abbreviations

Terms in alphabetical order

TERM	ABBREVIATION
Clinically Isolated Syndrome	CIS
Expanded Disability Status Scale	EDSS
Extreme Gradient Boosting	XGBoost, XGB
False Negative (Value)	FN
False Positive (Value)	FP
Fractional Anisotropy	FA
Functional MRI	fMRI
Gray Matter	GM
Healthy Subject	HS
K-Nearest Neighbor	KNN
Left hemisphere	lh
Multiple Sclerosis	MS
Magnetic Resonance Imaging	MRI
Patient with MS	PwMS
Primary progressive course	PPMS
Principal Component Analysis	PCA
Region(s) of interest	ROI(s)
Relapsing-Remitting Course	RRMS
Right Hemisphere	rh
Secondary progressive course	SPMS
Singular vector decomposition	SVD
Support vector classification	SVC
Support Vector Machine	SVM
True Negative (Value)	TN
True Positive (Value)	TP

8. Bibliography

Barabási, Albert-László. 2016. *Network Science*. Cambridge University Press. http://networksciencebook.com/.

Barkhof, Frederik. 2002. "The Clinico-Radiological Paradox in Multiple Sclerosis Revisited." *Current Opinion in Neurology* 15 (3): 239–45. https://doi.org/10.1097/00019052-200206000-00003.

Bassett, Danielle S., and Olaf Sporns. 2017. "Network Neuroscience." *Nature Neuroscience* 20 (3): 353–64. https://doi.org/10.1038/nn.4502.

Battiston, Federico, Jeremy Guillon, Mario Chavez, Vito Latora, and Fabrizio De Vico Fallani. 2018. "Multiplex Core–Periphery Organization of the Human Connectome." *Journal of The Royal Society Interface* 15 (146): 20180514. https://doi.org/10.1098/rsif.2018.0514.

Battiston, Federico, Vincenzo Nicosia, Mario Chavez, and Vito Latora. 2017. "Multilayer Motif Analysis of Brain Networks." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27 (4): 047404. https://doi.org/10.1063/1.4979282.

Behdenna, Abdelkader, Julien Haziza, Chloé-Agathe Azencott, and Akpéli Nordor. 2021. "PyComBat, a Python Tool for Batch Effects Correction in High-Throughput Molecular Data Using Empirical Bayes Methods." bioRxiv. https://doi.org/10.1101/2020.03.17.995431.

Bianconi, Ginestra. 2022. *Multilayer Networks: Structure and Function*. Oxford, New York: Oxford University Press.

Bove, Riley, Tanuja Chitnis, Bruce AC Cree, Mar Tintoré, Yvonne Naegelin, Bernard MJ Uitdehaag, Ludwig Kappos, et al. 2018. "SUMMIT (Serially Unified Multicenter Multiple Sclerosis Investigation): Creating a Repository of Deeply Phenotyped Contemporary Multiple Sclerosis Cohorts." *Multiple Sclerosis Journal* 24 (11): 1485–98. https://doi.org/10.1177/1352458517726657.

Brownlee, Jason. 2021. *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*.

Bullmore, Edward, and Olaf Sporns. 2009. "Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems." *Nature Reviews. Neuroscience* 10 (March): 186–98. https://doi.org/10.1038/nrn2575.

Casas-Roma, Jordi, Eloy Martinez-Heras, Albert Solé-Ribalta, Elisabeth Solana, Elisabet Lopez-Soley, Francesc Vivó, Marcos Diaz-Hurtado, et al. 2022. "Applying Multilayer Analysis to Morphological, Structural, and Functional Brain Networks to Identify Relevant Dysfunction Patterns." *Network Neuroscience* 6 (3): 916–33. https://doi.org/10.1162/netn_a_00258.

Chard, Declan T., Adnan A. S. Alahmadi, Bertrand Audoin, Thalis Charalambous, Christian Enzinger, Hanneke E. Hulst, Maria A. Rocca, et al. 2021. "Mind the Gap: From Neurons to Networks to Outcomes in Multiple Sclerosis." *Nature Reviews Neurology* 17 (3): 173–84. https://doi.org/10.1038/s41582-020-00439-8.

Chard, Declan, and S Anand Trip. 2017. "Resolving the Clinico-Radiological Paradox in Multiple Sclerosis." *F1000Research* 6 (October): 1828. https://doi.org/10.12688/f1000research.11932.1.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16 (June): 321–57. https://doi.org/10.1613/jair.953.

De Domenico, Manlio. 2020. "Multilayer Networks Illustrated," November. https://doi.org/10.17605/OSF.IO/GY53K.

———. 2022. *Multilayer Networks: Analysis and Visualization: Introduction to MuxViz with R*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-75718-2.

De Domenico, Manlio, Mason A. Porter, and Alex Arenas. 2015. “MuxViz: A Tool for Multilayer Analysis and Visualization of Networks.” *Journal of Complex Networks* 3 (2): 159–76. https://doi.org/10.1093/comnet/cnu038.

Fleischer, Vinzenz, Adriane Gröger, Nabin Koirala, Amgad Droby, Muthuraman Muthuraman, Pierre Kolber, Eva Reuter, Sven Meuth, Frauke Zipp, and Sergiu Groppa. 2016. “Increased Structural White and Grey Matter Network Connectivity Compensates for Functional Decline in Early Multiple Sclerosis.” *Multiple Sclerosis* 23 (May). https://doi.org/10.1177/1352458516651503.

Fleischer, Vinzenz, Angela Radetz, Dumitru Ciolac, Muthuraman Muthuraman, Gabriel Gonzalez-Escamilla, Frauke Zipp, and Sergiu Groppa. 2019. “Graph Theoretical Framework of Brain Networks in Multiple Sclerosis: A Review of Concepts.” *Neuroscience*, Non-invasive MRI windows on brain inflammation, 403 (April): 35–53. https://doi.org/10.1016/j.neuroscience.2017.10.033.

Fornito, Alex, Zalesky, Andrew, and Bullmore, Edward. n.d. *Fundamentals of brain network analysis*. 2016th ed. Elsevier Academic Press.

Fortin, Jean-Philippe, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, et al. 2018. “Harmonization of Cortical Thickness Measurements across Scanners and Sites.” *NeuroImage* 167 (February): 104–20. https://doi.org/10.1016/j.neuroimage.2017.11.024.

Fortin, Jean-Philippe, Drew Parker, Birkan Tunç, Takanori Watanabe, Mark A. Elliott, Kosha Ruparel, David R. Roalf, et al. 2017. “Harmonization of Multi-Site Diffusion Tensor Imaging Data.” *NeuroImage* 161 (November): 149–70. https://doi.org/10.1016/j.neuroimage.2017.08.047.

Gironés Roig, Jordi, Jordi Casas Roma, and Julià Minguillón Alfonso. 2017. *Minería de datos: modelos y algoritmos*. Tecnología. Barcelona: Editorial UOC.

Guillon, Jeremy, Mario Chavez, Federico Battiston, Yohan Attal, Valentina La Corte, Michel Thiebaut de Schotten, Bruno Dubois, Denis Schwartz, Olivier Colliot, and Fabrizio De Vico Fallani. 2019. “Disrupted Core-Periphery Structure of Multimodal Brain Networks in Alzheimer’s Disease.” *Network Neuroscience* 3 (2): 635–52. https://doi.org/10.1162/netn_a_00087.

Haider, Lukas, Tobias Zrzavy, Simon Hametner, Romana Höftberger, Francesca Bagnato, Günther Grabner, Siegfried Trattnig, Sabine Pfeifenbring, Wolfgang Brück, and Hans Lassmann. 2016. “The Topograpy of Demyelination and Neurodegeneration in the Multiple Sclerosis Brain.” *Brain* 139 (3): 807–15. https://doi.org/10.1093/brain/awv398.

Heuvel, Martijn P. van den, and Olaf Sporns. 2011. “Rich-Club Organization of the Human Connectome.” *Journal of Neuroscience* 31 (44): 15775–86. https://doi.org/10.1523/JNEUROSCI.3539-11.2011.

Hsu, Jung-Lung, Alexander Leemans, Chyi-Huey Bai, Cheng-Hui Lee, Yuh-Feng Tsai, Hou-Chang Chiu, and Wei-Hung Chen. 2008. “Gender Differences and Age-Related White Matter Changes of the Human Brain: A Diffusion Tensor Imaging Study.” *NeuroImage* 39 (2): 566–77. https://doi.org/10.1016/j.neuroimage.2007.09.017.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-1418-1.

Kennedy, Keith E., Nicole Kerlero de Rosbo, Antonio Uccelli, Maria Cellerino, Federico Ivaldi, Paola Contini, Raffaele De Palma, et al. 2023. “Multiscale Networks in Multiple Sclerosis.” bioRxiv. https://doi.org/10.1101/2023.02.26.530153.

Kocevar, Gabriel, Claudio Stamile, Salem Hannoun, François Cotton, Sandra Vukusic, Françoise Durand-Dubief, and Dominique Sappey-Marinier. 2016. “Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses.” *Frontiers in Neuroscience* 10. https://www.frontiersin.org/articles/10.3389/fnins.2016.00478.

Llufriu, Sara, Eloy Martinez-Heras, Elisabeth Solana, Nuria Sola-Valls, Maria Sepulveda, Yolanda Blanco, Elena H. Martinez-Lapiscina, et al. 2016. "Structural Networks Involved in Attention and Executive Functions in Multiple Sclerosis." *NeuroImage : Clinical* 13 (December): 288–96. https://doi.org/10.1016/j.nicl.2016.11.026.

Mandke, Kanad, Jil Meier, Matthew J. Brookes, Reuben D. O'Dea, Piet Van Mieghem, Cornelis J. Stam, Arjan Hillebrand, and Prejaas Tewarie. 2018. "Comparing Multilayer Brain Networks between Groups: Introducing Graph Metrics and Recommendations." *NeuroImage* 166 (February): 371–84. https://doi.org/10.1016/j.neuroimage.2017.11.016.

Martí-Juan, Gerard, Jaume Sastre-Garriga, Eloy Martinez-Heras, Angela Vidal-Jordana, Sara Llufriu, Sergiu Groppa, Gabriel Gonzalez-Escamilla, et al. 2023. "Using The Virtual Brain to Study the Relationship between Structural and Functional Connectivity in Patients with Multiple Sclerosis: A Multicenter Study." *Cerebral Cortex*, February, bhad041. https://doi.org/10.1093/cercor/bhad041.

Nabizadeh, Fardin, Soroush Masrouri, Elham Ramezannezhad, Ali Ghaderi, Amir Mohammad Sharafi, Soroush Soraneh, and Abdorreza Naser Moghadasi. 2022. "Artificial Intelligence in the Diagnosis of Multiple Sclerosis: A Systematic Review." *Multiple Sclerosis and Related Disorders* 59 (March): 103673. https://doi.org/10.1016/j.msard.2022.103673.

Pontillo, Giuseppe, Ferran Prados, Alle Meije Wink, Baris Kanber, Alvino Bisecco, Tommy A. A. Broeders, Arturo Brunetti, et al. 2022. "More than the Sum of Its Parts: Disrupted Core-Periphery of Multiplex Networks in Multiple Sclerosis." medRxiv. https://doi.org/10.1101/2022.12.17.22283623.

Schoonheim, Menno M., Tommy A.A. Broeders, and Jeroen J.G. Geurts. 2022. "The Network Collapse in Multiple Sclerosis: An Overview of Novel Concepts to Address Disease Dynamics." *NeuroImage : Clinical* 35 (July): 103108. https://doi.org/10.1016/j.nicl.2022.103108.

Schoonheim, Menno M., Kim A. Meijer, and Jeroen J. G. Geurts. 2015. "Network Collapse and Cognitive Impairment in Multiple Sclerosis." *Frontiers in Neurology* 6 (April): 82. https://doi.org/10.3389/fneur.2015.00082.

Seccia, Ruggiero, Silvia Romano, Marco Salvetti, Andrea Crisanti, Laura Palagi, and Francesca Grassi. 2021. "Machine Learning Use for Prognostic Purposes in Multiple Sclerosis." *Life* 11 (2): 122. https://doi.org/10.3390/life11020122.

Shu, Ni, Yunyun Duan, Mingrui Xia, Menno M. Schoonheim, Jing Huang, Zhuoqiong Ren, Zheng Sun, et al. 2016. "Disrupted Topological Organization of Structural and Functional Brain Connectomes in Clinically Isolated Syndrome and Multiple Sclerosis." *Scientific Reports* 6 (July): 29383. https://doi.org/10.1038/srep29383.

Solana, Elisabeth, Eloy Martinez-Heras, Jordi Casas-Roma, Laura Calvet, Elisabet Lopez-Soley, Maria Sepulveda, Nuria Sola-Valls, et al. 2019. "Modified Connectivity of Vulnerable Brain Nodes in Multiple Sclerosis, Their Impact on Cognition and Their Discriminative Value." *Scientific Reports* 9 (1): 20172. https://doi.org/10.1038/s41598-019-56806-z.

Solana, Elisabeth, Eloy Martinez-Heras, Elena H. Martinez-Lapiscina, Maria Sepulveda, Nuria Sola-Valls, Nuria Bargalló, Joan Berenguer, et al. 2018. "Magnetic Resonance Markers of Tissue Damage Related to Connectivity Disruption in Multiple Sclerosis." *NeuroImage: Clinical* 20 (January): 161–68. https://doi.org/10.1016/j.nicl.2018.07.012.

Sporns, Olaf. 2018. "Graph Theory Methods: Applications in Brain Networks." *Dialogues in Clinical Neuroscience* 20 (2): 111–21. https://doi.org/10.31887/DCNS.2018.20.2/osporns.

Vaiana, Michael, and Sarah Feldt Muldoon. 2020. "Multilayer Brain Networks." *Journal of Nonlinear Science* 30 (5): 2147–69. https://doi.org/10.1007/s00332-017-9436-8.

Welton, Thomas, Cris S. Constantinescu, Dorothee P. Auer, and Rob A. Dineen. 2020. "Graph Theoretic Analysis of Brain Connectomics in Multiple Sclerosis: Reliability and Relationship with Cognition." *Brain Connectivity* 10 (2): 95–104. https://doi.org/10.1089/brain.2019.0717.

Zhao, Yijun, Tong Wang, Riley Bove, Bruce Cree, Roland Henry, Hrishikesh Lokhande, Mariann Polgar-Turcsanyi, et al. 2020. "Ensemble Learning Predicts Multiple Sclerosis Disease Course in the SUMMIT Study." *Npj Digital Medicine* 3 (1): 1–8. https://doi.org/10.1038/s41746-020-00338-8.

Zurita, Mariana, Cristian Montalba, Tomás Labbé, Juan Pablo Cruz, Josué Dalboni da Rocha, Cristián Tejos, Ethel Ciampi, Claudia Cárcamo, Ranganatha Sitaram, and Sergio Uribe. 2018. "Characterization of Relapsing-Remitting Multiple Sclerosis Patients Using Support Vector Machine Classifications of Functional and Diffusion MRI Data." *NeuroImage. Clinical* 20: 724–30. https://doi.org/10.1016/j.nicl.2018.09.002.

9. Appendices

9.1. Graph measures definitions

Only definitions of measures used in this work are provided. All definitions are from this references (Fornito et al. 2016), (Bianconi 2022) and (Barabási, Albert-László 2016), though no explicit reference will be made in each definition

Strength

In an undirected network, the degree k_i of a node i is the number of edges incident to it. When network is weighted, there is an analog of node degree which is **node strength, s_i,** defined as the sum of the weights of the edges incident to node i

$$s_i = \sum_{j \neq i} \omega_{ij} \qquad (14)$$

Where ω_{ij} is the weight of the edge connecting nodes i and j.

In directed networks, a difference between edges leaving the node and edges arriving to the node can be made, and this way define an out-degree and in-degree: k_i^{out}, k_i^{in}

Clustering coefficient

The **local clustering** coefficient C_i of a node i of degree k_i measures the probability that two neighbors of node i are also connected. It is defined counting the number of triangles that node i is part of:

$$C_i = \begin{cases} \dfrac{2 \cdot \#number\ of\ triangles\ passing\ through\ 1}{k_i(k_i-1)} & for\ k_i > 1, \\ 0 & for\ k_i = 0{,}1 \end{cases} \quad (15)$$

Global clustering C is the average of the local clustering over all the nodes. In a network with N nodes:

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \quad (16)$$

An alternative definitions of global clustering is found in literature, quantity defined this way Is called **transitivity**:

$$T = 3\frac{\#\ triangles\ in\ the\ network}{\#\ distinct\ paths\ of\ length\ 2} \quad (17)$$

Both transitivity and clustering values range from zero to one. Values close to one indicate a network with large number of triangles and thus heavily interconnected.

Shortest Path

The shortest path between two nodes is the path with the fewest number of edges, and is often called distance and denoted by d_{ij}.

The average or **mean shortest path** is the average of shortest paths between any two nodes of the network. Therefore, in a network with N nodes:

$$\bar{\ell} = \frac{1}{N(N-1)}\sum_{i=i}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} d_{ij} \quad (18)$$

Another significant measure derivable from distance is the **diameter of the network**, defined as the maximum of the shortest distances between any two nodes of the network.

$$D = max(d_{ij}) \quad (19)$$

Path length in a *weighted* network is calculated as the *sum of the weights* of the edges in the path.

Efficiency

Global Efficiency is closely related to mean path length as is defined as the inverse of distances

$$E_{glob} = \frac{1}{N(N-1)} \sum_{i=i}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{d_{ij}} \quad (20)$$

A local efficiency can be defined also, for a node i

$$E_i = \frac{1}{N_{G_i}(N_{G_i}-1)} \sum_{j,h \in G_i} \frac{1}{d_{jh}} \quad (21)$$

Where G_i denotes the subgraph that comprises all nodes connected to i, but once removed node i and all its incident edges.

Modularity

In a network, a community is formed by a set of nodes that are more densely connected to each other than to the rest of the network.

Modularity defines a way of comparing the density of the edges within each community to the expected value in the hypothesis that a link between two generic nodes i and j occur with probability p_{ij}. Null hypothesis is that network is completely random, while preserving same degree distribution as the real graph. Taking all of this into account, modularity is thus defined:

$$Q = \frac{1}{\langle k \rangle N} \sum_{ij} (a_{ij} - p_{ij}) \delta_{g_i, g_j} \quad (22)$$

Where δ_{g_i, g_j} is the Kronecker delta function and a_{ij} are the elements of the adjacency matrix.

This definition can be extended to weighted networks

$$Q = \frac{1}{2W}\sum_{ij}(\omega_{ij} - p_{ij}^{\omega})\delta_{g_i,g_j} \quad (23)$$

Where W is the total weight of the unique edges of the network, and now p is the probability but given by strength ($p_{ij}^{\omega} = {}^{s_i s_j}/_{W}$).

Modularity is a measure that can evaluate the significance of a community.

Density

Density has a quite intuitive definition, is the ratio between the number of edges in a graph and the largest possible number of edges in that graph. In a undirected graph, maximum number of edges is N(N-1)/2 where N is the number of nodes, so density is:

$$D = 2\frac{\#\ edges}{N(N-1)} \quad (24)$$

This a concept not defined for weighted graphs.

Centrality (Closeness Centrality and PageRank)

In a common network, different nodes have different number of links that connect them to the rest of the network, therefore playing different roles in that network. It is reasonable to assume that highly connected vertices exert a more important role in the network.

If the goal is to rank the nodes in order of their “importance”, strength only gives a partial account of the picture. There are different measures aimed to this purpose, two of them are closeness centrality and PageRank

Closeness centrality measures how close a node is from the rest of the nodes in a graph. To do this one computes shortest distance d from that node to the rest of nodes, a node with short average path length is able to interact efficiently with many nodes, and also non central nodes will be able to reach that node easily. Formal definition of closeness centrality of a node i:

$$C_{C_i} = \frac{N-1}{\sum_{i\neq j} d_{ij}} \quad (25)$$

It is important to note how similar is this definition to the one for efficiency.

In weighted networks, this measure can be calculated using shortest weighted path. It is also important to note that if two nodes are disconnected, $d_{ij} = \infty$

Closeness centrality can be applied to multilayer networks, in **multilayer closeness centrality**, instead of distances confined to one layer, multilayer shortest-path is considered.

Although it seems fair to consider a node more important (i.e.: it has more centrality) if it has a link to a very important node, It has to be noted that it is likely high centrality nodes may have many links. As this measure was developed for webpages it is easy to understand with an example from that field: If a very important webpage contains a lot of URL links, the prestige of the pointed webpages is not the same as if the webpage contained only a very few and selected links. **PageRank centrality x_i** of node i satisfies:

$$x_i = \mu \sum_{j=1}^{N} a_{ij} \frac{1}{\kappa_j^{out}} x_j + \omega \qquad (26)$$

Where $\kappa_j^{out} = \max(k_j^{out}, 1)$, this way if a node has no out-degree there is no divide by zero problem[1], and μ is a parameter between 0 and 1. Last term is given by

$$\omega = \frac{1}{N} \sum_{j=1}^{N} [(1-\mu) + \mu \delta(k_j^{out}, 0] x_j \qquad (27)$$

With $\delta(x, y)$ indicating the Kronecker delta.

Multilayer PageRank Centrality

PageRank centrality can be extended to multilayer networks in order to evaluate the centrality of nodes considering all the layers across the network. It captures the effect of the centrality of a node in one layer on its centrality in another layer. A fine example is when a famous actor takes part in a political cause, his centrality in one layer (movies) influences his centrality in another layer (political). The **multilayer PageRank $X_i(q, n)$** of node i depends on two parameters $q, n \in [0,1]$ and satisfies

$$X_i(q,n) = \mu \sum_{j=1}^{N} (x_i)^q a_{ij}{}^{[\beta]} \frac{X_j(q,n)}{\kappa_j} + \omega \left(\frac{x_i}{\langle x \rangle} \right)^n \qquad (28)$$

Where κ_j and ω are given by

[1] In this case a_{ij} will be zero and result will be zero but having a zero in the denominator is avoided

$$\kappa_j = max\left(\sum_{r=1}^{N} a_{jr}^{[\beta]}(x_r)^q, 1\right) \quad (29)$$

$$\omega = \frac{1}{N}\sum_{j=1}^{N}\left[(1-\mu) + \mu\delta(\sum_{r=1}^{N} a_{jr}^{[\beta]}(x_r)^q, 0\right] X_j(q,n) \quad (30)$$

9.2. Libraries and packages

In the following tables all packages imported during this project are shown, along with their respective versions. It is worth noting that these table do not include dependent libraries that may have been automatically imported as requirements.

R 4.2.3	
Package	Version
igraph	1.4.2
ggplot2	3.4.2
muxViz	3.1
patchwork	1.1.2
tidyverse	2.0.0

Table 25. R packages and versions.

Python 3.10.10	
library	Version
Imbalanced-learn	0.10.1
Matplotlib	3.7.1
networkX	2.8.4
neuroCombat	0.12.12
numpy	1.23.5
nxviz	0.7.4
Pandas	1.5.3

scikit-learn	1.2.2
Scipy	1.10.0
Seaborn	0.12.2
statsmodels	0.13.5
xgboost	1.7.3

Table 26. Python libraries and versions

9.3. GitHub Repository

Here is the URL to the gitHub repository where code used in this work can be found. Along with the code, images produced by it and a HTML version of R markdown and jupyter notebook can also be found.

https://github.com/JoanGinard/Multilayer_MS