



HAL
open science

Graphes pour l'exploration des réseaux de neurones artificiels et de la connectivité cérébrale humaine

Lucrezia Carboni

► **To cite this version:**

Lucrezia Carboni. Graphes pour l'exploration des réseaux de neurones artificiels et de la connectivité cérébrale humaine. Réseau de neurones [cs.NE]. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALM060 . tel-04552217

HAL Id: tel-04552217

<https://theses.hal.science/tel-04552217>

Submitted on 19 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique

Spécialité : Mathématiques Appliquées

Unité de recherche : Laboratoire Jean Kuntzmann

Graphes pour l'exploration des réseaux de neurones artificiels et de la connectivité cérébrale humaine

GRAACE: GRaphs for Artificial neural networks and brAin Connectivity Exploration

Présentée par :

Lucrezia CARBONI

Direction de thèse :

Sophie ACHARD

CNRS Senior Researcher, Université Grenoble Alpes

Directrice de thèse

Michel DOJAT

DIRECTEUR DE RECHERCHE, INSERM DELEGATION AUVERGNE-RHONE-ALPES

Co-directeur de thèse

Rapporteurs :

BERTRAND THIRION

DIRECTEUR DE RECHERCHE, INRIA CENTRE DE SACLAY-ILE-DE-FRANCE

NICOLAS FARRUGIA

MAITRE DE CONFERENCES HDR, IMT ATLANTIQUE BREST

Thèse soutenue publiquement le **20 octobre 2023**, devant le jury composé de :

NADIA BRAUNER,

PROFESSEURE DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES

Présidente

SOPHIE ACHARD,

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES

Directrice de thèse

MICHEL DOJAT,

DIRECTEUR DE RECHERCHE, INSERM DELEGATION AUVERGNE-RHONE-ALPES

Co-directeur de thèse

BERTRAND THIRION,

DIRECTEUR DE RECHERCHE, INRIA CENTRE DE SACLAY-ILE-DE-FRANCE

Rapporteur

NICOLAS FARRUGIA,

MAITRE DE CONFERENCES HDR, IMT ATLANTIQUE BREST

Rapporteur

ANA MARQUES,

PRATICIENNE HOSPITALIERE HDR, CHU DE CLERMONT-FERRAND

Examinatrice

GIULIA PRETI,

DOCTEURE EN SCIENCES, CENTAI INSTITUTE S.P.A.

Examinatrice

Invités :

MARINA REYBOZ

INGENIEURE HDR, CEA CENTRE DE GRENOBLE

PAULO GONÇALVES

DIRECTEUR DE RECHERCHE, INRAE CENTRE GRENOBLE-RHONE-ALPES



Alla mia mamma e al mio papà

Acknowledgements

I like how the word 'acknowledgment' contains the word 'knowledge' within it, as if to thank is also to understand, to gain comprehension. And after these three years of research, perhaps I have grasped something myself.

I would like to express my sincere gratitude to the examiners and reviewers for dedicating their time, expertise, and effort to evaluate my work.

Your constructive critiques and thoughtful suggestions have undoubtedly strengthened the quality of this thesis, and I am genuinely grateful for your commitment. I want to deeply thank Michel and Sophie. Since we met three years ago, you have never failed to provide me with your trust, patient assistance, and encouraging support, pushing myself to always give more and go further. Without a doubt, without meeting you, I would be a different person today and wouldn't be where I am.

I would like to thank all my collaborators, whose contributions have significantly enriched and strengthened my research, Hanâ, Dwight, Francesco, Marion, Veronica, Martial, Simone, Paul, Marina, Youssef. I have had the privilege of conducting my research in two wonderful teams ready to engage in scientific debates, soccer tournaments, and incredible adventures.

A big thanks goes to people who showed me the way, inspiring, reassuring and guiding me. Among these, a huge thanks goes to Aurélien who share with me the office, the desks -my desk!-, the late working hours, my stress and this template. As well as to Masha who never stops bringing effervescent energy to my life bringing me to incredible adventures being the person you can count on. Then, to Dasha, silent presence you can always count on and to Meryem, for your words spoken from the heart.

I want to thank those who started with me and were true companions on this journey: Hanâ, for your authenticity; Minh Tri, for your adventurous spirit; Benjamin, for bringing me to Universal Studios, and of course Théo, who has always been there, a friend, a confident, the certain presence at coffee breaks and beyond.

A special thanks to Argheesh, for the friendships, the missed meeting points (too many bridges in Grenoble) and for always keeping the conversation going. As well as to Pierre, who -rarely- climbed with me, but discussed with me science, politics, theatres, and so many topics. And thanks to Konstantinos for loosing with me at our pasta competition.

And then, thanks to you all: to Sébastien, who taught my how to climb and fixed my bike so many times, Thomas, who managed to match me with the good people -or not-, Jean Côme, who explained me the greenwashing while preparing the perfect *tarte au citron*, to Marc who still needs to climb up la Bastille and to Clément for sailing with me the Route du Rhum. Thanks to Louise for her genuine spontaneity. Thanks to Yuchen who tried to teach me Chinese poems and how to play poker, even

if unsuccessfully. Thanks to Diego and Stenzel for sharing with me the office in the good moment, and the bad ones. Thank to Jacopo, and the babyfoot all'italian and Clothilde, who listened to my best french rehearsal, to Loan, Julien and Geoffroy for their irony. Thanks to Alicia, David, Trung Tin for their quiet ambition and to Bayan, Vaëa, and Chiara for their serenity and dedication.

I am grateful for knowing each and every one of you and I am sure we will debate ultimate Frisbee rules, once more. A big thanks goes to all the people who were ready to engage in crazy missions, soccer tournaments, toga parties and workshop organization from all my research labs and institutions, MIAI, LJK, GIN and Inria.

A big thank goes to all my friends for being *forever in Grenoble* or for nurturing our connections despite the many kilometers. Thanks for all our apéro, our parties, our hikes, our incredible road-trips and for sharing our thoughts, frustrations, and happiness! Thank you Zei, Johana, Aidan, Xhenis, Catarina, Tjana, Pedja, Carolina, Thibault, Kristi, Théo.

Thank you to all those who shared a roof with me — Leonardo, Rosetta, Quentin, Lucie, Arianna, Blandine, Maria Teresa, Giovanni, Pauline, Martina — for enduring my mess, my no-speaking-before-coffee moments, and for getting along so well despite our differences.

Thanks also to my lifelong (or so) friends, Ale, Silvia, Marco, Lucia, Chiara, Gianlu, Matteo, Veri, Fra, and Alex for always been by my side.

To Diletta and Carlotta, even though you're miles away, you're always there to tease, advise, argue, assist, and vent. Thanks to our parents who supported us without hesitation, allowing us to embark on this journey. And to all my family, a constant source of certainty and strength.

Abstract

The main objective of this thesis is to explore brain and artificial neural network connectivity from a graph-based perspective. While structural and functional connectivity analysis has been extensively studied in the context of the human brain, there is a lack of a similar analysis framework in artificial systems. To address this gap, this research focuses on two main axes.

In the first axis, the main objective is to determine a healthy signature characterization of the human brain resting state functional connectivity. A novel framework is proposed to achieve this objective, integrating traditional graph statistics and network reduction tools to determine healthy connectivity patterns. Hence, we build a graph pair-wise comparison and a classifier to identify pathological states and rank associated perturbed brain regions. Additionally, the generalization and robustness of the proposed framework are investigated across multiple datasets and variations in data quality.

The second research axis explores the benefits of brain-inspired connectivity exploration of artificial neural networks (ANNs) in the future perspective of more robust artificial systems development. A major robustness issue in ANN models is represented by catastrophic forgetting when the network dramatically forgets previously learned tasks when adapting to new ones. Our work demonstrates that graph modeling offers a simple and elegant framework for investigating ANNs, comparing different learning strategies, and detecting deleterious behaviors such as catastrophic forgetting. Moreover, we explore the potential of leveraging graph-based insights to effectively mitigate catastrophic forgetting, laying the foundations for future research and explorations in this area.

Résumé

L'objectif principal de cette thèse est d'explorer la connectivité cérébrale et celle des réseaux de neurones artificiels d'un point de vue de leur connectivité. Un modèle par graphes pour l'analyse de la connectivité structurelle et fonctionnelle a été largement étudié dans le contexte du cerveau humain, mais un tel cadre manque encore pour l'analyse des systèmes artificiels. Avec l'objectif d'intégrer l'analyse de la connectivité dans les systèmes artificiels, cette recherche se concentre sur deux axes principaux.

Dans le premier axe, l'objectif principal est de déterminer une caractérisation de la signature saine de la connectivité fonctionnelle de repos du cerveau humain. Pour atteindre cet objectif, une nouvelle méthode est proposée, intégrant des statistiques de graphe traditionnelles et des outils de réduction de réseau, pour déterminer des modèles de connectivité sains. Ainsi, nous construisons une comparaison en paires de graphes et un classifieur pour identifier les états pathologiques et identifier les régions cérébrales perturbées par une pathologie. De plus, la généralisation et la robustesse de la méthode proposée ont été étudiées sur plusieurs bases de données et variations de la qualité des données.

Le deuxième axe de recherche explore les avantages de l'intégration des études de la connectivité inspirée du cerveau aux réseaux de neurones artificiels (ANNs) dans la perspective du développement de systèmes artificiels plus robustes. Un problème majeur de robustesse dans les modèles d'ANN est représenté par l'oubli catastrophique qui apparaît lorsque le réseau oublie dramatiquement les tâches précédemment apprises lors de l'adaptation à de nouvelles tâches. Notre travail démontre que la modélisation par graphes offre un cadre simple et élégant pour étudier les ANNs, comparer différentes stratégies d'apprentissage et détecter des comportements nuisibles tels que l'oubli catastrophique. De plus, nous soulignons le potentiel d'une adaptation à de nouvelles tâches en contrôlant les graphes afin d'atténuer efficacement l'oubli catastrophique et jetant ainsi les bases de futures recherches et explorations dans ce domaine.

Contents

Contents	i
General Introduction	3
A Human Brain Connectivity	9
1 Context and Background	11
1.1 Functional Connectivity	13
1.1.1 Network-based methods for FC	15
1.2 State-of-the-art	17
1.2.1 PubMed Literature Search	19
1.2.2 Recall Graph-Theory	23
1.2.3 Classical Statistical Comparison	26
1.2.4 Classification Methods	31
1.3 Generative Networks	35
1.3.1 Erdős-Rényi model (ER)	35
1.3.2 Watts-Strogatz model (WS)	35
1.3.3 Barabási-Albert model	36
1.3.4 Degree sequence preserving model (DSP)	36
1.3.5 Economical preferential attachment model (EPA)	36
1.3.6 Economical clustering model (EC)	37
1.4 Summary and conclusion	37
2 Structural Pattern	39
2.1 Motivations	41
2.2 Nodal-statistics-based equivalence relation for graph collections	46
2.2.1 Structural equivalence for undirected unweighted graph . .	47
2.2.2 Structural equivalence for graph collections	51
2.2.3 General properties	54
2.3 Summary and conclusion	57
3 Applications	59
3.1 Databases	61

3.1.1	Preprocessing of rs-fMRI data	61
3.1.2	Adopted network inference protocol	62
3.2	(1) Synthetic Data and Real Data	65
3.2.1	Material	65
3.2.2	Results and Discussion	65
3.3	(2) Healthy Population Characterization	73
3.3.1	Material	73
3.3.2	Results and Discussion	73
3.4	(3) Discrimination of Patients Population	76
3.4.1	Comatose Patients	77
3.4.2	Parkinson Disease	83
3.5	(4) Reliability	89
3.5.1	Material and Method	89
3.5.2	Results and Discussion	91
3.6	Summary and conclusion	96
3.6.1	Limitations	97
3.6.2	Future work	98
B	Artificial Neural Networks	101
4	Context and Background	103
4.1	Artificial Neural Networks	105
4.1.1	Robust Artificial Neural Network	109
4.2	Continual Learning	110
4.2.1	Continual Learning Scenarios	112
4.2.2	Continual Learning Strategies	114
4.3	Network Science in ANNs	121
5	Graph-Based Analysis of ANN Connectivity	123
5.1	Introduction	126
5.2	Graph-based method for ANN connectivity analysis	126
5.2.1	Activation Network and Induced Graph	127
5.2.2	Graph statistics of interest	129
5.2.3	Graph-informed Features	129
5.2.4	Network Surgery: Nodal Role Identification	131
5.2.5	Graph-informed Weight Injection	131
5.3	Evaluation Metrics	133
5.4	Related Works and Limitations	134

6	Continual Learning Strategy Connectivity Characterization	137
6.1	Introduction	139
6.2	Material	139
6.2.1	Handwritten Digit Recognition	139
6.2.2	Face Emotion Recognition	141
6.3	Experiments	144
6.3.1	Experiment 1: Characterization	145
6.3.2	Experiment 2: Network Surgery	148
6.3.3	Experiment 3: Weights-injection	152
6.4	Summary and conclusion	157
6.4.1	Limitations	158
6.4.2	Future work	158
	General Conclusion	163
	List of Figures	I
	List of Tables	V
	List of Abbreviations	VII
C	Appendix	IX
A	Supplementary Material	XI
A.1	PubMed Query Results	XI
A.2	Bland-Altman Plots	XX
A.3	Structural Pattern under Uncertainty	XXV
A.3.1	Fuzzy Network Definition	XXV
A.3.2	Uncertainty in nodal-statistics-based equivalence	XXVII
A.4	Complete Network Surgery Results	XXIX
B	Related Work	XXXI
B.1	Network Embedding for Brain Connectivity	XXXI
B.2	Label-informed persistence diagrams comparison	XXXVI
B.3	Homological Information for Small-World Regime Refinement	LX

B.4 Adversarial attacks detection through ANN-induced-graphs	LXII
Compliance with Ethical Standards	LXXXVII
Published works	LXXXIX
Résumé français - French summary	XCI

General Introduction

The brain is certainly one of the most investigated human organs and still remains one of the most cryptic. Understanding how the brain works is one of the greatest challenges of current neuroscience with multiple implications across diverse disciplines such as: anatomy, psychology, clinical neuroscience, but also philosophy, ethics, and more contemporary domains such as neuromorphic computing, automated systems, and artificial intelligence.

In a clinical perspective, the progress of technology and the advancement of imaging techniques have given the possibility to obtain an accurate and precise reproduction of the brain anatomy, including neurons and blood vessels, together with a real-time monitoring of brain functionality with significant impacts on human health.

These advances manifest in a spectrum of applications. To name a few, they range from the application of artificial intelligence algorithm for the analysis of computed tomography scans for traumatic brain injury patients ([Brossard et al. 2021](#)), the integration of uncertainty estimation into sclerosis lesion segmentation ([Lambert et al. 2022](#)), to the pioneering development of magnetic resonance fingerprinting protocols ([Christen et al. 2014](#); [Delphin 2022](#)), or to the innovative use of microbubbles with ultrasound to breach the blood-brain barrier and design new drug transport ([Meairs et al. 2007](#)). Certainly, the use of advance methods to assist clinicians in their work is of particular interest to obtain accurate and early diagnosis, as it happens in breast cancer detection ([Yoon et al. 2023](#)), providing the development of new disease biomarkers.

However, an autonomic diagnosis tool alone does not equate to a comprehensive understanding of disease mechanisms, which remains of crucial importance for treatment refinement.

From this point of view, the emergence of **network neuroscience** holds great promise for its ability of providing a simple model for brain exploration. Certainly, the brain naturally acts as a network, both structurally and functionally and at different aggregation scales: synapse connections among single neurons, interconnections among neuronal regions, and cooperative connections among hemispheres. All these types of relations can be elegantly modeled as graphs on which well-founded mathematical definitions find a natural application. The use of such a graph model particularly for functional connectivity exploration can provide fascinating results and advance our comprehension of brain mechanisms under diverse states. When coming to the study of neurological disorders, network neuroscience can be crucial for identifying specific non-invasive biomarkers and for the definition of new therapies. For instance, in Alzheimer and Parkinson disease or for region mapping before surgical intervention ([Fox et al. 2010](#); [Du et al. 2018](#); [Zhang et al. 2021](#); [Oujamaa 2020](#); [Nandakumar et al. 2021](#)).

It is no coincidence that deep learning models simulating brain information processing are organized into networks. These artificial neural networks comprise multiple units or neurons and are constructed with increasingly complex and intricate structural architectures. Since their introduction, artificial neural networks (ANNs) have been developed to emulate the information processing, learning, and decision-making mechanisms of biological neural systems and have been studied with the aim of addressing

fundamental questions about how humans perceive the world, store and recall information, and how this information influences their behavior (Rosenblatt 1958; Bengio et al. 2015; Hassabis et al. 2017).

With this purpose, many research laboratories actively work towards creating ANN models that resemble the human brain more closely (Stanford Artificial intelligence Laboratory, Center for Brains, Minds, and Machines, Google Brain, etc.).

The integration of neuroscience and artificial intelligence extends beyond the development of brain-inspired neural network systems and bio-inspired algorithms (Hassabis et al. 2017). From neuroscience to artificial intelligence, the study of the brain can lead to the design of more robust, interpretable, and explainable artificial systems. In fact, many algorithms find inspiration in natural intelligence or are constrained to *learn* through human-features (Klyuzhin et al. 2022; Ilyas et al. 2019). On the opposite direction, contributions of artificial intelligence into neuroscience allows to *read the human mind*, with the development of recent models able to reconstruct images from human brain activity recorded during functional magnetic resonance imaging (Takagi et al. 2023; Nishimoto et al. 2011; Poldrack 2011; Mensch et al. 2021; Zhang et al. 2022b) or to finely understand how information is processed in different brain areas (Bashivan et al. 2019; Kanwisher et al. 2023).

In light of this productive integration and exchange between neuroscience and artificial intelligence, the Multidisciplinary Institute of Artificial Intelligence in Grenoble dedicates a research axis to the development of robust and understandable Neuromorphic systems by leveraging psychology, cognitive science, informatics, neuroscience, neuroimaging, mathematics, and statistics. It is within this overarching framework that this Ph.D. project finds its roots, spanning both the Neuroscience Institute of Grenoble's *Functional Neuroimaging and Brain Perfusion* team and the Laboratoire Jean Kuntzmann's *Statify* team, with collaboration with the Psychology and Neurocognition Laboratory *LPNC*.

Particularly, the general scope of this thesis is to provide a network-science perspective for neuroscience into the analysis of **human brain functional connectivity networks**, and for a brain-inspired **artificial neural networks robustness assessment**.

Both branches share the development of innovative network analysis tools allowing an original integration.

The former research branch, concerning human brain functional connectivity exploration, aims to ultimately establish a network signature of a population group, to enhance our comprehension of the underlying brain mechanisms implicated in pathological dysfunction and possibly refining the nosology of brain disorders. To achieve this objective, a novel framework is proposed, integrating established graph statistics and network reduction tools, to determine connectivity patterns. Hence, we build a

graph pair-wise comparison and a classifier to identify pathological states and rank associated perturbed brain regions.

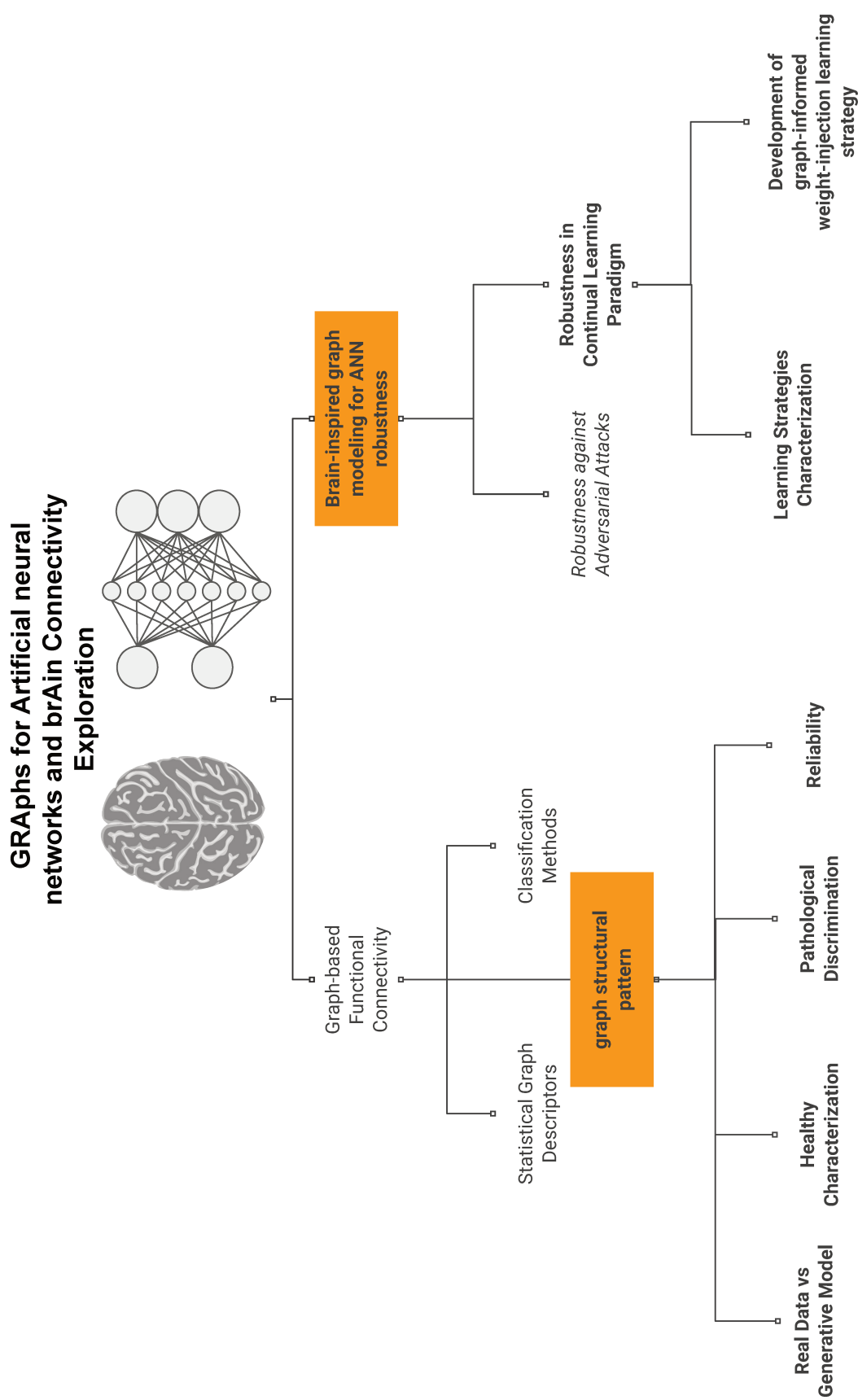
The latter research branch, concerning ANNs, specifically addresses artificial intelligence robustness issue by developing a brain-inspired framework of analysis to detect the occurrence of catastrophic forgetting. While the connectivity of the brain has been extensively studied and modeled, there is a limited number of works conducted in this area for ANNs and our proposal lays the foundation for future research and explorations in this area.

Following this main division, the present dissertation is organized into two main parts dedicated to these two distinct research axes.

The first part begins by contextualizing and providing background on functional brain connectivity graph modeling. It reviews the state-of-the-art in this domain and introduces the principal research questions. Chapter 1 also includes a concise recall of graph theory definitions. Chapter 2 introduces our main contribution, by giving its motivations and culminating in the definition of an innovative nodal equivalence relation for characterization of individual graph and graph collections at both global and local scale. Chapter 3 demonstrates the benefit of our proposal in four applications in human functional connectivity: characterizing generative synthetic data and real data, analyzing healthy subjects, developing a pathological classifier, and exploring data quality's influence on local characterization.

The second part of this manuscript is dedicated to artificial neural network systems. Chapter 4 introduces the robustness challenges in the development of artificial intelligence systems, focusing into the paradigm of continual learning and reviewing the existing strategies and scenarios in this context. Subsequently, Chapter 5 proposes a brain-inspired approach to analyse feedforward neural network by the development of a resting state graph model for ANNs. The final chapter, Chapter 6, applies our approach to characterize and explore different continual learning strategies on two architecture and image recognition tasks.

The appendix supplements our work with related studies, including investigations into data-quality considerations in functional brain connectivity in a topological data analysis fashion and the application of a graph-based model for the detection of adversarial attacks in ANN models.



In this schematic representation of the Ph.D. research, the main theoretical contributions are denoted by orange boxes. The studied subjects are highlighted in bold, italics indicates related contributions available in the appendix.

Part A

Human Brain Connectivity

— Chapter 1 —

Context and Background

Abstract

In this chapter, we delve into the analysis of functional connectivity (FC) studies, with a special focus on methods adopting a graph perspective. We show the diversity across studies and the resulting controversies on results. We introduce graph-theory concepts and graph descriptor commonly considered in functional connectivity studies. We emphasize the significance of interpretability and explainability in FC analysis, highlighting their role in understanding underlying physio-pathological mechanisms. We define these concepts in the context of the multilevel abstraction present in graph-based methods. Furthermore, we present generative network models that serve as effective tools for the comparative assessment of real data against synthetic counterparts. Finally, we identify the requirements for a comprehensive FC analysis framework, emphasizing interpretability, adaptability, group characterization, individual differences tracking, and local perturbation detection.

Contents

1.1	Functional Connectivity	13
1.1.1	Network-based methods for FC	15
1.2	State-of-the-art	17
1.2.1	PubMed Literature Search	19
1.2.2	Recall Graph-Theory	23
1.2.3	Classical Statistical Comparison	26
	Global Descriptors	26
	Global and Local Efficiency in FC: small-world brain networks.	27
	Local Descriptors	27
	Degree in FC: brain networks exhibit hubs.	28
	Clustering coefficient in FC: brain networks modularity.	28
	Centralities in FC networks: hubs detection.	29
1.2.4	Classification Methods	31
1.3	Generative Networks	35
1.3.1	Erdős-Rényi model (ER)	35
1.3.2	Watts-Strogatz model (WS)	35
1.3.3	Barabási-Albert model	36
1.3.4	Degree sequence preserving model (DSP)	36
1.3.5	Economical preferential attachment model (EPA)	36
1.3.6	Economical clustering model (EC)	37
1.4	Summary and conclusion	37

1.1 Functional Connectivity

The human brain represents only 2% of the total body mass, yet it is remarkably energy-demanding, consuming 20% of the body's energy even at rest (Rolfe et al. 1997; Fox et al. 2010). Neurons require oxygen and glucose for their activity, which are not stored directly in the cells, and instead supplied through blood flow. When neurons are activated, nearby capillaries experience increased blood flow to meet the energy demands. This results in a localized change of brain oxygenation levels which can be detected by functional magnetic resonance imaging (fMRI). Introduced in the early 1990s, fMRI captures the blood oxygen level-dependent (BOLD) signal as measure of brain activity (Ogawa et al. 1990; Mansfield et al. 1977; Pauling et al. 1936; Kwong 2012). The increase in the oxygenated hemoglobin alters the local magnetic properties of the blood, which can be detected by the fMRI scanner (Fig.1.1). The BOLD signal is commonly used to create activation maps, showing which brain regions are involved in specific tasks or responding to stimuli (Lv et al. 2018). Broadly, fMRI finds numerous applications in neuroscience and clinical studies, for instance in the detection of pathologies, or the design of new treatments (Fox et al. 2010; Wang et al. 2010; Bobholz et al. 2007).

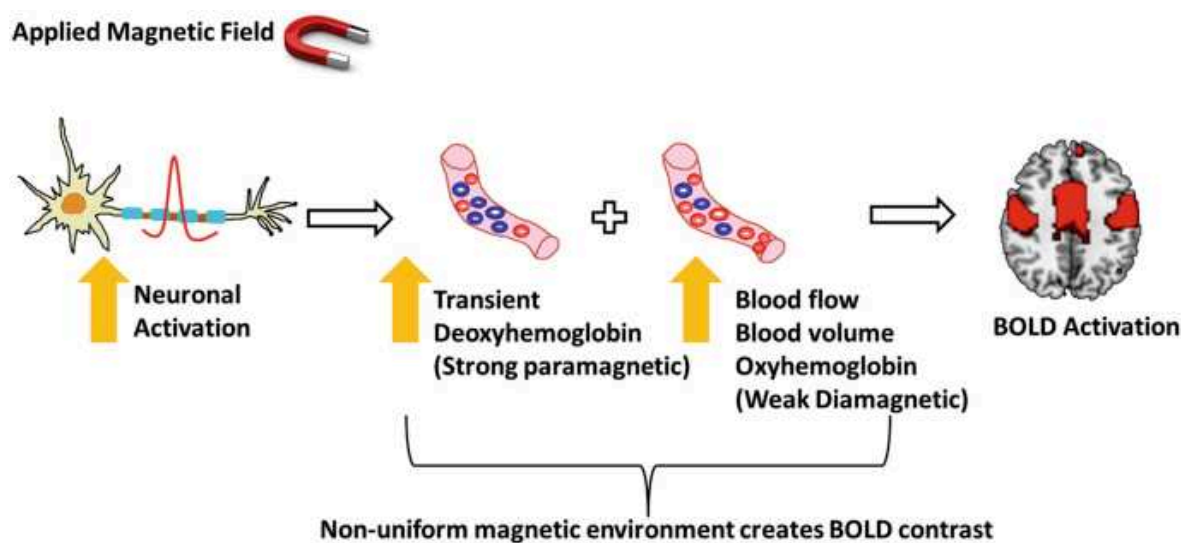


Figure 1.1 – Principle of fMRI: when neurons are activated, nearby capillaries respond by increasing blood flow to meet the energy demands. This results in a localized change in brain oxygenation levels, causing variations in the magnetic properties of blood. By quantifying these magnetic field changes, fMRI can measure the Blood Oxygen Level-Dependent (BOLD) signal, providing a non-invasive method to visualize and study brain activity. Adapted from Karunakaran et al. 2021.

fMRI allows to study brain activation during specific tasks and to acquire brain images while the scanned subject is at rest, letting the mind wandering and without performing any specific task. Resting state fMRI (rs-fMRI) has gained attention since it allows

to scan patients who may struggle with task-based instructions, such as those with neurological or psychiatric conditions or pediatric patients.

In the analysis of rs-fMRI we can distinguish two different approaches: functional segregation and functional integration¹ (Lv et al. 2018; Friston 2011). The former focuses on understanding the specific local functions of individual brain regions with the objective of associating to each region a particular function. On the other hand, functional integration analyses the brain in terms of the relationships or connectivity between different brain areas, considering the whole brain as an interconnected network. Our work adopts an integration point of view and explore the statistical dependencies or correlations between the activity of different brain regions. We call this functional connectivity (FC).

There are multiple metrics and methods currently used for functional connectivity studies: independent component analysis (ICA), seed-based FC analysis and graph-based analysis (Lv et al. 2018; Friston 2011; Van Den Heuvel et al. 2010).

ICA is a data-driven technique used to identify independent components in brain activity. It aims to separate the brain data into independent sources that represent distinct functional networks. While ICA can be effective in identifying individual components, it may not provide a comprehensive view of the whole brain network and its global properties (McKeown et al. 1998; Calhoun et al. 2009; Zuo et al. 2010; Varoquaux et al. 2010b; Beckmann et al. 2004; Calhoun et al. 2001).

Seed-based FC analysis involves selecting a specific brain region - named the *seed* - and examining its functional connectivity with other regions in the brain. This method is useful for investigating the connectivity of a specific region of interest. However, it focuses on pairwise connections and may not capture the complex interactions and global properties of the entire brain network (Biswal et al. 2010; Joel et al. 2011; Tang et al. 2021; Job et al. 2020; Bluhm et al. 2009; Fox et al. 2005; Fox et al. 2007).

Here, we assume a network perspective and adopt a graph-based analysis approach (Sporns 2016; Bullmore et al. 2009; Mheich et al. 2020; Bassett et al. 2017b; Smith et al. 2011; Vico Fallani et al. 2014). In this approach, the brain is modeled as a graph, a properly defined mathematical object which allows to encode the relationships among multiple units. Particularly, in our work, units represent brain regions and their relationships is encoded in edges.

This model offers several advantages, allowing to analyze the topology of brain networks, investigate local and global properties, and explore functional relationships be-

¹Note that here we refer to segregation and integration at the brain level. The same concepts can be found when considering a network model, see Def.1.13, Def.1.14 in Sec.1.2.3

tween brain regions. Within the network modelling for FC analysis, there is a diverse array of approaches. In the following section, we propose a categorization of the main existing methods.

1.1.1 Network-based methods for FC

Graph theory is a mathematical branch originated to solve complex system problems. Its origins track back to the famous Seven Bridges of Königsberg problem (Ráz 2018; Euler 1741), but its theoretical foundations and adaptability allow it to find applications in a variety of domains. In the context of brain imaging and specifically in resting state FC, it provides a mean to analyze the topology of brain networks. A network is a natural way to encode the pair-wise relations among in a set of units (the nodes or vertices). Graph theory is used to investigate both local properties (within specific brain regions) and global properties (across the entire brain) of functional brain networks. Regarding the temporal aspect of FC, two main modeling approaches are distinguished: static and dynamical models. In static models, the interaction between different brain regions is estimated over the entire available temporal length resulting in a single network per subject. On the other hand, dynamical models inferred multiple networks at different instant intervals to capture temporal changes in connectivity (Lurie et al. 2020; Varley et al. 2022). Our focus is dedicated to static model, but it can be extended for the investigation of dynamic FC.

A main issue of dynamical models is given by the choice of the number of instants to be considered for the estimation. A controversial point in static model is instead given by the fact that it implicitly assumes that the estimated dependency is static over time. The inference of network model requires crucial steps, particularly in defining units, the graph nodes and the links, the graph edges (Fig. 1.2). Defining units from fMRI data, where a unique time-series is extracted from each voxel, necessitates the selection of an aggregation method.

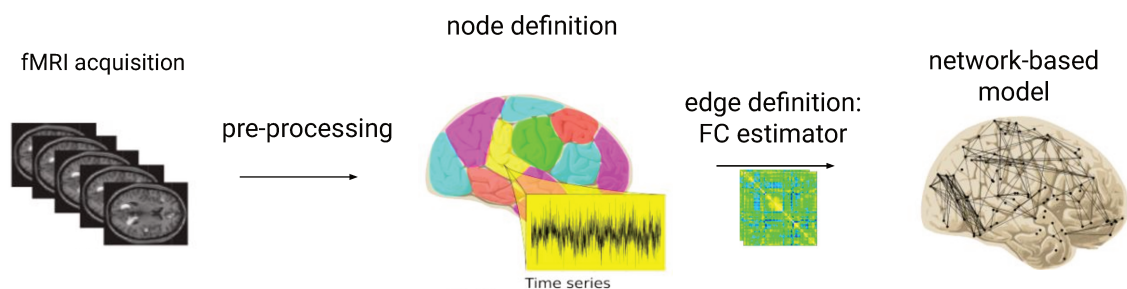


Figure 1.2 – Visualization of the required modeling steps to determine a FC network-based model from fMRI acquisition.

Two main approaches exist: anatomical-based and data-driven methods (Vico Fallani et al. 2014).

Anatomical-based methods use brain region parcellations to assign each voxel to a spe-

cific brain regions defined *a priori*. This approach enables the comparison of graph instances among different subjects while preserving regional identity, making it particularly useful for clinical studies. Another approach is to define vertices based on previously-specified coordinates and to include nearby voxels within the region, or even to consider each voxel as graph node.

On the other hand, data-driven approaches employ clustering algorithm or dictionary learning to define regions of interests, without prior anatomical information (Abraham et al. 2013; Varoquaux et al. 2011; Bhanot et al. 2019).

The second requirement for graph-modeling is given by the edge definition, corresponding to the choice of an estimator of FC interaction among brain regions. An established way to quantify the strength of interaction between a pair of regions is given by the Pearson correlation coefficient between signal of aggregated regions (Hlinka et al. 2011). The used of Pearson correlation implicitly assumes that the temporal order of the samples within each time series and their mutual interaction can be disregarded. Other methods adopt partial correlations, tangent space of covariance matrices, mutual information, etc (Smith et al. 2011; Dadi et al. 2019; Richiardi et al. 2013).

In the following, we adopt as FC estimator the Pearson correlation at a specific frequency scale, obtained by the application of the discrete wavelet transform (Achard et al. 2006; Achard et al. 2012). The discrete wavelet allows to decompose each fMRI time series into a set of compactly supported basis functions that are uniquely scaled in frequency and located in time. The use of wavelets considers the long-memory properties of fMRI time series and produces estimation with known variances based on the number of data points at each scale.

Recently, methods have been developed to simultaneously account for noise and intra-regional correlation impact, i.e. the correlation of signals within a region, on the inter-regional correlation estimation (Lbath et al. 2023; Achard et al. 2020).

After estimating functional connectivity using graph-based approaches, some authors apply a graph-filtering procedure to refine the network representation (Vico Fallani et al. 2014; De Vico Fallani et al. 2017; Achard et al. 2006; Achard et al. 2007). This step aims to select the most important weights and determine an unweighted graph, which can be crucial for better interpretation the underlying brain connectivity patterns.

Various filtering procedures exist, and they can vary across studies. Some methods are based on topological properties of the graph to identify the most significant connections (Bordier et al. 2017; Chen et al. 2008; Achard et al. 2006; Ferrarini et al. 2009). Others focus on determining a unique connected component within the graph, ensuring that all nodes are connected and form a coherent network (Bassett et al. 2006a). Proportional thresholding is another common technique, where a threshold is applied

to retain a certain percentage of the strongest edges, effectively sparsifying the graph and eliminating weaker connections (Achard et al. 2007; Bassett et al. 2009).

1.2 State-of-the-art

In the following section, we review existing methods to analyse FC brain networks. Specifically, we discuss works which consider undirected unweighted graph model extracted for each subject. As we mentioned, in a clinical perspective FC has the potential to serve as a non-invasive biomarker of pathological disease. Thus, we will discuss classification methods in the state-of-the-art which strive to assist clinicians in making diagnoses. Indeed, classification involves the process of categorizing items into predefined classes or categories based on their inherent characteristics or distinctive features.

We stress the importance in any FC network analysis of providing an improvement in our comprehension of the underlying physio-pathological mechanisms (Park et al. 2013; Vico Fallani et al. 2014). Indeed, FC analysis methods, especially when a classification task is included, should be evaluated in terms of their **interpretability** and **explainability**. We adopt these concepts in a broader sense with respect to their classical machine learning algorithm definition (Gilpin et al. 2018).

Within FC studies, any graph-based method introduces several levels of abstraction within its model: this commences from the brain, progresses through its FC network model, and culminates in the extraction of meaningful features from this graph (See Fig. 1.3). Hence, the interpretability and explainability of FC analysis framework's

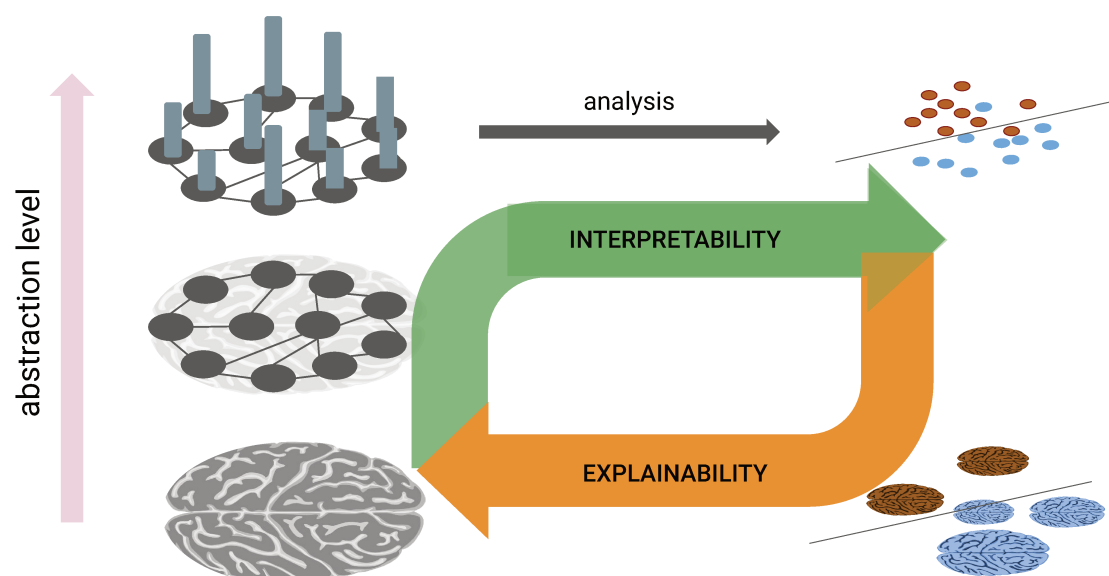


Figure 1.3 – Interpretability and explainability in FC studies.

should also encompass all levels of abstraction. We define these concepts as follows.

The FC method’s interpretability refers to a transparent and clear understanding of the entire analysis process: from the network model definition to the methodology used for the analysis itself. On the other hand, the FC method’s explainability corresponds to its effectiveness on illuminating the functioning of the brain across varying states or conditions, ultimately reconnecting with the brain level.

As illustrated in Figure 1.4, we categorize existing FC networks analysis frameworks into two main groups: classical statistical comparison and classification methods.

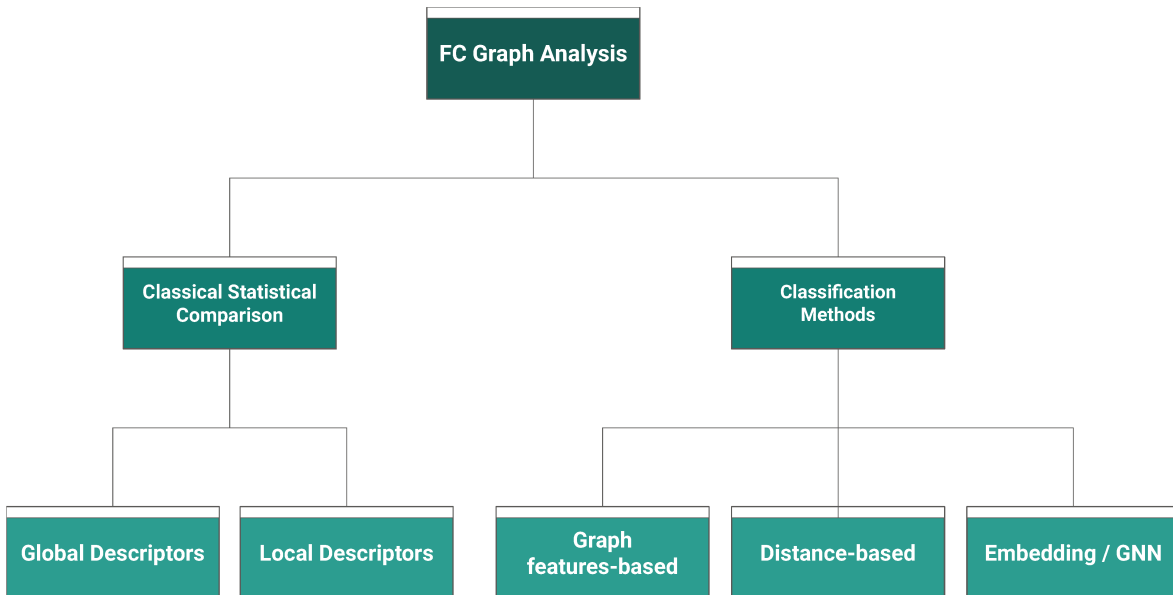


Figure 1.4 – Categorization of existing analysis methods for human brain FC graph-based models.

The former refers to a statistical comparison of established graph descriptors for distinguishing network groups (Bullmore et al. 2009; Wang et al. 2010). This statistical analysis can be performed for characterization of real data by comparing it with null models (Váša et al. 2022, Appendix B.3) or to compare different brain states, such as pathological subjects and healthy controls (Mheich et al. 2020). Graph descriptors can be extracted at different scales. At the global-level, a single graph descriptor is computed over the entire network, resulting in a single value per network, but multiple graph statistics can also be extracted per each node, and then compared in average or individually across groups. Formal definition of classical graph descriptors can be found in the next section.

On the other hand, classification methods refer to all approaches which have been developed to categorically classify subjects into different brain states. Classification methods have the advantage with respect to classical group comparison, to individually compare subjects. They are used for instance to predict if a network graph belongs to a pathological or healthy subject. Given the growing interest in artificial

intelligence, it is not surprising that such classification methods are based on machine learning techniques: systems which learn directly from data in order to take decision without specific instructions (Valliani et al. 2019; Du et al. 2018). These methods offer notable benefits, as they can uncover subtle patterns emerging directly from data that might not be apparent through traditional analyses. A commonly approach within classification methods employs machine learning algorithms trained on a set of extracted graph features (Richiardi et al. 2013; Craddock et al. 2009).

Other classification methods directly define a similarity score or a network pair-wise distance and apply it to cluster or classify different brain states (Wills et al. 2020). Finally, the development of network embedding and graph neural network (GNN) methods has also found captivating applications within network neuroscience. Numerous existing techniques have now embraced GNNs to classify different brain states (Bessadok et al. 2022a). Network embedding methods aim to represent nodes in a graph as low-dimensional vectors while preserving a specific proximity function. By transforming nodes into continuous vector representations, network embedding facilitates the application of existing machine learning algorithms on graph data. GNNs can also be interpreted as network embedding methods, since they provide a low-space representation of a graph into a latent space. With respect to classic network-embedding methods, the hidden representation is optimized for the specific classification task.

The diversity in FC analysis methods is accompanied by a multitude of studies that are grounded in varying FC estimation and acquisition parameters. To provide an overview of this diversity, we present in the following subsection the outcomes of a PubMed literature search.

1.2.1 PubMed Literature Search

A literature search has been conducted on PubMed on the 30th September 2021 using the query: *(resting state) graph fMRI connectivity (comparison OR prediction)*. The results were filtered from 2016 to 2021, two reviews paper were discarded, together with studies concerning small animals or task MRI, resulting in a total of 196 entries. Full results can be found in Tab. A.1 in Appendix A. Among the results 58% of the papers apply classical statistical methods, while 24% use a classification-based method (See Tab. 1.1)

To give an idea of the data variability in pre-processing and acquisition parameters, we extract magnetic field, repetition time, number of acquired volumes and scanning time, and eyes condition (open, closed, or cross-fixing). Concerning the network model after the pre-processing step, we select the considered frequency band and the chosen regions definition if given by pre-defined atlas. Finally, we collect the number of subjects, the analyzed group labels, and the data availability. These details were manually extracted from the text. To envisage an automatic extraction process, it is essential to establish a standardized format for presenting data, ensuring enhanced comparabil-

Table 1.1: Literature search has been conducted on PubMed on the 30th September 2021 using the query: *(resting state) graph fMRI connectivity (comparison OR prediction)*.

FC Graph Analysis	n (frequency)
Classical Statistical Comparison	114 (58%)
Classification Methods	47 (24%)
Both	22 (11%)
Other	13 (7%)

ity and facilitating searches in future works. In addition, in many cases, not all the needed information was included within the main text and must be sourced from supplementary materials or referenced publications or might not have been available at all.

Moreover, we included in our review the major used public fMRI dataset: UK BioBank (Sudlow et al. 2015), ABIDE (Di Martino et al. 2014), MJFOX (Loh et al. 2020), Human Connectome Project (Woolrich et al. 2001), iShare (Tsuchida et al. 2017), ADHD-200 (Bellec et al. 2017), and COBRE (Mayer et al. 2013; COBRE 2012). As

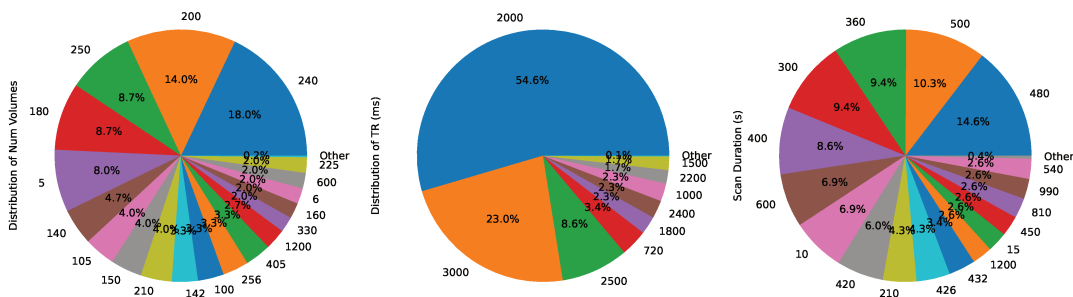


Figure 1.5 – Variability in the number of volumes, repetition time (TR), and scan acquisition time on the considered PubMed Literature Search.

we can observe in Fig. 1.5 and in Tab. A.1, there is a high parameter diversity both in data acquisition and pre-processing steps. Particularly, there exists a high variability in repetition time (TR), acquired number of volumes and scan duration, which, both for static and dynamic estimation of brain connectivity, will result in differences in the estimated networks. The range of acquired volumes spans from 100 to 1200, and this parameter introduces the most significant variability in the overall recorded results. Notably, the observed mode is 240, which corresponds to acquisition times of approximately 15 seconds and over 18 minutes for the extreme observed TR cases (minimum of 70 ms and maximum of 5000 ms). Given such substantial variations, it is reasonable to expect disparities in graph estimation and the conclusions drawn under these distinct parameters of acquisitions.

Besides, the eyes condition at the scan can not be discarded. Generally, there are three different possibilities: subjects are instructed to close their eyes, to keep their eyes open or to fix a colored cross in a screen. Studies have been conducted for assessing the difference in brain connectivity depending on the eye condition at scan ([Zou et al. 2015](#); [Patriat et al. 2013](#); [Barry et al. 2014](#); [Petro et al. 2022](#); [Yuan et al. 2014](#); [Agcaoglu et al. 2019](#)), demonstrating the need of a standard acquisition method to improve studies comparability and data pooling .

This diversity hampers study results comparison and multi-site functional connectivity data sharing. To cite a dramatic example, we report a review on Parkinson's Disease where different resting-state functional connectivity conditions results into contradictory conclusions. These results are reported in [Tab. 1.2](#) adapted from [Tessitore et al. 2019](#).

Table 1.2: Summary of the methods and results from the studies on Parkinson disease which adopt a graph-based analysis included in the review by Tessitore et al. 2019. PD: Parkinsonian patients HC: healthy controls.

Reference	Subjects	Main findings
Sang et al. 2015	26 early PD 30 HC	Decreased global efficiency in PD compared to HC. Increased nodal centrality in bilateral pallidum, inferior parietal lobule, and medial superior frontal gyrus, and decreased nodal centrality in caudate nucleus, supplementary motor areas, precentral gyrus, and middle frontal gyrus in PD compared to HC
Berman et al. 2016	19 PD 16 HC	Increased local efficiency in central executive network and salience network.
Fang et al. 2017	26 early PD 19 HC	Decreased nodal degree, global efficiency, local efficiency and characteristic path length within the Sensorimotor network (SMN) and visual network in PD compared to HC. Higher nodal degree, global efficiency and local efficiency, and lower characteristic path length within default mode network (DMN) and cerebellum in PD compared to HC. Lower cluster coefficient in thalamus and caudate nucleus in PD compared to HC
Suo et al. 2017	153 PD 81 HC	Decreased clustering coefficient, global efficiency, and local efficiency, and increased characteristic path length as well as decreased nodal centralities in the SMN, DMN, and temporal-occipital regions in PD compared to HC
De Schipper et al. 2018	107 PD 58 HC	Increased eigenvector centrality within frontoparietal regions in PD compared to HC. Increased connectivity in the SMN and VN in PD compared to HC
Hou et al. 2018	20 early akynetic PD 20 HC	Lower nodal centralities in the occipital lobe and areas of the limbic system and higher nodal centralities in frontal and temporal regions in PD compared to HC
Tuovinen et al. 2018	16 early PD 16 HC	At baseline, increased connectivity between cerebellum and SMN as well as decreased connectivity between motor regions and cingulate cortex in PD compared to HC. At 1.5 years follow-up, increased cerebellum connectivity within itself and to the caudate nucleus, thalamus and amygdala in PD compared to HC

In recent times, a consensus protocol for functional connectivity analysis has been established for the brain of rats (Grandjean et al. 2023), which provides a standardized and reliable methodology for analyzing functional brain networks in rats. Similarly, efforts are being made to reach a consensus protocol for human brain functional connectivity analysis, which would ensure the comparability and repeatability of results across different studies (Wang et al. 2023a; Botvinik-Nezer et al. 2020).

Concerning modeling steps for fMRI, we point out the work by Dadi et al. 2019 which proposes a comprehensive analysis of the impact of the main choices in terms of discrimination power. However, the work does not consider graph-based approaches for functional connectivity analysis.

Given the advantages of graph-based approaches in studying brain networks, future research needs to investigate the impact of different choices on graph representations of functional connectivity networks. This could involve exploring the influence of various graph construction methods, graph filtering procedures, and graph theoretical metrics on the discriminative power and interpretability of the resulting network representations. Such an analysis would contribute to a more comprehensive understanding of the methodological choices that impact the study of brain connectivity using graph-based approaches.

1.2.2 Recall Graph-Theory

In this section, we introduce the adopted graph notation, more details on graph-theory can be found in the works by Newman 2012; West et al. 2001; Van Steen 2010; Barabási 2013. A visualization of the main graph descriptors can be found in Fig. 1.6.

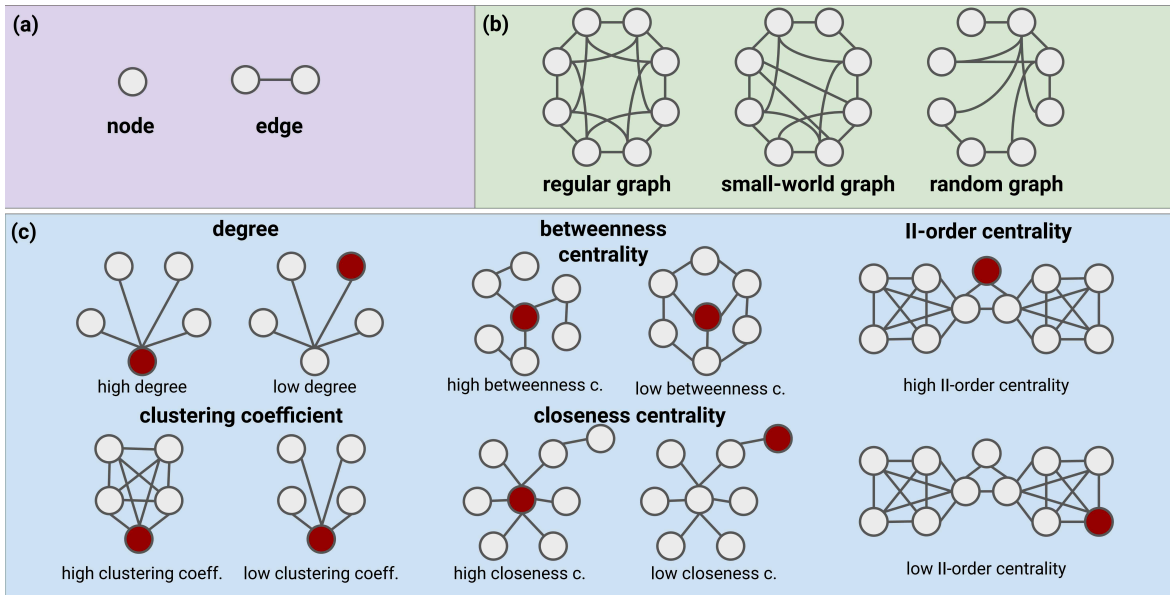


Figure 1.6 – Visualization of common graph descriptors classically considered in FC studies. (a) Standard definition of node and edges, (b) visualization of regular, small-world and random graph regime, (c) extreme example of the considered nodal-statistics. Adapted from [Jacunski et al. 2013](#)

Definition 1.1 (Graph). A graph \mathcal{G} is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Here, \mathcal{V} is a set of vertices or nodes. Each node represents a unit of the network we are considering.

$\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges among the nodes, encoding the presence of a special relationship among pairs of units.

Within the context of FC network modeling, each node represents a region of the brain.

Definition 1.2 (Weighted Graph). A graph \mathcal{G} is said to be weighted when it is equipped with a weight function $W : \mathcal{E} \rightarrow \mathbb{R}_0^+$ which associates to each edge a value. This value reflects the magnitude or strength of the encoded relationship.

In the following, we will deal with unweighted graphs: we can interpret unweighted graphs as special case of weighted graphs where each edge is assigned the same weight, which is equal to 1.

Definition 1.3 (Directed Graph). A graph \mathcal{G} is said to be directed if its edges $e \in \mathcal{E}$ are ordered pair of nodes (u, v) $u, v \in \mathcal{V}$, representing a connection from vertex u to vertex v .

Definition 1.4 (Undirected Graph). A graph \mathcal{G} is said to be undirected when each link represents a bi-directional relationship. $e \in \mathcal{E}$ is then a pair of nodes $e = \{u, v\}$ $u, v \in \mathcal{V}$ encoding the presence of a connection among the vertices u, v .

In this context of FC networks, the pairs of nodes are not ordered, thus, the consider graphs are undirected. In the following, the terms vertex, node, and unit are used interchangeably, as are the terms link and edge.

Definition 1.5 (Adjacency Matrix). The adjacency matrix A of an unweighted graph \mathcal{G} , is a binary matrix defined as follows:

$$A = (a_{uv})_{u, v \in \mathcal{V}} = \begin{cases} 1 & \text{if } \{u, v\} \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

For an undirected graph A is symmetric.

The adjacency matrix is one of the ways to represent a graph which directly encodes the presence of edges.

Definition 1.6 (Node Neighborhood). We define the nearest-neighborhood, first-order neighborhood, or simply the node neighborhood of node u as the set $\mathcal{N}(u)$, which includes all nodes connected to u .

We denote N the cardinality of \mathcal{V} .

Definition 1.7 (Complete Graph). A graph is complete if every pair of distinct vertices is connected by an edge. A complete graph of N nodes has exactly $\binom{N}{2} = \frac{N(N-1)}{2}$. Each node's neighborhood in a complete graph includes all the other nodes.

Definition 1.8 (Graph sparsity). We refer to graph sparsity as the ratio between the number of edges in the graph and the number of all possible edges among the set of nodes:

$$\text{Sparsity} = \frac{|\mathcal{E}|}{\binom{N}{2}}.$$

A graph whose Sparsity = 1 is the complete graph and vice versa.

Definition 1.9 (Path on a graph). In a graph, a path is an ordered sequence of vertices (v_1, v_2, \dots, v_k) where each consecutive pair of vertices is connected by an edge.

The length of a path in a graph is one less than the number of vertices it includes: $\text{Length}(v_1, v_2, \dots, v_k) = k - 1$. This corresponds to the number of edges a walker on node v_1 needs to cross to reach node v_k .

A path is said to be a cycle if it starts and ends at the same vertex, while traversing only distinct vertices in between.

Definition 1.10 (Minimum Path Length). Given a graph \mathcal{G} , the minimum path length between two vertices u and v is the smallest number of edges a walker must traverse to move from vertex u to vertex v .

A path from u to v of length the minimum path length between u and v is said a shortest path between u and v .

If no path exists between node u and v , we say node u is unreachable from v and conversely. In such case, we set their minimum path length to $+\infty$.

Definition 1.11 (Completely Connected Graph). A graph \mathcal{G} is said to be completely connected (or simply connected) if the minimum path length between every pair of distinct vertices is finite.

Definition 1.12 (Minimum Spanning Tree). A minimum spanning tree (MST) of a graph is a subset of its edges which identifies a sub-graph connecting all its vertices without any cycles and having the lowest possible total edge weight. A graph can admit many MSTs.

1.2.3 Classical Statistical Comparison

A classical statistical comparison analysis can concern global or local graph descriptors. As global-level statistic, we report the notions of global and local efficiencies which are commonly used to quantify the *small-worldness* of functional networks (Bassett et al. 2006b; Bassett et al. 2017a; Liao et al. 2017; Achard et al. 2007).

Global Descriptors

The concept of *small-world* network was originally introduced by Watts et al. 1998 and used to characterize networks which exhibit a good balance between short-distance connections among neighbors nodes and long-distance connections between non-proximal nodes.

Definition 1.13 (Global Efficiency). We define the global efficiency of a graph \mathcal{G} as:

$$E_{\text{global}}(\mathcal{G}) = \frac{1}{N(N-1)} \sum_{u \neq v \in \mathcal{V}} \frac{1}{L_{v,u}} \quad (1.1)$$

where $L_{u,v}$ is the minimum path-length between node u and v .

Definition 1.14 (Local Efficiency). The local efficiency of a graph \mathcal{G} is computed as the mean over all nodes of the node efficiency $E_{\text{nodal}}(u)$:

$$E_{\text{local}}(\mathcal{G}) = \frac{1}{N} \sum_{u \in \mathcal{V}} E_{\text{nodal}}(u) = \frac{1}{N} \sum_{u \in \mathcal{V}} \frac{1}{N_{\mathcal{N}(u)}(N_{\mathcal{N}(u)} - 1)} \sum_{j \neq k \in \mathcal{N}(u)} \frac{1}{L_{j,k}} \quad (1.2)$$

where $\mathcal{N}(u)$ is the set of nodes that are nearest-neighbors of the u and $N_{\mathcal{N}(u)} = |\mathcal{N}(u)|$.

Global efficiency in a network quantifies the network capacity for efficient information transfer across all nodes, potentially at high distance. The global efficiency reflects network information integration. On the other hand, local efficiency focuses on neighborhood communication and can be interpreted as a measure of network segregation, i.e. the presence of node groups highly connected among them with short-distance

paths (Sporns 2013; Deco et al. 2015). By considering global and local efficiency measures, networks can be categorized into three types: regular networks exhibit high local efficiency and low global efficiency, random networks show low local efficiency and high global efficiency, while small-world networks strike a balance between both efficiencies, positioning themselves between regular and random networks (Fig. 1.6 (b)).

Global and Local Efficiency in FC: small-world brain networks. Human brain FC networks studies - and more broadly various brain networks, even spanning different species- report a small-world topology (Watts et al. 1998; Achard et al. 2006; Bullmore et al. 2009; Wang et al. 2009; Bassett et al. 2010; Rubinov et al. 2015; Varoquaux et al. 2012). This organization is hypothesized to emerge from the evolutionary process to simultaneously optimize the cost of neuronal resources and the efficiency of information transmission (Bullmore et al. 2012; Samu et al. 2014). Notably, connections among anatomically adjacent brain regions are preferred to optimize the resource cost, while long-distance connections are required for faster information integration (Vértes et al. 2012; Chen et al. 2013b).

Finally, we refer to nodal statistics on a graph as any possible application on the set of nodes of the graph $s_{\mathcal{G}} : \mathcal{V} \rightarrow s(\mathcal{V})$, which is a function of the adjacency matrix. Given a nodal statistics, its graph average corresponds to a global graph descriptor and we denote it as

$$\bar{s}(\mathcal{G}) = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} s_{\mathcal{G}}(v) \quad (1.3)$$

Local Descriptors

In the following definitions (Def. 1.15-1.20), we report nodal statistics commonly used in functional connectivity studies (Hallquist et al. 2018; Richiardi et al. 2013; Mheich et al. 2020). These nodal statistics can be associated with small-world properties of human brain (Liao et al. 2017) or with its *modularity* structure (Bullmore et al. 2009). A supporting visualization can be found in panel (c) of Fig. 1.6.

Definition 1.15 (Degree). The degree of a node v represents the number of edges incident to the node

$$\text{deg}(v) = \sum_{u \in \mathcal{V}} a_{vu}.$$

Definition 1.16 (Degree distribution). The degree distribution $P(k)$ is the ratio between the number of nodes with degree equals to k and the total number of nodes N .

The degree and the degree distribution are among the fundamental graph descriptors that are commonly analyzed to understand the connectivity and structural characteristics of a network (Newman 2003; Newman et al. 2001; Broido et al. 2019; Albert et al.

2002).

Degree in FC: brain networks exhibit hubs. The nodal degree in FC networks is frequently employed to identify specific *hub* regions. These hubs are characterized by having a significantly high number of connections - or generally high centrality - in the network (Cole et al. 2010; Wang et al. 2010; Zuo et al. 2012). Numerous studies provide evidence for the presence of these hub regions in human brain networks, which serve to connect different parts of the brain (Power et al. 2013; Tomasi et al. 2011; Van den Heuvel et al. 2013). The presence of hubs is reflected in the heavy tail of the observed degree distribution of FC networks (Liao et al. 2017; Bassett et al. 2008; Achard et al. 2006; Bassett et al. 2017a; Heuvel et al. 2008; Eguiluz et al. 2005). A significant portion of studies on FC networks report degree-based hub detection results (Crossley et al. 2014; Guo et al. 2020; Hallquist et al. 2018; Saghayi et al. 2020b).

Definition 1.17 (Clustering coefficient). We define the clustering coefficient of a node v as a function of the number of triangles (i.e. group of three nodes) through the node itself.

$$Cc(v) = \frac{2T(v)}{\deg(v)(\deg(v) - 1)}.$$

The clustering coefficient can be considered an alternative measure of nodal efficiency, capturing network segregation by quantifying the presence of edges between pairs of node neighborhoods (Watts et al. 1998; Newman 2009).

Clustering coefficient in FC: brain networks modularity. As the clustering coefficient provides insights into how a node tends to form interconnected groups, its average serves as a quantification of the network's *modularity*. Modularity refers to the extent to which a network can be divided into node groups that have a high number of connections among themselves but few connections within each group. These groups are referred to as modules. The presence of modules in brain networks reflects the localization of information in densely clustered nodes. The brain's modular structure is counterbalanced by the presence of hubs, ensuring the maintenance of the small-world network property for efficient communication (Liao et al. 2017; Ferrarini et al. 2009; Sporns 2013). The brain's modularity structure has also been extensively studied (Laurienti et al. 2009; Ferrarini et al. 2009; Meunier et al. 2009; Chen et al. 2013a; Bhanot et al. 2023) in network neuroscience. Within these studies, the clustering coefficient and its average have been analyzed and applied to characterize different brain states and conditions (Eguiluz et al. 2005; Richards et al. 2018; Sala-Llonch et al. 2014; Bullmore et al. 2009; Saghayi et al. 2020b; Supekar et al. 2008; Zimmerman et al. 2018).

Definition 1.18 (Betweenness centrality). We define the betweenness centrality of a node given $\sigma(s, t)$, the number of shortest paths from node s to t , and $\sigma(s, t|v)$ the

number of shortest paths from s to t through v as

$$B(u) = \sum_{s,t \in \mathcal{V}} \frac{\sigma(s,t|u)}{\sigma(s,t)}.$$

Betweenness centrality is especially valuable when examining how information traverses a network model. Nodes with high betweenness centrality play a crucial role as bridges for information flow, and their identification can highlight significant nodes in the network (Freeman 1977).

Definition 1.19 (Closeness centrality). The closeness centrality measures the facility of connection of a node with respect to all the other nodes in the graph and it is defined as follows

$$Cs(v) = \frac{N-1}{\sum_{u \neq v} d(v,u)}$$

where $d(v,u)$ the shortest-path distance between v and u

Withing the notion of closeness centrality, a central node is a node which is close to all the other in the vertices set (Bavelas 1950).

Definition 1.20 (II-order centrality). The II-order or second-order centrality is defined as the standard deviation of return times of a simple random walk starting in node v

$$S(v) = \lim_{M \rightarrow \infty} \sqrt{\frac{1}{|M-1|} \sum_{k=1}^M \Xi_u(k)^2 - \left[\frac{1}{M-1} \sum_{k=1}^M \Xi_u(k) \right]^2}$$

with $\Xi_u(k)$ the k -th return time of the simple random walk starting in u to u .

The second-order centrality was introduced to reevaluate the importance of node bridges (Kermarrec et al. 2011), which facilitate information flow and connect distinct parts of a network, even if they are not strictly on the shortest paths (Fig. 1.6 (c)).

Centralities in FC networks: hubs detection. As previously mentioned, brain networks are characterized by the presence of hub regions. A hub can be defined as a vertex which occupies a central position among all other nodes in the network. This central position might simply correspond to regions exhibiting a high number of connections (Def. 1.15). However, the adopted definition of centrality impacts the identification of this central position. Different centrality measures may identify different hubs with varying properties (Sporns et al. 2007; Rubinov et al. 2010; Zuo et al. 2012). The selection of these statistics is fundamental for hub detection within FC networks, particularly when the aim is to compare changes in hubs across different brain states. This may require testing different nodal statistics separately to identify the one that most effectively distinguishes between diverse groups, based on the specific pathology or task under consideration (Achard et al. 2006; Oldham et al. 2019; Achard et al.

2012; Crossley et al. 2014; Power et al. 2013; Joyce et al. 2010; Drakesmith et al. 2015).

Finally, to conclude this section concerning statistical comparison method, we present the definition of hand-crafted index that provides motivations to our research. It's worth noting that various engineered indices have been presented in the literature. Notably, the Graphlet Degree Vector, which investigate the presence of sub-patterns within functional connectivity graphs to characterize healthy subjects (Finotelli et al. 2021b).

Here, we report the hub disruption index, introduced by Achard et al. 2012. This index allows to detect differences between groups in the identification of hubs and their re-organization (Fig.1.7). We reformulate the original definition by explicitly mentioning the dependence on s , the adopted centrality.

Definition 1.21 (s -nodal disruption). We notate a group of graphs having the same vertices set as $G = \{\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k) \text{ s.t. } \mathcal{V}_k = \mathcal{V}\}$ and, given a nodal statistics s , we consider the group mean nodal statistics

$$s_G(v) = \frac{1}{|G|} \sum_{\mathcal{G}_k \in G} s_{\mathcal{G}_k}(v).$$

We define the s -nodal disruption of the graph \mathcal{G} having same nodal set, with respect to the group of graphs G as

$$\kappa_s(v) = s(v) - s_G(v) \tag{1.4}$$

Definition 1.22 (Hub disruption index). Given the s -nodal disruption of a graph \mathcal{G} with respect to the group G , we define the hub disruption index of G , $\kappa_s(\mathcal{G})$, as the coefficient of the linear regression of $\kappa_s(v)$ as dependent variable and $s_G(v)$ as independent variable

$$\kappa_s(v) = \kappa_s(\mathcal{G})s_G(v) + \beta \tag{1.5}$$

with β the value of $\kappa_s(v)$ when the $s_G(v) = 0$. A toy example of the computation of $\kappa_s(\mathcal{G})$ can be found in Fig. 1.7.

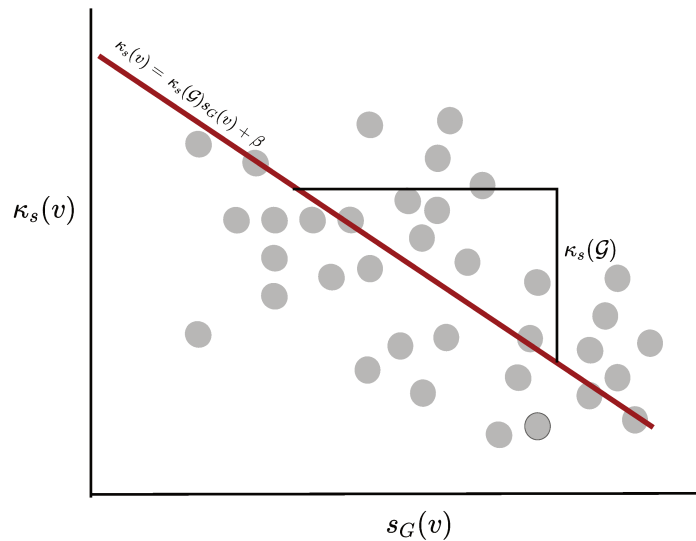


Figure 1.7 – Visualization of an example of computation of the hub disruption index of an individual graph \mathcal{G} associated to the nodal statistics s with respect to the graph collection \mathcal{G} . To compute the hub disruption index of the collection \mathcal{G} , with respect to the nodal statistics s , we compute $\kappa_s(v)$, i.e. the difference between the nodal statistics s of the graph \mathcal{G} and the mean nodal statistics of the group $s_G(v)$. Thus, this difference is plotted against the mean nodal statistics of the group $s_G(v)$. Finally, we determine the regression line $\kappa_s(v) = \kappa_s(\mathcal{G})s_G(v) + \beta$. Adapted from [Achard et al. 2012](#).

1.2.4 Classification Methods

We name graph-based classification methods all approaches which aim at predicting the brain state given the FC graph. These methods focus on optimizing the accuracy (i.e. the rate of correct predictions) of the classification and may overlook the interpretability aspect, essential in neuroscience and crucial for clinical applications.

While the development of a good classifier is fundamental for the true applicability of FC graph as biomarker of brain disease, research in this context lacks comparability across studies due to the variability of the pre-processing choices in graph inference and in dataset acquisition parameters (i.e. close or open eyes, scan length, etc.).

We propose to distinguish graph-based classification methods in three main groups (Fig. 1.4). The first group employs a classifier which takes as input graph-features.

A common machine learning application to network neuroscience ([Richiardi et al. 2013](#); [Bassett et al. 2012](#); [Casanova et al. 2012](#); [Cheng et al. 2015](#)) employs the previously defined graph-descriptors as input features of classifier algorithms. A schematic visu-

alization of this procedure can be found in Fig. 1.8.

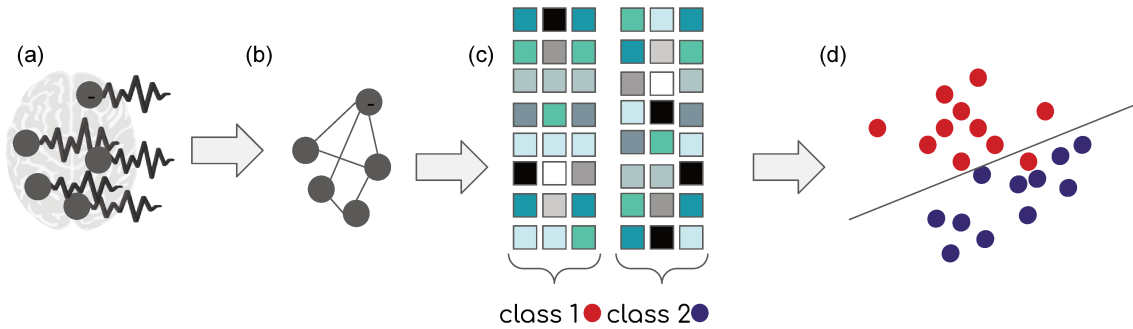


Figure 1.8 – Visualization of a classification procedure, adapted from [Richiardi et al. 2013](#). (a) A signal is collected over brain regions (b) the corresponding graph model is extracted (c) for each graph some features are extracted, graph belonging to different classes are expected to have different features (d) the features are used to distinguish between data points corresponding to the different classes.

While this kind of classifier is favored for its interpretability in relation to graph features, it is important to acknowledge that, when compared with counterparts that do not operate directly on the graph, they exhibit lower performance capabilities ([Lei et al. 2020](#)). Moreover, each study, according to the considered pathology or even the considered dataset, develop different classifiers based on different features ([Du et al. 2018](#); [Zanin et al. 2016](#)). For instance FC has been used for the design of automatic classification tools in various tasks concerning autism ([Graña et al. 2021](#); [Yang et al. 2022](#); [Iidaka 2015](#); [Bilgen et al. 2020](#)), schizophrenia ([Filippis et al. 2019](#); [Lei et al. 2020](#); [Cheng et al. 2015](#)), Parkinson disease ([Rubbert et al. 2019](#)), Alzheimer’s Disease, epilepsy ([Gholipour et al. 2022](#)), and other brain states ([Dai et al. 2022](#); [Shen et al. 2022](#); [Vergun et al. 2013](#); [Ball et al. 2016](#); [Lord et al. 2012](#); [Krämer et al. 2023](#); [Renard et al. 2021](#)).

Among the variety of dataset-specific classifier, a particular attention have received the graph Laplacian.

Definition 1.23 (Laplacian matrix). We recall the definition of the Laplacian matrix L :

$$L = D - A \quad (1.6)$$

where D is the degree matrix, a diagonal matrix where $D_u u$ represents the degree of node u , and A is the adjacency matrix.

In particular, different approaches determine the eigenvalues of the Laplacian matrix as input of a classifier. Such classification methods have proven to detect network patient alterations in Alzheimer disease or autism spectrum disorder ([Haan et al. 2012](#); [Mostafa et al. 2019](#); [Schirmer et al. 2021](#); [Mheich et al. 2020](#)).

The second category of classification methods under consideration relies on the use of a graph distance or a similarity score (Mheich et al. 2020; Wills et al. 2020). The use of graph distances and similarity scores determine an underlying metric space on graphs, which allows to compute the individual distance between different subjects. This permits the identification of clusters and the delineation of boundaries between these groups. Despite the existence of numerous graph metrics, including those that consider both structural and feature-based aspects simultaneously (Vayer et al. 2020; Thual et al. 2022), there is currently no conclusive evidence regarding an optimal distance metric to be employed in FC networks. Furthermore, there is a lack of graph-metrics specifically defined to handle brain graph.

Finally, the third group includes reduction tools such as network embedding or GNN methods (Cui et al. 2018; Hamilton 2020; Schieber et al. 2017; Zhang et al. 2022c). A network embedding is a dimensional reduction tool which maps a graph into a vector space by preserving specific graph features. Recently, the nodal embedding method `node2vec` (Grover et al. 2016), which associates to each node a vector, was applied to brain network characterization in Rosenthal et al. 2018. Embedding methods can be applied to define new network similarity (Nikolentzos et al. 2017) and being used for pathological discrimination (Carboni et al. 2021a; Lostar et al. 2020; Morris et al. 2017) or brain network evolution prediction (Göktaş et al. 2020).

GNNs (Cheung et al. 2020; Hamilton 2020; Bronstein et al. 2017), also find applications in network neuroscience (Bessadok et al. 2022a; Ménoret et al. 2017; Mhiri et al. 2020) as tool to process graph-structured data. The basic operation in GNN consists in aggregating feature vectors of the node neighborhood to update the node representation across different layers (See Fig. 1.9).

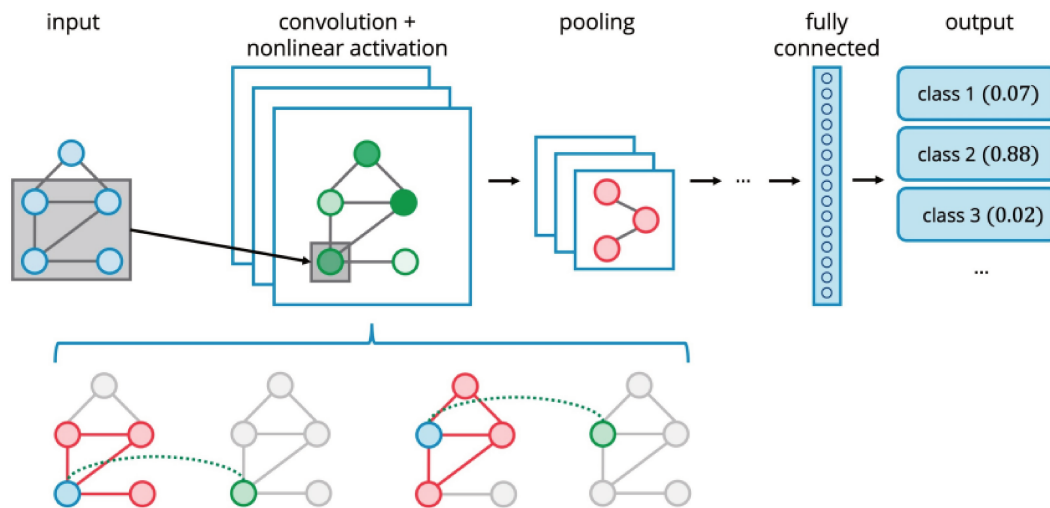


Figure 1.9 – Example of a GNN model for graph classification applied on an input. The model includes a graph convolutional layer, a pooling layer and a fully connected layer. A graph with node features is the input of the neural network model. As first layer for each vertex (shown in blue) the neighborhood information (shown in red) is first aggregated and then uses as input of a nonlinear activation function converting the aggregated information into a new node features (shown in green). This operation is repeated on all node of the graph. The pooling layer operates reducing the number node per graph. Finally, a fully connected layer outputs the probability of the graph to belong to three existing classes. Figure adapted from [Cheung et al. 2020](#).

This process is called message-passing and it is iterated to optimize a specific loss function. The graph-level representation can then be used for various tasks, such as node or graph classification, regression, or clustering. A GNN model can directly include the classifier as final layer, or instead learn a lower-space representation interpreted as network embedding such as in Auto Encoder architecture ([Banka et al. 2020](#)). For more details in the training of an artificial neural network, we refer to the Chapter 4. An important assumption of GNN model is permutation invariance, meaning that the predicted output does not depend on the chosen adjacency matrix ordering. However, in brain graph where a node represents specific brain region, the permutation invariance property may not be desirable and GNN models may need to incorporate node labels information, for instance through attention ([Veličković et al. 2017](#); [Vaswani et al. 2017](#)) or by learning adaptive node-wise aggregation scales ([Choi et al. 2022](#)).

While GNNs have shown promising results in FC ([Bessadok et al. 2022a](#); [Wang et al. 2022](#); [Li et al. 2021](#)), they have some drawbacks, including high computational costs for training and the need for large datasets. Moreover, their generalization to different datasets, especially from different centers or pathologies, can be challenging and require image harmonization ([Bottani et al. 2022](#); [Cackowski et al. 2023](#)). Additionally, the lack of explainability in GNNs is an area that needs further evaluation ([Kim et al. 2020a](#); [Agarwal et al. 2023](#); [El Ouahidi et al. 2022](#)).

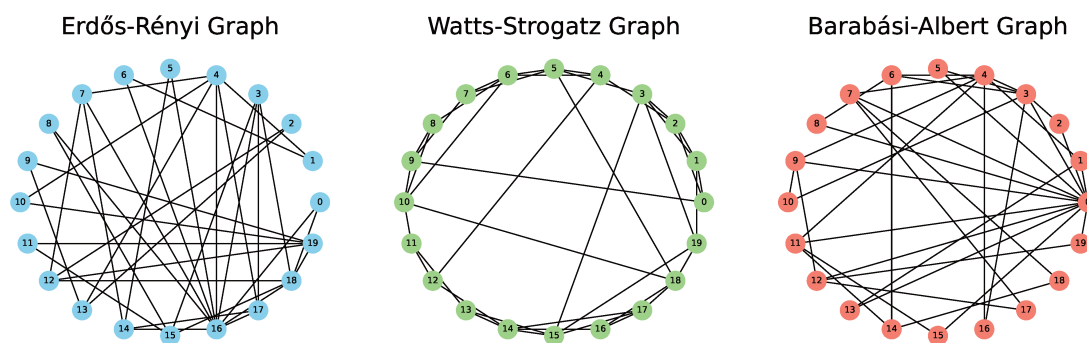


Figure 1.10 – Visualization of examples of ER, WS, BA graph.

1.3 Generative Networks

An important step for the understanding of brain functional connectivity can be achieved through the comparison of real data with synthetic ones. Here, we present the graph synthetic models we will further analyse in comparison with true data. A big branch of study in functional connectivity aims at finding explanation of brain functioning by comparing it to null models or determining its possible evolution justification (Bassett et al. 2018; Dichio et al. 2023b; Dichio et al. 2023a; Morgan et al. 2018; Vértes et al. 2012). In our proposal, we will consider generative networks to test whether the captured information is meaningful of the brain or derived by choices in network inferences, particularly when comparing real data with models which reproduce brain characteristics such as the observed degree distribution, its efficiency, or clustering coefficient.

1.3.1 Erdős-Rényi model (ER)

The Erdős-Rényi (ER) model generates a binomial graph $\mathcal{G}_{n,p}$ by the creation of edges among n nodes (Erdős et al. 1959). Each edge has a probability p of being created. The expected number of edges in $\mathcal{G}_{n,p}$ is then $p\binom{n}{2}$ and its sparsity ration equals p . For values of p close to 1, the graph tends to be the complete graph in which all possible edges are present.

1.3.2 Watts-Strogatz model (WS)

The Watts-Strogatz (WS) model generates a small-world graph $\mathcal{G}_{n,k,p}$ by connecting each node with its k neighbors nodes and then recombining each edge with probability p (Watts et al. 1998). In this case, the number of created edges is always $\frac{nk}{2}$, requiring an even value for k , which corresponds to a sparsity value of $\frac{nk}{2\binom{n}{2}} = \frac{k}{n-1}$. The p parameter, which regulates the probability of rewiring the edges, generates the regular graph ($p = 0$) in which all nodes have the same degree, and the completely random graph ($p = 1$) in which the expected number of edges are randomly distributed on the vertices set. We consider cases $p = 0.1, 0.5, 0.9$ and refer to the case $p = 0.5$ as the small-world model.

1.3.3 Barabási-Albert model

The Barabási-Albert model generates a graph $\mathcal{G}_{n,m}$ by favoring specific attachments (Barabási et al. 1999). It starts from a star graph of $m + 1$ nodes and attaches the $n - m - 1$ remaining nodes to the m existing nodes with high degree. In that case, the number of edges expected is given by the sum of the first m edges of the initial graph with $(n - m - 1)m$ edges created by attaching new nodes until the graph has n vertices leading to the sparsity value equals to $\frac{m(n-m)}{\binom{n}{2}}$. In this case, having fixed n and the level of sparsity l , there are two possible choices for m , corresponding to the solutions of

$$m^2 - mn + l \binom{n}{2} = 0.$$

The existence of real solutions to the previous equations is only guaranteed for $l \leq \frac{n^2}{4\binom{n}{2}}$ and in that case, it always has two positive solutions. We considered both cases, referring to BA1 and BA2, respectively for the lower and the highest root. Due to the constraints of existence of real solutions, all networks generated according to a Barabási-Albert model are sparse (Del Genio et al. 2011).

1.3.4 Degree sequence preserving model (DSP)

The degree sequence preserving (DSP) model is based on the configuration model (Barabási et al. 1999). In this case, we will build degree sequence preserving copy of graph coming from real dataset. The construction is also constrained by a given sparsity ration, thus, given the correlation matrix associated with a real graph and given a sparsity ratio, we threshold the correlation matrix to obtain a binary version with the number of edges corresponding to the fixed sparsity. Then, we extract the degree sequence and randomly generate a new graph that preserves the given degree sequence. Since the degree of each node is fixed, we obtain a synthetic graph which has the same sparsity as its real version. In such a way, for all sparsity values we considered, we obtain the synthetic graphs whose elements are the *model version* of the corresponding real graphs. An example of the simulated DSP networks is shown in Fig. 1.11.

1.3.5 Economical preferential attachment model (EPA)

The economical preferential attachment (EPA) model has been defined to reproduce functional brain networks (Vértes et al. 2012). The probability of observing a connection between two regions, u and region v is given by

$$p_{u,b} \propto (\deg(u) \deg(v))^\gamma (d_{u,v})^{-\eta}$$

where $\deg(u)$ is the degree of node u and $d_{u,v}$ is the Euclidean distance in anatomical space between u and v . Since we want to generate network at fixed sparsity, given a real network, we first extract its degree distribution. Next, we compute the $p_{u,v}$ of all possible pairs of nodes and then we select the edges with highest probability until we

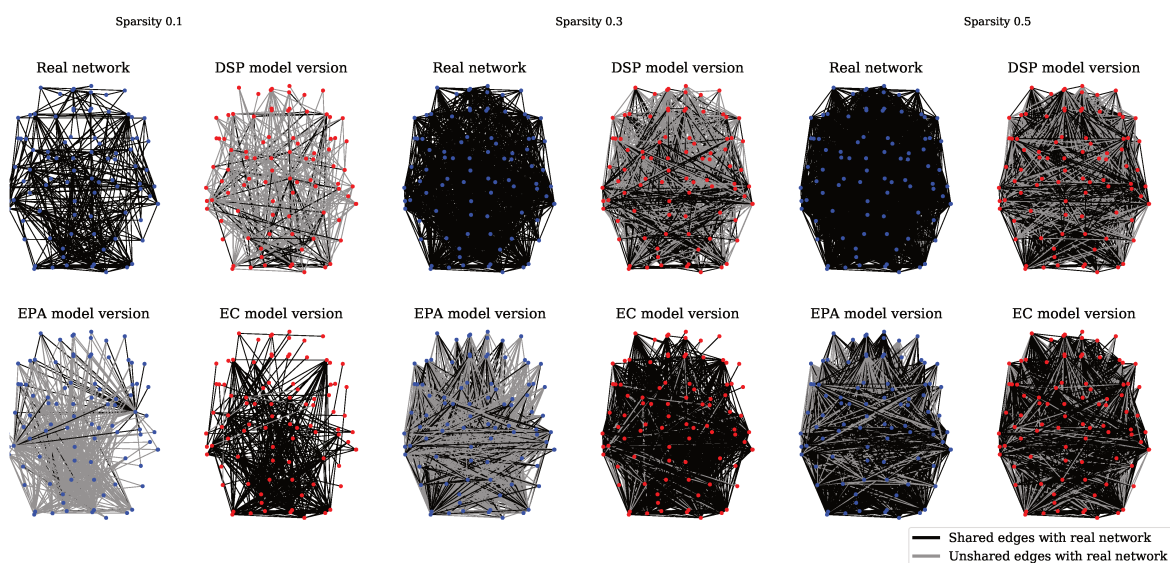


Figure 1.11 – Examples of real functional connectivity network of healthy subjects and their corresponding model versions at different sparsity values. DSP: Degree sequence preserving model; EPA: Economical preferential attachment model; EC: Economical clustering model.

reach an expected number of edges. To ensure connectivity, we also add the Minimum Spanning Tree as it is done in real data. γ, η are hyperparameters of the model and are fixed to better match the real data as in the work by [Vértes et al. 2012](#). An example of the simulated EPA networks is shown in Fig. 1.11.

1.3.6 Economical clustering model (EC)

The economical clustering (EC) model has also been proposed in the context of functional brain networks ([Vértes et al. 2012](#)). The probability of observing a connection between region u and region v is given by

$$p_{u,v} \propto (k_{u,v})^\gamma (d_{u,v})^{-\eta}$$

where $k_{u,v}$ is the number of nearest neighbors in common between nodes u and v , while $d_{u,v}$ is the Euclidean distance in anatomical space between u and v . For being able of tuning the sparsity of the model, we generate an EC model version of real network. Given a real network at a given sparsity ratio, we determine its $k_{u,v}$ and compute the $p_{u,v}$ of all possible node pairs. Finally, we select edges whose probability is higher until the expected number of edges is reached. Again, we ensure connectivity by adding missing edges from the Minimum Spanning Tree algorithm. The hyperparameters γ , and η are fixed according to the work by [Vértes et al. 2012](#). An example of the simulated EC networks is shown in Fig. 1.11.

1.4 Summary and conclusion

In summary, the analysis of functional connectivity studies from a graph perspective shows a high diversity across studies, leading to various questions and controversies in

the research field ([Hallquist et al. 2018](#); [Cwiek et al. 2022](#)). We have identified several requirements for the development of a general analysis framework for FC networks:

- the framework should take into account existing and interpretable classical network statistics;
- it should be adaptable to different type of pathologies or datasets, allowing for the tuning or selection of appropriate graph statistics to suit specific applications;
- the framework should enable group characterization, facilitating the comparison of network properties among different populations;
- it should also be capable of tracking subject individual differences, acknowledging the unique brain connectivity patterns exhibited by individuals;
- the framework should allow for local characterization, enabling the detection of local perturbations or anomalies within brain networks.

While we recognize the importance of developing classification methods for automatic diagnosis, our primary focus is not solely on competing for classification accuracy with existing methods. Instead, we aim to design an approach that is both adaptable and interpretable at group and individual level and at global and local network scale. This approach, presented in the next chapter, aims to bridge the gap between classical graph descriptors and advanced network classification techniques, ultimately providing an original characterization of healthy and pathological brain networks.

Chapter 2

Structural Pattern

Abstract

This chapter introduces a novel mathematical framework for functional connectivity network analysis. It begins by providing an overview of two prior studies that investigate nodal reorganization within functional connectivity networks. These explorations provide the motivation for the formulation of a graph structural pattern -the core definition of our work. Our newly introduced structural pattern finds also find inspirations in the concept of nodal automorphically equivalence relations prevalent in complex networks.

At the basis of our proposal is the introduction of a new nodal-statistics-based equivalence relation which allows to fill the gap between classical nodal statistics and network dimensional reduction tools.

This novel approach not only permits the combination of a multitude of nodal statistics for network analysis, but also enables the characterization of networks at both individual and group levels, global and local scales.

Contents

2.1	Motivations	41
2.2	Nodal-statistics-based equivalence relation for graph collections	46
2.2.1	Structural equivalence for undirected unweighted graph	47
2.2.2	Structural equivalence for graph collections	51
	Structural pattern comparison	51
	Nodes distinctiveness or similarity	54
2.2.3	General properties	54
	Properties of PC	54
	Relation with network ensembles entropy	55
	Properties of Orthogonality	56
	Properties of Correspondence of structural pattern score	57
2.3	Summary and conclusion	57

2.1 Motivations

Based on a paper written in collaboration with Michel Dojat and Sophie Achard, published in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). (Carboni et al. 2021a)

In several application scenarios which focus on complex network studies, being able to determine node roles has proven to be relevant (De Arruda et al. 2014; Weng et al. 2007; Borgatti et al. 2009; Finotelli et al. 2021a; Iacopini et al. 2020). Historically, the notion of node roles has been introduced in social science structural theory (Borgatti et al. 1992b) with at least two different conceptions: structural equivalence and structural isomorphism. According to the former, nodes are equivalent if they share exactly the same neighbors. For the latter, nodes are equivalent if there exists an automorphism which maps the first node to the second and *vice versa*. The nodal structural equivalence is related to the way nodes are connected with the other nodes in the graph and allows to take into accounts the existence of patterns and sub-networks.

In the context of social networks, the structural equivalence determines the role an agent recovers in a network, identifying peculiar node corresponding to influencers or grouping together agents with similar roles.

Recently, node roles analysis has been applied to various application domains such as web graphs (Meghabghab 2002), technological or biological networks (Rossi et al. 2014). Different algorithms have been proposed to detect structural equivalence classes in a single network by evaluating similarity metrics among nodes (Yu et al. 2021; Jeh et al. 2002; Chen et al. 2020).

In the case of brain connectivity networks, previous works characterized neurological deviations by looking for hub nodes in a collection of graphs (Achard et al. 2012). This research conducted by Achard et al. 2012 demonstrates a reconfiguration of these hub nodes among comatose patients compared to a control group. As shown in Fig. 2.1, high-connected nodes in healthy controls exhibit a decrease in the number of connections in comatose patients, while low-connected nodes in healthy controls show an increase number of connections in comatose patients.

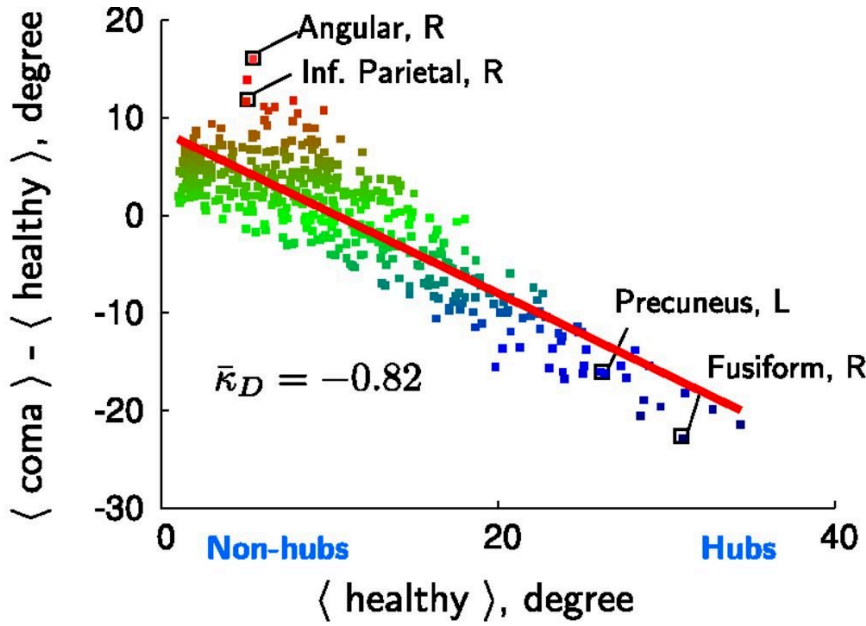


Figure 2.1 – Hub disruption index with respect to the degree of functional networks in comatose patients. The x-axis shows the mean degree of nodes in healthy volunteers, while the y-axis displays the difference in mean degree of nodes between the two groups. Abnormal hub nodes exhibit reduced degrees in comatose patients (e.g., precuneus) with respect to their high degree in the healthy group. Conversely, non-hub nodes display increased degrees in patients (e.g., angular gyrus), compared to their low healthy group degrees. The hub disruption index is quantified by the slope of the fitted (red) line, and node color indicates inter-group degree differences. Adapted from [Achard et al. 2012](#) with permission.

From this outcome, it becomes evident that the characterization of nodal roles within functional connectivity networks holds the promise of elucidating the neurological implications underlying brain connections. Furthermore, it highlights the necessity for extending the structural equivalence concept to encompass a collection of graphs.

From a mathematical point of view, the organization of nodes can be captured through the concept of graph automorphism, which is defined as follows:

Definition 2.1. An automorphism of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a bijection π between \mathcal{G} and itself such that:

$$\forall u, v \in \mathcal{V}, \{u, v\} \in \mathcal{E} \iff \{\pi(u), \pi(v)\} \in \mathcal{E}$$

If two vertices are connected in \mathcal{G} their images through the map π must be connected too. An obvious automorphism is the identity map, however not-obvious automorphism could exist.

Definition 2.2. Two nodes $v, w \in \mathcal{V}$ are structurally isomorphic or automorphically

equivalent $v \equiv w$ if there exists an automorphism $\pi : \mathcal{G} \rightarrow \mathcal{G}$ such that $\pi(v) = w$

An example of automorphically equivalent nodes is shown in Figure 2.2

Proposition 2.1. If two nodes are structurally isomorphic, so are their adjacent nodes.

$$v \equiv w \implies \forall u \in \mathcal{V} \text{ s.t. } \{v, u\} \in \mathcal{E} \exists u' \in \mathcal{V} \text{ s.t. } \{w, u'\} \in \mathcal{E}$$

for which it values:

$$u \equiv u'$$

Proof. By hypothesis we have $v \equiv w$, so it exists an automorphism π such that $w = \pi(v)$. By definition of automorphism:

$$\{v, u\} \in \mathcal{E} \iff \{\pi(v), \pi(u)\} = \{w, \pi(u)\} \in \mathcal{E}$$

Let u' equals to $\pi(u)$ then we have $\{w, u'\} \in \mathcal{E}$ for which it values $u \equiv u'$ via π . \square

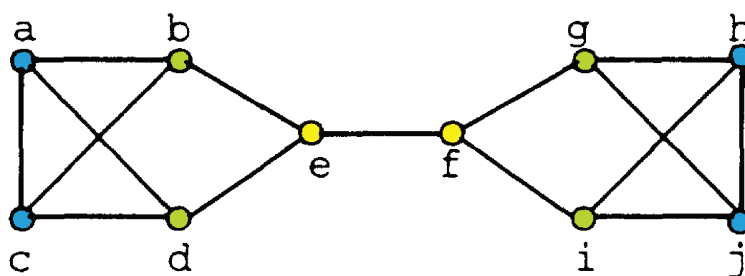


Figure 2.2 – Structural isomorphic nodes are shown in the same color. $a \equiv c \equiv h \equiv j$, $b \equiv d \equiv g \equiv i$, $e \equiv f$. Figure adapted from Borgatti et al. 1992a.

The detection of graph automorphisms goes under the graph isomorphism problem (Grohe et al. 2020; Babai 2016). Substantial advancement has been made in defining algorithmic tests for establishing graph isomorphism, with specific emphasis on the Weisfeiler-Lehman algorithmic tests for two graph to be isomorphic (Weisfeiler et al. 1968; Fürer 2017). However, its worst-case complexity can still be exponential in the number of nodes. The employment of a brute force approach to find nontrivial automorphisms necessitates assessing all possible permutations of the adjacency matrix. This process becomes easily intractable in real complex network applications, specifically when dealing with network collections.

In our pursuit of identifying node roles in FC graphs, we opted in our work Carboni et al. 2021a for `node2vec` embedding algorithm (Grover et al. 2016). This algorithm has demonstrated to effectively capture the structural equivalence of nodes. The `node2vec` algorithm learns a low-dimension nodal representation by sampling a particular random walk on the graph, thus determines a map such that *similar nodes* have *similar representation*. The `node2vec` approach reproduces the feature learning aspect of `word2vec`,

a technique in natural language processing. In `word2vec`, similar features vector are learned for words within a sentence possessing similar **semantic**, or **syntactic** characteristics (Mikolov et al. 2013, Tab. 2.1).

Table 2.1: Semantic and Syntactic definition.

	Definition
Semantic	refers to the inherent or inherent sense of a word, phrase, or symbol that relates to its real-world concepts, ideas, or objects.
Syntactic	pertains to the grammatical structure and arrangement of words within a sentence. It concerns the relationships between words based on how they are combined in a sentence.

When analysing a sentence, the position of a word can provide valuable insights. For instance, a word positioned near a verb might serve as an adverb, while a word preceding the verb could be its subject. Additionally, words that frequently co-occur in sentences tend to possess similar semantic meanings, whereas those that seldom co-occur might lack semantic connections.

The distinction between syntactic similarity and semantic similarity is translated into structural node similarity and homophily node similarity (Fig. 2.3). In the `node2vec`

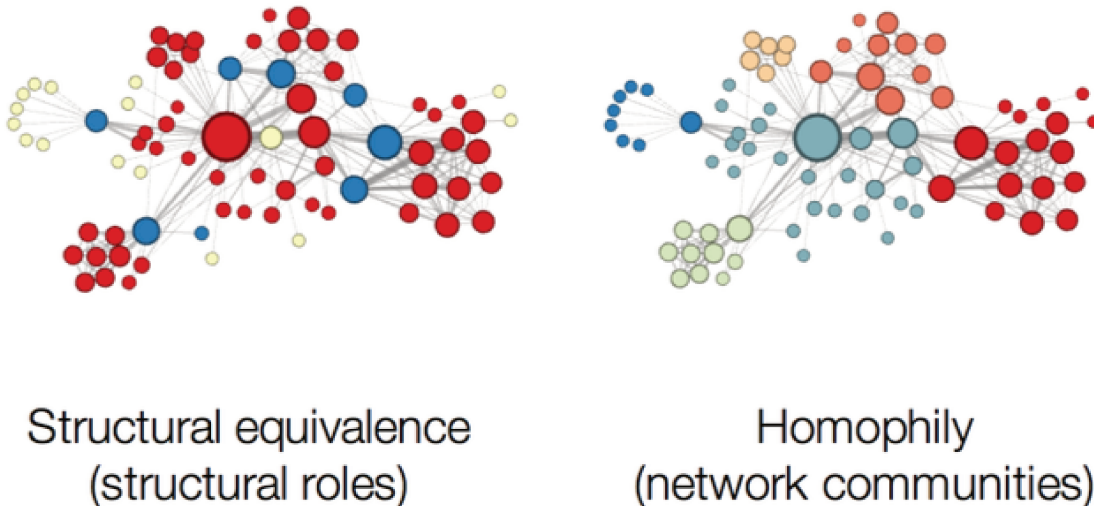


Figure 2.3 – Homophily and Structural Equivalence on *Les Misérables* network. Results of the nodal clustering on `node2vec` embedding space obtained with the DFS (left) and BFS (right) within *Les Misérables* network. Adapted from Grover et al. 2016.

algorithm, these two concepts are captured using distinct sampling strategies within the graph, as part of the `node2vec` random walk algorithm. Precisely, for each node, the algorithm computes a neighborhood set of a fixed number of similar nodes. The

breadth-first sampling (BFS) considers the neighborhood of a node as nodes which are immediate neighbors of the source. Whereas, the depth-first sampling DFS sequentially select nodes at increasing distances from the node itself. `node2vec` allows to smoothly interpolate between BFS and DFS by tuning its hyperparameters. This flexibility enables the algorithm to explore different structural aspects of the graph and capture varying degrees of syntactic and semantic similarities.

A visualization of the two differences sampling strategies is shown in Fig. 2.4. The resulting clustering on the embedding of *Les Misérables* network¹ (Min et al. 2016) in Fig. 2.3 shows the corresponding graph notions of semantic and syntactic: nodes high connected one to each other are detected by a pure BFS strategy, while nodes which occupy specific positions in relation with the other nodes are grouped together when DFS is used.

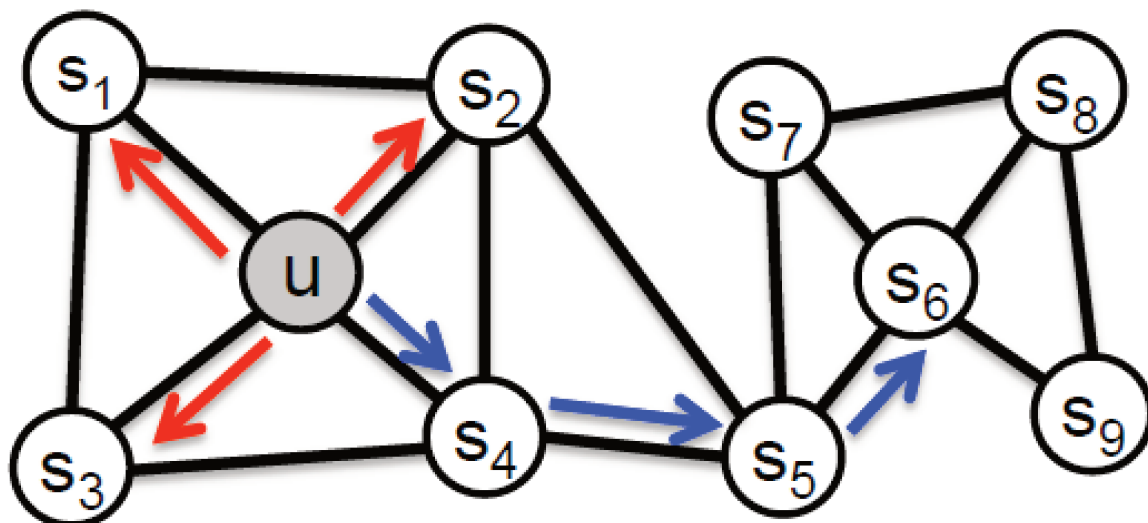


Figure 2.4 – `node2vec` sampling strategies. Generation of a neighborhood set of size 3 for node u , according to the two different sampling strategies. In red the BFS-neighborhood (s_1, s_2, s_3). In blue the DFS-neighborhood (s_4, s_5, s_6). Figure from Grover et al. 2016.

In our prior research (Carboni et al. 2021a; Carboni et al. 2021b), we employ the capabilities of `node2vec` in capturing nodal structural equivalence for the design of classification methods which differentiate comatose patients from healthy controls. This demonstrates the significance of considering nodal structural role for discerning different states of consciousness. However, the use of `node2vec`, as well as broader network embedding tools, comes with certain drawbacks. These include the necessity of meticulous hyperparameters tuning, dependence on massive dataset for training, and potential difficulties in providing comprehensive explanations and interpretations. Moreover, the

¹The network based on *Les Misérables* is constructed by representing characters as nodes and connecting together characters who appear together in the same scene.

inherent stochasticity of `node2vec` could impact the performances and comparability of the outcomes. To mitigate these challenges, we propose an alternative strategy, applicable when dealing with graphs of lower dimensions. This strategy involves direct operation in the graph space itself, favoring an integration between network science and neuroscience practitioners.

This new framework of analysis is inspired and motivated by the previous mentioned results. It does not require to learn low-space representation, but extract the graph structural pattern by means of classical nodal statistics.

2.2 Nodal-statistics-based equivalence relation for graph collections

Based on a paper written in collaboration with Sophie Achard and Michel Dojat and published in *Physical Review E*. Carboni et al. 2023b

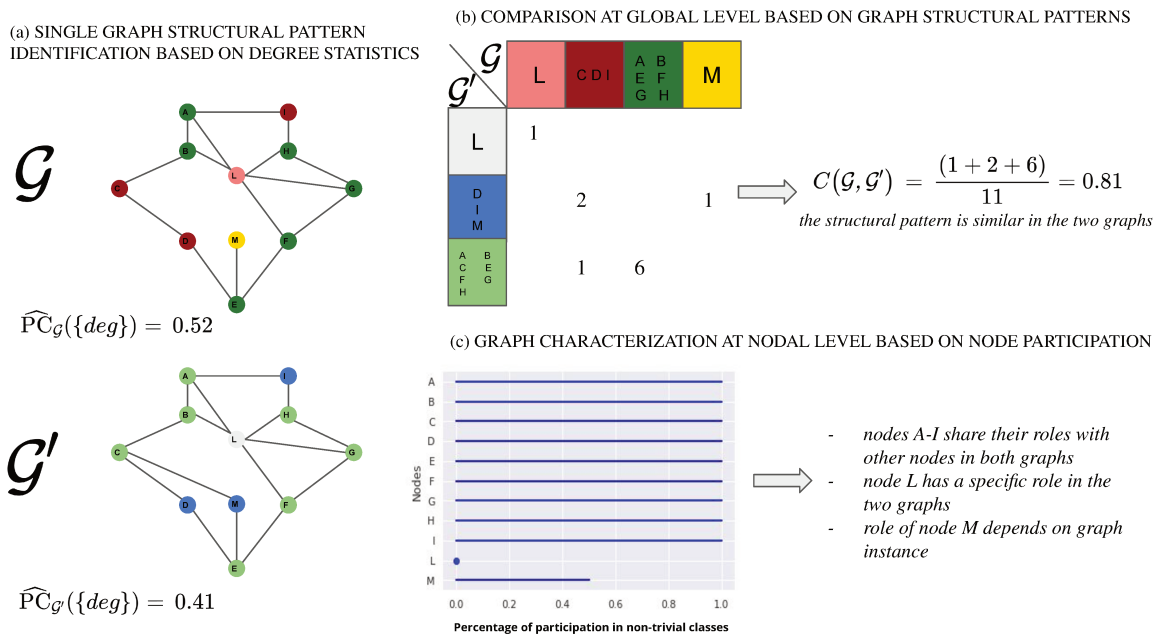


Figure 2.5 – Graphical visualization of the proposed nodal-statistics-based equivalence relation in graph collections. Global comparison and nodal characterization of graphs: (a) structural patterns associated with the same statistics are determined on the graphs, (b) the structural patterns are matched to compute a similarity value, (c) nodal participation in nontrivial classes is obtained for nodal characterization.

We examine a network collection defined on the same node-set, node role detection can provide meaningful information for collection characterization, possibly revealing a specific nodal partitioning. Indeed, in many real-world applications, the available graph set can potentially be characterized by specific node role classes (Kersting et al.

2016; Borgwardt et al. 2005; Cardillo et al. 2013; De Domenico et al. 2015; Hu et al. 2020). However, while many graph comparison metrics already exist (Wills et al. 2020; Schuld et al. 2020), there is no evidence of a method for comparing them, moreover, none of them directly address the detection of differences at the nodal level or it has been showed to be adapted for functional connectivity characterization.

The contributions of this work are fourfold. First, we define a structural equivalence relation on a graph node-set based on nodal statistics (any functions on the node-set). The proposed definition allows determining node role classes according to statistics values. The main innovation of this definition is given by the possibility of identifying the graph structural pattern based on an original combination of as many statistics as desired.

Second, we define two global measures of a statistics set which determine parsimony and heterogeneity of its elements. These measures only depend on the graph structure and can be used for statistics selection or graph complexity evaluation (Bianconi 2007). Third, we propose to compare graphs with the same vertices according to their structural patterns similarity. Indeed, thanks to the identification of node classes, we can compare different graph instances throughout the evaluation of the similarity of their structural patterns.

Finally, we propose a framework to determine node categories in a network group which allows to characterize the group at a nodal level and to discriminate nodes according to their role.

A visualization of our proposal is depicted in Fig. 2.5.

2.2.1 Structural equivalence for undirected unweighted graph

We propose to identify the graph structural pattern with the equivalence classes of a newly defined equivalence relation.

Definition 2.3. We propose to define an equivalence relation \sim_s , associated with a statistics s , on the nodes set \mathcal{V} of a graph as follows:

$$v \sim_s u \iff s(u) = s(v). \quad (2.1)$$

Definition 2.4. For a nodal statistics having as $s(\mathcal{V})$ a dense and continuous subset of \mathbb{R} , the equivalence is defined up to a fixed positive small ϵ : $v \sim_s u \iff |s(u) - s(v)| \leq \epsilon$ (Fig. 2.6).

As \sim_s is an equivalence relation on \mathcal{V} , it is possible to find its induced partition P on \mathcal{V} ,

Definition 2.5.

$$P_s = \frac{\mathcal{V}}{\sim_s} = \{[a]_{l, \sim_s} \quad \forall l \in s(\mathcal{V})\}, \quad (2.2)$$

which defines the structural pattern of \mathcal{G} associated with the statistics s ,

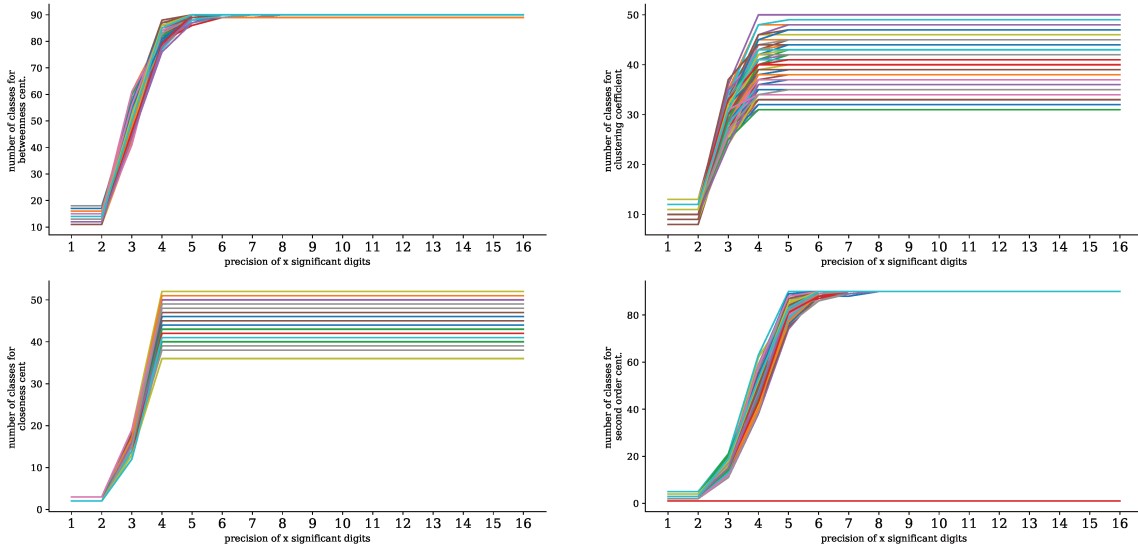


Figure 2.6 – ϵ choice effects in Erdős-Rényi graphs. The chosen ϵ corresponds to a number of significant digits to be used when comparing different nodal-statistics values. When the number of significant digits is higher the number of extracted classes increases. Depending on the considered statistics, the number of classes usually stabilizes around 4 or 5 significant digits.

and whose elements are the classes of equivalence $[a]_l, \forall l \in s(\mathcal{V})$,

$$[a]_{l, \sim_s} = [a] = \{b \in \mathcal{V} | a \sim_s b \iff s(a) = s(b) = l\}. \quad (2.3)$$

A necessary condition for two nodes to be automorphically equivalent is to belong to the same equivalence class.

Subsequently, we extend the equivalence relation associated with a statistics to any statistics collection $\mathcal{S} = \{s_i\}_{i=1, \dots, n}$, requiring that:

Definition 2.6.

$$a \sim_{\mathcal{S}} b \iff a \sim_{s_1} b, a \sim_{s_2} b, \dots, a \sim_{s_n} b. \quad (2.4)$$

Again, we can determine $P_{\mathcal{S}} = \{[a]_{\sim_{\mathcal{S}}}\}$, the induced partition by $\sim_{\mathcal{S}}$ on \mathcal{V} as the intersection of each class of the considered $\{s_i\}_{i=1, \dots, n}$. A visualization of the partitions associated with degree statistics is shown in Fig. 2.5 (a).

Since the automorphically equivalence relation preserves any nodal statistics, the nodal statistics-based equivalence relation associated with an infinity collection retrieves the automorphically equivalence. However, a finite nodal statistics collection with this property may also exist (Fig. 2.7).

We propose to compare statistics collection according to new defined global graph parameters which measure respectively parsimony and heterogeneity of its elements. These global parameters depend on the graph structure.

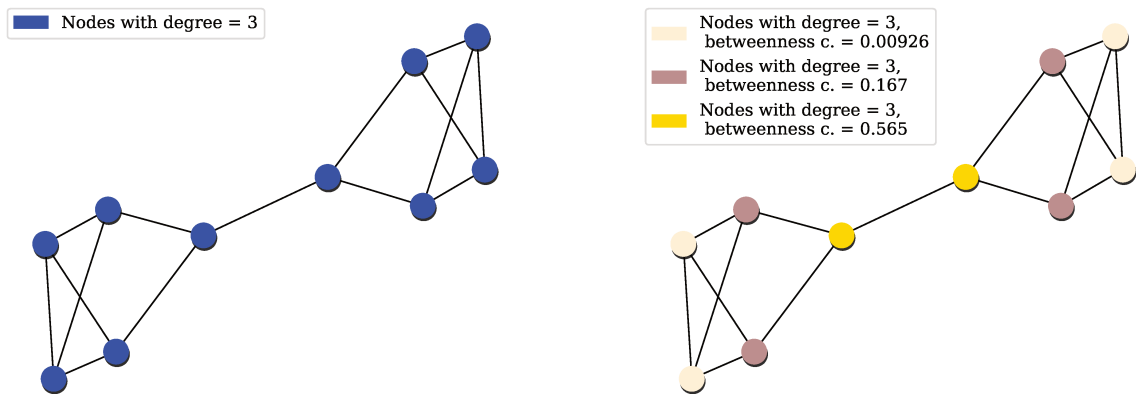


Figure 2.7 – Visualization of the structural pattern associated with different nodal statistics on a trivial graph whose nodes have the same degree. Left: Structural pattern associated with nodal degree. Right: Structural pattern associated with combination of betweenness centrality and degree. Colors correspond to different classes. In this toy example, the degree alone is not sufficient to reveal different equivalence classes and identifies a unique class. While, when two nodal statistics are considered, a nontrivial structural pattern appears.

Given $P_{\mathcal{S}}$, one can compute exactly the number of eligible automorphisms that map nodes into the same equivalence class, as it is computed below. Therefore, for each statistics collection on a graph \mathcal{G} , we can estimate how many permutations are prevented from being tested as being adjacency preserving in a brute force approach.

Definition 2.7. We introduce the power coefficient (PC) of a set \mathcal{S} for a graph $\mathcal{G} = (\mathcal{G}, \mathcal{E})$

$$\text{PC}_{\mathcal{G}}(\mathcal{S}) = \left| \log \frac{\#\{\text{permutations preserving } P_{\mathcal{S}}\}}{\#\{\text{permutations of } \mathcal{V}\}} \right| \quad (2.5)$$

with

$$\begin{aligned} \#\{\text{permutations preserving } P_{\mathcal{S}}\} &= \prod_{o \in P_{\mathcal{S}}} |o|! \\ \#\{\text{permutations of } \mathcal{V}\} &= |\mathcal{V}|!. \end{aligned}$$

The value $|\mathcal{V}|! e^{-\text{PC}}$ corresponds to an upper bound of the number of automorphisms of \mathcal{G} . Indeed, PC is increasing when more nodal statistics are combined together (Fig. 2.8). In the special case in which the permutations preserving $P_{\mathcal{S}}$ can be identified with the automorphisms of \mathcal{G} , PC can be interpreted as entropy of the network ensemble (Bianconi 2007) having \mathcal{G} topology. In all other cases, PC encodes the amount of information given by \mathcal{S} on the structure of \mathcal{G} and it is a parsimony measure for \mathcal{S} .

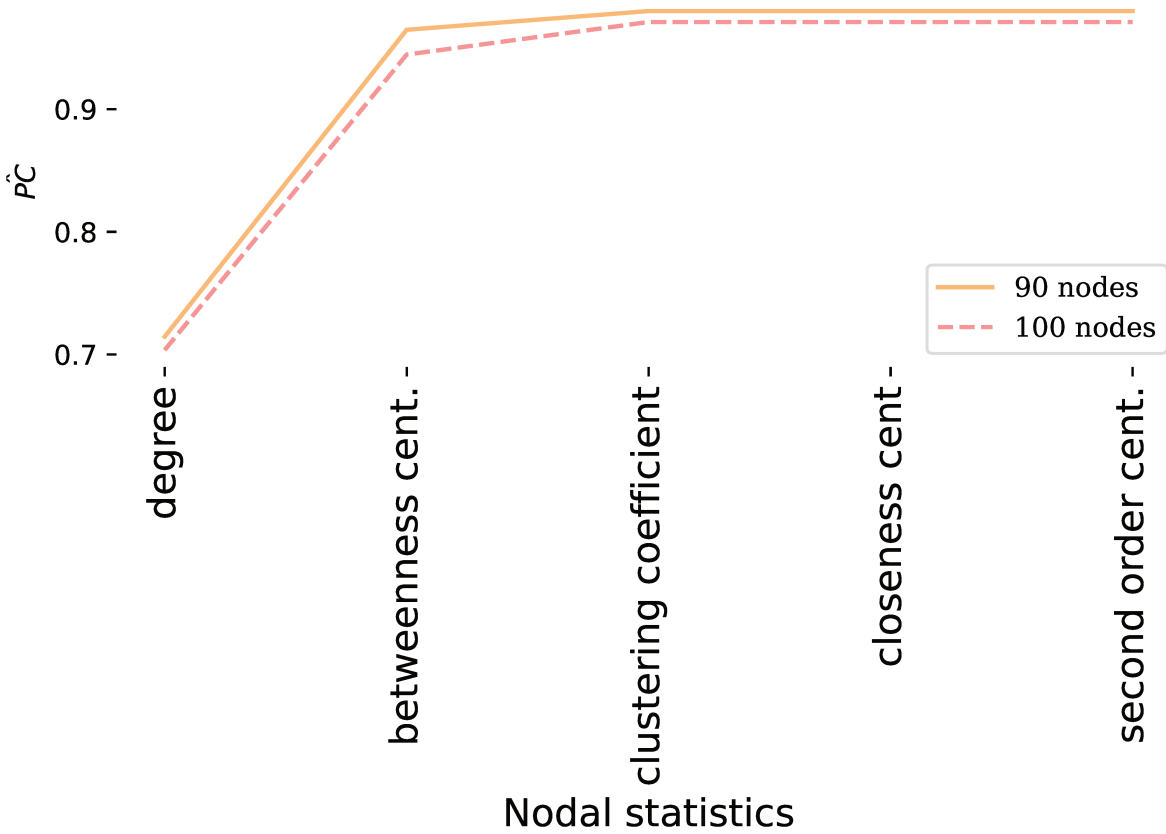


Figure 2.8 – Normalized power coefficient (\widehat{PC}) on nodal statistics incremental sets of two Erdős-Rényi graphs of 90 and 100 nodes: the more nodal statistics are considered the higher the \widehat{PC} .

Definition 2.8. Since PC takes values in $[0, \log \frac{1}{|\mathcal{V}|}]$, with an upper bound strictly depending on the number of nodes, we propose a normalized version of PC , $\widehat{PC} \in [0, 1]$:

$$\widehat{PC}_{\mathcal{G}}(\mathcal{S}) = \frac{PC_{\mathcal{G}}(\mathcal{S})}{\log |\mathcal{V}|} \quad (2.6)$$

$$= 1 - \frac{\log \#\{\text{permutations preserving } P_{\mathcal{S}}\}}{\log \#\{\text{permutations of } \mathcal{V}\}} \quad (2.7)$$

The higher the \widehat{PC} , the more the collection of statistics \mathcal{S} captures the heterogeneity of nodal structural roles in \mathcal{G} . Indeed, for a vertex-transitive graphs (i.e. all nodes are automorphically equivalent) $\widehat{PC}_{\mathcal{G}}(\mathcal{S}) = 0$ for all nodal statistics \mathcal{S} , while if it exists a collection $\bar{\mathcal{S}}$ s.t. $\widehat{PC}_{\mathcal{G}}(\bar{\mathcal{S}}) = 1$ then the graph \mathcal{G} does not admit nontrivial automorphisms. Hence, we introduce the notion of perfectly orthogonal statistics for heterogeneity evaluation of a collection elements. First, two nodal statistics are said to be perfectly orthogonal if their union-associated equivalence relation induces the trivial partition: all nodes belong to a single element set. Next, we extend the definition to any nodal statistics set: a nodal statistics collection is said to be perfectly orthogonal

if its induced partition is trivial.

Definition 2.9. An orthogonality measure for a given nodal statistics set on a graph can be assessed by computing the number of nodes in nontrivial classes on its associated partition:

$$O_{\mathcal{G}}(\mathcal{S}) = \frac{|\{v \in \mathcal{V} \text{ s.t. } \#[v]_{\sim_{\mathcal{S}}} \neq 1\}|}{|\mathcal{V}|} \quad (2.8)$$

$O_{\mathcal{G}}(\mathcal{S})$ is the ratio between the number of nodes in nontrivial classes and the total number of vertices and corresponds to an orthogonality score.

By definition, \mathcal{S} is perfectly orthogonal if and only if $O_{\mathcal{G}}(\mathcal{S}) = 0$.

2.2.2 Structural equivalence for graph collections

Structural pattern comparison

Graphs that have the same node set can be compared by evaluating the correspondence between their structural patterns. The node set constraint can be easily circumvented when two graphs do not share all the nodes, by including all nodes to the graphs vertices set and allowing the network to be composed of more connected components. Indeed, each network can be seen as the union of one strongly-connected component with as many single disconnected vertices as needed. We propose to compare structural patterns as follows. Let $\mathcal{G}, \mathcal{G}'$ be two graphs having same vertices \mathcal{V} and let \mathcal{S} be a statistics collection whose associated partitions are $P_{\mathcal{S}}, P'_{\mathcal{S}}$ on $\mathcal{G}, \mathcal{G}'$ respectively.

Definition 2.10. Given bijective mapping from $P_{\mathcal{S}}, P'_{\mathcal{S}}$ to an initial segment of the natural numbers as enumerations, let $c(v_i), c'(v_i)$ be the enumeration of the classes of v_i , the correspondence structural pattern score between $\mathcal{G}, \mathcal{G}'$ is defined as:

$$C(\mathcal{G}, \mathcal{G}') = \max_{\pi \in \Pi} \frac{1}{|\mathcal{V}|} \sum_{i=1}^{|\mathcal{V}|} \mathcal{X}(\pi(c(v_i)) = c'(v_i)) \quad (2.9)$$

where Π is the set of all coupling between the elements in $P_{\mathcal{S}}$ and the elements in $P'_{\mathcal{S}}$ and \mathcal{X} is the indicator function. A possible implementation of $C(\mathcal{G}, \mathcal{G}')$ in polynomial time is given by the Hungarian algorithm (Kuhn 1955) for assignment problems with a complexity $\mathcal{O}(\max\{|P_{\mathcal{S}}|, |P'_{\mathcal{S}}|\}^3)$ which in the worst case equals $\mathcal{O}(|\mathcal{V}|^3)$.

The correspondence structural pattern score can be applied for two different purposes: to evaluate structural pattern similarity between two graphs (Fig. 2.5 (b)) or to evaluate the similarity of structural patterns associated with different statistics collection on the same graph (Fig. 2.9). Since at least one class of $P_{\mathcal{S}}$ shares one element with one of the classes in $P'_{\mathcal{S}}$, $C(\mathcal{G}, \mathcal{G}') \geq \frac{1}{|\mathcal{V}|}$. As a consequence, even perfectly orthogonal statistics set of a graph can exhibit a correspondence pattern score higher than zero (Fig.2.9 (c)). If for every class in $P_{\mathcal{S}}$ there exists one class of $P'_{\mathcal{S}}$ having all and only its elements, then $P_{\mathcal{S}} = P'_{\mathcal{S}}$ and $C(\mathcal{G}, \mathcal{G}') = 1$. The opposite is also true: same partitions determine a correspondence structural pattern score equals to 1. More general prop-

erties of the defined global measures can be found in the Sec. [2.2.3](#).

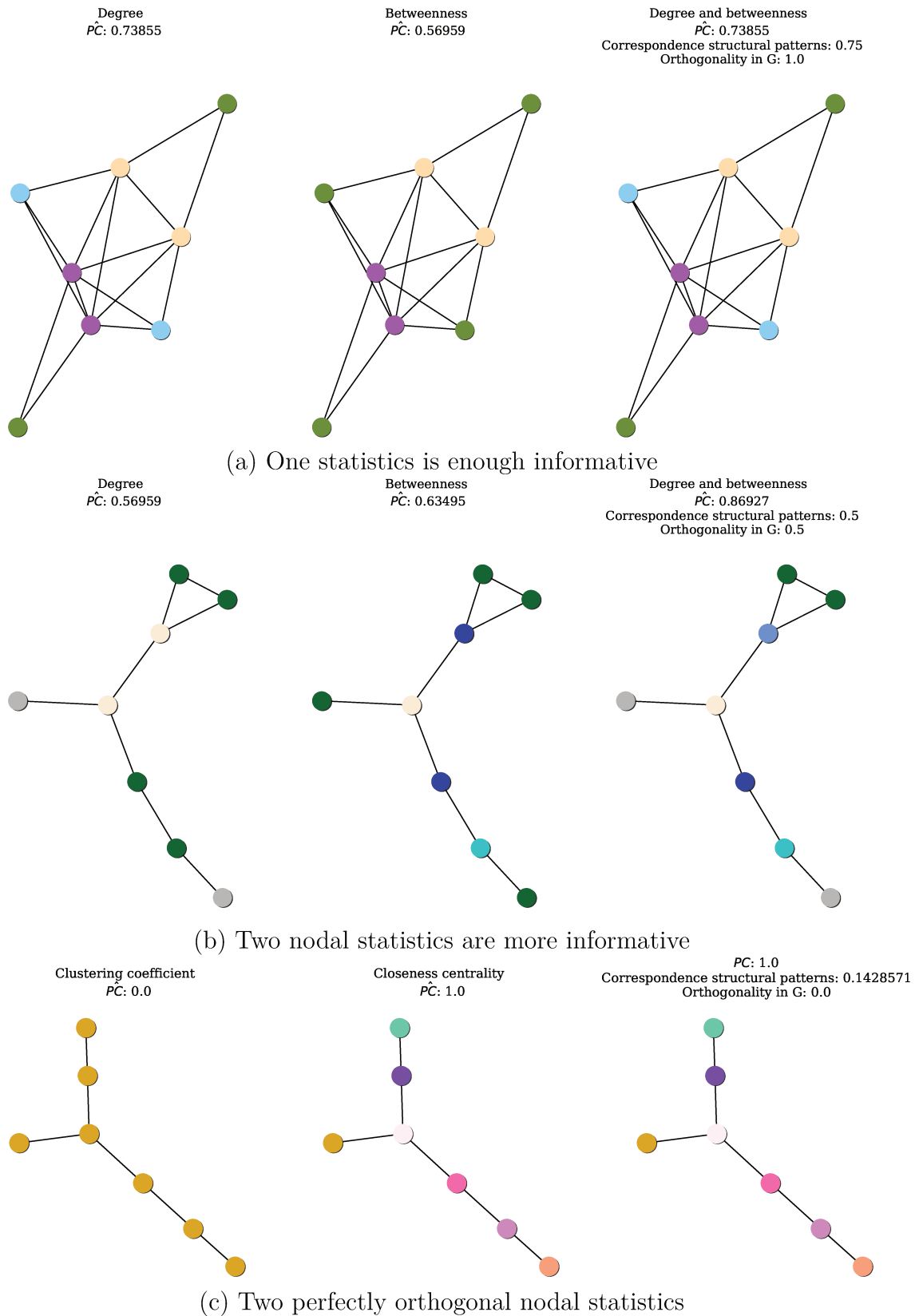


Figure 2.9 – Visualization of global metrics on three graphs (column) for three cases (row). (a) One statistics is sufficient: the \widehat{PC} value observed when considering two statistics (right) is equal to degree alone (left), meaning that betweenness provides no more useful information for determining the structural pattern. When we compare the two partitions with the degree statistics alone, we observe 75% of the nodes belonging to the same class and no trivial class (orthogonality equals to 1). This result can be interpreted in two different ways: more node statistics are needed to identify hub nodes, or the considered graph does not contain hub nodes. (b) Two statistics are more informative: to associate degree and betweenness improves the power coefficient. The identified patterns share half of nodes and their orthogonality is 0.5, meaning that their partition situates half of nodes in trivial classes. (c) Perfect orthogonality: The minimal orthogonality value is reached when one of the two compared structural pattern has only one class and the other contains trivial classes.

Nodes distinctiveness or similarity

Since eligible automorphisms can only map nodes within classes, a node in a trivial class (one element class) is always a fixed point. Thus, to provide a group characterization at nodal level, we propose to enumerate for each node its participation into nontrivial classes as a measure of the node's propensity not to be a fixed-point of admissible automorphisms. The more a node appears in nontrivial classes, the more it shows common properties with some other nodes in the graph. The persistence of a node to belong to a class in an entire graph group reveals the presence of shared properties among the group for the given node, i.e. hubs nodes, peripheral nodes, etc (Fig. 2.5 (c)). Thus,

Definition 2.11. given a graphs group $G = \{\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k) \text{ s.t. } \mathcal{V}_k = \mathcal{V}\}$, and a statistics collection \mathcal{S} we count the percentage of participation of each node of \mathcal{V} in nontrivial classes:

$$\forall v \in \mathcal{V} \quad \text{NPP}_G^{\mathcal{S}}(v) = \text{NPP}_G(v) = \frac{|\{\mathcal{G}_k \in G \text{ s.t. } \#[v]_{\sim_{\mathcal{S}}}^{\mathcal{G}_k} \neq 1\}|}{|G|} \quad (2.10)$$

with $[v]_{\sim_{\mathcal{S}}}^{\mathcal{G}_k}$ the class of v in \mathcal{G}_k in the partition induced by \mathcal{S} . In the following, with abuse of notation, we suppose \mathcal{S} fixed and avoid to explicitly repeat the dependency. A high percentage of participation means the node shares its role in many graph instances in the group, while at the opposite a node which does not share its role consistently shows a distinctiveness behavior in the considered graphs collection.

2.2.3 General properties

Properties of PC

Note that all the listed properties are true also for $\widehat{\text{PC}}$.

Observation 1. On the same graph the PC increases on increasing collections of nodal statistics (Fig. 2.8).

Observation 2 (PC for vertex-transitive graph). the PC of every nodal statistics collection equals zero for vertex-transitive graph.

Observation 3. If the PC of a nodal statistics collection equals the PC of one of its elements, then the correspondence structural patterns score of the structural pattern associated with the collection and the one of that element is 1.

Observation 4. If the PC of a graph equals 0 for one collection of statistics, then the graph does not admit nontrivial automorphisms.

Observation 5. If two graphs are isomorphic than their PC is the same for all statistics collection.

Relation with network ensembles entropy Our proposed method serves a specific purpose, focusing on introducing novel statistical tools adapted for the analysis of brain functional connectivity networks. However, its significance extends beyond its primary objective, as it also holds relevant implications for prior publications that aimed to characterize the complexity of networks (Bianconi 2007; Bianconi 2009; Bogacz et al. 2006). In particular, when adopting a statistical-mechanics approach, an ensemble of networks possessing specific structural properties can be effectively examined through its associated entropy. The number of eligible automorphisms of a graph corresponds to the number of permutations of its adjacency matrix. Following Bogacz et al. 2006, the partition function of the ensemble of a given topology $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\text{Aut}(\mathcal{G})$ the set of automorphism of \mathcal{G} , is defined as

Definition 2.12 (Partition function of the network ensemble).

$$Z(\mathcal{G} = (\mathcal{V}, \mathcal{E})) = \left| \frac{|\mathcal{V}|!}{|\text{Aut}(\mathcal{G})|} \right|. \quad (2.11)$$

We denote PC^* the PC computed for a collection of statistics whose equivalence relation corresponds to the automorphisms relation. Then, we have

$$\text{PC}^* = \left| \log \frac{1}{Z} \right| \quad (2.12)$$

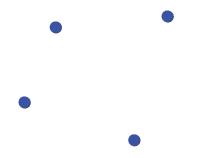
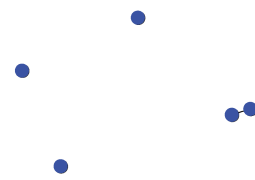
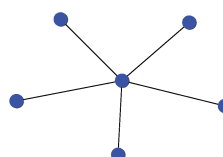
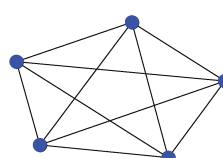
$$\text{PC}^* = |-\log Z| \quad (2.13)$$

$$\text{PC}^* = \log Z \quad (2.14)$$

$$\text{entropy} \propto \text{PC}^* \quad (2.15)$$

This is in line with the idea that a higher level of order in the graph structure is associated with lower entropy, indeed for vertex-transitive graphs the entropy reaches its minimal value at zero (Bianconi 2007; Bianconi 2009). A comparison with-in entropy and PC^* can be found in Table 2.2. Note that in the first and last examples the statistics collection choice does not affect the PC.

Table 2.2: Entropy and PC* on known graphs

Graph	Entropy	PC*
 <p>Trivial graph with n isolated nodes</p>	$\frac{1}{n} \binom{\log \frac{n(n-1)}{2}}{0} = 0$	$\forall \mathcal{S}, \quad \text{PC}(\mathcal{S}) = \text{PC}^*$ $\text{PC}^* = \log \frac{n!}{n!} = 0$
 <p>ER model $p \binom{n}{2} = 1$ 2 connected nodes and $n-2$ isolated nodes</p>	$\frac{1}{n} \binom{\log \frac{n(n-1)}{2}}{1} =$ $= \frac{1}{n} \log \frac{n(n-1)}{2}$	$\mathcal{S} = \{\text{deg}\}, \quad \text{PC}(\mathcal{S}) = \text{PC}^*$ $\text{PC}^* = \log \frac{(n-2)! 2}{n!} =$ $= \log \frac{2}{n(n-1)}$
 <p>Star graph with n nodes and $n-1$ edges</p>	$\frac{1}{n} \log \frac{n!}{(n-1)!} = \frac{1}{n} \log n$	$\mathcal{S} = \{\text{deg}\}, \quad \text{PC}(\mathcal{S}) = \text{PC}^*$ $\text{PC}^* = \log \frac{(n-1)!}{n!} =$ $= \log \frac{1}{n}$
 <p>Complete graph with n nodes having $n-1$ edges</p>	$\frac{1}{n} \log \frac{n!}{n!} = 0$	$\forall \mathcal{S}, \quad \text{PC}(\mathcal{S}) = \text{PC}^*$ $\text{PC}^* = \log \frac{n!}{n!} =$ $= 0$

Properties of Orthogonality

Observation 6. A nodal statistics whose induced partition is composed of classes having each one a unique element, is perfectly orthogonal with every nodal statistics.

Observation 7. If collection of statistics is perfectly orthogonal, all other collections having as subset that collection are perfectly orthogonal as well;

Observation 8. If a perfectly orthogonal statistics set exists on a graph, then the graph does not admit nontrivial automorphisms.

Properties of Correspondence of structural pattern score

Observation 9. All graphs defined on the same node set, having same degree sequence, have a correspondence of structural patterns associated with the degree statistics equals to 1.

Observation 10. The minimum values of structural pattern score is given by $\frac{1}{|V|}$

Observation 11. If on the same graph, the structural patterns score of different nodal statistics reaches the minimal value, then the nodal statistics are perfectly orthogonal.

2.3 Summary and conclusion

Graph models become largely used in real world applications and many nodal statistics have been proposed for node roles detection. However, the most informative statistics for graph comparison is highly dependent on the observed data and combining more statistics can be relevant.

We propose a mathematical framework with the specific purpose of providing new statistical tools for the analysis of brain functional connectivity networks.

We introduce a nodal-statistics-based equivalence relation and propose an innovative way to combine nodal statistics for graph structural pattern detection. We will use this notion to compare different graphs and characterize graph family defined on the same node set. As the equivalence relation depends on a collection of nodal statistics, we define a power coefficient and an orthogonality score to evaluate parsimony and heterogeneity of the collection. Such coefficients can be used as revisited measure of nodal statistics dependency.

We define a graph similarity based on node roles and a mathematical tool to detect nodes persistently different from others, by computing the percentage of participation in nontrivial classes. Interestingly, nodes which tend to persistently belong to trivial class are likely to play peculiar roles in the graphs, while at the opposite nodes with a high participation, appear to share similar property with other nodes.

— Chapter 3 —

Applications

Abstract

In this chapter, we present experiments demonstrating the benefit of a graph structural pattern approach in functional connectivity brain networks, both for characterization and in classification task. We compare real data with synthetic models, achieve consistent characterizations of healthy brain connectivity, enable pathological condition differentiation, and highlight the impact of time-series length on regional characterization. These findings underline the potential of our framework in comprehensive brain network analysis.

Contents

3.1	Databases	61
3.1.1	Preprocessing of rs-fMRI data	61
3.1.2	Adopted network inference protocol	62
3.2	(1) Synthetic Data and Real Data	65
3.2.1	Material	65
3.2.2	Results and Discussion	65
	Power Coefficient.	65
	Orthogonality and Correspondence of Structural Pattern.	67
	Nodal Percentage of Participation.	70
3.3	(2) Healthy Population Characterization	73
3.3.1	Material	73
3.3.2	Results and Discussion	73
	Orthogonality Score.	74
	Nodal Percentage of Participation for Regional Characterization.	75
	Discussion	76
3.4	(3) Discrimination of Patients Population	76
3.4.1	Comatose Patients	77
	Orthogonality coefficient	77
	Correspondence Structural Pattern	78
	Nodal Percentage of Participation	81
3.4.2	Parkinson Disease	83
	Orthogonality coefficient	84
	Correspondence of structural pattern score	85
	Nodal Percentage of Participation	86
3.5	(4) Reliability	89
3.5.1	Material and Method	89
3.5.2	Results and Discussion	91
3.6	Summary and conclusion	96
3.6.1	Limitations	97
3.6.2	Future work	98

In this chapter, we present experiments leading to key results regarding the application of a graph structural pattern characterization in functional connectivity brain networks. First, we provide a brief overview of the databases we utilized. To ensure the generalizability of our method and to demonstrate its adaptability, we made an effort to incorporate a wide range of data. However, as we elucidate in the previous chapter, it is crucial for the data to adhere to specific criteria to ensure the robustness of the results.

The four main experiments are organized as follows. First, a comparison between real data and synthetic generative models is provided. This serves as a proof of concept, illustrating that graph structural patterns have the capacity to capture information unique to brain connectivity networks.

In the second experiment, we detail how we achieved a consistency results across healthy subjects from various databases. This endeavor results in a comprehensive healthy characterization of brain connectivity.

Our third experiment pertains to a classification method for distinguishing between different pathological conditions. This method is effective even in discerning subtle differences among groups with varying symptomatology.

Lastly, we explore the influence of the length of time-series on regional characterization. We recommend the use of long acquisition time to effectively capture relevant patterns.

Overall, these results showcase the potential of utilizing graph structural patterns in analyzing functional connectivity brain networks, encompassing generalization across healthy subjects, discrimination between pathological conditions, and the significance of data length in regional characterization.

3.1 Databases

3.1.1 Preprocessing of rs-fMRI data

The data preprocessing is based on the resting state fMRI pipelines developed by V. Muñoz-Ramírez, M. Dojat, C. Delon-Martin and S. Achard at the Grenoble Institute of Neuroscience (Muñoz-Ramírez et al. 2021). The pre-processing requires an anatomical scan and the resting state acquisition. It encompasses several essential steps, as outlined below and summarized in Figure 3.1.

- Anatomical segmentation: the first step involves segmenting the anatomical data using the Statistical Parametric Mapping (SPM) software. This segmentation process categorizes each voxel into anatomical structures within the brain.
- Resting state realignment and conversion: in this step, the 4D resting state data is realigned, which ensures that the time series data for each voxel is temporally

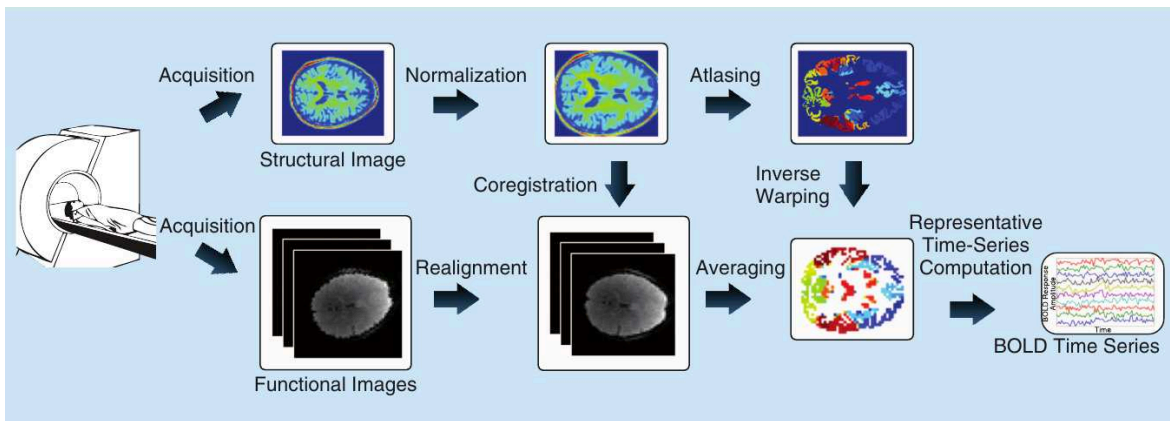


Figure 3.1 – Scheme for BOLD time-series extraction from fMRI acquisition. Figure adapted from [Richiardi et al. 2013](#).

aligned with the others. Additionally, the data is converted from a 4D volume to a series of 3D slices.

- Atlas registration: the adopted atlas is registered to the native space of the resting state data. This alignment allows for mapping brain regions onto the subject's volumes for subsequent analyses.
- Artifact detection: the pipeline includes a procedure to detect artifacts within the resting state data. Artifacts can arise from various sources, such as motion or physiological noise, and their identification is essential to ensure data quality. Points associated with a motion of more than 2mm are discarded.

These steps collectively form the preprocessing pipeline for the rs-fMRI data, aiming to enhance the quality and reliability of subsequent analyses. The use of this pipeline helps to align the anatomical and functional data, correct for potential artifacts, and prepare the data for further investigation into graph structural patterns and functional connectivity within the brain.

3.1.2 Adopted network inference protocol

The definition of the functional network is achieved throughout different phases shown in Fig. 3.2. First, the acquired fMRI data are aggregated over regions which are determined according to an anatomical labeling. Particularly, depending on the desired granularity, we adopt one of the following atlas for our analysis: AAL90, AAL16, and AICHA ([Tzourio-Mazoyer et al. 2002](#); [Rolls et al. 2020](#); [Joliot et al. 2015](#)). These atlases divide the brain into 90, 116 and 384 regions of interest (ROI), respectively. The choice of atlas depends on the specific research question and the desired resolution of analysis. For each ROI, we estimate a unique time series signal by averaging the fMRI time series over all voxels in each parcel, weighted by the proportion of gray matter in each voxel.

The following stage consists in the application of the discrete wavelet transform to the fMRI time series. Thanks to this procedure, each time series is decomposed into a set of compactly supported basis function, which are uniquely scaled in frequency and located in time. As a results, for each subject, different fMRI time series at distinct

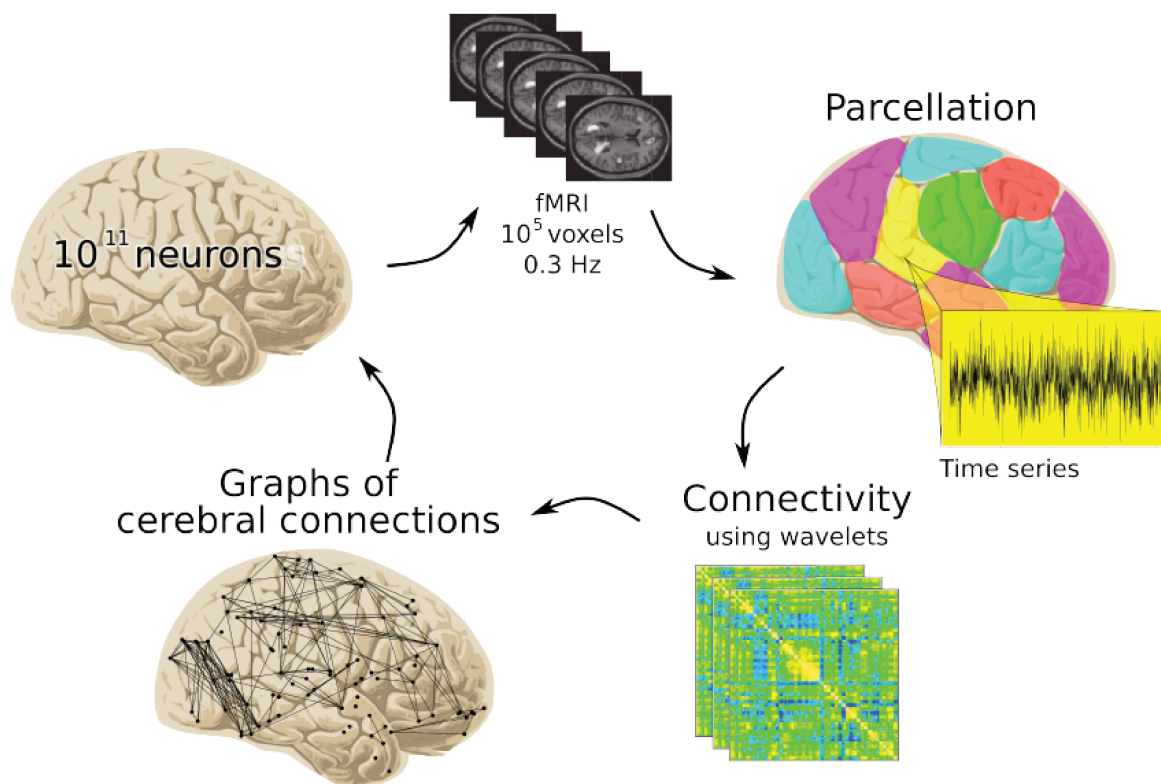


Figure 3.2 – Adopted framework for graph definition. Courtesy of Sophie Achard.

scales are at disposal. At each scale, which represents an interval of frequencies, the correlation among regions is estimated. Specifically, we estimate the correlation at the frequency band lower than 0.1Hz, since it has been observed that the resting state information activity is mainly captured at this frequency range (Biswal et al. 1995; Lowe et al. 1998; Cordes et al. 2000; Salvador et al. 2005).

Finally, we consider unweighted graphs obtained by graph filtering procedure on the absolute of the correlation matrix. A subject-dependent threshold is determined in order to obtain a fixed number of edges in all the estimated networks (Richiardi et al. 2013; Vidal et al. 2011; De Vico Fallani et al. 2017). The threshold guarantees a specific sparsity level in the final graph, ensuring that only the strongest connections in absolute value, are preserved independently on the correlation values (Achard et al. 2006; Fornito et al. 2013). The choice of the threshold is performed to extract graphs belonging to the *small-world regime* as shown in Fig. 3.3 and corresponds to low sparsity values. The small-world regime is defined in terms of global and local efficiency: when a network has global and local efficiencies comprise between the ones of a corresponding regular lattice and random networks. Moreover, to ensure connectivity, the edges of the Minimum Spanning Tree are always added. We refer to our work in Appendix B.2 where we make effort to avoid the selection of an *a priori* threshold and to Appendix B.3 where we refine the small-world regime definition by identify Watt-Strogatz hyperparameters model to fit global and local efficiencies of real data.

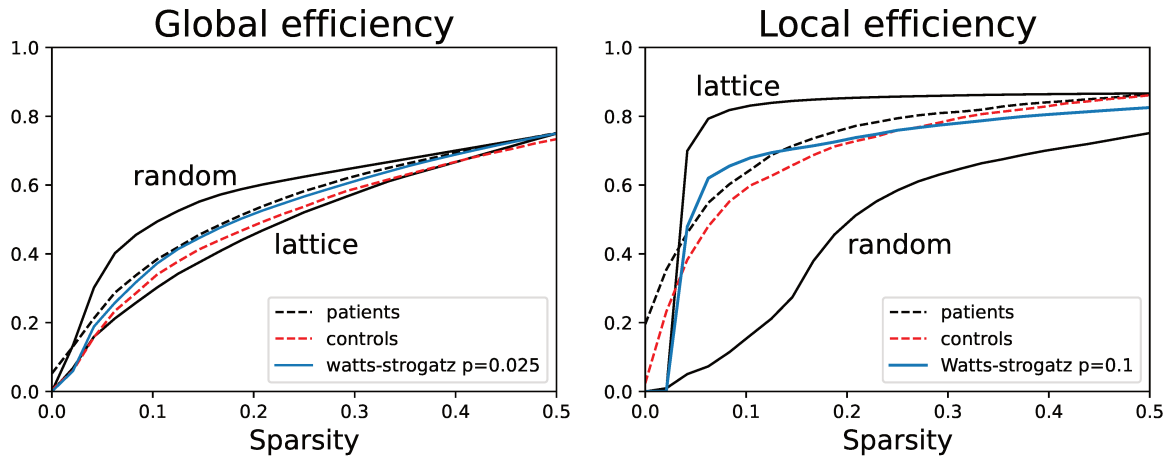


Figure 3.3 – Visualization of the small-world regime characterization of functional connectivity graphs.

Table 3.1: List of the studied databases including different classes of subjects. #Vol: Number of rs-fMRI volumes, HC: Healthy Controls, CO: Comatose patients, PD: Parkinsonian patients, PD-H: Parkinsonian patients with visual hallucinations, PD-I: Parkinsonian patients with visual illusions. Gin-Chuga: Grenoble Hospital, Chu-CF: Clermont-Ferrand Hospital. ChuStr: Strasbourg Hospital, HCP: Human Connectome Project, iShare: internet-based Students’ Health Research Enterprise, Gin-Chuga: Grenoble Hospital.

Provider	Subjects	Scanning Parameters				Age range years
		B0	TR (ms)	#Vol	Frequency Band	
HCP (Termenon et al. 2016)	HC(100×2)	3T	720	1200	0.043-0.087 Hz	20-43
iShare (Tsuchida et al. 2017)	HC(1814)	3T	850	1046	0.037-0.074 Hz	18-35
Gin-Chuga (Muñoz-Ramírez et al. 2019)	HC(11)-PD- <i>de novo</i> (13)	3T	1000	500	0.031-0.063 Hz	46-70, 51-70
ChuStr (Achard et al. 2012)	HC(20) - CO (17)	1.5T	3000	405	0.042-0.084 Hz	25-45, 21-82
Chu-CF (Marques et al. 2022)	HC(20) PD-H (17) PD-I (19)	3T	3000	200	0.041-0.083 Hz	60-78, 63-78, 61-75

We build networks from resting state fMRI (rs-fMRI) datasets from different databases (Table 3.1) which gather Healthy Controls (HC) and different pathologies. For dataset including patients with Parkinson Disease (PD), we consider a modified version of the AAL3 (Rolls et al. 2020), including the first 94 regions and the subcortical regions known to be affected in PD (substantia nigra, red nucleus, cingulate cortex, accumbens nucleus), in addition, we parcellate the Cerebellum in three regions per hemisphere (Posterior, Anterior, Inferior), resulting in a total of 106 regions. The same data and atlas are used in Muñoz-Ramírez et al. 2019. The iShare dataset was pre-processed similarly but with different software elements (Tsuchida et al. 2017).

Information about the main acquisition parameters and the populations are provided in Table 3.1, for more details we refer to the corresponding references.

3.2 (1) Synthetic Data and Real Data

Based on a paper written in collaboration with Sophie Achard and Michel Dojat and published in Physical Review E. (Carboni et al. 2023b)

In this section, we aim at providing interpretation of the different defined measures of interest of the previous chapter, namely the Power Coefficient and the Orthogonality score over a single graph, the Correspondence of Structural Pattern as pair-wise similarity score among networks and finally the local characterization of a graph collections through the nodal percentage of participation in nontrivial classes (NPP). For this extent, we evaluate our framework on synthetic data, as null-models for which we have a clear understanding of the underlying topological structure, and compare them with real data to show as a proof of concept that our proposal captures meaningful brain network information.

3.2.1 Material

We consider different generative graph models and compare them according to their sparsity level, defined as the ratio between the edge count on the graph and the edge count in a complete graph having the same nodes. In generating synthetic graphs, we constraint the number of nodes to 90 to align with the corresponding real dataset, AAL90 (Tzourio-Mazoyer et al. 2002). We examine ER, WS and BA1,BA2 models, together with a synthetic version of corresponding real networks coming from HCP: DSP conserving the same observed degree distribution, and EPA and EC models. As real data, we consider the 200 networks coming from HCP dataset (Tab. 3.1). When analyzing graph at fixed sparsity, we select 0.1 which guarantees that each extracted network belongs to small-world regime corresponding to global and local efficiencies comprised between the ones of ER graph and ones of the complete graph (Achard et al. 2007; Latora et al. 2001, Fig. 3.3). We consider the classical nodal statistics previously defined: degree, clustering coefficient, and the three centrality measures of betweenness, closeness, and second-order. We refer to the previous chapter for the definition of the different considered metrics (Chapter 2).

3.2.2 Results and Discussion

Power Coefficient.

In Fig. 3.4, we report the clustering coefficient \widehat{PC} on different generative models and real connectivity graphs with respect to their sparsity ratio. We can appreciate the different behavior of \widehat{PC} across models and sparsity levels.

As expected, for every nodal statistics the \widehat{PC} equals zero when computed over the complete graph (corresponding to the sparsity ratio of 1). In this case, all nodes are automorphically equivalent and it is not possible to extract any equivalence class.

In the Barabási-Albert model (BA1, BA2), we observe a monotone decreasing $\widehat{PC}(cc)$ with respect to the sparsity. Indeed, when the sparsity is low, we have few nodes of high clustering coefficient and many nodes of very low coefficient values. The number of automorphisms exchanging nodes of low values is then higher for small sparsity, while when the sparsity ratio increases the clustering coefficient values distribution tends to be less concentrated on the node set, identifying more classes and corresponding to higher \widehat{PC} . ER and WS show similar behavior especially for high sparsity values, while when the sparsity is low, WS tends to differ from ER model.

Regarding brain models, EPA fits correctly the HCP networks when the sparsity is higher than 0.7. EC and DSP curves follow the HCP curve tendencies, but with lower \widehat{PC} values. A possible explanation of this difference is the presence of hubs in HCP networks not well reproduced in the models. Indeed, a higher number of hubs will result in higher \widehat{PC} score.

Interestingly, the \widehat{PC} on the real data have the highest performance at all sparsity levels, reproducing the higher complexity of the real FC network topologies. When evaluating the \widehat{PC} of different measures on the same model, we can select for each

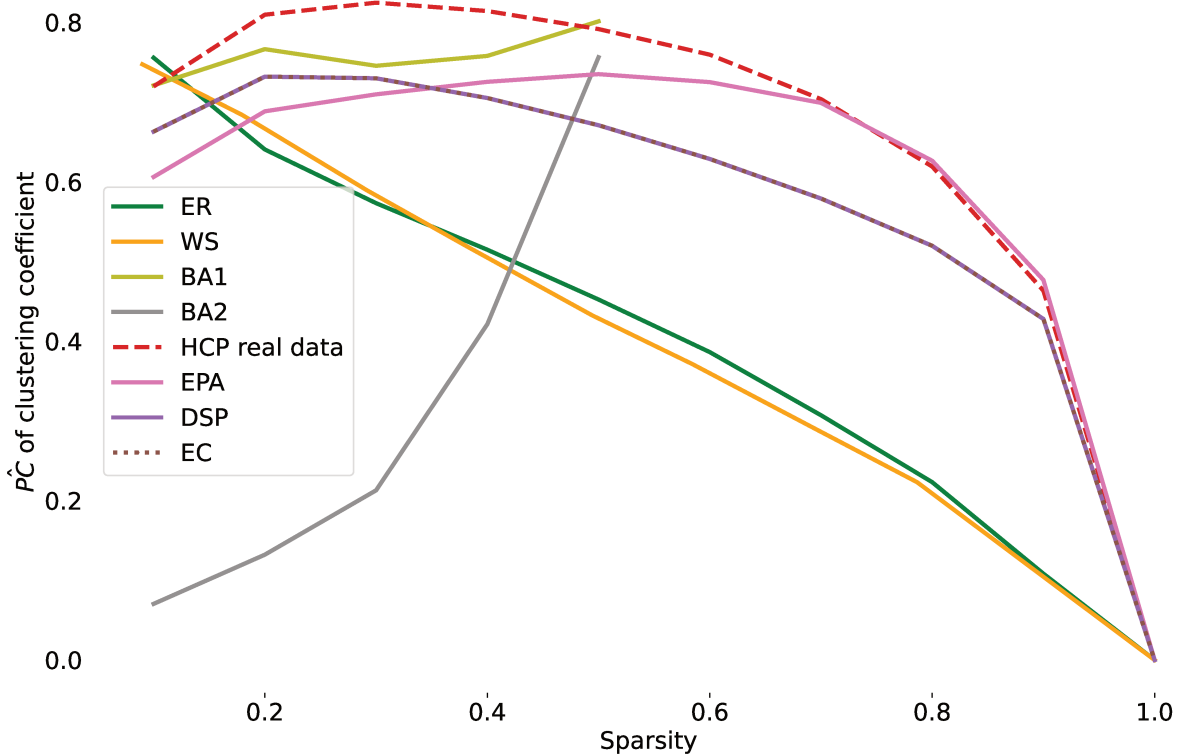


Figure 3.4 – Mean normalized power coefficient (\widehat{PC}) of clustering coefficient statistics on different generative models and real brain connectivity graphs (HCP) at different sparsity levels. ER: Erdős-Rényi; WS: Watts-Strogatz; BA1, BA2: Barabási-Albert; DSP: Degree sequence preserving model; EPA: Economical preferential attachment model; EC: Economical clustering model.

sparsity ratio which nodal statistics have the higher discriminative power on the node set. A same analysis can be performed over different nodal statistics, (see for instance Fig. 3.5).

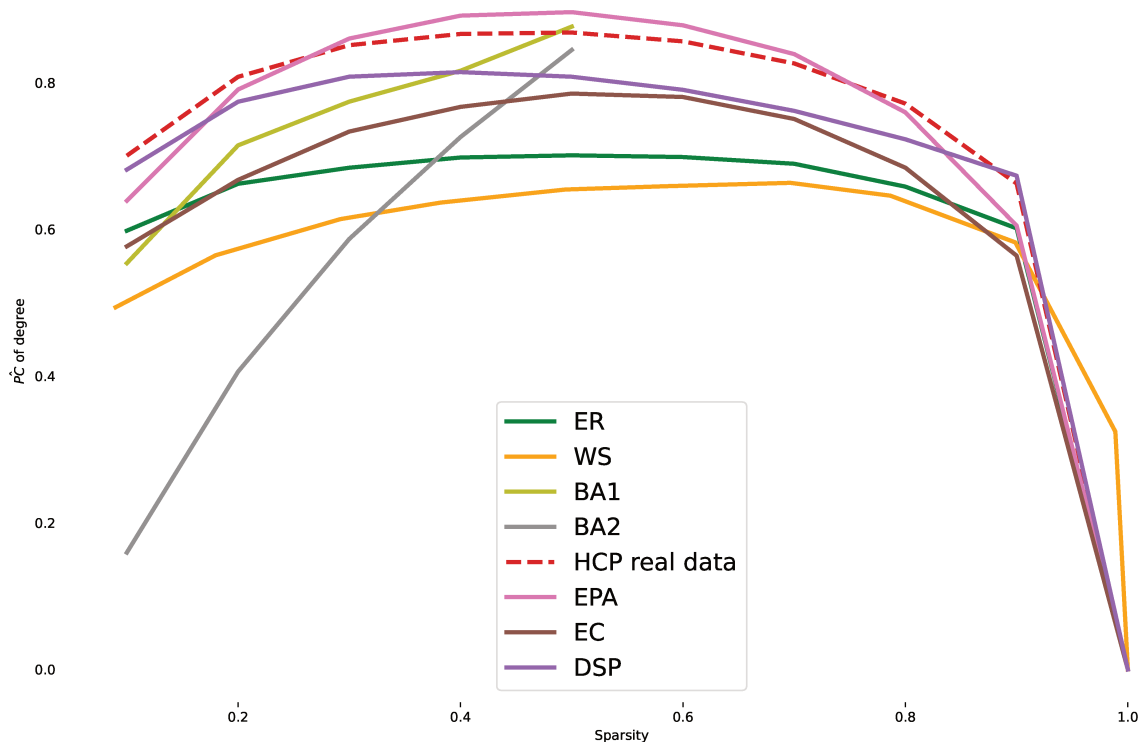


Figure 3.5 – Mean normalized power coefficient (\widehat{PC}) of degree statistics on different generative models and real brain connectivity graphs (HCP) at different sparsity levels. ER: Erdős-Rényi; WS: Watts-Strogatz; BA1, BA2: Barabási-Albert; DSP: Degree sequence preserving model; EPA: Economical preferential attachment model; EC: Economical clustering model. Interesting, the \widehat{PC} on the real data have the best performance at all sparsity levels. When evaluating the \widehat{PC} of different measures on the same model, we can select for each sparsity ratio which nodal statistics have the higher discriminative power on the node set.

Orthogonality and Correspondence of Structural Pattern.

In Fig. 3.6, we evaluate orthogonality and correspondence structural pattern of statistics pairs in WS and BA2 models at 0.1 sparsity. A visualization of their structural patterns is shown in Fig. 3.7.

In WS model, the degree shows high orthogonality values with all nodal statistics: many nodes that have same degree also share a second nodal statistics value. This is likely due to its degree distribution. Indeed, in a general WS graph $\mathcal{G}_{n,k,p}$ the degree distribution has a peak at the k values, meaning many nodes have approximately the

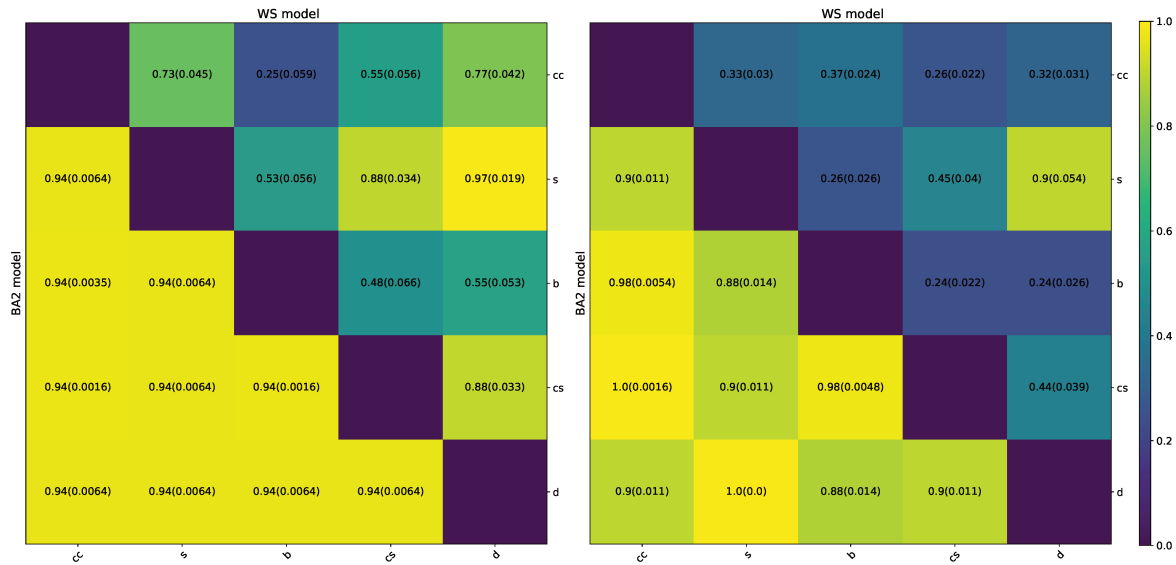


Figure 3.6 – Nodal statistics pair comparison on two models. Upper Triangular Matrix: WS (Watt-Strogatz model); Lower Triangular Matrix: BA2 (Barabási-Albert) model. Left: Orthogonality for a pair of statistics; Right: Correspondence structural pattern score for a pair of statistics. cc: clustering coefficient; b: betweenness centrality; d: degree; s: second-order centrality; cs: closeness centrality.

same degree (Albert et al. 2002). Consequently, the probability of retrieving high populated classes associated with degree is increased in this model.

Interestingly, the correspondence patterns scores between the degree and the other statistics are low except for the second-order centrality (Fig. 3.7, Top Right). Degree and second-order centrality capture different topological graph features (Kermarrec et al. 2011) and usually appear unrelated in complex networks. However, in the considered case, their induced partitions on the graph largely overlap. Indeed, they exhibit a strong negative correlation coefficient (-0.98 in average). Their high orthogonality and high correspondence scores reveal this correlation.

In WS model, the statistics pair, which shows the least correspondence pattern scores, is composed by degree and betweenness centrality: while trivial degree classes capture high connected nodes, the betweenness centrality better refines the class associated with the average degree value.

Completely different results are observed in BA2 model, for which the orthogonality of all considered statistics pair together with their correspondence scores appear close to 1.0. This shows how in preferential attachment model all statistics pairs determine almost the same structural patterns: few populated classes of high connected nodes and high populated class for the leaves. Indeed, for BA1,BA2 models a very high correspondence of structural patterns associated with single statistics is detected (Fig. 3.7 bottom).

We report results over the HCP dataset, by considering degree and betweenness centrality associate-equivalence relation. For this pair, low orthogonality and correspondence patterns scores are observed both on WS model and real data (Fig. 3.6, Fig. 3.8).

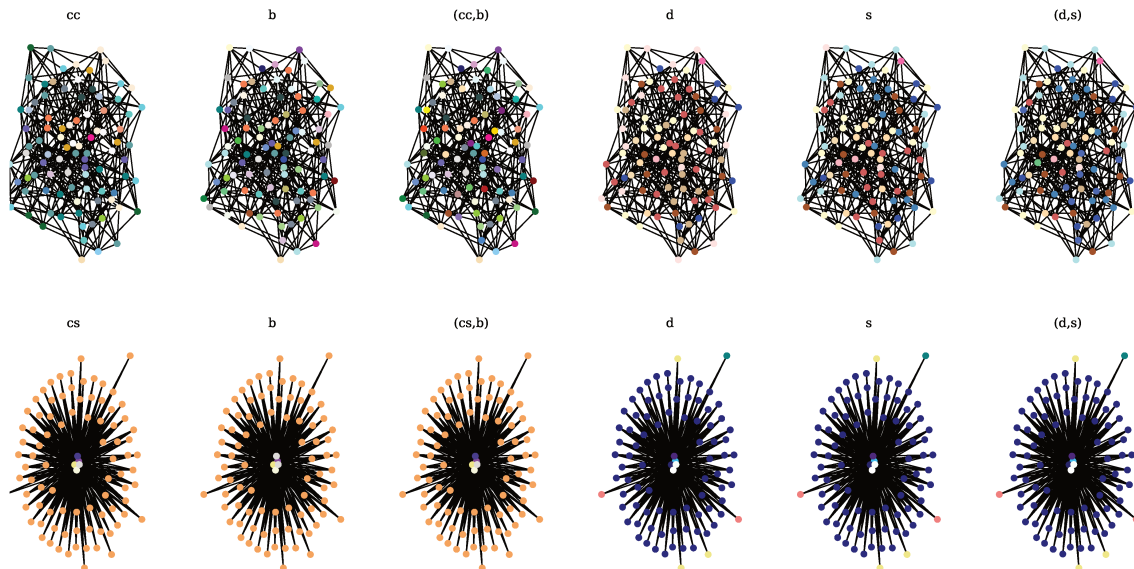


Figure 3.7 – Structural patterns associated with different statistics on Watt-Strogatz (top) and Barabási-Albert (bottom) models. cc: Clustering coefficient; b: Betweenness centrality; d: degree; s: Second-order centrality; cs: Closeness centrality.

In Fig. 3.9, we compare the correspondence structural pattern score distribution for generative models and HCP datasets. The observed ER and WS distribution values are lower compared to real data. Moreover, when considering a dataset combining half HCP real networks and half ER networks, we observe a reduction in the structural pattern comparison values and an increase in the variance. Interestingly, while HCP data and WS model both exhibit small world properties, their score distributions are very distinct, meaning that the computed structural pattern score on real data comes from peculiar brain properties which are not completely captured by networks of small world regime.

For brain connectivity models, EC and EPA have similar distributions, but when compared to real data they exhibit lower values. Instead, the DSP show higher variance in comparison to HCP and a non-Gaussian behavior.

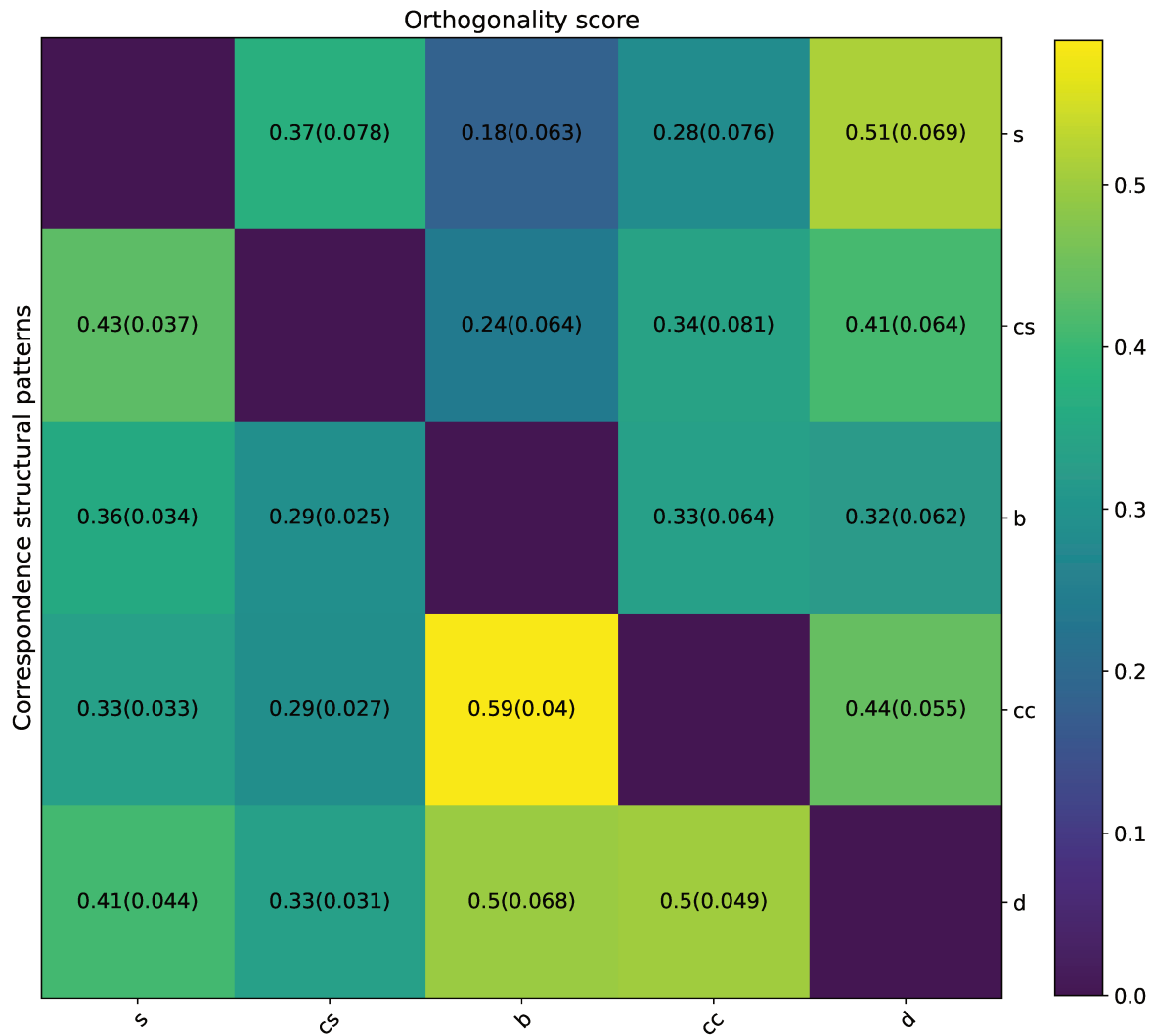


Figure 3.8 – Average value of orthogonality and correspondence structural pattern scores (SD) of nodal statistics pairs on HCP dataset. Upper Diagonal: Orthogonality score, Lower Diagonal: Correspondence structural patterns score. CC: clustering coefficient; B: betweenness centrality; D: degree; S: second-order centrality CS: closeness centrality.

Nodal Percentage of Participation.

Finally, we compare nodal participation in models and real HCP dataset (Fig. 3.10). As expected, the values of percentage of participation for EC and ECA brain models, are higher than real ones, due to the spatial relations that constrain the role of each brain region. Thus, same nodes play the same role in many samples of the generated datasets. On the contrary, DSP provides a lower bound in the percentage of participation of real data. Indeed, constraining graphs to only keep same degree sequence, increases the node role variability in the group.

A well-known brain property is the presence in the two hemispheres of homotopic regions: the right and left hemispheres are approximately mirror images of each other.

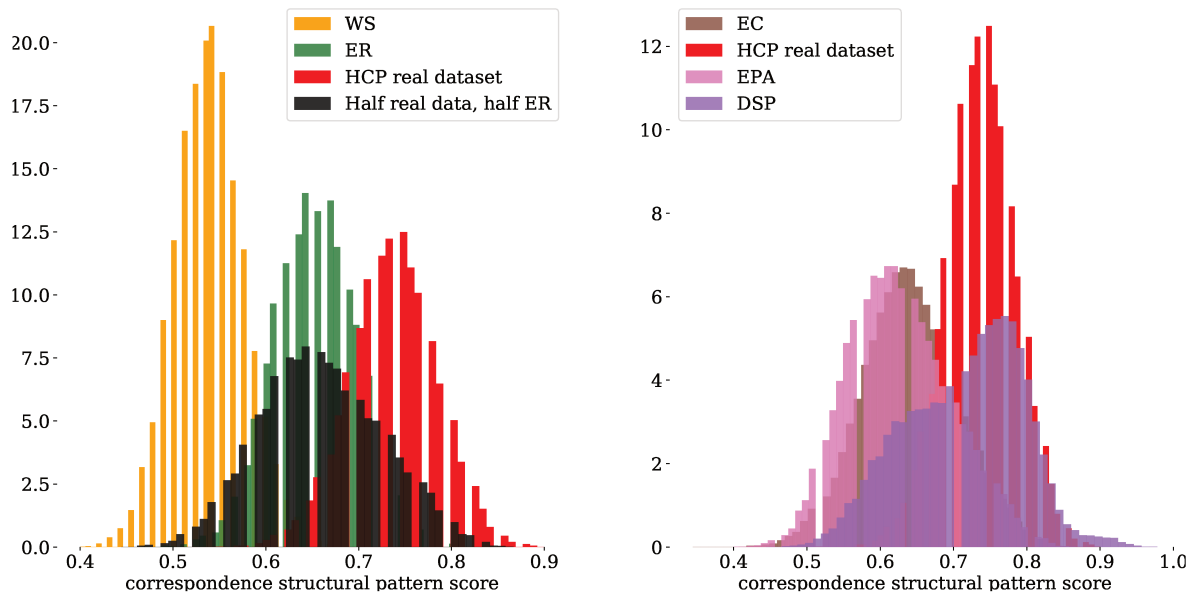


Figure 3.9 – Correspondence structural pattern score distribution on the considered model and real data. Left: WS: Watt-Strogatz model, ER: Erdős-Rényi model and HCP; dataset composed of 100 samples randomly chosen from HCP dataset and 100 from ER model. Right: Brain models and HCP data. DSP: Degree sequence preserving model; EPA: Economical preferential attachment model; EC: Economical clustering model.

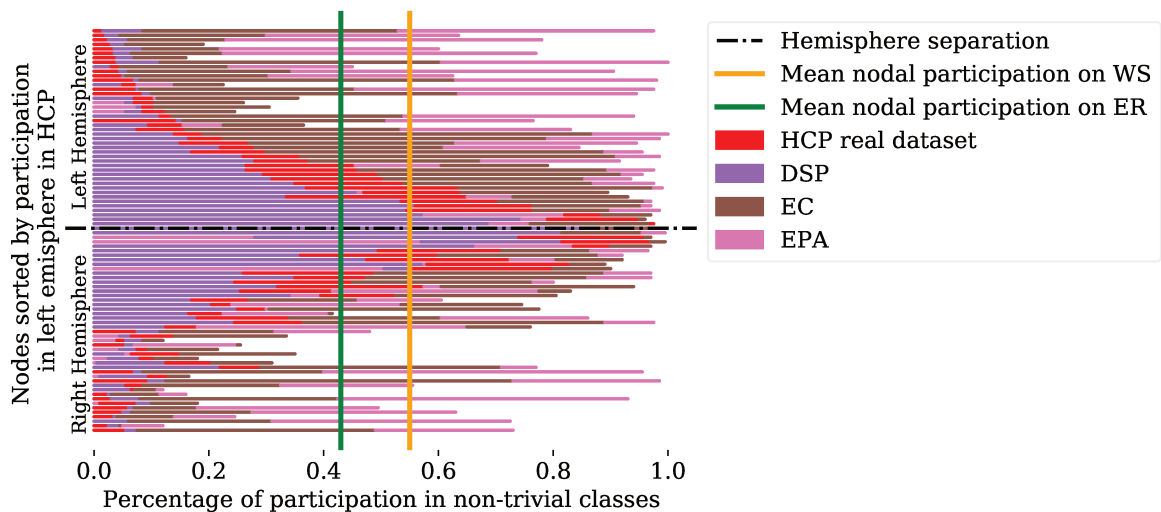


Figure 3.10 – Nodal participation in nontrivial classes for real (HCP), and synthetic datasets. ER: Erdős-Rényi model, WS: Watt-Strogatz model, DSP: Degree sequence preserving model; EPA: Economical preferential attachment model; EC: Economical clustering model.

That means the same region can be found in both hemispheres. An interesting result is the high number of observed classes to which belong both the homotopic regions (38% in average on HCP - Table 3.2). Again, when analyzing the entire data, we find that the nodal participation of brain regions is symmetrical: pairs of homotopic

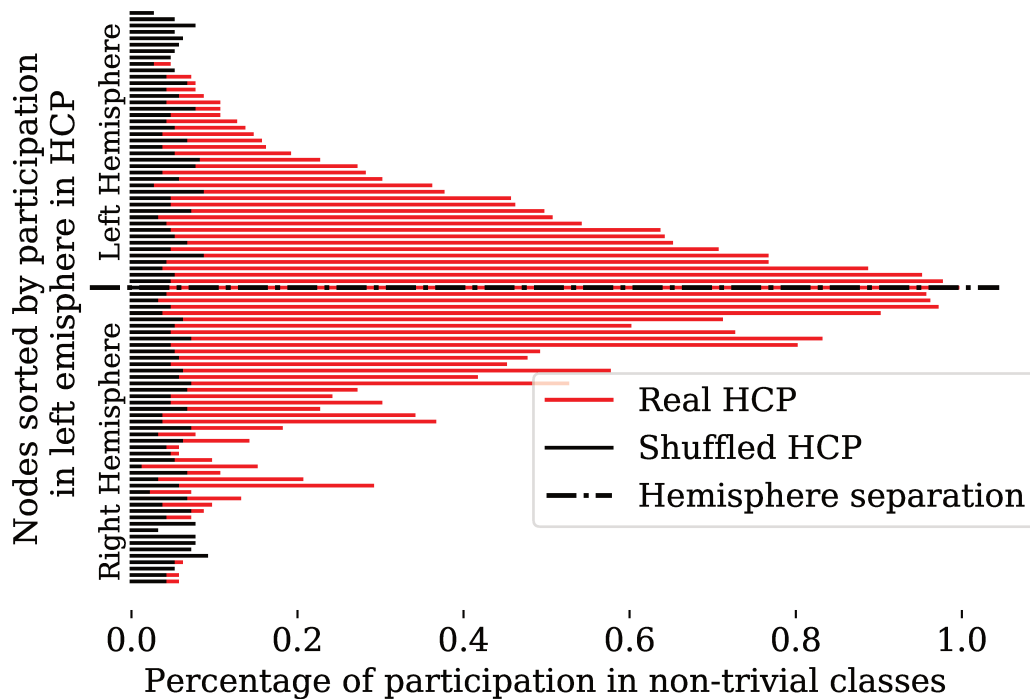


Figure 3.11 – Nodal participation in nontrivial class in HCP dataset and a shuffled version. Nodes labels are sorted according to the percentage of participation of left hemisphere regions. The symmetry reveals the expected hemisphere similarity in the participation of analogue regions. The percentage of participation of each node is also compared with a shuffled HCP dataset, where each real network is re-ordered by a random shuffle of the adjacency matrix, preserving the degree distribution. In this way, we expect that non-zero percentage of participation is simply due to chance. The participation indices for this random dataset appear to be significantly lower than the ones observed in the real HCP. However, even if closer to 0, all nodes appear to participate at least in one nontrivial class. Thus, when, for the real data, we observe a high participation index, we can conclude that the node is likely to share its equivalence role with some other nodes in the graph. At the same time, when a node does not have a positive percentage of participation, we expect the node to be unique, consistently in all networks and so to retrieve regions associated with unique functions.

regions have similar percentage of participation in nontrivial classes (Fig. 3.10). This property is still present in the brain models that integrate the brain geometry in their construction, such as DSP (Fig. 3.10). The percentage of participation of each node is also compared with a shuffled HCP dataset, where each real network is re-ordered by a random shuffle of the adjacency matrix, preserving the degree distribution. In this way, we expect that non-zero percentage of participation is simply due to chance. The results in Fig. 3.11 reveals how the NPP distribution on real data follows a very distinct behavior with respect its shuffled versions.

Table 3.2: Ratio of nodes in nontrivial classes in the same class of their homotopical in HCP dataset.

AVG	SD	MIN	MAX
0.38	0.11	0.13	0.67

3.3 (2) Healthy Population Characterization

In this section, we aim at providing the characterization of the healthy group both at global and local level. Indeed, there exists an inter-individual variability under same healthy conditions on the estimated network inference and network estimated on the same subject at different moment may result in different estimation. Here, we explore the ability of our proposed metrics to generalize to different datasets when healthy conditions are met. Particularly, we select healthy subjects coming from different databases and compare them in terms of orthogonality score and nodal percentage of participation.

3.3.1 Material

We gather Healthy Controls (HC) from four databases HCP, iShare, Gin-Chuga and ChuStr. The HCP database provides a total of 200 networks, each corresponding to 100 subjects who were scanned on different occasions. The age range of this population spans from 20 to 43 years. The networks in this dataset are computed based on the longest acquisition time among our databases of interest, amounting to a total of 1200 time-points in the time-series. The iShare dataset corresponds to the largest dataset in our collection, comprising 1814 scans of subjects aged between 18 to 35 years. It is important to note that, unlike the other scans where subjects lay down with open eyes (or fixing a cross displayed on a screen), the iShare acquisitions were obtained with subjects' eyes closed. Finally, we consider 20 HC subjects, aged 25 to 45, collected by ChuStr provider and 11 subjects of the higher age range (46-70) from the Gin-Chuga dataset.

3.3.2 Results and Discussion

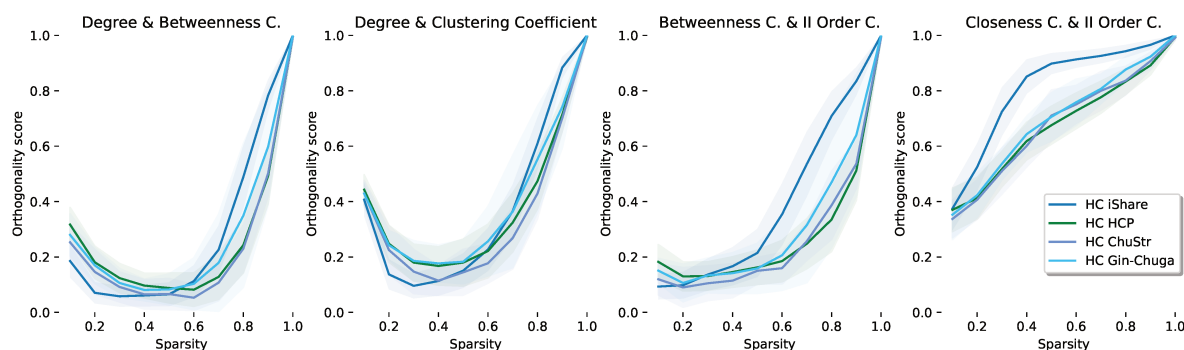


Figure 3.12 – Average orthogonality score curves of pairs of nodal statistics on healthy controls (HC) over sparsity. Networks are computed using AAL90 atlas. C.: Centrality.

Orthogonality Score.

We evaluate the mean of the orthogonality values over all samples of adjacency matrices at different sparsity values for HC obtained using AAL90. We remind that the orthogonality can be interpreted as a measure of redundancy and independence between different nodal statistics over a specific network. Specifically, across brain network analyses it can provide information on the underlying structure of the graph and justifying which structural properties are the predominant ones.

Differences are observed, particularly within the context of the iShare dataset. A distinct iShare mean orthogonality curve is reported with respect to closeness centrality and II order centrality, which exhibit notably elevated orthogonality coefficients in relation to other pairs of statistical measures (Fig. 3.12). Indeed, these two nodal statistics show a negative correlation on the extracted iShare graphs and their associated structural pattern partitions appear to highly overlap, resulting in high orthogonality score. Healthy brain connectivity networks extracted from iShare do not display nodes that serve as alternative bridge connections linking different modules of the network. An example of node which provide alternative bridge connections is illustrated in the visualization of second-order centrality in Fig. 1.6). This dissimilar behavior could be originated by the difference in scanning conditions, particularly eye conditions at scan. Previous research has indicated the presence of differences in functional connectivity, particularly in visual and auditory networks (Agcaoglu et al. 2019; Patriat et al. 2013), that might arise due to these variations. These discrepancies could potentially lead to distinct estimations of the network’s modularity structure, ultimately reflected in the structural patterns highlighted by the second-order centrality. However, further investigations into the impact of eye conditions on the structural patterns are necessary.

To statistically validate the presence of different orthogonality curve, we employed ANOVA test to the orthogonality values across dataset at different sparsity. The results demonstrated significant differences ($p < 0.05$) among the considered groups for closeness centrality and second-order centrality.

No statistical difference is reported for the other nodal statistics pair. To assess this HC consistency results, we built a classifier to perform a multi-class discrimination by detecting the provenance of the datasets. As it happens in conventional machine learning practices in the field of network neuroscience, we interpret the orthogonality score as global graph-descriptor and adopt it as feature of a K-Means algorithm (Richiardi et al. 2013).

To evaluate the performance of the classifier, we apply a 10-fold stratified cross-validation approach. In this approach, the entire set of available subjects was divided into 10 groups of equal size, while maintaining the proportion of each class within each provider. During each run of the experiment, one group is used as the test set, while the remaining groups are used as the training set. The evaluation results are

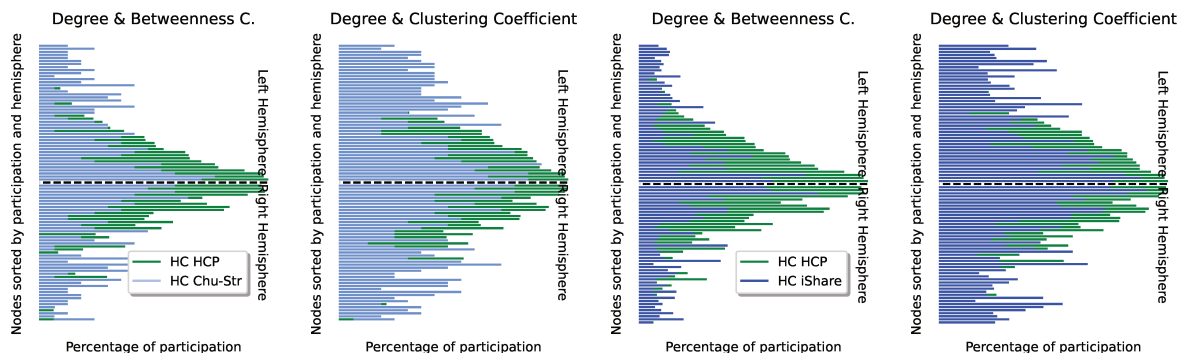


Figure 3.13 – Nodal percentage of participation for healthy controls datasets. C.: Centrality. Left hemisphere nodes are sorted by participation from the lowest to the highest values in HC HCP. Right hemispheres nodes are sorted in order to occupy symmetrical positions of the corresponding nodes in the left hemisphere with respect to the dotted line.

reported in terms of mean accuracy. The K-Means algorithm was set to identify 4 clusters, corresponding to the number of different data providers. It results that it is not possible to distinguish the provenance of data in a multi-centre HC dataset (Tab. 3.3), while instead it is eventually possible to discriminate the pathology (Section 3.4). This demonstrates their possible pooling without the need of data harmonization beforehand.

Table 3.3: Mean Accuracy on center prediction performed on the healthy control datasets.

	Degree	Betweenness C.	Clustering Coeff.	Closeness C.
Betweenness C.	0.23			
Clustering Coeff.	0.48	0.42		
Closeness C.	0.50	0.32	0.31	
II Order C.	0.36	0.42	0.17	0.28

Nodal Percentage of Participation for Regional Characterization.

Finally, we compare the nodal percentage of participation in nontrivial classes (NPP) in HC dataset in order to obtain a regional characterization of the healthy groups of comparable age distribution. The NPP shown in Fig. 3.13 are computed for a sparsity equals to 0.1. Even if the regional values are slightly different, the ranking in the percentages of participation in HC groups are largely overlapping. Indeed, when fitting the NPP of HC of different groups, we observe high correlations score, especially for degree and betweenness centrality: HCP - iShare 0.8, HCP - Chu-Str 0.7, iShare - Chu-Str 0.7 (Tab. 3.4). Interestingly, the symmetry left-right hemispheres is well captured by all the chosen pairs of nodal statistics.

Table 3.4: Regression parameters between nodal participation of different HC groups for different pairs of statistics.

	Degree - betweenness C.		Degree - Clustering Coefficient	
	R-value	Standard Error	R-value	Standard Error
HCP - iShare	0.8	0.036	0.7	0.047
HCP - Chu-Str	0.7	0.044	0.5	0.050
iShare - Chu-Str	0.7	0.071	0.6	0.094

Discussion

We show that healthy control populations, with different sizes and acquired in different centers, provide similar orthogonality curves; the minimum of the curve is always reached at the same sparsity value, except for iShare. This exhibited difference is probably due to the use of a different preprocessing chain and a change in the absolute values of the estimated correlations. Indeed, iShare correlations matrices have lower values leading to reach the minimum orthogonality values for lower sparsity. Particularly, in iShare dataset, selecting higher levels of sparsity determines the estimation of spurious edges since they are associated with very low correlation. Additionally, the convexity of the orthogonality curves can be used to evaluate the complexity of the graphs structure at different sparsity levels. Indeed, as the sparsity increases to 1, the graphs become much more close to a complete graph in which all edges are observed. In this case, all nodes are equivalent, independently of the considered pair of nodal statistics. On the contrary, for very low sparsity levels, different pairs of nodal statistics capture the same graph information and show no-minimal orthogonality scores. When the orthogonality curve reaches the minimum, the pair of statistics is as close as possible - for the considered statistics set - to the perfect orthogonality on the given graph structure. At this sparsity level, the combination of statistics places many nodes in trivial class, capturing node specific behaviors and revealing the presence of particular structural properties on the extracted graph.

3.4 (3) Discrimination of Patients Population

In this section, our focus is on demonstrating the advantages of our proposal within the context of pathological discrimination and characterization. Specifically, we begin by applying our proposal to comatose patients, driven by the insights gained from previous findings related to nodal re-organization, which served as a foundation for establishing the structural pattern.

Subsequently, we validate that a comparison utilizing the scores derived from the structural patterns effectively distinguishes between comatose patients and healthy individuals. Furthermore, we utilize the nodal participation percentage to detect regional perturbations of the patients group versus the healthy controls.

As a second application, we extend a similar framework to Parkinsonian patients. This serves a dual purpose: to classify recently diagnosed patients from the healthy control

group and to identify subtle distinctions among patients.

We conduct a comparison with existing state-of-the-art classification techniques. Finally, we present a synthesis of our observations from a neuroscientific interpretation perspective.

3.4.1 Comatose Patients

Based on a communication written in collaboration with Michel Dojat and Sophie Achard accepted for XXIXème Colloque Francophone de Traitement du Signal et des Images - GRETSI 2023 (Carboni et al. 2023c)

To demonstrate the advantages of our proposed framework in real-world data scenarios, we begin by examining the functional networks of individuals in comatose state. Data are provided by Chu-Str and counts 20 HC and 17 comatose patients (CO). Among the comatose patients, twelve experienced cardiac and respiratory arrest, two patients experienced hypoglycemia, two others fell into a coma after a gaseous embolism incident, and one patient had extracranial artery dissection.

The task of distinguishing between subjects with differing levels of consciousness through the analysis of functional networks poses a multifaceted challenge that is not readily apparent. Previous studies have encountered difficulties in detecting significant changes in graph metrics when contrasting comatose patients with their healthy counterparts. Nonetheless, as previously mentioned, it has been observed that disruptions in hub connectivity yield discriminatory outcomes in comatose patient classification (Achard et al. 2012).

Orthogonality coefficient

Our analysis commences with an assessment of the variations in orthogonality scores across diverse levels of sparsity. Demonstrating the efficacy of orthogonality as a comprehensive global index of the underlying graph structure topology, we unveil its capability to effectively differentiate between the healthy controls group and patients. In contrast, the utilization of the corresponding standard metric fails to provide a distinct discriminatory insight. This observation underscores the superiority of orthogonality as a discerning feature, enabling clearer discrimination and deeper understanding of the underlying graph structure differences between these two distinct groups. In Fig. 3.14, we show the difference in employing the orthogonality coefficient of pair of nodal statistics with respect to the use of the corresponding global average in detecting differences across the groups.

Drawing inspiration from conventional machine learning practices in the field of network neuroscience, we interpret the orthogonality score as global graph-descriptor and adopt it as feature of K-Means algorithm (Richiardi et al. 2013). This integration allows us to effectively cluster orthogonality curves that share similarities, ultimately enabling the grouping of coherent patterns. Moreover, this approach empowers us to predict the

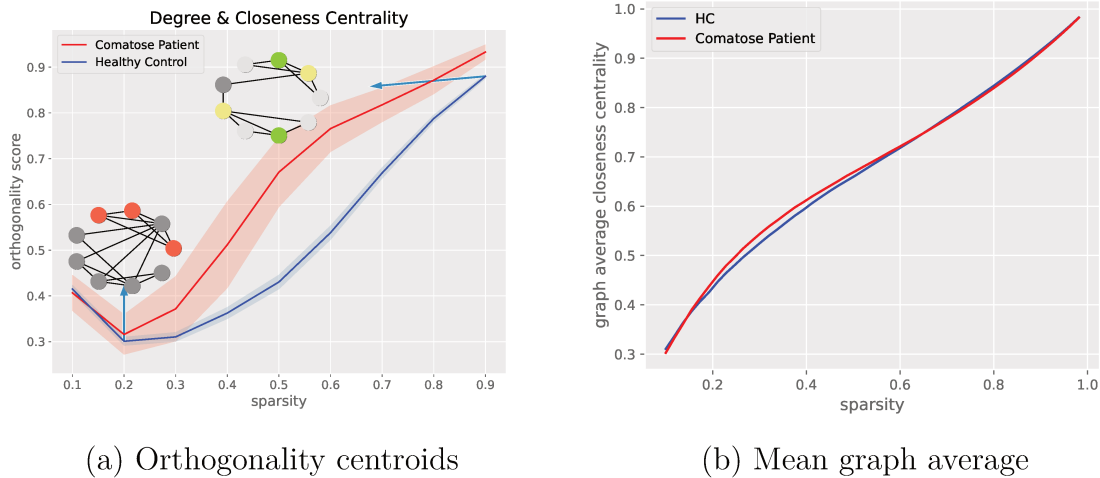


Figure 3.14 – (a) Visualization of the centroid of the orthogonality curves obtained with K-Means clustering. Left: CO centroid and HC centroid obtained considering degree and closeness centrality. Examples of graph with a given orthogonality values of the considered pair of statistics are also shown with node colors correspond to different classes. Grey is used for node in trivial classes. (b) Visualization of the mean graph average closeness centrality at different sparsity levels per group.

classes of new samples based on the insights garnered from the clustered orthogonality curves.

We perform a binary discrimination task, differentiating between patients and a control group by the use of the orthogonality coefficient of pair of nodal statistics. This facilitates the interpretation in terms of the predominant substructures within graph topology.

As depicted in Fig. 3.15, it becomes evident that the combination of closeness centrality and either degree or II order centrality can be effectively employed to distinguish comatose patients from the HC group. These results suggest that, for Comatose patients, connections among different groups of high-interconnected nodes in the graph structure are provided by nodes having a high degree, while in Healthy Controls, links between these groups, which do not involve high connected nodes, may exist (Fig. 3.14).

Correspondence Structural Pattern

We determine the distribution of the correspondence structural pattern score associated with different nodal statistics in three different groups: a group of only healthy controls, a group including all the comatose patients and healthy controls, and a group including only the comatose patients.

Subsequently, we employ a Support Vector Machine (SVM), utilizing the correspondence structural pattern score as a kernel function to effectively discriminate between functional networks of comatose patients and those of healthy controls. Fig. 3.16,

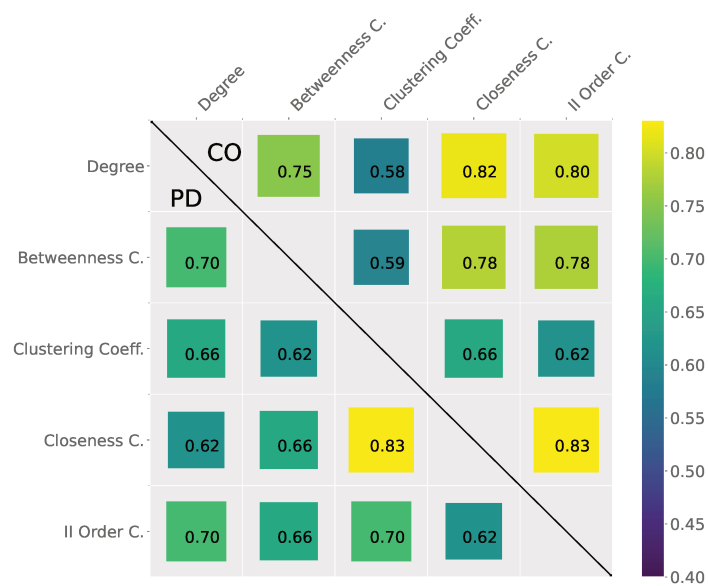


Figure 3.15 – Accuracy results in patients classification 7-fold cross validation over the orthogonality curves per pairs of nodal statistics. Upper Matrix: Comatose patient discrimination. Lower Matrix: Parkinsonian patient discrimination. C.: centrality, Clust. Coeff.: Clustering Coefficient.

displays the distributions of the structural pattern scores within these three groups of interest. Results are shown for the equivalence relation associated with the degree and the clustering coefficient in combination or as single statistics. For the three equivalence relations, a discernible separation is observed between the distributions of healthy controls and comatose patients; with the former group consistently exhibiting a higher correspondence score. This distinction is statistically significant and has been validated through the Z-test with a p-value of less than 0.001 for all the three cases. Additionally, significant difference was also observed in the distributions of the entire dataset encompassing both comatose patients and healthy controls (CO+HC) and the group of solely healthy controls (HC).

Interestingly, while a significant difference exists between the CO+HC and CO groups when considering the structural pattern associated with other nodal statistics combinations, this is not the case with the clustering coefficient-associated structural pattern. In this case, the correspondence scores between comatose subjects are similar to the values obtained when they are pooled with HC, indicating a notable heterogeneity within the comatose subgroup when considering the clustering coefficient.

Moreover, we can appreciate the benefit of combining together more nodal statistics for the identification of finer structural patterns. This combinations results in structural pattern scores vary from 0.2 to 0.5 when considering only the degree and 0.4 to 0.8 when degree and clustering coefficient are considered simultaneously. Remarkably, even with this expanded range, the discernible and statistically significant differentiation between comatose patients and healthy controls remains consistently preserved.

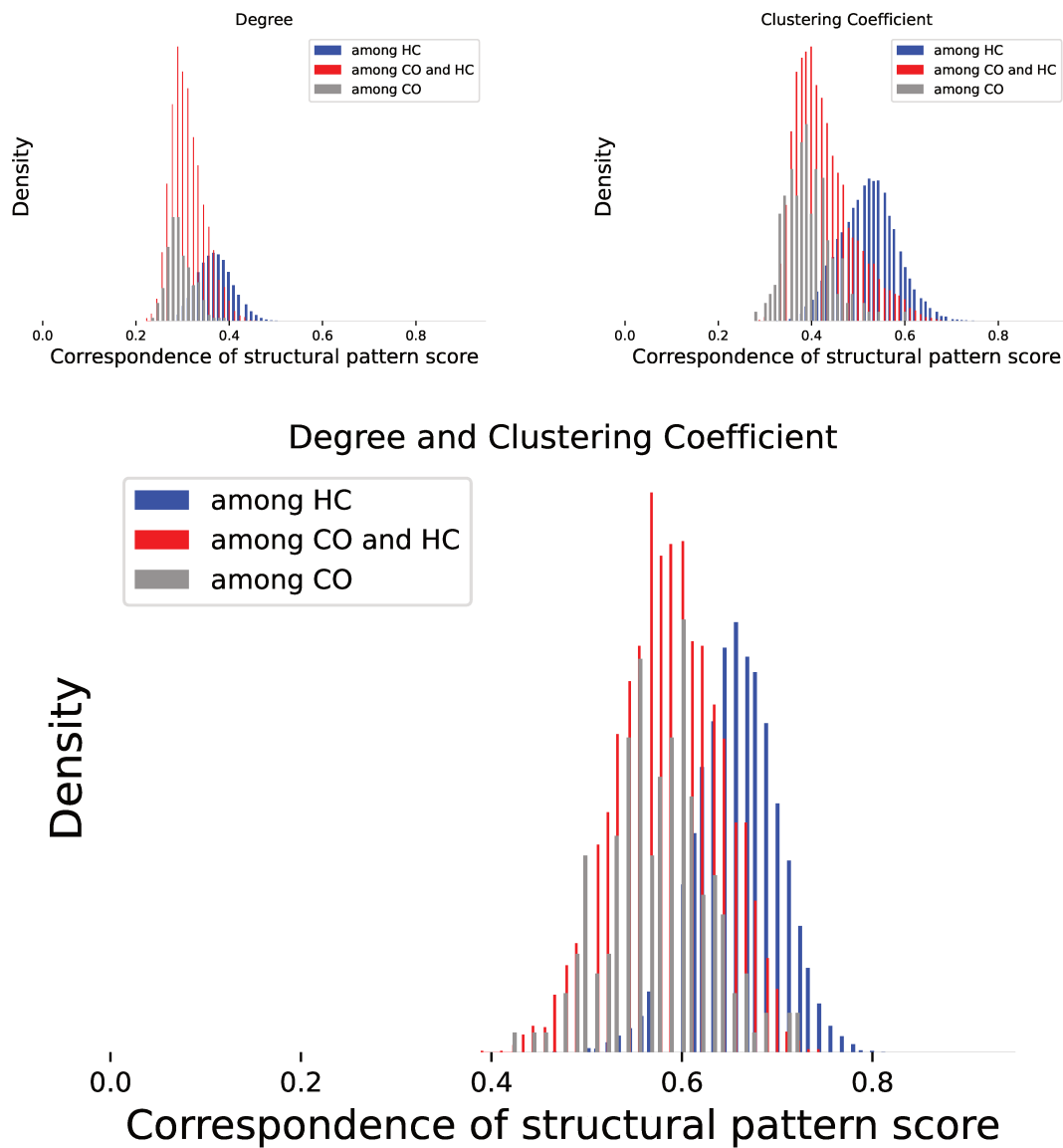


Figure 3.16 – Distribution of correspondence structural patterns score in the considered three groups of subjects.

Finally, Fig. 3.17 presents a graphical representation of the outcomes of the discrimination between comatose patients and healthy controls when considering the structural pattern associated with degree, clustering coefficient and their combinations. The discrimination reaches a perfect score in a 5-cross-validation procedure.

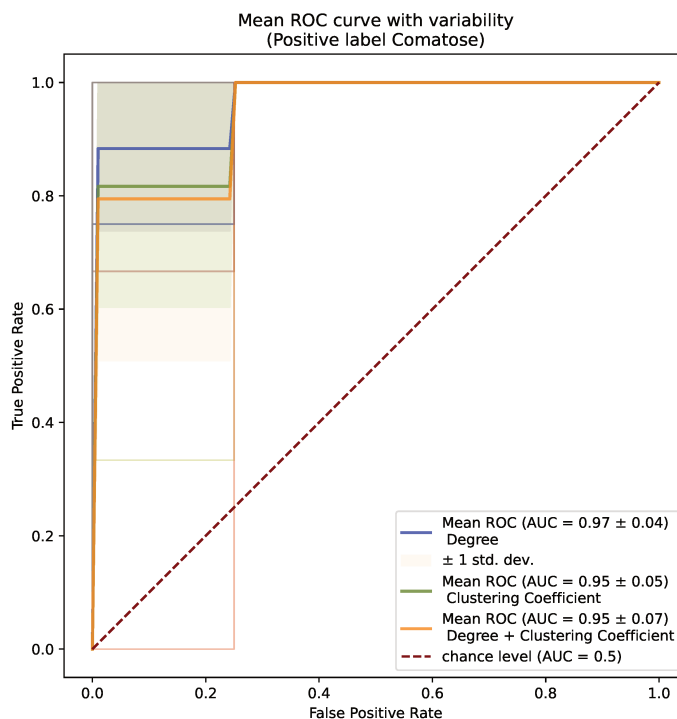


Figure 3.17 – ROC curve of discrimination between comatose network and healthy controls in a 5 cross-validation considering structural pattern associated with degree, clustering coefficient and their combinations.

Nodal Percentage of Participation

Concerning the regional characterization of the patients group, we compare node percentage of participation scores of comatose patients with healthy controls in Fig. 3.18.

When juxtaposed with the Human Connectome Project (HCP) scores, it becomes evident that the percentage participation in comatose patients exhibits reduced variability. In the context of comatose patients, nearly every node tends to cluster around the mean percentage of 0.26 ± 0.17 , whereas the HCP dataset displays a wider distribution with a mean percentage of 0.32 ± 0.30 (See Tab. 3.5). This tighter clustering of nodes around the mean in comatose graphs makes it more challenging to discern a distinct hierarchy in node behavior within this context as it would happen in random dataset (Fig. 3.19).

Table 3.5: Statistics on nodal percentage of participation.

	AVG	SD	MIN	MAX
HCP	0.32	0.30	0.015	0.98
Comatose	0.26	0.17	0.0	0.76

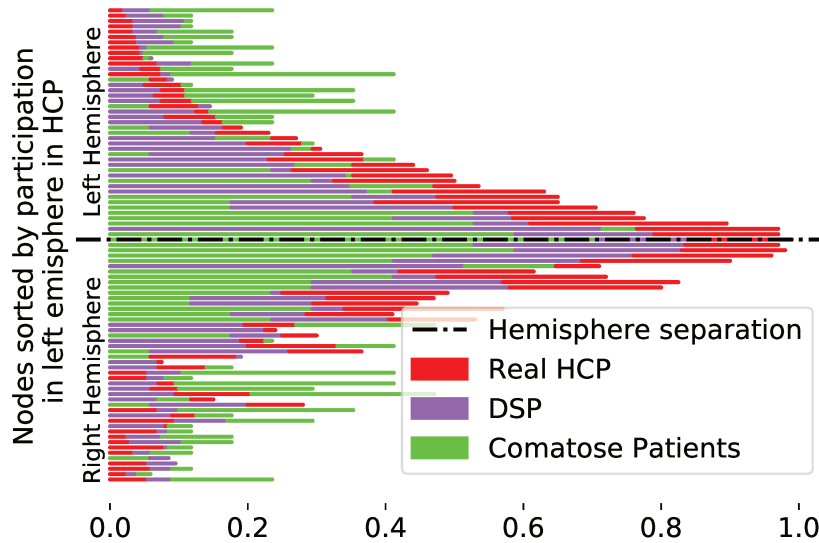


Figure 3.18 – Nodal participation in nontrivial classes associated with degree and betweenness centrality for comatose patients, healthy controls (HCP dataset) and DSP model. Left hemisphere nodes are sorted by participation in HCP, right hemisphere nodes occupy symmetrical positions of their corresponding left hemisphere nodes with respect to the dotted line.

As shown in Fig. 3.20, the NPP of each region differs from HC and comatose patients (correlation of NPP for degree and closeness centrality: 0.2, for closeness centrality and II order centrality: -0.1). In all cases, the maximum nodal comatose participation value is lower than the corresponding value in HC.

Since the low number of NPP in the patient group, can be due either to the presence of many trivial classes in patient networks, or to the fact that the nodes in nontrivial classes are not consistently be the same in the group, we evaluate the number of nodes in nontrivial classes per graph, and we report comparable number in controls and patients (Tab. 3.6). Hence, the difference in the percentage of participation indicates the presence of higher structural patterns variability in patients.

Table 3.6: Nodes in nontrivial class per graph

	AVG	SD	MIN	MAX
HCP	0.32	0.062	0.17	0.52
Comatose	0.26	0.067	0.13	0.41

Finally, we rank the regions according to their perturbation with respect to the participation in healthy controls. This allows to determine which brain regions have higher change in their structural between the healthy control group and the patient one. For Parietal Inferior, Angular and Fusiform gyri (already hubs in Achard et al. 2012) and Hippocampus appears to be particularly impacted.

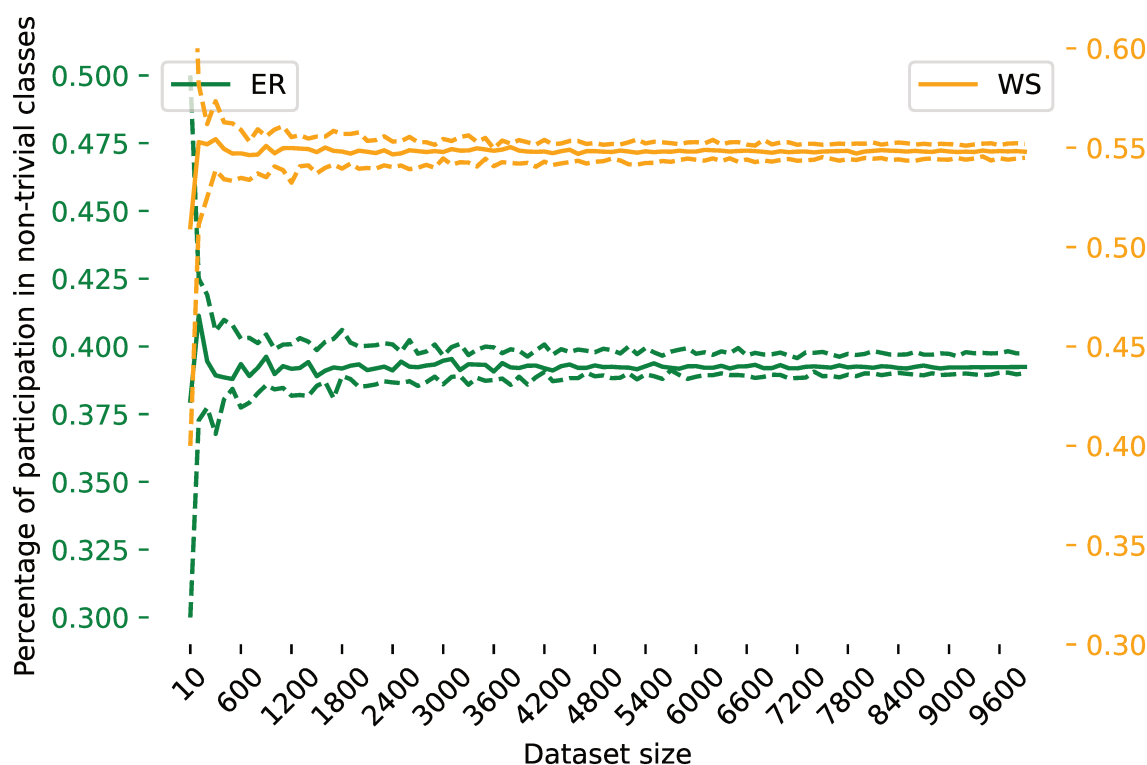


Figure 3.19 – Mean participation of a node in datasets of different size. Mean percentage of participation in nontrivial classes for a single node in 90 nodes ER (Erdős-Rényi) and WS (Watt-Strogatz) model at 0.1 sparsity level. The results indicate that the mean percentage of participation stabilizes respectively at 0.43 and 0.55 for ER and WS models. WS is obtained with 0.5 edge rewiring probability. Dots lines are first and third quartile.

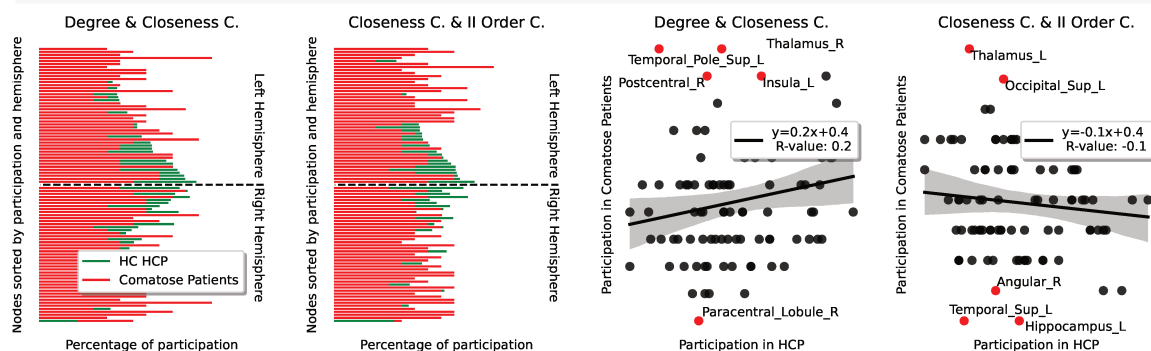


Figure 3.20 – Nodal percentage of participation in HC and CO for degree and closeness centrality.

3.4.2 Parkinson Disease

Parkinson's disease (PD) is a neurodegenerative disorder affecting the aging population. It manifests with a wide range of non-motor symptoms, such as hallucinations, apathy, depression etc. (Jankovic 2003; Maciéas-García et al. 2022; Moro et al. 2020). The study of functional connectivity networks in PD has shown promising results for

early detection (Long et al. 2012; Muñoz-Ramírez et al. 2019) and abnormalities investigation in diverse symptomatology (Gao et al. 2016; Marques et al. 2022; Tessitore et al. 2019). In this section, we consider three different PD patients reflecting the first stages progression of the disease: *de novo* patients (H&Y score ≤ 2 , stage I-II (Bhidayasiri et al. 2012)) with normal vision, patients ($2 < \text{H\&Y} \leq 3$, stage III) with visual illusions or with visual hallucinations. In the quest to differentiate these PD patients from healthy controls (HC, age-matched), we demonstrate the existence of nodal re-organization in PD by applying a classifier based on the similarity of the graph structural patterns. Moreover, our classifier can differentiate between PD and HC groups and between patients experiencing visual illusions and visual hallucinations. Finally, we can detect brain regions with abnormal behavior in the different groups, giving a local characterization of the disease symptoms. Our classifier has comparable performances with other current methods, with the added value of providing interpretation and explanation of the exhibited differences among groups both at the global and regional levels.

Orthogonality coefficient

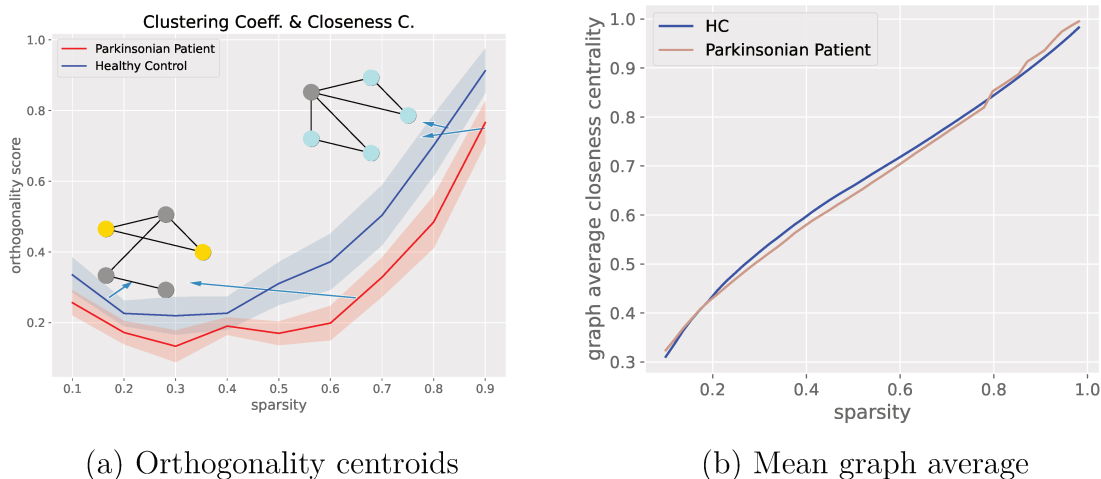


Figure 3.21 – (a) Visualization of the centroid of the orthogonality curves obtained with K-Means clustering. Left: CO centroid and PD centroid obtained considering clustering coefficient and closeness centrality. Examples of graph with a given orthogonality values of the considered pair of statistics are also shown with node colors correspond to different classes. Grey is used for node in trivial classes. (b) Visualization of the mean graph average closeness centrality at different sparsity levels per group.

In our initial application concerning Parkinson’s disease, we replicate the orthogonality curve classification experiments as outlined in previous section. We systematically investigate the orthogonality coefficient’s behavior across different sparsity levels and nodal statistics pairs. Our focus centers on assessing the discriminative capability of each pairwise combination of nodal statistics in the context of distinguishing between

de novo PD patients and a control group of healthy individuals within the comparable age range, as provided by the Gin-Chuga database.

Throughout this investigation, the higher discrimination power emerges when combining between closeness centrality and clustering coefficient (Fig. 3.15). When investigating the different group orthogonality curves (reported in Fig. 3.21) a mean orthogonality is observed in Parkinsonian patients in comparison to the healthy control group. This indicates a discernible distinction in the distribution of triplets of fully connected nodes between the patient and control networks and nodes occupying central position. These findings provide compelling evidence of potential network-level variations that may be indicative of the underlying physiological differences between the two groups.

Correspondence of structural pattern score

As a second step in PD exploration, we evaluate the possibility of comparing networks by using the correspondence structural pattern score. This evaluation spans a range of sparsity levels, ranging from 0.1 to 0.9, with incremental steps of 0.1. These sparsity levels are associated with the considered nodal statistics and their pair combinations. Our approach involves the training of a SVM classifier, where the correspondence structural pattern score serves as a pivotal input.*¹.

To benchmark, our classification results are compared with a selection of methods that have been used in the literature in a classification task of functional brain connectivity networks. We select some representative methods with the constraint of relying on a prefixed-atlas correlation estimation and being applied in disease classification tasks. A first class of approaches directly uses correlation matrices or second-order matrices (i.e. tangent Pearson) as input of an SVM or other types of classifier (Rish et al. 2009; Heinzle et al. 2012; Dadi et al. 2019; Varoquaux et al. 2010a; Schirmer et al. 2021). We implement this first class (M1) by selecting the best classifier in accuracy among (SVMs, Random forest, Ridge classifier, K-nearest neighbor and naive Bayes classifiers) fed by the extracted matrices. As a second class of approaches (M2), we consider a graph-based method relying on the normalized Laplacian spectrum computation. The eigenvalues of each graph are used as input-features of the SVM classifier. Similar methods have been proposed to detect network patient alterations (Haan et al. 2012; Mostafa et al. 2019; Mheich et al. 2020; Schirmer et al. 2021). Next, we select the HyperConnectome AutoEncoder HCAE (M3) as representative of graph neural network (GNN)-based approaches (Banka et al. 2020). (M3) learns embedding features from a brain graph (one graph per subject) and has been applied for brain state classification (Bessadok et al. 2022b). Finally, we consider a method (M4) that has been proposed for the classification of early PD patients. We implement a modified version of the algorithm proposed by Long et al. 2012 which only relies on functional connectivity

¹The code is made publicly available at <https://gricad-gitlab.univ-grenoble-alpes.fr/carbonil/regional-differentiation-based-on-graph-nodal-statistics-for-functional-brain-connectivity-networks-characterization/>

correlation matrices, instead of requiring multimodal information. Classification results are reported in Tab. 3.7, we recall the definition of sensitivity and specificity.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3.1)$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (3.2)$$

In our method, we indicate the best combination of sparsity level and nodal statistics (NS).

It is worth noticing that the best performances of classification for PD vs HC are obtained at low sparsity range, which corresponds to the small-word network regime.

Table 3.7: Classification results for repeated stratified fold cross validation maximizing the global accuracy. K has been fixed to 8 for having at least three test samples and results are averaged over 100 runs (SD). Sens: sensitivity, Spec: specificity, Acc: accuracy. SD: Standard deviation. Sp: Sparsity, NS: Nodal statistics, CL Coeff: Clustering Coefficient, C: centrality, Bet: Betweenness, Cl: Closeness. In our method, we indicate the best combination of sparsity level and nodal statistics (NS).

Task		Method					Sp.	NS
		M1 (SD)	M2 (SD)	M3 (SD)	M4 (SD)	Ours (SD)		
HC vs PD- <i>de novo</i>	Spec	0.42 (0.44)	0.40 (0.44)	0.35 (0.48)	0.47 (0.45)	0.63 (0.43)	0.3	CL. Coeff. (A)
	Sens	0.86 (0.28)	0.85 (0.30)	0.64 (0.48)	0.73 (0.36)	0.95 (0.15)		
	Acc	0.65 (0.25)	0.63 (0.25)	0.42 (0.15)	0.61 (0.27)	0.79 (0.21)		
HC vs (PD-H,PD-I)	Spec	0.19 (0.35)	0.49 (0.44)	0.0011 (0.030)	0.19 (0.36)	0.19 (0.34)	0.3	Bet. C.+ Cl. C. (B1)
	Sens	0.81 (0.25)	0.75 (0.25)	0.99 (0.051)	0.72 (0.28)	0.94 (0.15)		
	Acc	0.61 (0.20)	0.67 (0.21)	0.66 (0.11)	0.54 (0.23)	0.67 (0.17)		
PD-I vs PD-H	Spec	0.67 (0.35)	0.66 (0.33)	0.41 (0.37)	0.69 (0.36)	0.44 (0.38)	0.1	CL. Coeff. (C)
	Sens	0.59 (0.33)	0.24 (0.34)	0.69 (0.32)	0.51 (0.35)	0.85 (0.24)		
	Acc	0.63 (0.24)	0.61 (0.24)	0.56 (0.22)	0.59 (0.23)	0.66 (0.21)		

Our classification results highlight the presence of a nodal reorganization associated with different nodal statistics in the presence of Parkinson’s Disease and during its progression in early stages. The overall classification scores obtained for all the methods are not really high reflecting the difficulties to discriminate the patients at this stage of the pathology progression ($HY < 3$). It is worth mentioning that when the duration of scanning is longer (Grenoble), the classification scores are improved for all the methods (HC vs PD-*de novo*). Possibilities to improve accuracy results may come from the use of a multimodal approach taking into account changes in structural MRI (Long et al. 2012).

Nodal Percentage of Participation

Finally, to obtain an interpretable characterization at the nodal level of a the group of brain connectivity networks $\mathbf{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_m\}$ of different subjects, we propose to evaluate the persistence of distinctive node role in the group. Specifically, we identify nodes whose role is highly distinctive as the nodes in a trivial class: a class with no other node. On the other hand, a node whose role is shared with at least another node

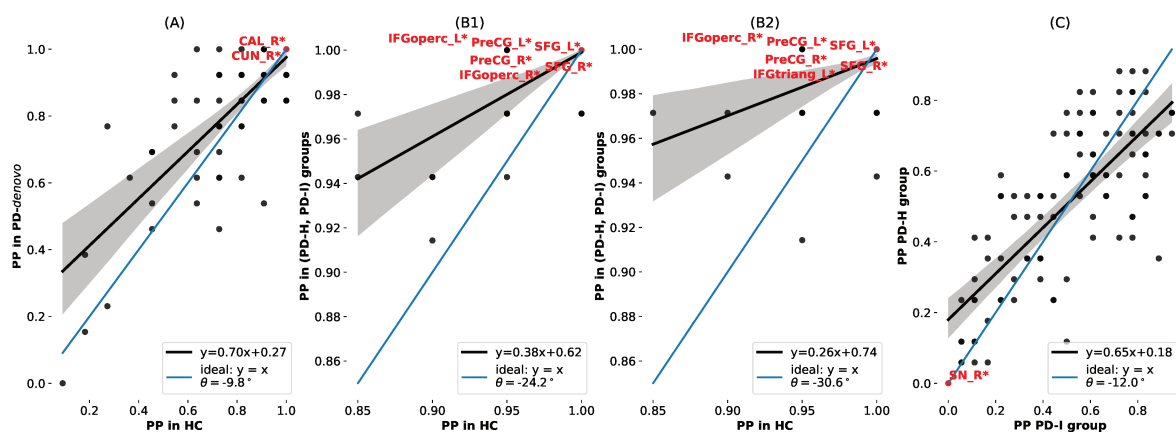


Figure 3.22 – Visualization of the nodal PP in the different subject groups. A selection of significant perturbed regions is colored in red. The blue line indicates the hypothetical equivalence of the populations under comparison. Brain regions are annotated according to [Rolls et al. 2020](#), description can be found in the text.

in the graph belongs to a nontrivial class. The node persistence to have a distinctive role in the group is computed as the nodal percentage of participation in trivial class (PP):

$$\forall v_i \in \mathcal{V} \quad \text{PP}_{\mathbf{G}}(v_i) := \frac{|\{\mathcal{G}_k \in \mathbf{G} \text{ s.t. } \#[v]_{\mathcal{G}_k} = 1\}|}{m} \quad (3.3)$$

with $m = |\mathbf{G}|$ the number of subjects in \mathbf{G} . The PP provides a summary of the structural patterns of a network group at the nodal (regional) level. Differences in the participation in each group can be explicitly computed, allowing to track any change in the regions between different groups. In an *ideal* case where two groups of subjects have the same structural patterns, we expect their PPs to be equal, meaning that each region equally shares its role in the structural pattern of the graphs in the two groups. Thus, the difference between groups can be estimated by linear regression model coefficients and using the residuals to rank the local perturbation given by each region v_i . Moreover, we quantify the nodal disruption by computing the angle between the ideal regression line having a slope coefficient of 1 and the estimated one as in ([Achard et al. 2012](#)). Note that the nodal PP is a complementary measure with respect to the NPP which quantifies the node persistence in **nontrivial** classes.

To determine which regions show significant differences in the PP between two groups, we perform a permutation test with Bonferroni correction, controlling for multiple comparisons across all regions. We prefer to avoid false discovery as we are able to find many disrupted regions. This analysis identifies the regions that are most significantly perturbed in their nodal PP between the two groups. And finally, we estimate the angle between the regressed line and the ideal one (the identity), as a measure of the global nodal disruption index ([Achard et al. 2006](#)). Results are reported in Figure 3.22, and a full list of significantly different regions can be found in Tab. 3.8 and Tab. 3.9.

The use of the nodal percentage of participation (PP), reveals altered subcortical and cortical regions in accordance with the PD pathology. By comparing *de novo* PD patients with HC, we detect 17 significantly altered regions among which the middle

Table 3.8: List of regions significantly different in HC vs PD *de novo*.

MFG_L	Frontal_Mid_2_L	MOG_R	Occipital_Mid_R
IFGoperc_R	Frontal_Inf_Oper_R	FFG_R	Fusiform_R
ROL_R	Rolandic_Oper_R	SPG_L	Parietal_Sup_L
SFGmedial_L	Frontal_Sup_Medial_L	SMG_L	SupraMarginal_L
MCC_L	Cingulate_Post_L	ANG_R	Angular_R
CAL_R	Calcarine_R	STG_R	Temporal_Sup_R
CUN_R	Cuneus_R	MTG_R	Temporal_Mid_R
SOG_L/R	Occipital_Sup_L/R	ITG_R	Temporal_Inf_R

Table 3.9: List of significantly different regions in HC vs (PD-I, PD-H).

PreCG_L/R	Precentral_L/R	CAL_L/R	Calcarine_L/R	SFG_L/R	Frontal_Sup_2_L/R
CUN_L/R	Cuneus_L/R	IFGoperc_R	Frontal_Inf_Oper_R	LING_L/R	Lingual_L/R
IFGtriang_L/R	Frontal_Inf_Tri_L/R	SOG_L/R	Occipital_Sup_L/R	IFGorb_L/R	Frontal_Inf_Orb_2_L/R
MOG_L/R	Occipital_Mid_L/R	ROL_R	Rolandic_Oper_R	IOG_L/R	Occipital_Inf_L/R
SMA_L/R	Supp_Motor_Area_L/R	FFG_L/R	Fusiform_L/R	OLF_L	Olfactory_L
PoCG_L/R	Postcentral_L/R	SFGmedial_R	Frontal_Sup_Medial_R	SPG_L/R	Parietal_Sup_L/R
PFCventmed_L/R	Frontal_Med_Orb_L/R	IPG_L/	Parietal_Inf_L	REC_L/R	Rectus_L/R
SMG_L	SupraMarginal_L	OFCmed_L/R	OFCmed_L/R	ACC_L/R	Cingulate_Ant_L/R
ANG_L/R	Angular_L/R	OFCant_L/R	OFCant_L/R	PCUN_L/R	Precuneus_L/R
OFCpost_R	OFCpost_R	PCS_R	Paracentral_Lobule_R	OFClat_L/R	OFClat_L/R
PUT_R	Putamen_R	INS_L/R	Insula_L/R	HES_L	Heschl_L
ACCmid_L	Cingulate_Mid_L/R	STG_L/R	Temporal_Sup_L/R	MCC_L/R	Cingulate_Post_L/R
TPOsup_L/R	Temporal_Pole_Sup_L/R	HIP_L	Hippocampus_L	MTG_L/R	Temporal_Mid_L/R
PHG_L/R	ParaHippocampal_L/R	TPOmid_L/R	Temporal_Pole_Mid_L/R	AMYG_R	Amygdala_R
ITG_L	Temporal_Inf_R	CERant_L/R	Cerebellum_Ant_L/R	THAL_L	Thalamus_L
N_Acc_R	Nucleus_Accumbens_R				

frontal gyrus and the posterior cingulate in the Default Mode Network, whose disruption is associated with cognitive decline, and calcarine cortex and cuneus which, in accordance with the literature, show progressive connectivity deterioration during the premotor phase (Tessitore et al. 2019; Muñoz-Ramírez et al. 2019). 11 regions are located in the right hemisphere and 6 in the left one, with only the occipital superior gyrus being affected in both hemispheres, suggesting an asymmetrical distribution of abnormal regions in *de novo* patients which can be related with the lateralisation of PD. When comparing HC with PD patients with visual symptoms, more regions are impacted probably due to the disease’s progression. We select two combinations of nodal statistics and sparsity that reach the same global discriminative power (B1) and (B2). We detect a significant PP difference in 86 regions for the first combination and 77 regions in the second. In accordance with the literature (Gao et al. 2016; Tessitore et al. 2019; Marques et al. 2022), we report a significant difference in the anterior cingulate cortex (ACC), the nucleus accumbens (N_ACC), the cerebellum anterior (CERant), thalamus (THAL), the amygdala (AMYG) and the Supplementary Motor Area (SMA). Interestingly, the detected perturbed regions largely overlap in the two combinations and greater disruption coefficients θ s are observed when comparing HC against PD patients with visual abnormalities compared to *de novo* PD patients (Fig. 3.22).

Our study contributes to the understanding of PD progression, detecting multiple regions associated with the disease at an early stage, a more advanced stage, and with

different visual symptoms.

3.5 (4) Reliability

The primary objective of this section is to provide recommendations for determining the minimum time-series length to conduct robust regional-level analyses in functional connectivity. To achieve this goal, we focus on healthy control (HC) subjects with extended time-series scans, specifically those containing 1000 or more data points coming from iShare and HCP. Our approach involves a subsampling of these lengthy time-series to extract shorter versions, each comprising n consecutive data points. For each of these shortened time-series, we proceed to estimate a new connectivity matrix.

Moreover, we will consider an *attacked* version: where a predetermined number of randomly selected edges from each real network undergo re-wiring. The re-wiring is considered valid only if the obtained network is still completely connected.

This process allows us to compare the nodal statistics at the local level across these diverse estimations. By analyzing the variations in nodal statistics resulting from different time-series lengths, we gain insights into the impact of the number of acquired volumes on functional connectivity patterns. Ultimately, this investigation aims to suggest the appropriate time-series length required to ensure the robustness and reliability of regional-level functional connectivity analyses.

3.5.1 Material and Method

To assess the difference across different estimations, differing in time-length, we evaluate Intra-Class Correlation Coefficient (ICC) and Bland-Altman plot over the nodal values and their group average (Koo et al. 2016; Altman et al. 1983; Giavarina 2015; McGraw et al. 1996).

These analyses quantify the reliability and agreement between measurements derived from network estimated with different time-series lengths.

The former, it has been defined as a different form of Pearson correlation coefficient and it is used to evaluate the reliability of different raters or test-retest consistency. In our evaluation, as suggested by McGraw et al. 1996; Koo et al. 2016, we consider the ICC defined for a two-way mixed-effects models for absolute agreement, corresponding to ICC(A,1) following the notation by McGraw et al. 1996. In fact, we assume in our model that the choice of a set of time-series lengths determines a *raters population* whose reliability needs to be assessed. Reporting the result of measurements into a table where each column corresponds to a measurement obtained with a different points number in the time-series, and each row corresponds to a distinct brain region subject to measurement evaluation, the ICC(A,1) is given by the following equation:

$$\text{ICC}(A, 1) = \frac{\text{MS}_R - \text{MS}_E}{\text{MS}_R + (k - 1)\text{MS}_E + \frac{k}{n}(\text{MS}_C - \text{MS}_E)} \quad (3.4)$$

where MS_R = mean square for rows (across raters); MS_E = mean square for error, MS_C = mean square for columns; n = number of subjects; k = number of raters.

We interpret the reliability score according with the work by [Koo et al. 2016](#) as reported in [Table 3.10](#).

Table 3.10: Reliability interpretation based on ICC values.

ICC	Reliability
$ICC \leq 0.5$	poor
$0.5 < ICC \leq 0.75$	moderate
$0.75 < ICC \leq 0.9$	good
$0.9 < ICC$	excellent

It is important to highlight that while the ICC traditionally assumes that all raters are equivalent, in our specific context, these raters are distinguishable due to the variations in time-series lengths. Consequently, it becomes crucial to consider not only the ICC but also to complement its evaluation with the Bland-Altman (BA) plot ([Bland et al. 1986](#)). By doing so, we can gain a more comprehensive understanding of the agreement and disagreement patterns between measurements from different time-series lengths. The presence of any noticeable trends or discrepancies, as indicated by the BA plot, can provide further insights into the nature of measurements and potential biases introduced by varying time-series lengths.

The BA graphical approach can compare each measurement to a supposed gold-standard method. Specifically, we assume the measurement performed on the network estimated on the complete time-series as the gold-standard reference. However, it is worth noticing that a disagreement between measurements obtained from estimations with varying volume numbers in the acquisition does not necessarily indicate which of the two measurements should be favored.

In situations of perfect agreement, the mean difference between measurements should ideally be equal to zero and not show any tendency with respect to the average of the two methods. We evaluate the BA plot according to the criteria proposed by [Giavarina 2015](#). Particularly, we determine if there exists a systematic or proportional bias when comparing a measurement against its gold-standard by finding a regression line of the difference of the estimated methods over their mean. A poor agreement is deducted if it is possible to find a well-fitted regression line.

We conduct our analysis on the iShare and HCP datasets. Within this framework, we generate updated dataset denoted as $iShare_n$ and HCP_n , each corresponding to new versions of adjacency matrices derived from time-series containing n consecutive points.

3.5.2 Results and Discussion

As an initial experiment, we assess the variation in the number of edges within the estimated network (counting 400 edges) when utilizing a reduced number of data points. Results are reported in Fig. 3.23. We observe a noticeable trend where the network estimations converge more closely to those obtained using the complete time-series length as we incrementally increase the number of data points used in the estimation process. This gradual convergence suggests that a larger number of data points yields to more robust network estimations. In the right panel of Fig. 3.23, we evaluate the difference on estimated networks by employing a sliding window encompassing 200 data points at different starting position. This window is iteratively positioned at various starting points within the entirety of the time-series length. The outcomes of this analysis show higher difference with the *standard* network when inferring from the initial segment of the time-series. However, as we progress through the time-series, the network estimation stabilizes and gradually aligns itself at an average of 200 edges difference with respect to the final estimation.

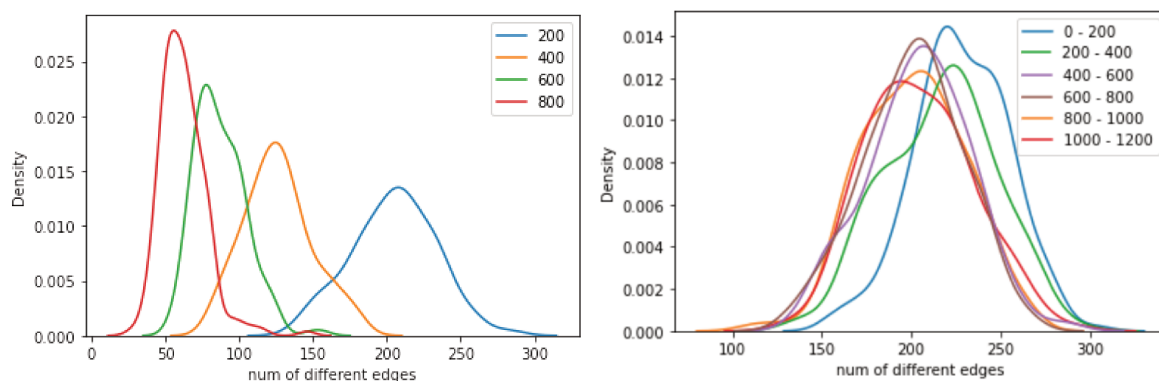


Figure 3.23 – Left: Number of edges discrepancies in observed between the network estimated using the complete available time-series length and its respective subsamples of 200, 400, 600, 800 points. Right: Number of edges discrepancies identified between the network estimated using the complete available time-series length and a subsample of 200 data points originating from various starting positions within the entire available time-series.

As second experiment, we assess the disparities in the degree distribution between network estimations based on shorter time-series versions and their complete counterparts. This disparity is evaluated using the Kolmogorov–Smirnov (KS) test, a statistical tool that quantifies discrepancies in the empirical degree distribution between two samples: the standard network version and the abbreviated version with fewer data points. The KS test facilitates a comprehensive comparison by scrutinizing the degree distribution of the graph derived from the entire time-series against the degree distribution of the graph estimated using a subset of n data points. This approach enables us to discern and quantify the degree of variation in the network’s structural characteristics introduced by the use of shorter time-series segments. Moreover, for comparison, we analyse the attacked network version where a specific number of randomly sampled

edges are subjected to re-sampling. The results of this analysis are shown in Tab. 3.11 for varying time-series length estimations and for the attacked versions.

Table 3.11: Number of significant rejects of the null hypothesis at 0.05 in degree comparison. HCP comprises 200 networks, iShare 1814 networks.

num of points	HCP	iShare	num of attacks	HCP	iShare
200	6	21	50	28	5
400	0	4	80	34	8
600	0	1	110	49	8
800	0	0	140	53	10
			170	55	18
			200	58	11
			230	56	21

Our findings reveal intriguing insights into the stability of the degree distribution of FC networks or the influence of the network inference process on its computation. Indeed, in the majority of instances, the hypothesis of a consistent degree distribution could not be statistically rejected. On the other hand, when the FC networks undergo re-sampling, we observe an increase in the number of times the hypothesis is rejected. This observation seems to imply that despite the use of estimations with shorter lengths, resulting in different edges estimation, the global properties of the FC network are largely preserved and effectively captured even at short acquisitions.

Finally, we assess the absolute agreement concerning nodal statistics at the regional level, employing both ICC and BA plots for a comprehensive evaluation. We evaluate the regional group mean of nodal statistics s_G on the available datasets and their corresponding shorter versions and the NPP associated with each nodal statistics. The ICC reliability results are shown in Figg. 3.24, 3.25, while Fig. 3.26 shows the BA plots concerning the clustering coefficient (other nodal statistics results can be found in Figg. A.2-A.5 in Appendix A).

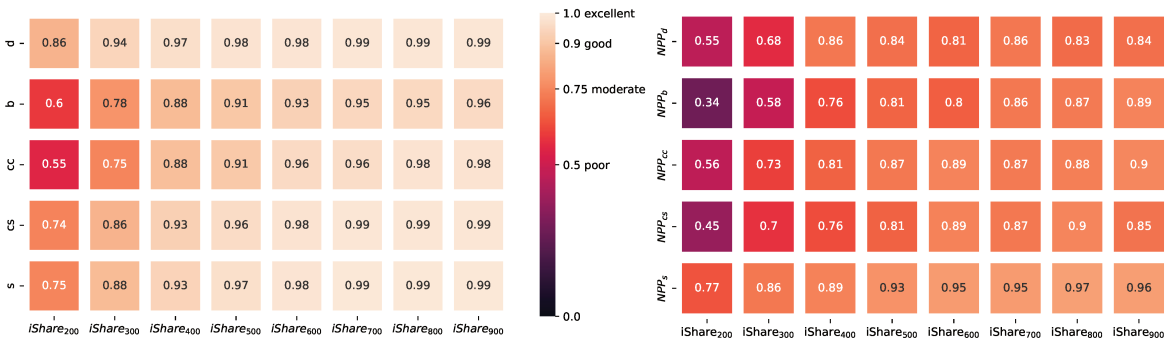


Figure 3.24 – Reliability of nodal metrics in iShare dataset.

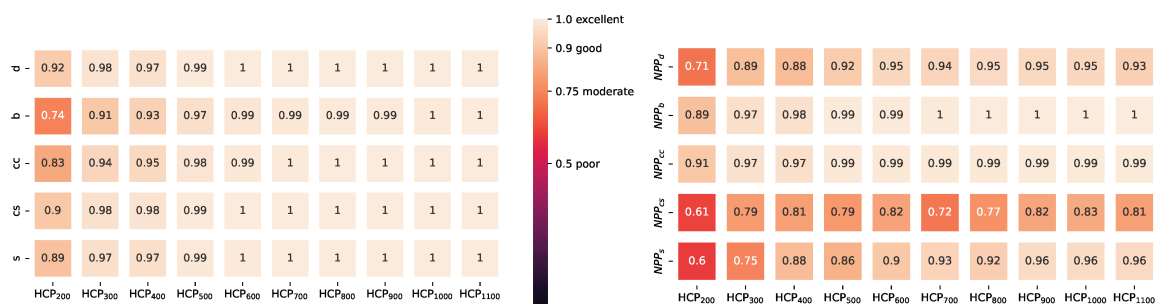


Figure 3.25 – Reliability of nodal metrics in HCP dataset.

As a general tendency, when poor reliability is observed in the group average, poor reliability is also reported in the associated corresponding NPP. In a broader context, we note that the iShare dataset tends to exhibit comparatively lower reliability in contrast to the HCP dataset. This discrepancy could potentially be attributed to the disparity in the number of subjects between the two datasets. The greater population size in the iShare dataset may introduce heightened variability.

Excellent levels of reliability are evident within the HCP dataset, notably emerging within 600-700 points in the group average estimation. Concerning the NPP, we observe excellent scores for betweenness centrality and clustering coefficient, mirroring their average trend from around 700 points. Conversely, second-order centrality demonstrates excellent reliability after 900 points. Closeness centrality, despite reaching the good score of 0.83, surpasses this threshold. Interestingly, the degree also never surpass the 0.95 score which, even if corresponding to good reliability, may be expected to reach excellent score. Yet, it is worth highlighting that even a minor alteration in degree statistics can yield varying estimations of the corresponding structural pattern equivalence classes, in contrast to what happen with continuous nodal statistics where the equivalence is granted up to a certain tolerance. This distinction could potentially elucidate the higher reliability scores observed for continuous nodal statistics in the NPP analysis.

Similarly, when focusing on the iShare dataset, analogous trends emerge in terms of average nodal statistics, with reliability scores achieving good and excellent levels across all statistics commencing around the 500 points. In the context of NPP, results exhibit slightly diminished reliability, with scores higher than 0.9 only for second-order centrality at 900 points, while other cases hover between moderate and good.

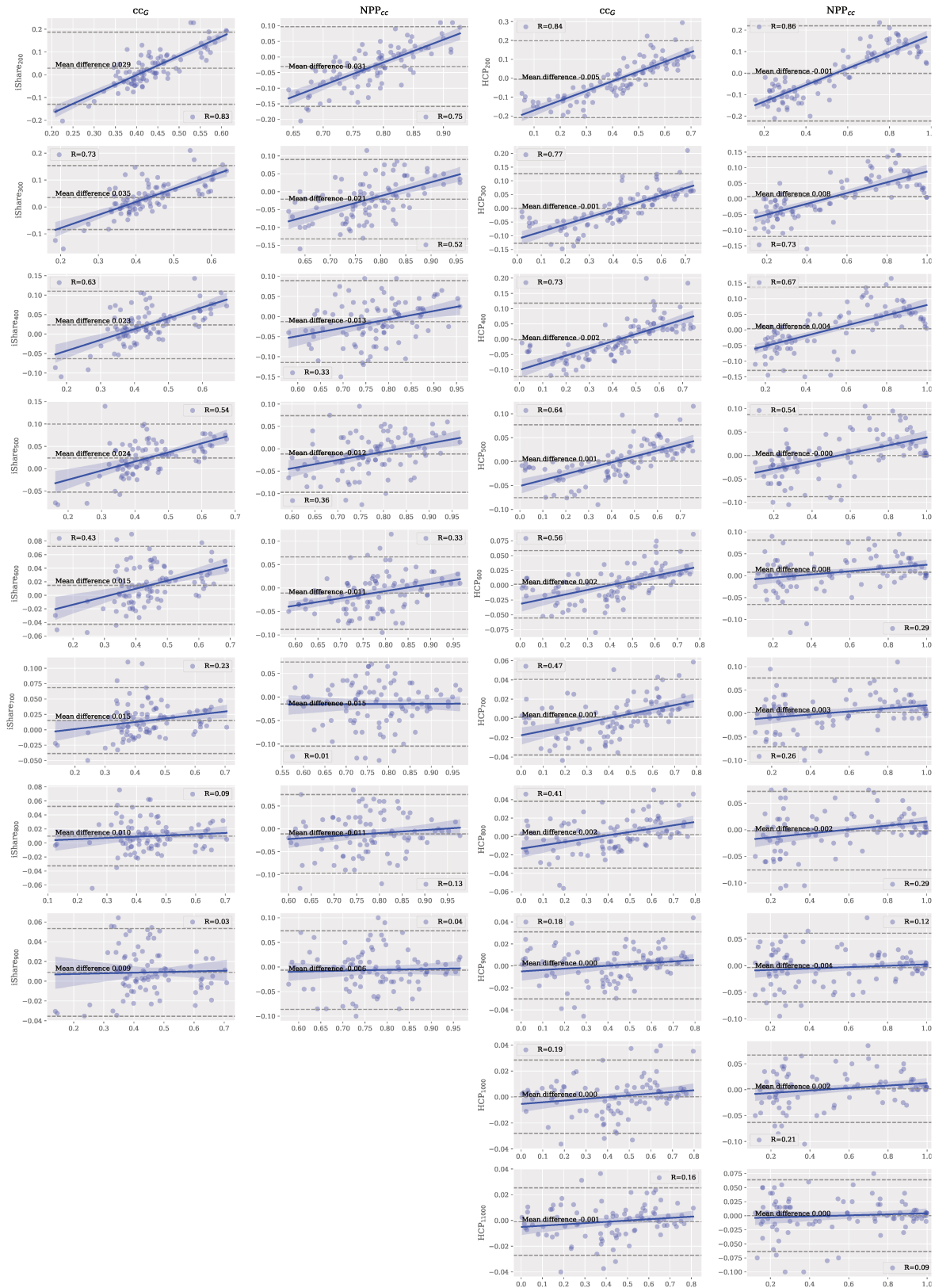


Figure 3.26 – Bland-Altman plot of clustering coefficient in iShare and HCP dataset.

Yet, considering only the ICC results can be misleading. It seems to report generally good reliability already when considering only 200 points. Finer analysis conducted with the Bland-Altman plots reveals additional insights.

For instance, we consider the case of clustering coefficient for which the corresponding BA plots are displayed in Fig. 3.26 in both datasets and different configurations.

Despite showing an excellent ICC score (0.91) for the iShare dataset group average at 500 points, a closer examination of the corresponding BA plot reveals the tendency to overestimate the group average in regions exhibiting higher average difference scores. This trend persists until at least 800 time-series points are employed for the network estimation, highlighting the complexity of the reliability assessment. Interestingly, our proposed NPP seems to stabilize before at 700 points.

A similar pattern is discernible in the HCP dataset, where biases gradually fade away after 800 points for both group average and NPP estimations. This underscores the importance of considering complementary methodologies, such as BA plots, to gain a more comprehensive understanding of the reliability landscape together with the necessity for extended data acquisition when local FC characterization is aimed.

In Fig. 3.27, a visual representation of the NPP is presented for various nodal statistics, regressed over the standard dataset. The depicted curves exhibit a gradual convergence to the final estimated values as the considered time-series lengths increase. Again, this observation emphasizes the importance of extended time-series in achieving more accurate and stable estimations.

A comparative analysis is then conducted by juxtaposing these results with the corresponding attacked versions showcased in the last row of Fig. 3.27. Evidently, the estimation of randomly spurious edges results in very different curves, showing that the employment of shorter time-series lengths is capable of capturing meaningful FC information, but requires longer estimation for stabilizing.

Consequently, we advocate for the adoption of time-series consisting of at least 800 data points to ensure better consistency and reliability in FC studies, in lines with previous studies (Termenon et al. 2016; Noble et al. 2019).

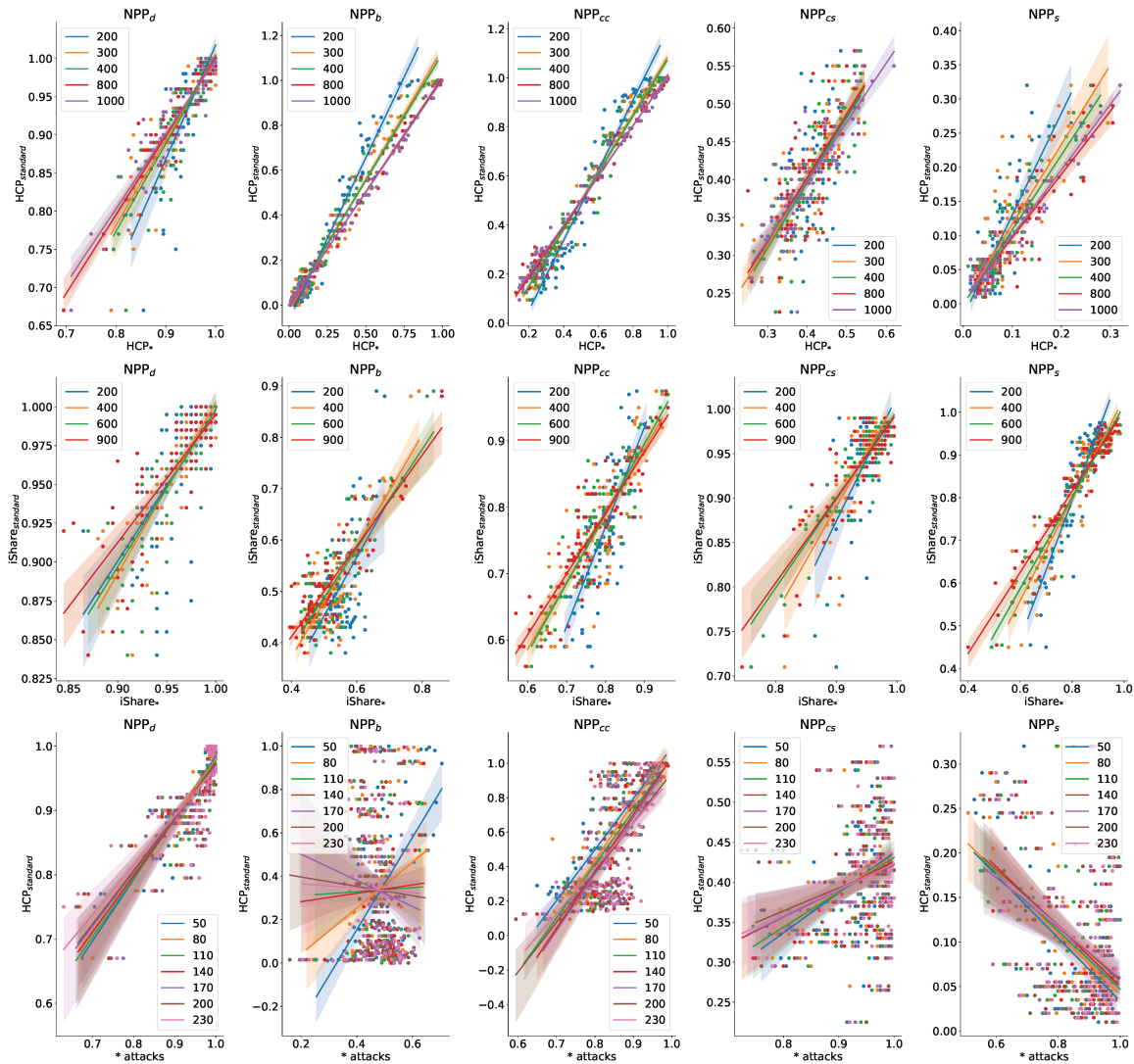


Figure 3.27 – Nodal percentage of participation associated with different nodal statistics on the standard dataset and regressed over its corresponding shorter or attacked versions.

3.6 Summary and conclusion

In summary, our work introduces an innovative framework that combines classical network statistics with network reduction and classification techniques. Drawing inspiration from established concepts in social network science, our framework aims to detect an originally defined structural pattern of a network and employ it for comparison and characterization within a graph collection. These structural patterns are based on single or combination of classical nodal statistics. Our proposal to combine multiple nodal statistics offers an original way to simultaneously capture the intricate information within the node space of a network. The introduction of novel global graph descriptors for nodal statistics combination has the potential to put some order in the plethora of existing graph statistics (see also Appendix A), originally redefining the concept of redundancy (Rubinov 2022) in the orthogonality score.

Notably, the use of our novel graph orthogonality score uncovers subtle distinctions between network groups and, particularly when applied to pathological and healthy control networks, it detects differences which would be overlooked by the associated nodal-statistics in conventional methods. Simultaneously, it preserves consistency among healthy controls coming from various databases.

In order to measure individual similarity, built upon these structural patterns, we define the correspondence structural pattern score. Our findings reveal high correspondence between real networks in comparison to synthetic generative models, demonstrating the capability of structural pattern characterization to capture meaningful functional connectivity (FC) information. The versatility of our framework is investigated through its application across various pathologies, effectively discerning even between nuanced symptomatic differences in Parkinsonian patients.

While our result as classification scores may not surpass other existing methods in FC, the significance of our proposal lies in its ability to identify globally affected underlying structures within pathologies and to detect anomalies at the regional level. Notably, the definition of the nodal participation enables the ranking of perturbed regions when comparing different functional connectivity network groups, detecting the homotopy brain property and emphasizing the advantages of local characterization.

It is interesting to observe how different pathological conditions are better distinguished by different nodal statistics combination, revealing the need of considering a variety of graph properties (Fig. 3.15). This phenomenon suggests that various pathological contexts and neurophysiological mechanisms impact the brain’s functional connectivity in distinctive ways.

Finally, our investigation into the reliability of regional measurement advocates for the importance of acquiring long time-series data when seeking precise local characterizations. Particularly, reliability and accuracy of edge estimation play a crucial role in drawing meaningful conclusions from network data. Even a small deviation or error in edge estimation can lead to significant differences in the resulting network structural pattern and properties. Thus, we propose an extension of our framework to address uncertainty in edge estimation, as outlined in Appendix A.3. While the theoretical foundations are robust, empirical experimentation is essential to validate the potential benefits of this extended framework and it is object of on-going work.

3.6.1 Limitations

Our work, while yielding valuable insights, faces several limitations primarily rooted in the nature of the data employed in our experiments. It is important to observe that, especially when applying classification methods, the dataset sizes are relatively small (Helweggen et al. 2023). In light of this, our evaluation heavily relies on cross-validation techniques, enabling us to assess the model’s performance across multiple train-test splits. Furthermore, a crucial aspect already highlighted in our literature review, pertains to the reliance on specific acquisition hyper-parameters for our results. The extension of our method’s applicability to datasets originating from various sources

with varying data quality remains a challenge that necessitates future evaluation and validation.

Moreover, although our framework exhibits good results in classifying distinct conditions, other existing FC classification methods, mainly not-based on a graph estimation, reports better classification accuracy. Emphasizing our primary objective, our framework excels in identifying and characterizing the inherent structural changes rather than solely prioritizing state-of-the-art classification scores. Our emphasis lies in the interpretability and utility of the underlying structural patterns rather than in classification performance. Finally, we focus on a predetermined collection of nodal statistics and their pairwise combinations. The selection of these nodal statistics was influenced by prior research, and the inclusion of pairs of nodal statistics can facilitate the results interpretation. However, it is worth noting that the framework does offer the flexibility to investigate all conceivable combinations and relevant nodal statistics.

3.6.2 Future work

In moving towards the practical integration of FC graph-based methods for clinical applications, it is imperative to establish a golden standard for the entire pipeline—from fMRI data acquisition to network estimation and similarity definition. The current multi-steps process involving graph modeling and subsequent analysis presents challenges to research comparability and inhibits the realization of FC as a reliable disease biomarker.

Indeed, a real FC biomarker application in clinical contexts might still be some distance away and its realization is not guaranteed. Yet, as it may be the case for Parkinson Disease, embracing a multimodal approach could be the key (Long et al. 2012). Just as a clinician considers a patient’s complete medical history and various diagnostic tests rather than relying solely on a single data point, mathematical discrimination models should leverage all pertinent modalities available. Integrating multiple data streams can enrich the accuracy and robustness of the analysis, potentially yielding more powerful diagnostic tools.

Furthermore, the emergence of end-to-end learning frameworks that commence with raw image data and culminate in predictive outcomes holds promise for improved accuracy. This approach may garner greater interest in clinical practice, potentially diminishing the current emphasis on network science methodologies. However, research on this shift may focus on model explainability.

Considering our analysis framework, it can be extended specifically to handle equivalence over the graphs in a family. In this case, the group version nodal statistics assigns a value (or an interval) to each node in the vertices-set, such as the average per node of the statistics across multiple graph instances (or its first-third quartile interval). Then, we could introduce the corresponding nodal equivalence relation whose nodes are equivalent if their average of nodal statistics is the same (or fall in the same interval). In this case, the definition of the structural pattern corresponds to *an average structural*

pattern of a virtual average graph. The graph family version of power coefficient and orthogonality corresponds to the traditional definition on this average graph. The ability of the average structural pattern to characterize the group of graphs could be explored.

Finally, the generality of our proposal enables application across diverse domains. The concept of structural patterns of our methodology, has broader applicability in other different fields such as social science, protein analysis, urban networks, and more generally in network structured data. The flexibility of our proposal will allow an application beyond neuroimaging, extending into various domains where structural patterns may hold significance.

Part B

Artificial Neural Networks

— Chapter 4 —

Context and Background

This chapter provides an overview of the context, background, and related works concerning the applications of network science to artificial neural networks (ANNs). Initially, ANNs are formally defined, accompanied by an introduction to the adopted notation. Thus, continual learning paradigm is introduced in relation with the development of robust artificial systems. Different learning scenarios and existing continual learning strategies are presented and categorized. Particularly, we propose to distinguish continual learning strategies which are biologically inspired from those algorithmic-based. Finally, the chapter briefly revisits network science applications for designing sparse or more effective artificial neural networks, as well as existing research related to our work.

Contents

4.1	Artificial Neural Networks	105
	Definition	105
	Feedforward Artificial Neural Network	106
	Classification Evaluation	107
4.1.1	Robust Artificial Neural Network	109
4.2	Continual Learning	110
4.2.1	Continual Learning Scenarios	112
4.2.2	Continual Learning Strategies	114
4.3	Network Science in ANNs	121

4.1 Artificial Neural Networks

The history of artificial neural networks can be traced back to the 1940s when McCulloch and Pitts introduced the first formal neuron (McCulloch et al. 1943). This first neuron model was based on the idea that complex inputs could be simplified and categorized into a binary decision. Later, in 1949, Hebb introduced the concept of synaptic modification, suggesting that learning occurs through the strengthening or weakening of connections among neurons (Hebb 1949). Then, in 1958, Rosenblatt developed the perceptron, the first artificial neuron (Rosenblatt 1958). The perceptron consisted of multiple input combined together to produce a single output.

In the following years, advancements continued to be made in the field of neural networks. In 1986, Rumelhart introduced the multilayer perceptron, which allowed the construction of more complex networks by combining many perceptrons in hierarchical structure composed of multiple layers (Rumelhart et al. 1986).

Definition

From a mathematical point of view, we define an artificial neuron as a parametric function $\phi_{\mathbf{w},b}$ which transforms an input vector x into an output value y . Similarly, an artificial neural network (ANN) is a mathematical function that processes information from the input to the output by applying a hierarchical combination of artificial neurons. The way the neurons (or units) are combined together determines the ANN architecture. Such architecture can be easily represented as a graph, whose vertices are the units and the oriented links represent a linear transformation applied to the output of the first connected unit and being the input of the second connected unit (See Fig. 4.2).

The architecture determines the number of layers (list of neurons having the same input), the number of units per layer, and the presence and orientation of edges among the units. The nonlinear transformation function $\sigma_{\mathbf{w},b}$ associated to each unit is called *activation function*. The activation function mimics the stimulation of a biological neuron by transforming multiple inputs into an output value. Common examples of activation functions are the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ or the Rectified Linear Unit (ReLU) function $\sigma(x) = \max(0, x)$ shown in Fig. 4.1.

It has been proven that an artificial neural network can approximate any continuous function with infinity precision level (Hornik et al. 1989). This property of ANN models, named **expressiveness**, together with their adaptability to different tasks has allowed ANN to find applications in multiple domains reaching or even exceeding human-like performances (Esteva et al. 2017; Silver et al. 2016; Rajpurkar et al. 2018). According to their architecture, artificial neural network can be grouped in different structure classes: Feedforward Neural Network, Convolutional Neural Network, Recurrent Neural Network, Boltzmann Neural Network, Spiking Neural Network or combinations of this modular architecture. In our work, we mainly consider feedforward

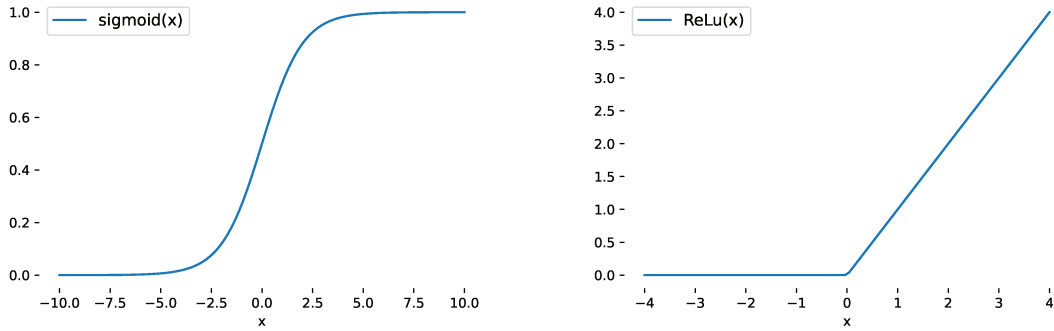


Figure 4.1 – Examples of common activation functions

ANNs, but see [Bengio et al. 2015](#); [Hassabis et al. 2017](#) for reviews of different ANNs and applications. In feedforward ANNs, neurons are organised in a sequence of layers oriented from the input to the output, making the architecture graph representation a multipartite graph.

Feedforward Artificial Neural Network

We denote l the l -th layer of a feedforward neural network, where $l = 0$ represents the input layer. The depth of the neural network is denoted L , which corresponds to the total number of layers in the architecture. Each layer l consists of a varying number of units, indicated by $H(l)$, and each unit is associated with an activation function σ . The connections between layer l and $l + 1$ correspond to a linear transformation A^l , which performs a weighted sum. The synaptic weights of the neural network are determined by the set of coefficients in each linear transformation. In particular, we represent the weight associated with the edge connecting the i -th unit of layer $l - 1$ and the j -th unit of layer l as $\mathbf{w}_{i,j}^{l-1,l}$, which is the coefficient of the output of the i -th unit of layer $l - 1$ in the linear operator A^l .

The artificial neural network, denoted as $f_{\mathbf{W}}$, processes an input sample $x \in \mathbb{R}^{H(0)}$ by applying a sequence of linear and nonlinear transformations, as shown in Equation 4.1. Each unit i in layer l is associated with an activation value u_i^l , calculated using Equation 4.2, where b_i is a fixed bias term, \mathbf{w} and b are the parameters of the artificial neuron function and $u_i^0 = x_i$, where x_i is the i -th entry of the input vector x .

$$f_{\mathbf{W}}(x) = \sigma^L \left(A^L \sigma^{L-1} (A^{L-1} \dots A^1 \sigma^1 (A^0 x) \dots) \right) \quad (4.1)$$

$$u_i^l = \sigma^l (A^{l-1} x)_i = \sigma^l \left(\sum_j \mathbf{w}_{j,i}^{l-1,l} u_j^{l-1} + b_i \right) \quad (4.2)$$

The set of parameters \mathbf{W} of the neural network are *learned* through the training or learning phase of the neural network. Here, we consider a **supervised** setting where inputs and their expected outputs are provided to the neural network ([LeCun et al. 2015](#)). A classical example is to perform a classification task: associating to each

input a label (i.e. an image labeled by category). The neural network processes a set of inputs and produces the corresponding outputs. During the training phase, the objective is to obtain as output label the desired one, (or as close as possible). This is achieved by optimizing an objective function. Such a function measures the distance (the *error*) between the NN outputs and the desired ones, thus the goal of a supervised classification learning is to learn the prediction function $f : \mathcal{X} \rightarrow \mathcal{Y}$ given a set of K labeled data $\{(x^k, y^k)\}_{k=1}^K$, $x^k \in \mathcal{X}, y^k \in \mathcal{Y}$. This corresponds to the optimization problem:

$$\min_{f_{\mathbf{w}} \in \mathcal{F}} \frac{1}{K} \sum_{k=1}^K \mathcal{L}(y^k, f_{\mathbf{w}}(x^k)) \quad (4.3)$$

where \mathcal{X}, \mathcal{Y} are respectively the data space and the label space, \mathcal{F} is the set of function between \mathcal{X}, \mathcal{Y} , the pair (x^k, y^k) are the data points and their labels, \mathcal{L} is a **loss function** which computes the error between y^k and $f(x^k)$. Given $f_{\mathbf{w}}$ the parametric function associated to a considered artificial network, the optimization problem reduces in finding the optimal weights

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{w}} \frac{1}{K} \sum_{k=1}^K \mathcal{L}(y^k, f_{\mathbf{w}}(x^k)). \quad (4.4)$$

In the following, when it is not needed to precisely identify the synaptic weights in a layer, we will enumerate the parameters of \mathbf{W} simply as \mathbf{W}_j , adopting an enumeration which does not take into account the hierarchical structure of the ANN.

A solution to Equation 4.4 can be achieved by gradient descent and *backpropagation* (Cauchy et al. 1847; Rumelhart et al. 1986), which applies the chain rule of derivative from the output to the input in order to determine the partial derivative of the loss function with respect to each synaptic weight in the neural network. Each weight $\mathbf{w}_{i,j}^{l-1,l}$ is updated iteratively during the training process by subtracting the product of the learning rate η and the partial derivative of the loss function \mathcal{L} with respect to the weight as $\mathbf{w}_{i,k}^{l-1,l} \leftarrow \mathbf{w}_{i,j}^{l-1,l} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}_{i,j}^{l-1,l}}$. In practice, the iteration is repeated until convergence and grants reaching a global optimum if \mathcal{L} is convex.

Classification Evaluation

In a classification task of supervised learning, the evaluation metric associated with the performance of the ANN is typically given in terms of accuracy.

Definition 4.1. The **accuracy** in a classification task is given by the ratio of the number of correct classifications and the total number of tested samples.

$$\text{Accuracy} = \frac{\text{number of correctly classified samples}}{\text{total number of classified samples}} \quad (4.5)$$

Adam Optimizer. From an implementation point of view, instead of updating the weights at every single sample, a commonly used learning algorithm is stochastic gradient descent which employs small batches of randomly sampled training examples to

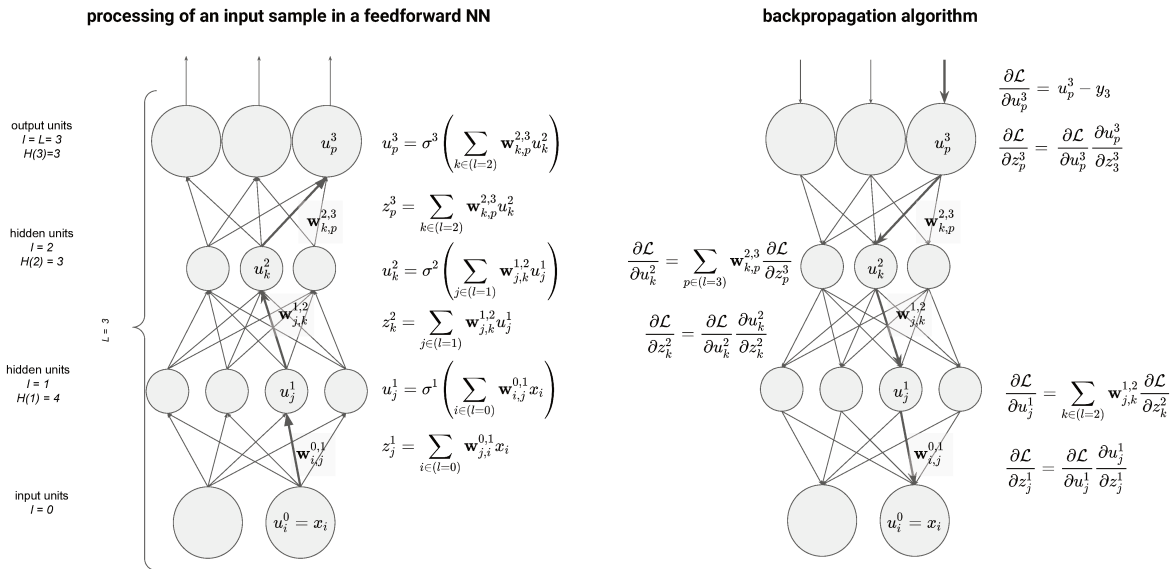


Figure 4.2 – Visualization of the backpropagation algorithm in a feedforward neural network. Figure adapted from [LeCun et al. 2015](#).

estimate the averaged gradient of those samples and update the weights accordingly ([Robbins et al. 1951](#)). In the implementation of stochastic gradient descent, an important step refers to the selection of an appropriate learning rate η : a too small learning rate can lead to slow convergence, while a too large rate may hinder convergence and cause the loss function to fluctuate around the minimum. To overcome these issue new methods have been proposed to adapt the learning rate automatically in the learning process ([Sun et al. 2019](#); [Yang et al. 2020b](#)) or even to define unit-base learning rates ([Blier et al. 2019](#)). In the following, we will use Adam ([Kingma et al. 2014](#)) as optimization algorithm which determines an adaptive learning rate through the estimation of mean and variance of the gradient. The Adam optimizer grants a quick convergence and a faster training phase ([Kingma et al. 2014](#)). A pseudo-code can be found in [Algorithm 1](#).

Algorithm 1 Adam: Stochastic Optimization, adapted from [Kingma et al. 2014](#).

Require: η : Stepsize

Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

```

1:  $m_0 \leftarrow 0$  ▷ Initialize 1st moment vector
2:  $v_0 \leftarrow 0$  ▷ Initialize 2nd moment vector
3:  $t \leftarrow 0$  ▷ Initialize timestep
4: while  $\theta_t$  not converged do
5:    $t \leftarrow t + 1$ 
6:    $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$  ▷ Get gradients w.r.t. stochastic objective at timestep  $t$ 
7:    $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$  ▷ Update biased first moment estimate
8:    $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$  ▷ Update biased second raw moment estimate
9:    $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$  ▷ Compute bias-corrected first moment estimate
10:   $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$  ▷ Compute bias-corrected second raw moment estimate
11:   $\theta_t \leftarrow \theta_{t-1} - \eta \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$  ▷ Update parameters
12: end while
13: return  $\theta_t$  ▷ Resulting parameters

```

4.1.1 Robust Artificial Neural Network

Despite the expressiveness of neural network model, a major challenge in training artificial neural networks is represented by their **generalization** ability, i.e. the ability to associate good output to new input samples which are unseen during the training phase. In some cases indeed, the model may be affected by the overfitting phenomenon, where the learned function is a very good approximation on the training dataset, but failed when applied to the test data (see Fig. 4.3).

In the development of a **robust** neural network model, the generalization issue remains a challenge in two different settings: static generalization ability, where the neural network is expected to be able to generalize to unseen sample extracted from the same distribution (or from distribution-shift) inside a given fixed class set; dynamic generalization ability, where the neural network is expected to generalize to previously unseen tasks by re-using previously learned information and after a new training session. The desiderata of a robust neural network model are listed in Table 4.1.

In the following chapters, we consider the dynamic generalization ability, where a neural network model is trained in sequential manners (for static generalization ability in ANN see for instance [Neyshabur et al. 2018](#); [Novak et al. 2018](#); [Arbel et al. 2023](#); [Pitas et al. 2019](#)). Moreover, our contribution to adversarial attacks robustness is reported in the Appendix B.4.

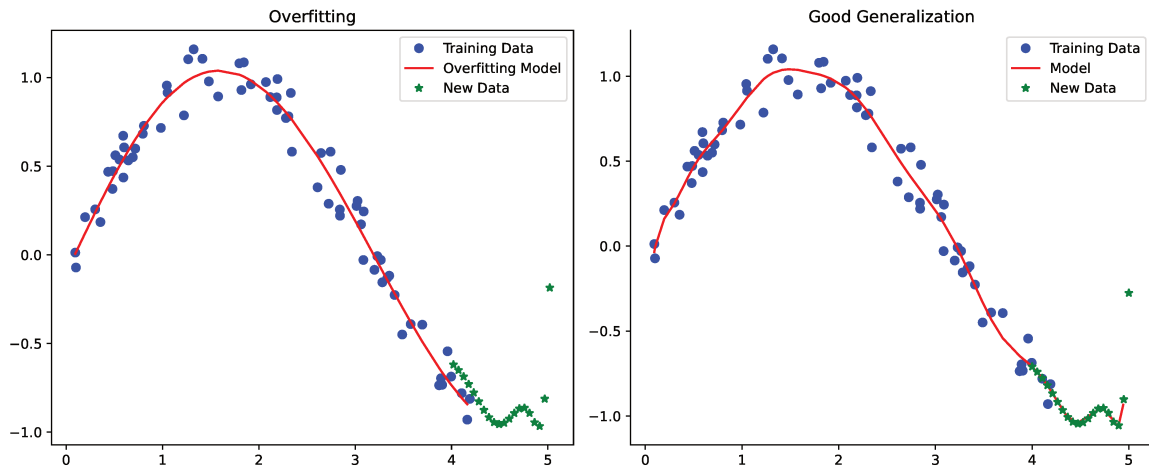


Figure 4.3 – Visualization of two cases where the predicted model does not generalised to new sample due to overfitting (left) and a good generalization case (right).

Table 4.1: Desiderata of a robust neural network systems.

Property	Definition
Static generalization ability	the learned function should be able to generalize to unpreviously seen samples belonging to the same distribution or class-set
Dynamic generalization ability	the learned function should be able, after a new learning phase, to generalize to unpreviously seen samples from a different class, without forgetting the previously learned information
Robustness against adversarial attacks	the learned function should not be fooled by slightly perturbed inputs (Goodfellow et al. 2015)

4.2 Continual Learning

One of the main debates surrounding artificial neural networks in the fields of cognitive science and artificial intelligence revolves around the disparities in behavior between ANNs and biological brains across different scenarios where ANN fail to reproduce crucial aspects of human intelligence (Hassabis et al. 2017; LeCun et al. 2015; Serre 2019). Indeed, ANNs are blamed to be too simplistic models (Doerig et al. 2023; Marcus et al. 2019; Marcus 2020) of natural intelligence since they exhibit contrasting behaviors compared to humans in various contexts: they require massive-size data to learn, they fail in generalization, they can be fooled by adversarial attacks, they are unable to learn in sequence and to transfer learning in different tasks.

Among the others, a main challenge in the development of a more-human-like artificial agent is represented by the need of learning in a continuous manner. Indeed, while

natural intelligence is faced by a continuous adaptation to new information, artificial systems are usually trained under the assumption of a static environment (De Lange et al. 2021; Hadsell et al. 2020). Natural intelligence is able to have a continual adaptation to new tasks across a lifetime and even to re-use previously learned ability for faster learning or consolidation of previously seen tasks. An entire research program about **lifelong learning** aims at reproducing this process in artificial systems. This is the opposite of the static-assumption, where the training of a neural network is performed over a training dataset and the learned function is assumed to generalize to unpreviously seen samples. However, in real life, agents are continuously faced with new information, reinforcing the need of developing a model which can continuously learn and adapt to new tasks. A first attempt to implement lifelong learning in artificial systems is represented by a continual learning paradigm, where neural networks are trained sequentially over different datasets. See Tab. 4.2 for a proper definition of sequential and related learning paradigms.

Table 4.2: Glossary of different learning paradigms. For more details see for instance Chen et al. 2018.

Learning Paradigm	Definition
Offline Learning	the training phase consists of a unique learning session, where all the samples in the training data are presented to the NN.
Online / Incremental / Continual / Sequential Learning	the training phases consists of multiple learning sessions, where new samples for training appear across learning sessions.
Transfer Learning	the training phase consists of at least two sessions where the learned information of the first phase is transfer to the second phase, yet at the end of the last learning session the model is evaluated only on its performance of the last-seen dataset.
Curriculum Learning	the training phases consists of multiple learning sessions where the order is set <i>a priori</i> in order to go from the easiest task to more difficult ones.
Reinforcement Learning	an agent is trained based on a system of rewards to perform a particular task.
Metalearning	the agent is trained to learn the best learning strategy.
Lifelong Learning	a special case of online learning which also implies a faster learning while going through the learning sessions.

One major issue is the occurrence of the **catastrophic forgetting** phenomenon when neural networks are trained in sequential sessions. This phenomenon manifests when a neural network model, as it adapts to new tasks, loses its capability to execute the previously learned ones.

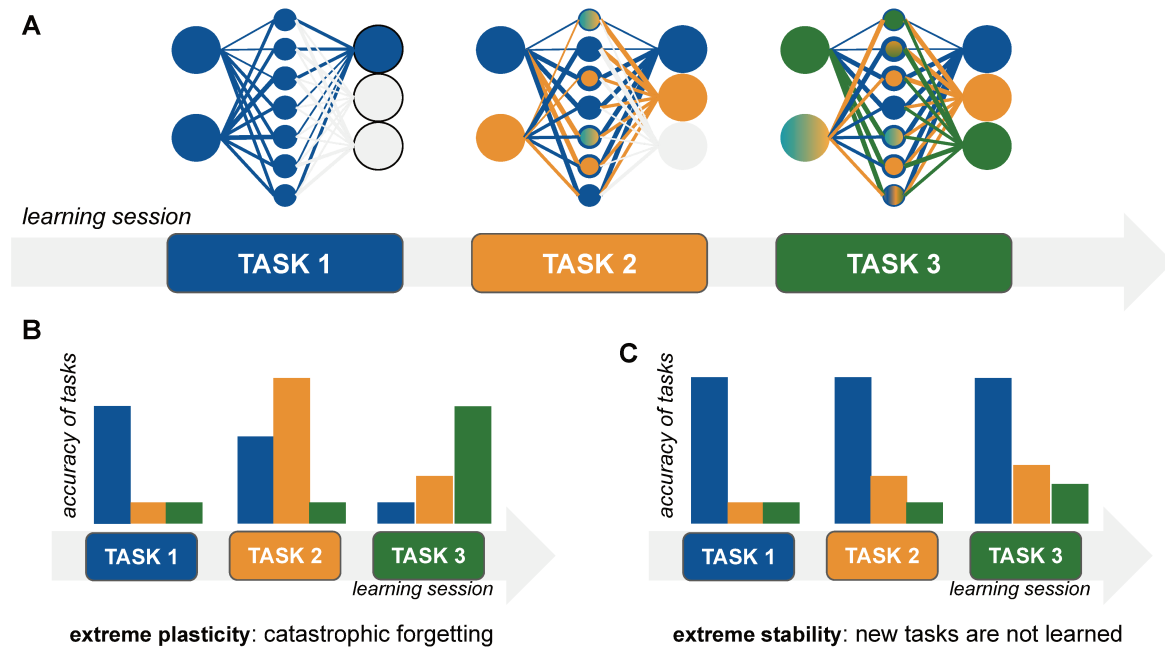


Figure 4.4 – Schematic visualization of the sequential learning framework of three tasks. (A) The same ANN is trained sequentially to perform three different tasks. While the architecture does not change across the learning sessions, the associated weights are updated at the end of each session. (B) Example of performances of a model affected by catastrophic forgetting. The Task 1 performance at the end of the third learning session, has deteriorated. This corresponds to ANN where plasticity property is stronger than its stability. (C) Example of performance of a model which perfectly learns Task 1, at the end of all the learning sessions. However, the model does not learn new tasks. This corresponds to an ANN where the stability property is stronger than the plasticity. (Adapted from Figure 2 in Hadsell et al. 2020)

As shown in Fig. 4.4, a sequentially-trained model can result in two extreme cases: an extreme stability case and an extreme plasticity one. **Stability** and **plasticity** are properties of human intelligence which refer respectively to the ability of maintaining previous learned information and adapting to new one. A robust continual learning procedure is expected to find the good balance between the two extremes requiring enough plasticity for the integration of new knowledge, but also enough stability preventing the forgetting of previous knowledge. Determining the good trade-off between these properties goes under the name of the stability-plasticity dilemma (Abraham et al. 2005; Mermillod et al. 2013).

4.2.1 Continual Learning Scenarios

The training of a neural network model in sequential learning sessions includes different learning scenarios which are usually categorized in three different types (Hsu et al. 2018; Ven et al. 2018; Mainsant et al. 2022; Ke et al. 2022; Wang et al. 2023b). In general, a sequential learning framework implies that the neural network model sees at different learning session $t = 1, \dots, T$, a different dataset $\mathcal{D}_t = \{(x_t^k, y_t^k)\}$ including

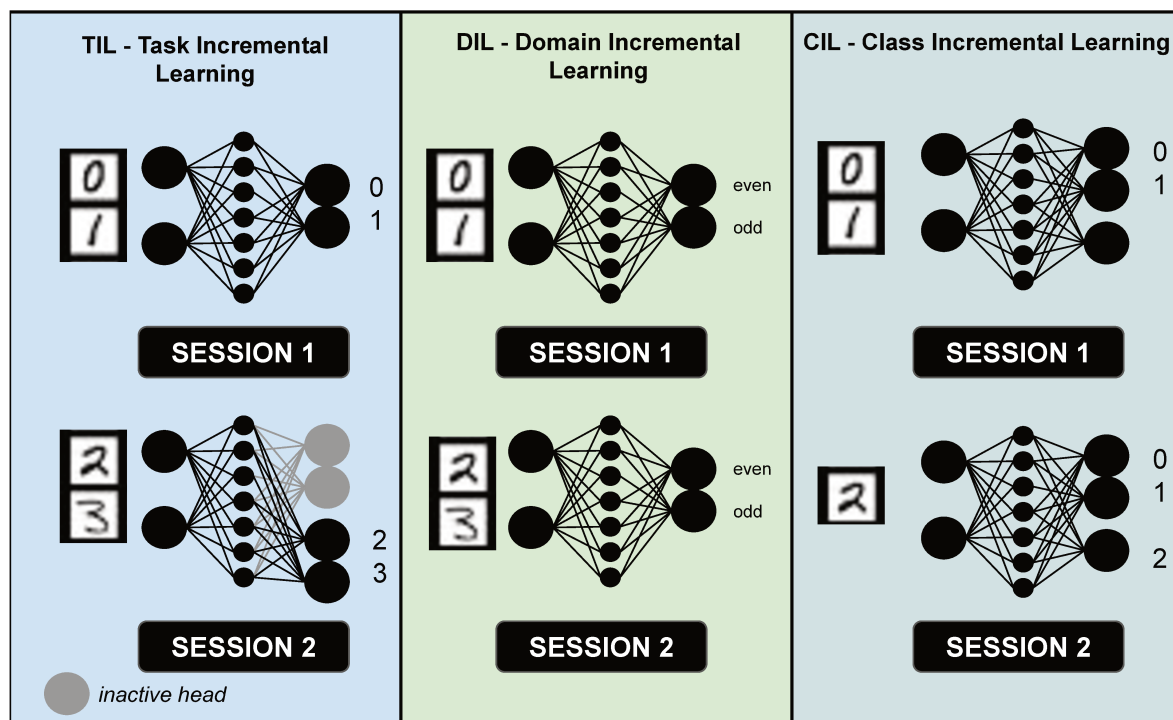


Figure 4.5 – Visualization of three examples of the different continual learning scenarios on the MNIST dataset. TIL refers to the learning of different tasks at each session. DIL refers to the learning over different datasets of the same task in sequential learning session. CIL concerns the learning of different class at each learning session. Adapted from Hsu et al. 2018.

pairs of sample x_t^k and their corresponding label y_t^k . We distinguish three continual learning scenarios: task incremental learning, domain incremental learning and class incremental learning. A visualization of the three different cases can be found in Fig. 4.5.

In a **task incremental learning (TIL)**, the model is trained in sequential learning sessions to perform different tasks. In this configuration, at test time the learning session identifiers are provided. Across the learning session, the model is not entirely trained on the new task, since the output layer (i.e. the classifier) is specific of the learning session. Yet, the first layers are shared and sequentially trained. This means that after learning all tasks T , at test time we apply the shared model trained sequentially up to the T -th learning session, with the output layer specific task t . This approach is called multi-head configuration. In a TIL scenario, not only the class labels are distinguished at each learning session, but different tasks can be performed at each session: for instance a classification task can be followed by a regression task and so on. As an example of TIL, consider a model that, with each new task, learns to classify two classes. For instance, at $t = 1$, the model learns to distinguish handwritten digit images of 0 and 1, then, at $t = 2$, it is trained to classify 2 and 3, and so on. Since we have access to the learning session identifier, the goal of TIL is to learn the function $f : \mathcal{X} \times \mathcal{T} \rightarrow \mathcal{Y}$ with \mathcal{X} the input space of all tasks, and \mathcal{Y} the label space of all the tasks. In this scenario, the label spaces of all tasks learned so far (\mathcal{Y}_t) is a subset or

equals \mathcal{Y} .

Domain incremental learning (DIL) refers to the sequential learning on different dataset of a same task: the class labels are the same across the learning sessions. In this scenario, task identifiers are not provided. In DIL the model is trained subsequently on new instances of the same class. Across the learning session the entire model is updated and trained over and over. This approach contrasts with the process in a multi-headed configuration and is referred to as a single-headed setup. In DIL the learned function is $f : \mathcal{X} \rightarrow \mathcal{Y}$, in this scenario \mathcal{Y}_t is constant across learning session and equals to \mathcal{Y} . As an example of DIL, consider a model trained to classify two classes: even and odd numbers from handwritten digit images. At the learning session $t = 1$ the model is trained on 0 and 1, then, at $t = 2$, it is trained using 2 and 3 and so on.

Finally, the **class incremental learning (CIL)** refers to a sequential learning where at each learning session classes which were not seen before, are learned. Session identifiers are not provided, and we have that $\mathcal{Y}_t \cap \mathcal{Y}_{t'} = \emptyset, \forall t \neq t'$ and \mathcal{Y} corresponds to the union of all $\cup_{t=1}^T \mathcal{Y}_t$. The objective of CIL is also to learn the function $f : \mathcal{X} \rightarrow \mathcal{Y}$. This is the most difficult scenarios since while training on task $t + 1$ the model needs to define the boundaries with the previously seen samples without re-using the dataset \mathcal{D}_{t-1} . This results in the so-called Inter-task Class Separation problem identified in [Kim et al. 2022](#). As an example, we can think at a model trained to recognize the ten handwritten digits trained sequentially for each digit.

4.2.2 Continual Learning Strategies

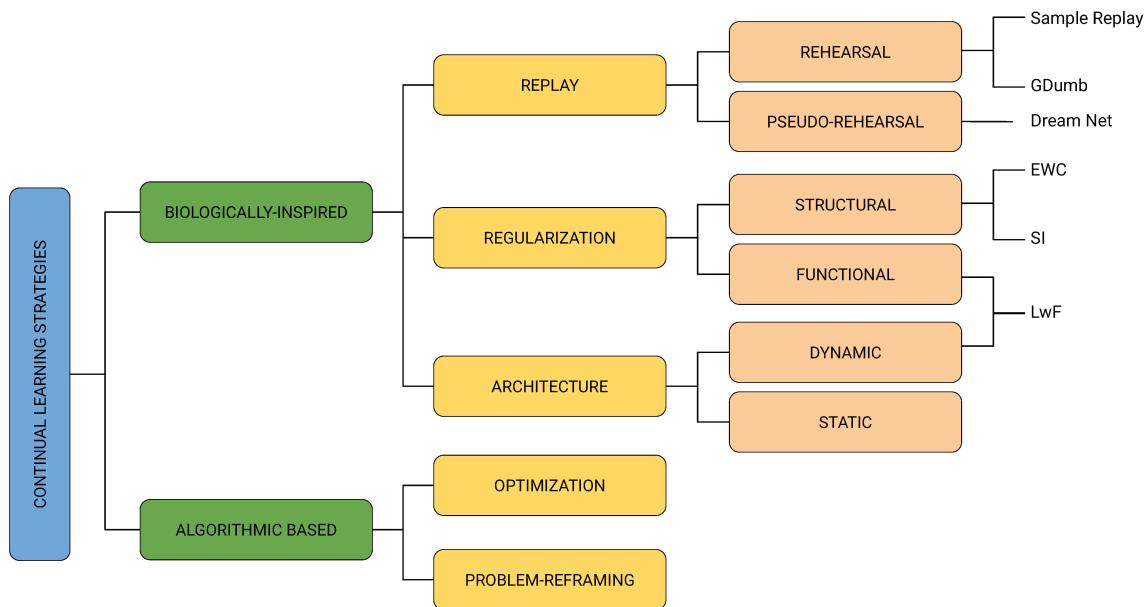


Figure 4.6 – Overview of existing continual learning strategies. Branches are not exclusive and can intersect.

Across the years, multiple techniques and approaches have been proposed to face the challenge of continual learning.

As we are mainly interested in ANN behavior as brain-model, we concentrate into brain-inspired approaches which includes **regularization method**, **replay methods** and **architecture based methods** (Chang et al. 2021; Hadsell et al. 2020; Parisi et al. 2019; De Lange et al. 2021; Aimone et al. 2009; Kempermann et al. 2004; Draelos et al. 2017). A visualization of our proposed categorization can be found in Fig.4.6. Our categorization is based on the framework proposed by Wang et al. 2023b, but it also distinguishes between biologically-inspired strategies and algorithmic-based approaches. The latter category includes optimization-based strategies that directly operate in the optimization programs, as well as problem-reframing-based strategies that involve the use of self-supervised learning or pre-training the model prior to commencing the continual training sessions.

The first family has been inspired by the replay mechanism observed in rodents and humans (McClelland et al. 1995; O’Reilly et al. 2014; Liu et al. 2019c), where neural activity patterns are replayed during sleep as a means of memory consolidation (Fig. 4.7 (b)).

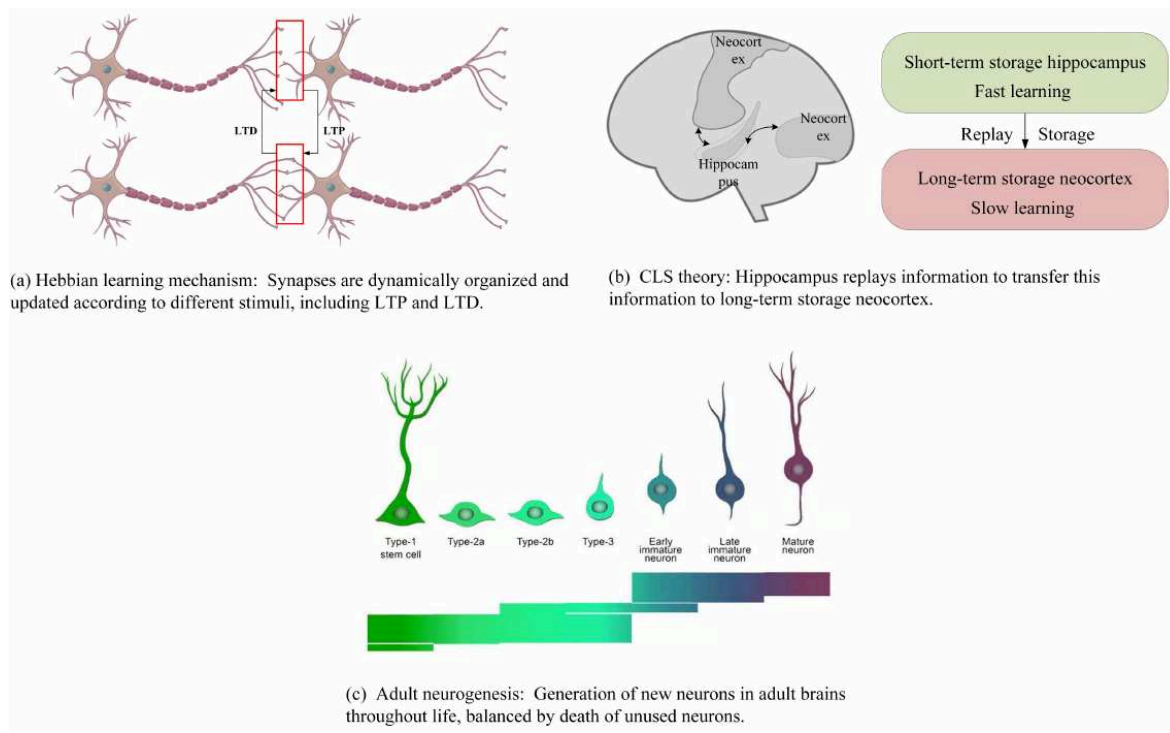


Figure 4.7 – Brain-inspiration of continual learning strategies. The human brain adopts different continual learning strategies which have served as inspiration in the development of ANN continual learning strategies. Adapted from Chang et al. 2021.

For ANNs, two implementations have been introduced: the rehearsal approach where some of the previously seen samples are reused with the current samples to learn, and the pseudo-rehearsal approach where artificially generated new examples are introduced to represent previously learned knowledge. We consider two rehearsal approaches: **Sample Replay** that stores randomly previously seen samples (Lomonaco et al. 2021)

and **GDumb** that selects the stored samples by asymptotically balancing the class distribution (Prabhu et al. 2020). Both methods require to fix a memory budget corresponding to the maximum number of samples that can be stored and re-use across learning session. **GDumb** algorithm allocates an equal number of samples for each learning session. When facing a new class, it gradually removes samples from the memory belonging to the class with the maximum number of samples, substituting them with samples from the new class. A visualization of this process is shown in Fig. 4.8.

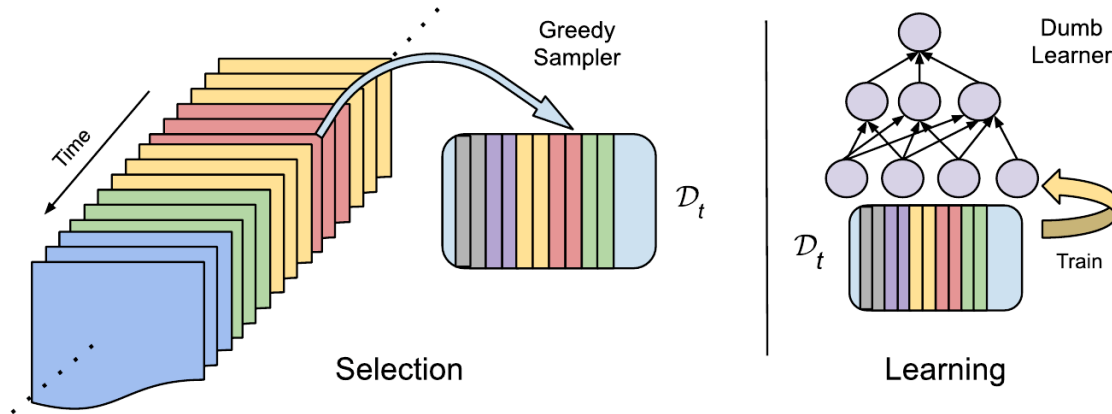


Figure 4.8 – Visualization of GDumb learning strategy process. Adapted from Prabhu et al. 2020.

For the pseudo-rehearsal approach, instead, we evaluate the **Dream Net** strategy (Mainsant et al. 2021) developed by our collaborators in Grenoble. Dream Net requires an output layer composed of several neurons corresponding to the input (Auto-associative or Auto-encoder part) and several neurons corresponding to the number of classes (Hetero-associative or part). Dream Net architecture learning process can be divided into three phases:

1. The *Learning Net* at each learning session, learns real features from the class of the session and pseudo-features from the previously learned classes.
2. The *Learning Net* at the end of the learning session transfers its weights to *Memory Net*
3. *Memory Net* captures the learned function using a re-injection sampling procedure. The re-injection sampling procedure consists of the following steps: inject a random noise input vector and re-inject the replication vector obtained at the output of the auto-associative part of *Memory Net* at its input and so on. At each re-injection, Auto and Hetero associative outputs of *Memory Net* are conserved to create pseudo-examples. After several re-injection, a pseudo-examples database is obtained that contains pseudo-features and corresponding pseudo-labels obtained after each re-injection (data from the first inference is not kept).

A visualization scheme of the Dream Net learning procedure can be found in Figure 4.9.

The regularization methods have been developed to retain the most *important weights* while learning new classes. Their bio-inspiration relies on the hypothesis that continual learning relies on task-specific synaptic consolidation, making certain synapses less plastic and stable over time (Clopath 2012). For instance, experiments with mice demonstrate that a strengthening of excitatory synapses occurs at new skill acquisition (Yang et al. 2009), leading to an increased volume of specific spines. The increased volume persists despite the subsequent learning of new tasks and is associated with the persistence of performance of the very first learned task several months later. When these spines are removed the task is forgotten (Cichon et al. 2015; Hayashi-Takagi et al. 2015, panel (a) Fig. 4.7). Among the ANNs regularization methods, we consider Elastic-Weight-Consolidation (**EWC**), Synaptic Intelligence (**SI**) and Learning without Forgetting (**LwF**) strategies (Kirkpatrick et al. 2017; Zenke et al. 2017; Li et al. 2017b). EWC and SI are *structural* regularization methods (Li et al. 2017b) that constrain relevant weights to stay close to their old values. The major difference between these two approaches is given by the estimation of important weights. EWC relies on an offline estimation of the Fisher information matrix while SI proposes an online computation of the importance of a synapse being proportional to the product of its weight and the activity of the post-synaptic neurons.

EWC penalizes excessive movements in learning new task from the just obtained optimal weights. At each learning session t , a regularization term is added to the loss function $\mathcal{L}_{t\mathbf{w}}$ constraining the new learned weights to stay close to the previous learned one:

$$\mathcal{L}_{t\mathbf{w}} = \mathcal{L}_{|\mathcal{D}_t} + \sum_c \frac{\lambda}{2} F_c(\mathbf{w}_c - \hat{\mathbf{w}}_{c_{t-1}})^2 \quad (4.6)$$

where $\mathcal{L}_{t\mathbf{w}}$ is the loss function at the learning session t with weights parameters \mathbf{w} , $\mathcal{L}_{|\mathcal{D}_t}$ is the loss function evaluated on the dataset associated to the learning session t , c is an enumeration of the weights, λ is the regularization hyperparameter assessing the importance of previously seen task with respect to the new one, F_c is the Fisher matrix at session $t - 1$ corresponding to an estimation of the importance of each parameter in solving the optimization problem over \mathcal{D}_{t-1} and $\hat{\mathbf{w}}_{t-1,c}$ is the optimal c -th weight determined at session $t - 1$. The Fisher information matrix

$$F = \mathbb{E} \left[\left(\frac{\partial}{\partial \mathbf{w}} \log f(\mathcal{X}; \mathbf{w}) \right)^2 \middle| \mathbf{w} \right]$$

is computed at the end of each learning session. SI proposes a regularization parameter in the loss function of each learning step by introducing a surrogate loss whose minimum is reached for the previously determined optimal weights parameters. Specifically, the actual loss function becomes

$$\mathcal{L}_{\mathbf{w}_t} = \mathcal{L}_{|\mathcal{D}_t} + \psi \sum_c \Omega_c^t (\mathbf{w}_c - \hat{\mathbf{w}}_{c_{t-1}})^2 \quad (4.7)$$

where ψ is an hyperparameter corresponding to the influence of the previously learned tasks, $\hat{\mathbf{w}}_{c_{t-1}}$ is the previously determined optimal parameters and Ω_c^t is the proposed

new estimation of parameters importance which is updated online as follows:

$$\Omega_c^t = \sum_{t' < t} \frac{\omega_c^{t'}}{(\Delta_c^{t'})^2 + \xi} \quad (4.8)$$

$\omega_c^{t'}$ is the parameter specific contribution to changes in the total loss, obtained by summing at each weight update iteration of the training phase the product of the individual gradient $\frac{\partial \mathcal{L}}{\partial \mathbf{w}_c}$ and the parameter update. This corresponds to write the change of the loss function over the entire training phase as sum of infinitesimal changes across each iteration per individual parameter. $\Delta_c^{t'}$ corresponds instead to the difference between the parameter value at two consecutive iterations, which is simply given by the gradient descent rule, and ξ is introduced to ensure a non-zero denominator.

While EWC and SI directly operate by regularizing the model parameters, *functional* regularization approaches, such as the LwF strategy, penalize changes in the input-output function of the neural network. This is achieved by constraining the predictions obtained by applying the current model to remain close to the predictions obtained by applying the previously learned model to samples from the current dataset \mathcal{D}_t .

Finally, the third class of methods which dynamically update the network architecture are inspired by neurogenesis happening in adult mammalian brain when learning new information (Kempermann et al. 2004; Aimone et al. 2009 (c) panel Fig. 4.7). These methods can dynamically increase the architecture model by adding neurons, layers or sub-networks, or instead freezing part of the model while learning over the other in static conditions. Architecture-based approaches have the main issue of scalability, since the architecture continues to expand while more tasks are added (Parisi et al. 2018; Yoon et al. 2017; Draelos et al. 2017). The LwF strategy requires neurogenesis: the model architecture is updated by the addition of a node in the output layer corresponding to the new observed class, together with the inclusion of all the connections from the last hidden layer to the output one.

The LwF loss function can be separated in two parts: \mathcal{L}_{old} and \mathcal{L}_{new} with the influence of previously seen tasks being weighted by the hyperparameter α_0 . The equation 4.9 is given by :

$$\mathcal{L}_t = \alpha_0 \mathcal{L}_{old}|_{\mathcal{D}_t}(y_i^t, f^{t-1}(x_i^t)) + \mathcal{L}_{new}|_{\mathcal{D}_t}(y_i^t, f^t(x_i^t)) \quad (4.9)$$

The difference between the current model and the previous one is given by the addition of a node in the output layer corresponding to the new observed class, together with the inclusion of all the new parameters from the previous layer and the output one. \mathcal{L}_{old} is computed using the parameters learned during the session $t - 1$ for the sample belonging to the current dataset \mathcal{D}_t and corresponds to

$$\mathcal{L}_{old}(f^{t-1}(x_i^t), y_i^t) = - \sum_{t'=1}^t \Upsilon(f^{t-1}(x_i^t)) \log \Upsilon(f^t(x_i^t)) \quad (4.10)$$

which is the sum over the available labels of the difference between the recorded output from previous model and the current output, rescaled by the Υ function as

$$\Upsilon(f^k(x_i^k)) = \frac{(f^k(x_i^k))^{1/\zeta}}{\sum_{t'=1}^k f^{t'}(x_i^{t'})} \quad (4.11)$$

where ζ is a hyperparameter regulating the smoothness of the classes probability distribution. \mathcal{L}_{new} is a chosen loss function computed for the dataset \mathcal{D}_t .

Other architecture-based strategies propose to freeze part of the neural network to force the appearance of specialized sub-networks (Rusu et al. 2016; Turner et al. 2021; Draelos et al. 2017).

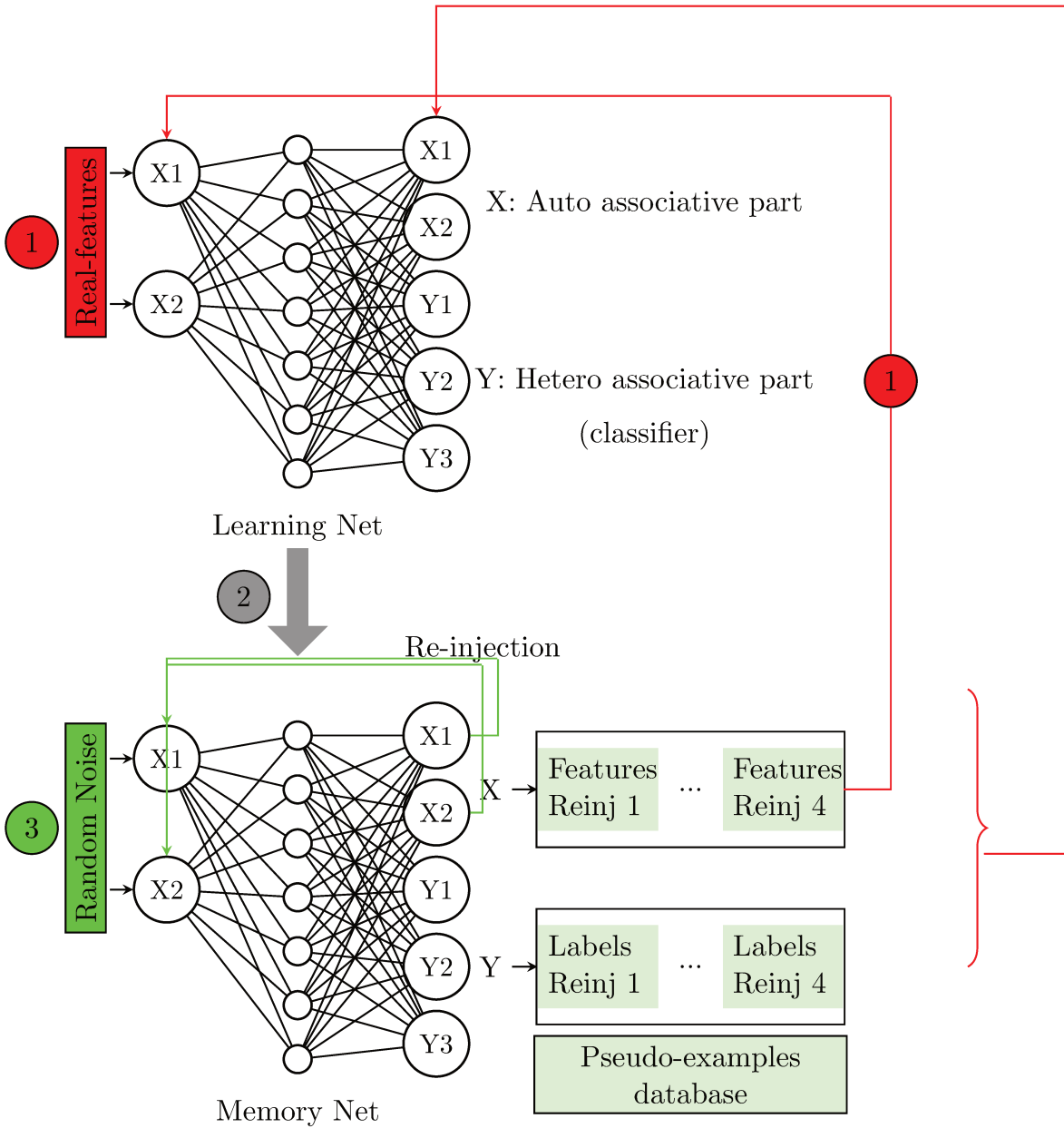


Figure 4.9 – Architecture scheme of Dream Net model. The Dream Net model learning procedure can be divided into 3 phases indicated as 1, 2, and 3.

4.3 Network Science in ANNs

Since the development of ANNs and over the years, there has been an interest in using insights from cognitive neuroscience to improve their performances (Marblestone et al. 2016; Hassabis et al. 2017; Khacef et al. 2018; McCulloch et al. 1943; Hebb 2005; McClelland et al. 1986). Particularly, artificial intelligence practitioners can find inspiration from biological intelligence and natural mechanisms for the development of new models (Hassabis et al. 2017).

However, despite the seminal bio-inspiration in ANNs design, many tools introduced for brain information processing investigation have not been fully used to study ANN behaviors. Among these, graph modeling has shown to be a powerful framework widely used in neuroscience to study brain structural and functional connectivity (Petersen et al. 2015; Bullmore et al. 2009; Wang et al. 2010; Sporns 2022; Barabási 2013).

In this section, we review related works about network science applications in ANN models (comprehensive reviews be found in Chung et al. 2023; Kaviani et al. 2021). Most part of the previous works in this direction concerns the *structure* of the ANN itself, i.e. its architecture, to achieve faster convergence and better results. Large part of the literature focuses on Hopfield neural network (Hopfield 1982; Stauffer et al. 2003; Kaviani et al. 2021), which is a particular ANN composed of a set of neurons all connected in a regular structure and which simulates memory. On the other hands, concerning feed forward neural network, many works explore the use of architectures characterized from a network science point of view. For instance, various works have explored the possibility of connecting neurons in ANN to build small-world, scale-free or sparse ANN instead of traditional feed forward network with the objective of improving performances, having faster convergence or reducing the model dimension.

Small-world neural network was first proposed in Simard et al. 2005. The author trained different layered neural network after rewiring a different number of connections between neurons making the architecture transit from a regular structure to a random one, passing by the small-world architecture and evaluate the effect of introducing long-jump links between neurons of different layers. The influence of complex topology in artificial neural network performance was later on explored in different applications, but remain a controversial topic (Erkaymaz et al. 2012; Erkaymaz et al. 2014; Jiang 2009; Zheng et al. 2010). Scale-free topology was instead applied and evaluated across multiple datasets in Mocanu et al. 2016.

Regarding work for network characterization, recently complex network tools have been used for the analysis of both trained and untrained ANNs. In La Malfa et al. 2021; La Malfa et al. 2022, the authors apply complex network theory to deep ANNs of different types (feed forward neural network, convolutional NN, etc.). Their ANNs characterization is based on metrics distribution over weighted directed graphs.

A motif discovery process has been proposed in Zambra et al. 2020, where the authors

characterize multi-layer perceptrons during the learning phase by detecting different patterns of connection of groups of four or five units. Their method is used to compare different weights initialization in multi-class classification. While the proposed framework has promising results, the motif search can become computationally expensive while the dimension of ANN increases (Masoudi-Nejad et al. 2012; Patra et al. 2020). In their application, the ANN size appears quite small (30-20 units and 3 layers).

Next, the work of Scabini et al. 2021 extends complex network techniques to detect different neuron types and to relate their presence to the performance of fully connected neural networks. Their framework considers only weighted graphs and an Offline learning setting.

A different graph model definition is proposed in Hanczar et al. 2020, where the ANN decision process for a binary classification task is based on the definition of a *relevance network* per class, obtained through the computation of the layer-wise relevance propagation (LRP) score (Bach et al. 2015). While LRP has been defined and it is mainly used to determine input features contributing to the final classification, the authors originally propose to use the class relevance network as a human-understandable decision process. Their approach associates each unit with the final decision, providing a human-understandable decision process, but requires expert knowledge insertion.

Finally, a new approach has been proposed in Corneanu et al. 2019 by the definition of a *functional network* obtained by the computation of Pearson correlation among activation units. Multiple instances of functional networks across training epochs are then compared using topological metrics in order to assess the model evolution during the learning process. In their approach, the ANN model is considered as a structural network and its weights and architecture are not considered.

In the next chapters, we present our collaborative work on the use of graph to study ANN connectivity and explore different continual learning strategies.

Chapter 5

Graph-Based Analysis of ANN Connectivity

Based on a paper written in collaboration with Dwight Nwagwe, Marion Mainsant, Raphael Bayle, Marina Reyboz, Martial Mermillod, Michel Dojat, and Sophie Achard. Under Review. [Carboni et al. 2023d](#)

In this chapter, we present our original graph-based analysis model for ANN connectivity. This work aims to bridge the gap between neuroscience and artificial neural networks by introducing a graph modeling framework, a powerful tool widely used in studying brain structural and functional connectivity. Our approach, largely inspired by a brain framework analysis, introduced the definition of ANN functional connectivity at rest and a corresponding induced-graph definition. Thus, we introduce graph-informed features for the characterization architecture at different learning sessions and graph-informed pruning and weight-injection methods. Finally, we present the adopted evaluation metric and their interpretation in terms of plasticity and stability of artificial models. This proposal lays the foundation for the development of a graph-features-informed learning strategy. At the end of the chapter, we discuss relevant related works and limitations.

Contents

5.1	Introduction	126
5.2	Graph-based method for ANN connectivity analysis	126
5.2.1	Activation Network and Induced Graph	127
5.2.2	Graph statistics of interest	129
5.2.3	Graph-informed Features	129
5.2.4	Network Surgery: Nodal Role Identification	131
5.2.5	Graph-informed Weight Injection	131
5.3	Evaluation Metrics	133
5.4	Related Works and Limitations	134

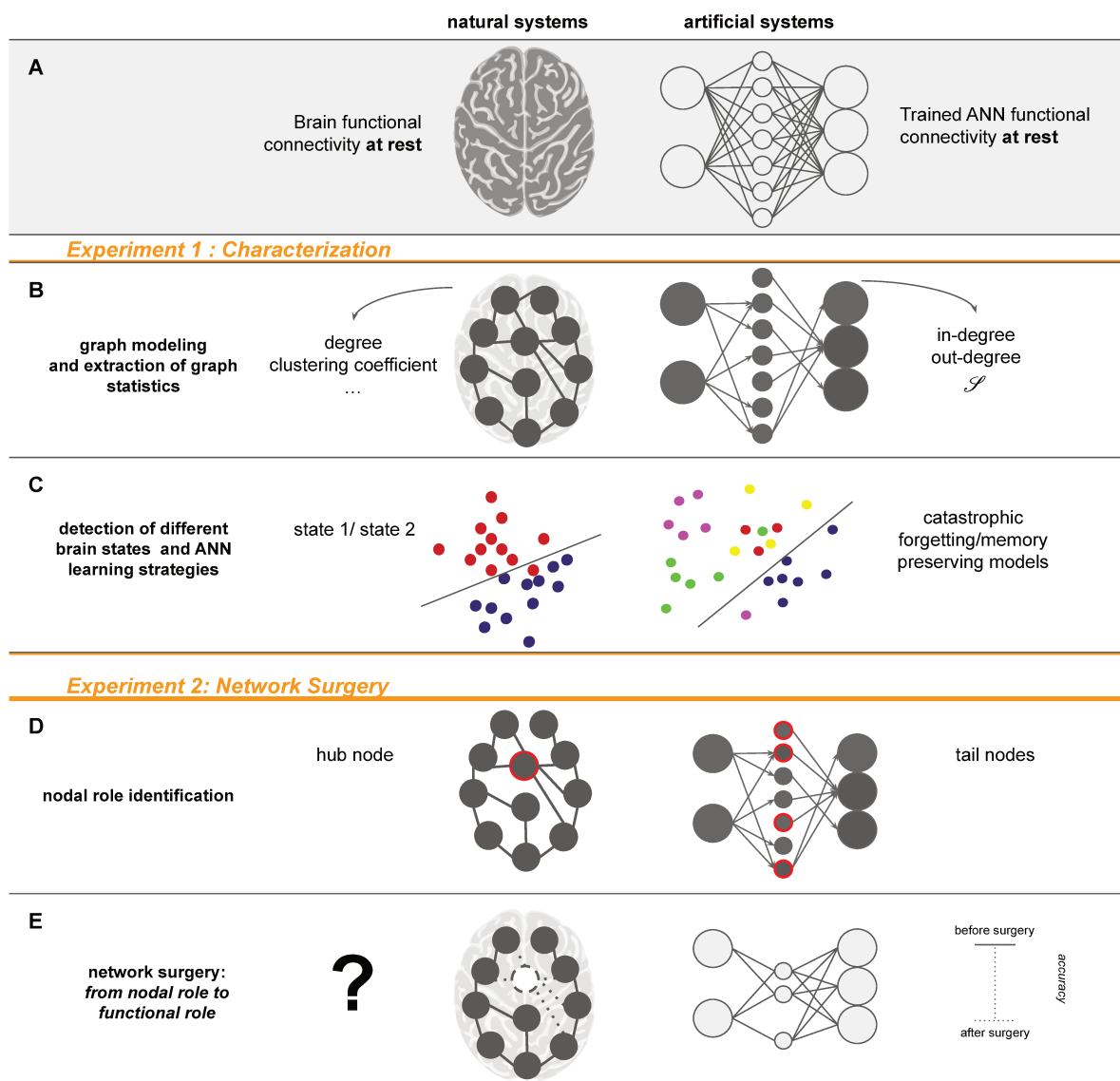


Figure 5.1 – Overview of the brain-inspired perspective of the proposed graph-based analysis of ANNs. (A) Inspired by the concept of brain functional connectivity at rest, we propose to investigate the equivalent functional connectivity at rest of trained artificial systems. (B) Graph modeling can be applied for both brain and ANN connectivity modeling. Thus, different graph statistics can be extracted to characterize the resting state network. (C) Similarly to brain functional connectivity studies aiming at differentiating between different brain states (for instance healthy vs pathological conditions), by the use of the considered graph statistics, we show how the graph statistics can be used as features for the identification of ANN in good learning conditions or affected by the catastrophic forgetting phenomenon, or even to identify the learning strategy used. (D) At a finer level of a single-state characterization, nodal roles are investigated to determine the presence of nodes having specific roles in the network. In natural systems, nodal role discovery aims at identifying the hub nodes (nodes with a high number of connections). In artificial systems, we propose to consider the group of units belonging to the tail of the considered graph statistics distribution. (E) While in the brain functional role identification can only be achieved by the identification of nodal role in the connectivity network, in ANNs we can perform a network surgery by turning off specific sets of units to evaluate the change in the system performance. This allows associating the network nodal role with a functional role.

5.1 Introduction

Despite the seminal bio-inspiration in ANNs design, many tools introduced for brain information processing investigation have not been fully used to study ANN behavior. Among these, graph modeling has shown to be a powerful framework widely used in neuroscience to study brain structural and functional connectivity (Petersen et al. 2015; Bullmore et al. 2009; Wang et al. 2010; Sporns 2022; Barabási 2013). In this chapter, we propose to explore how such a framework can be elegantly and interestingly used to investigate the ANN properties and particular behaviors. A conceptual visualization of our proposal can be found in Fig. 5.1.

As a case of study, we concentrate on the sequential learning process where ANNs are trained on an ordered series of tasks as shown in Fig. 4.4 (Buzzega et al. 2020; Hadsell et al. 2020). This process mimics how the brain continuously learns and adapts to new tasks (Milgram et al. 1987; Pascual-Leone et al. 2005). As neuroscientific literature reports, brain connectivity changes are associated with new learning tasks (Casimo 2018; Vico Fallani et al. 2010; Zouridakis et al. 2007), we explore, using graph modeling, the corresponding ANN connectivity changes. In particular, we investigate how specific graph statistics are modified during a continual learning framework and the conditions that led to catastrophic forgetting (i.e. the performance of previously learned tasks dramatically decreases when new tasks are learned in a sequential manner by stopping training on task A while beginning training on task B, Fig. 4.4 Panel B).

An overview of the brain-inspired perspective of the proposed graph-based analysis of ANNs is proposed at the beginning of the chapter.

In the following, we refer to an architecture, a learning strategy, a fixed order of training sessions, and the general classification task associated with the database as a **configuration**.

All our graph modeling analysis of ANNs across sequential learning sessions aims to extract information on how the network adapts to new tasks and how it preserves knowledge of previous tasks from a connectivity point of view. This can potentially provide insights into the plausible neural mechanisms underlying continual learning in the biological neural networks and in reverse inspire the design of more efficient and biologically-plausible continual learning artificial systems.

5.2 Graph-based method for ANN connectivity analysis

From a trained ANN, we build a corresponding graph model starting from its architecture: the units (i.e nodes) present in each layer and the presence and orientation of edges among the units. Indeed, we extract the *activation network at rest* (see Def. 5.1), similar to a brain resting-state connectivity analysis (van den Heuvel et al. 2010),

by feeding to a trained ANN an input sample of 1-entries. Then, we perform a graph filtering procedure to determine the most active units and the strongest connections (**B** panel Fig. 5.1). This results in an induced graph that is oriented from the input layer to the output layer, and its structure is encoded by its adjacency matrix $A = (a_{nm})$.

For each configuration, we extract as many induced resting-state graphs as learning sessions.

Unlike biological connectivity graphs, the feedforward ANN is a partite graph with a fixed structure where edges exist only between consecutive layers.

5.2.1 Activation Network and Induced Graph

We consider an ANN model to be uniquely identified by an architecture graph and a synaptic weight function which determines the parameters of each artificial neuron by associating a weight to each edge.

The weights are uniquely determined at the end of the training process, we denote $\mathbf{w}_{i,j}^{l-1,l}$ the weight associated with the edge connecting the i -th node of layer $l - 1$ and the j -th node of layer l . Note that the input layer corresponds to $l = 0$.

At each input sample x processed by the ANN model, we can associate an activation network defined by the computation of the sequence of linear and nonlinear transformations applied to the input and associate to each unit, its activation value σ and to each edge, the resulting value of the corresponding transformation. In biological terms, the activation network can be thought of as the map of the brain response to a stimulus x . In particular, each unit i in layer l is associated with an activation u_i^l :

$$u_i^l = \sigma\left(\sum_j \mathbf{w}_{j,i}^{l-1,l} u_j^{l-1} + b_i\right) \quad (5.1)$$

where σ is the activation function and b is a fixed bias term, \mathbf{w} and b are the parameters of the artificial neuron function.

Definition 5.1. Inspired by the concept of brain resting-state connectivity analysis, where the brain activation map is determined at mind-wandering, we define the **activation network at rest** feeding to the ANN a vector of ones to simulate the process of a *non-task-related-information*.

With this procedure, the ANN is not engaged in any specific cognitive task, and we can assess the intrinsic functional organization of the artificial system. The activation network at rest results in a set of artificial neurons that are spontaneously active and functionally connected with each other in the absence of true external inputs. The vector of ones can serve as a simple way to activate the network at rest, by providing a constant input to all the nodes, yet different choices (random noise, periodic signals, average of the considered data, ..) can be envisaged in future explorations. Thus, for the units in the input layer, we fix $u_i^0 = 1$. The weights of the edges in the following layers of the activation network at rest are given by

$$\tilde{w}_{i,j}^{l-1,l} = \mathbf{w}_{i,j}^{l-1,l} u_i^{l-1} \quad (5.2)$$

Hence, for the first hidden layer, we have $\tilde{w}_{i,j}^{0,1} = \mathbf{w}_{i,j}^{0,1} u_i^0 = \mathbf{w}_{i,j}^{0,1}$ and for the subsequent layer, we simply apply (5.2) again with u_i^1 the output of the first layer. Even if for the first layer $\tilde{w}_{i,j}^{0,1}$ equals in value $\mathbf{w}_{i,j}^{0,1}$, they formally correspond to two very distinct concepts, being \mathbf{w} the parametric synaptic weights of the artificial neural network function and \tilde{w} the weights function associated to the edges of a graph.

Finally, we perform a graph filtering procedure for both the nodal features - to determine the most active units - and edges weights - to determine the strongest connections. As final results of this procedure, we obtain a binary directed graph, i.e. every two nodes are either connected either disconnected with edges direction going from the input layer to the output. Note that the number of nodes and edges in the activation network only depends on the architecture graph of the model which is fixed for all configurations. The graph filtering procedure can be performed by choosing the number of units to extract in each layer, and the desired total number of edges with respect to a given criterion (for instance to observe particular graph properties). This corresponds to determine a weight threshold WT such that the number of edges whose graph weights are greater in absolute value of WT equals the chosen graph sparsity. Similarly, on the nodal features, we can filter out units whose activation is not greater than an activation threshold AT.

Definition 5.2. We define the **induced graph** obtained by filtering the activation network at rest of a trained neural network as $\mathcal{G}(\text{NN})$

$$\mathcal{G}(\text{NN}) = \mathcal{G} = (\mathcal{V}, \mathcal{E}) : \quad (5.3)$$

$$\mathcal{V} = \{(l, i) \text{ s.t. } i \in \{1, \dots, H(l)\}, l = 0 \text{ or } (l \in \{1, \dots, L\} \text{ and } u_i^l > \text{AT})\} \quad (5.4)$$

$$A = (a_{(l-1,i),(l,j)})_{\substack{(l-1,i),(l,j) \in \mathcal{V} \\ l \in \{1, \dots, D\}, \\ i \in \{1, \dots, H(l-1)\}, \\ j \in \{1, \dots, H(l)\}}} = \begin{cases} +1 & \text{if } |\tilde{w}_{i,j}^{l-1,l}| > \text{WT} \\ 0 & \text{otherwise} \end{cases} \quad (5.5)$$

where A is the adjacency matrix (Fig. 5.2), L the depth of the ANN and $H(l)$ the number of units in layer l .

Note that after applying the graph-filtering procedure, some nodes can be disconnected from the resulting graph and the existence of a unique connected component can be guaranteed by the addition of one single edge. However, in our simulations, a unique connected component was always observed with some disconnected units whose weights were not significant in absolute values to be included in the resulting induced graph.

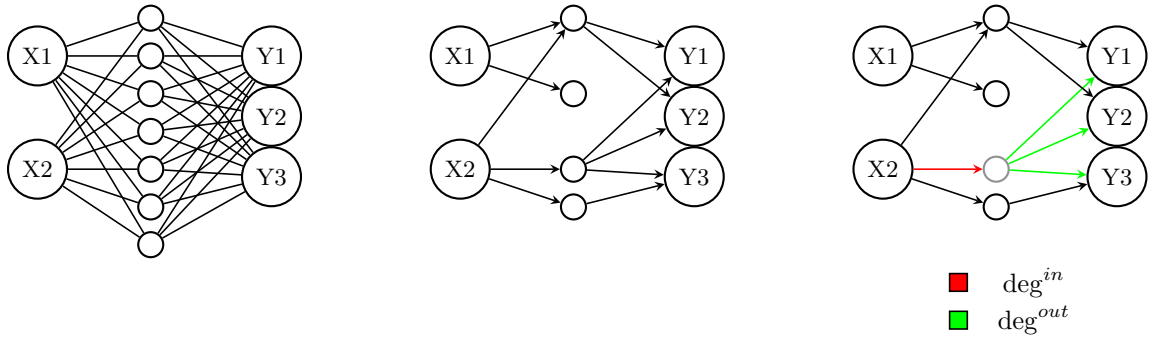


Figure 5.2 – A toy example of an induced graph. Left: NN architecture graph. Center: The induced graph at rest. Right: Visualization of the statistics of interest for the gray node.

5.2.2 Graph statistics of interest

The fixed structure of the graphs we analyze only allows edges between two consecutive layers. As a result, graph statistics like diameter or centrality measures, which detect the graph shape, are not expected to provide valuable additional insights when comparing the same architecture trained with different strategies. Hence, we focus on in-degree and out-degree of units in the hidden layers, defined in Eq. 5.7. These statistics respectively count the number of incoming and outgoing edges, capturing the amount of information flow through each unit (See Fig. 5.2). Thus, the value in minus out degree can be used to rank nodes in the hidden layer. In the following, we omit the dependency on l and only indicate with i the node (l, i) .

Definition 5.3.

$$\deg_{\text{in}}(i) = \sum_{j \in \{1, \dots, H(l-1)\}} a_{(l-1, j), (l, i)} \quad (5.6)$$

$$\deg_{\text{out}}(i) = \sum_{j \in \{1, \dots, H(l+1)\}} a_{(l, i), (l+1, j)} \quad (5.7)$$

In particular, we consider the quantity

$$\mathcal{S}(i) = \deg_{\text{in}}(i) - \deg_{\text{out}}(i) \quad (5.8)$$

and we distinguish two types of nodes: nodes belonging to the interquartile range of \mathcal{S} and nodes in the tail of \mathcal{S} . An example, with \mathcal{S} having average in zero, can be visualized in Fig. 5.3

5.2.3 Graph-informed Features

We enumerate the T learning sessions, and notate *base* the first one, at the end of each training session we determine an induced graph. Thus, for each configuration (an architecture, a learning strategy, a database, and the learning order), we extract T induced graphs. In the following we include the subscript t corresponding to the

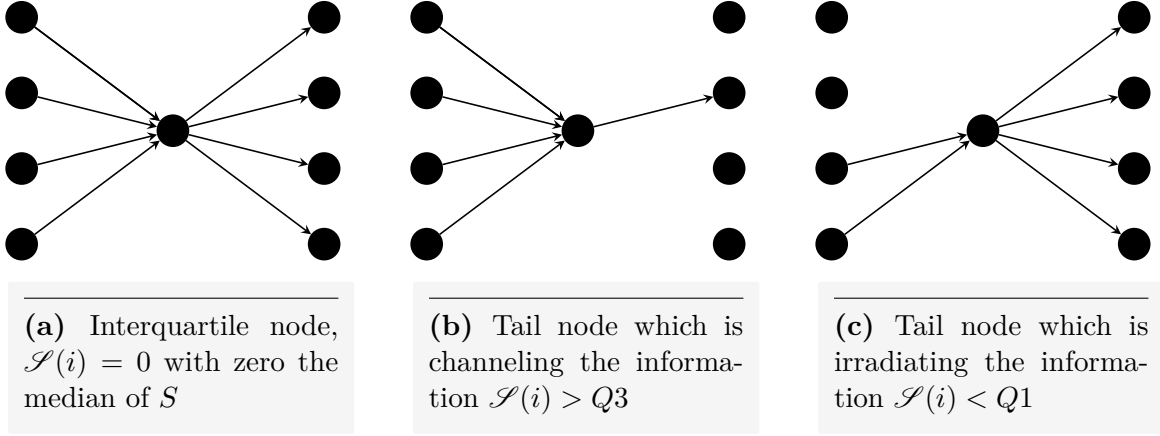


Figure 5.3 – Examples of interquartile and tail nodes.

learning session, for the induced graph extracted at session t we consider the statistics \mathcal{S}_t . Moreover, we define the following graph features.

Definition 5.4. We define the Maximum Unit Change in Equation 5.9.

$$\text{Maximum Unit Change} = \max_{\substack{i \in \{1, \dots, H(l)\} \\ t \in \{1, \dots, T\}}} |\mathcal{S}_t(i) - \mathcal{S}_{t-1}(i)| \quad (5.9)$$

where $H(l)$ is the number of units in layer l .

Next, we determined the number of units in the tail of S_t at each learning session, given by v_t in Equation 5.10.

$$v_t = \sum_{i \in \{1, \dots, H(l)\}} \chi_{\{j | \mathcal{S}_t(j) < Q1_t\}}(i) + \chi_{\{j | \mathcal{S}_t(j) > Q3_t\}}(i) \quad (5.10)$$

$Q1_t, Q3_t$ are the first and third quartiles of $\mathcal{S}_t(j)$ distribution.

In general, we can expect $\mathcal{S}_t(i)$ to be positive. Yet, a node hosting a negative flow may exist. It corresponds to a neuron receiving little information and spreading it irrespective to all output units. Vice versa, a very high $\mathcal{S}_t(i)$ value represents a neuron aggregating multiple information into a few output units. The extracted graph feature Maximum Unit Change represents the maximum consecutive change in the flow of one hidden neuron across two learning sessions. Each v_t quantifies the number of units in the hidden layer whose flow does not belong to its interquartile range.

Definition 5.5. Next, we define the Consecutive Tail Persistence as the percentage of tail units in the tails of consecutive learning sessions (i.e. in the tail of both \mathcal{S}_{t-1} and

\mathcal{S}_t).

Definition 5.6. Finally, we define the Tail Persistence as the percentage of units which are in the tail of \mathcal{S}_1 after the learning session of the *base* task and in the final distribution \mathcal{S}_T after all learning sessions.

5.2.4 Network Surgery: Nodal Role Identification

We are interested in determining the behavior of hidden units according to their value in $\mathcal{S}_t(i)$. In particular, for each model and configuration NN, we notate the ordered list of trained networks as $\text{NN} = (\text{NN}^1, \dots, \text{NN}^t, \dots, \text{NN}^T)$. At each learning session t , we extract the associated induced graph and compute for each hidden unit i the statistics $\mathcal{S}_t(i)$. Thus, we compute the first and third quartiles $Q1_t, Q3_t$. We denote $\mathbf{w}_{i,j}^{l,l+1}(t)$ the synaptic weights between the i -th unit of layer l and the j -th unit of layer $l+1$ of the model NN^t obtained after the t -learning session, L is the depth of the NN and H_A the number of units in the auto-associative part of the output layer, if any. We propose the following definitions.

Definition 5.7. We define a version $\text{NN}^\tau = (\text{NN}^{1,\tau}, \dots, \text{NN}^{T,\tau})$ which only preserves the weights of units outside the interquartile in-out degree interval, namely in the *tail*.

$$\text{NN}^{t,\tau} : \mathbf{w}_{(i),(l+1,j)}^{t,\tau} = \begin{cases} 0 & l = L - 1 \text{ and } H_A \leq j \leq H_A + T \text{ and} \\ & i \in \{k \mid Q1_t < \mathcal{S}_t(k) < Q3_t\} \\ \mathbf{w}_{i,j}^{l,l+1}(t) & \text{otherwise} \end{cases} . \quad (5.11)$$

Definition 5.8. Second, a complementary version of the previous one, which discards the weights of units in the tail. We notate this version $\text{NN}^{\bar{\tau}} = (\text{NN}^{1,\bar{\tau}}, \text{NN}^{2,\bar{\tau}}, \dots, \text{NN}^{T,\bar{\tau}})$. $\text{NN}^{t,\bar{\tau}}$:

$$\mathbf{w}_{(i),(l+1,j)}^{t,\bar{\tau}} = \begin{cases} & l = L - 1 \text{ and} \\ \mathbf{w}_{i,j}^{l,l+1}(t) & H_A \leq j \leq H_A + T \text{ and} \\ & i \in \{k \mid Q1_t < \mathcal{S}_t(k) < Q3_t\} \\ 0 & \text{otherwise} \end{cases} . \quad (5.12)$$

A schematic visualization of these pruned versions is shown in Fig. 5.4.

5.2.5 Graph-informed Weight Injection

We propose three procedures across consecutive sessions to recall previously learned classes: a forward weights injection without fine-tuning, backward weights injection without fine-tuning and a weight constraint with fine-tuning.

We denote $\tau^1, \tau^2, \dots, \tau^T$ the hidden units in the tails of the distribution of $\mathcal{S}_t(i)_{t=1,\dots,T}$. Given a NN, we define three new ordered lists, as follows:

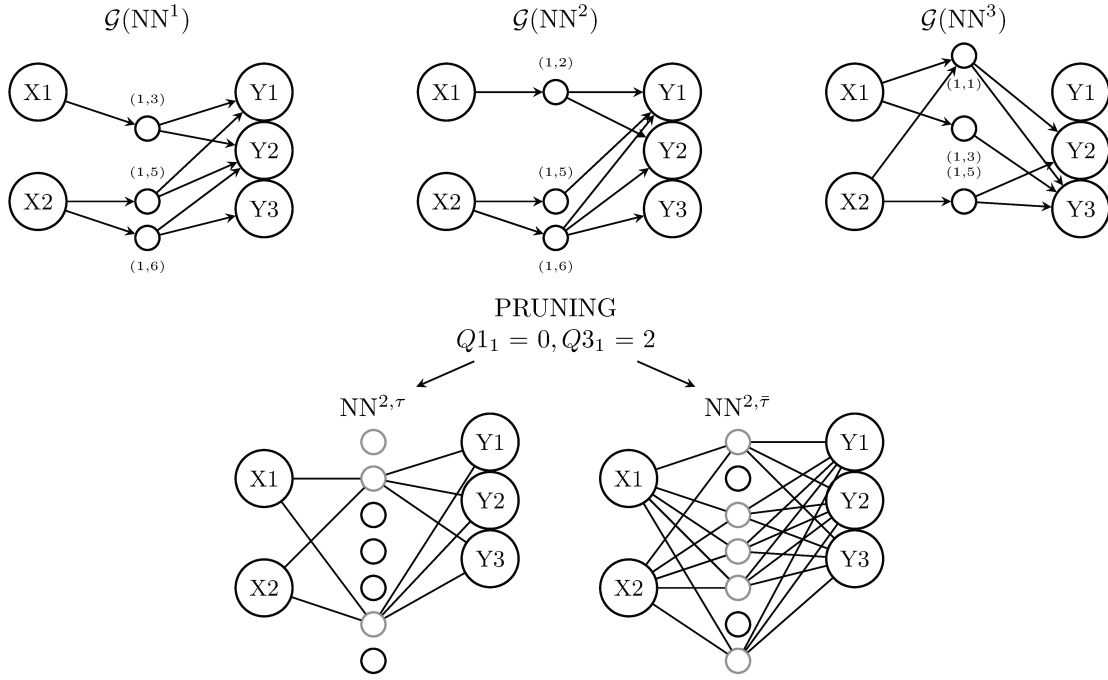


Figure 5.4 – Schematic visualization of the pruning versions. Gray units correspond to the not-pruned ones.

Definition 5.9. The forward weights injection procedure

$$\text{NN}^{\text{FWI}} = (\text{NN}^{1,\text{FWI}}, \dots, \text{NN}^{T,\text{FWI}})$$

$$\text{NN}^{t,\text{WI}} : \mathbf{w}_{(i),(l+1,j)}^{t,\text{WI}} = \begin{cases} \mathbf{w}_{i,j}^{l,l+1}(t-1) & (i) \in \tau^{t-1} \\ \mathbf{w}_{i,j}^{l,l+1}(t) & \text{otherwise} \end{cases} .$$

Definition 5.10. the backward weights injection procedure

$$\text{NN}^{\text{BWI}} = (\text{NN}^{1,\text{BWI}}, \dots, \text{NN}^{T,\text{BWI}})$$

$$\forall t > 1, \text{NN}^{t,\text{BWI}} : \mathbf{w}_{(i),(l+1,j)}^{t,\text{BWI}} = \begin{cases} \mathbf{w}_{i,j}^{l,l+1}(t) & (i) \in \tau^t \\ \mathbf{w}_{(i),(l+1,j)}^{t-1,\text{BWI}} & \text{otherwise} \end{cases} .$$

Definition 5.11. the weights injection with fine-tuning procedure is defined as

$$\text{NN}^{\text{WIF}} = (\text{NN}^{1,\text{WIF}}, \dots, \text{NN}^{T,\text{WIF}})$$

in which at each step the hidden units in the tail are detected and their weights frozen, before training in the new task as shown in Fig. 5.5.

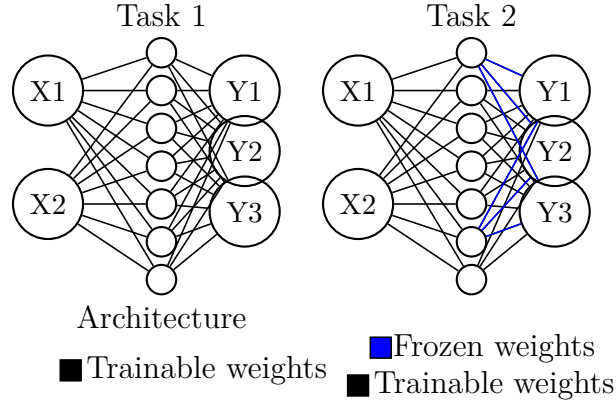


Figure 5.5 – Weights injection with fine-tuning procedure. At each learning session, the weights among hidden units in the tails and the output units in the hetero associative part are frozen. Thus, the model is trained for the new task.

5.3 Evaluation Metrics

Since the occurrence of catastrophic forgetting is related to the stability-plasticity dilemma, we propose to evaluate each architecture and strategy on these two properties by the estimation of the following metrics based on the accuracy of the different learned classes. These metrics have been proposed in [Kemker et al. 2018](#), here we originally interpret them with the stability and plasticity properties

Definition 5.12.

$$stability := \Omega_{\text{base}} = \frac{1}{T-1} \sum_{t=2}^T \frac{\alpha_{\text{base},t}}{\alpha_{\text{ideal}}} \quad (5.13)$$

$$plasticity := \Omega_{\text{new}} = \frac{1}{T-1} \sum_{t=2}^T \alpha_{\text{new},t} \quad (5.14)$$

where T corresponds to the total number of learning sessions, $\alpha_{\text{new},t}$ is the test accuracy for the class immediately learned at session t , $\alpha_{\text{base},t}$ is the test accuracy of the class learned during the first session (base set) after t new learning sessions and α_{ideal} is the offline method accuracy on the base set, which can be assumed to be the ideal performance. To determine α_{ideal} we train the architecture in one single learning session with a unique dataset including the entire classes.

In continual learning, we notate local accuracy the accuracy of the last learned class, and global accuracy at each learning session as the accuracy of all the seen classes.

We quantify the *stability* with Ω_{base} that measures the ability to retain the class learned during the first session (base session), after learning the successive sessions; and the

plasticity with Ω_{new} that measures the performance in learning a new task (See B-C panels in Fig. 4.4). Unless a model outperforms the offline model accuracy on the base set α_{ideal} , *stability* and *plasticity* vary between $[0, 1]$. Note that the occurrence of the catastrophic forgetting phenomenon is directly quantified by the *stability* metric: low *stability* implies forgetting.

5.4 Related Works and Limitations

Induced-graph definition. The proposed graph-based analysis is based on a unique induced graph definition, while for its estimation it does not require any input data, by essentially propagating the weights themselves across the different layers, the meaning of the induced graph is not explored with respect to the learned function. Other ways to define the graph need to be explored in future work, together with the exploration of the relation of the extracted induced graph and the neural network model from a statistical-based point of view. In particular, in our work in adversarial attacks (Appendix B.4), we explored a different way to define the induced graph, by associating to each sample a thresholded network where each edge (i.e. each synaptic weight) is associated to its influence on the output. This is achieved by applying the layerwise relevance propagation LRP (Bach et al. 2015, Montavon et al. 2019). LRP offers the possibility of a more detailed analysis of the information flow per sample within the neural network from a graph perspective. Alternatively, an average graph could be used to associate each configuration and learning session with a unique activation graph. However, the representative power of such a graph might be a concern.

Similarly, other graph definition, such as the functional network proposed in Corneanu et al. 2019 also merit comparison and exploration. These alternative approaches could provide valuable insights into the behavior and performance of the neural network in different scenarios.

Relation with hard-pruning. Our thresholding procedure to determine the induced graph is similar to a magnitude-based pruning process (Kim et al. 2015; Han et al. 2015b; Han et al. 2015a), where low magnitude weights are turned-off in neural network models to obtain smaller number of parameters. The main difference is given by the fact that our process requires the propagation of an input vector from the input to the output. Yet, given the choice of employing a ones-vector as input, the weights extracted from the first layer are obtained by selecting the strongest weight-magnitude, as done in hard-pruning (Kim et al. 2015; Han et al. 2015b; Han et al. 2015a). Our approach - similarly to hard-pruning techniques - implicitly adopts weight magnitude as predictor of weight to the output. Previous works (Olden et al. 2002; Olden et al. 2004) demonstrate that the use of weight values for assessing variable contributions in neural networks reach high performance with low computational cost.

Graph-statistics. The adopted method remains general and can be adapted to take

into account different choices of graph-statistics. As a first preliminary exploration, we consider a single graph-statistics given by the combination of in- and out-degree. Yet, many other nodal statistics exist and need to be considered in future works. Specifically, the choice of the statistics \mathcal{S} is based on the concept of network flow, indeed intuitively it captures the amount of information which is processed by each unit in the network.

In the next chapter we applied our proposed graph modeling to the ANNs for the exploration of different continual learning strategies on two classification tasks.

— Chapter 6 —

Continual Learning Strategy Connectivity Characterization

Based on a paper written in collaboration with Dwight Nwagwe, Marion Mainsant, Raphael Bayle, Marina Reyboz, Martial Mermillod, Michel Dojat, and Sophie Achard. In preparation for submission Carboni et al. 2023d

In this chapter we propose as case of study the characterization of various learning strategies on continual learning scenario. Particularly, we consider two different classification task and related architectures. We perform three experiments to thoroughly assess the graph-based framework's ability to recognize different deleterious behaviors arising from various learning strategies in continual learning scenarios. Moreover, we introduce the notion of plasticity/stability-critical unit set and design the pruning experiment to obtain a fine-grained analysis at the unit level of the considered learning strategies. Our graph modeling manage to identify units with slight norm changes in weights across learning sessions, without requiring input data processing.

Contents

6.1	Introduction	139
6.2	Material	139
6.2.1	Handwritten Digit Recognition	139
	Database	139
	Architecture and Learning Strategies	140
6.2.2	Face Emotion Recognition	141
	Database	141
	Architecture and Learning Strategies	142
6.3	Experiments	144
6.3.1	Experiment 1: Characterization	145
	Handwritten Digit Recognition task	145
	Face Emotion Recognition task	147
6.3.2	Experiment 2: Network Surgery	148
	Handwritten Digit Recognition	151
	Face Emotion Recognition task	152
6.3.3	Experiment 3: Weights-injection	152
	Handwritten Digit Recognition task	154
	Face Emotion Recognition task	154
6.4	Summary and conclusion	157
6.4.1	Limitations	158
6.4.2	Future work	158

6.1 Introduction

In this chapter, we explore the previously proposed framework for the characterization of sequential learning process of ANNs. This process mimics how the brain continuously learns and adapts to new tasks (Milgram et al. 1987; Pascual-Leone et al. 2005). As neuroscientific literature reports, brain connectivity changes are associated with new learning tasks (Casimo 2018; Vico Fallani et al. 2010; Zouridakis et al. 2007), we explore, using graph modeling, the corresponding ANN connectivity changes. In particular, we investigate how the specific graph statistics is modified during a continual learning framework and the conditions that lead to catastrophic forgetting.

We consider two simple ANNs architectures (an input layer, a hidden layer, and an output layer) sequentially trained in a simple *handwritten digit recognition task* and in a more complex *face emotion recognition task* using respectively the MNIST database (Deng 2012; LeCun et al. 1989) and the FER+ database (Goodfellow et al. 2013a; Barsoum et al. 2016).

In both cases, we train the ANN architectures using different learning strategies (Lomonaco et al. 2021) in various learning session orders. In particular, we compare brain-inspired learning strategies specifically developed to reduce catastrophic forgetting occurrence (McClelland et al. 1995; French 1999). To ensure comparability of architecture, we focus on the analysis of changes occurring over a fixed graph structure, thus we do not consider neurogenesis strategies which require a change in the architecture structure as new tasks are learned.

To achieve a baseline comparison, we will also consider **Finetune** and **Cumulative** strategies. In the former nothing is done to avoid catastrophic forgetting, while in the latter the architecture is subsequently trained using all previously seen training data up to the task of the current session as it happens in Offline training.

6.2 Material

6.2.1 Handwritten Digit Recognition

Database

Despite being a universal task, digit handwriting is influenced by individual uniqueness in the formation and appearance of the digits (Jain et al. 2008). Educated humans gain expertise in recognizing handwritten digits all along their existence, from early school training, when such an ability is acquired, to adult life during which the ability is continuously refined to adapt to recognize distorted samples or more personal style (Legault et al. 1992).

Training an artificial system to be competitive with humans in a handwritten digit recognition task is a fundamental step in human-machine interaction (LeCun et al. 1995; Ciregan et al. 2012; Kumar et al. 2018; Niu et al. 2012; Pashine et al. 2021). In this field, the MNIST database Deng 2012 is widely used for benchmarking various

image recognition algorithms. The database contains a total of 60,000 training and 10,000 test images of handwritten digits, each of which is 28x28 pixels in size, written by more than 500 different writers (LeCun et al. 1995). The digits range from 0 to 9 as grayscale images.

Architecture and Learning Strategies

Different ANN architectures and algorithms of varying complexity have been proposed to tackle the classification of the MNIST dataset (Baldominos et al. 2019; Ciregan et al. 2012; Jarrett et al. 2009). Due to the objective of our case of study, we used the MNIST database in a simple feedforward architecture defined as follows: an input layer of 784 units, a hidden layer of 512 units, and the output layer with 10 output neurons. Thus, the images were flattened before being fed into the ANNs. Comparable architectures are used in the literature for learning strategies evaluation (Goodfellow et al. 2013b; Kirkpatrick et al. 2017; Zenke et al. 2017; Lomonaco et al. 2021).

We trained this architecture using different learning strategies (hyperparameters details are reported in Table 6.1):

- Sample Replay (Lomonaco et al. 2021)
- GDumb (Prabhu et al. 2020)
- SI (Zenke et al. 2017)
- EWC (Kirkpatrick et al. 2017)
- LwF (Li et al. 2017b)
- Cumulative
- Finetune

Table 6.1: Hyperparameters of the considered learning strategies in MNIST settings.

Hyperparameters			
Units in [input, hidden, output] [784, 512, 10]	Activation functions [relu, softmax]	Optimizer Adam	Loss function Cross-entropy

For each architecture and learning strategy, we tested different randomly orders of learning sessions. In Table 6.2, we report for our simulations, the achieved results in *stability,plasticity*, last global accuracy Ω_{all} ($\Omega_{\text{all}} = \frac{1}{T-1} \sum_{i=2}^T \frac{\alpha_{\text{all},i}}{\alpha_{\text{ideal}}}$) as defined in Kemker et al. 2018.

Table 6.2: Complete list of the trained configuration results on the handwritten digit recognition database MNIST. Acc: global accuracy after training on all sessions.

Order	Strategy	Acc.	stab.	plas.	Order	Strategy	Acc.	stab.	plas.
(A) 1036457928	Sample Replay	0.80	0.77	0.90	⋮	⋮	⋮	⋮	⋮
	GDumb	0.23	0.00	0.32	(F) 941752386	Sample Replay	0.73	1.00	0.80
	EWC	0.12	0.14	1.00		GDumb	0.40	0.84	0.42
	LwF	0.15	0.21	0.99		EWC	0.20	0.43	0.99
	SI	0.13	0.19	1.00		LwF	0.26	0.68	0.98
	Finetune	0.11	0.08	1.00		SI	0.24	0.53	0.99
	Cumulative	0.72	1.00	0.59		Finetune	0.12	0.15	1.00
(B) 2813047569	Sample Replay	0.60	0.94	0.55		Cumulative	0.71	1.00	0.45
	GDumb	0.54	0.42	0.63	(G) 9480712653	Sample Replay	0.70	0.89	0.74
	EWC	0.31	0.03	0.99		GDumb	0.44	0.00	0.55
	LwF	0.46	0.11	0.97		EWC	0.24	0.00	0.99
	SI	0.45	0.12	0.98		LwF	0.31	0.00	0.91
	Finetune	0.14	0.00	1.00		SI	0.29	0.00	0.96
	Cumulative	0.68	1.00	0.59		Finetune	0.13	0.00	1.00
(C) 1204673985	Sample Replay	0.82	0.56	0.88		Cumulative	0.65	1.01	0.55
	GDumb	0.41	0.00	0.53	(H) 6180534972	Sample Replay	0.76	0.97	0.75
	EWC	0.15	0.00	1.00		GDumb	0.42	0.89	0.44
	LwF	0.23	0.00	0.98		EWC	0.18	0.46	0.99
	SI	0.22	0.00	0.99		LwF	0.30	0.78	0.97
	Finetune	0.11	0.00	1.00		SI	0.25	0.72	0.99
	Cumulative	0.80	1.00	0.65		Finetune	0.13	0.14	1.00
(D) 4816203957	S. Sample Replay	0.68	0.91	0.72		Cumulative	0.70	1.00	0.60
	GDumb	0.48	0.24	0.57	(I) 2784903165	Sample Replay	0.64	0.97	0.70
	EWC	0.25	0.00	0.99		GDumb	0.47	0.62	0.56
	LwF	0.37	0.07	0.98		EWC	0.23	0.38	0.99
	SI	0.32	0.02	0.99		LwF	0.37	0.52	0.91
	Finetune	0.17	0.00	0.99		SI	0.29	0.52	0.96
	Cumulative	0.73	1.00	0.62		Finetune	0.12	0.10	1.00
(E) 6031925487	Sample Replay	0.78	1.01	0.80		Cumulative	0.76	0.98	0.70
	GDumb	0.46	0.64	0.53	(J) 2536790418	Sample Replay	0.69	0.98	0.73
	EWC	0.26	0.18	0.99		GDumb	0.47	0.69	0.51
	LwF	0.31	0.36	0.98		EWC	0.21	0.17	0.98
	SI	0.29	0.31	0.99		LwF	0.33	0.49	0.95
	Finetune	0.17	0.00	1.00		SI	0.24	0.28	0.98
	Cumulative	0.70	1.02	0.52		Finetune	0.14	0.00	1.00
⋮	⋮	⋮	⋮	⋮		Cumulative	0.83	0.99	0.68

6.2.2 Face Emotion Recognition

Database

In human social interaction, the ability to identify other beings feeling and emotions is crucial, particularly to adapt the individual’s behavior. Emotion recognition is mainly achieved, but not exclusively, by decoding non-verbal information and in particular facial expressions. Many social cognition studies focus on understanding emotion recognition in humans, for instance detecting specific region contributions (in particular in the amygdala Adolphs et al. 1994; Adolphs et al. 1995; Calder 1996; Breiter et al. 1996; Morris et al. 1996; Davis et al. 2001; Pessoa et al. 2010; Anderson et al. 2000) or defining functional activation at different emotional faces processing (Fusar–Poli et al. 2009; Gómez et al. 2020; Liao et al. 2021; Underwood et al. 2021).

Automatic systems and ANNs for facial expression recognition have also been introduced (Kumari et al. 2015; Mehta et al. 2018; Li et al. 2020). Available databases mainly cover the six basic emotions: Anger, Fear, Sadness, Disgust, Surprise and Happiness (Ekman 1992a; Ekman 1992b; Ekman et al. 1978) and a Neutral emotional state. Here, we considered the Fer+ databases (Goodfellow et al. 2013a; Barsoum et al. 2016). This database contains 35,685 grayscale 48x48 pixels images with all the basic emotions and covering all ages, gender, and ethnicity, the labels are provided by 10 crowd taggers.

Architecture and Learning Strategies

We follow the work of Mainsant et al. 2021 which introduces an ANN allowing continual learning without the requirement of neurogenesis, nor the necessity of an oracle about learned or data to be learned, as well as data privacy issues for facial emotion recognition.

In a preliminary step, we performed a feature extraction employing a pre-trained ResNet50 model provided by Wang et al. 2018a. A visualization of ResNet50 architecture can be found in Fig. 6.1. The feature vectors used as input of our NN architectures were obtained after the flatten operation of ResNet (indicated by a red rectangle in Fig. 6.1).

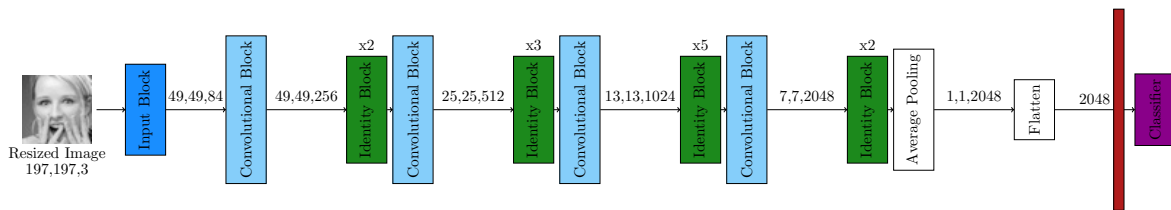


Figure 6.1 – ResNet50 model architecture. The features vector taken as the input of the last fully connected layers is extracted in correspondence of the red box. The Convolutional Blocks extract features changing the input dimensions. The Identity Blocks extract features without changing the input dimensions.

The 2048-feature vectors are then fed into the following fixed architecture:

- input layer of 2048 units corresponding to the size of features extracted from images;
- hidden layer with 1000 neurons;
- output layer.

We evaluated Dream Net together with the other learning strategies (as listed in Table 6.3). Note that the methods do not entirely share the same architecture: indeed Dream Net requires an output layer composed of several neurons corresponding to the input (Auto-associative or Auto-encoder part) and several neurons corresponding to the number of classes (Hetero-associative or part). While this part is used both for training and evaluation of the model, the Dream Net activation network at rest is obtained by discarding all the units coming from the auto-associative part and their

related edges, so that only the part of the ANN directly involved in the classification task is retained. This guarantees a fair comparison with the other learning strategies and architectures.

Table 6.3: Hyperparameters of the considered learning strategies on Fer+ settings.

	Strategy	Hyperparameters		
	Units in [input, hidden, output]	Activation functions	Optimizer	Loss function
Cumulative*	[2048, 1000, 2055]	[relu, sigmoid]	Adam*	Binary Cross-entropy
Finetune*	[2048, 1000, 2055]	[relu, sigmoid]	Adam	Binary Cross-entropy
Dream Net	[2048, 1000, 2055]	[relu, sigmoid]	Adam	Binary Cross-entropy
Cumulative	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy
GDumb	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy
Sample Replay	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy
EWC	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy
LwF	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy
SI	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy
Finetune	[2048, 1000, 7]	[relu, sigmoid]	Adam	Binary Cross-entropy

Dream Net architecture was sequentially trained with different learning strategies and in different learning sessions. We tested 7 choices of emotion orders so that each class is learned in the first position, followed by random order choice.

The different learning configurations are listed in Tab. 6.4, with the learning orders notated by A: angry, D: disgust, F: fear, H: happy, S: sad, Su: surprise, and N: neutral. The model setting performance evaluation is based on the last global accuracy and on the metrics proposed in [Kemker et al. 2018](#).

Table 6.4: List of the trained configuration results on the face emotion recognition database Fer+. Acc: global accuracy after training on all sessions.

Order	Strategy	Acc.	stab.	plas.	Order	Strategy	Acc.	stab.	plas.
(A) ADFHSSuN	Cumulative	0.99	1.01	0.67	(E) SNDFSuAH	Cumulative	0.77	0.52	0.50
	Cumulative*	0.76	1.03	0.81		Cumulative*	0.77	1.00	0.85
	Dream Net	0.71	0.81	0.84		Dream Net	0.71	0.99	0.90
	Sample Replay	0.57	0.52	0.95		Sample Replay	0.37	0.60	0.33
	GDumb	0.61	0.57	0.79		GDumb	0.37	0.80	0.25
	EWC	0.10	1.15	0.00		EWC	0.099	0.00	0.17
	LwF	0.34	0.00	0.83		LwF	0.34	0.23	0.17
	SI	0.34	0.00	1.00		SI	0.34	0.23	0.00
	Finetune*	0.02	0.00	1.00		Finetune	0.26	0.00	1.00
	Finetune	0.032	0.19	0.17		Finetune*	0.36	0.00	1.00
(B) DAHSNSuF	Cumulative	0.76	1.00	0.40	(F) SuDFNHAS	Cumulative	0.76	0.33	0.57
	Cumulative*	0.76	1.03	0.84		Cumulative*	0.76	1.00	0.86
	Dream Net	0.77	1.00	0.82		Dream Net	0.71	0.93	0.92
	Sample Replay	0.54	0.54	0.26		Sample Replay	0.41	0.68	0.30
	GDumb	0.52	0.53	0.25		GDumb	0.34	0.66	0.41
	EWC	0.099	0.00	0.17		EWC	0.099	0.00	0.17
	LwF	0.34	0.19	0.17		LwF	0.34	0.20	0.17
	SI	0.34	0.19	0.17		SI	0.34	0.20	0.00
	Finetune	0.032	0.00	1.00		Finetune	0.13	0.00	1.00
	Finetune*	0.34	0.00	1.00		Finetune*	0.13	0.00	1.00
(C) FHSuDNSA	Cumulative	0.78	0.86	0.39	(G) NDSuFHSA	Cumulative	0.76	0.77	0.68
	Cumulative*	0.76	1.56	0.81		Cumulative*	0.76	1.00	0.83
	Dream Net	0.51	1.31	0.32		Dream Net	0.56	0.79	0.58
	Sample Replay	0.57	0.68	0.14		Sample Replay	0.34	0.84	0.20
	GDumb	0.62	0.81	0.27		GDumb	0.34	0.77	0.43
	EWC	0.099	0.00	0.17		EWC	0.099	0.51	0.17
	LwF	0.34	0.22	0.17		LwF	0.34	0.61	0.00
	SI	0.34	0.22	0.00		SI	0.34	0.20	0.17
	Finetune	0.099	0.00	1.00		Finetune	0.098	0.00	1.00
	Finetune*	0.099	0.00	1.00		Finetune*	0.098	0.00	1.00
(D) HADNSFSu	Cumulative	0.76	1.01	0.67					
	Cumulative*	0.76	0.99	0.81					
	Dream Net	0.71	1.00	0.86					
	Sample Replay	0.56	0.54	0.35					
	GDumb	0.43	1.00	0.40					
	EWC	0.099	0.00	0.17					
	LwF	0.34	0.24	0.33					
	SI	0.34	0.24	0.17					
	Finetune	0.12	0.00	1.00					
	Finetune*	0.34	0.00	1.00					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

6.3 Experiments

Following the standard human brain functional connectivity analysis framework, we performed two types of experiments on the induced graph $\mathcal{G}(NN)$ of various NNs: a global setting recognition and a nodal role identification. For the former, we assumed that the induced graphs of a NN contained enough information to identify its learning settings. For the latter, we associated nodal graph statistics with a functional property of the artificial system. In particular, we performed a network surgery, by removing hidden units according to their statistics value. Thus, we evaluated the *stability* and the

plasticity performance: if removing the units had a strong negative effect, thus the units are considered as critical. In particular, we defined a set of units as stability-critical (resp. plasticity) if by pruning all its units, we observed negative changes in terms of stability (resp. plasticity) performance. If the evaluation performance in the pruned version is increased, then we defined the unit set as stability/plasticity-inhibitorial.

6.3.1 Experiment 1: Characterization

At a global level, we show how this brain-inspired framework can be used to extract interpretable statistics which allow the detection of configurations affected by catastrophic forgetting. Moreover, the same statistics can be used to group together configurations that apply the same learning strategies. Similarly to human brain studies where functional connectivity allows to discriminate across brain states, the first objective of our analysis was to demonstrate that a graph-based connectivity analysis of ANNs could discriminate architectures according to different learning strategies. Particularly, we explored the possibility of correctly identifying models whose learning was affected by catastrophic forgetting. To this extent, we considered the induced graphs of each configuration in different learning settings. We determined the threshold to guarantee in and out degrees statistics followed a similar power-law distribution (an example is shown in Fig. 6.2).

In this line, we first propose a general learning strategy characterization **Experiment 1 (Characterization)**. For each configuration, we determine the number of hidden units in the tail of \mathcal{S}_t together with the Maximum Unit Change.

We characterize the different learning strategies by comparing these graph-based features and by using them for their identification in a reduced space.

Moreover, we validate the use of the degree statistics by investigating the relationship among units in the tail and the norm changes of their synaptic weights across the different learning sessions.

Given the $T + 1$ features per configuration, we determined the principal components and perform a K-Means Clustering algorithm in the reduced space. Thus, we evaluated the consensus score by computing the adjusted rand index where the true labels are given by the learning strategies or by two classes *affected by catastrophic forgetting*, *not affected* which corresponds to the configurations where $stability < 0.5$, $stability \geq 0.5$.

Handwritten Digit Recognition task

We observe similar distribution across the different learning strategies, except for the Finetune model. The Maximum Unit Change value is under 100 in continual learning strategies and Cumulative, but equals to 179.7 ± 34.0 for the Finetune model (Fig. 6.3 Top Left). The tail persistence units shows the peculiar behavior of Cumulative and Finetune models for which the information is subsequently overwritten: their consecutive persistence is severely lower compared to other strategies and only less than 40%

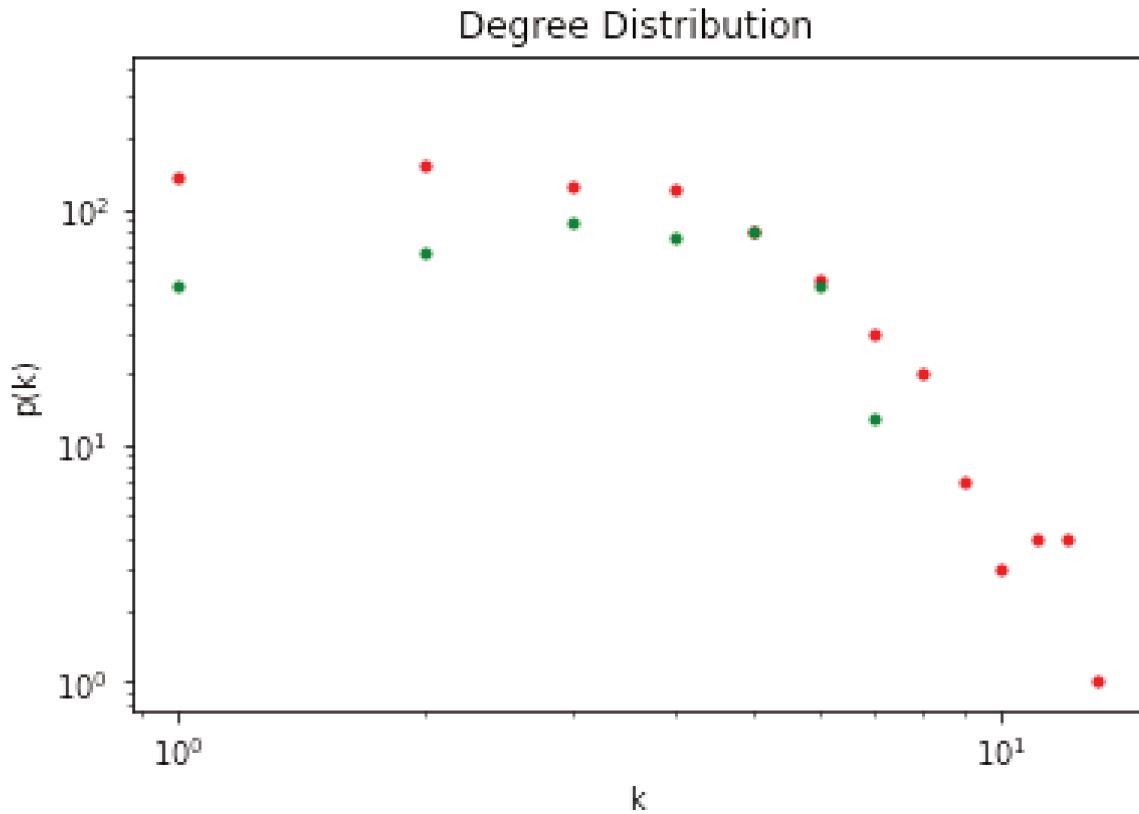


Figure 6.2 – Example of in (green) and out (red) degree distributions in the induced graph of an activation network at rest.

of the tail units persists from the learning of the basal task. For all other strategies, the percentage reaches 80% with the exceptional case of the GDumb an average of 98%. Visualization in a reduced space of the graph-features vector reveals how these graph-based extracted features are able to identify how Finetune and other strategies are affected by catastrophic forgetting (Fig. 6.3 Bottom Left). A simple clustering algorithm in the reduced space groups together all regularization learning strategies, replays and Cumulative reaching an overall consensus score of 0.71 (Fig. 6.4).

In Fig. 6.5 we explore the relationship between the nodal statistics and the consecutive norm change of weights in the trained ANN. It is interesting to observe a slight change in norm across consecutive steps even in non-regularization methods. Additionally, we should notice that in the MNIST dataset, there is a large number of units that receive as input zeros from the border of the flattened images, these units do not update their synaptic weights across the learning process. The statistics \mathcal{S} is able to retrieve the majority of units whose synaptic weights norm does not change across consecutive steps. This is highly valuable for a resting-state analysis which defines induced graphs without any input data.

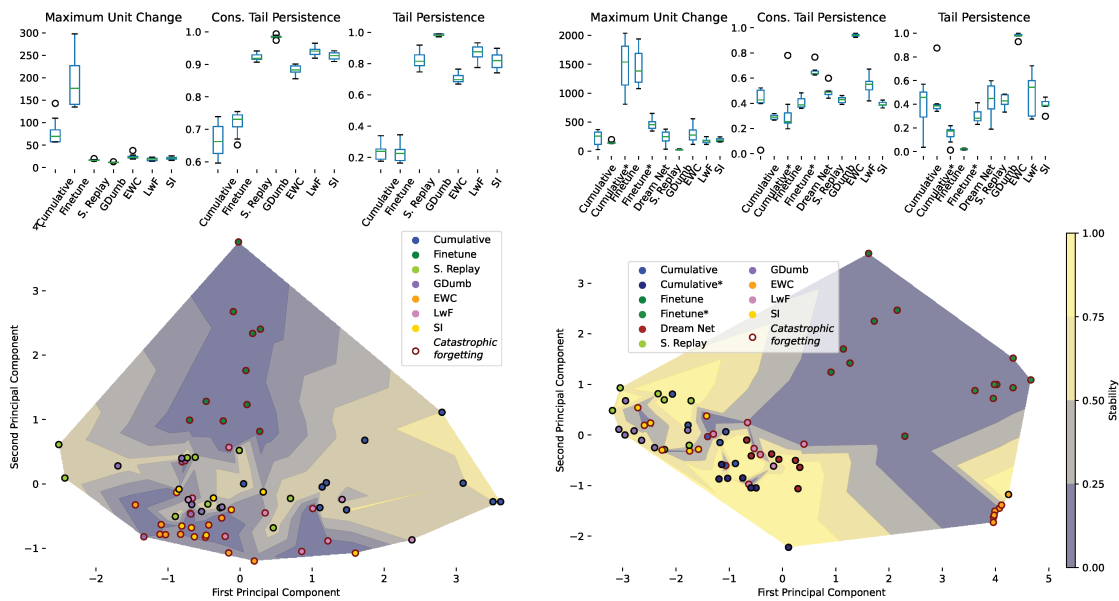


Figure 6.3 – Results of Experiment 1 (Characterization). **Left:** Handwritten Digit Recognition task **Right:** Face Emotion Recognition task. **Top:** Distribution of the graph-based features by strategy. Cons. Tail Persistence: Consecutive Tail Persistence percentage of units in the tail at consecutive learning sessions, Tail Persistence: Percentage of units being in the tail at the first and last learning session. **Bottom:** Visualization of the reduced space of the graph-based features extracted by each sequence of ANNs trained with different strategies. A point corresponds to a unique configuration given by the ANN architecture, the learning strategy, and the fixed-order learning. Points are colored by their stability performances. Each red circle indicates catastrophic forgetting. Similarly, the plasticity performances for all strategies are reported in the appendix. S.Replay: Sample Replay, EWC: Elastic-Weight-Consolidation, SI: Synaptic Intelligence, LwF: Learning without Forgetting.

Face Emotion Recognition task

The same experiment was performed in the Face Emotion Recognition task. In this more difficult task, we considered two different architectures. Since Dream Net requires an auto-hetero associative architecture, we introduced two different versions for Cumulative and Finetune with or without an auto-associative part in the output of the NN. In general, the introduction of auto-associative neurons gives better results already in the standard Cumulative configuration.

As in the handwritten digit recognition task, we report a strong Maximum Unit Change for the Finetune strategy (Top Right Fig. 6.3). The tail persistence results show a consecutive tail persistence in Dream Net strategy of about 0.7, with lower values for Cumulative, Finetune, Sample Replay and GDumb. EWC shows the highest consecutive tail persistence, but the other regularization strategies have in average less persistence with respect to Dream Net. Not surprisingly, the percentages in tail persistence have the lowest average for the Finetune configurations and the highest for EWC. Finally, the visualization in a reduced space of the graph features reveals the possibility of identifying the robust learning strategies (Bottom Right Fig. 6.3), a

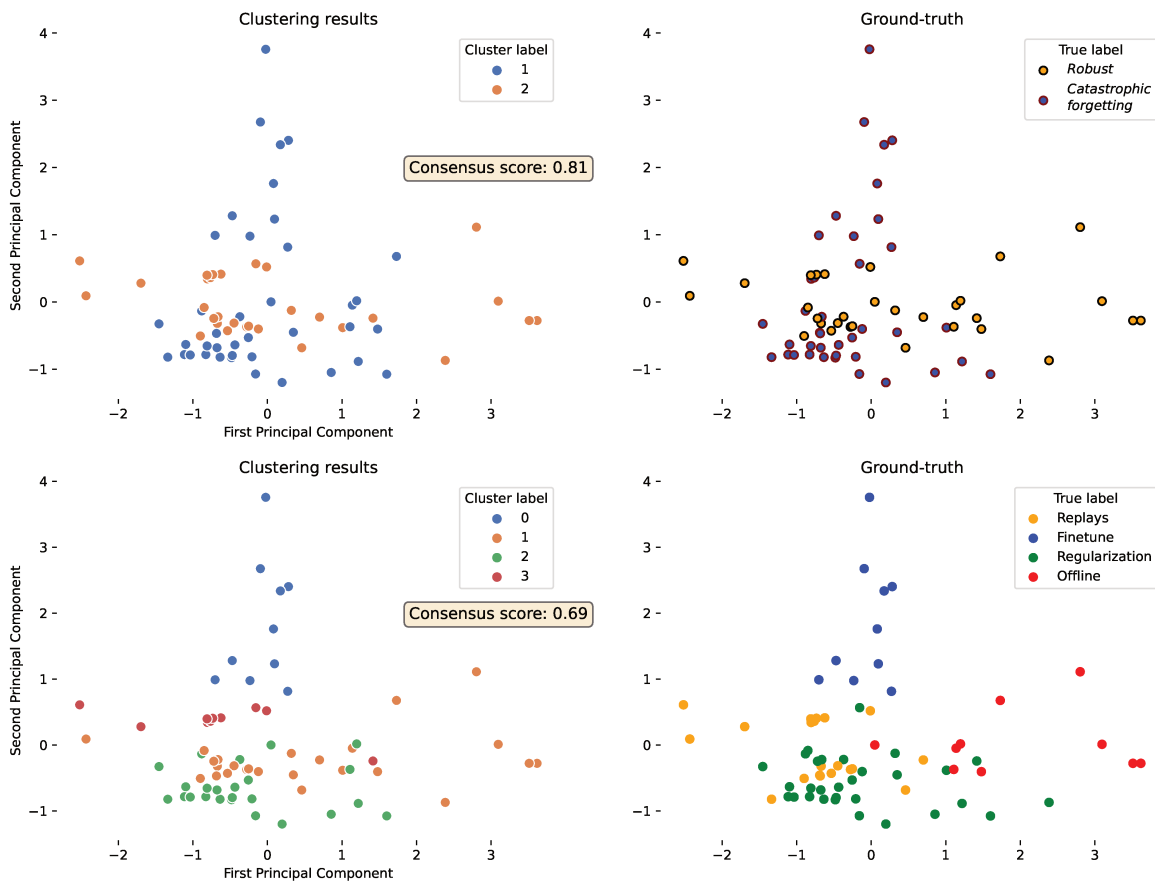


Figure 6.4 – Visualization of the cluster labels (Left) obtained by applying KMeans (Top: $K=2$, Bottom: $K=4$) in the reduced space of the Handwritten Recognition task and their ground truth (Right).

simple K-Means algorithm reaches a consensus score of 0.77 (Fig. 6.6). Interestingly, the Finetune configuration, which is dramatically affected by catastrophic forgetting, appears to be moved away with respect to the other clusters.

6.3.2 Experiment 2: Network Surgery

Similar to what happens in the human brain, where each node can be associated with a different role given a graph nodal statistics (Carboni et al. 2023b), we assume that each unit behaves differently at fostering or inhibiting the plasticity and stability performance of the ANN. We hypothesize that such properties are unevenly distributed across the units of an ANN and we aim at identifying units that contribute differently to the stability and plasticity performances of the ANN. Specifically, we distinguish two subsets of \mathcal{S}_i : units that belong to the middle quartiles and units which belong to the tail. As it happens in natural intelligence, where high volume spines persistence is associated with the memory of the corresponding task (Yang et al. 2009), we hypothesize that units that persist in having extreme \mathcal{S} values can be associated with high stability. In contrast, nodes whose \mathcal{S} values strongly change can be associated with the plasticity property.

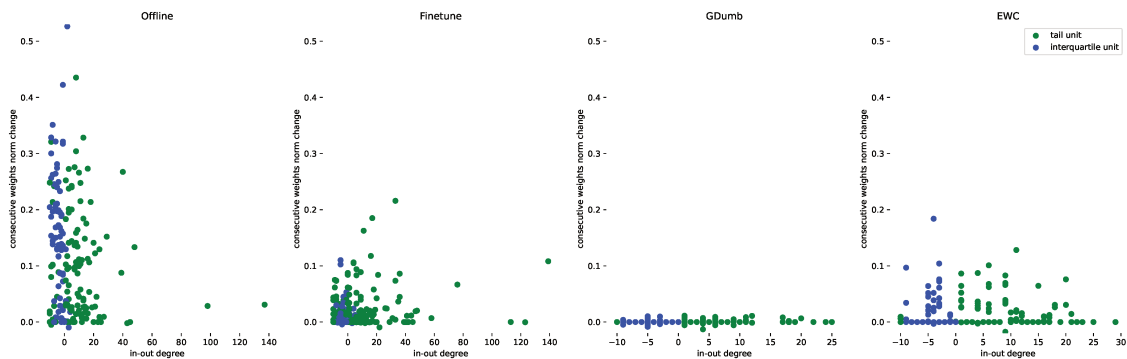


Figure 6.5 – Visualization of the relation between \mathcal{S} and the norm changes of the weights

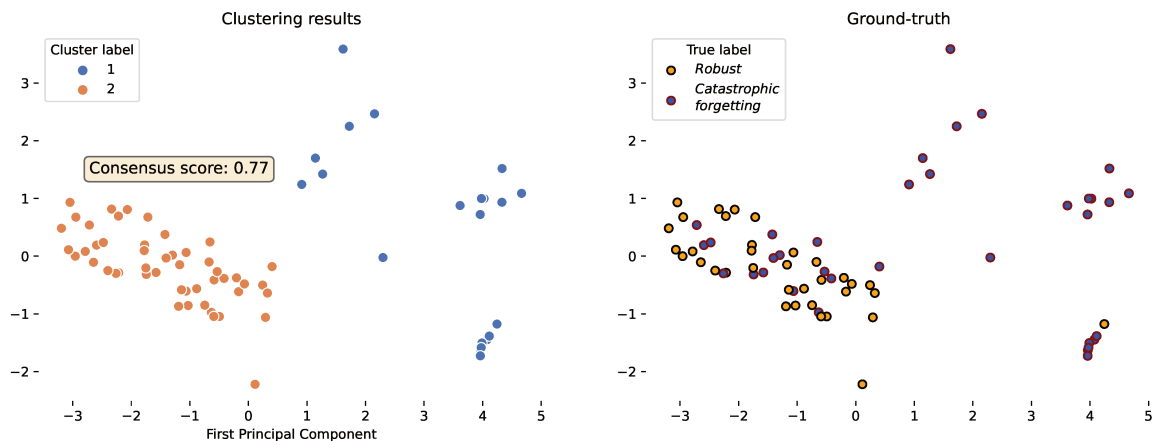


Figure 6.6 – Visualization of the cluster labels (Left) obtained by applying K-Means with $K=2$ in the reduced space of the Face Emotion Recognition task and their ground truth (Right).

In the **Experiment 2 (Network Surgery)**, we simulate the removal of the increased-volume spines (Cichon et al. 2015; Hayashi-Takagi et al. 2015) by pruning (i.e. turning off) the units in ANN. The objective is to associate the interquartile units and the tail units with the stability or plasticity properties of the ANN configuration by evaluating the change in the performance with respect to the standard model (when no unit is turned off). In artificial systems, it is indeed possible to perform such a *network surgery* in order to associate a functional role to a unit following a graph-statistics-based nodal role.

Inspired by the proposal in Zhang et al. 2022a, we assume that if a set of units is pruned and the performance changes significantly in terms of *stability*, then the pruned units are critical for the *stability* of the ANN model. The same applies to the *plasticity*. Given the number of units in an ANN, testing the pruning of each unit becomes rapidly unfeasible, thus we consider the set of units in the interquartile and in the tail distribution of \mathcal{S} . Hence we propose two pruned ANN versions, one which only preserves the

synaptic weights of units in the tail of the distribution of \mathcal{S} , and another version that nullifies the weights of the tail units. Since the distribution of \mathcal{S}_t changes across the learning session, we define a copy of each trained ANN in the order sequence, with each pruned copy at the t -th learning session obtained by looking at the \mathcal{S}_t distribution.

These pruned ANNs ordered by learning sessions are respectively denoted NN^τ and $\text{NN}^{\bar{\tau}}$, while the standard sequence (i.e. where no pruning is applied) is notated NN .

We evaluate the difference between the *stability/plasticity* of the standard model and its corresponding pruned version: a strong difference corresponds to a critical stability/plasticity unit (i.e. when the units are pruned the performance decreases dramatically), while a low positive difference corresponds to a robust unit (i.e. the pruned units do not affect the performance). Finally, a negative difference identifies plasticity/stability inhibitory units, i.e. units whose synaptic weights have a negative effect on its plasticity/stability.

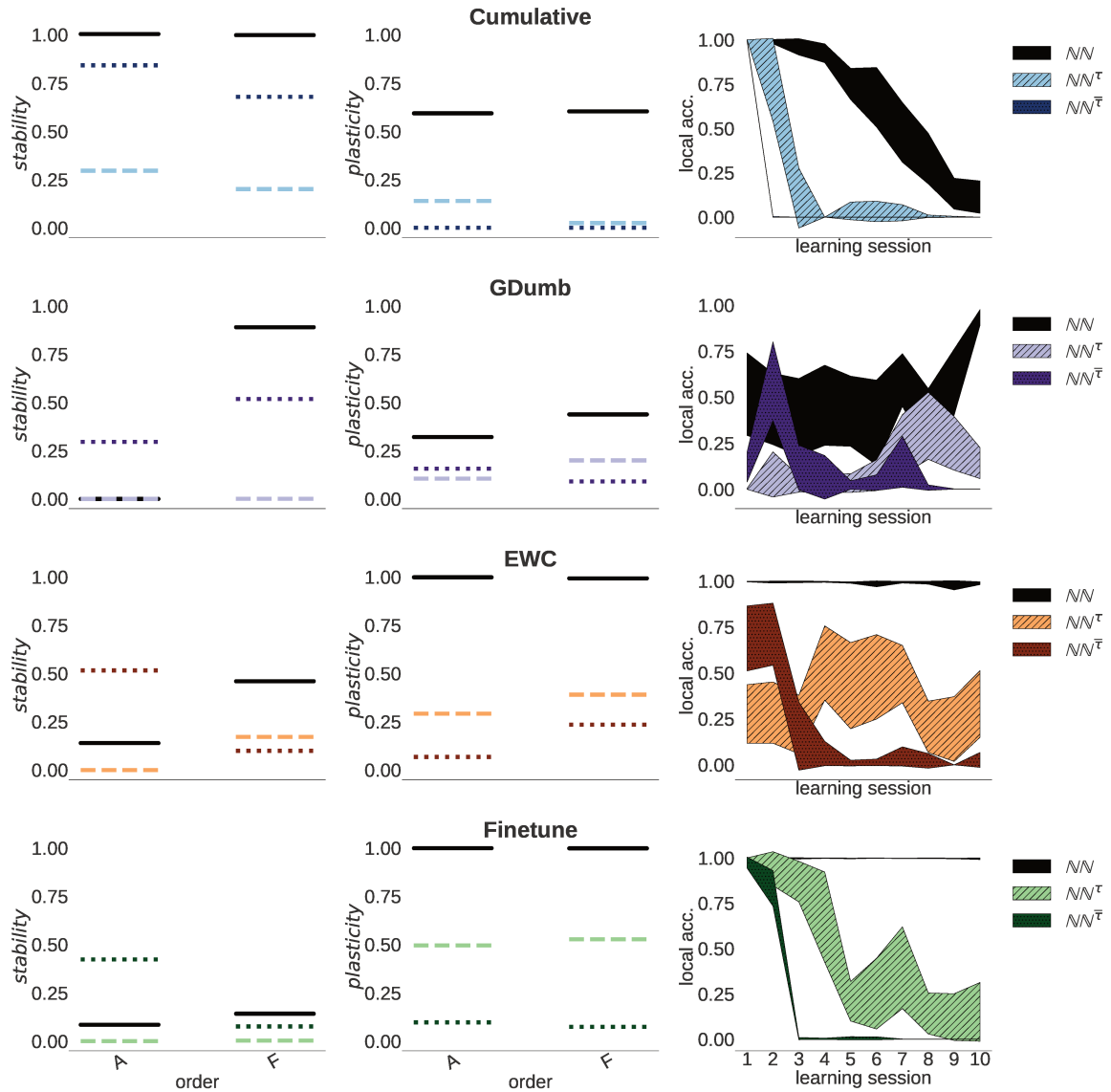


Figure 6.7 – Results of Experiment 2 (Network Surgery) for two orders of Handwritten Digit Recognition. *Stability* (Left) and *Plasticity* (Center) performances for all configurations on handwritten digit recognition task for the standard model NN (plain black line) and their pruned versions $\text{NN}^\tau, \text{NN}^{\bar{\tau}}$ (colored dashed lines). The performance is reported for two different orders (A, F , see Tab. 6.2) of learning sessions. **Right**: average local accuracy performances for the last learned class across learning sessions.

Handwritten Digit Recognition

We observe different behaviors depending on the learning strategy (see Fig. 6.7, Fig. A.7). The Cumulative learning strategy reveals how in offline learning the stability and plasticity properties are shared in the hidden units: pruning of both sets equally affects the performance. On the contrary, the Finetune model, whose stability is extremely

low in the standard version, shows strongly improved performance for the pruned sets. Indeed, we can identify the tail units as stability-inhibitory: pruning them has a beneficial effect in recalling the basal task, reaching a maximum *stability* performance of almost 0.46 for the first training session on the class corresponding to the digit 4 (Fig. A.7). The GDumb approach behaves similarly to the Finetune or the Cumulative models depending on the standard performance: it exhibits a beneficial effect of pruning the tail when the standard performance is poor and critical behavior in general pruning when the standard performance is higher. A stronger effect is observed when pruning the interquartile range units.

Concerning plasticity, a common general behavior across all learning strategies is a stronger plasticity-critical behavior of the tail units: when their weights are set to zero the plasticity performance is dramatically reduced. The pruning of the interquartile range also affects the plasticity, but with less effect, especially for the Finetune model. We report in details the local accuracy at each learning session (Right Fig. 6.7). We can notice how in the Finetune model, the NN^T has positive local accuracy and for a few learning sessions has comparable results with the standard version. In the regularization methods, a transition between the best local accuracy pruned performance exists, with the interquartile having better results at the beginning of the learning sessions and the tail in the following phases.

Face Emotion Recognition task

The Cumulative model distributes the stability properties across all units independently by the distribution of \mathcal{S} . This leads to a dramatic decrease of the stability performance for the pruning versions. Concerning the plasticity property, we detect a higher criticality for the units in the interquartile range. For Dream Net, all pruned settings have very poor results for new tasks, with the extreme case of the model with tailing removing NN^T that has a zero Ω_{new} . Different results are observed for the Finetune model, where the pruning versions, keeping only the weights in the tail, reach the same Ω_{new} that the standard one.

On the contrary of Cumulative and Dream Net, where none pruned copy reaches similar performances, for Finetune, the NN^T is a good pruned copy of the complete standard model.

6.3.3 Experiment 3: Weights-injection

The weights injection experiment aims at validating the identified stability/plasticity relation at the hidden unit level and assessing the biological assumptions.

In particular, we assume that injecting the weight of a set that is plasticity-critical for the learning of a task into its consecutive trained architecture can ensure the ability to perform such a task after the learning of new tasks. More generally, we aim at verifying if the injection of stability-critical units in a model with low stability can

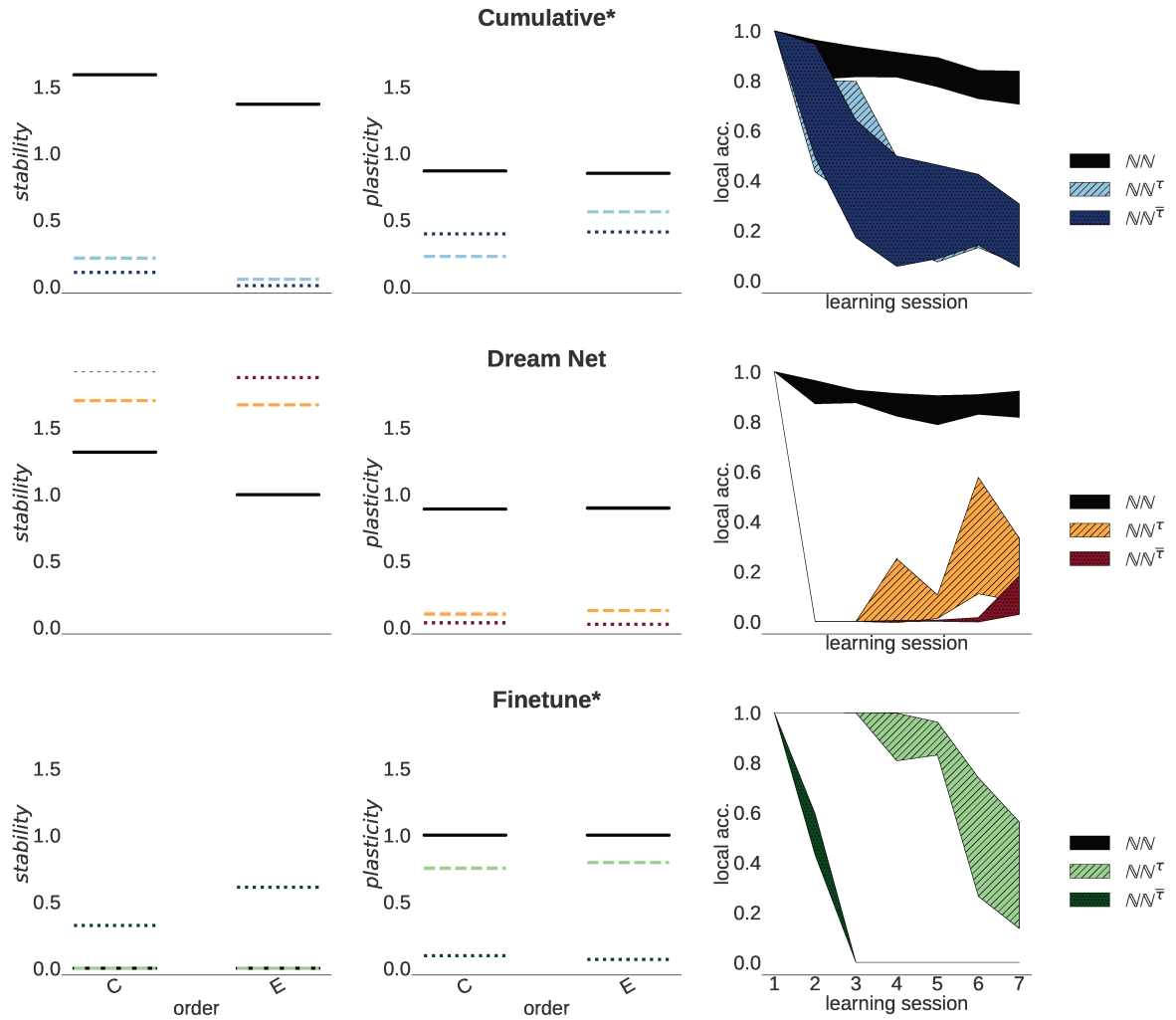


Figure 6.8 – Results of Experiment 2 (Network Surgery) for two orders of Face Emotion Recognition Task. *Stability* (**Left**) and *Plasticity* (**Center**) performances for all strategies on face emotion recognition task for the standard model \mathcal{NN} (plain black line) and their pruned versions $\mathcal{NN}^\tau, \mathcal{NN}^{\bar{\tau}}$ (colored dashed lines). The performance is reported for two different orders (C, E , see Tab. 6.4) of learning sessions. **Right**: average local accuracy performances for the last learned class across learning sessions.

have a positive effect and similarly for plasticity.

This weights injection experiments can be seen as a new learning strategy for continual learning which employ graph-information. Note that the first two procedures FWI and BWI do not require any training and are simply obtained by considering the model snapshots of the previous task and the current one. While FWI only injects the weights from the previous one into the current model by looking at which hidden units were in the tail in the previous step, the BWI procedure keeps all the weights from the previous task, except those corresponding to units that are currently in the tail, for which the weights from the current trained networks are used.

The evaluation of these injections is reported in terms of difference in performance between the receiver model, which goes through the weights injection process, and the standard model that doesn't receive any weight injections. The performance is measured on stability and plasticity metrics. If the performance difference is positive, it indicates that injecting weights has positively impacted the model's performance. Conversely, a negative difference suggests a detrimental effect.

This experimental approach introduces a novel learning strategy for continual learning that leverages graph-information about the neural network's architecture. The weights-injection procedure is related to the concept of transfer incremental learning where transfer learning methods (Weiss et al. 2016; Zhuang et al. 2020; Boukli Hacene et al. 2018) are applied in continual learning scenarios (Masana et al. 2022; Boukli Hacene et al. 2018).

Handwritten Digit Recognition task

A clear increase in the plasticity is observed when fine-tuning is used to retrain the model (NN^{WIF}), even if the number of trainable parameters is reduced with respect to the Finetune learning strategy. An interesting result is observed in terms of stability especially in the Finetune and rarely in some configurations of the EWC, LwF and SI strategies, where a slightly beneficial effect is observed by injecting the weights in the sequential model Fig. 6.9. In the Replay methods and Cumulative, where a fine-tuning procedure that preserves the tail units gives better results in the plasticity.

Face Emotion Recognition task

Concerning the stability, the weights injection procedure without tuning, outperforms the standard for Finetune and Dream Net. Results in plasticity show that only with a fine-tuning procedure, we are able to learn new tasks. Note that the FWI injection procedure in Finetune, significantly affects the plasticity performance, but never reaching the zero (Fig. 6.10).

In the Finetune model, the weight injection procedure without fine-tuning increases the standard performance, indeed when evaluating the global accuracy results in the Finetune model and its weights-injection setting, we observe that when learning a few tasks the weight injection significantly outperforms the standard version (Fig. 6.10), losing however the adaptability to new tasks.

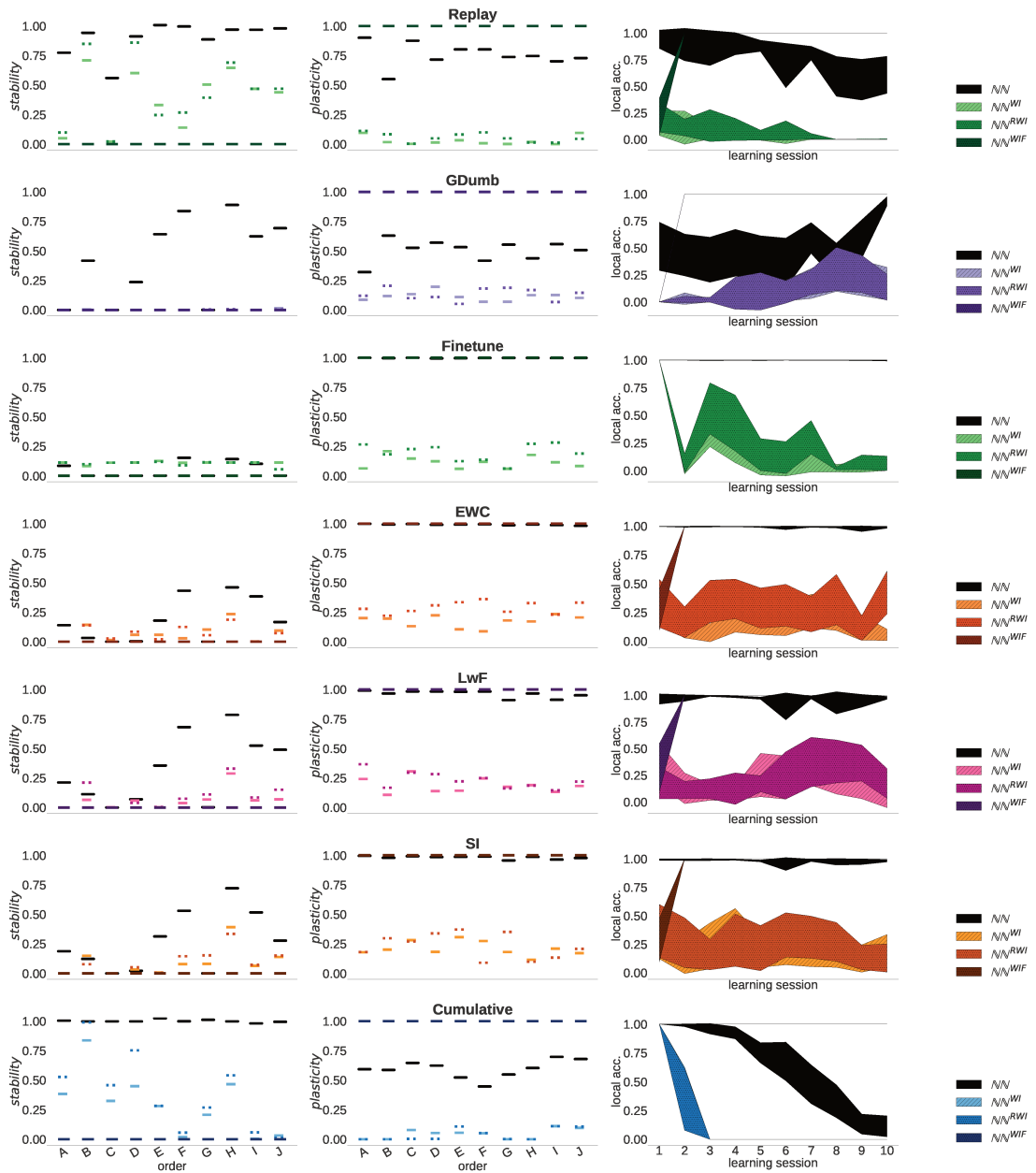


Figure 6.9 – Results of Experiment 3 (Weights-injection) on MNIST configurations.

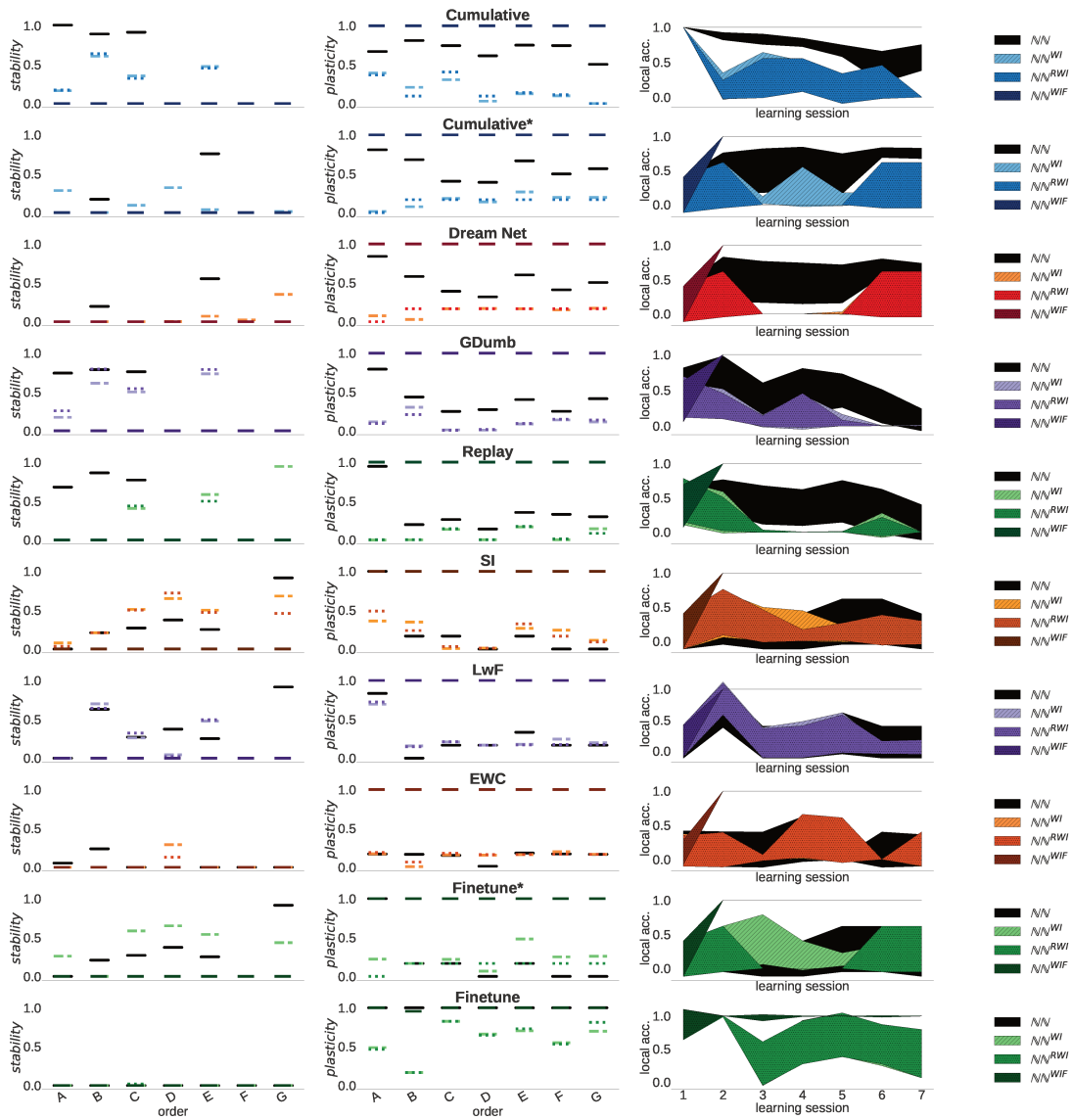


Figure 6.10 – Results of Experiment 3 (Weights-injection) on Fer+ configurations.

6.4 Summary and conclusion

In this study, we leverage a novel research framework in which ANNs are studied starting from their connectivity properties. This framework enables us to integrate the biological inspiration of ANNs into their analyzing tool by proposing a way to fill the gap between brain connectivity studies and the analysis of the information flow in ANN. Utilizing this research framework, we concentrate our analysis on the catastrophic forgetting issue. Our objective was to determine the relationship between existing learning strategies that alleviates the catastrophic forgetting phenomenon and general graph connectivity features. We showed that a simple graph-induced definition and the extraction of interpretable graph features are important indicators of the stability properties of an ANN model being enough to detect the learning strategies applied to the same ANN architecture learning to perform the same task.

This study investigates the utility of implementing memory consolidation or task-specific synaptic consolidation, by distributing unequally the stability and plasticity properties in the ANN units. We report good stability in online learning strategies in the presence of a persistence of tail units with strong synaptic weights across the learning sessions. We showed in the weight-injection experiment how a Finetune model can retrieve previous information when such persistence is brutally forced with the FWI technique. While such persistence can be expected for the regularization methods, it surprisingly appears to be present in replay methods too, suggesting that both methods can induce similar distribution in the strength of the connection in the hidden units. Surprisingly, replay methods achieve the same result of constraining synaptic weight updates across learning sessions, without requiring hyperparameter tuning, unlike regularization strategies. Our graph modeling and the chosen graph statistics are able to detect the units whose weights slightly change in norm across the learning session. This is highly valuable for our graph modeling at rest, which does not estimate the importance of weights given true input data, but whose only relation with data is given by the determined weights after training.

We observed that the low performance of regularization methods is associated with the similarity of behavior between the Finetune strategy.

When coming to more complicated tasks, major differences are revealed between the Cumulative and the Dream Net replay methods. First, Dream Net and Cumulative have opposite behavior in the network surgery experiment: any pruning technique of the Cumulative configuration destroys the stability, while for Dream Net the pruning increases stability at the cost of diminishing plasticity. This suggests that the use of a dropout technique in the training phase of Dream Net, can result in better stability results.

The results in the pruning experiments capture the difference between multiple tasks learning in one session and sequential learning: while offline learning automatically distributes the connections and their strength across the hidden units for the different tasks, sequential learning imposes strong task-specific synaptic weights on a few connections which are continuously overwritten and substituted. Thus, catastrophic

forgetting is alleviated when these connections are not erased but slightly adapted to the subsequent tasks.

In the Finetune model, the proposed graph statistics detect the lottery ticket winner (i.e. a pruned version of the model with the same performance as the entire network and less redundancy (Frankle et al. 2018)). We found that the model which only preserves the weights of the tail units reaches the same performance of the standard one, despite the number of preserved units. This is in line with pruning literature results which determine a very small subnetwork having almost the same performance as the complete model (Wolinski 2020; Tanaka et al. 2020). Remarkably, we prove that the high plasticity of a Finetune model is strongly related by units in the tail of the distribution and that a simple weights-injection across sequential models can enhance its stability.

Yet, none of the weights-injection procedure can be suitable for the design of new continual learning strategy and more works is needed for the design of a graph-informed learning strategy overcoming catastrophic forgetting.

6.4.1 Limitations

Besides laying down the foundation for a graph-statistics-based study of the learning process in ANN, we mainly show empirical findings of post-training ANN model characterization. However, the training process in an ANN is not negligible and by definition highly dependent on the dataset used for training (Ramayachitra et al. 2014; Ali et al. 2019; Djolonga et al. 2021; Song et al. 2022). This is the major weakness preventing results generalization of a post-training analysis: results may change when different tasks or datasets are used, or even by testing a different learning order. We tackle some generalization induced by the order of learning sessions, by randomly testing different orders since testing all possible orders becomes quickly intractable.

In addition, our proposal is only applied to a feed-forward ANN model having a unique hidden layer. In deeper neural networks, we may observe different results depending on the considered hidden layers, as different robustness was observed associated with different layers (Zhang et al. 2022a).

6.4.2 Future work

Many studies have already introduced the need for more complex neural network architectures, giving both brain-inspired motivations and graph-based topological requirements (Sussillo et al. 2009; Mocanu et al. 2016; Mocanu et al. 2018; Hasson et al. 2020; Liu et al. 2021; Kaviani et al. 2021), hence indicating possible future research directions to extend the present work to different architectures.

In addition to existing work that promotes the graph-based approach to define new

ANN architectures (Elsken et al. 2019; Leijnen et al. 2020) we introduce for the first time, the graph-based analysis of the connectivity at rest of ANNs.

We have used graph modeling and graph-based statistics to analyze trained ANNs in a continual learning framework. By studying different models which differently address the catastrophic forgetting issue, we show that it is possible to identify models with different performances based on simple features extracted from a binary graph obtained considering the strongest weights connection. We propose to identify critical hidden units according to the performance of pruned version model which turns off neurons according to their in-out degree values in the induced graph. The results show that the selected statistics and choice of hidden unit sets in the tail or interquartile range can be used to identify critical unit sets for the stability and plasticity of the corresponding configuration.

Finally, as it happens in the brain, we conclude that the hidden units in ANN are not homogeneous in recalling previously learned information or at adapting to newly learned information. These results lay down the foundations to study how the learning process in ANN using graph-theory tools, providing insights into the occurrence of catastrophic forgetting and the presence of a stability-critical set of neurons.

General Conclusion

General Conclusion

To conclude, this work explores the integration and synergistic collaboration across various multidisciplinary fields. We propose an integration between artificial systems practitioners, neuroscientists and network science researchers. Particularly, our objective is to propose a graph-based analysis of artificial neural network and brain connectivity. This translates into twofold objectives: on one side the proposal of an innovative way to characterize functional connectivity network groups, on the other side, the extension of a brain-inspired method for artificial neural networks analysis. As we have pointed out, a more stringent integration and collaboration between these domains have significant promise for advancing our understanding for both brain function and pathology and artificial neural network robustness.

Throughout this work, we have stressed the importance of establishing a gold-standard analysis framework for functional brain connectivity with graph-based approach. Potentially, such a framework will serve as a common ground, facilitating consistent interpretation of results and enhancing reliability. Furthermore, it is crucial for the neuroimaging community to prioritize ensuring the quality of data and establishing standardized protocols for acquiring resting-state fMRI data. This effort is essential to lay a strong foundation for the development of functional connectivity studies research. Despite the multitude of existing functional connectivity methods and studies, we stress the importance of the definition of a method which can detect the regional organization of brain network. We have identified the criteria a functional connectivity analysis method should fulfil: consider interpretable classical network statistics; demonstrate adaptability to diverse pathologies or datasets, allowing the customization of graph statistics to match specific application cases; provide a framework for characterizing groups; possess the capability to account for individual differences among subjects; and enable localized characterization, to facilitate the identification of regional disruptions. We make a significant contribution in this regard. We introduce a novel framework that leverages classical network statistics to identify structural patterns within graphs. We assess the adaptability of this proposal across various pathologies and by collecting data from multiple databases. We show that distinct pathologies necessitate the employment of specific nodal statistics to accurately differentiate pathological networks from healthy controls. Finally, we detect regional perturbation in line with literature results.

Coming to the integration of brain-inspired methodology to artificial neural network connectivity studies, we make efforts for adapting a brain analysis framework to the analysis of artificial neural networks. This approach is based on the definition of an induced-graph associated to a trained neural network architecture. Applied to continual learning strategies, our method effectively detects catastrophic forgetting through simple topological information. This approach has the potential to help in define more

robust artificial systems, more closely emulating the human brain connectivity. From this standpoint, it is crucial to embrace relevant metrics for evaluating robustness, for instance specifically focusing on plasticity and stability within the context of continual learning paradigms. Within the proposed experiments, we cast doubt upon the traditional concept of exchangeable units in artificial neural network. By evaluating the model's performance across a range of pruning scenarios, we manage to connect specific units to particular class predictions.

Although there is additional space for exploration, for instance for the use of induced-graph topological characteristics in the design of new continual learning strategies, our contribution effectively proposes an integration of network science principles into artificial neural networks analysis in a brain-inspired perspective.

Bibliography

- Abdallah, C. G., C. L. Averill, A. E. Ramage, L. A. Averill, S. Goktas, S. Nemati, J. H. Krystal, J. D. Roache, P. A. Resick, S. Young-McCaughan, A. L. Peterson, P. Fox and STRONG STAR Consortium, Saliency network disruption in U.s. army soldiers with posttraumatic stress disorder, *Chronic Stress (Thousand Oaks)* **3**, 247054701985046 (2019).
- Abdelnour, F., M. Dayan, O. Devinsky, T. Thesen and A. Raj, Functional brain connectivity is predictable from anatomic network's Laplacian eigen-structure, *Neuroimage* **172**, 728 (2018).
- Abdelnour, F., M. Dayan, O. Devinsky, T. Thesen and A. Raj, Algebraic relationship between the structural network's Laplacian and functional network's adjacency matrix is preserved in temporal lobe epilepsy subjects, *Neuroimage* **228**, 117705 (2021).
- Abraham, A., E. Dohmatob, B. Thirion, D. Samaras and G. Varoquaux, "Extracting brain regions from rest fMRI with total-variation constrained dictionary learning", *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16*, (2013).
- Abraham, W. C. and A. Robins, Memory retention – the synaptic stability versus plasticity dilemma, *Trends in Neurosciences* **28**, 73 (2005).
- Achard, S. and E. Bullmore, Efficiency and cost of economical brain functional networks, *PLoS Comput Biol* **3**, e17 (2007).
- Achard, S., J.-F. Coeurjolly, P. L. de Micheaux and J. Richiardi, Robust correlation for aggregated data with spatial characteristics, arXiv preprint arXiv:2011.08269 (2020).
- Achard, S., C. Delon-Martin, P. E. Vértes, F. Renard, M. Schenck, F. Schneider, C. Heinrich, S. Kremer and E. T. Bullmore, Hubs of brain functional networks are radically reorganized in comatose patients, *Proceedings of the National Academy of Sciences* **109**, 20608 (2012).
- Achard, S., R. Salvador, B. Whitcher, J. Suckling and E. Bullmore, A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs, *Journal of Neuroscience* **26**, 63 (2006).
- Adhikari, M. H., C. D. Hacker, J. S. Siegel, A. Griffa, P. Hagmann, G. Deco and M. Corbetta, Decreased integration and information capacity in stroke measured by whole brain models of resting state activity, *Brain* **140**, 1068 (2017).
- Adolphs, R., D. Tranel, H. Damasio and A. Damasio, Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala, *Nature* **372**, 669 (1994).
- Adolphs, R., D. Tranel, H. Damasio and A. R. Damasio, Fear and the human amygdala, *Journal of neuroscience* **15**, 5879 (1995).

- Agarwal, C., O. Queen, H. Lakkaraju and M. Zitnik, Evaluating explainability for graph neural networks, *Scientific Data* **10**, 144 (2023).
- Agcaoglu, O., T. W. Wilson, Y.-P. Wang, J. Stephen and V. D. Calhoun, Resting state connectivity differences in eyes open versus eyes closed conditions, *Human brain mapping* **40**, 2488 (2019).
- Aimone, J. B., J. Wiles and F. H. Gage, Computational Influence of Adult Neurogenesis on Memory Encoding, *Neuron* **61**, 187 (2009).
- Albert, R. and A.-L. Barabási, Statistical mechanics of complex networks, *Reviews of modern physics* **74**, 47 (2002).
- Ali, H., M. N. M. Salleh, R. Saedudin, K. Hussain and M. F. Mushtaq, Imbalance class problems in data mining: A review, *Indonesian Journal of Electrical Engineering and Computer Science* **14**, 1560 (2019).
- Altman, D. G. and J. M. Bland, Measurement in Medicine: The Analysis of Method Comparison Studies, *Journal of the Royal Statistical Society. Series D (The Statistician)* **32**, 307 (1983).
- Amiri, S., J. Mehvari-Habibabadi, N. Mohammadi-Mobarakeh, S. S. Hashemi-Fesharaki, M. M. Mirbagheri, K. Elisevich and M.-R. Nazem-Zadeh, Graph theory application with functional connectivity to distinguish left from right temporal lobe epilepsy, *Epilepsy Res.* **167**, 106449 (2020).
- Anderson, A. K. and E. A. Phelps, Expression without recognition: contributions of the human amygdala to emotional communication, *Psychological Science* **11**, 106 (2000).
- Arbel, M., R. Menegaux and P. Wolinski, *Rethinking Gauss-Newton for learning over-parameterized models* (2023).
- Audrain, S., A. J. Barnett and M. P. McAndrews, Language network measures at rest indicate individual differences in naming decline after anterior temporal lobe resection, *Hum. Brain Mapp.* **39**, 4404 (2018).
- Babai, L., “Graph isomorphism in quasipolynomial time”, *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, (2016).
- Bach, S., A. Binder, G. Montavon, F. Klauschen, K.-R. Müller and W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, *PloS one* **10**, e0130140 (2015).
- Bachmann, C., H. I. L. Jacobs, P. Porta Mana, K. Dillen, N. Richter, B. von Reutern, J. Dronse, O. A. Onur, K.-J. Langen, G. R. Fink, J. Kukolja and A. Morrison, On the extraction and analysis of graphs from resting-state fMRI to support a correct and robust diagnostic tool for Alzheimer’s disease, *Front. Neurosci.* **12**, 528 (2018).
- Baldominos, A., Y. Saez and P. Isasi, A survey of handwritten character recognition with mnist and emnist, *Applied Sciences* **9**, 3169 (2019).
- Ball, G., P. Aljabar, T. Arichi, N. Tusor, D. Cox, N. Merchant, P. Nongena, J. Hajnal, A. Edwards and S. Counsell, Machine-learning to characterise neonatal functional connectivity in the preterm brain, *NeuroImage* **124**, 267 (2016).
- Banka, A., I. Buzi and I. Rekik, “Multi-View Brain HyperConnectome AutoEncoder For Brain State Classification”, *PRIME*, (2020).

- Bansal, R. and B. S. Peterson, Use of random matrix theory in the discovery of resting state brain networks, *Magn. Reson. Imaging* **77**, 69 (2021).
- Barabási, A.-L., Network science, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20120375 (2013).
- Barabási, A.-L. and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
- Barry, R., F. M. De Blasio and A. E. Cave, Eyes-closed vs. eyes-open EEG in young and older adults (2014).
- Barsoum, E., C. Zhang, C. C. Ferrer and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution”, *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, (2016).
- Bashivan, P., K. Kar and J. J. DiCarlo, Neural population control via deep image synthesis, *Science* **364** (2019).
- Bassett, D. S., E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger and A. Meyer-Lindenberg, Hierarchical organization of human cortical networks in health and schizophrenia, *Journal of Neuroscience* **28**, 9239 (2008).
- Bassett, D. S. and E. T. Bullmore, Small-world brain networks revisited, *The Neuroscientist* **23**, 499 (2017).
- Bassett, D. S., E. T. Bullmore, A. Meyer-Lindenberg, J. A. Apud, D. R. Weinberger and R. Coppola, Cognitive fitness of cost-efficient brain functional networks, *Proceedings of the National Academy of Sciences* **106**, 11747 (2009).
- Bassett, D. S., D. L. Greenfield, A. Meyer-Lindenberg, D. R. Weinberger, S. W. Moore and E. T. Bullmore, Efficient physical embedding of topologically complex information processing networks in brains and computer circuits, *PLoS computational biology* **6**, e1000748 (2010).
- Bassett, D. S., A. Meyer-Lindenberg, S. Achard, T. Duke and E. Bullmore, Adaptive reconfiguration of fractal small-world human brain functional networks, *Proceedings of the National Academy of Sciences* **103**, 19518 (2006).
- Bassett, D. S., B. G. Nelson, B. A. Mueller, J. Camchong and K. O. Lim, Altered resting state complexity in schizophrenia, *Neuroimage* **59**, 2196 (2012).
- Bassett, D. S. and O. Sporns, Network neuroscience, *Nature neuroscience* **20**, 353 (2017).
- Bassett, D. S., P. Zurn and J. I. Gold, On the nature and use of models in network neuroscience, *Nature Reviews Neuroscience* **19**, 566 (2018).
- Bassett, D. S. and E. Bullmore, Small-world brain networks, *The neuroscientist* **12**, 512 (2006).
- Bavelas, A., Communication patterns in task-oriented groups, *The journal of the acoustical society of America* **22**, 725 (1950).
- Beckmann, C. F. and S. M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging, *IEEE transactions on medical imaging* **23**, 137 (2004).
- Behfar, Q., S. K. Behfar, B. von Reutern, N. Richter, J. Dronse, R. Fassbender, G. R. Fink and O. A. Onur, Graph theory analysis reveals resting-state compensatory mechanisms in healthy aging and prodromal Alzheimer’s disease, *Front. Aging Neurosci.* **12**, 576627 (2020).

- Bellec, P., C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies and R. C. Craddock, The neuro bureau ADHD-200 preprocessed repository, *Neuroimage* **144**, 275 (2017).
- Ben Simon, E., A. Maron-Katz, N. Lahav, R. Shamir and T. Hendler, Tired and mis-connected: A breakdown of brain modularity following sleep deprivation, *Hum. Brain Mapp.* **38**, 3300 (2017).
- Bengio, Y., D.-H. Lee, J. Bornschein, T. Mesnard and Z. Lin, Towards biologically plausible deep learning, arXiv preprint arXiv:1502.04156 (2015).
- Berman, B. D., J. Smucny, K. P. Wylie, E. Shelton, E. Kronberg, M. Leehey and J. R. Tregellas, Levodopa modulates small-world architecture of functional brain networks in Parkinson’s disease, *Movement Disorders* **31**, 1676 (2016).
- Bessadok, A., M. A. Mahjoub and I. Rezik, Graph neural networks in network neuroscience, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- Bessadok, A., M. A. Mahjoub and I. Rezik, Graph neural networks in network neuroscience, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- Bhanot, A., G. Becq, S. Achard and E. L. Barbier, “Analyse de la connectivité fonctionnelle en IRMf : graphes ou ICA ?”, *Proceedings of the [Full Conference Name Abbreviation] [Year]*, (2023).
- Bhanot, A., C. Meillier, F. Heitz and L. Harsan, “Online dictionary learning for single-subject fMRI data unmixing”, *2019 27th European Signal Processing Conference (EUSIPCO)*, (2019).
- Bharath, R. D., G. Chaitanya, R. Panda, K. Raghavendra, S. Sinha, A. Sahoo, S. Gohel, B. B. Biswal and P. Satishchandra, Reduced small world brain connectivity in probands with a family history of epilepsy, *Eur. J. Neurol.* **23**, 1729 (2016).
- Bhidayasiri, R., D. Tarsy, R. Bhidayasiri and D. Tarsy, Parkinson’s disease: Hoehn and Yahr scale, *Movement Disorders: A Video Atlas: A Video Atlas 4* (2012).
- Bianconi, G., The entropy of randomized network ensembles, *EPL (Europhysics Letters)* **81**, 28005 (2007).
- Bianconi, G., Entropy of network ensembles, *Physical Review E* **79**, 036114 (2009).
- Bilgen, I., G. Guvercin and I. Rezik, Machine learning methods for brain network classification: Application to autism diagnosis using cortical morphological networks, *Journal of Neuroscience Methods* **343**, 108799 (2020).
- Billings, J. C. W., G. J. Thompson, W.-J. Pan, M. E. Magnuson, A. Medda and S. Keilholz, Disentangling multispectral functional connectivity with wavelets, *Front. Neurosci.* **12**, 812 (2018).
- Biswal, B., F. Zerrin Yetkin, V. M. Houghton and J. S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI, *Magnetic resonance in medicine* **34**, 537 (1995).
- Biswal, B. B., M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, et al., Toward discovery science of human brain function, *Proceedings of the national academy of sciences* **107**, 4734 (2010).

- Bland, J. M. and D. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, *The lancet* **327**, 307 (1986).
- Blier, L., P. Wolinski and Y. Ollivier, *Learning with Random Learning Rates* (2019).
- Blommaert, J., A. Radwan, C. Sleurs, C. Maggen, M. van Gerwen, V. Wolters, D. Christiaens, R. Peeters, P. Dupont, S. Sunaert, K. Van Calsteren, S. Deprez and F. Amant, The impact of cancer and chemotherapy during pregnancy on child neurodevelopment: A multimodal neuroimaging analysis, *EClinicalMedicine* **28**, 100598 (2020).
- Bluhm, R., P. Williamson, R. Lanius, J. Théberge, M. Densmore, R. Bartha, R. Neufeld and E. Osuch, Resting state default-mode network connectivity in early depression using a seed region-of-interest analysis: Decreased connectivity with caudate nucleus, *Psychiatry and clinical neurosciences* **63**, 754 (2009).
- Bobholz, J. A., S. M. Rao, A. J. Saykin and N. Pliskin, Clinical use of functional magnetic resonance imaging: reflections on the new CPT codes, *Neuropsychology review* **17**, 189 (2007).
- Bogacz, L., Z. Burda and B. Waclaw, Homogeneous complex networks, *Physica A: Statistical Mechanics and its Applications* **366**, 587 (2006).
- Bordier, C., C. Nicolini and A. Bifone, Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold, *Frontiers in neuroscience* **11**, 441 (2017).
- Borgatti, S. P. and M. G. Everett, Notions of position in social network analysis, *Sociological methodology* **1** (1992).
- Borgatti, S. P., A. Mehra, D. J. Brass and G. Labianca, Network analysis in the social sciences, *Science* **323**, 892 (2009).
- Borgatti, S. P. and M. G. Everett, Notions of Position in Social Network Analysis, *Sociological Methodology* **22**, 1 (1992).
- Borgwardt, K. M., C. S. Ong, S. Schönauer, S. Vishwanathan, A. J. Smola and H.-P. Kriegel, Protein function prediction via graph kernels, *Bioinformatics* **21**, i47 (2005).
- Bottani, S., N. Burgos, A. Maire, A. Wild, S. Ströer, D. Dormont, O. Colliot and {APPRIMAGE Study Group}, Automatic quality control of brain T1-weighted magnetic resonance images for a clinical data warehouse, *Medical Image Analysis* **75**, 102219 (2022).
- Botvinik-Nezer, R., F. Holzmeister, C. F. Camerer, A. Dreber, J. Huber, M. Johannesson, M. Kirchler, R. Iwanir, J. A. Mumford, R. A. Adcock, et al., Variability in the analysis of a single neuroimaging dataset by many teams, *Nature* **582**, 84 (2020).
- Boukli Hacene, G., V. Gripon, N. Farrugia, M. Arzel and M. Jezequel, Transfer Incremental Learning Using Data Augmentation, *Applied Sciences* **8** (2018).
- Breiter, H. C., N. L. Etcoff, P. J. Whalen, W. A. Kennedy, S. L. Rauch, R. L. Buckner, M. M. Strauss, S. E. Hyman and B. R. Rosen, Response and habituation of the human amygdala during visual processing of facial expression, *Neuron* **17**, 875 (1996).

- Brenner, E. K., B. M. Hampstead, E. C. Grossner, R. A. Bernier, N. Gilbert, K. Sathian and F. G. Hillary, Diminished neural network dynamics in amnesic mild cognitive impairment, *Int. J. Psychophysiol.* **130**, 63 (2018).
- Broido, A. D. and A. Clauset, Scale-free networks are rare, *Nature communications* **10**, 1017 (2019).
- Bronstein, M. M., J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, Geometric Deep Learning: Going beyond Euclidean data, *IEEE Signal Processing Magazine* **34**, 18 (2017).
- Brossard, C., B. Lemasson, A. Attyé, J.-A. De Busschère, J.-F. Payen, E. L. Barbier, J. Grèze and P. Bouzat, Contribution of CT-scan analysis by artificial intelligence to the clinical care of TBI patients, *Frontiers in Neurology* **12**, 666875 (2021).
- Bullmore, E. and O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, *Nature reviews neuroscience* **10**, 186 (2009).
- Bullmore, E. and O. Sporns, The economy of brain network organization, *Nature reviews neuroscience* **13**, 336 (2012).
- Buzzega, P., M. Boschini, A. Porrello, D. Abati and S. Calderara, Dark experience for general continual learning: a strong, simple baseline, *Advances in neural information processing systems* **33**, 15920 (2020).
- Cackowski, S., E. L. Barbier, M. Dojat and T. Christen, Imunity: a generalizable vae-gan solution for multicenter mr image harmonization, *Medical Image Analysis* **88**, 102799 (2023).
- Calder, A. J., Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear, *Cognitive Neuropsychology* **13**, 699 (1996).
- Calhoun, V. D., T. Adali, G. D. Pearlson and J. J. Pekar, A method for making group inferences from functional MRI data using independent component analysis, *Human brain mapping* **14**, 140 (2001).
- Calhoun, V. D., J. Liu and T. Adali, A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data, *Neuroimage* **45**, S163 (2009).
- Campabadal, A., A. Abos, B. Segura, M. Serradell, C. Uribe, H. C. Baggio, C. Gaig, J. Santamaria, Y. Compta, N. Bargallo, C. Junque and A. Iranzo, Disruption of posterior brain functional connectivity and its relation to cognitive impairment in idiopathic REM sleep behavior disorder, *NeuroImage Clin.* **25**, 102138 (2020).
- Cao, M., Y. He, Z. Dai, X. Liao, T. Jeon, M. Ouyang, L. Chalak, Y. Bi, N. Rollins, Q. Dong and H. Huang, Early development of functional network segregation revealed by connectomic analysis of the preterm human brain, *Cereb. Cortex* **27**, 1949 (2017).
- Carboni, L., S. Achard and M. Dojat, "Network embedding for brain connectivity", *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, (2021).
- Carboni, L., S. Achard and M. Dojat, "Stationary distribution of node2vec random walk for brain connectivity", *CCS21 Conference on Complex Systems, October 2021, Lyon, France*, (2021).

- Carboni, L., S. Achard and M. Dojat, “Nodal-statistics-based structural pattern detection for graph collections characterization”, *CCS22 Conference on Complex Systems, October 2022, Palma de Mallorca, Spain*, (2022).
- Carboni, L., S. Achard and M. Dojat, “Différencier les réseaux de connectivité cérébrale fonctionnelle de repos : apport de la combinaison de statistiques nodales de graphes”, *SFRMBM - 6ème congrès de la Société Française de Résonance Magnétique en Biologie et Médecine*, (2023).
- Carboni, L., M. Dojat and S. Achard, Nodal statistics-based equivalence relation for graph collections, *Physical Reviews E* **107**, 014302-1 (2023).
- Carboni, L., M. Dojat and S. Achard, “Nodal equivalence relation for brain connectivity analysis”, *XXIXème Colloque Francophone de Traitement du Signal et des Images (GRETSI 2023)*, (2023).
- Carboni, L., D. Nwaigwe, M. Mainsant, R. Bayle, M. Reyboz, M. Mermillod, M. Dojat and S. Achard, Exploring Continual Learning Strategies in Artificial Neural Networks through Graph-Based Analysis of Connectivity: Insights from a Brain-Inspired Perspective (2023).
- Cardillo, A., J. Gómez-Gardenes, M. Zanin, M. Romance, D. Papo, F. d. Pozo and S. Boccaletti, Emergence of network features from multiplexity, *Scientific reports* **3**, 1 (2013).
- Carotenuto, A., H. Wilson, B. Giordano, S. P. Caminiti, Z. Chappell, S. C. R. Williams, A. Hammers, E. Silber, P. Brex and M. Politis, Impaired connectivity within neuromodulatory networks in multiple sclerosis and clinical implications, *J. Neurol.* **267**, 2042 (2020).
- Cary, R. P., S. Ray, D. S. Grayson, J. Painter, S. Carpenter, L. Maron, O. Sporns, A. A. Stevens, J. T. Nigg and D. A. Fair, Network structure among brain systems in adult ADHD is uniquely modified by stimulant administration, *Cereb. Cortex* **27**, 3970 (2017).
- Casanova, R., C. Whitlow, B. Wagner, M. Espeland and J. Maldjian, Combining graph and machine learning methods to analyze differences in functional connectivity across sex, *The open neuroimaging journal* **6**, 1 (2012).
- Casimo, K., Spontaneous and task-related changes in resting state connectivity (PhD thesis, University of Washington).
- Cauchy, A. et al., Méthode générale pour la résolution des systèmes d'équations simultanées, *Comp. Rend. Sci. Paris* **25**, 536 (1847).
- Chang, Y., W. Li, J. Peng, B. Tang, Y. Kang, Y. Lei, Y. Gui, Q. Zhu, Y. Liu and H. Li, Reviewing continual learning from the perspective of human-level intelligence, *arXiv preprint arXiv:2111.11964* (2021).
- Chen, B., Abnormal cortical region and subsystem complexity in dynamical functional connectivity of chronic schizophrenia: A new graph index for fMRI analysis, *J. Neurosci. Methods* **311**, 28 (2019).
- Chen, B., Abnormal cortical region and subsystem complexity in dynamical functional connectivity of chronic schizophrenia: A new graph index for fMRI analysis, *J. Neurosci. Methods* **311**, 28 (2019).
- Chen, G., H.-Y. Zhang, C. Xie, G. Chen, Z.-J. Zhang, G.-J. Teng and S.-J. Li, Modular reorganization of brain resting state networks and its independent vali-

- dation in Alzheimer's disease patients, *Frontiers in human neuroscience* **7**, 456 (2013).
- Chen, Y., S. Wang, C. C. Hilgetag and C. Zhou, Trade-off between multiple constraints enables simultaneous formation of modules and hubs in neural systems, *PLoS computational biology* **9**, e1002937 (2013).
- Chen, Y., J. Pu, X. Liu and X. Zhang, Gaussian mixture embedding of multiple node roles in networks, *World Wide Web* **23**, 927 (2020).
- Chen, Z. J., Y. He, P. Rosa-Neto, J. Germann and A. C. Evans, Revealing modular architecture of human brain structural networks by using cortical thickness from MRI, *Cerebral cortex* **18**, 2374 (2008).
- Chen, Z., X. Hu, Q. Chen and T. Feng, Altered structural and functional brain network overall organization predict human intertemporal decision-making, *Human Brain Mapping* **40**, 306 (2019).
- Chen, Z. and B. Liu, *Lifelong machine learning* vol. 1 (Springer).
- Cheng, H., S. Newman, J. Goñi, J. S. Kent, J. Howell, A. Bolbecker, A. Puce, B. F. O'Donnell and W. P. Hetrick, Nodal centrality of functional network in the differentiation of schizophrenia, *Schizophrenia research* **168**, 345 (2015).
- Cheung, M., J. Shi, O. Wright, L. Y. Jiang, X. Liu and J. M. Moura, Graph signal processing and deep learning: Convolution, pooling, and topology, *IEEE Signal Processing Magazine* **37**, 139 (2020).
- Chiominto, S., F. Pappone, H. Lbath, L. Carboni and S. Achard, "Homological information to refine small-world regime of brain functional connectivity networks", *CCS22 Conference on Complex Systems, October 2022, Palma de Mallorca, Spain*, (2022).
- Chirumamilla, V. C., V. Fleischer, A. Droby, T. Anjum, M. Muthuraman, F. Zipp and S. Groppa, Functional connectivity analysis using whole brain and regional network metrics in MS patients, *Annu Int Conf IEEE Eng Med Biol Soc* **2016**, 4039 (2016).
- Chockanathan, U., A. M. DSouza, A. Z. Abidin, G. Schifitto and A. Wismüller, Identification and functional characterization of HIV-associated neurocognitive disorders with large-scale Granger causality analysis on resting-state functional MRI, *Proc. SPIE* **10575** (2018).
- Chockanathan, U., A. M. DSouza, A. Z. Abidin, G. Schifitto and A. Wismüller, Automated diagnosis of HIV-associated neurocognitive disorders using large-scale Granger causality analysis of resting-state functional MRI, *Comput. Biol. Med.* **106**, 24 (2019).
- Choi, I., G. Wu and W. H. Kim, "How Much to Aggregate: Learning Adaptive Node-Wise Scales on Graphs for Brain Networks", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2022).
- Christen, T., N. Pannetier, W. W. Ni, D. Qiu, M. E. Moseley, N. Schuff and G. Zaharchuk, MR vascular fingerprinting: A new approach to compute cerebral blood volume, mean vessel radius, and oxygenation maps in the human brain, *Neuroimage* **89**, 262 (2014).
- Chung, D. and I. Sohn, Neural Network Optimization Based on Complex Network Theory: A Survey, *Mathematics* **11**, 321 (2023).

- Cichon, J. and W.-B. Gan, Branch-specific dendritic Ca²⁺ spikes cause persistent synaptic plasticity, *Nature* **520**, 180 (2015).
- Ciregan, D., U. Meier and J. Schmidhuber, “Multi-column deep neural networks for image classification”, *2012 IEEE conference on computer vision and pattern recognition*, (2012).
- Clopath, C., Synaptic consolidation: an approach to long-term learning, *Cognitive neurodynamics* **6**, 251 (2012).
- COBRE (2012), *The Mind Research Network and the University of New Mexico funded by a National Institute of Health Center of Biomedical Research Excellence (COBRE)*.
- Cole, M. W., S. Pathak and W. Schneider, Identifying the brain’s most globally connected regions, *Neuroimage* **49**, 3132 (2010).
- Collantoni, E., P. Meneguzzo, M. Solmi, E. Tenconi, R. Manara and A. Favaro, Functional connectivity patterns and the role of 5-HTTLPR polymorphism on network architecture in female patients with anorexia nervosa, *Front. Neurosci.* **13**, 1056 (2019).
- Cope, T. E., T. Rittman, R. J. Borchert, P. S. Jones, D. Vatansever, K. Allinson, L. Passamonti, P. Vazquez Rodriguez, W. R. Bevan-Jones, J. T. O’Brien and J. B. Rowe, Tau burden and the functional connectome in Alzheimer’s disease and progressive supranuclear palsy, *Brain* **141**, 550 (2018).
- Cordes, D., V. M. Haughton, K. Arfanakis, G. J. Wendt, P. A. Turski, C. H. Moritz, M. A. Quigley and M. E. Meyerand, Mapping functionally related regions of brain with functional connectivity MR imaging, *American journal of neuroradiology* **21**, 1636 (2000).
- Cordova, M., K. Shada, D. V. Demeter, O. Doyle, O. Miranda-Dominguez, A. Perrone, E. Schifsky, A. Graham, E. Fombonne, B. Langhorst, J. Nigg, D. A. Fair and E. Feczko, Heterogeneity of executive function revealed by a functional random forest approach across ADHD and ASD, *NeuroImage Clin.* **26**, 102245 (2020).
- Corneanu, C. A., M. Madadi, S. Escalera and A. M. Martinez, “What does it mean to learn in deep networks? And, how does one detect adversarial attacks?”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2019).
- Craddock, R. C., P. E. Holtzheimer III, X. P. Hu and H. S. Mayberg, Disease state prediction from resting state functional connectivity, *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **62**, 1619 (2009).
- Crossley, N. A., A. Mechelli, J. Scott, F. Carletti, P. T. Fox, P. McGuire and E. T. Bullmore, The hubs of the human connectome are generally implicated in the anatomy of brain disorders, *Brain* **137**, 2382 (2014).
- Cui, P., X. Wang, J. Pei and W. Zhu, A survey on network embedding, *IEEE Transactions on Knowledge and Data Engineering* **31**, 833 (2018).
- Cwiek, A., S. M. Rajtmajer, B. Wyble, V. Honavar, E. Grossner and F. G. Hillary, Feeding the machine: Challenges to reproducible predictive modeling in resting-state connectomics, *Network Neuroscience* **6**, 29 (2022).

- Dadi, K., M. Rahim, A. Abraham, D. Chyzyk, M. Milham, B. Thirion, G. Varoquaux, A. D. N. Initiative, et al., Benchmarking functional connectome-based predictive models for resting-state fMRI, *NeuroImage* **192**, 115 (2019).
- Dai, P., T. Xiong, X. Zhou, Y. Ou, Y. Li, X. Kui, Z. Chen, B. Zou, W. Li, Z. Huang, et al., The alterations of brain functional connectivity networks in major depressive disorder detected by machine learning through multisite rs-fMRI data, *Behavioural Brain Research* **435**, 114058 (2022).
- Dall'Acqua, P., S. Johannes, L. Mica, H.-P. Simmen, R. Glaab, J. Fandino, M. Schwendinger, C. Meier, E. J. Ulbrich, A. Müller, H. Baetschmann, L. Jäncke and J. Hänggi, Functional and structural network recovery after mild traumatic brain injury: A 1-year longitudinal study, *Front. Hum. Neurosci.* **11**, 280 (2017).
- Davis, M. and P. J. Whalen, The amygdala: vigilance and emotion, *Molecular psychiatry* **6**, 13 (2001).
- De Arruda, G. F., A. L. Barbieri, P. M. Rodriguez, F. A. Rodrigues, Y. Moreno and L. da Fontoura Costa, Role of centrality for the identification of influential spreaders in complex networks, *Physical Review E* **90**, 032812 (2014).
- De Domenico, M., V. Nicosia, A. Arenas and V. Latora, Structural reducibility of multilayer networks, *Nature communications* **6**, 1 (2015).
- De Lange, M., R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh and T. Tuytelaars, A continual learning survey: Defying forgetting in classification tasks, *IEEE transactions on pattern analysis and machine intelligence* **44**, 3366 (2021).
- De Micco, R., F. Agosta, S. Basaia, M. Siciliano, C. Cividini, G. Tedeschi, M. Filippi and A. Tessitore, Functional connectomics and disease progression in drug-naïve Parkinson's disease patients, *Mov. Disord.* **36**, 1603 (2021).
- De Schipper, L. J., A. Hafkemeijer, J. Van der Grond, J. Marinus, J. M. Henselmans and J. J. Van Hilten, Altered whole-brain and network-based functional connectivity in Parkinson's disease, *Frontiers in neurology* **9**, 419 (2018).
- De Vico Fallani, F., V. Latora and M. Chavez, A topological criterion for filtering information in complex brain networks, *PLoS computational biology* **13**, e1005305 (2017).
- Deco, G., G. Tononi, M. Boly and M. L. Kringelbach, Rethinking segregation and integration: contributions of whole-brain modelling, *Nature Reviews Neuroscience* **16**, 430 (2015).
- Del Genio, C. I., T. Gross and K. E. Bassler, All scale-free networks are sparse, *Physical review letters* **107**, 178701 (2011).
- Delphin, A., *Développements IRM Fingerprint pour la mesure de l'oxygénation cérébrale* (PhD thesis, Université Grenoble Alpes).
- Deng, L., The mnist database of handwritten digit images for machine learning research, *IEEE Signal Processing Magazine* **29**, 141 (2012).
- Deng, Y., K. Liu, L. Shi, Y. Lei, P. Liang, K. Li, W. C. W. Chu, D. Wang and Alzheimer's Disease Neuroimaging Initiative, Identifying the alteration patterns of brain functional connectivity in progressive mild cognitive impairment patients: A longitudinal whole-brain voxel-wise degree analysis, *Front. Aging Neurosci.* **8**, 195 (2016).

- Desikan, R. S., F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al., An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest, *Neuroimage* **31**, 968 (2006).
- Di Martino, A., C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, et al., The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, *Molecular psychiatry* **19**, 659 (2014).
- Dichio, V. and F. De Vico Fallani, Statistical models of complex brain networks: a maximum entropy approach, *Reports on Progress in Physics* (2023).
- Dichio, V. and F. D. V. Fallani, The exploration-exploitation paradigm for networked biological systems, arXiv preprint arXiv:2306.17300 (2023).
- Diez, I., L. Ortiz-Terán, B. Williams, R. Jalilianhasanpour, J. P. Ospina, B. C. Dickerson, M. S. Keshavan, W. C. LaFrance Jr, J. Sepulcre and D. L. Perez, Corticolimbic fast-tracking: enhanced multimodal integration in functional neurological disorder, *J. Neurol. Neurosurg. Psychiatry* **90**, 929 (2019).
- Djulonga, J., J. Yung, M. Tschannen, R. Romijnders, L. Beyer, A. Kolesnikov, J. Puigcerver, M. Minderer, A. D'Amour, D. Moldovan, et al., "On robustness and transferability of convolutional neural networks", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021).
- Doerig, A., R. P. Sommers, K. Seeliger, B. Richards, J. Ismael, G. W. Lindsay, K. P. Kording, T. Konkle, M. A. Van Gerven, N. Kriegeskorte, et al., The neuroconnectionist research programme, *Nature Reviews Neuroscience* **1** (2023).
- Dong, G., L. Yang, C.-s. R. Li, X. Wang, Y. Zhang, W. Du, Y. Han and X. Tang, Dynamic network connectivity predicts subjective cognitive decline: the Sino-Longitudinal Cognitive impairment and dementia study, *Brain imaging and behavior* **14**, 2692 (2020).
- Douw, L., M. Quaak, S. M. D. D. Fitzsimmons, S. J. de Wit, Y. D. van der Werf, O. A. van den Heuvel and C. Vriend, Static and dynamic network properties of the repetitive transcranial magnetic stimulation target predict changes in emotion regulation in obsessive-compulsive disorder, *Brain Stimul.* **13**, 318 (2020).
- Draeos, T. J., N. E. Miner, C. C. Lamb, J. A. Cox, C. M. Vineyard, K. D. Carlson, W. M. Severa, C. D. James and J. B. Aimone, "Neurogenesis deep learning: Extending deep networks to accommodate new classes", *2017 international joint conference on neural networks (IJCNN)*, (2017).
- Drakesmith, M., K. Caeyenberghs, A. Dutt, S. Zammit, C. J. Evans, A. Reichenberg, G. Lewis, A. S. David and D. K. Jones, Schizophrenia-like topological changes in the structural connectome of individuals with subclinical psychotic experiences, *Human brain mapping* **36**, 2629 (2015).
- Du, Y., Z. Fu and V. D. Calhoun, Classification and prediction of brain disorders using functional connectivity: promising but challenging, *Frontiers in neuroscience* **12**, 525 (2018).

- Duncan, E. S. and S. L. Small, Increased modularity of resting state networks supports improved narrative production in aphasia recovery, *Brain Connect.* **6**, 524 (2016).
- Eguiluz, V. M., D. R. Chialvo, G. A. Cecchi, M. Baliki and A. V. Apkarian, Scale-free brain functional networks, *Physical review letters* **94**, 018102 (2005).
- Ekman, P., Are there basic emotions?, *Psychological Review* **99(3)** (1992).
- Ekman, P., Facial expressions of emotion: an old controversy and new findings, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **335**, 63 (1992).
- Ekman, P. and W. V. Friesen, Facial action coding system, *Environmental Psychology & Nonverbal Behavior* (1978).
- El Ouahidi, Y., H. Tessier, G. Lioi, N. Farrugia, B. Padeloup and V. Gripon, “Pruning Graph Convolutional Networks to select meaningful graph frequencies for fMRI decoding”, *2022 30th European Signal Processing Conference (EUSIPCO)*, (2022).
- Eldaief, M. C., S. McMains, R. M. Hutchison, M. A. Halko and A. Pascual-Leone, Reconfiguration of intrinsic functional coupling patterns following circumscribed network lesions, *Cereb. Cortex* **27**, 2894 (2017).
- Elsken, T., J. H. Metzen and F. Hutter, Neural architecture search: A survey, *The Journal of Machine Learning Research* **20**, 1997 (2019).
- Erdeniz, B., E. Serin, Y. İbadi and C. Taş, Decreased functional connectivity in schizophrenia: The relationship between social functioning, social cognition and graph theoretical network measures, *Psychiatry Res. Neuroimaging* **270**, 22 (2017).
- Erdős, P. and A. Rényi, On Random Graphs I, *Publicationes Mathematicae Debrecen* **6**, 290 (1959).
- Erkaymaz, O., M. ÖZER and N. Yumuşak, Impact of small-world topology on the performance of a feed-forward artificial neural network based on 2 different real-life problems, *Turkish Journal of Electrical Engineering and Computer Sciences* **22**, 708 (2014).
- Erkaymaz, O., M. Özer and N. Yumuşak, Performance analysis of a feed-forward artificial neural network with small-world topology, *Procedia Technology* **1**, 291 (2012).
- Esteva, A., B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *nature* **542**, 115 (2017).
- Euler, L., *Solutio problematis ad geometriam situs pertinentis*, *Commentarii academiae scientiarum Petropolitanae* 128 (1741).
- Fan, L., H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, et al., The human brainnetome atlas: a new brain atlas based on connectional architecture, *Cerebral cortex* **26**, 3508 (2016).
- Fan, X., Y. Wu, L. Cai, J. Ma, N. Pan, X. Xu, T. Sun, J. Jing and X. Li, The differences in the whole-brain functional network between Cantonese-Mandarin bilinguals and Mandarin monolinguals, *Brain Sci.* **11**, 310 (2021).

- Fang, J., H. Chen, Z. Cao, Y. Jiang, L. Ma, H. Ma and T. Feng, Impaired brain network architecture in newly diagnosed Parkinson's disease based on graph theoretical analysis, *Neuroscience letters* **657**, 151 (2017).
- Ferrarini, L., I. M. Veer, E. Baerends, M.-J. van Tol, R. J. Renken, N. J. van der Wee, D. J. Veltman, A. Aleman, F. G. Zitman, B. W. Penninx, et al., Hierarchical functional modularity in the resting-state human brain, *Human brain mapping* **30**, 2220 (2009).
- Figueroa-Jimenez, M. D., C. Cañete-Massé, M. Carbó-Carreté, D. Zarabozo-Hurtado, M. Perú-Cebollero, J. G. Salazar-Estrada and J. Guàrdia-Olmos, Resting-state default mode network connectivity in young individuals with Down syndrome, *Brain Behav.* **11**, e01905 (2021).
- Filippis, R. de, E. A. Carbone, R. Gaetano, A. Bruni, V. Pugliese, C. Segura-Garcia and P. De Fazio, Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review, *Neuropsychiatric disease and treatment* 1605 (2019).
- Finotelli, P., O. Dipasquale, I. Costantini, A. Pini, F. Baglio, G. Baselli, P. Dulio and M. Cercignani, Exploring resting-state functional connectivity invariants across the lifespan in healthy people by means of a recently proposed graph theoretical model, *PLoS One* **13**, e0206567 (2018).
- Finotelli, P., C. G. Forlim, L. Klock, A. Pini, J. Bächle, L. Stoll, P. Giemsa, M. Fuchs, N. Schoofs, C. Montag, et al., New graph-theoretical-multimodal approach using temporal and structural correlations reveals disruption in the thalamo-cortical network in patients with schizophrenia, *Brain Connectivity* **9**, 760 (2019).
- Finotelli, P., C. Piccardi, E. Miglio and P. Dulio, A Graphlet-Based Topological Characterization of the Resting-State Network in Healthy People, *Frontiers in neuroscience* **15** (2021).
- Finotelli, P., C. Piccardi, E. Miglio and P. Dulio, A graphlet-based topological characterization of the resting-state network in healthy people, *Front. Neurosci.* **15**, 665544 (2021).
- Fitzsimmons, S. M. D. D., L. Douw, O. A. van den Heuvel, Y. D. van der Werf and C. Vriend, Resting-state and task-based centrality of dorsolateral prefrontal cortex predict resilience to 1 Hz repetitive transcranial magnetic stimulation, *Hum. Brain Mapp.* **41**, 3161 (2020).
- Fornito, A., A. Zalesky and M. Breakspear, Graph analysis of the human connectome: promise, progress, and pitfalls, *Neuroimage* **80**, 426 (2013).
- Fox, M. D. and M. Greicius, Clinical applications of resting state functional connectivity, *Frontiers in systems neuroscience* **4**, 1443 (2010).
- Fox, M. D. and M. E. Raichle, Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging, *Nature reviews neuroscience* **8**, 700 (2007).
- Fox, M. D., A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen and M. E. Raichle, The human brain is intrinsically organized into dynamic, anti-correlated functional networks, *Proceedings of the National Academy of Sciences* **102**, 9673 (2005).

- Frankle, J. and M. Carbin, The lottery ticket hypothesis: Finding small, trainable neural networks, arXiv preprint arXiv:1803.03635 (2018).
- Freeman, L. C., A set of measures of centrality based on betweenness, *Sociometry* **35** (1977).
- French, R. M., Catastrophic forgetting in connectionist networks, *Trends in cognitive sciences* **3**, 128 (1999).
- Friston, K. J., Functional and effective connectivity: a review, *Brain connectivity* **1**, 13 (2011).
- Fujimoto, U., A. Ogawa, T. Osada, M. Tanaka, A. Suda, N. Hattori, K. Kamagata, S. Aoki and S. Konishi, Network centrality reveals dissociable brain activity during response inhibition in human right ventral part of inferior frontal cortex, *Neuroscience* **433**, 163 (2020).
- Fulong, X., K. Spruyt, L. Chao, Z. Dianjiang, Z. Jun and H. Fang, Resting-state brain network topological properties and the correlation with neuropsychological assessment in adolescent narcolepsy, *Sleep* **43** (2020).
- Fürer, M., “On the combinatorial power of the Weisfeiler-Lehman algorithm”, *International Conference on Algorithms and Complexity*, (2017).
- Fusar-Poli, P., A. Placentino, F. Carletti, P. Landi, P. Allen, S. Surguladze, F. Benedetti, M. Abbamonte, R. Gasparotti, F. Barale, et al., Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies, *Journal of Psychiatry and Neuroscience* **34**, 418 (2009).
- Gao, L.-l. and T. Wu, The study of brain functional connectivity in Parkinson’s disease, *Translational neurodegeneration* **5**, 1 (2016).
- Gaudio, S., G. Olivo, B. Beomonte Zobel and H. B. Schiöth, Altered cerebellar-insular-parietal-cingular subnetwork in adolescents in the earliest stages of anorexia nervosa: a network-based statistic analysis, *Transl. Psychiatry* **8**, 127 (2018).
- Gerchen, M. F. and P. Kirsch, Combining task-related activation and connectivity analysis of fMRI data reveals complex modulation of brain networks, *Hum. Brain Mapp.* **38**, 5726 (2017).
- Gholipour, T., X. You, S. M. Stufflebeam, M. Loew, M. Z. Koubeissi, V. L. Morgan and W. D. Gaillard, Common functional connectivity alterations in focal epilepsies identified by machine learning, *Epilepsia* **63**, 629 (2022).
- Giavarina, D., Understanding bland altman analysis, *Biochemia medica* **25**, 141 (2015).
- Gilpin, L. H., D. Bau, B. Z. Yuan, A. Bajwa, M. Specter and L. Kagal, “Explaining explanations: An overview of interpretability of machine learning”, *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, (2018).
- Gilson, M., N. E. Kouvaris, G. Deco, J.-F. Mangin, C. Poupon, S. Lefranc, D. Rivière and G. Zamora-López, Network analysis of whole-brain fMRI dynamics: A new framework based on dynamic communicability, *Neuroimage* **201**, 116007 (2019).
- Göktaş, A. S., A. Bessadok and I. Reşik, “Residual Embedding Similarity-Based Network Selection for Predicting Brain Network Evolution Trajectory from a Single Observation”, *Predictive Intelligence in Medicine*, (2020).

- Gómez, A., O. L. Quintero, N. Lopez-Celani and L. F. Villa, “Emotional Networked maps from EEG signals”, *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, (2020).
- Goodfellow, I. J., D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al., “Challenges in representation learning: A report on three machine learning contests”, *International conference on neural information processing*, (2013).
- Goodfellow, I. J., M. Mirza, D. Xiao, A. Courville and Y. Bengio, *An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks* (2013), URL: <https://arxiv.org/abs/1312.6211>.
- Goodfellow, I. J., J. Shlens and C. Szegedy, “Explaining and Harnessing Adversarial Examples”, *3rd International Conference on Learning Representations*, (2015).
- Graña, M. and M. Silva, Impact of machine learning pipeline choices in autism prediction from functional connectivity data, *International journal of neural systems* **31**, 2150009 (2021).
- Grandjean, J., G. Desrosiers-Gregoire, C. Anckaerts, D. Angeles-Valdez, F. Ayad, D. A. Barrière, I. Blockx, A. Bortel, M. Broadwater, B. M. Cardoso, et al., A consensus protocol for functional connectivity analysis in the rat brain, *Nature neuroscience* **26**, 673 (2023).
- Grohe, M. and P. Schweitzer, The graph isomorphism problem, *Communications of the ACM* **63**, 128 (2020).
- Grover, A. and J. Leskovec, “node2vec: Scalable feature learning for networks”, *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, (2016).
- Guo, M., Y. Ren, H. Yu, H. Yang, C. Cao, Y. Li and G. Fan, Alterations in degree centrality and functional connectivity in Parkinson’s disease patients with freezing of gait: a resting-state functional magnetic resonance imaging study, *Frontiers in neuroscience* **14**, 582079 (2020).
- Haan, W. de, W. M. van der Flier, H. Wang, P. F. Van Mieghem, P. Scheltens and C. J. Stam, Disruption of functional brain networks in Alzheimer’s disease: what can we learn from graph spectral analysis of resting-state magnetoencephalography?, *Brain connectivity* **2**, 45 (2012).
- Hadsell, R., D. Rao, A. A. Rusu and R. Pascanu, Embracing Change: Continual Learning in Deep Neural Networks, *Trends in Cognitive Sciences* **24**, 1028 (2020).
- Hahn, G., G. Zamora-López, L. Uhrig, E. Tagliazucchi, H. Laufs, D. Mantini, M. L. Kringelbach, B. Jarraya and G. Deco, Signature of consciousness in brain-wide synchronization patterns of monkey and human fMRI signals, *Neuroimage* **226**, 117470 (2021).
- Hallquist, M. N. and F. G. Hillary, Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world, *Network Neuroscience* **3**, 1 (2018).
- Hamilton, W. L., Graph Representation Learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning* **14**, 1 (2020).

- Han, K., S. B. Chapman and D. C. Krawczyk, Cognitive training reorganizes network modularity in traumatic brain injury, *Neurorehabilitation and neural repair* **34**, 26 (2020).
- Han, L., Z. Na, L. Chunli, C. Yuchen, Z. Pengfei, W. Hao, C. Xu, Z. Peng, W. Zheng, Y. Zhenghan, et al., Baseline functional connectivity features of neural network nodes can predict improvement after sound therapy through adjusted narrow band noise in tinnitus patients, *Frontiers in neuroscience* **13**, 614 (2019).
- Han, S., H. Mao and W. J. Dally, Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding, arXiv preprint arXiv:1510.00149 (2015).
- Han, S., J. Pool, J. Tran and W. Dally, Learning both weights and connections for efficient neural network, *Advances in neural information processing systems* **28** (2015).
- Hanczar, B., F. Zehraoui, T. Issa and M. Arles, Biological interpretation of deep neural network for phenotype prediction based on gene expression, *BMC bioinformatics* **21**, 1 (2020).
- Hart, M. G., S. J. Price and J. Suckling, Connectome analysis for pre-operative brain mapping in neurosurgery, *Br. J. Neurosurg.* **30**, 506 (2016).
- Hassabis, D., D. Kumaran, C. Summerfield and M. Botvinick, Neuroscience-inspired artificial intelligence, *Neuron* **95**, 245 (2017).
- Hasson, U., S. A. Nastase and A. Goldstein, Direct fit to nature: An evolutionary perspective on biological and artificial neural networks, *Neuron* **105**, 416 (2020).
- Hayashi-Takagi, A., S. Yagishita, M. Nakamura, F. Shirai, Y. I. Wu, A. L. Loshbaugh, B. Kuhlman, K. M. Hahn and H. Kasai, Labelling and optical erasure of synaptic memory traces in the motor cortex, *Nature* **525**, 333 (2015).
- He, H., Q. Yu, Y. Du, V. Vergara, T. A. Victor, W. C. Drevets, J. B. Savitz, T. Jiang, J. Sui and V. D. Calhoun, Resting-state functional network connectivity in prefrontal regions differs between unmedicated patients with bipolar and major depressive disorders, *Journal of affective disorders* **190**, 483 (2016).
- He, X., G. E. Doucet, D. Pustina, M. R. Sperling, A. D. Sharan and J. I. Tracy, Presurgical thalamic hubness predicts surgical outcome in temporal lobe epilepsy, *Neurology* **88**, 2285 (2017).
- Hebb, D. O., *The organization of behavior: A neuropsychological theory* (Science editions).
- Hebb, D. O., *The organization of behavior: A neuropsychological theory* (Psychology press).
- Heinzle, J., M. A. Wenzel and J.-D. Haynes, Visuomotor functional network topology predicts upcoming tasks, *Journal of Neuroscience* **32**, 9960 (2012).
- Helwegen, K., I. Libedinsky and M. P. van den Heuvel, Statistical power in network neuroscience, *Trends in Cognitive Sciences* **27**, 282 (2023).
- Herzberg, M. P., K. J. McKenzie, A. S. Hodel, R. H. Hunt, B. A. Mueller, M. R. Gunnar and K. M. Thomas, Accelerated maturation in functional connectivity

- following early life stress: Circuit specific or broadly distributed?, *Dev. Cogn. Neurosci.* **48**, 100922 (2021).
- Heuvel, M. P. van den, C. J. Stam, M. Boersma and H. H. Pol, Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain, *Neuroimage* **43**, 528 (2008).
- Hlinka, J., M. Paluš, M. Vejmelka, D. Mantini and M. Corbetta, Functional connectivity in resting-state fMRI: is linear correlation sufficient?, *Neuroimage* **54**, 2218 (2011).
- Ho, T. C., M. D. Sacchet, C. G. Connolly, D. S. Margulies, O. Tymofiyeva, M. P. Paulus, A. N. Simmons, I. H. Gotlib and T. T. Yang, Inflexible functional connectivity of the dorsal anterior cingulate cortex in adolescent major depressive disorder, *Neuropsychopharmacology* **42**, 2434 (2017).
- Hojjati, S. H., A. Ebrahimzadeh and A. Babajani-Feremi, Identification of the early stage of Alzheimer's disease using structural MRI and resting-state fMRI, *Frontiers in neurology* **10**, 904 (2019).
- Hojjati, S. H., A. Ebrahimzadeh, A. Khazaei, A. Babajani-Feremi and Alzheimer's Disease Neuroimaging Initiative, Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI, *Comput. Biol. Med.* **102**, 30 (2018).
- Holla, B., J. Biswal, V. Ramesh, V. Shivakumar, R. D. Bharath, V. Benegal, G. Venkatasubramanian, P. K. Chand and P. Murthy, Effect of prefrontal tDCS on resting brain fMRI graph measures in Alcohol Use Disorders: A randomized, double-blind, sham-controlled study, *Prog. Neuropsychopharmacol. Biol. Psychiatry* **102**, 109950 (2020).
- Honnorat, N., T. D. Satterthwaite, R. E. Gur, R. C. Gur and C. Davatzikos, sGraSP: A graph-based method for the derivation of subject-specific functional parcellations of the brain, *J. Neurosci. Methods* **277**, 1 (2017).
- Hopfield, J. J., Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences* **79**, 2554 (1982).
- Hornik, K., M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural networks* **2**, 359 (1989).
- Hou, X., P. Liu, H. Gu, M. Chan, Y. Li, S.-L. Peng, G. Wig, Y. Yang, D. Park and H. Lu, Estimation of brain functional connectivity from hypercapnia BOLD MRI data: Validation in a lifespan cohort of 170 subjects, *Neuroimage* **186**, 455 (2019).
- Hou, Y., Q. Wei, R. Ou, J. Yang, W. Song, Q. Gong and H. Shang, Impaired topographic organization in cognitively unimpaired drug-naive patients with rigidity-dominant Parkinson's disease, *Parkinsonism & Related Disorders* **56**, 52 (2018).
- Hrybouski, S., I. Cribben, J. McGonigle, F. Olsen, R. Carter, P. Seres, C. R. Madan and N. V. Malykhin, Investigating the effects of healthy cognitive aging on brain functional connectivity using 4.7 T resting-state functional magnetic resonance imaging, *Brain Struct. Funct.* **226**, 1067 (2021).

- Hsu, Y.-C., Y.-C. Liu, A. Ramasamy and Z. Kira, Re-evaluating continual learning scenarios: A categorization and case for strong baselines, arXiv preprint arXiv:1810.12488 (2018).
- Hu, W., M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta and J. Leskovec, Open graph benchmark: Datasets for machine learning on graphs, arXiv preprint arXiv:2005.00687 (2020).
- Huang, S., K. Wakaizumi, B. Wu, B. Shen, B. Wu, L. Fan, M. N. Baliki, G. Zhan, A. V. Apkarian and L. Huang, Whole-brain functional network disruption in chronic pain with disc herniation, *Pain* **160**, 2829 (2019).
- Huang, Z., H. H. Davis IV, A. Wolff, G. Northoff, et al., Thalamo-sensorimotor functional connectivity correlates with world ranking of olympic, elite, and high performance athletes, *Neural Plasticity* **2017** (2017).
- Iacopini, I., G. Di Bona, E. Ubaldi, V. Loreto and V. Latora, Interacting Discovery Processes on Complex Networks, *Phys. Rev. Lett.* **125**, 248301 (2020).
- Iidaka, T., Resting state functional magnetic resonance imaging and neural network classified autism and control, *Cortex* **63**, 55 (2015).
- Ilyas, A., S. Santurkar, D. Tsipras, L. Engstrom, B. Tran and A. Madry, Adversarial examples are not bugs, they are features, *Advances in neural information processing systems* **32** (2019).
- Iordan, A. D., K. A. Cooke, K. D. Moored, B. Katz, M. Buschkuehl, S. M. Jaeggi, J. Jonides, S. J. Peltier, T. A. Polk and P. A. Reuter-Lorenz, Aging and network properties: Stability over time and links with learning during working memory training, *Front. Aging Neurosci.* **9**, 419 (2017).
- Jacunski, A. and N. Tatonetti, Connecting the dots: applications of network medicine in pharmacology and disease, *Clinical Pharmacology & Therapeutics* **94**, 659 (2013).
- Jain, G. and J. Ko, Handwritten digits recognition, Multimedia Systems, Project Report, University of Toronto 1 (2008).
- Jankovic, J., Pathophysiology and clinical assessment of parkinsonian symptoms and signs. in, *Handbook of Parkinson's disease* (CRC Press).
- Jarrett, K., K. Kavukcuoglu, M. Ranzato and Y. LeCun, "What is the best multi-stage architecture for object recognition?", *2009 IEEE 12th International Conference on Computer Vision*, (2009).
- Jeh, G. and J. Widom, "Simrank: a measure of structural-context similarity", *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, (2002).
- Jiang, F., Optimisation de la Topologie de Grands Réseaux de Neurones (PhD thesis, Colorado State University).
- Job, A., C. Jaroszynski, A. Kavounoudias, A. Jaillard and C. Delon-Martin, Functional connectivity in chronic nonbothersome tinnitus following acoustic trauma: a seed-based resting-state functional magnetic resonance imaging study, *Brain connectivity* **10**, 279 (2020).
- Joel, S. E., B. S. Caffo, P. C. Van Zijl and J. J. Pekar, On the relationship between seed-based and ICA-based measures of functional connectivity, *Magnetic resonance in medicine* **66**, 644 (2011).

- Joliot, M., G. Jobard, M. Naveau, N. Delcroix, L. Petit, L. Zago, F. Crivello, E. Mellet, B. Mazoyer and N. Tzourio-Mazoyer, AICHA: An atlas of intrinsic connectivity of homotopic areas, *J Neurosci Methods* **254**, 46 (2015).
- Joyce, K. E., P. J. Laurienti, J. H. Burdette and S. Hayasaka, A new measure of centrality for brain networks, *PloS one* **5**, e12200 (2010).
- Kang, J., F. D. Bowman, H. Mayberg and H. Liu, A depression network of functionally connected regions discovered via multi-attribute canonical correlation graphs, *Neuroimage* **141**, 431 (2016).
- Kanwisher, N., M. Khosla and K. Dobs, Using artificial neural networks to ask ‘why’ questions of minds and brains, *Trends in Neurosciences* **46**, 240 (2023).
- Karunakaran, K., M. Wolfer and B. B. Biswal, Review of Resting-State Functional Connectivity Methods and Application in Clinical Populations. in, *Brain Network Dysfunction in Neuropsychiatric Illness: Methods, Applications, and Implications* (Springer).
- Kato, S., E. Bagarinao, H. Isoda, S. Koyama, H. Watanabe, S. Maesawa, D. Mori, K. Hara, M. Katsuno, M. Hoshiyama, S. Naganawa, N. Ozaki and G. Sobue, Effects of head motion on the evaluation of age-related brain network changes using resting state functional MRI, *Magn. Reson. Med. Sci.* **20**, 338 (2021).
- Kaviani, S. and I. Sohn, Application of complex systems topologies in artificial neural networks optimization: An overview, *Expert Systems with Applications* **180**, 115073 (2021).
- Ke, Z. and B. Liu, Continual learning of natural language processing tasks: A survey, *arXiv preprint arXiv:2211.12701* (2022).
- Kemker, R., M. McClure, A. Abitino, T. Hayes and C. Kanan, “Measuring catastrophic forgetting in neural networks”, *Proceedings of the AAAI Conference on Artificial Intelligence*, (2018).
- Kempermann, G., L. Wiskott and F. H. Gage, Functional significance of adult neurogenesis, *Current opinion in neurobiology* **14**, 186 (2004).
- Kermarrec, A.-M., E. Le Merrer, B. Sericola and G. Trédan, Second order centrality: Distributed assessment of nodes criticality in complex networks, *Computer Communications* **34**, 619 (2011).
- Kersting, K., N. M. Kriege, C. Morris, P. Mutzel and M. Neumann, *Benchmark Data Sets for Graph Kernels* (2016).
- Kesler, S. R., A. Rao, D. W. Blayney, I. A. Oakley-Girvan, M. Karuturi and O. Palesh, Predicting long-term cognitive outcome following breast cancer with pre-treatment resting state fMRI and random forest machine learning, *Front. Hum. Neurosci.* **11**, 555 (2017).
- Khacef, L., N. Abderrahmane and B. Miramond, “Confronting machine-learning with neuroscience for neuromorphic architectures design”, *2018 International Joint Conference on Neural Networks (IJCNN)*, (2018).
- Kharabian Masouleh, S., S. Herzig, L. Klose, E. Roggenhofer, H. Tenckhoff, T. Kaiser, A. Thöne-Otto, M. Wiese, T. Berg, M. L. Schroeter, D. S. Margulies and A. Villringer, Functional connectivity alterations in patients with chronic hepatitis C virus infection: A multimodal MRI study, *J. Viral Hepat.* **24**, 216 (2017).

- Kim, B.-H. and J. C. Ye, Understanding graph isomorphism network for rs-fMRI functional connectivity analysis, *Frontiers in neuroscience* **14**, 630 (2020).
- Kim, G., C. Xiao, T. Konishi, Z. Ke and B. Liu, A theoretical study on solving continual learning, *Advances in Neural Information Processing Systems* **35**, 5065 (2022).
- Kim, S.-H., M.-W. Lee, M.-J. Kang, S. G. Lee, J.-G. Lee and C.-W. Mun, Comparison analysis between the medication efficacy of the milnacipran and functional connectivity of neural networks in fibromyalgia patients, *Brain Sci.* **10**, 295 (2020).
- Kim, Y.-D., E. Park, S. Yoo, T. Choi, L. Yang and D. Shin, Compression of deep convolutional neural networks for fast and low power mobile applications, *arXiv preprint arXiv:1511.06530* (2015).
- Kingma, D. P. and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- Kirkpatrick, J., R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al., Overcoming catastrophic forgetting in neural networks, *Proceedings of the national academy of sciences* **114**, 3521 (2017).
- Kirshenbaum, J. S., R. Chahal, T. C. Ho, L. S. King, A. J. Gifuni, D. Mastrovito, S. M. Coury, R. L. Weisenburger and I. H. Gotlib, Correlates and predictors of the severity of suicidal ideation in adolescence: an examination of brain connectomics and psychosocial characteristics, *J. Child Psychol. Psychiatry* **63**, 701 (2022).
- Klooster, D. C. W., S. L. Franklin, R. M. H. Besseling, J. F. A. Jansen, K. Caeyenberghs, R. Duprat, A. P. Aldenkamp, A. J. A. de Louw, P. A. J. M. Boon and C. Baeken, Focal application of accelerated iTBS results in global changes in graph measures, *Hum. Brain Mapp.* **40**, 432 (2019).
- Klyuzhin, I. S., Y. Xu, A. Ortiz, J. L. Ferres, G. Hamarneh and A. Rahmim, Testing the ability of convolutional neural networks to learn radiomic features, *Computer Methods and Programs in Biomedicine* **219**, 106750 (2022).
- Kong, Y., S. Gao, Y. Yue, Z. Hou, H. Shu, C. Xie, Z. Zhang and Y. Yuan, Spatio-temporal graph convolutional network for diagnosis and treatment response prediction of major depressive disorder from functional connectivity, *Hum. Brain Mapp.* **42**, 3922 (2021).
- Koo, T. K. and M. Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, *Journal of chiropractic medicine* **15**, 155 (2016).
- Koyama, M. S., C. Kelly, Z. Shehzad, D. Penesetti, F. X. Castellanos and M. P. Milham, Reading networks at rest, *Cerebral cortex* **20**, 2549 (2010).
- Krämer, C., J. Stumme, L. da Costa Campos, C. Rubbert, J. Caspers, S. Caspers and C. Jockwitz, Classification and prediction of cognitive performance differences in older age based on brain network patterns using a machine learning approach, *Network Neuroscience* **7**, 122 (2023).
- Kuhn, H. W., The Hungarian method for the assignment problem, *Naval research logistics quarterly* **2**, 83 (1955).

- Kumar, N. and H. Beniwal, Survey on handwritten digit recognition using machine learning, *International Journal of Computer Sciences and Engineering* **6**, 96 (2018).
- Kumari, J., R. Rajesh and K. Pooja, Facial expression recognition: A survey, *Procedia computer science* **58**, 486 (2015).
- Kwong, K. K., Record of a single fMRI experiment in May of 1991, *Neuroimage* **62**, 610 (2012).
- La Malfa, E., G. La Malfa, C. Caprioli, G. Nicosia and V. Latora, Deep Neural Networks as Complex Networks, arXiv preprint arXiv:2209.05488 (2022).
- La Malfa, E., G. La Malfa, G. Nicosia and V. Latora, "Characterizing Learning Dynamics of Deep Neural Networks via Complex Networks", *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, (2021).
- Lambert, B., F. Forbes, S. Doyle, A. Tucholka and M. Dojat, *Improving Uncertainty-based Out-of-Distribution Detection for Medical Image Segmentation* (2022).
- Latora, V. and M. Marchiori, Efficient behavior of small-world networks, *Physical review letters* **87**, 198701 (2001).
- Laurienti, P., C. Hugenschmidt and S. Hayasaka, Modularity maps reveal community structure in the resting human brain, *Nature Precedings* **1** (2009).
- Lbath, H., A. Petersen, W. Meiring and S. Achard, *Clustering-Based Inter-Regional Correlation Estimation* (2023).
- LeCun, Y., Y. Bengio and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
- LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural computation* **1**, 541 (1989).
- LeCun, Y., L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, et al., "Comparison of learning algorithms for handwritten digit recognition", *International conference on artificial neural networks*, (1995).
- Lee, J., E. Park, A. Lee, W. H. Chang, D.-S. Kim and Y.-H. Kim, Recovery-related indicators of motor network plasticity according to impairment severity after stroke, *European Journal of Neurology* **24**, 1290 (2017).
- Legault, R., C. Y. Suen and C. Nadal, Difficult Cases in Handwritten Numeral Recognition. in H. S. Baird, H. Bunke and K. Yamamoto (ed. by), *Structured Document Image Analysis* (Springer Berlin Heidelberg).
- Lei, D., W. H. Pinaya, T. van Amelsvoort, M. Marcelis, G. Donohoe, D. O. Mothersill, A. Corvin, M. Gill, S. Vieira, X. Huang, et al., Detecting schizophrenia at the level of the individual: relative diagnostic value of whole-brain images, connectome-wide functional connectivity and graph-based metrics, *Psychological medicine* **50**, 1852 (2020).
- Leijnen, S. and F. v. Veen, The neural network zoo, *Multidisciplinary Digital Publishing Institute Proceedings* **47**, 9 (2020).
- Leming, M., L. Su, S. Chattopadhyay and J. Suckling, Normative pathways in the functional connectome, *Neuroimage* **184**, 317 (2019).

- Levakov, G., J. Faskowitz, G. Avidan and O. Sporns, Mapping individual differences across brain network structure to function and behavior with connectome embedding, *Neuroimage* **242**, 118469 (2021).
- Li, L., M. Zhi, Z. Hou, Y. Zhang, Y. Yue and Y. Yuan, Abnormal brain functional connectivity leads to impaired mood and cognition in hyperthyroidism: a resting-state functional MRI study, *Oncotarget* **8**, 6283 (2017).
- Li, S. and W. Deng, Deep facial expression recognition: A survey, *IEEE transactions on affective computing* (2020).
- Li, X., Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola and J. S. Duncan, BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis, *Medical Image Analysis* **74**, 102233 (2021).
- Li, Z. and D. Hoiem, Learning without forgetting, *IEEE transactions on pattern analysis and machine intelligence* **40**, 2935 (2017).
- Liao, X., A. V. Vasilakos and Y. He, Small-world human brain networks: perspectives and challenges, *Neuroscience & Biobehavioral Reviews* **77**, 286 (2017).
- Liao, Z., T. Banaschewski, A. L. Bokde, S. Desrivières, H. Flor, A. Grigis, H. Garavan, P. Gowland, A. Heinz, B. Ittermann, et al., Similarity and stability of face network across populations and throughout adolescence and adulthood, *NeuroImage* **244**, 118587 (2021).
- Lin, P., Y. Yang, J. Jovicich, N. De Pisapia, X. Wang, C. S. Zuo and J. J. Levitt, Static and dynamic posterior cingulate cortex nodal topology of default mode network predicts attention task performance, *Brain imaging and behavior* **10**, 212 (2016).
- Lin, S.-Y., C.-C. Lee, Y.-S. Chen and L.-W. Kuo, Investigation of functional brain network reconfiguration during vocal emotional processing using graph-theoretical analysis, *Soc. Cogn. Affect. Neurosci.* **14**, 529 (2019).
- Lindner, P., P. Flodin, M. Budhiraja, I. Savic, J. Jokinen, J. Tiihonen and S. Hodgins, Associations of psychopathic traits with local and global brain network topology in young adult women, *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 1003 (2018).
- Liu, L., H. Zhang, J. Wu, Z. Yu, X. Chen, I. Rekik, Q. Wang, J. Lu and D. Shen, Overall survival time prediction for high-grade glioma patients based on large-scale brain functional networks, *Brain imaging and behavior* **13**, 1333 (2019).
- Liu, R., H. Chen, R. Qin, Y. Gu, X. Chen, J. Zou, Y. Jiang, W. Li, F. Bai, B. Zhang, et al., The altered reconfiguration pattern of brain modular architecture regulates cognitive function in cerebral small vessel disease, *Frontiers in Neurology* **10**, 324 (2019).
- Liu, S., T. Van der Lee, A. Yaman, Z. Atashgahi, D. Ferraro, G. Sokar, M. Pechenizkiy and D. C. Mocanu, "Topological insights into sparse neural networks", *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III*, (2021).
- Liu, Y., H. Wang, Y. Duan, J. Huang, Z. Ren, J. Ye, H. Dong, F. Shi, K. Li and J. Wang, Functional brain network alterations in clinically isolated syndrome and

- multiple sclerosis: a graph-based connectome study, *Radiology* **282**, 534 (2017).
- Liu, Y., R. J. Dolan, Z. Kurth-Nelson and T. E. Behrens, Human replay spontaneously reorganizes experience, *Cell* **178**, 640 (2019).
- Liu, Z., J. Zhang, K. Zhang, J. Zhang, X. Li, W. Cheng, M. Li, L. Zhao, W. Deng, W. Guo, et al., Distinguishable brain networks relate disease susceptibility to symptom expression in schizophrenia, *Human brain mapping* **39**, 3503 (2018).
- Loh, A. et al. (2020), *A disease-specific functional connectome of Parkinson's disease patients*. Version 1, Zenodo.
- Lomonaco, V., L. Pellegrini, A. Cossu, A. Carta, G. Graffieti, T. L. Hayes, M. De Lange, M. Masana, J. Pomponi, G. M. Van de Ven, et al., "Avalanche: an end-to-end library for continual learning", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021).
- Long, D., J. Wang, M. Xuan, Q. Gu, X. Xu, D. Kong and M. Zhang, Automatic Classification of Early Parkinson's Disease with Multi-Modal MR Imaging, *PLOS ONE* **7**, 1 (2012).
- Long, Q., S. Bhinge, V. D. Calhoun and T. Adali, Graph-theoretical analysis identifies transient spatial states of resting-state dynamic functional network connectivity and reveals dysconnectivity in schizophrenia, *J. Neurosci. Methods* **350**, 109039 (2021).
- Lord, A., D. Horn, M. Breakspear and M. Walter, Changes in Community Structure of Resting State Functional Connectivity in Unipolar Depression, *PLOS ONE* **7**, e41282 (2012).
- Lostar, M. and I. Reikik, *Deep Hypergraph U-Net for Brain Graph Embedding and Classification* (2020).
- Lowe, M., B. Mock and J. Sorenson, Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations, *Neuroimage* **7**, 119 (1998).
- Luppi, A. I., R. L. Carhart-Harris, L. Roseman, I. Pappas, D. K. Menon and E. A. Stamatakis, LSD alters dynamic integration and segregation in the human brain, *Neuroimage* **227**, 117653 (2021).
- Lurie, D. J., D. Kessler, D. S. Bassett, R. F. Betzel, M. Breakspear, S. Kheilholz, A. Kucyi, R. Liégeois, M. A. Lindquist, A. R. McIntosh, et al., Questions and controversies in the study of time-varying functional connectivity in resting fMRI, *Network neuroscience* **4**, 30 (2020).
- Lv, H., Z. Wang, E. Tong, L. M. Williams, G. Zaharchuk, M. Zeineh, A. N. Goldstein-Piekarski, T. M. Ball, C. Liao and M. Wintermark, Resting-state functional MRI: everything that nonexperts have always wanted to know, *American Journal of Neuroradiology* **39**, 1390 (2018).
- Maciás-García, P., R. Rashid-López, Á. J. Cruz-Gómez, E. Lozano-Soto, F. Sanmartino, R. Espinosa-Rosso and J. J. González-Rosa, Neuropsychiatric symptoms in clinically defined Parkinson's disease: an updated review of literature, *Behavioural Neurology* **2022** (2022).

- Mainsant, M., M. Mermillod, C. Godin and M. Reyboz, “A study of the Dream Net model robustness across continual learning scenarios”, *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, (2022).
- Mainsant, M., M. Solinas, M. Reyboz, C. Godin and M. Mermillod, “Dream Net: a privacy preserving continual learning model for face emotion recognition”, *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, (2021).
- Makovac, E., M. Mancini, S. Fagioli, D. R. Watson, F. Meeten, C. L. Rae, H. D. Critchley and C. Ottaviani, Network abnormalities in generalized anxiety pervade beyond the amygdala-pre-frontal cortex circuit: Insights from graph theory, *Psychiatry Res. Neuroimaging* **281**, 107 (2018).
- Mandelli, M. L. et al., Healthy brain connectivity predicts atrophy progression in non-fluent variant of primary progressive aphasia, *Brain* **139**, 2778 (2016).
- Mansfield, P. and A. A. Maudsley, Medical imaging by NMR, *The British journal of radiology* **50**, 188 (1977).
- Mao, Y., Z. Liao, X. Liu, T. Li, J. Hu, D. Le, Y. Pei, W. Sun, J. Lin, Y. Qiu, J. Zhu, Y. Chen, C. Qi, H. Su and E. Yu, Disrupted balance of long and short-range functional connectivity density in Alzheimer’s disease (AD) and mild cognitive impairment (MCI) patients: a resting-state fMRI study, *Ann. Transl. Med.* **9**, 65 (2021).
- Marblestone, A. H., G. Wayne and K. P. Kording, Toward an Integration of Deep Learning and Neuroscience, *Frontiers in Computational Neuroscience* **10**, 94 (2016).
- Marcus, G., The next decade in AI: four steps towards robust artificial intelligence, arXiv preprint arXiv:2002.06177 (2020).
- Marcus, G. and E. Davis, *Rebooting AI: Building artificial intelligence we can trust* (Vintage).
- Marques, A., N. L. Taylor, D. Roquet, S. Beze, C. Chassain, B. Pereira, C. O’callaghan, S. J. Lewis and F. Durif, Structural and functional correlates of hallucinations and illusions in Parkinson’s Disease, *Journal of Parkinson’s disease* **12**, 397 (2022).
- Masana, M., X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov and J. Van De Weijer, Class-incremental learning: survey and performance evaluation on image classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 5513 (2022).
- Masoudi-Nejad, A., F. Schreiber and Z. R. M. Kashani, Building blocks of biological networks: a review on major network motif discovery algorithms, *IET systems biology* **6**, 164 (2012).
- Maximo, J. O., D. L. Murdaugh, S. O’Kelley and R. K. Kana, Changes in intrinsic local connectivity after reading intervention in children with autism, *Brain Lang.* **175**, 11 (2017).
- Mayer, A. R., D. Ruhl, F. Merideth, J. Ling, F. M. Hanlon, J. Bustillo and J. Canive, Functional imaging of the hemodynamic sensory gating response in schizophrenia, *Human brain mapping* **34**, 2302 (2013).
- Mazrooyisebdani, M., V. A. Nair, C. Garcia-Ramos, R. Mohanty, E. Meyerand, B. Hermann, V. Prabhakaran and R. Ahmed, Graph theory analysis of func-

- tional connectivity combined with machine learning approaches demonstrates widespread network differences and predicts clinical variables in temporal lobe epilepsy, *Brain Connect.* **10**, 39 (2020).
- McClelland, J. L., B. L. McNaughton and R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review* **102**, 419 (1995).
- McClelland, J. and D. Rumelhart, *Parallel distributed processing: Explorations in the microstructure of cognition* (1986).
- McCulloch, W. S. and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *The bulletin of mathematical biophysics* **5**, 115 (1943).
- McGraw, K. O. and S. P. Wong, Forming inferences about some intraclass correlation coefficients. *Psychological methods* **1**, 30 (1996).
- McKeown, M. J., S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell and T. J. Sejnowski, Analysis of fMRI data by blind separation into independent spatial components, *Human brain mapping* **6**, 160 (1998).
- Meairs, S. and A. Alonso, Ultrasound, microbubbles and the blood–brain barrier, *Progress in biophysics and molecular biology* **93**, 354 (2007).
- Meghabghab, G., Discovering authorities and hubs in different topological web graph structures, *Information processing & management* **38**, 111 (2002).
- Mehta, D., M. F. H. Siddiqui and A. Y. Javaid, Facial emotion recognition: A survey and real-world user experiences in mixed reality, *Sensors* **18**, 416 (2018).
- Ménoret, M., N. Farrugia, B. Padeloup and V. Gripon, “Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction”, *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, (2017).
- Mensch, A., J. Mairal, B. Thirion and G. Varoquaux, Extracting representations of cognition across neuroimaging studies improves brain decoding, *PLoS computational biology* **17**, e1008795 (2021).
- Mermillod, M., A. Bugajska and P. Bonin, *The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects* (2013).
- Meunier, D., S. Achard, A. Morcom and E. Bullmore, Age-related changes in modular organization of human brain functional networks, *Neuroimage* **44**, 715 (2009).
- Mheich, A., F. Wendling and M. Hassan, Brain network similarity: methods and applications, *Network Neuroscience* **4**, 507 (2020).
- Mhiri, I., A. Ben Khalifa, M. Mahjoub and I. Rekik, Brain Graph Super-Resolution for Boosting Neurological Disorder Diagnosis using Unsupervised Multi-Topology Connectional Brain Template Learning, *Medical Image Analysis* **65**, 101768 (2020).
- Mijalkov, M., E. Kakaei, J. B. Pereira, E. Westman, G. Volpe and Alzheimer’s Disease Neuroimaging Initiative, BRAPH: A graph theory software for the analysis of brain connectivity, *PLoS One* **12**, e0178798 (2017).
- Mikolov, T., K. Chen, G. Corrado and J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).

- Milgram, N. W., C. M. MacLeod and T. L. Petit, *Neuroplasticity, learning, and memory*. (Alan R. Liss).
- Min, S. and J. Park, Narrative as a Complex Network: A Study of Victor Hugo's Les Misérables. in, *Proceedings of HCI Korea* ().
- Mirakhorli, J., H. Amindavar and M. Mirakhorli, A new method to predict anomaly in brain network based on graph deep learning, *Rev. Neurosci.* **31**, 681 (2020).
- Mocanu, D. C., E. Mocanu, P. H. Nguyen, M. Gibescu and A. Liotta, A topological insight into restricted boltzmann machines, *Machine Learning* **104**, 243 (2016).
- Mocanu, D. C., E. Mocanu, P. Stone, P. H. Nguyen, M. Gibescu and A. Liotta, Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science, *Nature communications* **9**, 2383 (2018).
- Montavon, G., A. Binder, S. Lapuschkin, W. Samek and K.-R. Müller (2019), "Layer-Wise Relevance Propagation: An Overview", *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen and K.-R. Müller (ed. by), Springer International Publishing, 193.
- Morgan, S. E., S. R. White, E. T. Bullmore and P. E. Vértes, A Network Neuroscience Approach to Typical and Atypical Brain Development, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **3**, 754 (2018).
- Moro, E., E. Bellot, S. Meoni, P. Pelissier, R. Hera, M. Dojat, V. Coizet and S. C. S. Group, Visual dysfunction of the superior colliculus in de novo parkinsonian patients, *Annals of neurology* **87**, 533 (2020).
- Morris, C. and I. Rekik, "Autism Spectrum Disorder Diagnosis Using Sparse Graph Embedding of Morphological Brain Networks", *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*, (2017).
- Morris, J. S., C. D. Frith, D. I. Perrett, D. Rowland, A. W. Young, A. J. Calder and R. J. Dolan, A differential neural response in the human amygdala to fearful and happy facial expressions, *Nature* **383**, 812 (1996).
- Mostafa, S., L. Tang and F.-X. Wu, Diagnosis of autism spectrum disorder based on eigenvalues of brain networks, *IEEE Access* **7**, 128474 (2019).
- Muñoz-Ramírez, V., M. Dojat, C. Delon-Martin and S. Achard, *rs_graph_processing* (2021).
- Muñoz-Ramírez, V., F. Forbes, J. Arbel, A. Arnaud and M. Dojat, "Quantitative MRI characterization of brain abnormalities in de novo Parkinsonian patients", *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (2019).
- Nandakumar, N., K. Manzoor, S. Agarwal, J. J. Pillai, S. K. Gujar, H. I. Sair and A. Venkataraman, "A multi-scale spatial and temporal attention network on dynamic connectivity to localize the eloquent cortex in brain tumor patients", *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 2021, Proceedings*, (2021).
- Neudorf, J., C. Ekstrand, S. Kress and R. Borowsky, Brain structural connectivity predicts brain functional complexity: Diffusion tensor imaging derived

- centrality accounts for variance in fractal properties of functional magnetic resonance imaging signal, *Neuroscience* **438**, 1 (2020).
- Neumann, L., N. Wulms, V. Witte, T. Spisak, M. Zunhammer, U. Bingel and T. Schmidt-Wilcke, Network properties and regional brain morphology of the insular cortex correlate with individual pain thresholds, *Hum. Brain Mapp.* **42**, 4896 (2021).
- Newman, M., *Networks: An introduction*. 2010: Oxford university press, *Artif. Life* **18**, 241 (2012).
- Newman, M. E., The structure and function of complex networks, *SIAM review* **45**, 167 (2003).
- Newman, M. E., Random graphs with clustering, *Physical review letters* **103**, 058701 (2009).
- Newman, M. E., S. H. Strogatz and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, *Physical review E* **64**, 026118 (2001).
- Neyshabur, B., Z. Li, S. Bhojanapalli, Y. LeCun and N. Srebro, Towards understanding the role of over-parametrization in generalization of neural networks, *arXiv preprint arXiv:1805.12076* (2018).
- Nikolentzos, G., P. Meladianos and M. Vazirgiannis, “Matching node embeddings for graph similarity”, *Thirty-First AAAI Conference on Artificial Intelligence*, (2017).
- Nishimoto, S., A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu and J. L. Gallant, Reconstructing visual experiences from brain activity evoked by natural movies, *Current biology* **21**, 1641 (2011).
- Niu, X.-X. and C. Y. Suen, A novel hybrid CNN–SVM classifier for recognizing handwritten digits, *Pattern Recognition* **45**, 1318 (2012).
- Noble, S., D. Scheinost and R. T. Constable, A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis, *Neuroimage* **203**, 116157 (2019).
- Novak, R., Y. Bahri, D. A. Abolafia, J. Pennington and J. Sohl-Dickstein, Sensitivity and generalization in neural networks: an empirical study, *arXiv preprint arXiv:1802.08760* (2018).
- Nwaigwe, D., L. Carboni, M. Mermillod, S. Achard and M. Dojat, Graph-based methods coupled with specific distributional distances for adversarial attack detection (2023).
- O’Reilly, R. C., R. Bhattacharyya, M. D. Howard and N. Ketz, Complementary learning systems, *Cognitive science* **38**, 1229 (2014).
- Ogawa, A., Time-varying measures of cerebral network centrality correlate with visual saliency during movie watching, *Brain Behav.* **11**, e2334 (2021).
- Ogawa, S., T.-M. Lee, A. R. Kay and D. W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences* **87**, 9868 (1990).
- Olden, J. D. and D. A. Jackson, Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks, *Ecological modelling* **154**, 135 (2002).

- Olden, J. D., M. K. Joy and R. G. Death, An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, *Ecological modelling* **178**, 389 (2004).
- Oldham, S. and A. Fornito, The development of brain network hubs, *Developmental cognitive neuroscience* **36**, 100607 (2019).
- Openneer, T. J. C., J.-B. C. Marsman, D. van der Meer, N. J. Forde, S. E. A. Akkermans, J. Naaijen, J. K. Buitelaar, A. Dietrich and P. J. Hoekstra, A graph theory study of resting-state functional connectivity in children with Tourette syndrome, *Cortex* **126**, 63 (2020).
- Osadchiy, V., E. A. Mayer, R. Bhatt, J. S. Labus, L. Gao, L. A. Kilpatrick, C. Liu, K. Tillisch, B. Naliboff, L. Chang and A. Gupta, History of early life adversity is associated with increased food addiction and sex-specific alterations in reward network connectivity in obesity, *Obes. Sci. Pract.* **5**, 416 (2019).
- Oujamaa, L., Evolution topologique des hubs dans l'état de conscience altérée post-traumatique: un marqueur de récupération fonctionnelle (PhD thesis, Université Grenoble Alpes).
- Paldino, M. J., W. Zhang, Z. D. Chu and F. Golriz, Metrics of brain network architecture capture the impact of disease in children with epilepsy, *NeuroImage Clin.* **13**, 201 (2017).
- Parisi, G. I., X. Ji and S. Wermter, On the role of neurogenesis in overcoming catastrophic forgetting, arXiv preprint arXiv:1811.02113 (2018).
- Parisi, G. I., R. Kemker, J. L. Part, C. Kanan and S. Wermter, Continual lifelong learning with neural networks: A review, *Neural Networks* **113**, 54 (2019).
- Park, H.-J. and K. Friston, Structural and functional brain networks: from connections to cognition, *Science* **342**, 1238411 (2013).
- Pascual-Leone, A., A. Amedi, F. Fregni and L. B. Merabet, The plastic human brain cortex, *Annu. Rev. Neurosci.* **28**, 377 (2005).
- Pashine, S., R. Dixit and R. Kushwah, Handwritten Digit Recognition using Machine and Deep Learning Algorithms, *CoRR* **abs/2106.12614** (2021).
- Patra, S. and A. Mohapatra, Review of tools and algorithms for network motif discovery in biological networks, *IET systems biology* **14**, 171 (2020).
- Patriat, R., E. K. Molloy, T. B. Meier, G. R. Kirk, V. A. Nair, M. E. Meyerand, V. Prabhakaran and R. M. Birn, The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated, *Neuroimage* **78**, 463 (2013).
- Pauling, L. and C. D. Coryell, The magnetic properties and structure of the hemochromogens and related substances, *Proceedings of the National Academy of Sciences* **22**, 159 (1936).
- Pessoa, L. and R. Adolphs, Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance, *Nature reviews neuroscience* **11**, 773 (2010).
- Petersen, A., J. Zhao, O. Carmichael and H.-G. Müller, Quantifying individual brain connectivity with functional principal component analysis for networks, *Brain Connect.* **6**, 540 (2016).

- Petersen, S. and O. Sporns, Brain Networks and Cognitive Architectures, *Neuron* **88**, 207 (2015).
- Petro, N. M., L. R. Ott, S. H. Penhale, M. P. Rempe, C. M. Embury, G. Picci, Y.-P. Wang, J. M. Stephen, V. D. Calhoun and T. W. Wilson, Eyes-closed versus eyes-open differences in spontaneous neural dynamics during development, *NeuroImage* **258**, 119337 (2022).
- Pezoulas, V. C., M. Zervakis, S. Michelogiannis and M. A. Klados, Resting-state functional connectivity and network analysis of cerebellum with respect to IQ and gender, *Front. Hum. Neurosci.* **11**, 189 (2017).
- Pitas, K., A. Loukas, M. Davies and P. Vandergheynst, Some limitations of norm based generalization bounds in deep neural networks, arXiv preprint arXiv:1905.09677 (2019).
- Poldrack, R. A., Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding, *Neuron* **72**, 692 (2011).
- Power, J. D., B. L. Schlaggar, C. N. Lessov-Schlaggar and S. E. Petersen, Evidence for hubs in human functional brain networks, *Neuron* **79**, 798 (2013).
- Prabhu, A., P. H. Torr and P. K. Dokania, “Gdumb: A simple approach that questions our progress in continual learning”, *European conference on computer vision*, (2020).
- Prajapati, R. and I. A. Emerson, Global and regional connectivity analysis of resting-state function MRI brain images using graph theory in Parkinson’s disease, *Int. J. Neurosci.* **131**, 105 (2021).
- Puig, J. et al., Resting-state functional connectivity magnetic resonance imaging and outcome after acute stroke, *Stroke* **49**, 2353 (2018).
- Raimondo, S. and M. De Domenico, Measuring topological descriptors of complex networks under uncertainty, *Physical Review E* **103**, 022311 (2021).
- Rajpurkar, P., J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, et al., Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, *PLoS medicine* **15**, e1002686 (2018).
- Ramyachitra, D. and P. Manikandan, Imbalanced dataset classification and solutions: a review, *International Journal of Computing and Business Research (IJCBR)* **5**, 1 (2014).
- Rangaprakash, D., M. N. Dretsch, J. S. Katz, T. S. Denney Jr and G. Deshpande, Dynamics of segregation and integration in directional brain networks: Illustration in soldiers with PTSD and neurotrauma, *Front. Neurosci.* **13**, 803 (2019).
- Ray, S., B. B. Biswal, A. Aya, S. Gohel, A. Srinagesh, C. Hanson and S. J. Hanson, Modeling causal relationships among brain areas in the mesocorticolimbic system during resting-state in cocaine users utilizing a graph theoretic approach, *J. Alcohol. Drug Depend.* **5** (2017).
- Rätz, T., Euler’s Königsberg: the explanatory power of mathematics, *European Journal for Philosophy of Science* **8**, 331 (2018).
- Regner, M. F., N. Saenz, K. Maharajh, D. J. Yamamoto, B. Mohl, K. Wylie, J. Tregellas and J. Tanabe, Top-down network effective connectivity in abstinent substance dependent individuals, *PLoS One* **11**, e0164818 (2016).

- Renard, F., C. Heinrich, M. Bouthillon, M. Schenck, F. Schneider, S. Kremer and S. Achard, A covariate-constraint method to map brain feature space into lower dimensional manifolds, *Network Neuroscience* **5**, 252 (2021).
- Ribeiro de Paula, D. et al., A method for independent component graph analysis of resting-state fMRI, *Brain Behav.* **7**, e00626 (2017).
- Richards, T. L., V. W. Berninger, K. Yagle, R. D. Abbott and D. Peterson, Brain's functional network clustering coefficient changes in response to instruction (RTI) in students with and without reading disabilities: Multi-leveled reading brain's RTI, *Cogent psychology* **5**, 1424680 (2018).
- Richiardi, J., S. Achard, H. Bunke and D. Van De Ville, Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience, *IEEE Signal processing magazine* **30**, 58 (2013).
- Rish, I., B. Thyreau, B. Thirion, M. Plaze, M.-l. Paillere-martinot, C. Martelli, J.-l. Martinot, J.-b. Poline and G. Cecchi, "Discriminative Network Models of Schizophrenia", *Advances in Neural Information Processing Systems*, (2009).
- Robbins, H. and S. Monro, A stochastic approximation method, *The annals of mathematical statistics* **400** (1951).
- Roberts, G., A. Lord, A. Frankland, A. Wright, P. Lau, F. Levy, R. K. Lenroot, P. B. Mitchell and M. Breakspear, Functional dysconnection of the inferior frontal gyrus in young people with bipolar disorder or at genetic high risk, *Biol. Psychiatry* **81**, 718 (2017).
- Rodriguez, C. I., V. M. Vergara, V. D. Calhoun, D. D. Savage, D. A. Hamilton, C. D. Tesche and J. M. Stephen, Disruptions in global network segregation and integration in adolescents and young adults with fetal alcohol spectrum disorder, *Alcohol. Clin. Exp. Res.* **45**, 1775 (2021).
- Rolfe, D. and G. C. Brown, Cellular energy utilization and molecular origin of standard metabolic rate in mammals, *Physiological reviews* **77**, 731 (1997).
- Rolls, E. T., C. C. Huang, C. P. Lin, J. Feng and M. Joliot, Automated anatomical labelling atlas 3, *Neuroimage* **206**, 116189 (2020).
- Rosenblatt, F., The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review* **65**, 386 (1958).
- Rosenthal, G., F. Váša, A. Griffa, P. Hagmann, E. Amico, J. Goñi, G. Avidan and O. Sporns, Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes, *Nature communications* **9**, 1 (2018).
- Rossi, R. A. and N. K. Ahmed, Role discovery in networks, *IEEE Transactions on Knowledge and Data Engineering* **27**, 1112 (2014).
- Rubbert, C., C. Mathys, C. Jockwitz, C. J. Hartmann, S. B. Eickhoff, F. Hoffstaedter, S. Caspers, C. R. Eickhoff, B. Sigl, N. A. Teichert, et al., Machine-learning identifies Parkinson's disease patients based on resting-state between-network functional connectivity, *The british journal of radiology* **92**, 20180886 (2019).
- Rubinov, M., Circular and unified analysis in network neuroscience (2022).
- Rubinov, M. and O. Sporns, Complex network measures of brain connectivity: uses and interpretations, *Neuroimage* **52**, 1059 (2010).

- Rubinov, M., R. J. Ypma, C. Watson and E. T. Bullmore, Wiring cost and topological participation of the mouse brain connectome, *Proceedings of the National Academy of Sciences* **112**, 10032 (2015).
- Rumelhart, D. E., G. E. Hinton and R. J. Williams, Learning representations by back-propagating errors, *nature* **323**, 533 (1986).
- Rusu, A. A., N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu and R. Hadsell, Progressive neural networks, arXiv preprint arXiv:1606.04671 (2016).
- Ryan, J. P., H. T. Karim, H. J. Aizenstein, N. L. Helbling and F. G. S. Toledo, Insulin sensitivity predicts brain network connectivity following a meal, *Neuroimage* **171**, 268 (2018).
- Sadeh, N., J. M. Spielberg, M. W. Logue, J. P. Hayes, E. J. Wolf, R. E. McGlinchey, W. P. Milberg, S. A. Schichman, A. Stone and M. W. Miller, Linking genes, circuits, and behavior: network connectivity as a novel endophenotype of externalizing, *Psychol. Med.* **49**, 1905 (2019).
- Saghayi, M., J. Greenberg, C. O'Grady, F. Varno, M. A. Hashmi, B. Bracken, S. Matwin, S. W. Lazar and J. A. Hashmi, Brain network topology predicts participant adherence to mental training programs, *Netw. Neurosci.* **4**, 528 (2020).
- Saghayi, M., J. Greenberg, C. O'Grady, F. Varno, M. A. Hashmi, B. Bracken, S. Matwin, S. W. Lazar and J. A. Hashmi, Brain network topology predicts participant adherence to mental training programs, *Network Neuroscience* **4**, 528 (2020).
- Sako, W., T. Abe, T. Furukawa, R. Oki, S. Haji, N. Murakami, Y. Izumi, M. Harada and R. Kaji, Differences in the intra-cerebellar connections and graph theoretical measures between Parkinson's disease and multiple system atrophy, *J. Neurol. Sci.* **400**, 129 (2019).
- Sala-Llonch, R., C. Junqué, E. M. Arenaza-Urquijo, D. Vidal-Piñeiro, C. Valls-Pedret, E. M. Palacios, S. Domènech, A. Salvà, N. Bargalló and D. Bartrés-Faz, Changes in whole-brain functional networks and memory performance in aging, *Neurobiology of aging* **35**, 2193 (2014).
- Salvador, R., J. Suckling, C. Schwarzbauer and E. Bullmore, Undirected graphs of frequency-dependent functional connectivity in whole brain networks, *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 937 (2005).
- Samu, D., A. K. Seth and T. Nowotny, Influence of wiring cost on the large-scale architecture of human cortical connectivity, *PLoS computational biology* **10**, e1003557 (2014).
- Sang, L., J. Zhang, L. Wang, J. Zhang, Y. Zhang, P. Li, J. Wang and M. Qiu, Alteration of brain functional networks in early-stage Parkinson's disease: a resting-state fMRI study, *PloS one* **10**, e0141815 (2015).
- Sato, J. R. et al., Association between abnormal brain functional connectivity in children and psychopathology: A study based on graph theory and machine learning, *World J. Biol. Psychiatry* **19**, 119 (2018).

- Scabini, L. F. and O. M. Bruno, Structure and Performance of Fully Connected Neural Networks: Emerging Complex Network Properties, arXiv preprint arXiv:2107.14062 (2021).
- Schaefer, A., R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff and B. T. Yeo, Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI, *Cerebral cortex* **28**, 3095 (2018).
- Schieber, T. A., L. Carpi, A. Díaz-Guilera, P. M. Pardalos, C. Masoller and M. G. Ravetti, Quantification of network structural dissimilarities, *Nature Communications* **8**, 13928 (2017).
- Schipper, L. J. de, A. Hafkemeijer, J. van der Grond, J. Marinus, J. M. L. Henselmans and J. J. van Hilten, Altered whole-brain and network-based functional connectivity in Parkinson's disease, *Front. Neurol.* **9**, 419 (2018).
- Schirmer, M. D., A. Venkataraman, I. Rekik, M. Kim, S. H. Mostofsky, M. B. Nebel, K. Rosch, K. Seymour, D. Crocetti, H. Irzan, et al., Neuropsychiatric disease classification using functional connectomics-results of the connectomics in neuroimaging transfer learning challenge, *Medical image analysis* **70**, 101972 (2021).
- Schuld, M., K. Brádler, R. Israel, D. Su and B. Gupt, Measuring the similarity of graphs with a Gaussian boson sampler, *Physical Review A* **101**, 032314 (2020).
- Sen, B., K. R. Cullen and K. K. Parhi, Classification of adolescent major depressive disorder via static and dynamic connectivity, *IEEE J. Biomed. Health Inform.* **25**, 2604 (2021).
- Serre, T., Deep learning: the good, the bad, and the ugly, *Annual review of vision science* **5**, 399 (2019).
- Shaw, S. B., M. C. McKinnon, J. Heisz and S. Becker, Dynamic task-linked switching between brain networks - A tri-network perspective, *Brain Cogn.* **151**, 105725 (2021).
- Sheffield, J. M., G. Repovs, M. P. Harms, C. S. Carter, J. M. Gold, A. W. MacDonald 3rd, J. D. Ragland, S. M. Silverstein, D. Godwin and D. M. Barch, Evidence for accelerated decline of functional brain network efficiency in schizophrenia, *Schizophr. Bull.* **42**, 753 (2016).
- Shen, X., F. Tokoglu, X. Papademetris and R. T. Constable, Groupwise whole-brain parcellation from resting-state fMRI data for network node identification, *Neuroimage* **82**, 403 (2013).
- Shen, Y., Q. Lu, T. Zhang, H. Yan, N. Mansouri, K. Osipowicz, O. Tanglay, I. Young, S. Doyen, X. Lu, X. Zhang, M. E. Sughrue and T. Wang, Use of machine learning to identify functional connectivity changes in a clinical cohort of patients at risk for dementia, *Frontiers in Aging Neuroscience* **14**, 962319 (2022).
- Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of Go with deep neural networks and tree search, *nature* **529**, 484 (2016).

- Simard, D., L. Nadeau and H. Kröger, Fastest learning in small-world neural networks, *Physics Letters A* **336**, 8 (2005).
- Simpson, S. L., M. Bahrami and P. J. Laurienti, A mixed-modeling framework for analyzing multitask whole-brain network data, *Netw. Neurosci.* **3**, 307 (2019).
- Sitoleux, P., L. Carboni, H. Lbath and S. Achard, Multiscale and multi-density comparison of functional brain networks through label-informed persistence diagrams (2023).
- Smith, S. M., K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey and M. W. Woolrich, Network modelling methods for fMRI, *Neuroimage* **54**, 875 (2011).
- Smith, S. M. et al., Advances in functional and structural MR image analysis and implementation as FSL, *NeuroImage* **23**, S208 (2004).
- Song, H., M. Kim, D. Park, Y. Shin and J.-G. Lee, Learning from noisy labels with deep neural networks: A survey, *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- Song, Y., T. M. Epalle and H. Lu, Characterizing and predicting autism spectrum disorder by performing resting-state functional network community pattern analysis, *Front. Hum. Neurosci.* **13**, 203 (2019).
- Sourty, M., L. Thoraval, D. Roquet, J.-P. Armspach, J. Foucher and F. Blanc, Identifying dynamic functional connectivity changes in dementia with Lewy bodies based on product hidden Markov models, *Front. Comput. Neurosci.* **10**, 60 (2016).
- Sporns, O., Network attributes for segregation and integration in the human brain, *Current opinion in neurobiology* **23**, 162 (2013).
- Sporns, O., *Networks of the Brain* (MIT press).
- Sporns, O., Graph theory methods: applications in brain networks, *Dialogues in clinical neuroscience* (2022).
- Sporns, O., C. J. Honey and R. Kötter, Identification and classification of hubs in brain networks, *PloS one* **2**, e1049 (2007).
- Stauffer, D., A. Aharony, L. da Fontoura Costa and J. Adler, Efficient Hopfield pattern recognition on a scale-free neural network, *The European Physical Journal B-Condensed Matter and Complex Systems* **32**, 395 (2003).
- Storti, S. F., I. Boscolo Galazzo, S. Montemezzi, G. Menegaz and F. B. Pizzini, Dual-echo ASL contributes to decrypting the link between functional connectivity and cerebral blood flow, *Hum. Brain Mapp.* **38**, 5831 (2017).
- Subramanian, S., K. Rajamanickam, J. S. Prakash, M. Ramachandran and for Alzheimer's Disease Neuroimaging Initiative (ADNI), Study on structural atrophy changes and functional connectivity measures in Alzheimer's disease, *J. Med. Imaging (Bellingham)* **7**, 016002 (2020).
- Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, et al., UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, *PLoS medicine* **12**, e1001779 (2015).
- Sun, S., Z. Cao, H. Zhu and J. Zhao, A survey of optimization methods from a machine learning perspective, *IEEE transactions on cybernetics* **50**, 3668 (2019).

- Sun, Y., Z. Dai, J. Li, S. L. Collinson and K. Sim, Modular-level alterations of structure–function coupling in schizophrenia connectome, *Human brain mapping* **38**, 2008 (2017).
- Sun, Y., J. Li, J. Suckling and L. Feng, Asymmetry of hemispheric network topology reveals dissociable processes between functional and structural brain connectome in community-living elders, *Frontiers in aging neuroscience* **9**, 361 (2017).
- Suo, X., D. Lei, N. Li, L. Cheng, F. Chen, M. Wang, G. J. Kemp, R. Peng and Q. Gong, Functional brain connectome and its relation to Hoehn and Yahr stage in Parkinson disease, *Radiology* **285**, 904 (2017).
- Supekar, K., V. Menon, D. Rubin, M. Musen and M. D. Greicius, Network analysis of intrinsic functional brain connectivity in Alzheimer’s disease, *PLoS computational biology* **4**, e1000100 (2008).
- Surampudi, S. G., J. Misra, G. Deco, R. S. Bapi, A. Sharma and D. Roy, Resting state dynamics meets anatomical structure: Temporal multiple kernel learning (tMKL) model, *Neuroimage* **184**, 609 (2019).
- Sussillo, D. and L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks, *Neuron* **63**, 544 (2009).
- Takagi, Y. and S. Nishimoto, “High-resolution image reconstruction with latent diffusion models from human brain activity”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2023).
- Tanaka, H., D. Kunin, D. L. Yamins and S. Ganguli, Pruning neural networks without any data by iteratively conserving synaptic flow, *Advances in neural information processing systems* **33**, 6377 (2020).
- Tang, F., D. Zhu, W. Ma, Q. Yao, Q. Li and J. Shi, Differences changes in cerebellar functional connectivity between mild cognitive impairment and Alzheimer’s disease: a seed-based approach, *Frontiers in Neurology* **12**, 645171 (2021).
- Taya, F., S. I. Dimitriadis, A. Dragomir, J. Lim, Y. Sun, K. F. Wong, N. V. Thakor and A. Bezerianos, Fronto-parietal subnetworks flexibility compensates for cognitive decline due to mental fatigue, *Human brain mapping* **39**, 3528 (2018).
- Taya, F., J. de Souza, N. V. Thakor and A. Bezerianos, Comparison method for community detection on brain networks from neuroimaging data, *Applied Network Science* **1**, 1 (2016).
- Termenon, M., A. Jaillard, C. Delon-Martin and S. Achard, Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project, *Neuroimage* **142**, 172 (2016).
- Tessitore, A., M. Cirillo and R. De Micco, Functional connectivity signatures of Parkinson’s disease, *Journal of Parkinson’s disease* **9**, 637 (2019).
- Thual, A., Q. H. TRAN, T. Zemsanova, N. Courty, R. Flamary, S. Dehaene and B. Thirion, “Aligning individual brains with fused unbalanced Gromov Wasserstein”, *Advances in Neural Information Processing Systems*, (2022).
- Tomasi, D. and N. D. Volkow, Functional connectivity hubs in the human brain, *Neuroimage* **57**, 908 (2011).

- Tomasi, D. G., E. Shokri-Kojori and N. D. Volkow, Brain network dynamics adhere to a power law, *Front. Neurosci.* **11**, 72 (2017).
- Trofimova, A., J. L. Smith, V. Ahluwalia, J. Hurtado, R. K. Gore and J. W. Allen, Alterations in resting-state functional brain connectivity and correlations with vestibular/Ocular-Motor Screening measures in postconcussion vestibular dysfunction, *J. Neuroimaging* **31**, 277 (2021).
- Tsuchida, A., A. Laurent, F. Crivello, L. Petit, M. Joliot, A. Pepe, N. Beguedo, M.-F. Gueye, V. Verrecchia, V. Nozais, L. Zago, E. Mellet, S. Debette, C. Tzourio and B. Mazoyer, The MRi-Share database: brain imaging in a cross-sectional cohort of 1870 university students, *Neuroscience & Biobehavioral Reviews* **77**, 286 (2017).
- Tumati, S., J.-B. C. Marsman, P. P. De Deyn, S. Martens, A. Aleman and Alzheimer's Disease Neuroimaging Initiative, Functional network topology associated with apathy in Alzheimer's disease, *J. Affect. Disord.* **266**, 473 (2020).
- Tuovinen, N., K. Seppi, F. de Pasquale, C. Müller, M. Nocker, M. Schocke, E. R. Gizewski, C. Kremser, G. K. Wenning, W. Poewe, et al., The reorganization of functional architecture in the early-stages of Parkinson's disease, *Parkinsonism & related disorders* **50**, 61 (2018).
- Turner, D., P. J. S. Cardoso and J. M. F. Rodrigues, Modular Dynamic Neural Network: A Continual Learning Architecture, *Applied Sciences* **11** (2021).
- Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer and M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, *Neuroimage* **15**, 273 (2002).
- Underwood, R., E. Tolmeijer, J. Wibroe, E. Peters and L. Mason, Networks underpinning emotion: A systematic review and synthesis of functional and effective connectivity, *NeuroImage* **243**, 118486 (2021).
- Vaghi, M. M., P. E. Vértes, M. G. Kitzbichler, A. M. Apergis-Schoute, F. E. van der Flier, N. A. Fineberg, A. Sule, R. Zaman, V. Voon, P. Kundu, E. T. Bullmore and T. W. Robbins, Specific frontostriatal circuits for impaired cognitive flexibility and goal-directed planning in obsessive-compulsive disorder: Evidence from resting-state functional connectivity, *Biol. Psychiatry* **81**, 708 (2017).
- Valliani, A. A.-A., D. Ranti and E. K. Oermann, Deep learning and neurology: a systematic review, *Neurology and therapy* **8**, 351 (2019).
- Van Den Heuvel, M. P. and H. E. H. Pol, Exploring the brain network: a review on resting-state fMRI functional connectivity, *European neuropsychopharmacology* **20**, 519 (2010).
- Van den Heuvel, M. P. and O. Sporns, Network hubs in the human brain, *Trends in cognitive sciences* **17**, 683 (2013).
- van den Heuvel, M. P. and H. E. Hulshoff Pol, Exploring the brain network: A review on resting-state fMRI functional connectivity, *European Neuropsychopharmacology* **20**, 519 (2010).
- Van Steen, M., Graph theory and complex networks, *An introduction* **144**, 1 (2010).
- Vancea, R., K. Simonyan, M. Petracca, M. Brys, A. Di Rocco, M. F. Ghilardi and M. Inglese, Cognitive performance in mid-stage Parkinson's disease: functional

- connectivity under chronic antiparkinson treatment, *Brain Imaging Behav.* **13**, 200 (2019).
- Varley, T. F. and O. Sporns, Network analysis of time series: Novel approaches to network neuroscience, *Frontiers in Neuroscience* **15**, 787068 (2022).
- Varoquaux, G., A. Gramfort, F. Pedregosa, V. Michel and B. Thirion, “Multi-subject dictionary learning to segment an atlas of brain spontaneous activity”, *Information Processing in Medical Imaging: 22nd International Conference, IPMI 2011, Kloster Irsee, Germany, July 3-8, 2011. Proceedings 22*, (2011).
- Varoquaux, G., A. Gramfort, J. B. Poline and B. Thirion, Markov models for fMRI correlation structure: is brain functional connectivity small world, or decomposable into networks?, *Journal of physiology-Paris* **106**, 212 (2012).
- Varoquaux, G., A. Gramfort, J.-B. Poline and B. Thirion, Brain covariance selection: better individual functional connectivity models using population prior, *Advances in neural information processing systems* **23** (2010).
- Varoquaux, G., S. Sadaghiani, P. Pinel, A. Kleinschmidt, J.-B. Poline and B. Thirion, A group model for stable multi-subject ICA on fMRI datasets, *Neuroimage* **51**, 288 (2010).
- Váša, F. and B. Mišić, Null models in network neuroscience, *Nature Reviews Neuroscience* **23**, 493 (2022).
- Váša, F., J. Seidlitz, R. Romero-Garcia, K. J. Whitaker, G. Rosenthal, P. E. Vértes, M. Shinn, A. Alexander-Bloch, P. Fonagy, R. J. Dolan, P. B. Jones, I. M. Goodyer, NSPN consortium, O. Sporns and E. T. Bullmore, Adolescent tuning of association cortex in human structural brain networks, *Cereb. Cortex* **28**, 281 (2018).
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, “Attention is All you Need”, *Advances in Neural Information Processing Systems*, (2017).
- Vatansever, D., T. Karapanagiotidis, D. S. Margulies, E. Jefferies and J. Smallwood, Distinct patterns of thought mediate the link between brain functional connectomes and well-being, *Netw. Neurosci.* **4**, 637 (2020).
- Vayer, T., L. Chapel, R. Flamary, R. Tavenard and N. Courty, Fused Gromov-Wasserstein Distance for Structured Objects, *Algorithms* **13** (2020).
- Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio, Graph attention networks, arXiv preprint arXiv:1710.10903 (2017).
- Ven, G. M. van de and A. S. Tolias, “Three continual learning scenarios”, *NeurIPS Continual Learning Workshop*, (2018).
- Vergara, V. M., Q. Yu and V. D. Calhoun, A method to assess randomness of functional connectivity matrices, *J. Neurosci. Methods* **303**, 146 (2018).
- Vergun, S., A. S. Deshpande, T. B. Meier, J. Song, D. L. Tudorascu, V. A. Nair and V. Prabhakaran, Characterizing functional connectivity differences in aging adults using machine learning on resting state fMRI data, *Frontiers in computational neuroscience* **7**, 38 (2013).
- Vértes, P. E., A. F. Alexander-Bloch, N. Gogtay, J. N. Giedd, J. L. Rapoport and E. T. Bullmore, Simple models of human brain functional networks, *Proceedings of the National Academy of Sciences* **109**, 5868 (2012).

- Vico Fallani, F. de, F. Baluch, L. Astolfi, D. Subramanian, G. Zouridakis and F. Babiloni, Structural organization of functional networks from EEG signals during motor learning tasks, *International Journal of Bifurcation and Chaos* **20**, 905 (2010).
- Vico Fallani, F. de, J. Richiardi, M. Chavez and S. Achard, Graph analysis of functional brain networks: practical issues in translational neuroscience, *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130521 (2014).
- Vidal, M., M. E. Cusick and A.-L. Barabási, Interactome networks and human disease, *Cell* **144**, 986 (2011).
- Wagner, G., F. de la Cruz, S. Köhler, F. Pereira, S. Richard-Devantoy, G. Turecki, K.-J. Bär and F. Jollant, Connectomics-based functional network alterations in both depressed patients with suicidal behavior and healthy relatives of suicide victims, *Sci. Rep.* **9**, 14330 (2019).
- Wang, H.-T., S. L. Meisler, H. Sharmarke, N. Clarke, N. Gensollen, C. J. Markiewicz, F. Paugam, B. Thirion and P. Bellec, A reproducible benchmark of denoising strategies in resting-state fMRI connectivity using fMRIPrep and Nilearn, *bioRxiv* 2023 (2023).
- Wang, H., R. Zhu, Z. Dai, S. Tian, J. Shao, X. Wang, Y. Sun, Z. Chen, X. Li, Z. Yao, et al., Aberrant functional connectivity and graph properties in bipolar II disorder with suicide attempts, *Journal of affective disorders* **275**, 202 (2020).
- Wang, J., L. Wang, Y. Zang, H. Yang, H. Tang, Q. Gong, Z. Chen, C. Zhu and Y. He, Parcellation-dependent small-world brain functional networks: A resting-state fMRI study, *Human brain mapping* **30**, 1511 (2009).
- Wang, J., X. Zuo and Y. He, Graph-based network analysis of resting-state functional MRI, *Frontiers in systems neuroscience* **4**, 16 (2010).
- Wang, J., L. Xiao, T. W. Wilson, J. M. Stephen, V. D. Calhoun and Y.-P. Wang, Examining brain maturation during adolescence using graph laplacian learning based fourier transform, *Journal of neuroscience methods* **338**, 108649 (2020).
- Wang, L., X. Zhang, H. Su and J. Zhu, A comprehensive survey of continual learning: Theory, method and application, *arXiv preprint arXiv:2302.00487* (2023).
- Wang, L., W. Yuan, L. Zeng, J. Xu, Y. Mo, X. Zhao and L. Peng, Dementia analysis from functional connectivity network with graph neural networks, *Information Processing & Management* **59**, 102901 (2022).
- Wang, X., J. Huang, J. Zhu, M. Yang and F. Yang, “Facial expression recognition with deep learning”, *Proceedings of the 10th International Conference on Internet Multimedia Computing and Service*, (2018).
- Wang, Y. F., L. J. Zheng, Y. Liu, Y. B. Ye, S. Luo, G. M. Lu, D. Gong and L. J. Zhang, The gut microbiota-inflammation-brain axis in end-stage renal disease: perspectives from default mode network, *Theranostics* **9**, 8171 (2019).
- Wang, Z., K. Zeljic, Q. Jiang, Y. Gu, W. Wang and Z. Wang, Dynamic network communication in the human functional connectome predicts perceptual variability in visual illusion, *Cerebral Cortex* **28**, 48 (2018).

- Watts, D. J. and S. H. Strogatz, Collective dynamics of ‘small-world’ networks, *nature* **393**, 440 (1998).
- Weisfeiler, B. and A. Leman, The reduction of a graph to canonical form and the algebra which appears therein, *nti, Series* **2**, 12 (1968).
- Weiss, K., T. M. Khoshgoftaar and D. Wang, A survey of transfer learning, *Journal of Big data* **3**, 1 (2016).
- Welton, T., C. S. Constantinescu, D. P. Auer and R. A. Dineen, Graph theoretic analysis of brain connectomics in multiple sclerosis: Reliability and relationship with cognition, *Brain Connect.* **10**, 95 (2020).
- Weng, C.-Y., W.-T. Chu and J.-L. Wu, “Movie analysis based on roles’ social network”, *2007 IEEE International Conference on Multimedia and Expo*, (2007).
- West, D. B. et al., *Introduction to graph theory* vol. 2 (Prentice hall Upper Saddle River).
- Wills, P. and F. G. Meyer, Metrics for graph comparison: a practitioner’s guide, *PloS one* **15**, e0228728 (2020).
- Wirsih, J., B. Ridley, P. Besson, V. Jirsa, C. Bénar, J.-P. Ranjeva and M. Guye, Complementary contributions of concurrent EEG and fMRI connectivity for predicting structural connectivity, *Neuroimage* **161**, 251 (2017).
- Wolinski, P., *Structural Learning of Neural Networks* (PhD thesis, Université Paris-Saclay).
- Woolrich, M. W., B. D. Ripley, M. Brady and S. M. Smith, Temporal autocorrelation in univariate linear modeling of fMRI data, *Neuroimage* **14**, 1370 (2001).
- Xiang, Y., J. Wang, G. Tan, F.-X. Wu and J. Liu, Schizophrenia identification using multi-view graph measures of functional brain networks, *Front. Bioeng. Biotechnol.* **7**, 479 (2019).
- Xiao, L., A. Zhang, B. Cai, J. M. Stephen, T. W. Wilson, V. D. Calhoun and Y.-P. Wang, Correlation guided graph learning to estimate functional connectivity patterns from fMRI data, *IEEE Trans. Biomed. Eng.* **68**, 1154 (2021).
- Xie, X., C. Cai, P. F. Damasceno, S. S. Nagarajan and A. Raj, Emergence of canonical functional networks from the structural connectome, *Neuroimage* **237**, 118190 (2021).
- Xu, X., W. Li, J. Mei, M. Tao, X. Wang, Q. Zhao, X. Liang, W. Wu, D. Ding and P. Wang, Feature selection and combination of information in the functional brain connectome for discrimination of mild cognitive impairment and analyses of altered brain patterns, *Frontiers in aging neuroscience* **12**, 28 (2020).
- Xu, X., W. Li, M. Tao, Z. Xie, X. Gao, L. Yue and P. Wang, Effective and accurate diagnosis of subjective cognitive decline based on functional connection and graph theory view, *Frontiers in neuroscience* **14**, 577887 (2020).
- Yang, C., N. Luo, M. Liang, S. Zhou, Q. Yu, J. Zhang, M. Zhang, J. Guo, H. Wang, J. Yu, Q. Cui, H. Chen and Q. Gao, Altered brain functional connectivity density in fast-ball sports athletes with early stage of motor training, *Front. Psychol.* **11**, 530122 (2020).
- Yang, G., F. Pan and W.-B. Gan, Stably maintained dendritic spines are associated with lifelong memories, *Nature* **462**, 920 (2009).

- Yang, L. and A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, *Neurocomputing* **415**, 295 (2020).
- Yang, X., N. Zhang and P. Schrader, A study of brain networks for autism spectrum disorder classification using resting-state functional connectivity, *Machine Learning with Applications* **8**, 100290 (2022).
- Yin, Y., Z. Wang, Z. Zhang and Y. Yuan, Aberrant topographical organization of the default mode network underlying the cognitive impairment of remitted late-onset depression, *Neurosci. Lett.* **629**, 26 (2016).
- Yoon, J., E. Yang, J. Lee and S. J. Hwang, Lifelong learning with dynamically expandable networks, arXiv preprint arXiv:1708.01547 (2017).
- Yoon, J. H., F. Strand, P. A. Baltzer, E. F. Conant, F. J. Gilbert, C. D. Lehman, E. A. Morris, L. A. Mullen, R. M. Nishikawa, N. Sharma, et al., Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis: A Systematic Review and Meta-Analysis, *Radiology* **307**, e222639 (2023).
- Yu, W., S. Iranmanesh, A. Haldar, M. Zhang and H. Ferhatosmanoglu, RoleSim*: Scaling axiomatic role-based similarity ranking on large graphs, *World Wide Web* **1** (2021).
- Yuan, B.-K., J. Wang, Y.-F. Zang and D.-Q. Liu, Amplitude differences in high-frequency fMRI signals between eyes open and eyes closed resting states, *Frontiers in human neuroscience* **8**, 503 (2014).
- Zambra, M., A. Maritan and A. Testolin, Emergence of Network Motifs in Deep Neural Networks, *Entropy* **22** (2020).
- Zanin, M., D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik and S. Boccaletti, Combining complex networks and data mining: why and how, *Physics Reports* **635**, 1 (2016).
- Zenke, F., B. Poole and S. Ganguli, “Continual learning through synaptic intelligence”, *International Conference on Machine Learning*, (2017).
- Zhang, C., N. D. Cahill, M. R. Arbabshirani, T. White, S. A. Baum and A. M. Michael, Sex and age effects of functional connectivity in early adulthood, *Brain connectivity* **6**, 700 (2016).
- Zhang, C., S. Bengio and Y. Singer, Are all layers created equal?, *Journal of Machine Learning Research* **23**, 1 (2022).
- Zhang, J., A. Kucyi, J. Raya, A. N. Nielsen, J. S. Nomi, J. S. Damoiseaux, D. J. Greene, S. G. Horowitz, L. Q. Uddin and S. Whitfield-Gabrieli, What have we really learned from functional connectivity in clinical populations?, *NeuroImage* **242**, 118466 (2021).
- Zhang, W., V. Muravina, R. Azencott, Z. D. Chu, M. J. Paldino, et al., Mutual information better quantifies brain network architecture in children with epilepsy, *Computational and mathematical methods in medicine* **2018** (2018).
- Zhang, Y., N. Farrugia and P. Bellec, Deep learning models of cognitive processes constrained by human brain connectomes, *Medical Image Analysis* **80**, 102507 (2022).
- Zhang, Z., P. Cui and W. Zhu, Deep Learning on Graphs: A Survey, *IEEE Transactions on Knowledge and Data Engineering* **34**, 249 (2022).

- Zhao, S., D. Rangaprakash, P. Liang and G. Deshpande, Deterioration from healthy to mild cognitive impairment and Alzheimer's disease mirrored in corresponding loss of centrality in directed brain networks, *Brain informatics* **6**, 1 (2019).
- Zhao, Y., L. Song, J. Ding, N. Lin, Q. Wang, X. Du, R. Sun and Z. Han, Left anterior temporal lobe and bilateral anterior cingulate cortex are semantic hub regions: Evidence from behavior-nodal degree mapping in brain-damaged patients, *Journal of Neuroscience* **37**, 141 (2017).
- Zheng, P., W. Tang and J. Zhang, A simple method for designing efficient small-world neural networks, *Neural Networks* **23**, 155 (2010).
- Zheng, W., Y. Ge, S. Ren, W. Ran, X. Zhang, W. Tian, Z. Chen, W. Dou and Z. Wang, Abnormal static and dynamic functional connectivity of resting-state fMRI in multiple system atrophy, *Aging (Albany NY)* **12**, 16341 (2020).
- Zhu, D., T. Yuan, J. Gao, Q. Xu, K. Xue, W. Zhu, J. Tang, F. Liu, J. Wang and C. Yu, Correlation between cortical gene expression and resting-state functional network centrality in healthy young adults, *Human Brain Mapping* **42**, 2236 (2021).
- Zhu, J., Q. Zeng, Q. Shi, J. Li, S. Dong, C. Lai and G. Cheng, Altered brain functional network in subtypes of Parkinson's disease: A dynamic perspective, *Frontiers in Aging Neuroscience* **13**, 710735 (2021).
- Zhu, L., H. Shu, D. Liu, Q. Guo, Z. Wang and Z. Zhang, Apolipoprotein E ϵ 4 specifically modulates the hippocampus functional connectivity network in patients with amnesic mild cognitive impairment, *Front. Aging Neurosci.* **10**, 289 (2018).
- Zhuang, F., Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, A comprehensive survey on transfer learning, *Proceedings of the IEEE* **109**, 43 (2020).
- Zhuang, K., W. Yang, Y. Li, J. Zhang, Q. Chen, J. Meng, D. Wei, J. Sun, L. He, Y. Mao, X. Wang, D. Vatansever and J. Qiu, Connectome-based evidence for creative thinking as an emergent property of ordinary cognitive operations, *Neuroimage* **227**, 117632 (2021).
- Zimmerman, B., M. Finnegan, S. Paul, S. Schmidt, Y. Tai, K. Roth, Y. Chen and F. T. Husain, Functional brain changes during mindfulness-based cognitive therapy associated with tinnitus severity, *Front. Neurosci.* **13**, 747 (2019).
- Zimmerman, B. J., I. Abraham, S. A. Schmidt, Y. Baryshnikov and F. T. Husain, Dissociating tinnitus patients from healthy controls using resting-state cyclicity analysis and clustering, *Network Neuroscience* **3**, 67 (2018).
- Zou, Q., B.-K. Yuan, H. Gu, D. Liu, D. J. Wang, J.-H. Gao, Y. Yang and Y.-F. Zang, Detecting static and dynamic differences between eyes-closed and eyes-open resting states using ASL and BOLD fMRI, *PloS one* **10**, e0121757 (2015).
- Zouridakis, G., F. Baluch, I. Stevenson and D. Subramanian, "Spatiotemporal Profiles of Brain Activation During Learning and Strategy Formulation", *Joint Meeting of the 6th International Symposium on Noninvasive Functional*

Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, (2007).

Zuo, X.-N., R. Ehmke, M. Mennes, D. Imperati, F. X. Castellanos, O. Sporns and M. P. Milham, Network centrality in the human functional connectome, *Cerebral cortex* **22**, 1862 (2012).

Zuo, X.-N., C. Kelly, J. S. Adelstein, D. F. Klein, F. X. Castellanos and M. P. Milham, Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach, *Neuroimage* **49**, 2163 (2010).

List of Figures

1.1	Principle of fMRI.	13
1.2	Visualization of network-based modeling step.	15
1.3	Interpretability and explainability in FC studies.	17
1.4	Categorization of existing analysis methods for FC graph-based models	18
1.5	Variability in PubMed Literature Search	20
1.6	Visualization of common graph descriptors	24
1.7	Example of hub disruption index computation	31
1.8	Scheme of a classification procedure	32
1.9	Example of GNN model in graph classification	34
1.10	Visualization of examples of ER, WS, BA graph.	35
1.11	Examples of real FC networks and their corresponding model versions .	37
2.1	Degree hub disruption index of FC network of comatose patients	42
2.2	Structural isomorphism visualization	43
2.3	Homophily and Structural Equivalence on <i>Les Misérables</i> network	44
2.4	<code>node2vec</code> sampling strategies	45
2.5	Graphical visualization of the proposed nodal-statistics-based equivalence relation in graph collections	46
2.6	ϵ choice effects in Erdős-Rényi graphs.	48
2.7	Example of the structural pattern associated with different statistics . .	49
2.8	Normalized power coefficient (\widehat{PC}) on nodal statistics incremental sets .	50
2.9	Visualization of the proposed global metrics on three different cases . .	53
3.1	Scheme for BOLD time-series extraction	62
3.2	Adopted framework for graph definition	63
3.3	Small-world regime on FC networks	64
3.4	Mean normalized power coefficient (\widehat{PC}) of clustering coefficient statistics	66
3.5	Mean normalized power coefficient (\widehat{PC}) of degree statistics.	67
3.6	Nodal statistics pair comparison on WS and BA2	68
3.7	Structural patterns on Watt-Strogatz and Barabási-Albert models	69
3.8	Orthogonality and correspondence structural pattern scores on HCP dataset.	70
3.9	Correspondence structural pattern distribution	71
3.10	Nodal participation in nontrivial classes on HCP and synthetic datasets	71
3.11	Nodal participation in nontrivial class in HCP dataset and a shuffled version	72
3.12	Orthogonality score curves of pairs of nodal statistics on healthy controls (HC) over sparsity	73
3.13	Comparison of Nodal percentage of participation for HC datasets.	75
3.14	Orthogonality curves and closeness centrality average on CO and HC . .	78
3.15	Accuracy results in patients classification	79
3.16	Distribution of correspondence structural patterns score on CO and HC	80

3.17	ROC curve of discrimination between CO and HC	81
3.18	Nodal participation in nontrivial classes for CO, HCP and DSP model	82
3.19	Mean participation of a node in datasets of different size	83
3.20	Nodal percentage of participation in HC and CO	83
3.21	Orthogonality curve and closeness centrality average on PD and HC	84
3.22	Visualization of the nodal PP in the different subject groups	87
3.23	Edges discrepancies using shorter time-series for network estimation	91
3.24	Reliability of nodal metrics in iShare dataset.	92
3.25	Reliability of nodal metrics in HCP dataset.	93
3.26	Bland-Altman plot of clustering coefficient statistics	94
3.27	Nodal percentage of participation on standard, shorter and attacked dataset version	96
4.1	Examples of common activation functions	106
4.2	Visualization of the backpropagation algorithm in a feedforward neural network	108
4.3	Overfitting and good generalization model	110
4.4	Scheme of the sequential learning framework of three tasks	112
4.5	Three continual learning scenarios	113
4.6	Overview of existing continual learning strategies	114
4.7	Brain-inspiration of continual learning strategies	115
4.8	Visualization of GDumb learning strategy	116
4.9	Architecture scheme of Dream Net model	120
5.1	Overview of the brain-inspired perspective of the proposed graph-based analysis of ANNs	125
5.2	Toy example of induced graph	129
5.3	Examples of interquartile and tail nodes	130
5.4	Schematic visualization of the pruning versions	132
5.5	Weights injection with fine-tuning procedure	133
6.1	ResNet50 model architecture	142
6.2	In and out-degree distribution of induced graph	146
6.3	Results of Experiment 1 (Characterization)	147
6.4	Clustering results on ANNs for Handwritten Recognition task	148
6.5	Visualization of the relation between \mathcal{S} and the norm changes of the weights	149
6.6	Clustering results on ANNs for Face Emotion Recognition task	149
6.7	Results of Experiment 2 (Network Surgery) for two orders of Handwritten Digit Recognition	151
6.8	Results of Experiment 2 (Network Surgery) for two orders of Face Emotion Recognition Task	153
6.9	Results of Experiment 3 (Weights-injection) on MNIST configurations	155
6.10	Results of Experiment 3 (Weights-injection) on Fer+ configurations	156
A.1	Atlas influence on orthogonality curves	XX
A.2	Bland-Altman plot of the degree nodal statistics.	XXI

A.3	Bland-Altman plot of the betweenness centrality statistics.	XXII
A.4	Bland-Altman plot of the closeness centrality statistics.	XXIII
A.5	Bland-Altman plot of II order centrality statistics.	XXIV
A.6	Fuzzy network analysis visualization	XXVI
A.7	Results of Experiment 2 (Network Surgery) for all orders and strategies on MNIST task	XXIX
A.8	Results of Experiment 2 (Network Surgery) for all orders and strategies on Face Emotion Recognition Task	XXX

List of Tables

1.1	Literature search on PubMed	20
1.2	Summary of the methods and results from the studies on Parkinson Disease	22
2.1	Semantic and Syntactic definition	44
2.2	Entropy and PC* on known graphs	56
3.1	List of the studied databases including different classes of subjects. . .	64
3.2	Ratio of nodes in nontrivial classes in the same class of their homotopical in HCP dataset.	73
3.3	Mean Accuracy on center prediction performed on the healthy control datasets.	75
3.4	Regression parameters between nodal participation of different HC groups for different pairs of statistics	76
3.5	Statistics on nodal percentage of participation	81
3.6	Nodes in nontrivial class per graph	82
3.7	Classification results on PD	86
3.8	List of regions significantly different in HC vs PD <i>de novo</i>	88
3.9	List of significantly different regions in HC vs (PD-I, PD-H)	88
3.10	Reliability interpretation based on ICC values	90
3.11	Number of significant rejects of the null hypothesis at 0.05 in degree comparison	92
4.1	Desiderata of a robust neural network systems	110
4.2	Glossary of different learning paradigms	111
6.1	Hyperparameters of the considered learning strategies in MNIST settings.140	
6.2	Results the trained configuration on MNIST	141
6.3	Hyperparameters of the considered learning strategies on Fer+	143
6.4	Results the trained configuration on Fer+	144
A.1	Query Results on Pubmed	XI
A.2	Accuracy results when considering single nodal statistics or their complete combination	XIX

List of Abbreviations

ANN	Artificial Neural Network
FC	Functional connectivity
fMRI	Functional Magnetic Resonance Imaging
rs-fMRI	resting-state fMRI
TR	Repetition time
ICA	independent components analysis
GNN	Graph Neural Network
HC	Healthy Controls
PD	Parkinsonian patients
d	Degree
cc	Clustering coefficient
b	Betweenness centrality
cs	Closeness centrality
s	Second-order centrality
ER	Erdős-Rényi model
WS	Watts-Strogatz model
BA1 / BA2	Barabási-Albert model
DSP	Degree sequence preserving model
EPA	Economical preferential attachment model
EC	Economical clustering model
PC	Power Coefficient
NPP	Nodal percentage of participation in nontrivial classes
ROI	Regions of interest
CO	Comatose patients
PD-H	Parkinsonian patients with visual hallucinations
PD-I	Parkinsonian patients with visual illusions
AVG	Average
SD	Standard Deviation

SVM	Support Vector Machine
PP	percentage of participation in trivial classes
ICC	Intra-Class Correlation Coefficient
BA	Bland-Altman
TIL	Task incremental learning
DIL	Domain incremental learning
CIL	Class incremental learning
EWC	Elastic-Weight-Consolidation
SI	Synaptic Intelligence
LwF	Learning withouth Forgetting
LRP	Layer-wise relevance propagation

Part C

Appendix

A

A.1 PubMed Query Results

Table A.1: Query Results on Pubmed. MF: magnetic field, Num Vol: Number of Acquired Volumes, ScD: Scan Duration, Fr: resting-state considered frequency, EC: eyes condition, Num S: Number of Subjects

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Abdallah et al. 2019	3	3000	200	10'			ROIs
Abdelnour et al. 2018	3	2000	199	6'38"	0.01-0.08	OX	AAL90
Abdelnour et al. 2021	3	2000	199	6'38"	0.01-0.08	O	AAL90
Adhikari et al. 2017	3	2500	768	30'	0.009-0.08	O	DK ²
Amiri et al. 2020	3	3000	330	16'30"	0.01-0.1	C	AAL90
Audrain et al. 2018	3	2000	180	6'	0.008-0.09	C	Brainnetome ³
Bachmann et al. 2018	3	3000	140	7'	<0.02		
Bansal et al. 2021	3	2200			0.01-0.1	C	DK
Behfar et al. 2020	3	3000	155	7'		O	
Ben Simon et al. 2017	3	2500		0	0.01-0.08	OX	
Bharath et al. 2016	3	3000	185	9'40"		C	Craddock
Billings et al. 2018	3	645	900	9'40	wavelet	O	voxel
Blommaert et al. 2020		1700	254	7'			

²Desikan-Killiany, Desikan et al. 2006

³Fan et al. 2016

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Brenner et al. 2018	3	2000	210	8'	<0.15		HAO ⁴
Campabadal et al. 2020	3	2500		10'	<0.01	C	
Cao et al. 2017	3	1500	210	5'15"	0.01-0.10	C ⁵	voxel
Carotenuto et al. 2020	3	3000	240	12'	0.008-0.09		MNI ⁶
Cary et al. 2017	3	2000	600	20'	<0.08	O	Power
Chen 2019a	3	720	4666	58'		O	ROIs
Chen 2019b	3	2000		8'	0.009-0.08	O	Power
Chen et al. 2019	3	2000	240	8'	0.01-0.08	C	AAL90
Chirumamilla et al. 2016		3000					AAL116
Chockanathan et al. 2019	3	1650	250	412'5	0.038-0.076		AAL16
Chockanathan et al. 2018	3	1650	250	412'5		C	
Collantoni et al. 2019	1.5	2009	250	8'22"	0.005 – 0.1	C	ROIs
Cope et al. 2018	3	2430	269	11'			
Cordova et al. 2020	3	2500	120	5'			ROIs
Dall'Acqua et al. 2017	3	2220	140	5'19"	0.01-1	C	AAL90
De Micco et al. 2021	3	1508	240	6'	0.01-0.08	C	ROIs
Schipper et al. 2018	3	2200	200	7'29"	<0.01	C	
Deng et al. 2016	3	3000	137	6'51"	0.01-0.08		voxel
Diez et al. 2019		5000		6'40"		O	
Dong et al. 2020	3	1900	216	7'12"	0.01-0.08		

⁴Harvard-Oxford atlas [Smith et al. 2004](#)⁵natural sleep⁶Montreal Neurological Institute

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Douw et al. 2020	3	1800	200	6'	0.06-0.12	C	Brainnetome
Duncan et al. 2016	3	1500	200	5'	0.01-0.1		
Eldaief et al. 2017	3	2500			<0.08	O	
Erdeniz et al. 2017	1.5	2640		11'	0.01-0.08	OX	AAL90
Fan et al. 2021	3	2000	240	8'	0.01-0.08	C	AAL90
Figuroa-Jimenez et al. 2021	3	2000	220	6-9'		OX	
Finotelli et al. 2018	1.5	2500	160	6'06"	<0.01	C	HOA ⁷
Finotelli et al. 2019	3	2000		5'	0.01-1	C	AAL90
Fitzsimmons et al. 2020	3	2100					ROIs
Fujimoto et al. 2020	3	1000	360	6'		OX	
Fulong et al. 2020	3	2030	240	8'32	0.01-0.1	O	
Gaudio et al. 2018	1.5	3560	80	6'44"	0.01-0.08		AAL116
Gerchen et al. 2017	3	1370		5'08"			
Gilson et al. 2019	3	2400	244	9'46"	0.01-0.1		
Hahn et al. 2021	3	2080	1505	52'17"	0.01-0.1		AAL90
Han et al. 2020	3	2000	204	6'56"	0.009-0.08	C	Power
Han et al. 2019	3	2000	200	8'	0.01-0.08		Power
Hart et al. 2016	3	2420	269	10'51"	wavelets		
He et al. 2016	3	2000	225	7'30"		O	
He et al. 2017				5'	wavelet 0.05-0.1		AAL90 / HOA

⁷Harvard-Oxford Atlas

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Herzberg et al. 2021	3	2500	140	5'57"	0.009–0.08		
Ho et al. 2017	3	2000	256	8'32"	0.009–0.08		
Hojjati et al. 2019	3	3000	140	6'	0.01–0.08	O	Dosenbach
Hojjati et al. 2018	3	3000	140	7'	0.01–0.08	O	AAL90 /Dosen- bach
Holla et al. 2020	3	2000	303	10'	0.01–0.09	O	Power
Honnorat et al. 2017		3000	120	6'	0.01–0.08		
Hou et al. 2019	3	2000	154	5'08"	0.01–0.1		
Hrybouski et al. 2021	4.7	3000	200	10'	<0.1		
Huang et al. 2019	3	2500	230	9'35"	0.008–0.01		ROIs
Huang et al. 2017	3	1000	360	6'	0.01–0.1	C	Voxel
Hou et al. 2019	3	2000	242	8'	0.01–0.1		
Iordan et al. 2017	3	2000	235	7'50	0.01–0.1	OX	Power
Kang et al. 2016	3		150	7'30"			Voxel/AAL90
Kato et al. 2021		2500	198	8'15	0.01–0.1	C	
Kesler et al. 2017	3	2000	216	7'12"	0.008–0.09	C	AAL90
Kharabian Masouleh et al. 2017	3	2000	300	10'	0.005–0.1	C	
Kim et al. 2020b	3	3000		7'30"			ROIs
Kirshenbaum et al. 2022	3	2000	180	6'		C	Brainnetome
Klooster et al. 2019	3	2000		10'12"	0.1–0.01		
Kong et al. 2021	3	2000	240	8'	0.01–0.08		

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Lee et al. 2017	3	3000	100	5'	0.009-0.08	C	ROIs
Leming et al. 2019	3	3000	256	8'56"		C	AAL116
Levakov et al. 2021	⁸ 3 7	645 1000		10' 0	0.008-0.08 0.0005		Schaefer
Li et al. 2017a	3	2000	240	8'	0.01-0.08	C	voxel /seed based
Lin et al. 2016	4	2500		10'	0.009-0.08	C	seed-based
Lin et al. 2019	3	3000	100	5'	0.01-0.1		Power
Lindner et al. 2018	3	2500			0.008-0.09	OX	AAL90
Liu et al. 2019a	3	2000	240	8'	0.01-0.08		AAL116
Liu et al. 2019b	3	2000	230	7'40	0.01-0.08	C	ROIs
Liu et al. 2017	3	2000	180	6'	0.06-0.1	OX	AAL90
Liu et al. 2018		2000	200	6'40	0.01-0.08	C	AAL90
Long et al. 2021	3	2000	150	5'		OX	
Luppi et al. 2021	3	2000		7'20" x3 ⁹	0.008-0.09	C	Schaefer ¹⁰ / Brain- netome
Makovac et al. 2018	1.5	2520		5'	wavelets 0.05-0.1		
Mandelli et al. 2016	3	2000	240	480	0.008-0.15	C	seed-based ROIs
Mao et al. 2021		2000	210	420	0.01-0.08	C	
Maximo et al. 2017	3	1000	419	6'59"	0.008-0.08	O	KsRN ¹¹
Mazrooyisebdani et al. 2020	3	2600		5'	0.009-0.08	C	ROIs
Mijalkov et al. 2017	3	2400	210	8'29	0.01-0.08		

⁸two different datasets are considered in the same paper

⁹x3: the scan is repeated 3 times

¹⁰Schaefer et al. 2018

¹¹Koyama's Reading Network Koyama et al. 2010

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Mirakhorli et al. 2020	3	3000	140	7'			AAL2
Neudorf et al. 2020	3	720	1200	14'24"			Shen ¹² / voxels
Neumann et al. 2021	3	2500	200	8'20	0.008-0.1	¹³	Voxels
Ogawa 2021	3	720		0	0.0005		
Openneer et al. 2020	3	2300	215	8'24"	0.008-0.08	OX	
Osadchiy et al. 2019	3	2000		10'	0.008-0.08	C	ROIs
Paldino et al. 2017	3	2000	300	10'	<0.01		variable ¹⁴
Petersen et al. 2016	1.5	2000	240	8'	0.01-0.08	NSI ¹⁵	seed based
Pezoulas et al. 2017	3	720	91	01'05"		OX	
Prajapati et al. 2021	3		123				Dosenbach
	3	2400	220	8'48"	0.01-0.1		
Puig et al. 2018	1.5	2500	240	10'	0.01-0.08		
Rangaprakas et al. 2019	3	600	1000	10'	0.01-0.1	O	Craddock
Ray et al. 2017	3	2000				OC	ROIs
Regner et al. 2016	3	2000	150	5'	0.008-0.15	C	ICA
Ribeiro de Paula et al. 2017	3	2000				C	DK
Roberts et al. 2017	3	2000	188	6'16		C	ROIs
Rodriguez et al. 2021	3	2000	300	5'	<0.15		
Ryan et al. 2018	3	1500			0.01-0.1	OX	Seed

¹²Shen et al. 2013¹³no specific instructions, keep their head still and avoid movements¹⁴The number of nodes in each patient's network ranged from 511 to 841 (mean: 684; standard deviation: 68)¹⁵no specific instructions

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Sadeh et al. 2019	3	3000		6'		O	
Saghayi et al. 2020a	3			5'			HOA
Sako et al. 2019	3	2000			0.01-0.08	C	
Sato et al. 2018	1.5	2000	180	6'	0.01-0.1	OX	seed based
Sen et al. 2021		2000		6'		C	ROIs
Shaw et al. 2021	3	2000			0.008-0.09		
Sheffield et al. 2016	3	2300		60'	0.009-0.08		Power
Simpson et al. 2019	1.5	3000	200	6'	0.00765-0.068	OX	AAL90
Song et al. 2019		2000			0.01-0.08		
Sourty et al. 2016	3	3000	121	6'03"		O	ICA
Storti et al. 2017	3	2200	256	9'23"	0.008-0.1	C	ROIs
Subramanian et al. 2020	3	3000	200	10'	0.01-0.1	OX	
Sun et al. 2017a	3	2000	240	8'	0.01-0.1	OX	AAL90
Sun et al. 2017b	3	2550	210	9'	0.01-0.1		AAL90
Surampudi et al. 2019	3	720	1200	14'24"	0.01-0.08	OX	
Taya et al. 2016	3	2000	163	5'26"	0.01-0.1	O	AAL90
Taya et al. 2018	3	2000	163	5'	0.06-0.012	O	HOA
Tomasi et al. 2017,	3	720	1200	14'24"	0.01-0.08	OX	ROIs
Trofimova et al. 2021	3	2000	300	10'	0.008-0.09		ROIs
Tumati et al. 2020	3	3000	140	7'	0.009-0.08	C	7 different subnetworks
Vaghi et al. 2017	3	2470		10'	0.049-0.101	OX	AFNI

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Vancea et al. 2019					0.009-0.08	O	ROIs
Váša et al. 2018	3	3000	104	5'12"			clustering
Vatansever et al. 2020	3	3000	180	9'			
Vergara et al. 2018	3	2000	152	5'04"	0.01-0.15	OX	
Wagner et al. 2019	3	2090	285	9'55"	0.01-0.08	OX	ROIs
Wang et al. 2020a	3	3000	133	6'45"	0.01-0.08	C	Power
Wang et al. 2020b		3000		6'20"	0.01-0.1	OX	
Wang et al. 2019	3	2000	250	8'20"	0.01-0.08	C	BN246 ¹⁶
Wang et al. 2018b					<0.0167		ROIs
Welton et al. 2020	3	2200	191	7'			HOA
Wirsich et al. 2017	3	3600		21'	0.04-0.09	C	Destrieux
Xiang et al. 2019	4,7				0.01-0.1	O	Brainnetome
Xiao et al. 2021	3	3000	124	6'	0.01-0.1	OX	Power
Xie et al. 2021	3	720	1200	14'24"	0.043-0.087	OX	DK ¹⁷
Xu et al. 2020b	3	2000	240	8'	0.01-0.1	C	AAL90
Xu et al. 2020a	3	2000	240	8'06"	0.01-0.1	C	Power/ AAL90
Yang et al. 2020a		2000	266	8'52"	0.01-0.08	C	
Yin et al. 2016	1.5	3000	132	7'06"	0.01-0.08	C	ROIs
Zhang et al. 2016	3	720	1200	14'24"		OX	AAL116
Zhang et al. 2018	3	2000		10'	<0.01	C	freesurer

¹⁶Wang et al. 2019¹⁷desikan-killiany

	MF (T)	TR (ms)	Num Vol.	ScD (min)	Fr (Hz)	EC ¹	Atlas
Zhao et al. 2019	3	3000	140	7'			Craddock
Zhao et al. 2017	1.5	2000	120	4'	0.01–0.1	C	AAL90
Zheng et al. 2020	3	2000		6'	0.01–0.08	C	
Zhu et al. 2021a	3	720	1200	14'24"	0.01–0.08	OX	Brodmann
Zhu et al. 2021b.	3	2400	200	8'40"	<0.15		ICA
Zhu et al. 2018	3	2000		8'	0.01–0.1	C	Voxels
Zhuang et al. 2021	3	2000	242	8'04	0.01–0.1	C	AAL90 ¹⁸
Zimmerman et al. 2019	3	2000	304	10'	0.008–0.08		HOA

Supplementary Results on the Orthgonality

Table A.2: Accuracy results obtained when comparing the orthogonality curves of single nodal statistics or over the combination of all the five considered statistics. Discr. Task: Discrimination task.

Discr. Task	Degree	Betweenness C.	Clustering Coeff.	Closeness C.	II Order C.	All
CO	0.78	0.81	0.58	0.82	0.56	0.79
PD	0.62	0.62	0.80	0.62	0.66	0.75
Data-center	0.35	0.53	0.36	0.31	0.31	0.25

Atlas On the HCP and iShare dataset, we compare the effect of the atlas choice on the orthogonality curves in Fig. A.1. The orthogonality of degree and clustering coefficient seems invariant with respect to the atlas choice, while closeness centrality and II order centrality orthogonality approaches one, meaning that for finer atlas the modularity structure of the graph for every sparsity is stronger. Interestingly, the orthogonality curves of the two datasets appear to be consistent: minimum and maximum reached at the same sparsity level. That reveals the orthogonality is capturing the presence of some graph substructures proper of the considered dataset. Interestingly, the orthogonality curves of the two datasets appear to be consistent: minimum and maximum reached at the same sparsity level. That reveals the orthogonality is capturing the presence of some graph substructures proper of the considered dataset.

¹⁸re-parcellated into 1024 regions equal volume size

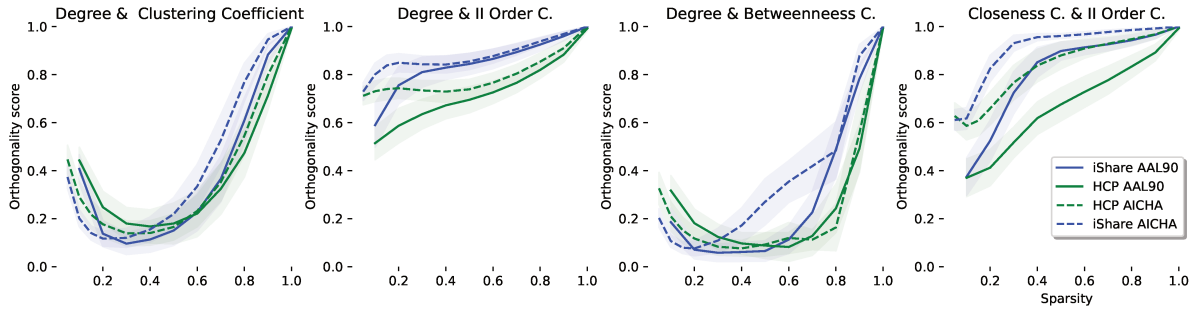


Figure A.1 – Orthogonality sensitivity to the atlas choice (AAL90 and AICHA with 384 regions) in HCP and iShare datasets.

Single and complete combination of nodal statistics To evaluate the benefit of the combination of nodal statistics, we reproduce the classification tasks over the orthogonality score on a single statistics and on the collection of all the considered statistics. Results are shown in Table A.2. Depending on the statistics, the discrimination is still possible when considering the orthogonality, while instead when using graph average of a single statistics without taking into account its associated structural patterns, a group discrimination is not possible (Fig. 3.14 and 3.21 panels (b)). The lower accuracy when using the combination of all the nodal statistics represents the fact that increasing the nodal statistics in the collection determine a too fine structural patterns making it harder (even if not impossible) to perform the discrimination.

A.2 Bland-Altman Plots

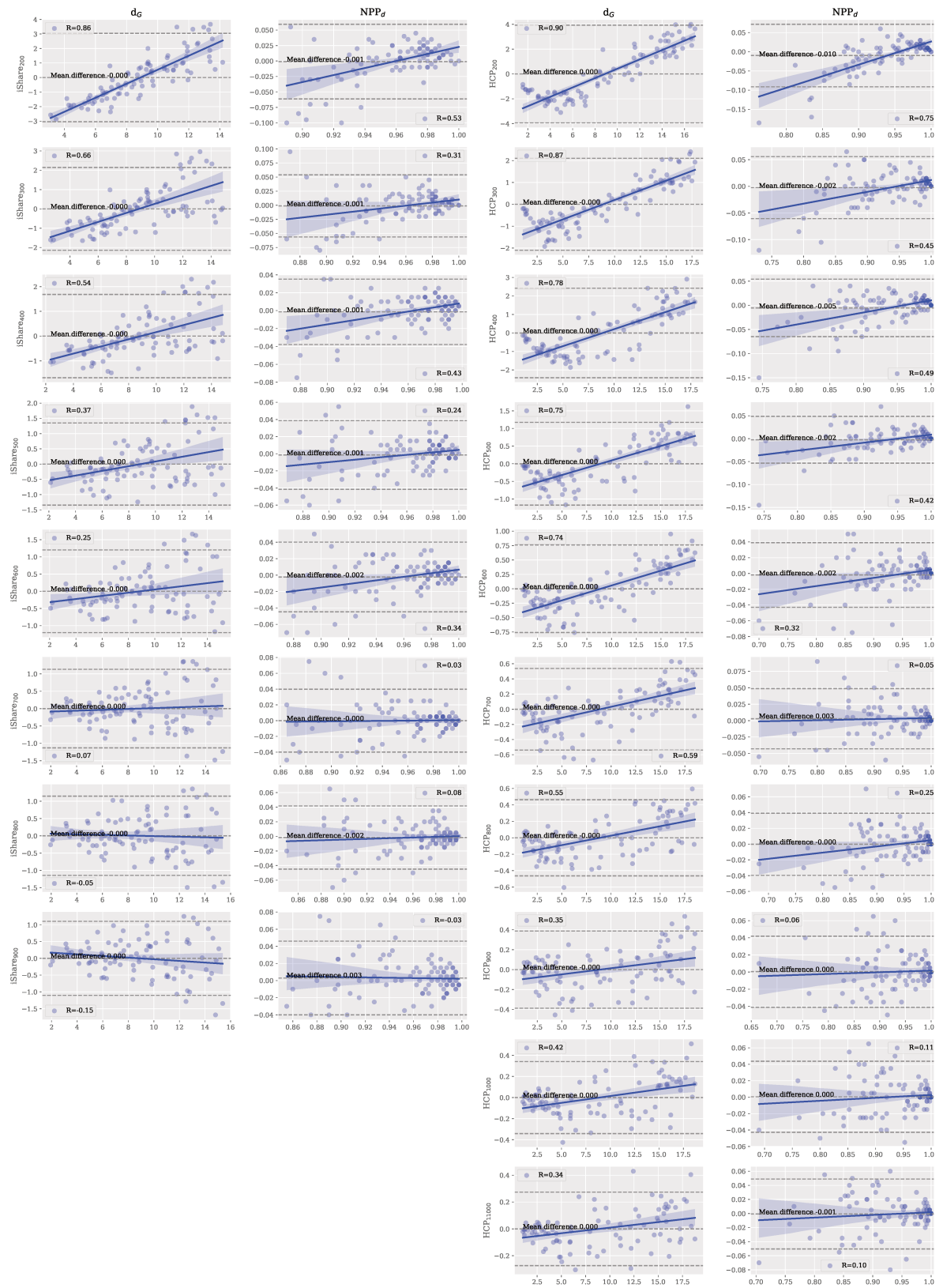


Figure A.2 – Bland-Altman plot of the degree nodal statistics.

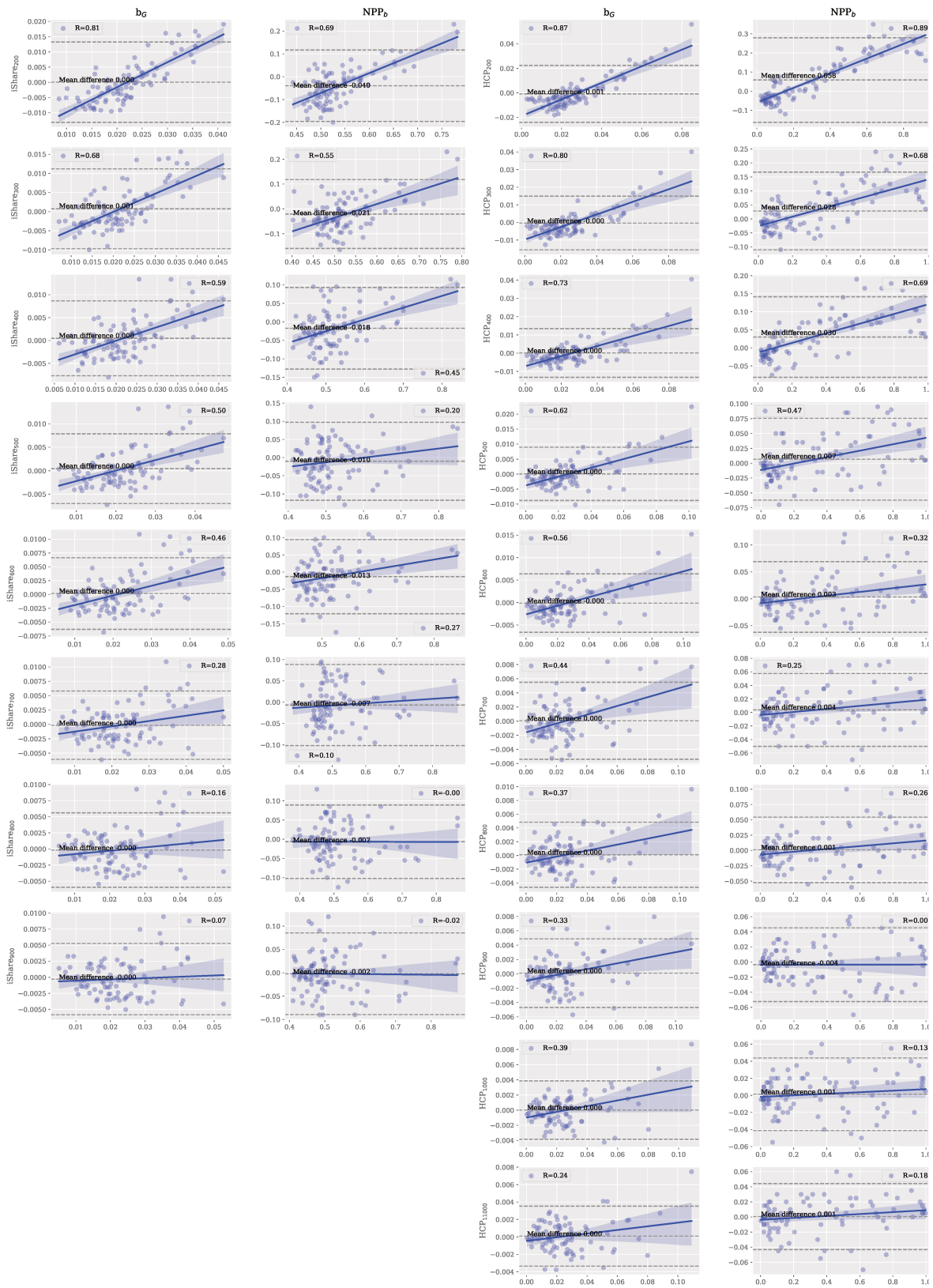


Figure A.3 – Bland-Altman plot of the betweenness centrality statistics.

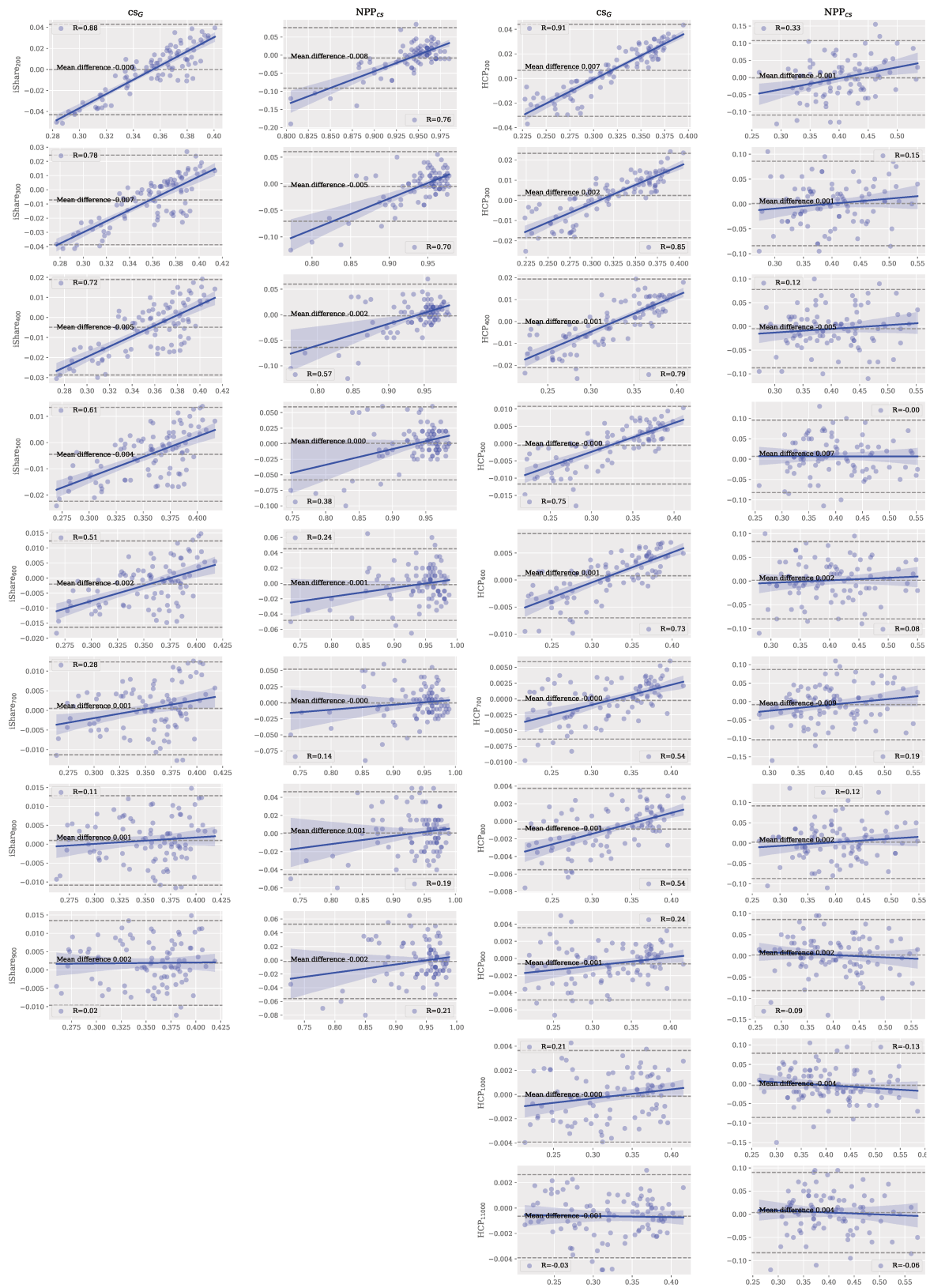


Figure A.4 – Bland-Altman plot of the closeness centrality statistics.

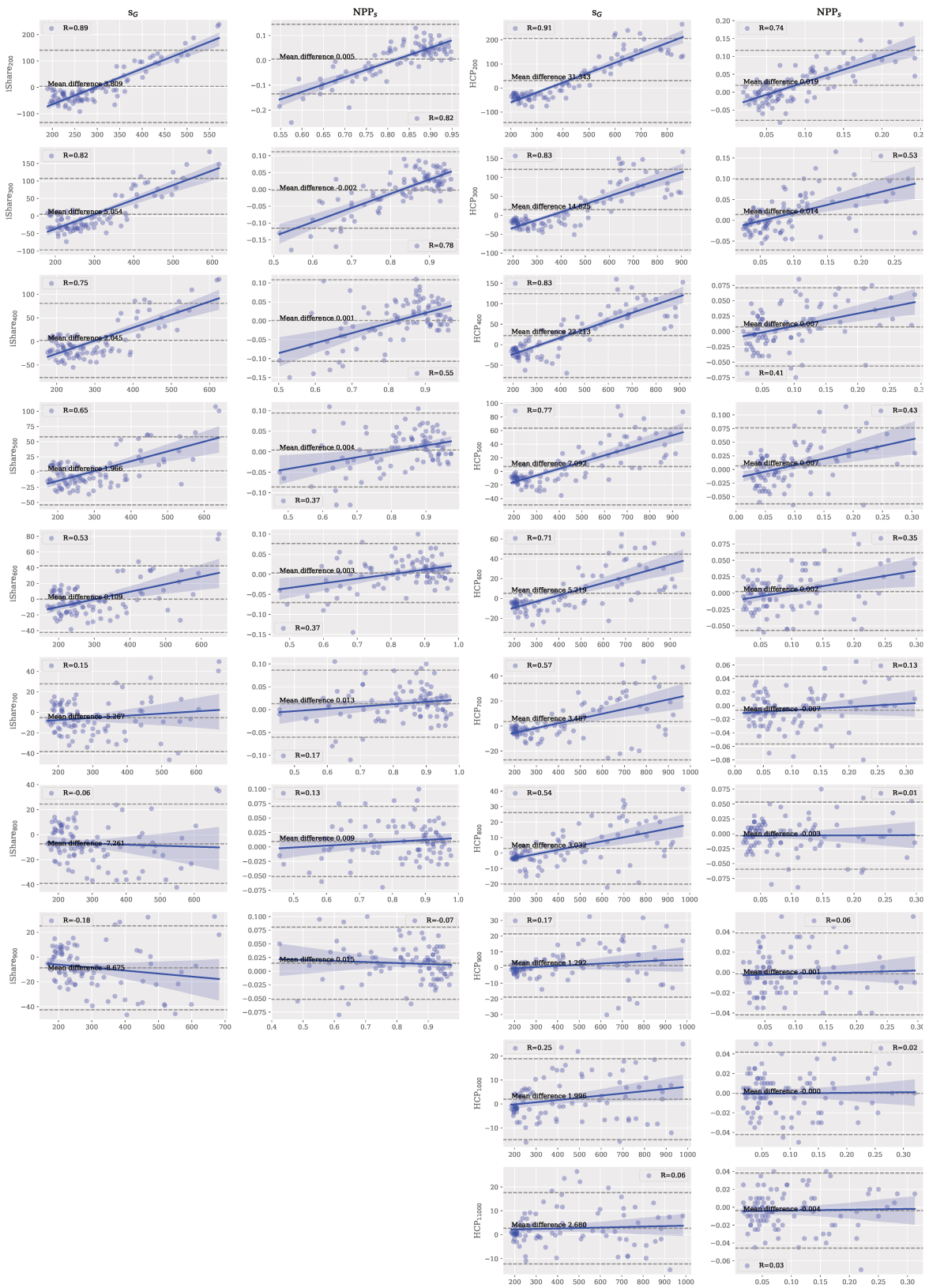


Figure A.5 – Bland-Altman plot of II order centrality statistics.

A.3 Structural Pattern under Uncertainty Graph Estimation

In network analysis, the reliability and accuracy of edge estimation play a crucial role in drawing meaningful conclusions from network data. As we showed in our previous chapter even a small deviation or error in edge estimation can lead to significant differences in the resulting network structure and properties.

In this section, we propose an extension of our proposal to take into account uncertainty in edge estimation. In our structural equivalence definition the implications of even a single spurious edge on the overall estimation of equivalence classes underline the need for a rigorous approach to address this uncertainty.

To address this challenge, we extend our proposal into the realm of fuzzy networks, adopting the framework introduced by [Raimondo et al. 2021](#). This extension allows us to encapsulate the inherent uncertainty in edge estimation within our structural equivalence analysis.

To consolidate our findings and showcase the real-world applicability of our approach, we present a first attempt of application, revealing the benefit of considering uncertainty in practical network equivalence analysis scenarios.

A.3.1 Fuzzy Network Definition

The work of [Raimondo et al. 2021](#) proposes to replace traditional nodal statistics with their uncertainty estimation. In particular, they replace the adjacency matrix with a weighted matrix where each entry corresponds to the probability of existence of the corresponding edge.

$$P = \begin{bmatrix} \pi_{1,1} & \dots & \pi_{1,N} \\ \vdots & \ddots & \vdots \\ \pi_{N,1} & \dots & \pi_{N,N} \end{bmatrix} \quad (1.1)$$

The corresponding associated graph can be interpreted as a complete graph including all possible edges among nodes each one weighted by its probability of existence $\pi_{i,j}$. Let us consider a single node in the fuzzy representation of the complex network (Fig. [A.6](#)). The node v is connected to $N - 1$ edges, and each edge is associated with an independent probability of being present. Due to this uncertainty, the conventional definition of node degree, which represents the number of edges connected to a node, no longer applies. This is because the node can have all possible degrees simultaneously, each with a certain probability.

Definition A.1 (Degree of a node in a fuzzy network). Following [Raimondo et al. 2021](#), we define the degree of a node in a fuzzy network as a random variable described by its probability distribution, which depends on the probabilities $\pi_{u,v}$. The probability of node v having degree $d_v = k$ corresponds to the probability of having k successes in a sequence of $N - 1$ independent Bernoulli trials with success probabilities

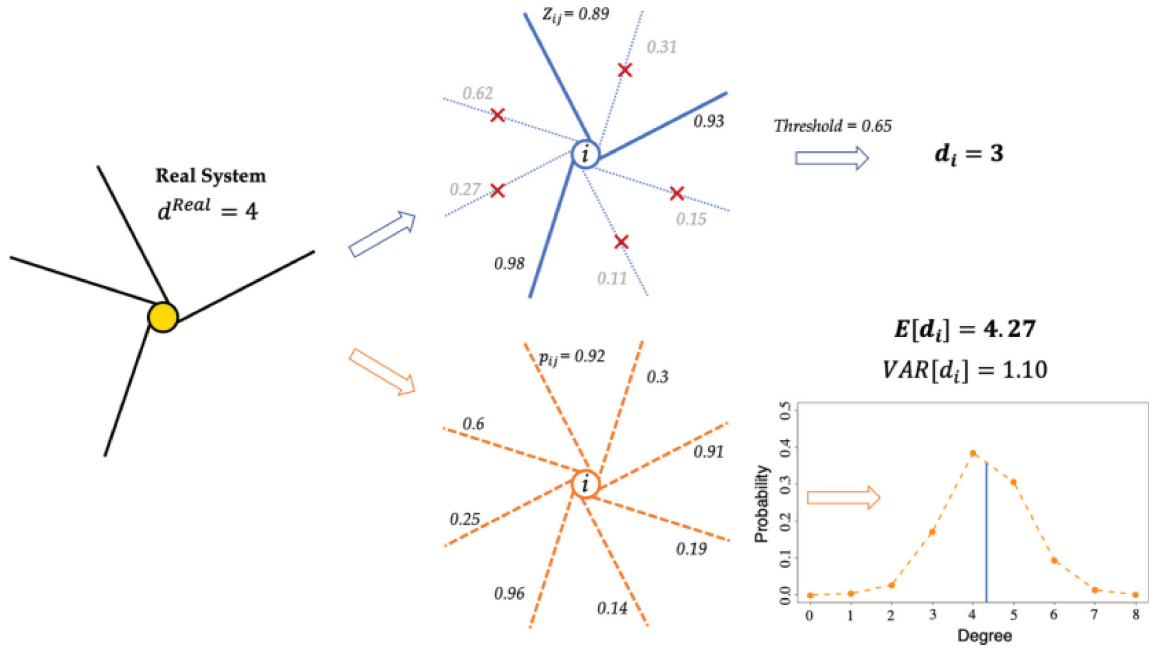


Figure A.6 – Fuzzy network analysis visualization. Visualization of the difference resulting analysis on the degree of a node when considered as node of thresholded graph or when inserted in a fuzzy representation. Reproduced with permission from Raimondo et al. 2021

$\pi_{v1}, \pi_{v2}, \dots, \pi_{vN}$ (we exclude the possibility of having the edge e_{vv}):

$$e_{ij} | \pi_{ij} \sim \text{Bernoulli}(\pi_{ij}), \quad d_i = \sum_{j=1}^N e_{ij} | \pi_{ij}. \quad (1.2)$$

In general, each edge incident to node v has a different probabilities of existence, but in the case π_{vu} were all equal for all possible choice of u , then the probability distribution of the d_v would be a Binomial distribution. The probability of having k successful trials out of a total of $N - 1$ can be written as:

$$P(d_i = k) = \sum_{S \in F_k} \prod_{j \in S} \pi_{ij} \prod_{l \in \bar{S}} (1 - \pi_{il}), \quad (1.3)$$

where F_k is the set of all subsets of k edges that can be selected from $e_{v1}, e_{v2}, e_{v3}, \dots, e_{vN}$. For example, if $N = 3$, then $F_2 = \{e_{v1}, e_{v2}\}, \{e_{v1}, e_{v3}, e_{v2}, e_{v3}\}$. \bar{S} is the complement of S , i.e., the set $\bar{S} = \{e_{i1}, e_{i2}, e_{i3}, \dots, e_{iN}\} \setminus S$. This distribution is the so-called Poisson-Binomial distribution, and it represents the node degree distribution, having mean and variance

$$\mu_{d_i} = \sum_j \pi_{ij}, \quad \sigma_{d_i}^2 = \sum_j \pi_{ij}(1 - \pi_{ij}). \quad (1.4)$$

When the graph structure is inferred from noisy signal, the use of an entire distribution for each node provides more information than the usual definition of degree, allowing to consider uncertainty.

Similarly, we can define the other nodal statistics in terms of probability distributions. We refer to [Raimondo et al. 2021](#) for the complete details.

Definition A.2 (Clustering coefficient of a node in a fuzzy network). In a standard network, the clustering coefficient of a node is defined as a function of the number of triangles including the node itself (See Def. 1.17), normalized with respect to the node degree.

In a network of N nodes, each node may be tied to t closed triangles in $\binom{N-1}{t}$ distinct configurations. The probability of each configuration c is given as a function of the probability of existence of each edge involved in the corresponding triangle:

$$q_c = P\left(\bigcap_{i,j \in S_c} e_{ij}\right) \cdot \left(1 - P\left(\bigcap_{i,j \in \bar{S}_c} e_{ij}\right)\right)$$

where S_c is the event set of all pairs of nodes associated to the configuration c , and \bar{S}_c is its complementary set. Again, assuming that the existence of each edge is an independent random variable as in (Equation 1.2), the configurations are mutually disjoint events. Thus, the probability that the node i has a clustering coefficient equals to k is given by:

$$P(cc_i = k) = \sum_{c \in \tilde{C}_i} q_c = \sum_{c \in \tilde{C}_i} q_c, \quad (1.5)$$

where \tilde{C}_i represents the set of configurations in which node i has a clustering coefficient equal to k .

While all nodal statistics can be potentially adapted from deterministic graphs to fuzzy networks, the primary challenge in effectively determining their distribution for each node arises from the vast number of potential event configurations that must be considered. This becomes especially intractable for nodal statistics that involve higher orders of neighborhood information, such as centrality measures. Keeping track of all these configurations can quickly become unmanageable.

In the following, we extend the notion of structural equivalence associated with a nodal statistics interpreted as a random variable.

A.3.2 Nodal-statistics-based equivalence relation with uncertainty

Here, we extend the notion of node structural equivalence to the degree in a fuzzy network. We re/use the notation introduced in the previous chapter. Let s be the degree as random variable and let u, v be two nodes of the graph, we first determine the probability of $u \sim_s v$, $\mathbb{P}(u \sim_d v)$.

$$\mathbb{P}(u \sim_d v) = \mathbb{P}(\deg(u) = \deg(v)) = \mathbb{P}(d_u = d_v) = \sum_k \mathbb{P}(d_u = k, d_v = k) = \quad (1.6)$$

we re-write eq. 1.6 as sum of independent events as:

$$= \sum_k \left(\mathbb{P}(d_u = k, d_v = k | e_{uv} = 1) \mathbb{P}(e_{uv} = 1) + \mathbb{P}(d_u = k, d_v = k | e_{uv} = 0) \mathbb{P}(e_{uv} = 0) \right) = \quad (1.7)$$

$$= \sum_k \left(\left(\mathbb{P}(d_u = k | e_{uv} = 1) \mathbb{P}(d_v = k | e_{uv} = 1) \mathbb{P}(e_{uv} = 1) \right) + \left(\mathbb{P}(d_u = k | e_{uv} = 0) \mathbb{P}(d_v = k | e_{uv} = 0) \mathbb{P}(e_{uv} = 0) \right) \right) = \quad (1.8)$$

$$= \sum_k \left(\mathbb{P}(d_u = k | e_{uv} = 1) \mathbb{P}(d_v = k | e_{uv} = 1) \pi_{uv} + \mathbb{P}(d_u = k | e_{uv} = 0) \mathbb{P}(d_v = k | e_{uv} = 0) (1 - \pi_{uv}) \right) = \quad (1.9)$$

we denote F_k^u the set of all subsets of k edges selected from e_{i1}, \dots, e_{iN} excluding e_{iu}

$$\mathbb{P}(u \sim_d v) = \sum_k \left(\pi_{uv} \sum_{S \in F_{k-1}^v} \sum_{T \in F_{k-1}^u} \prod_{j \in S} \pi_{uj} \prod_{l \in T} \pi_{vl} \prod_{p \in \bar{S}} (1 - \pi_{up}) \prod_{q \in \bar{T}} (1 - \pi_{vp}) \right) \quad (1.10)$$

$$+ (1 - \pi_{uv}) \sum_{S \in F_k^v} \sum_{T \in F_k^u} \prod_{j \in S} \pi_{uj} \prod_{l \in T} \pi_{vl} \prod_{p \in \bar{S}} (1 - \pi_{up}) \prod_{q \in \bar{T}} (1 - \pi_{vp}) \quad (1.11)$$

Equations 1.10 line gives the probability of two nodes being degree equivalent. This probability results in a pair-wise relation among nodes which, unlike what happened when the edges are supposed to be stably reconstructed, can not be directly extended in its corresponding induced partition. Indeed, such construction requires to find all possible 2^N subsets of \mathcal{V} which can be equivalence classes and associated them with their corresponding probability.

Our proposal, is thus to extend the structural equivalence for fuzzy network by considering equivalent nodes whose nodal-statistics-distribution are close enough given a distribution metric, as follows:

Definition A.3 (Structural Equivalence with nodal statistics s distribution). Given a distance d among distributions, we say two nodes of a fuzzy network u, v to be structurally equivalent $u \sim_s v$ if and only if

$$u \sim_s v \iff d(s(v), s(u)) \leq \epsilon \quad (1.12)$$

where $s(v)$ indicates the distribution of the statistics s for node v . The choice of d requires to properly evaluate which type of relation we would like to capture.

All the definitions proposed in chapter 2, follows directly from Def. A.3. This allows to take into account an uncertainty of nodal statistics, and results in the definition of a similar way to compare graph structural patterns and characterize collection of fuzzy networks.

A.4 Complete Network Surgery Results

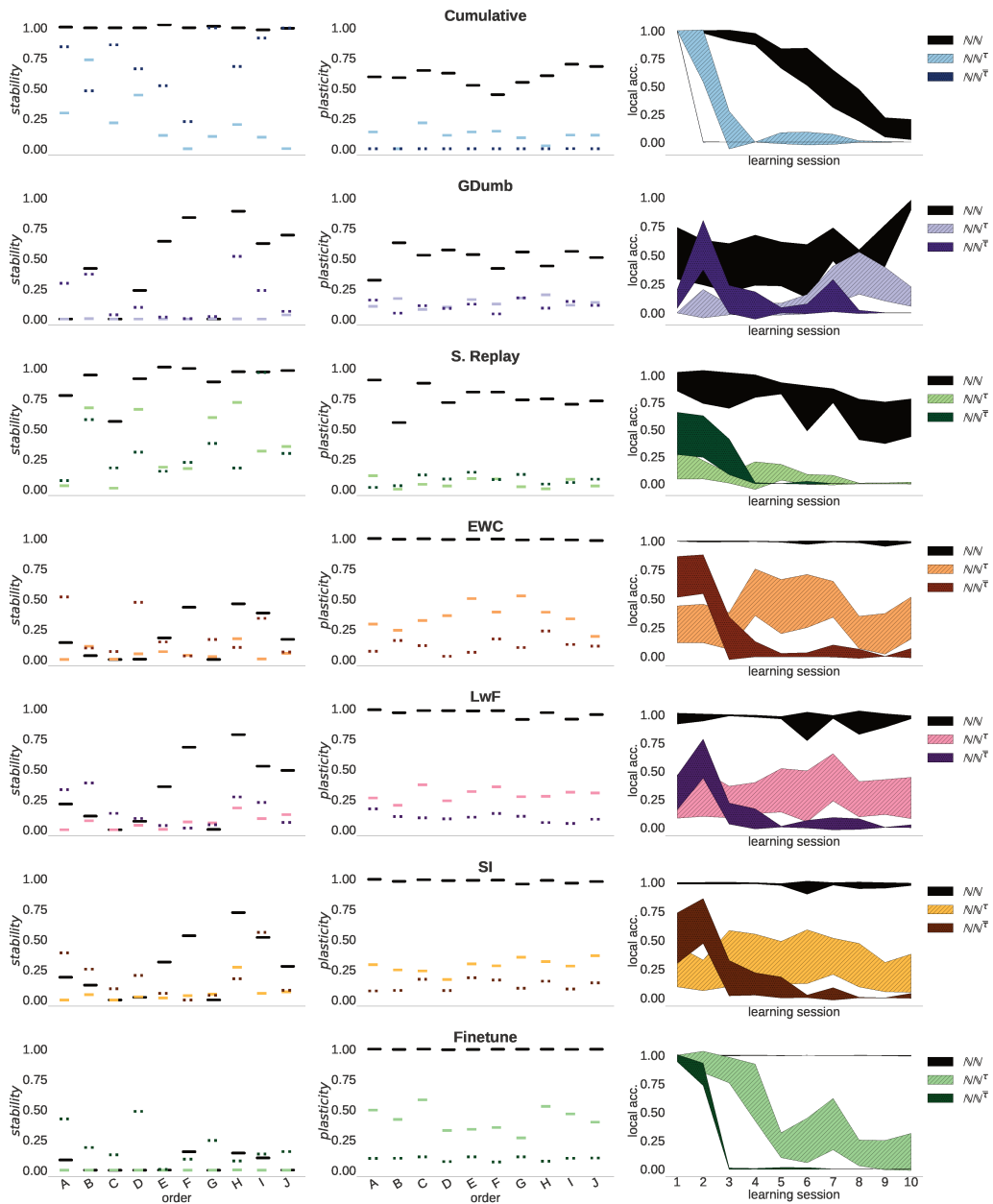


Figure A.7 – Network surgery results for all the considered orders and strategies on MNIST task.

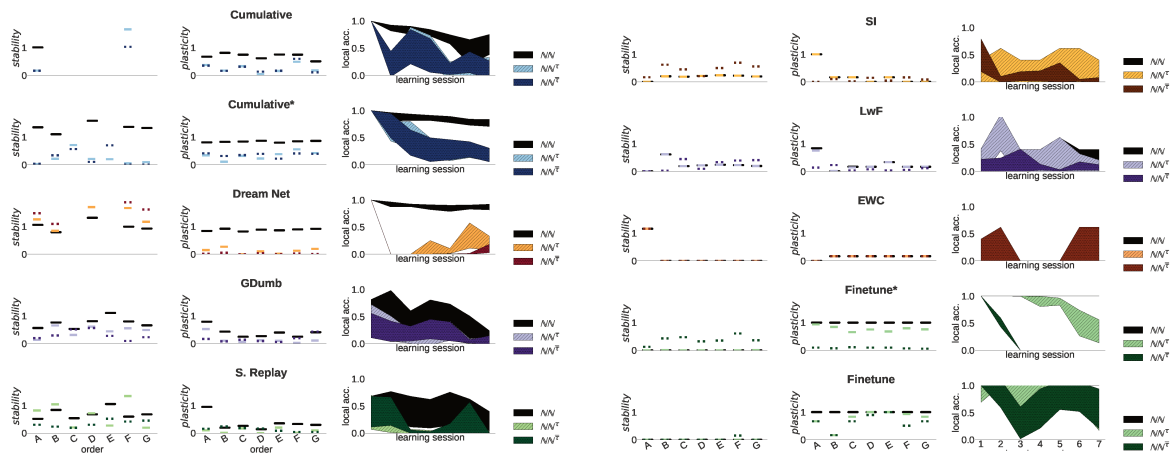


Figure A.8 – Results of Experiment 2 (Network Surgery) on all strategies and orders on Face Emotion Recognition Task. *Stability* (Left) and *Plasticity* (Center) performances on the configurations on face emotion recognition task in their pruned versions. For all different orders (see Tab. 6.4) of learning sessions, we report the performance of the standard model NN in plain black line. The change in the performance when the model is pruned is reported for the two different pruning versions and colored. (Right): average local accuracy performances at different learning sessions. For the standard NN and the pruned models NN^T , $NN^{\bar{}}$ the accuracy for the last learned class across learning sessions is reported with its confidence interval for all studied configurations.

B

B.1 Network Embedding for Brain Connectivity

NETWORK EMBEDDING FOR BRAIN CONNECTIVITY

Lucrezia Carboni^{1,2} *Sophie Achard*¹ *Michel Dojat*²

¹ Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France

² Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN, 38000 Grenoble, France

ABSTRACT

In Neurosciences, networks are currently used for representing the brain connections system with the purpose of determining the specific characteristics of the brain itself. However, discriminating between a healthy human brain network and a pathological one using common network descriptors could be misleading. For this reason, we explored network embedding techniques with the purpose of brain connectivity networks comparison. We proposed first the definition of representative graph for healthy brain connectivity. Then, two classification procedures through embedding are introduced, achieving good accuracy results in different datasets. Moreover, the intriguing power of this technique is given by the possibility of visualizing networks in a low-dimensional space, facilitating the interpretation of the differences between networks under diverse conditions e.g. normal or pathological.

Index Terms— Network classification, MRI, low dimensional space, graph comparison.

1. INTRODUCTION

In various contexts, network embedding technique has been developed for performing analysis on a single graph. This technique has been proven to work for different applications, such as node classification or link prediction. An efficient network embedding algorithm is capable of capturing the relevant features of the graph and reproduce them in a low-dimensional Euclidean space.

In the context of Neurosciences, networks are used to represent the set of connections between brain regions with eventually the aim of distinguish pathological versus healthy states. However, such a distinction could be missed based on the network descriptors commonly used. Indeed there is no clear evidence of a best measure to be used in the discrimination of different brain states [1]. As a results each process focuses on specific descriptors which differ from pathology to pathology or requires hand-crafted index to reveal a dissimilarity (i.e. [2]). In this study, we investigate the use of network embedding with the goal of applying to human brain connectivity network classification. Moreover, the intriguing power of the embedding in our domain is given by the possibility to visualize the connectivity network in a low-dimensional space,

which allows to determine which brain regions and properties are peculiar of the given network and are fundamentals for the normal vs abnormal classification process, providing interestingly new knowledge about the pathological conditions.

2. MATERIAL AND METHODS

2.1. Data: brain connectivity networks definition

The definition of brain connectivity networks was achieved through out different phases proposed in [3]. First, the functional MR images (fMRI) acquired at rest were aggregated over 90 regions defined according to the anatomical labeling in [4]. For each parcel, a unique time series signal was determined by averaging the fMRI time series over all voxels of the parcel, weighted by the proportion of gray matter in each voxel, obtained by individual structural image tissue segmentation, to take into account for the partial volume effect. The following stage consisted in the application of the discrete wavelet transform to each mean fMRI time series. As a result, for each subject, different fMRI time series at distinct scales, associated to different intervals of frequencies, were at our disposal. Since it has been observed that the resting state information activity is mainly captured at a frequency lower than 0.1Hz, the correlation among regions was estimated only for this frequency band. Finally, by thresholding the correlation matrix, a binary matrix was defined. This matrix corresponded to the adjacency matrix of the brain connectivity network, whose nodes were given by the parcels. The threshold was tuned for each subject in order to compute a binary matrix with a given non-zeros entries number; in our experiments it has been set to 400. Table 1 indicates the resting state datasets we used in our experiments extracted from different databases.

Dataset name	Number of Data		
	Total	Class 1	Class 2
HCP test retest [5]	100×2	-	-
Coma [2]	37	20 Control	17 Comatose
Young&Elderly [6]	26	15 Young	11 Elderly

Table 1. List of the databases used including different classes of subjects. HCP: Human Connectome Project.

2.2. Nodal embedding

Network embedding can be seen as a dimensionality reduction tool which maps a network into a vector space. We focussed on nodal embedding, namely a mapping function which maps a graph into a bag-of-vector where each vector is associated to a single vertex of the graph.

Among the available methods, we selected `node2vec` embedding algorithm [7] which has been proven of being able to capture the structural equivalence of nodes. `node2vec` is based on the Skip-gram architecture which is used to learn a features representation of words based on their context [8]. In the case of network, the concept of context is translated into the one of *neighborhood*. Precisely, `node2vec` defines a flexible notion of a node's neighborhood depending on specific characteristics we are interested in, such as those obtained by structural relations or by similar relationships between the neighborhoods. Note that neighborhood is not defined based on a unique similarity function, but can be defined via two searching strategy. For each vertex, the algorithm computes a neighborhood set of a given number of nodes. The breadth-first sampling (BFS) considers the neighborhood of a node as nodes which are immediate neighbors of the source. Whereas, the DFS neighborhood is composed by nodes which are sequentially sampled at increasing distances from the node itself. `node2vec` allows to smoothly interpolate between BFS and DFS.

In synthesis, the `node2vec` embedding function is determined by the following parameters: d dimension of the embedding space, N number of random walks per node used to estimate the proximity matrix, L random walk length, k size of the neighborhood set for each node, p return parameter and q in-out parameter controlling the sampling strategy (small $p \rightarrow$ BFS, small $q \rightarrow$ DFS).

Setting the parameters' values referring to a single graph's embedding would result in the necessity of using different parameters for each single graph. However, in this way we loose the possibility of a fair graphs comparison. Because our purpose is to capture the class features for different graphs, we propose to set these parameters according to the results obtained for a graph comparison for which we know the ground truth. To clarify, we would like to perform a graph embedding for which the computed similarity index for the same class is maximized, while the similarity index between different classes is minimized. The obtained embedding, thereby, would eventually be able to capture the shared characteristic of a group of graphs. Since all the graphs in our datasets are comparable in terms of the number of nodes and edges, we chose to tune the parameters based on the HCP dataset. This dataset provides test and retest data for the same subjects. Thus, using this unique dataset for the tuning, we expected to maximize the similarity index between graphs of the same subject, forcing the embedding to capture the characteristics which were relevant in brain connectivity. Indeed, a parameters configuration which estimates similar embedding

for similar graphs, was assumed to capture the intrinsic nature of the corresponding brain connectivity network.

The final parameters' values were $d = 3$, $N = 20$, $L = 2 \times$ network diameter, $k = 3$, $p = 1$, $q = 2$.

2.3. Pyramid graph matching kernel

In our project, we are interested in comparing networks which belong to different classes, healthy vs pathological subjects or for patients sub-typing. This comparison could be directly performed in the network space; however, our purpose was to explore the potentiality of embedding in the analysis of brain connectivity. With this goal, we found the work [9] newsworthy. In particular, inspired by the pyramid match kernel used in computer vision, the authors have designed an equivalent version for graphs, named pyramid matching graph kernel. The main idea of the algorithm was to count the matching of vectors in the embedding space at different resolution levels. We reproduced their setting using `node2vec` as embedding algorithm. In the embedded space, given a number of level, a grid of cells having increasing size in one dimension for each level was computed. Two vectors corresponding to the same region in the brain were matched if they belonged to the same cell. Each matching was weighted according to the dimension of the corresponding region. Counting the number of matchings between two embedded graphs allowed to compute their similarity.

2.4. Representative graph

Previous works have defined, for a class, an average graph in the network space, taking into account the average of the adjacency matrices or selecting edges present in all the graphs of the class. We propose an innovative approach using a network embedding algorithm. It requires labeled graphs with labeled nodes. Note that all graphs have the same number of nodes. Having at our disposal a labeled set of graphs, we considered all the networks which belong to the same class. For each node, we considered the set of the embedding vectors obtained through the mapping of all graphs of the class. Then, we averaged components by components and computed the barycenter. The output of this process was a set of vectors, where each vector was associated to a given node in the graph. This output can be interpreted as a virtual embedding of a representative graph. In fact, we were implicitly assuming that the vector position in the embedding space should be related to the label associated to the node. Moreover, since the embedding was supposed to capture the structural equivalence of nodes, we hypothesized that for individuals belonging to the same class, same brain regions exhibited the same connectivity pattern.

2.5. Classification procedure

We propose to compare two classification procedures through embedding, standard SVM vs a graph representation of a class

	C1 - MA	C1 - SD	C2 - MA	C2 - SD
COMA	0.83	0.019	0.75	0.025
Y&E	0.59	0.045	0.62	0.051

Table 2. Classification performances over all the datasets through the two different classification procedures. MA - mean accuracy and SD - standard deviation computed on 100 experiments.

(see Fig.1). For both procedures, we computed the embedding of all networks using `node2vec`. As a result, each graph was converted into a set of vectors. We randomly determined a subset of the dataset as the training set. The remaining data formed the candidates for classification (testing set). The training set was partitioned into two classes.

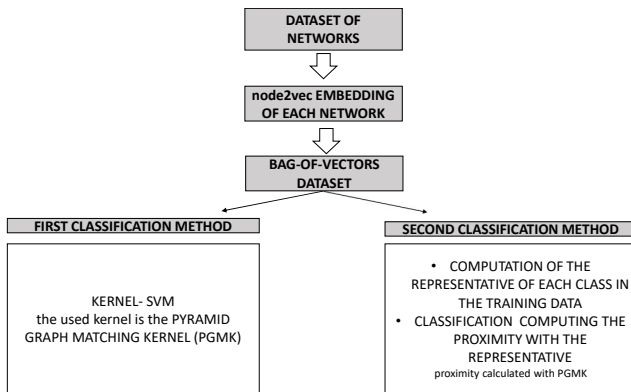


Fig. 1. Classification procedure scheme

2.5.1. First classification procedure: C1

We calculated the gram matrix using the pyramid graph matching kernel. Then, we used kernel-SVM to predict the labels on the candidates in the testing set. The kernel used was the pyramid graph matching kernel.

2.5.2. Second classification procedure: C2

This procedure used the average graph as a representative for each class for the comparison. The average embedding graph was computed for each class. To classify the testing set, we computed the pyramid graph matching similarity index between each candidate and the two graph representatives of the classes. We predicted the graph label, according to the more similar graph.

3. EXPERIMENTS AND RESULTS

First, we compared the representative networks obtained from the different groups of controls coming from the different datasets. The first three graphics in Fig. 2 is illustrating the robustness of the construction of the representative graphs obtained from the three different datasets acquired in different

conditions, but preprocessed using the same methods. Second, we ran the two classification procedures 100 times for the datasets. Each time the embedding and the training sets were computed in a repeated random sub-sampling validation framework. We observed a slightly variability in terms of accuracy depending on the datasets.

4. DISCUSSION

Network representation is largely used in Neurosciences for structural or functional brain connectivity studies. In this paper we explored the applicability of the network embedding approach for brain connectivity investigation. We calibrated the embedding method parameters based on one dataset considered as the *ground truth*. It is important to mention that this was not done to improve performances but to favor a certain pattern on the data. Besides, the dataset used for calibration was indeed very different in term of acquisition conditions (different centers, different MR scanners) than the other datasets we considered then for classification performances testing.

Our preliminary results demonstrate, using datasets coming from different databases, that a network embedding approach, combined with standard classification methods, provides good performances for separating two different families of graphs representative of pathological states (brain trauma, coma) versus healthy conditions or normal aging connectivity changes (young vs elderly). The performances were very similar for all the datasets, representative of very different clinical situations, with the classification method that used a graph representative of a class of subjects versus the more standard SVM method. This demonstrates that our definition of the graph representative of a class was valid. However, we notice that SVM provided the best performances for the coma dataset being able of capturing the dissimilarity between the two classes without any prior selection of graph descriptors neither requiring the definition of a specific measure [2]. The performances remained appreciable for Young&Elderly.

Our classification scores are good, but not excellent. However, we consider that our procedure should not be only evaluated on its classification results. Indeed, the major interest of network embedding is to facilitate the interpretation of the differences in a low-dimensional space between networks (see Fig. 2). This is highly valuable in Neurosciences where networks may be representative of brain connectivity under different conditions e.g. normal or abnormal. Thus, even whether previous works may report better classification performances, we assume that a network embedding procedure provides an additional value: the possibility of determining a signature for the class.

The use of global descriptors for graph characterization has been proven not to be enough for detecting some differences between graphs [2]. Our achievement shows the power of our embedding method for graphs comparison in capturing the meaningful features of the network. Even if further analysis

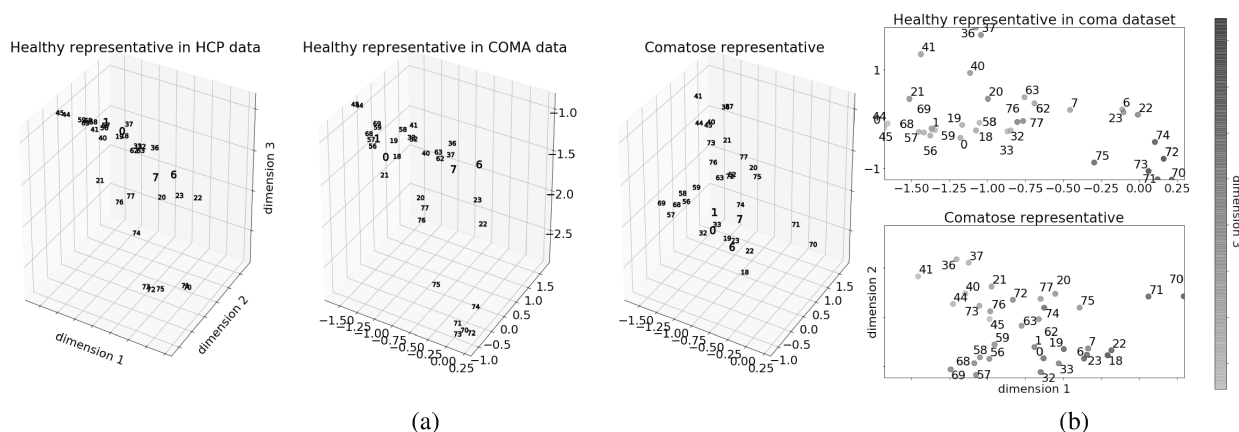


Fig. 2. (a) Visualization in a 3D space of a subset of regions in different representative embeddings. Qualitatively, we can appreciate the closeness of the healthy representatives extracted from different datasets. While, major differences could be visualized with respect to the right picture representing the comatose class. Numbers refer to the region labels in [4]. (b) Projection over the first two dimensions of the embedded vectors in the healthy control (*top*) and comatose (*bottom*) representatives in COMA data. Points' marker are colored according to the values of the third coordinate.

needs to be conducted, we interpret the good performances obtained as an indicator of the type of graph property preserved in the embedding.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available by the HCP open source. Ethical approval was not required as confirmed by the license attached with the open data. The use of the other datasets was approved by the ethical committee consulted for each study and all the concerned participants or next of skin signed the informed consent form.

6. ACKNOWLEDGEMENTS

LC was supported by MIAI@Grenoble Alpes (ANR 19-P3IA-003). The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

- [1] Michael N Hallquist and Frank G Hillary, "Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world," *Network Neuroscience*, vol. 3, no. 1, pp. 1–26, 2018.
- [2] Sophie Achard, Chantal Delon-Martin, Petra E Vértes, Félix Renard, Maleka Schenck, Francis Schneider, Christian Heinrich, Stéphane Kremer, and Edward T Bullmore, "Hubs of brain functional networks are radically reorganized in comatose patients," *Proceedings of the National Academy of Sciences*, vol. 109, no. 50, pp. 20608–20613, 2012.
- [3] Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and ED Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *Journal of Neuroscience*, vol. 26, no. 1, pp. 63–72, 2006.
- [4] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [5] Maïté Termenon, Assia Jaillard, Chantal Delon-Martin, and Sophie Achard, "Reliability of graph analysis of resting state fmri using test-retest dataset from the human connectome project," *Neuroimage*, vol. 142, pp. 172–187, 2016.
- [6] Sophie Achard and Ed Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Comput Biol*, vol. 3, no. 2, pp. e17, 2007.
- [7] Aditya Grover and Jure Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.
- [8] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [9] Giannis Nikolentzos, Polykarpos Meladianos, and Michalis Vazirgiannis, "Matching node embeddings for graph similarity," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

B.2 Multiscale and multi-density comparison of functional brain networks through label-informed persistence diagrams

Multiscale and multi-density comparison of functional brain networks through label-informed persistence diagrams

Paul Sitoleux¹, Lucrezia Carboni^{1,2}, Hanâ Lbath¹ and Sophie Achard¹

¹ University Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France

² Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN

E-mail: sophie.achard@univ-grenoble-alpes.fr

11 August 2023

Abstract. *Objective.* Functional magnetic resonance imaging (fMRI) allows the construction of functional brain networks, offering a tool to probe the organization of neural activity. The objective of this study is to investigate persistent homology in such networks. Specifically, we evaluate the importance of node label information and the necessity for multiscale approaches. *Approach.* In this work, persistent homology (PH) is first applied to a density-based filtration. In addition, we propose a procedure to extract label information from persistent homology summaries of labeled graphs. We then investigate its ability to discriminate between real data and surrogate data generated from null models, as well as between healthy subjects and comatose patients. *Main results.* We first show our proposed label-informed approach outperforms standard PH-based methods in the aforementioned discrimination tasks. Furthermore, the performance of our proposed procedure is comparable to that of standard matrix distance or thresholding-based methods, which intrinsically account for node label information. *Significance.* Understanding network organization is critical to the identification of brain dysfunctions related to pathological conditions, as well as probing the influence of age difference, states of consciousness, or neurological disorders, with eventual applications to diagnosis and long-term prognosis. The multiscale nature of functional networks provides a significant challenge for their analysis, which is frequently done by studying binary graphs resulting from the application of an arbitrary threshold. PH offers a framework for bypassing this issue. We outline in this work some of the limitations of PH and show the importance of multiscale and multi-density approaches.

Keywords: : resting-state functional magnetic resonance imaging (rs-fMRI), brain connectivity networks, generative models of networks, graph distance, persistent homology

Submitted to: *J. Neural Eng.*

1. Introduction

In recent years, efforts to understand the human brain have been made by modeling it as a complex network [1]. Advances in magnetic resonance imaging technology have allowed to introduce structural and functional networks [2]. They have mainly been characterized through the various tools and concepts from graph theory and network sciences, such as small-worldness [3], degree distribution, rich-club coefficient, etc.

A widely used approach to define functional brain networks requires parcellating the rs-fMRI signal with a predefined anatomical atlas, giving us the N nodes of the graph [4]. For each node, the rs-fMRI time series is estimated as the average of voxel activity in the corresponding parcellated region. Taking a similarity measure for each pair of nodes yields a dense adjacency matrix, to which one generally applies a threshold, in order to obtain a sparse binary graph [5, 6, 7, 3]. An important issue with this approach is the choice of a threshold. In the literature, this choice is based on correlation distribution properties, the average degree, connectivity [8], graph regime, multiple testing, or to optimize a metric such as discrimination accuracy, etc. To avoid this, analyses can be conducted at multiple density levels [9]. However, there is a lack of attention to multi-density approaches exploring and summarizing information in functional brain networks.

Moreover, purely graph-based methods are often limited to global or local statistics, only presenting information at the scale of the whole graph or of a single node neighborhood [10]. Persistent homology is a topological data analysis (TDA) approach that produces multiscale summaries from a point cloud or a distance matrix. Persistent homology tracks homological features, in short, the number of connected components and the number of topological cavities (equivalent to a circle or a hollow sphere in dimensions 1 and 2, or to higher dimensional cavities), by building simplices. This is a way to look at structures in the graph that are orders higher than the dyadic relationships the edges of the graph represent [11]. The interest in higher-order interactions has taken off in recent years, whether in the broader context of complex systems modeling [12, 13] or in structural [14] or functional brain networks [15, 16, 17], with notable findings in aging and neurodegenerative disorders. Another argument that has been advanced in favor of persistent homology is that it is easily applicable to biological data, which is often highly dimensional and lacks a natural concept of distance [18]. For instance, in neuroscience network data it has been applied to differentiate healthy subjects and subjects with neural disorders [19, 20, 21], study the influence of psychotropic substances and sedatives [22, 23] or to analyze neuronal network simulations [24, 25].

One of the main goals of functional connectivity network studies is detecting and recognizing brain dysfunctions related to pathological conditions. Indeed, functional networks offer a unique way of extracting new noninvasive biomarkers [26] that provide a powerful understanding of physio-pathological mechanisms. Recognizing a given pathology by analysis of the functional connectivity requires graph comparison distances,

similarities functions, or statistical tools (efficiency, small-worldness, global or nodal statistics [27], spectral graph analysis, edit distances...) [28]. For this purpose, it is imperative to be able to precisely quantify whole-brain variations that can arise in different cohorts, probing the influence of age differences, states of consciousness, or neurological disorders. To show the benefit of our proposed framework in real data applications, we consider functional networks of comatose patients. Discriminating between subjects within different states of consciousness using functional networks is a difficult task: previous work did not detect significant modifications in graph metrics between comatose patients and healthy controls, but what appears to be a reorganization of network hubs [29]. Here, we propose a new distance among brain networks that combines homological features and regional information.

We are particularly interested in applications of TDA in functional brain networks. To that end, we consider null models which generate surrogate correlation matrices as benchmarks of real data. This permits the identification of relevant features captured by a persistent homology approach in brain connectivity. Null models are ubiquitous in network neuroscience [30]. For instance, they allow one to test the statistical significance of graph features of empirical networks against a null hypothesis. Furthermore, null models which can be tuned to reproduce a specific graph layout offer a tool for comparing graph distances and understanding which properties they capture.

In this work, we present applications of persistent homology that allow insights into functional brain connectivity networks without requiring an arbitrary threshold choice to obtain binary graphs. Persistent homology is newly applied to a density-adjacency matrix to probe its ability to discriminate between real data and surrogate data generated from null models, as well as between a set of healthy subjects and comatose patients. However, while persistent homology summaries are multiscale, they forego critical information: the identity of nodes and edges. Indeed, these summaries, such as the persistence diagram or Betti curves, do not depend on the labeling of nodes. We introduce distances taking into account the node labeling, and explore them in the above-mentioned tasks.

This paper explores four main directions: (i) a novel density-based filtration, (ii) label-informed distance on persistence diagrams, useful in non-permutation invariant settings, (iii) application to null-models and comatose-control discrimination, and (iv) gain insights on the construction of null models from real brain data.

2. Materials and methods

2.1. Data

We consider a set of resting-state functional networks from the Human Connectome Project (HCP) [31] to investigate the general properties of topological features of brain graphs. To investigate how significant these features are, attempts will be made to cluster a second dataset of 20 healthy controls and 17 patients in a comatose state [29] (8

patients were dropped due to excessive head motion). The patients were between 21 and 82 years old, had an initial Wessex Head Injury Matrix (WHIM) score between 1 and 37 (on a scale that goes from 0, meaning deep coma, to 62, meaning full recovery), and were scanned between 3 and 32 days after the acute medical event. Twelve of them fell into a coma following a cardiac and respiratory arrest, two after going into hypoglycemia, two after a gaseous embolism and one after an extracranial arterial dissection. Six months following their fall into a coma, 9 patients had died, 5 remained in a vegetative state and 3 made a recovery.

Data were corrected for head motion and co-registered with the subject’s structural MRI images. This allowed mapping fMRI images to a customized parcellation template. The data were not spatially smoothed and, unless specified, did not undergo any cerebrospinal fluid, white matter, or gray matter regression. The process needed to go from raw blood-oxygen-level-dependent (BOLD) image sequences to a meaningful functional connectivity graph comprises multiple steps, each with its set of assumptions and consequences on the final network. For example, in the case of global signal regression, that is, regressing over average of the whole image, it has been established that one has to make a choice between discarding some neural signals and keeping non-neural nuisances [32].

2.2. Functional connectivity network construction

The measurements were parcellated into $N = 90$ regions using the Automated Anatomical labeling (AAL) template [33]. For each region, the time series were estimated by averaging over all voxels, weighted by the proportion of gray matter in each voxel (estimated through structural MRI) and corrected for head motion. Correlations were estimated between Daubechie’s wavelets coefficient between all pairs of time series. Analysis was restrained to the frequency interval 0.04-0.08 Hz. See [29] for additional details.

2.3. Functional connectivity network representation

A graph $G = (V, E)$ is a collection of nodes or vertices V and edges linking those vertices, i.e., each element in E is an element of $V \times V$. There is a range of ways to represent a graph, here it will be represented by its adjacency matrix, usually denoted A . An adjacency matrix is a symmetric matrix where each off-diagonal element a_{ij} is a weight, representing a chosen attribute of the relation between nodes i and j .

Using the correlation matrix between the regional wavelet coefficients C , we construct the graph adjacency matrix A :

$$A = 1 - |C|. \quad (1)$$

To neutralize the effect of inter-subject fluctuations in the values of the correlations, an alternative matrix can be defined by considering the density-based adjacency matrix \tilde{A} , computed using the level of edge-appearance as follows:

$$\tilde{a}_{ij} = \frac{2u_{ij}}{N(N-1)}, \quad i \neq j, \quad (2)$$

where \tilde{a}_{ij} is the normalized index of the u_{ij}^{th} off-diagonal element of matrix A , ordered according to the absolute value of the correlations. In this new matrix, the order between each off-diagonal element is respected, but these new values are spaced at a constant interval $\frac{2}{N(N-1)}$.

2.4. Null models for functional connectivity

Null models are an ubiquitous tool in network neuroscience [30]. By offering a way to generate networks according to a simplified model, they allow benchmarking of the properties of empirical networks against the null hypothesis provided by a given model.

Comparing null models and empirical functional connectivity allows for the appreciation of null model properties and improved understanding of the information captured by various distances. Here, four null models for correlation networks are presented, each conserving some features from an initial functional connectivity matrix [30]. First, we consider the Zalesky matching algorithm, which generates correlation matrices with a given average correlation and variance between correlations [34]. This is achieved through generating an $N + 1$ normally distributed random vectors of length T $x_i, y \sim \mathcal{N}(0, \mathbf{I})$ ($i \in \{1, \dots, N\}$), with \mathbf{I} the T -dimensional identity matrix. Then, the values of x_i are repeatedly adjusted as $x_i \leftarrow x_i + ay$ until the desired average correlation between the vectors is obtained. The process is repeated with a different number of time points T until the correlation variances also match.

As a second model, we consider a spatiotemporal one which generates time series imitating spatial and temporal autocorrelation from the regional time series of a given examination of a subject [35]. Here, the model is applied to the regional wavelet coefficients.

The last two null models are a phase randomization model, generating new time series where each regional time series has the same power spectrum, and an Erdős-Rényi model, where each correlation value is randomly distributed.

2.5. Persistent homology

Persistent homology is a mathematical formalism from the larger field of topological data analysis, that allows the production of multiscale summaries of point cloud or graph distance matrices [36], such as Betti numbers and persistent diagrams.

Starting from a distance matrix, simplicial complexes are built, using a rule denominated *filtration*, at each possible scale in the matrix. At each step, the appearance and disappearance of topological features are tracked. These features correspond to the dimension of p^{th} homology group [37], with p a non-negative integer. Essentially, the number of 0-dimensional features is the number of connected components, and higher

dimensional ones represent *topological cavities*: 1-dimensional features are circle-like, 2-dimensional features are hollow sphere-like, and so on.

For a set of $k + 1$ vertices $\{v_0, v_1, \dots, v_k\}$, the k -dimensional abstract simplex or k -simplex is the set containing all nonempty subsets of vertices. A simplex σ^k is a *face* of a simplex σ^n , $k < n$ means that each vertex of σ^k is a vertex of σ^n . For example, 0-simplex is a single point, a 1-simplex is an edge linking two points, a 2-simplex is a triangle with its three edges and three points as 1 and 0-faces, a 3-simplex is a tetrahedron, etc.

A *simplicial complex* K is a finite set of properly joined simplicies, i.e., each intersection of two simplicies is either a face of both or empty, and where each face of a member of K is also a member of K . The *dimension* of K is the largest positive integer r such that K has an r -simplex. Figure 1 presents a 3-dimensional simplicial complex

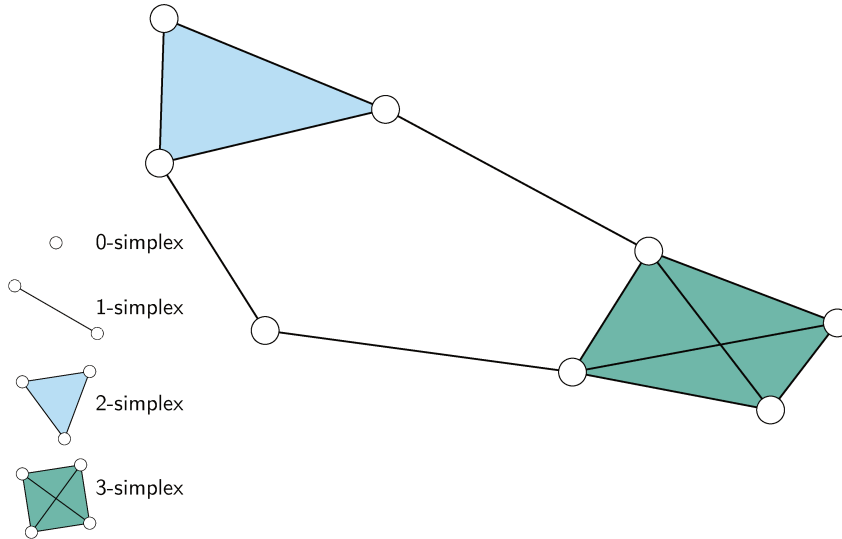


Figure 1: A simplicial complex of dimension 3.

For a given simplicial complex, the number of k -dimensional *holes* or *features* is called the k^{th} *Betti number*, denoted β_k . Formally, β_k is the dimension of the k^{th} homology group H_k , which is itself the quotient of the k^{th} cycle group Z_k with the k^{th} boundary group of the given simplicial complex:

$$\beta_k = \dim(H_k) = \dim\left(\frac{Z_k}{B_k}\right). \quad (3)$$

In short, this means that β_k counts the number of $k + 1$ -dimensional volumes that are enclosed by, at least, a k -dimensional cycle that does not correspond to the boundary of a simplex of the given simplicial complex, with β_0 counting the number of connected components of the simplicial complex. Therefore, in Figure 1 the simplicial complex has $\beta_0 = 1$, since it has a unique connected component, $\beta_1 = 1$ with the non-filled central cycle 1 and $\beta_k = 0$ for $k \geq 2$.

In persistent homology, there is some freedom in how simplices are built. For its lower computational cost relative to the other options, in this work, and in most persistent homology applications, the *Vietoris-Rips* filtration is used. In this setting, a simplex is in the relevant simplicial complex if all its 1-faces or edges are. Edges are successively added to the simplicial complex when the corresponding value in the adjacency matrix is equal to or lower than the given filtration parameter. In this work, instead of using correlation values, the homology groups are computed at each scale of the density-based adjacency matrix.

A way to present this topological information is to plot the *Betti curves* of the Betti numbers against the filtration values, which can be either inter-correlations or density levels. Another way is to plot the *death* values of features against their *birth* values. This yields a *persistence diagram* in the eighth plane delimited by the $b = d$ and $d = 0$ lines.

Let us consider the following example. The first row of Figure 2 presents simplicial complexes with $\beta_0 = 1$ and $\beta_1 = 0$. In all those 3 cases, there exists at least one cycle, but each corresponds to the boundary of a simplex. For the second row, each simplicial complex has $\beta_1 = 1$. The first two have a unique representative: ABCD and ABCDEFGHIJ, while the last one has multiples, with two of them disjoint: ACEGI and BDFHJ. This illustrates both the robustness of persistent homology with regard to the general structure and the limitations of its discriminative power.

From the examples given here, it also seems that $\beta_1 > 0$ is a signature of a relatively low clustering coefficient (i.e., the proportion of possible triangles). This shows how these features could be significant. Taking the analogy of friendship networks, it means that there is a cycle of length 4 with some gap with A and C both being friends with B and D, but neither A and C nor B and D are. This might indicate either a pure coincidence or some hidden organizational principle.

2.6. Comparing persistence diagrams

We first recall various distances between persistence diagrams.

The p -Wasserstein distance between measures $\mu, \nu, p \in [1, \infty[$ is

$$W_p(\mu, \nu) = \left(\int_{\mathcal{X} \times \mathcal{X}} c(x, y)^p d\pi(x, y) \right)^{1/p} \quad (4)$$

where $\pi(x, y)$ is the optimal coupling between μ and ν under cost $c(\cdot, \cdot)^p$. In these works, the focus is mainly on the 1-Wasserstein distance.

Another common option is the *bottleneck distance*, which is the $p \rightarrow \infty$ limit of the p -Wasserstein distance:

$$\text{Bo}(\mu, \nu) = \lim_{p \rightarrow +\infty} W_p(\mu, \nu). \quad (5)$$

For a, b two discrete distributions, the *optimal transport distance* can be expressed as

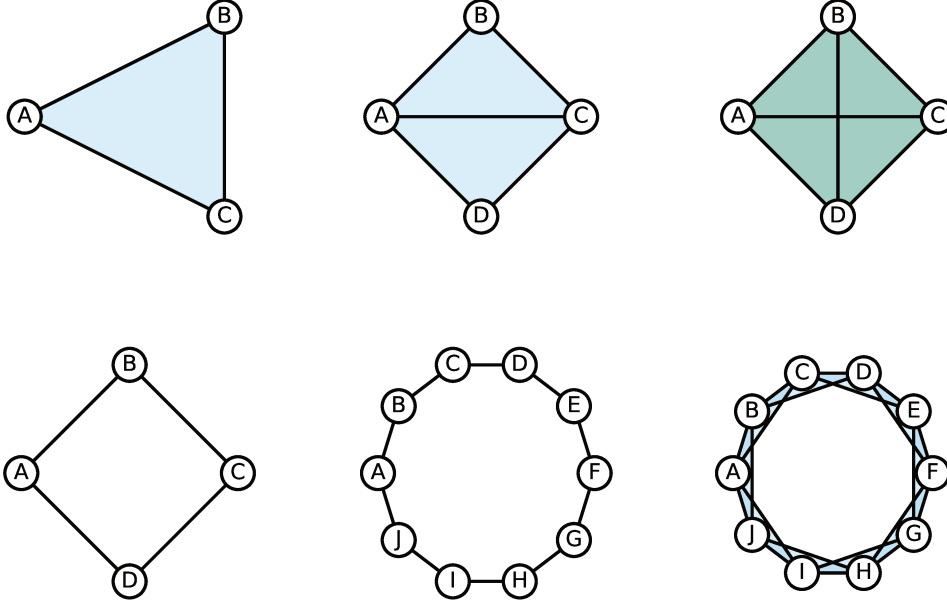


Figure 2: Examples of graphs with different Betti numbers. On the top row, they have $\beta_1 = 0$ while those on the bottom have $\beta_1 = 1$.

$$d_{\text{OT}}(a, b) = \gamma_{ab}^* = \operatorname{argmin}_{\gamma \in \mathbb{R}^{m \times n}} \sum_{i,j} \gamma_{ij} M_{ij} \quad (6)$$

such that $\gamma \mathbf{I} = a, \gamma^T \mathbf{I} = b, \gamma \geq 0$ and with M the cost matrix defining the cost of moving the histogram bin a_i to bin b_j , \mathbf{I} the identity matrix of the relevant dimension.

We now introduce a label-informed distance on persistence diagrams, by designing a novel cost matrix. First, we list edges $\{\{e_k[i]\}_{i \in \{1, \dots, n_k\}}\}$ which correspond to the birth or death of the k^{th} homological feature in the persistence diagrams, where n_k is the number features. Next we choose a simple binary cost matrix defined as follows:

$$M_{ij}^* = \begin{cases} 0 & \text{if } e[i] = e[j] \\ 1 & \text{otherwise} \end{cases} \quad (7)$$

We define the *edge comparison optimal transport* distance as the optimal transport distance d_{OT} where $M = M^*$:

$$d_{e, k \dots}(a, b) = \operatorname{argmin}_{\gamma \in \mathbb{R}^{m \times n}} \sum_{i,j} \gamma_{ij} M_{ij}^*, \quad (8)$$

where $k \dots$ denotes the included homological features. It should be noted that this is an approach that is similar to the fused Gromov-Wasserstein distance [38, 39], applied to a labeled persistence diagram instead of a labeled graph. Fused Gromov-Wasserstein

Multiscale comparison of functional brain networks

9

takes into account both labels and graph structure by taking a weighted sum of two terms representing each aspect. The label term is equivalent to the distance introduced here for a general cost matrix M .

In the following, the persistence diagrams and the Wasserstein distances between them are computed using `giotto-tda`, with an error tolerance of 0.01 [40]. The optimal transport distance d_{OT} between the persistence diagram is computed with the POT package [41].

2.7. Comparing edges of labeled graphs

A simple distance between labeled graphs would be to look at the *Frobenius* norm of the difference of adjacency matrices P, Q .

$$d_F(P, Q) = \frac{1}{2} \frac{\|P - Q\|_F}{\|P\|_F + \|Q\|_F} \quad (9)$$

with $\|P\|_F = \left[\sum_{i,j} p_{ij}^2 \right]^{1/2}$ the Frobenius norm of matrix P (with a normalization factor to facilitate its interpretation).

In the context of thresholded graphs, node labels can be used to compare graphs. For \mathcal{A}, \mathcal{B} two sets, the *overlap* similarity is

$$\text{Overlap}(\mathcal{A}, \mathcal{B}) = \frac{|\mathcal{A} \cap \mathcal{B}|}{\min(|\mathcal{A}|, |\mathcal{B}|)} \quad (10)$$

and the overlap distance

$$d_O(\mathcal{A}, \mathcal{B}) = 1 - \text{Overlap}(\mathcal{A}, \mathcal{B}). \quad (11)$$

In our setting, \mathcal{A}, \mathcal{B} correspond to two sets of node label pairs, associated to edges in the thresholded graph.

2.8. Quantitative group comparison

In order to have a more quantitative understanding of the various distances presented here we evaluate them in a group comparison task. The distance matrix structure makes it complicated to find groups if one only considers pairwise distances. To counteract this issue, the distance matrix will be treated as if it were a matrix of observations and fed to clustering algorithms. While debatable, this approach allows exploring this kind of group structure using simple and proven clustering algorithms: KMeans, hierarchical clustering with complete linkage, and spectral clustering on the nearest-neighbors graph [42].

The performance of the clustering algorithms will be evaluated using the *adjusted mutual information* (AMI):

$$\text{AMI}(X; Y) = \frac{I(X; Y) - \mathbb{E}[I(X; Y)]}{\max(H(X), H(Y)) - \mathbb{E}[I(X; Y)]} \quad (12)$$

where $H(X)$ is the Shannon entropy of the random variable X and $I(X;Y)$ the mutual information between random variables X and Y . The AMI takes values 1 for a perfect clustering and has an expected value of 0 for a random assignment of labels [43].

3. Results

3.1. Density-based adjacency matrix

In this section, we first compare the correlation- and density-based adjacency matrices on real-world data. The plot of the inter-subject Frobenius distance for the coma cohort is depicted in Figure 3. In the upper diagonal, which corresponds to the density-based matrices, a clear separation between the comatose and healthy subjects is apparent. However, the two groups cannot be distinguished when using the correlation-based adjacency matrices.

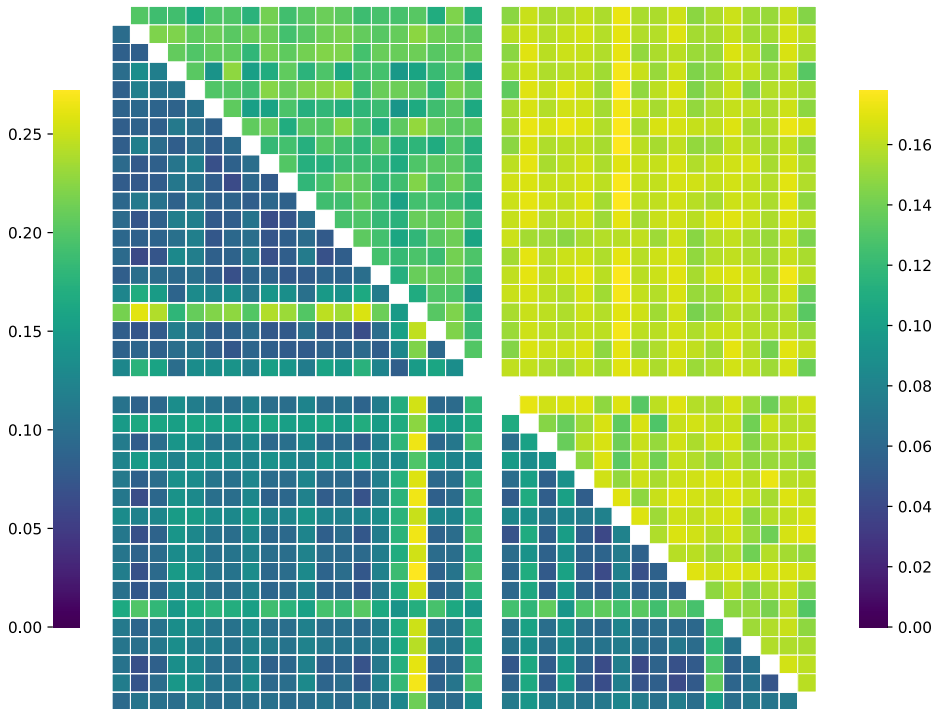


Figure 3: Frobenius distances on the coma dataset. Lower diagonal: normalized Frobenius norm for correlation adjacency matrices A . Upper diagonal: normalized Frobenius norm for density-based adjacency matrices \tilde{A} .

By varying the density of the graphs, one can also determine an optimal density threshold value at which groups can be clustered. To that end, we apply clustering algorithms on the overlap distance. Moreover, comparing different densities provide insights on the underlying structure of different graph groups. For healthy subjects and comatose patients, the results are presented in Figure 4a, and in Figure 4b for

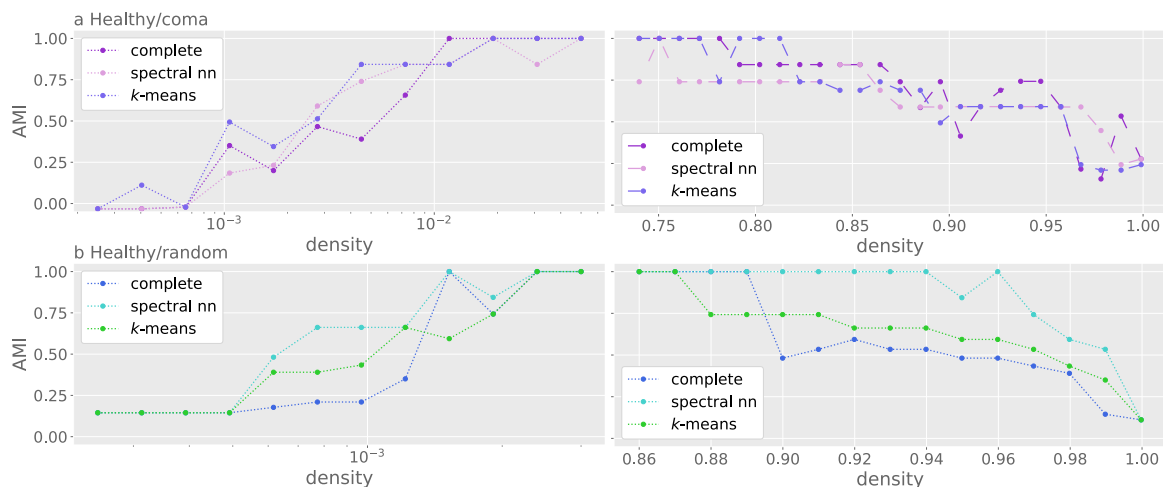


Figure 4: Adjusted mutual information score (AMI) for density edge overlaps: (a) 20 healthy controls and 17 patients in a comatose state (b) 20 healthy controls and 17 random matrices (symmetric, with normal iid off-diagonal elements).

random graphs compared with healthy controls and comatose patients. Random graph adjacency matrices are generated as symmetric matrices, with normal independent and identically distributed (iid) off-diagonal elements.

At lower densities, it is obviously impossible to cluster between groups. Starting from a density of $\sim 10^{-3}$ —which corresponds to a four-edge graph—the different chosen algorithms improve and, eventually, cluster groups perfectly for all problems with a relatively low density. Here, the fMRI measurements of the comatose patients stray markedly away from the healthy controls and appear collectively closer to each other than to the random graphs (see Table 1). But when individually comparing these two groups with random graphs, more edges are required to discriminate comatose patients from random graphs than healthy subjects from random graphs.

When comparing functional brain graphs to random graphs, this gives an estimate of the scale at which the structure starts to fade away. For non-random graphs, it gives the scales at which the variations get drowned in the noise.

3.2. Betti numbers

Figure 5 presents the Betti curves of the different rs-fMRI correlation matrices and their null models using the density-based filtration, which does not require any thresholding step. The latter are presented in the first two rows.

3.2.1. Null models. Both Erdős-Rényi and phase randomization models exhibit fast percolation, reaching a single connected component at low density (Figure 5a). The β_1 and β_2 curves also have similar behavior, with a localized peak where the maximal value is slightly lower for the phase randomization model (Figure 5b and Figure 5c). The Zalesky and spatiotemporal models exhibit slightly different behavior, with a slower

Table 1: Graph density thresholds at which each clustering algorithm starts and stops to cluster the relevant groups perfectly. coma+/coma- correspond to the smallest/highest density threshold which perfectly identifies healthy and comatose subjects. randh+/randh- are analogously defined for random graphs and healthy subjects. randc+/randc- similarly corresponds to random graphs and comatose subjects.

	<i>k</i> -means	hierarchical	spectral
coma+	$(1.3 \pm 0.1) \cdot 10^{-2}$	$(8.8 \pm 0.3) \cdot 10^{-3}$	$(1.68 \pm 0.01) \cdot 10^{-2}$
coma-	0.81	0.79	0.75 ± 0.02
randh+	$(2.3 \pm 0.3) \cdot 10^{-3}$	$(1.7 \pm 0.3) \cdot 10^{-3}$	$(1.4 \pm 0.3) \cdot 10^{-3}$
randh-	0.87	0.90	0.97
randc+	$(1.3 \pm 0.1) \cdot 10^{-2}$	$(1.0 \pm 0.1) \cdot 10^{-2}$	$(4.9 \pm 0.1) \cdot 10^{-2}$
randc-	0.79	0.70	0.75

decrease in the β_0 curve (Figure 5d), particularly in the case of the spatiotemporal model. The β_1 and β_2 features are on average more present in the Zalesky graphs than in the spatiotemporal graphs, and tend to appear at lower densities in the former than in the latter (Figure 5e and Figure 5f).

Betti numbers hence seem to capture some characteristics that are specific to each null model.

3.2.2. Real data. Betti curves highlight the importance of the preprocessing of brain graphs, and synthesize how it influences the dependencies of the large number of random variables at play. Here, we compare three preprocessing strategies. The first (HCP) corresponds to only taking the wavelet correlation between time series, while the other ones include a regression on white matter and cerebrospinal fluid (WM & CSF), with the last adding global signal regression (GSR), which are common preprocessing step. For β_0 , additional signal regressions make the curve fall faster, meaning that fewer regions are poorly connected to the main connected component (Figure 5g). For β_1, β_2 , the curves follow each other at low and high densities but reach higher maxima (Figure 5h and Figure 5i). However, it seems that there is an important variability in the number of features, making it hard to reach conclusions that could be generalized over the entire cohort.

For the healthy subject/comatose patient cohort, there are limited deviations between the mean Betti curve, and they are restricted to some density intervals (Figure 5j, Figure 5k, Figure 5l). However, the shape of the confidence interval for the coma group indicates that there are some patients who present Betti numbers noticeably higher than that of the group average and are probably driving it up.

Finally, it seems Betti numbers mostly capture preprocessing differences, and are not sufficient to clearly identify the different groups.

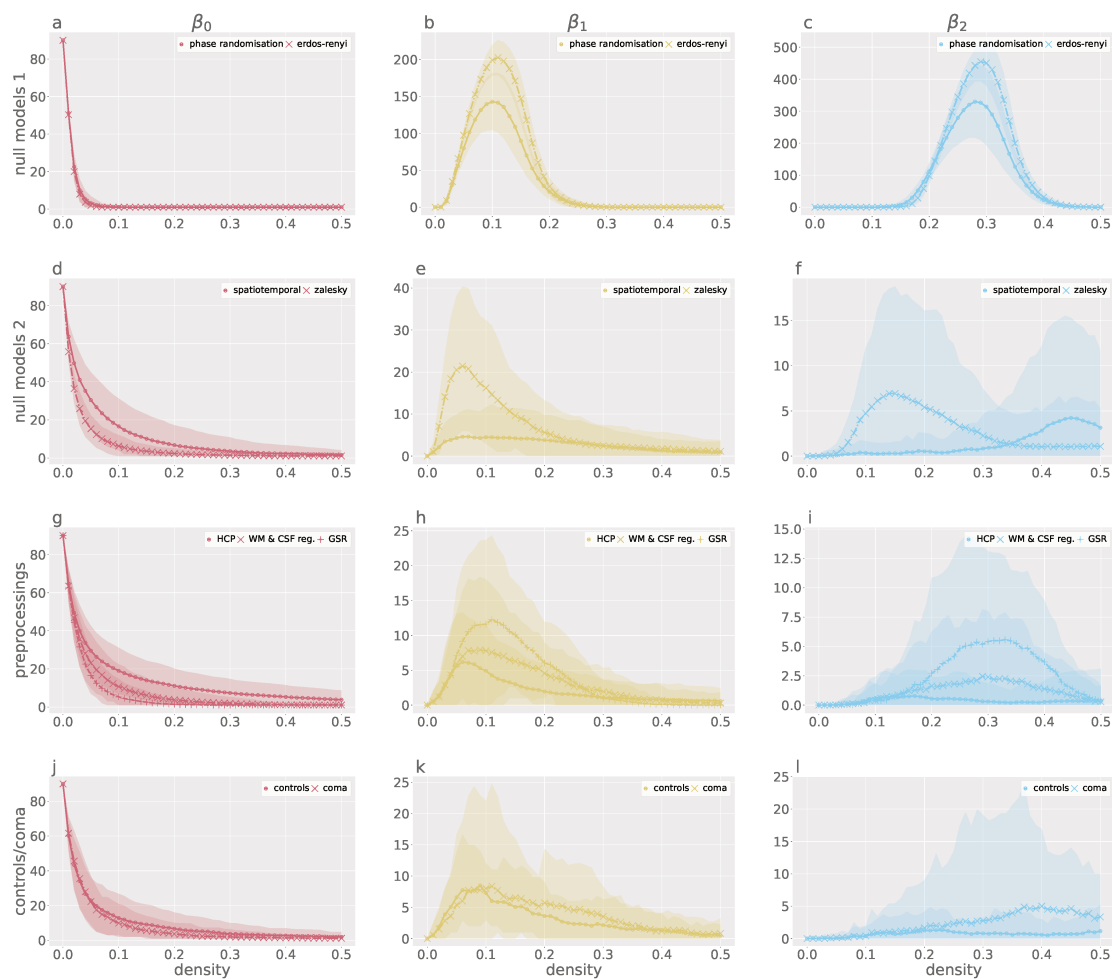


Figure 5: From left to right: Betti curves for β_0 , β_1 and β_2 . (a) (b) (c) Null model Betti curves: phase randomization and Erdős-Rényi (d) (e) (f) Null model Betti curves: spatiotemporal and Zalesky (g) (h) (i) Empirical functional connectivity in HCP subjects with different preprocessings, (j) (k) (l) healthy subjects and comatose patients. Shaded areas give a 95% confidence interval around the mean β_i value at the given density.

3.3. Persistence diagrams

With the mean Betti curves having shown their ability to capture some variability between the different classes of empirical and simulated networks, this section explores the possibilities offered by standard optimal transport distances between persistence diagrams.

Examples of persistence diagrams are presented in Figure 6. The Erdős-Rényi model and the phase randomization model show a large number of β_1 , β_2 features, clear separation between the different feature scales where the β_1 or β_2 features dominate, and important variations in the lifetime of the different features. The example of an

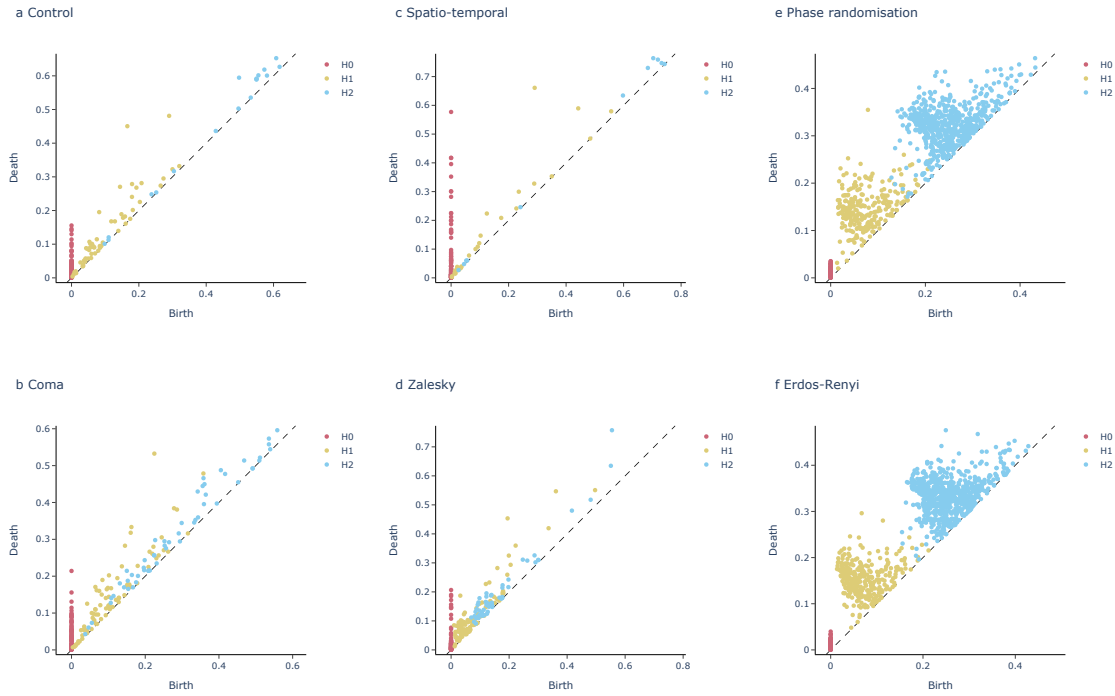


Figure 6: persistence diagrams (a) of functional connectivity of a healthy subject (b) of a comatose patient (c) of a spatiotemporal null model (d) of a Zalesky null model (e) of a phase randomization null model and (f) of an Erdős-Rényi model.

HCP subject, comatose patient, spatiotemporal and Zalesky model all exhibit similar behavior: fewer features with most of them having a short lifetime, and fuzzier separation between β_1 - and β_2 -dominated scales.

Furthermore, we consider the distance between persistence diagrams without label information (cf. Figure 7). For the healthy subject and comatose patient cohort, the 1-Wasserstein distance does not capture the difference between the two groups. Most subjects and patients seem equidistant to each other, with a handful of outliers being a greater distance away from the main group. The bottleneck distance presents a similar behavior.

Meanwhile, for the null models, the Wasserstein distance clearly discriminates between three groups: phase randomization, Erdős-Rényi and a third group including healthy subjects, spatiotemporal, and Zalesky models. The bottleneck distance only separates the latter from a single group containing both phase randomization and Erdős-Rényi models.

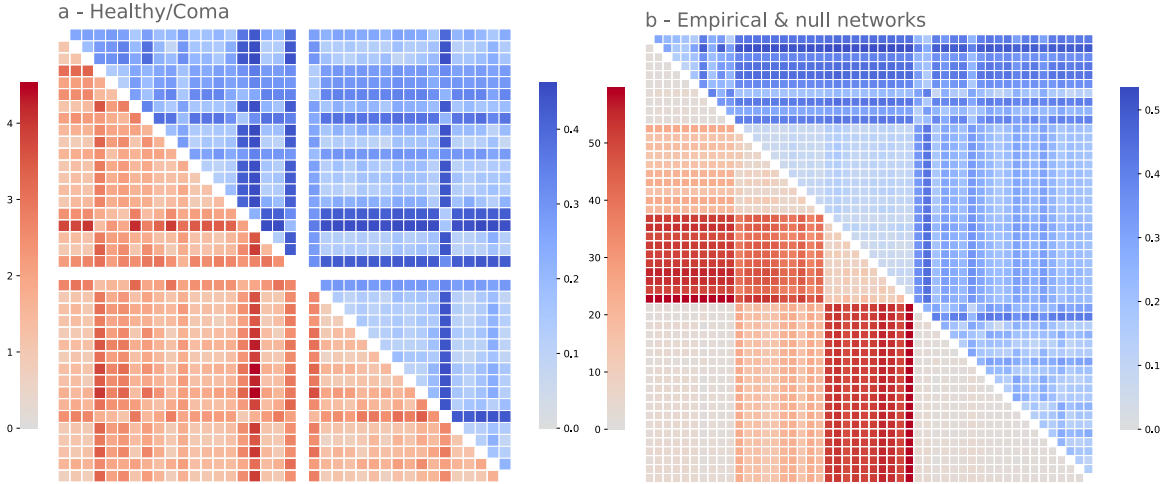


Figure 7: **Upper triangle:** Bottleneck distance. **Lower triangle:** 1-Wasserstein distance. (a) Distance matrices for 20 healthy subjects and 17 comatose patients and (b) 10 healthy subjects with 10 realizations of each of the four null models (phase randomization, Erdős-Rényi, spatiotemporal, Zalesky).

3.4. Node-label information

It appears that the persistence diagram alone is not enough to identify fine differences in true functional network data, unlike a simple overlap approach combined with density-based thresholding. Hence, taking into account the node ordering is essential to effectively compare functional networks. Thus, we evaluate our proposed node-label informed distance on the persistence diagrams.

First, we apply the overlap distance to the coma dataset and to null models at a density of 0.05 (approximately the percolation density in random graphs, but this remains an arbitrary choice). Results are reported in the upper diagonal of the matrices in Figure 8. A simple inspection of the distance matrix reveals a high separation between comatose and healthy subjects. Similarly, in the comparison of null models and real data, the overlap distance clearly groups apart healthy subjects from the null models. Nevertheless, this distance does not differentiate between the various null models.

Then, we compute the edge comparison optimal transport distance which is reported in the lower diagonal of matrices in Figure 8. The distance is obtained by considering H_0 , H_1 , and H_2 homological features. A clear separation between the healthy subjects and the comatose patients is apparent. Similarly to the overlap distance, the optimal transport edge comparison distance separates real data from null models. Moreover, contrary to the overlap distance, our proposal clusters together the phase randomization and Erdős-Rényi models.

Finally, we quantitatively compare all the considered distances by their reached adjusted mutual information score (AMI) for the clustering task in the coma dataset (See Table 2).

The overlap, the Frobenius and our proposed edge comparison optimal transport

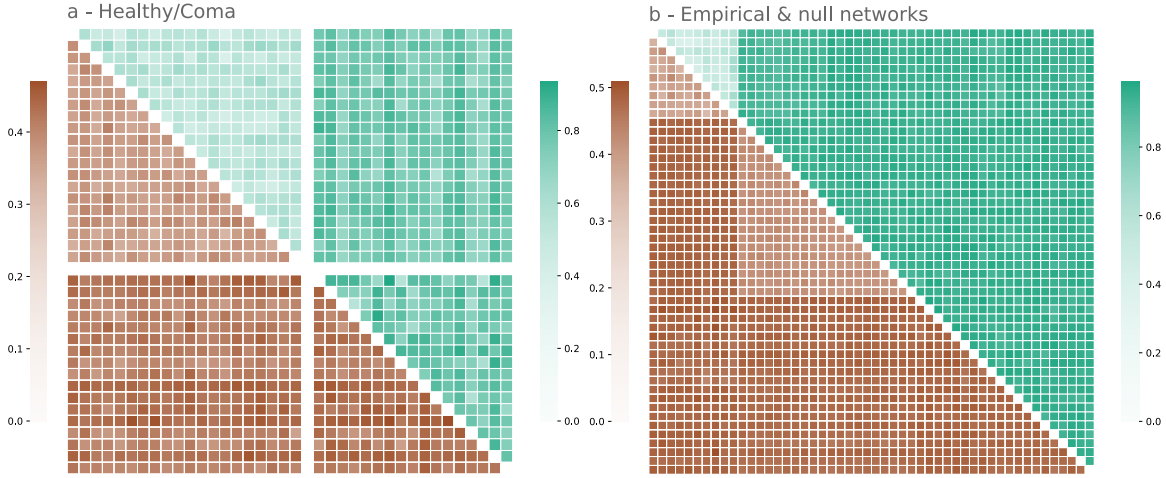


Figure 8: Edge comparison optimal transport distance $d_{e,012}$ between edges of the persistence diagrams (with $\beta_0, \beta_1, \beta_2$ features, lower triangle) and overlap distances between edges at density 0.05 (upper triangle), (a) of 20 healthy subjects and 17 comatose patients (left) and (b) 10 healthy subjects with 10 realizations of each of the 4 null models (right, phase randomization, Erdős-Rényi, spatiotemporal, Zalesky).

Table 2: Adjusted mutual information score of different distances for the clustering of the healthy subjects and comatose patients. The distances are $d_{O,\tau=0.05}$ the edge-overlap distance for thresholded matrices at density level 0.05, d_F the Frobenius-distance, W_p the p -Wasserstein distance between persistence diagrams, Bo the Bottleneck distance between persistence diagrams, $d_{e,k_1k_2k_3}$ the edge comparison optimal transport distance on the persistence diagram with homology features $H_{k_1}, H_{k_2}, H_{k_3}$.

	k -means	hierarchical	spectral
$d_{O,\tau=0.05}$	1	1	1
d_F	1	0.59	1
W_1	0.04	0.04	0.03
W_2	0.06	0.06	0.06
Bo	0.06	0.06	0.12
$d_{e,012}$	1	0.32	1
$d_{e,0}$	0.74	0.32	0.74
$d_{e,1}$	0.16	0.06	0.07
$d_{e,2}$	0.05	-0.02	0.05
$d_{e,01}$	0.84	0.41	1
$d_{e,12}$	0.00	0.00	0.25

$d_{e,012}$ $d_{e,01}$ distances all succeed in realizing a perfect clustering of the healthy-coma cohort. However, the standard Wasserstein and Bottleneck distances do not manage to

cluster the data, with an AMI in the order of 0.05.

Node label information is hence crucial to high-quality persistence diagram comparison.

4. Discussion

4.1. Threshold

As far as the authors know, this work is the first attempt to use graph-density weighted persistence diagrams for the analysis of dense weighted graphs. This original approach is grounded in our results presented in Figure 3 where the Frobenius distance of correlation matrices fails to detect any difference between healthy subjects and patients. It seems that in functional brain connectivity correlation rank is more informative than its value. Commonly, functional connectivity matrices are thresholded before performing downstream analysis. A common task in neuroscience research is to differentiate multiple classes. Thus, the threshold can be defined to optimize the clustering separation as it is done in Table 1. The inter-subject comparison appears easier when simply defining a distance through the intersection of the edge sets of thresholded, binary graphs. Yet, this exclusively requires a supervised approach. Indeed, in these cases, the density threshold needs to be learned on data with known class assignments. The benefit of this approach is also highlighted by the perfect AMI score of the overlap distance-based clustering reported in Table 2. This is the only distance to have a perfect score in all three considered algorithms. However, it should be noted that the distance d_O was applied to a thresholded matrix, where the density was chosen to maximize the AMI. This means these results might not hold in general settings where the threshold choice is not optimized for the AMI, or in unsupervised cases.

4.2. Graph-comparison and label information

Recent developments in the study of complex systems and the increase of data that can be naturally modeled as networks have led to an explosion of graph-centric problems. From this arises the need to develop approaches to compare graphs, for example, in order to establish the properties of a given molecule, proteins, etc. Despite the great interest in graph distances, relatively few methods are engineered to take into account edge and node labels, compared with the proliferation of permutation invariant distances [44]. While introducing labeled and weighted edges brings an additional layer to the task, it allows leveraging alternative approaches in order to quantify graph similarity.

The traditional persistent homology approach allows the comparison of brain functional connectivity networks without requiring a threshold choice or supervised data. Indeed, one of the main arguments for the use of persistent homology is its multiscale approach. However, classical persistent distances fail to capture the variability between the comatose patients and healthy subjects (Figure 7). This might be due to the permutation invariant representation which does not take into account node labeling.

This is especially precarious for functional brain networks, where nodes are brain regions and are obviously distinct from each other.

Moreover, standard methods might be limited to discriminating groups with radically different underlying graph structures, as it happens for null models in Figure 7. This issue can be partially solved by including in the persistence diagram distance the label information of the edges associated with the appearance and disappearance of each feature.

This allows a considerable increase in the AMI score, reaching a perfect separation of the coma and control groups. While it offers an appealing baseline, it remains somewhat limited from an interpretation standpoint. Nevertheless, the H_0 persistence might be interpreted as an implicit node-wise threshold, where one keeps roughly only the most significant edge for each node.

4.3. Real Data

Our reported persistent homology results in real data with different preprocessing reveal the effect of preprocessing on the Betti numbers, and thus on high-order graph structure (Fig. 5). This highlights the importance of documenting the applied framework for inferring brain networks from fMRI time series. This also suggests that studies on the effect of preprocessing on connectivity require more attention. Our application to the coma cohort is highly valuable since discriminating between subjects within different states of consciousness is a difficult task. With the exception of the Frobenius and traditional persistence diagram distances, all the considered distances manage to perfectly cluster patients and controls (Figure 7, Table 2). Interestingly, the pairwise distance between patients and controls is approximately the same as the distance between patients. It would seem that healthy controls are organized along a similar pattern while comatose patients are not.

4.4. Null models

A qualitative comparison of persistence diagrams shows that the Zalesky and the spatiotemporal model can roughly reproduce the persistent homology of empirical functional brain networks (Figure 6). The persistence diagrams of Phase randomization and Erdős-Rényi models are similar one to another, but markedly different from the others.

This visual inspection is reinforced by the results on the considered distances (Figures 7, 8). Interestingly, when considering the W_1 distance the real data are grouped together with the Zalesky and spatiotemporal model. This means that real data birth and death feature distribution can be approximated by considering a dominant Gaussian signal and adjusting the noise of individual regions in order to match the correlation distribution or by capturing limited spatial and temporal auto-correlations from the fMRI data. Meanwhile, the amplitude of the persistence diagrams appears to vary across real data and these two models. This leads to high bottleneck distances and

prevents grouping them all together. The opposite is observed for phase randomization and Erdős-Rényi models: they are grouped together by bottleneck distance and not by W_1 . Moreover, they are also gathered by the edge comparison optimal transport, suggesting they might produce similar label feature distribution.

Both considered label-based distances differentiate the spatiotemporal and Zalesky null models from empirical data, but not from each other (Figure 8). This might be expected for the latter, since it does not take into account any label information to generate surrogate matrices, but is more surprising for the former. Hence, these label-dependent distances demonstrate that even null models that input some kind of spatial information do not reproduce label-dependent behavior.

None of the null models we consider manage to reproduce the real data label organization. From a quality control perspective, this helps ensure that the inferred networks carry meaningful information and are not overcome by noise.

4.5. Limitations and perspectives

For somewhat large networks, persistent homology stays limited to the lower dimensions, as the computational cost for computing higher-order features increases exponentially as it requires finding an arrangement of cliques. Although in this case, as the dependencies seem strong, finding high-order features would be surprising, it cannot yet be ruled out. Furthermore, in functional brain networks, first and second-order homology features seem to be short-lived, limiting the interest in persistent homology, where longer-lived features are the signature of particular topological invariants and are the main attribute that is targeted by persistent homology. In short, persistent homology appears to capture some of the texture of functional brain networks but does not uncover larger-scale organization.

The approach of building simplicial complexes using correlation matrices has been criticized as a heavily limited framework for the exploration of higher-order interactions [45]. An alternative would be to consider information measures able to capture synergistic behavior and, either use them to investigate general properties of the system [46] or to build simplicial complexes using the tricorrelation and upwards. The second option is strongly limited by the high combinatorial price to pay, but could be implemented on a restricted set of nodes in the graph. Additionally, approaches based on density could solve one of the big issues in studying higher-order interactions: the lack of an equivalent to correlations. Indeed, there is no equivalent to the Cauchy-Schwartz inequality for trilinear and higher dimensional forms, and hence, for example, the coskewness of 3 random variables is not bounded and its usefulness for building a filtration is not straightforward. If instead, one looks at the triangle density (or 2-simplex density), one would be able to obtain two-dimensional homology that truly takes into account triadic dynamics (and this stays true for higher-order interactions). In practice, this could prove beneficial, as it should yield a finer understanding of the dynamics. Nevertheless, it stays limited by the exponential growth of the number k -

simplices when k increases.

It should also be noticed that these works have been done on a restricted cohort. While this has not been investigated in this setting, it is becoming more and more apparent that small sample sizes can have calamitous consequences on the confidence in results of functional MRI studies [47]. This presents multiple challenges, notably the need for analysis procedures that scale well and for an increase in data availability. One can easily see how problematic this is in the case of comatose patients, as their ability to give informed consent is at best uncertain [48, 49].

5. Conclusion

To the best of our knowledge, our proposed graph-density-based filtration is the first attempt to include density levels in weighted graph persistent filtration, instead of correlation values. Moreover, we propose to include label information when comparing non-permutation invariant applications. Particularly, this is required in functional brain networks where nodes are associated with brain regions and are not perfectly exchangeable. We evaluate our approach on both real data and null models.

While standard persistent homology fails to discriminate healthy controls from comatose patients, our label-informed distance addresses this weakness. In this task, our distance is comparable to simple overlap or matrix distances. Furthermore, our proposal does not require the choice of an arbitrary threshold and is more informative.

Finally, these label-dependent distances show that node information included in some null models does not constrain the model enough to be close to the real data. This suggests new objectives in the design of null models for brain connectivity.

Although the application of persistent homology for the comparison of functional brain networks remains limited, it offers some understanding of the role of multiscale approaches and of higher-order interactions. It also helps to bring new insights about the importance of integrating node label information for quantitative functional brain network comparison.

Acknowledgments

This work was partly funded by French National Research Agency project ANR-20-NEUC-0003-02, and by MIAI@Grenoble Alpes (ANR 19-P3IA-003).

References

- [1] Sporns O 2022 *Dialogues in clinical neuroscience*
- [2] Bullmore E and Sporns O 2009 *Nature reviews neuroscience* **10** 186–198
- [3] Achard S and Bullmore E 2007 *PLOS Computational Biology* **3** 1–10 URL <https://doi.org/10.1371/journal.pcbi.0030017>
- [4] Stanley M, Moussa M, Paolini B, Lyday R, Burdette J and Laurienti P 2013 *Frontiers in computational neuroscience* **7** 169
- [5] Bassett D S and Bullmore E 2006 *The neuroscientist* **12** 512–523

- [6] Bordier C, Nicolini C and Bifone A 2017 *Frontiers in neuroscience* **11** 441
- [7] Theis N, Rubin J, Cape J, Iyengar S, Gur R E, Gur R C, Roalf D R, Pogue-Geile M F, Almasry L, Nimgaonkar V L *et al.* 2021 *bioRxiv* 2021–10
- [8] Couto C M V, Comin C H and Costa L d F 2017 *Mol. Biosyst.* **13** 2024–2035
- [9] Kartun-Giles A P and Bianconi G 2019 *Chaos, Solitons & Fractals: X* **1** 100004
- [10] Sizemore A E, Phillips-Cremins J E, Ghrist R and Bassett D S 2019 *Network Neuroscience* **3** 656–673 URL https://doi.org/10.1162/netn_a00073
- [11] Torres L, Blevins A S, Bassett D and Eliassi-Rad T 2021 *SIAM Review* **63** 435–485 URL <https://doi.org/10.1137/20m1355896>
- [12] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J G and Petri G 2020 *Physics Reports* **874** 1–92 URL <https://doi.org/10.1016/j.physrep.2020.05.004>
- [13] Battiston F, Amico E, Barrat A, Bianconi G, de Arruda G F, Franceschiello B, Iacopini I, Kéfi S, Latora V, Moreno Y, Murray M M, Peixoto T P, Vaccarino F and Petri G 2021 *Nature Physics* **17** 1093–1098 URL <https://doi.org/10.1038/s41567-021-01371-4>
- [14] Andjelković M, Tadić B and Melnik R 2020 *Scientific Reports* **10** URL <https://doi.org/10.1038/s41598-020-74392-3>
- [15] Gatica M, Cofré R, Mediano P A, Rosas F E, Orio P, Diez I, Swinnen S P and Cortes J M 2021 *Brain Connectivity* **11** 734–744 URL <https://doi.org/10.1089/brain.2020.0982>
- [16] Herzog R, Rosas F E, Whelan R, Fittipaldi S, Santamaria-Garcia H, Cruzat J, Birba A, Moguilner S, Tagliazucchi E, Prado P and Ibanez A 2022 *Neurobiology of Disease* **175** 105918 URL <https://doi.org/10.1016/j.nbd.2022.105918>
- [17] Gatica M, Rosas F E, Mediano P A M, Diez I, Swinnen S P, Orio P, Cofré R and Cortes J M 2022 *PLOS Computational Biology* **18** e1010431 URL <https://doi.org/10.1371/journal.pcbi.1010431>
- [18] Carlsson G 2009 *Bulletin of the American Mathematical Society* **46** 255–308 URL <https://doi.org/10.1090/s0273-0979-09-01249-x>
- [19] Lee H, Kang H, Chung M K, Kim B N and Lee D S 2012 *IEEE Transactions on Medical Imaging* **31** 2267–2277 URL <https://doi.org/10.1109/tmi.2012.2219590>
- [20] Chung M K, Hanson J L, Ye J, Davidson R J and Pollak S D 2015 *IEEE Transactions on Medical Imaging* **34** 1928–1939 URL <https://doi.org/10.1109/tmi.2015.2416271>
- [21] Caputi L, Pidnebesna A and Hlinka J 2021 *NeuroImage* **238** 118245 URL <https://doi.org/10.1016/j.neuroimage.2021.118245>
- [22] Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer P J and Vaccarino F 2014 *Journal of The Royal Society Interface* **11** 20140873 URL <https://doi.org/10.1098/rsif.2014.0873>
- [23] Varley T F, Denny V, Sporns O and Patania A 2021 *Royal Society Open Science* **8** 201971 URL <https://doi.org/10.1098/rsos.201971>
- [24] Reimann M W, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, Dłotko P, Levi R, Hess K and Markram H 2017 *Frontiers in Computational Neuroscience* **11** URL <https://doi.org/10.3389/fncom.2017.00048>
- [25] Bardin J B, Spreemann G and Hess K 2019 *Network Neuroscience* **3** 725–743 URL https://doi.org/10.1162/netn_a00080
- [26] Hallett M, de Haan W, Deco G, Dengler R, Di Iorio R, Gallea C, Gerloff C, Grefkes C, Helmich R C, Kringelbach M L *et al.* 2020 *Clinical Neurophysiology* **131** 1621–1651
- [27] Carboni L, Dojat M and Achard S 2023 *Physical Review E* **107** 014302
- [28] Mheich A, Wendling F and Hassan M 2020 *Network Neuroscience* **4** 507–527
- [29] Achard S, Delon-Martin C, Vértes P E, Renard F, Schenck M, Schneider F, Heinrich C, Kremer S and Bullmore E T 2012 *Proceedings of the National Academy of Sciences* **109** 20608–20613 URL <https://doi.org/10.1073/pnas.1208933109>
- [30] Váša F and Mišić B 2022 *Nature Reviews Neuroscience* **23** 493–504 URL <https://doi.org/10.1038/s41583-022-00601-9>

- [31] Essen D V, Ugurbil K, Auerbach E, Barch D, Behrens T, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss S, Penna S D, Feinberg D, Glasser M, Harel N, Heath A, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen S, Prior F, Schlaggar B, Smith S, Snyder A, Xu J and Yacoub E 2012 *NeuroImage* **62** 2222–2231 URL <https://doi.org/10.1016/j.neuroimage.2012.02.018>
- [32] Liu T T, Nalci A and Falahpour M 2017 *NeuroImage* **150** 213–229 URL <https://doi.org/10.1016/j.neuroimage.2017.02.036>
- [33] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B and Joliot M 2002 *NeuroImage* **15** 273–289 URL <https://doi.org/10.1006/nimg.2001.0978>
- [34] Zalesky A, Fornito A and Bullmore E 2012 *NeuroImage* **60** 2096–2106 URL <https://doi.org/10.1016/j.neuroimage.2012.02.001>
- [35] Shinn M, Hu A, Turner L, Noble S, Preller K H, Ji J L, Moujaes F, Achard S, Scheinost D, Constable R T, Krystal J H, Vollenweider F X, Lee D, Anticevic A, Bullmore E T and Murray J D 2021 URL <https://doi.org/10.1101/2021.06.01.446561>
- [36] Chazal F and Michel B 2021 *Frontiers in Artificial Intelligence* **4** URL <https://doi.org/10.3389/frai.2021.667963>
- [37] Croom F H 1978 *Basic Concepts of Algebraic Topology* (Springer New York) URL <https://doi.org/10.1007/978-1-4684-9475-4>
- [38] Titouan V, Courty N, Tavenard R, Laetitia C and Flamary R 2019 Optimal transport for structured data with application on graphs *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research vol 97)* ed Chaudhuri K and Salakhutdinov R (PMLR) pp 6275–6284 URL <https://proceedings.mlr.press/v97/titouan19a.html>
- [39] Vayer T, Chapel L, Flamary R, Tavenard R and Courty N 2020 *Algorithms* **13** 212 URL <https://doi.org/10.3390/a13090212>
- [40] Tauzin G, Lupo U, Tunstall L, Pérez J B, Caorsi M, Medina-Mardones A, Dassatti A and Hess K 2020 giotto-tda: A topological data analysis toolkit for machine learning and data exploration (*Preprint* 2004.02551)
- [41] Flamary R, Courty N, Gramfort A, Alaya M Z, Boisbunon A, Chambon S, Chapel L, Corenflos A, Fatras K, Fournier N, Gautheron L, Gayraud N T, Janati H, Rakotomamonjy A, Redko I, Rolet A, Schutz A, Seguy V, Sutherland D J, Tavenard R, Tong A and Vayer T 2021 *Journal of Machine Learning Research* **22** 1–8 URL <http://jmlr.org/papers/v22/20-451.html>
- [42] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E 2011 *Journal of Machine Learning Research* **12** 2825–2830
- [43] Vinh N X, Epps J and Bailey J 2010 *Journal of Machine Learning Research* **11** 2837–2854 URL <http://jmlr.org/papers/v11/vinh10a.html>
- [44] Tantardini M, Ieva F, Tajoli L and Piccardi C 2019 *Scientific Reports* **9** URL <https://doi.org/10.1038/s41598-019-53708-y>
- [45] Rosas F E, Mediano P A M, Luppi A I, Varley T F, Lizier J T, Stramaglia S, Jensen H J and Marinazzo D 2022 *Nature Physics* URL
- [46] Rosas F E, Mediano P A M, Gastpar M and Jensen H J 2019 *Physical Review E* **100** URL
- [47] Marek S, Tervo-Clemmens B, Calabro F J, Montez D F, Kay B P, Hatoum A S, Donohue M R, Foran W, Miller R L, Hendrickson T J, Malone S M, Kandala S, Feczko E, Miranda-Dominguez O, Graham A M, Earl E A, Perrone A J, Cordova M, Doyle O, Moore L A, Conan G M, Uriarte J, Snider K, Lynch B J, Wilgenbusch J C, Pengo T, Tam A, Chen J, Newbold D J, Zheng A, Seider N A, Van A N, Metoki A, Chauvin R J, Laumann T O, Greene D J, Petersen S E, Garavan H, Thompson W K, Nichols T E, Yeo B T T, Barch D M, Luna B, Fair D A and Dosenbach N U F 2022 *Nature* URL <https://doi.org/10.1038/s41586-022-04492-9>
- [48] Bruni T, Graham M, Norton L, Gofton T, Owen A M and Weijer C 2019 *Journal of medical ethics* **45** 299–303

Multiscale comparison of functional brain networks

23

- [49] White T, Blok E and Calhoun V D 2020 *Human Brain Mapping* **43** 278–291 URL <https://doi.org/10.1002/hbm.25120>

B.3 Homological Information for functional connectivity network small-world regime refinement

Homological information to refine small-world regime of brain functional connectivity networks

Simone Chiominto¹, Francesco Pappone¹, Hanâ Lbath², Lucrezia Carboni^{2,3}, and Sophie Achard²

¹ Dipartimento di Scienze Matematiche, Politecnico di Torino, 10129 Torino, IT

² Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, FR

³ Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, FR

Topological information extraction from network-shaped data is a powerful tool, especially for the characterization of complex structures [6]. Particularly, simple homological features such as Betti numbers of the graph clique complex (Figure 1) can be used to identify synthetic networks obtained by different models [3].

In this work, we consider brain functional connectivity networks with the goal of refining previous results on their small-world properties from a topological perspective.

In [1], brain functional networks have been characterized based on their global and local efficiencies, defined as follows:

$$E_{\text{global}} := \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{L_{i,j}},$$

$$E_{\text{local}} := \frac{1}{N_{G_i}(N_{G_i}-1)} \sum_{j \neq k \in G_i} \frac{1}{L_{j,k}},$$

where G is an undirected unweighted graph, N is the number of nodes of G , $L_{i,j}$ is the minimum path length between nodes i and j and G_i is the subgraph defined by the set of nodes that are nearest-neighbours of the i th node. By definition $i \notin G_i$.

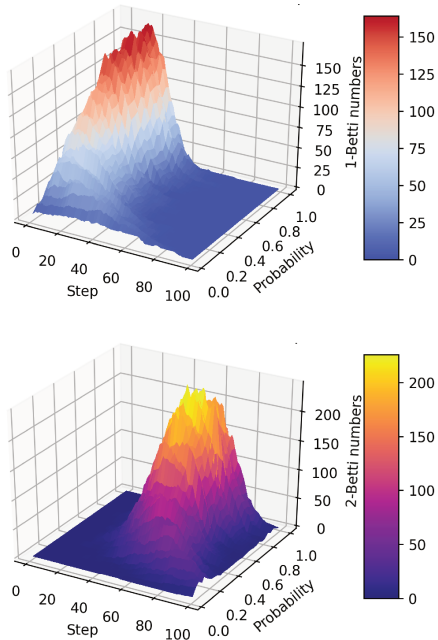


Fig. 1: First (top) and second (bottom) dimensional Betti numbers on small-world graph simulations for different rewiring probabilities at different filtration steps.

In particular, the brain networks efficiency values are computed as a function of a cost variable:

$$\text{cost} := \frac{\text{no. of links}}{N(N-1)}$$

Results demonstrate the existence of a small-world regime, namely a range of costs in which brain networks have both global and local efficiency between those of the extreme cases of Erdos-Renyi [2] and lattice graphs.

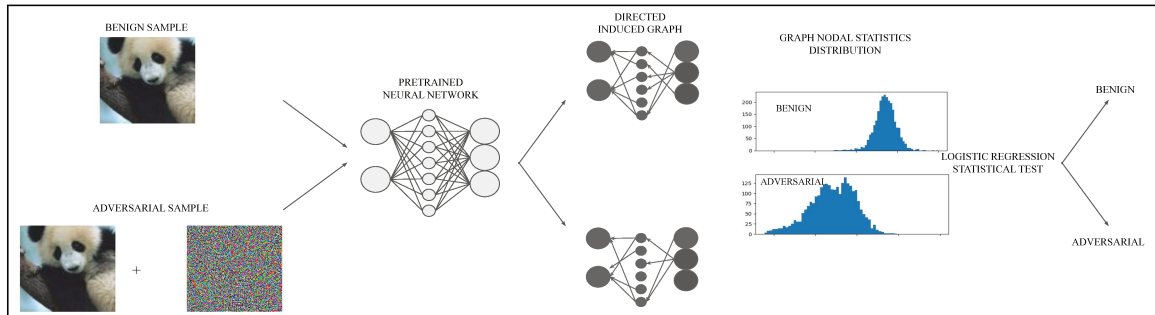
We refine the small-world regime both by providing a characterization in terms of topological features and by inferring the rewiring probability of brain networks under the assumption of Watts-Strogatz generative model [9].

The first characterization is achieved by the computation of the first Betti number on a clique complex filtration associated to a graph obtained at a fixed cost. We prove that applying our proposed approach to a range of costs allows extraction of meaningful topological features from the underlying graphs. We are indeed able to statistically differentiate patients in a comatose state from healthy control subjects.

We then propose to train a Graph Neural Network (GNN) [5] to estimate the rewiring probability associated to synthetically-generated Watts-Strogatz networks at fixed cost values, using exclusively topological information from the neighborhood complex [4] of each node as features. We demonstrate the importance of the topological information by comparing our results with a null model of GNN trained on the same dataset without any topological features. Hence, we use the trained GNN to infer the rewiring probability of real-world brain functional connectivity networks.

-
- [1] S. Achard and E. Bullmore. Efficiency and cost of economical brain functional networks. *PLoS computational biology*, 3(2):e17, 2007.
 - [2] P. Erdős and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
 - [3] D. Horak, S. Maletić, and M. Rajković. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, Mar 2009.
 - [4] A. P. Kartun-Giles and G. Bianconi. Beyond the clustering coefficient: A topological analysis of node neighbourhoods in complex networks. *Chaos, Solitons & Fractals: X*, 1:100004, 2019.
 - [5] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks.
 - [6] F. A. Santos, E. P. Raposo, M. D. Coutinho-Filho, M. Copelli, C. J. Stam, and L. Douw. Topological phase transitions in functional brain networks. *Physical Review E*, 100(3):032414, 2019.
 - [7] G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and K. Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration, 2020.
 - [8] N. Verma, E. Boyer, and J. Verbeek. Dynamic filters in graph convolutional networks. *CoRR*, abs/1706.05206, 2017.
 - [9] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.

B.4 Graph-based methods coupled with specific distributional distances for adversarial attack detection



Visualization of our proposed framework to detect adversarial attacks.

Graph-based methods coupled with specific distributional distances for adversarial attack detection

Dwight Nwaigwe^{a,c}, Lucrezia Carboni^{a,c}, Martial Mermillod^b, Sophie Achard^a, Michel Dojat^c

^a*Univ. Grenoble Alpes, Inria, CNRS, LJK, Grenoble, France*

^b*Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, Grenoble, France*

^c*Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN, Grenoble, France*

Abstract

Artificial neural networks are prone to being fooled by carefully perturbed inputs which cause an egregious misclassification. These *adversarial* attacks have been the focus of extensive research. Likewise, there has been an abundance of research in ways to detect and defend against them. We introduce a novel approach of detection and interpretation of adversarial attacks from a graph perspective. For an image, benign or adversarial, we study how a neural network's architecture can induce an associated graph. We study this graph and introduce specific measures used to predict and interpret adversarial attacks. We show that graphs-based approaches help to investigate the inner workings of adversarial attacks.

Keywords: artificial neural network, graph theory, deep learning, machine learning, adversarial, Wasserstein, bio-inspired

1. Introduction

Artificial neural networks (ANN) are known to be prone to misclassifying carefully perturbed inputs [14]. These perturbed inputs, called adversarial, have been at the forefront of research in the machine learning community for the past decade. There is a lot of interest in creating new adversarial detection and defense methods, especially as this has consequence for a variety of real-world domains that rely on ANN for classification [8], [13], [31].

But among the known methods it is apparent that few of them, as diverse as they are, study adversarial attacks from a graph theory perspective. The

objective of this paper is the exploration of adversarial attacks using graph-based methods. Indeed, the ANN structure can be described by a graph. In the most basic example, if one considers a standard feedforward ANN then, in a graphical representation, the neurons are associated to vertices/nodes and the weights between them are associated to edges. One may take this representation as inspiration for studying ANN from a graph perspective, although we stress that there is more than one way to obtain a graph from an ANN.

In [17], the authors provide a survey of the history of interactions between neuroscience and artificial intelligence and they note how much of the modern success in artificial intelligence can be traced to the understanding of or inspiration by biological systems. There is a line of research in neuroscience that studies the brain using elements of graph theory [4], and this provides some motivation for the use of graph-theoretic approaches to studying ANN.

In this document, we study the detection of adversarial examples using graphs. Given an input to the neural network, we compute an associated sparse graph. From this graph, we then use a combination of selected edges, an importance measure, and degree of nodes to predict if the input is adversarial. In one of our approaches, logistic regression is used. Our second approach is statistical, being based on Wasserstein distance applied to degree of nodes. Lastly, we interpret the relative strength of attacks through our graph-based approach. An advantage of our detection methods is that they include a thresholding step which is non differentiable, thereby precluding gradient masking [28] and making it difficult to make adaptive attacks. As part of our studies we also provide benchmarks.

2. Background and related work

There have been some efforts in interpreting ANN in graph-theoretic ways. The authors of [32] study the existence and properties of *motifs*, clusters of neurons in ANN which appear often. In [18], they interpret ANN as a graph and study how MNIST and CIFAR datasets exhibit different distributions under defined quantities (e.g. node input strength, neuron strength). In [9], a topological study of ANN is made via its *functional graph*, a graph obtained from correlations of neurons. Other work [26],[23],[20] apply a similar topological view to studying ANN. Despite relating graphs to ANN, none of these articles demonstrate using graphs to detect adversarial examples, nor do they provide statistics on detection. An interesting use of graphs

occurs in [6] where they are used to evaluate the robustness of an ANN, as opposed to adversarial detection. In [21] (“LID”), [19], [16], and [12], logistic regression is used to classify an input as benign or adversarial based on certain features, none of which are graph related. Statistical approaches can be found in [11] (“RSA”) and [29], also neither of which use graph methods. In [11], the distances between class prototypes are used to determine if an input is adversarial, while in [29], the authors claim that adding noise to images affects the logits in such a way that adversarial inputs can be detected. Our methods extend and complement the previous methods by showing the power of graph theory perspectives from either a logistic regression or a pure statistics perspective. We also compare our methods with LID and RSA.

3. Graph generation and quantities of interest

To compute the associated graph \mathcal{G} for a neural network and input pair, we use layerwise relevance propagation [2], [24]. This algorithm allows one to assign quantities to neurons which can be interpreted as an indicator of the influence that a neuron has on the output. We assume our graph to be directed. Following the notation in [24] for the LRP- $\alpha\beta$ rule, signals are propagated from the output layer towards the input layers. For a neuron k in layer $\ell + 1$ that is connected to neuron i in layer ℓ , the propagated signal from k to i is defined to be

$$R_{i,k}^{\ell,\ell+1} = R_k^{\ell+1} \left(\alpha \frac{a_i \max(w_{ik}, 0)}{\epsilon + \sum_h a_h \max(w_{hk}, 0)} - \beta \frac{a_i \min(w_{ik}, 0)}{\epsilon + \sum_h a_h \min(w_{hk}, 0)} \right) \quad (1)$$

where R_k^ℓ is the relevance of neuron k , a_i is the activation of neuron i in layer ℓ ; w_{hk} is the weight between neurons h, k ; ϵ is a small parameter; and $\alpha - \beta = 1$. The relevance of a neuron k in layer l is given by

$$R_k^\ell = \sum_i R_{k,i}^{\ell,\ell+1}. \quad (2)$$

To start the algorithm, one assigns the relevance of the output neurons of the neural network to be equal to the neural network output. Upon completion of the algorithm, we rank the pairwise-relevance scores $\{R_{i,k}^{\ell,\ell+1}\}$ in descending order and keep the top 1%. Our thresholding is inspired by [4]. These edges become the edges in our induced graph \mathcal{G} . One can compute various quantities from \mathcal{G} . One such quantity is given by

$$I(v_i) = \sum_{j \neq i} \frac{1}{2^{d(v_i, v_j)}} \quad (3)$$

where $\{v_i\}_i$ is the set of nodes and $d(v_i, v_j)$ is the distance between vertices v_i, v_j . We note that for the distance between adjacent nodes we use (1), and the distance between any pair of nodes is given by the path that corresponds to the shortest sum of distances of adjacent nodes. An intuitive meaning of (3) is that it gives more importance to a vertex that has many neighbors and short distances to them. This equation is inspired by closeness centrality [3] which is given by

$$C(v_i) = \frac{1}{\sum_{j \neq i} d(v_i, v_j)}. \quad (4)$$

A difference between (3) and (4) is that the former is monotone in the cardinality of $\{v_i\}_i$. For bipartite graphs, or “stacks” of bipartite graphs (one can think of multi-layer perceptrons in this fashion) a measure of closeness centrality tends not be useful, hence the motivation for (3).

Another quantity of interest is the degree of a vertex, specifically which is defined to be the difference between out degree and in degree:

$$\deg(v) = \deg_{out}(v) - \deg_{in}(v). \quad (5)$$

Our last quantity of interest are the values of certain edges of \mathcal{G} . This allows us to incorporate some of \mathcal{G} 's topology. The edges we use are those that correspond to the last two layers of the original neural network. We only use these edges because using all edges would require a data structure of size $O(n_1 n_2, \dots, n_l)$, where n_i is the number of nodes in layer i . Clearly, this requires an extensive amount of memory when a sufficient number of n_i is large. One can see that in general, when using graph data, it is preferable, at least from a memory standpoint, to use a quantity whose size is much smaller than $O(n_1 n_2, \dots, n_l)$, for instance a dataset whose size is $O(|V|)$, where V is the set of nodes. In fact, our use of degree and node importance (4) as computed for each node meets this constraint.

In [15], the neurons just before the softmax layer are studied, which has a similarity with our study of edge relevance. In that article, the authors use the said neurons to compare robustness of non-human primate and human vision with regards to adversarial images. This lends further a (biological)

motivation in our use of edge relevance for the edges connecting the penultimate to the output layer. Since we apply a threshold to the edges of \mathcal{G} , there are nodes of \mathcal{G} which are not adjacent to an edge. More generally, the edges among the set $\{\mathcal{G}_i\}_i$, need not be the same, where $\{\mathcal{G}_i\}_i$ represents a set of graphs induced from the same architecture. To enforce consistency of representation for the relevance of edges adjacent to the output layer, we create a weighted adjacency matrix of the same dimension as the adjacency matrix for nodes in the last two layers. The relevance values that are above the threshold are recorded as is, and those below this percentile are set to 0. The matrix is then flattened into a vector. This flattened vector is our third quantity of interest, and its nonzero components are given by (1), assuming that component is greater than the threshold.

Lastly, we note that it would be very difficult to create an adaptive attack to counter the methodology proposed here since our detection methods involve graph thresholding, a nondifferentiable operation.

Table 1: Summary of relevant graph statistics. Edge relevance is restricted to last layer.

FORMULA	NAME
$R_{i,k}^{\ell,\ell+1} = R_k^{\ell+1} \left(\alpha \frac{a_i \max(w_{ik},0)}{\epsilon + \sum_h a_h \max(w_{hk},0)} - \beta \frac{a_i \min(w_{ik},0)}{\epsilon + \sum_h a_h \min(w_{hk},0)} \right)$	EDGE RELEVANCE
$I(v_i) = \sum_{j \neq i} \frac{1}{2^{d(v_i, v_j)}}$	NODE IMPORTANCE
$\text{deg} = \text{deg}_{out}(v) - \text{deg}_{in}(v)$	DEGREE

4. A statistical test based on Wasserstein distances

The Wasserstein-1 distance between two probability distributions p and q defined on a measurable metric space \mathcal{X} is given by

$$\mathcal{W}(p, q) = \min_{\pi(x,y) \in \Pi} \int \|x - y\|_1 d\pi(x, y) \quad (6)$$

where Π is the set of all measures on $(\mathcal{X}, \mathcal{X})$ whose marginal distributions are given by p and q . In the case when p consists of one sample x and q

consists of discrete samples $(y_i)_{i=1}^N$, then

$$\mathcal{W}(\delta_x, q) = \frac{1}{N} \sum_i^N \|x - y_i\|_1. \quad (7)$$

where δ_x is the distribution with support at x . Wasserstein distances have been applied to machine learning in several ways. In [7], Wasserstein distances are used to compress data into a small dimensional subspace while maintaining a large distance from adversarial distributions. Other work [30] uses Wasserstein distances to create adversarial attacks.

Our goal in using Wasserstein distances is different than that in the examples mentioned. Our goal is to apply Wasserstein differences for benign and adversarial graph statistics in order to classify an input as benign or adversarial. The statistic we are concerned with is degree.

Let $\hat{\mathcal{B}}_i$ denote the empirical distribution of degree in the case when benign inputs are correctly classified as belonging to class i . Similarly, let $\hat{\mathcal{A}}_i$ denote the empirical distribution that corresponds to perturbed inputs which the model incorrectly classifies as belonging to class i , and whose unperturbed image is correctly classified. For instance, since we are concerned with degree, the domain of the distribution function $\hat{\mathcal{B}}_i$ is a vector whose dimension is equal to the number of nodes in the induced graphs. If for some input, the model outputs class i , we would like to know if the output was generated by a random variable with distribution \mathcal{B}_i or with distribution \mathcal{A}_i where the lack of a hat denotes the true distribution. As before, we first construct the graph \mathcal{G} for the sample and compute a sample degree vector, which we denote by the random variable \mathbf{Z} . For a yet to be defined subset of nodes \mathcal{S} , we define the following Wasserstein Sums Ratio (WSR) quantity:

$$\text{WSR}(\mathcal{S}, \hat{\mathcal{A}}_i, \hat{\mathcal{B}}_i, \mathbf{Z}, i) = \frac{\sum_{j \in \mathcal{S}} \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{B}}_i^j)}{\sum_{j \in \mathcal{S}} \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{A}}_i^j)} \quad (8)$$

where the j in $\hat{\mathcal{A}}_i^j$ refers to the empirical distribution for node j , and similarly for $\hat{\mathcal{B}}_i^j$. Equation (8) says that for each node that belongs to \mathcal{S} , we compute Wasserstein-1 distances node-wise from the sample to the empirical distributions and we sum over the node indices, and compute the ratio. If the ratio is less than some threshold, we classify the input as benign, otherwise as adversarial. It may occur that the denominator of (8) is equal to 0, thus, in this case, a small term is added to the numerator and denominator.

This can happen if the empirical distributions $\{\hat{\mathcal{A}}_i^j\}_{j \in \mathcal{S}}$ only have support at a point. Lastly, we note that we could have also computed the Wasserstein distance in \mathbb{R}^N , where N is the number of nodes in \mathcal{G} . However, that is a more involved procedure. Using (7), we can write (8) as

$$\text{WSR}(\mathcal{S}, \hat{\mathcal{A}}, \hat{\mathcal{B}}, \mathbf{Z}, i) = \frac{\frac{1}{N_{\hat{\mathcal{B}}_i^j}} \sum_{j \in \mathcal{S}} \sum_{k=1}^{N_{\hat{\mathcal{B}}_i^j}} \|\mathbf{Z}_j - y_i^j(k)\|_1}{\frac{1}{N_{\hat{\mathcal{A}}_i^j}} \sum_{j \in \mathcal{S}} \sum_{k=1}^{N_{\hat{\mathcal{A}}_i^j}} \|\mathbf{Z}_j - x_i^j(k)\|_1} \quad (9)$$

where $y_i^j(k)$ is a sample from $\hat{\mathcal{B}}_i^j$ and $x_i^j(k)$ is a sample from $\hat{\mathcal{A}}_i^j$, and $N_{\hat{\mathcal{B}}_i^j}$ is the number of samples in $\hat{\mathcal{B}}_i^j$, respectively for $\hat{\mathcal{A}}_i^j$. Lastly, we make the set \mathcal{S} as follows: we calculate

$$\Delta_i^j := \mathbb{E}X_i^j - \mathbb{E}Y_i^j \quad (10)$$

where X_i^j has distribution $\hat{\mathcal{A}}_i^j$ and Y_i^j has distribution $\hat{\mathcal{B}}_i^j$ and \mathbb{E} is expected value. We then create the set

$$\mathcal{S} := \{j : \Delta_i^j < 0 \text{ for all } i\}. \quad (11)$$

The set \mathcal{S} identifies nodes where the mean of the benign distribution is greater than the adversarial distribution for all classes. Should it happen that $\hat{\mathcal{A}}_i^j$ is empty for some j (we have experienced this only for one combination of model and attack), one may create a placeholder version of it by setting each entry to a very large negative value (the large negative value has the effect of removing the index j from consideration when making the set \mathcal{S}). Algorithm 1 shows adversarial detection using WSR.

Algorithm 1 Adversarial detection using WSR (variant 1)

Input: neural network \mathcal{NN} , image I ; τ , \mathcal{S} , $\hat{\mathcal{A}}_i^j$; $\hat{\mathcal{B}}_i^j$ for all i and j
 $i \leftarrow \mathcal{NN}(I)$
 compute \mathcal{G} from I and \mathcal{NN}
 compute node degree \mathbf{z} from \mathcal{G}
 $val \leftarrow \text{WSR}(\mathcal{S}, \hat{\mathcal{A}}, \hat{\mathcal{B}}, \mathbf{z}, i)$
 if $val < \tau$ then classify I as benign, otherwise classify I as adversarial.

The way we construct \mathcal{S} has the tendency to pick nodes that generalize well across all classes at the expense of nodes that specialize. In an alternative

algorithm, we propose to use the specialized nodes. For a given output that is classified as class i , we use $\mathcal{S}_i = \{j : \Delta_i^j < 0\}$. This can result in a more accurate test using our approach, but at the expense of a little longer computation since there are more nodes to use for computations. The algorithm is shown in Algorithm 2.

Algorithm 2 Adversarial detection using WSR (variant 2)

Input: neural network \mathcal{NN} , image I ; τ_i , \mathcal{S}_i , $\hat{\mathcal{A}}_i^j$; $\hat{\mathcal{B}}_i^j$ for all i and j
 $i \leftarrow \mathcal{NN}(I)$
 compute \mathcal{G} from I and \mathcal{NN}
 compute node degree \mathbf{z} from \mathcal{G}
 $val \leftarrow \text{WSR}(\mathcal{S}_i, \hat{\mathcal{A}}, \hat{\mathcal{B}}, \mathbf{z}, i)$
 if $val < \tau_i$ then classify I as benign, otherwise classify I as adversarial.

5. Consistency

We would like to analyze under what conditions (8) is a faithful predictor. We treat the case of a finite-width ANN with sufficiently many neurons. A finite-width ANN has the property that the degree distribution has compact support, which implies that the Wasserstein distance between an empirical degree distribution and true distribution is bounded, and the Wasserstein distance is continuous with respect to $\|\cdot\|_\infty$. We begin our proof of consistency by showing that given a real-valued random variable X ; an empirical distribution \hat{F}_n of some other real-valued random variable with true distribution F ; a function G (whose arguments are a random variable and a distribution) that is uniformly continuous in the second argument with respect to $\|\cdot\|_\infty$; and bounded, that

$$\mathbb{E}_X G(X, \hat{F}_n) \xrightarrow{a.s.} \mathbb{E}_X G(X, F) \quad (12)$$

as $n \rightarrow \infty$. To prove (12), it is sufficient to show that

$$G(X, \hat{F}_n) \xrightarrow{a.s.} G(X, F) \quad \forall x. \quad (13)$$

Under identical and independently distributed (iid) assumptions, the Glivenko-Cantelli lemma states that $\|\hat{F}_n - F\|_\infty \xrightarrow{a.s.} 0$. This combined with the uniform continuity of G in the second argument with respect to $\|\cdot\|_\infty$ proves (13). To prove (12), we let $h_n(x) = G(x, \hat{F}_n)$ and $h(x) = G(x, F)$. From (13)

we have $h_n(x) \xrightarrow{a.s.} h(x)$ for all x as $n \rightarrow \infty$. We may combine this with the boundedness assumption to use the Lebesgue dominated convergence theorem, resulting in $\lim_{n \rightarrow \infty} \mathbb{E}_X h_n(X) = \mathbb{E}_X \lim_{n \rightarrow \infty} h_n(X) = \mathbb{E}_X h(X)$ almost surely.

We now begin to analyze (8), and we start by supposing that our random variable \mathbf{Z} corresponds to the benign case. Let

$$\begin{aligned} U_{j,i}^b &= \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{B}}_i^j) \\ U_{j,i}^a &= \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{A}}_i^j). \end{aligned} \tag{14}$$

For additional simplicity, let us assume that quantities defined in (14) are iid over the index j . The iid assumption implies that

$$\mathbf{E}_{\mathbf{z}_j} \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{B}}_i^j) =: \mathbf{E}U_i^b$$

and

$$\mathbf{E}_{\mathbf{z}_j} \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{A}}_i^j) =: \mathbf{E}U_i^a$$

for all i . By equation (12), the results we obtain going forward will hold for the population distribution in high probability assuming our empirical distributions have enough samples. By the weak law of large numbers,

$$\left| \frac{\sum_{j=1}^{|\mathcal{S}|} \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{B}}_i^j)}{|\mathcal{S}|} - \mathbf{E}U_i^b \right| < \epsilon_1 \text{ as } |\mathcal{S}| \rightarrow \infty$$

Similarly,

$$\left| \frac{\sum_{j=1}^{|\mathcal{S}|} \mathcal{W}(\delta_{\mathbf{z}_j}, \hat{\mathcal{A}}_i^j)}{|\mathcal{S}|} - \mathbf{E}U_i^a \right| < \epsilon_2 \text{ as } |\mathcal{S}| \rightarrow \infty.$$

Then (8) is equal to

$$\begin{aligned} \frac{\sum_{j=1}^{|\mathcal{S}|} U_{j,i}^b}{\sum_{j=1}^{|\mathcal{S}|} U_{j,i}^a} &= \frac{|\mathcal{S}| \mathbf{E}U_i^b + |\mathcal{S}| \epsilon_1}{|\mathcal{S}| \mathbf{E}U_i^a + |\mathcal{S}| \epsilon_2} \\ &= \frac{\mathbf{E}U_i^b + \epsilon_1}{\mathbf{E}U_i^a + \epsilon_2} \\ &\rightarrow \frac{\mathbf{E}U_i^b}{\mathbf{E}U_i^a} \text{ as } |\mathcal{S}| \rightarrow \infty \end{aligned} \tag{15}$$

where ϵ_1 and ϵ_2 are $o(|\mathcal{S}|)$. If we consider the case when \mathbf{Z} is adversarial, we get a similar limit as in (15). Thus for consistency, we need the two limits to not be equal, thus we write

$$\frac{\mathbf{E}U_i^b}{\mathbf{E}U_i^a} < \frac{\mathbf{E}V_i^b}{\mathbf{E}V_i^a} \quad (16)$$

where we use V to denote adversarial quantities. This is equivalent to $\mathbf{E}U_i^b \mathbf{E}V_i^a < \mathbf{E}U_i^a \mathbf{E}V_i^b$. This is a realistic assumption for distributions with different means. A classification threshold, τ , is then picked such that

$$\frac{\mathbf{E}U_i^b}{\mathbf{E}U_i^a} < \tau < \frac{\mathbf{E}V_i^b}{\mathbf{E}V_i^a}. \quad (17)$$

An interesting example of (16) is the case in which $\mathbf{E}U_i^b = \mathbf{E}V_i^a$ and $\mathbf{E}U_i^a = \mathbf{E}V_i^b$ and where all terms do not equal 1. In this instance, (8) in the benign case will be the inverse of that in the adversarial case. Furthermore, neither ratio will equal 1. This happens when adversarial distributions are simply shifts of benign distributions.

6. Experimental details

6.1. Architectures

We experiment with five models, two of which are detailed in Tables 2-3 while the other three are VGG-19, InceptionResnetV2, and MobileNet. The last layers of VGG-19, InceptionResnetV2, and MobileNet are preloaded from Keras, and their last layers are replaced with three custom, fully-connected layers, with output sizes 4096, 1000, and 10, respectively, and trained with ImageNet weights. With respect to the last three models, we only compute graph-based quantities from these layers. For models 1 and 2, we use all layers.

6.2. Datasets

We trained our models on MNIST, CIFAR-10, and SVHN datasets. For each model we created adversarial examples using the Adversarial Robustness Toolbox [27]. For CIFAR-10 and SVHN, all images were enlarged to (224, 224, 3). Images were preprocessed using built-in Keras layers that handle input preprocessing.

Table 2: Architecture of Model 1

LAYER TYPE	OUTPUT SIZE	ACTIVATION FUNCTION
FULLY CONNECTED	300	ReLU
FULLY CONNECTED	200	ReLU
FULLY CONNECTED	150	ReLU
FULLY CONNECTED	150	ReLU
FULLY CONNECTED	100	SIGMOID
FULLY CONNECTED	10	SOFTMAX

Table 3: Architecture of Model 2

LAYER TYPE	OUTPUT SIZE	ACTIVATION FUNCTION
CONV	3 FILTERS, KERNEL SIZE (4,4)	IDENTITY
MAXPOOL	POOL SIZE=(2,2), STRIDES=(2,2)	ReLU
CONV	3 FILTERS, KERNEL SIZE (4,4)	IDENTITY
MAXPOOL	POOL SIZE=(2,2), STRIDES=(2,2)	ReLU
FULLY CONNECTED	100	ReLU
FULLY CONNECTED	10	SOFTMAX

6.3. Attacks

We consider fast gradient sign method, [14], projected gradient descent [22], untargeted Carlini-Wagner L2 [5], DeepFool [25], Square [1], and Auto [10] attacks. Fast gradient sign method attacks were clipped when perturbations were outside a ball of radius 10% in the ℓ^∞ norm. Projected gradient descent attacks were crafted using the same norm but with up to a 5% perturbation; the number of iterations was 40 except for InceptionResnetV2, MobileNet, and VGG19, in which 10 were used. Square and Auto attacks had the same norm and perturbation as projected gradient descent attacks. Optimization was done using ADAM with learning rate 0.01. For each attack we generated 10,000 adversarially perturbed images from 10,000 original

(test data) images. In creating training data for the detection methods we introduce, approximately 14,000 samples were used, and the methods were compared on approximately 6,000 samples. For RSA the numbers are approximately the same. For LID, we used approximately 6,000 training and test samples each, with the exception of models 1 and 2 in which we used approximately 7,000 training and 3,000 test samples.

6.4. Hyperparameters

The values of ϵ and α in our implementation of LRP- $\alpha\beta$ are 2 and 10^{-7} , respectively. In our implementation of RSA we use $M = 8$, $K = 16$, and the layer used is the third from the output layer. For creating noisy samples in the algorithm in LID, we use Gaussian noise of zero mean and variance 0.05. Also in our implementation of LID, we only use the last 10 layers for computational ease.

7. Results and discussion

7.1. Comparison of logistic regression approaches

In Tables 4a, 4b, and 4c we report the specificity (percentage benign samples that are correctly detected) and sensitivity (percentage adversarial samples that are correctly detected). One can see that the various graph statistics considered here can be strong sensitive and specific predictors of adversarial attacks in the case of using logistic regression. Among Mobilenet, Inception-ResnetV2 and VGG19, degree seems to slightly be the best predictor among our statistics. From the tables, we see that the worst performance occurs for Carlini-Wagner and Deepfool attacks. These two attacks are known to be among the most difficult to detect, so our results are consistent with this belief. In particular, for VGG19 and Carlini-Wagner, our classifier is able to almost always detect benign samples, but detects almost no adversarial examples.

Among models 1 and 2, degree is significantly the best predictor, while edge relevance for Model 2 is a poor predictor across all attacks, being unable to detect adversarial images. This is because the edge relevance for benign and adversarial samples are equal to 0. The largest relevances for Model 2 are found in layers closer to the input layer. During the thresholding process, the relevances for the edges corresponding to the output layer are set to 0 because they are relatively small. Lastly, in comparison to LID, our results are superior across almost all model/attack combinations.

		Attack						
		FGSM	PGD	CW2	DF	Sq	AA	
model	MobileNet	degree	99.64/98.10	99.63/99.23	82.21/90.74	80.23/91.34	94.08/93.22	100/99.63
		node importance	99.52/99.35	99.46/100	64.50/99.43	66.78/93.91	91.86/93.39	99.89/100
		edge relevance	100/99.98	99.71/99.61	85.27/90.06	87.75/89.24	100/99.85	100/99.86
		LID	23.49/85.93	12.01/97.90	22.30/79.84	35.59/68.10	24.47/98.60	7.02/87.35
model	InceptionResNetV2	degree	100/99.95	100/99.92	84.05/94.31	73.88/86.22	96.18/98.50	99.66/100
		node importance	100/100	100/100	65.76/88.13	44.12/96.08	89.30/99.36	100/99.96
		edge relevance	99.96/99.70	100/99.70	80.36/78.76	80.96/74.43	99.21/96.59	100/99.78
		LID	40.13/66.23	76.58/44.09	21.08/73.91	72.97/28.78	67.79/41.59	31.60/75.93
model	VGG19	degree	100/99.86	99.92/99.73	99.77/04.20	97.55/99.08	98.98/98.21	99.16/100
		node importance	99.96/100	99.96/100	100/0.7	94.88/99.35	98.24/98.28	97.97/100
		edge relevance	99.73/99.72	99.96/99.96	100/0	99.66/96.82	99.66/97.21	100/99.92
		LID	79.66/6.57	99.39/42.71	41.77/57.55	99.73/0.88	98.78/0.11	22.14/76.08

(a) Cifar-10

		Attack						
		FGSM	PGD	CW2	DF	Sq	AA	
model	MobileNet	degree	100/100	100/99.77	77.25/91.87	55.44/92.96	99.66/99.04	100/100
		node importance	100/100	99.66/100	55.61/90.53	54.39/88.69	89.30/99.36	100/99.96
		edge relevance	100/99.79	100/99.85	99.59/76.24	79.85/74.90	99.21/96.96	100/99.81
		LID	57.12/40.29	69.78/37.14	77.47/23.03	92.97/7.03	88.45/7.71	85.23/37.52
model	InceptionResNetV2	degree	99.88/99.84	98.63/100	76.00/91.52	75.29/92.63	87.35/97.11	99.41/100
		node importance	100/100	100/99.96	45.68/93.81	89.52/94.36	89.52/94.36	100/100
		edge relevance	99.96/99.68	100/99.60	74.06/76.25	92.95/90.58	92.95/90.58	100/99.92
		LID	92.14/4.10	69.78/37.14	77.17/23.02	53.32/48.55	29.87/59.27	10.75/61.80
model	VGG19	degree	100/99.93	100/100	97.67/29.68	98.80/99.18	96.97/99.57	99.96/100
		node importance	100/99.94	100/99.96	98.84/9.68	99.34/99.39	98.64/99.73	100/100
		edge relevance	100/99.88	100/99.81	100/0	99.96/98.37	100/97.83	100/99.92
		LID	67.79/40.38	2.26/85.83	42.24/47.81	4.14/92.37	7.01/98.72	6.47/29.00

(b) SVHN

		Attack						
		FGSM	PGD	CW2	DF	Sq	AA	
model	Model 2	degree	99.00/95.45	99.90/95.45	98.05/99.19	97.26/99.20	99.86/65.44	99.97/95.34
		node importance	84.80/24.54	92.79/12.46	52.75/62.37	34.22/73.53	97.71/5.53	94.50/8.04
		edge relevance	100/0	100/0	100/0	100/0	100/0	100/0
		LID	81.94/13.43	88.19/10.41	74.90/21.72	71.79/22.27	52.77/50.96	16.10/87.53
model	Model 1	degree	95.96/98.91	94.76/84.19	6.58/88.76	95.51/96.71	98.63/71.97	96.02/84.48
		node importance	81.67/96.19	96.09/54.83	0.14/99.70	88.14/96.57	100/2.09	93.87/60.39
		edge relevance	90.34/86.50	89.31/85.25	49.55/42.19	95.68/93.58	100/0	89.47/88.57
		LID	50.60/49.75	72.07/25.38	47.20/55.53	71.79/22.27	52.77/50.96	26.71/72.12

(c) MNIST

Table 4: Comparison between logistic regression methods. First and second quantities in each entry are benign and adversarial detection rate, respectively. FGSM, PGD, CW2, DF, Sq, and AA represent fast gradient sign method, projected gradient descent, Carlini-Wagner L2, Square, and Auto attacks, respectively. Values are percentages.

7.2. Comparison of statistical approaches

Tables 5a, 5b, and 5c show results in terms of AUROC (area under receiver operating characteristic curve) for various detection methods. In almost all cases, WSR2 provides more accurate predictions than WSR1. Fur-

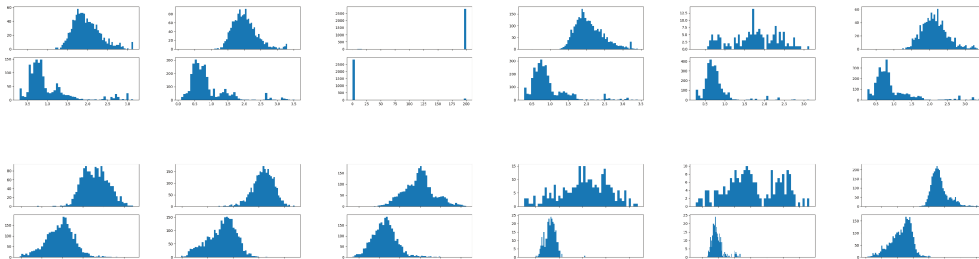


Figure 1: Empirical distributions for WSR1 for Model 2 (top row) and WSR2 for InceptionResnetV2 (bottom row). The top panel of each subplot shows WSR computed for adversarial examples, and the bottom subplot shows the computation for benign examples. FGSM, PGD, CW2, DF, Sq, and AA represent fast gradient sign method, projected gradient descent, Carlini-Wagner L2, Deepfool, Square, and Auto attacks, respectively. For Model 2 and CW2, values above 200 are set to 200 for ease of display. Note that the benign and adversarial plots for Model 2 tend to agree with the remark made in section 1 about inverses.

ther, both WSR variations outperform RSA. Model 1, in comparison to the other models, performs somewhat poorly under WSR1. This seems to be due to Model 1 having the least number of neurons, making the corresponding $|\mathcal{S}|$ relatively small. On this note, we can also see from the tables that model/attack pairs with small $|\mathcal{S}|$ tend to have worse results under WSR. This is particularly noticeable in the case of Carlini-Wagner and Deepfool attacks under WSR1; this lower performance was also noted in our results using logistic regression.

We can use WSR and logistic regression in a complementary way. For instance, graph-based quantities generated from VGG19 and Carlini-Wagner attacks tend to be poorly classified with logistic regression. In contrast, WSR2 performs well in this case, and it can be used in place of logistic regression.

We considered using equation 5 from [29] as a baseline, perhaps in place of RSA, but chose not to because of the extremely large time needed for the source code to run, and secondly, our initial results suggested that this method gives poor accuracy, near 50%, which is much lower than the numbers the article states. In our effort to increase accuracy we experimented with

different hyperparameters, including noise, but to no avail. This calls into question the usefulness and robustness of using equation 5 in [29].

7.3. Nodal analysis

The distributions of node quantities is highly dependent on the model and attack. From the tables it can be seen that AUROC for WSR decreases as the strength of the attack increases (we consider a partial order of increasing attack strength to be: Fast Gradient Sign Method, Projected Gradient Descent, and Carlini-Wagner L2). We can relate this observation to how the cardinality of $|\mathcal{S}|$ varies with model/attack. The cardinality of $|\mathcal{S}|$ can be seen in Tables 6a,6b, and 6c. For CIFAR-10 and SVHN datasets, we observe that the cardinality tends to be a lot smaller for Carlini-Wagner L2 and DeepFool attacks, and it seems to explain the lower accuracy achieved by WSR on these attacks. We recall that from section 5, the accuracy of WSR increases with $|\mathcal{S}|$.

We also note that in some cases the benign distribution of WSR and the adversarial distribution of WSR are centered at points which are close to inverses. This seems to be the case for Model 2, as shown in figure 1. This is in agreement with an earlier remark in Section 5 about equation (8) having inverse values under benign and adversarial examples, assuming the benign and adversarial test statistics have the same distributions up to a shift.

		Attack						
		FGSM	PGD	CW2	DF	Sq	AA	
model	MobileNet	WSR1	99.98	100	79.78	73.56	92.98	100
		WSR2	100	94.47	98.00	92.11	99.25	100
		RSA	51.31	68.56	84.94	48.79	97.44	85.74
model	InceptionResNetV2	WSR1	99.64	99.00	81.46	64.23	84.88	100
		WSR2	97.26	99.05	95.69	88.54	89.93	99.98
		RSA	94.67	98.89	99.06	98.59	98.72	95.07
model	VGG19	WSR1	100	99.98	74.83	98.78	97.08	99.98
		WSR2	99.87	99.97	97.36	99.32	99.70	100
		RSA	69.08	54.28	71.70	73.77	76.58	63.78

(a) Cifar-10

		Attack						
		FGSM	PGD	CW2	DF	Sq	AA	
model	MobileNet	WSR1	100	100	81.62	78.95	99.29	100
		WSR2	100	100	95.20	90.26	99.80	100
		RSA	80.12	63.12	87.69	82.83	76.48	82.27
model	InceptionResNetV2	WSR1	100	100	78.35	80.59	92.12	100
		WSR2	99.86	99.91	93.84	94.29	96.81	100
		RSA	51.77	61.81	56.35	60.25	58.44	58.19
model	VGG19	WSR1	100	100	76.33	99.65	99.54	100
		WSR2	100	100	94.57	99.68	99.87	100
		RSA	80.69	57.27	76.79	79.63	81.77	58.23

(b) SVHN

		Attack						
		FGSM	PGD	CW2	DF	Sq	AA	
model	Model 2	WSR1	95.16	95.37	96.48	95.44	91.77	95.54
		WSR2	96.75	96.23	96.73	96.08	91.74	96.68
		RSA	66.81	62.81	58.54	55.95	68.33	63.12
model	Model 1	WSR1	95.53	81.06	41.25	82.48	89.3	83.6
		WSR2	96.36	94.40	39.65	97.81	99.48	94.36
		RSA	71.80	72.10	51.64	89.60	96.34	71.85

(c) MNIST

Table 5: Comparison of AUROC for statistical detection methods. FGSM, PGD, CW2, DF, Sq, and AA represent fast gradient sign method, projected gradient descent, Carlini-Wagner L2, Deepfool, Square, and Auto attacks, respectively. WSR1 and WSR2 are WSR variants 1 and 2 respectively. Values are percentages.

model		Attack					
		FGSM	PGD	CW2	DF	Sq	AA
MobileNet	WSR1	877	734	51	68	405	79
InceptionResnetV2	WSR1	378	484	109	180	240	120
VGG19	WSR1	180	776	34	925	692	578

(a) Cifar-10

model		Attack					
		FGSM	PGD	CW2	DF	Sq	AA
MobileNet	WSR1	422	636	78	97	551	287
InceptionResnetV2	WSR1	754	1269	109	160	922	496
VGG19	WSR1	945	1253	63	1379	929	843

(b) SVHN

model		Attack					
		FGSM	PGD	CW2	DF	Sq	AA
Model 2	WSR1	129	105	50	34	52	115
Model 1	WSR1	33	10	7	5	4	10

(c) MNIST

Table 6: Cardinality of \mathcal{S} by model and attack. FFGSM, PGD, CW2, DF, Sq, and AA represent fast gradient sign method, projected gradient descent, Carlini-Wagner L2, Deepfool, Square, and Auto attacks, respectively.

8. Conclusion

We have demonstrated that neural network architectures can be interpreted in a graph context from which we can use the statistics of graph-based quantities to detect adversarial attacks. We introduced three measures that we applied to our graphs and used them as predictors of adversarial attack. We showed that this approach can produce high detection performances with logistic regression. We also studied the distributions of node degree using a statistical test based on Wasserstein distances. We find it intriguing that a sparse graph encodes sufficient information about inputs to a neural network. We hope that the perspective introduced here will provide a different way of understanding adversarial attacks.

9. Acknowledgments

L. Carboni and D. Nwaigwe are the recipients of a grant from MIAI@Grenoble Alpes (ANR 19-P3IA-003).

References

- [1] Maksym Andriushchenko, Francesco Croce, Nicolas Flammarion, and Matthias Hein. Square attack: A query-efficient black-box adversarial attack via random search. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXIII*, volume 12368 of *Lecture Notes in Computer Science*, pages 484–501. Springer, 2020.
- [2] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE*, 10(7):e0130140, July 2015.
- [3] Alex Bavelas. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.
- [4] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, March 2009.
- [5] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 39–57, 2017.
- [6] Wuxinlin Cheng, Chenhui Deng, Zhiqiang Zhao, Yaohui Cai, Zhiru Zhang, and Zhuo Feng. Spade: A spectral method for black-box adversarial robustness evaluation. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1814–1824. PMLR, 18–24 July 2021.
- [7] Anoop Cherian and Shuchin Aeron. Representation learning via adversarially-contrastive optimal transport. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2020.
- [8] Dan Cireşan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338, 2012. Selected Papers from IJCNN 2011.

- [9] Ciprian A. Corneanu, Meysam Madadi, Sergio Escalera, and Aleix M. Martinez. What does it mean to learn in deep networks? And, how does one detect adversarial attacks? In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4752–4761, 2019.
- [10] Francesco Croce and Matthias Hein. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 2206–2216. PMLR, 2020.
- [11] Nathan Drenkow, Neil Fendley, and Philippe Burlina. Attack agnostic detection of adversarial examples via random subspace analysis. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022, Waikoloa, HI, USA, January 3-8, 2022*, pages 2815–2825. IEEE, 2022.
- [12] Reuben Feinman, Ryan R. Curtin, Saurabh Shintre, and Andrew B. Gardner. Detecting adversarial samples from artifacts. *arXiv preprint ARXIV.1703.00410*, 2017.
- [13] Samuel G. Finlayson, John D. Bowers, Joichi Ito, Jonathan L. Zittrain, Andrew L. Beam, and Isaac S. Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, 2019.
- [14] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *3rd International Conference on Learning Representations*, 2015.
- [15] Chong Guo, Michael Lee, Guillaume Leclerc, Joel Dapello, Yug Rao, Aleksander Madry, and James Dicarlo. Adversarially trained neural representations are already as robust as biological neural representations. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8072–8081. PMLR, 17–23 Jul 2022.
- [16] Paula Harder, Franz-Josef Pfreundt, Margret Keuper, and Janis Keuper. SpectralDefense: Detecting Adversarial Attacks on CNNs in the Fourier

- Domain. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Shenzhen, China, Jul 2021. IEEE.
- [17] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95(2):245–258, July 2017.
- [18] Emanuele La Malfa, Gabriele La Malfa, Claudio Caprioli, Giuseppe Nicosia, and Vito Latora. Deep neural networks as complex networks. *arXiv preprint ARXIV.2209.05488*, 2022.
- [19] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems*, pages 7167–7177, 2018.
- [20] Shiwei Liu, Tim van der Lee, Anil Yaman, Zahra Atashgahi, Davide Ferraro, Ghada Sokar, Mykola Pechenizkiy, and Decebal Constantin Mocanu. Topological insights into sparse neural networks. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part III*, volume 12459 of *Lecture Notes in Computer Science*, pages 279–294. Springer, 2020.
- [21] Xingjun Ma, Bo Li, Yisen Wang, Sarah M. Erfani, Sudanthi N. R. Wijewickrema, Grant Schoenebeck, Dawn Song, Michael E. Houle, and James Bailey. Characterizing adversarial subspaces using local intrinsic dimensionality. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.
- [22] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.
- [23] Decebal Constantin Mocanu, Elena Mocanu, Phuong H. Nguyen, Madeleine Gibescu, and Antonio Liotta. A topological insight into

- restricted Boltzmann machines. *Machine Learning*, 104(2-3):243–270, September 2016.
- [24] Grégoire Montavon, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. *Layer-Wise Relevance Propagation: An Overview*, pages 193–209. Springer International Publishing, Cham, 2019.
- [25] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: A simple and accurate method to fool deep neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2574–2582, 2016.
- [26] Gregory Naitzat, Andrey Zhitnikov, and Lek-Heng Lim. Topology of deep neural networks. *Journal of Machine Learning Research*, 21(1), June 2022.
- [27] Maria-Irina Nicolae, Mathieu Sinn, Minh Ngoc Tran, Beat Buesser, Ambrish Rawat, Martin Wistuba, Valentina Zantedeschi, Nathalie Baracaldo, Bryant Chen, Heiko Ludwig, Ian Molloy, and Ben Edwards. Adversarial robustness toolbox v1.2.0. *CoRR*, 1807.01069, 2018.
- [28] Nicolas Papernot, Patrick D. McDaniel, Ian J. Goodfellow, Somesh Jha, Z. Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, AsiaCCS 2017, Abu Dhabi, United Arab Emirates, April 2-6, 2017*, pages 506–519. ACM, 2017.
- [29] Kevin Roth, Yannic Kilcher, and Thomas Hofmann. The odds are odd: A statistical test for detecting adversarial examples. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5498–5507. PMLR, 2019.
- [30] Eric Wong, Frank R. Schmidt, and J. Zico Kolter. Wasserstein adversarial examples via projected sinkhorn iterations. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6808–6817. PMLR, 2019.

- [31] Han Xu, Yao Ma, Hao-Chen Liu, Debayan Deb, Hui Liu, Ji-Liang Tang, and Anil K. Jain. Adversarial Attacks and Defenses in Images, Graphs and Text: A Review. *International Journal of Automation and Computing*, 17(2):151–178, April 2020.
- [32] Matteo Zambra, Amos Maritan, and Alberto Testolin. Emergence of network motifs in deep neural networks. *Entropy*, 22(2), 2020.

Compliance with Ethical Standards

This research study was conducted retrospectively using human subject data made available by the different provider centers. The use of the datasets was approved by the ethical committee consulted for each study and the concerned participants or next of skin signed the informed consent form.

Published works

Articles and Conference Communications

- L. Carboni, M. Dojat and S. Achard, Nodal statistics-based equivalence relation for graph collections, *Physical Reviews E* **107**, 014302-1 (2023)
- L. Carboni, M. Dojat and S. Achard, “Nodal equivalence relation for brain connectivity analysis”, *XXIXème Colloque Francophone de Traitement du Signal et des Images (GRETSI 2023)*, (2023)
- L. Carboni, S. Achard and M. Dojat, “Network embedding for brain connectivity”, *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, (2021)
- L. Carboni, D. Nwaigwe, M. Mainsant, R. Bayle, M. Reyboz, M. Mermillod, M. Dojat and S. Achard, Exploring Continual Learning Strategies in Artificial Neural Networks through Graph-Based Analysis of Connectivity: Insights from a Brain-Inspired Perspective (2023)
- D. Nwaigwe, L. Carboni, M. Mermillod, S. Achard and M. Dojat, Graph-based methods coupled with specific distributional distances for adversarial attack detection (2023)
- P. Sitoleux, L. Carboni, H. Lbath and S. Achard, Multiscale and multi-density comparison of functional brain networks through label-informed persistence diagrams (2023)

Abstracts

- L. Carboni, S. Achard and M. Dojat, “Différencier les réseaux de connectivité cérébrale fonctionnelle de repos : apport de la combinaison de statistiques nodales de graphes”, *SFRMBM - 6ème congrès de la Société Française de Résonance Magnétique en Biologie et Médecine*, (2023) - Poster presentation
- L. Carboni, S. Achard and M. Dojat, “Nodal-statistics-based structural pattern detection for graph collections characterization”, *CCS22 Conference on Complex Systems, October 2022, Palma de Mallorca, Spain*, (2022) - Poster presentation
- L. Carboni, S. Achard and M. Dojat, “Stationary distribution of node2vec random walk for brain connectivity”, *CCS21 Conference on Complex Systems, October 2021, Lyon, France*, (2021) - Oral presentation
- S. Chiominto, F. Pappone, H. Lbath, L. Carboni and S. Achard, “Homological information to refine small-world regime of brain functional connectivity networks”, *CCS22 Conference on Complex Systems, October 2022, Palma de Mallorca, Spain*, (2022) - Oral presentation

Awards [MOMI22 Best Poster award 2022](#)

Poster title: [Human Brain Functional Network Characterization](#)

Résumé français - French Summary

Chapitre 1: Context and Background

Dans ce chapitre, nous plongeons dans l'analyse des études de connectivité fonctionnelle (CF), en mettant particulièrement l'accent sur les méthodes adoptant un modèle de graphe. Nous mettons en évidence la diversité qui prévaut au sein de ces études, générant des interprétations variées et suscitant des résultats contradictoires. Nous commençons par introduire les concepts clés de la théorie des graphes ainsi que les descripteurs fréquemment utilisés dans les études de connectivité fonctionnelle. Dans ce contexte, nous soulignons l'importance cruciale de l'interprétabilité et de l'explicabilité dans l'analyse des réseaux de connectivité cérébrales. Ces éléments jouent un rôle fondamental pour éclairer les mécanismes physiopathologiques sous-jacents régissant l'activité cérébrale. De plus, nous présentons des modèles de réseau générateurs qui se révèlent être des outils efficaces pour l'évaluation comparative de données réelles par rapport à leurs équivalents synthétiques. Enfin, nous identifions les exigences d'un cadre d'analyse complet de la connectivité cérébrale. En mettant en avant ces aspects essentiels, notre objectif est de tracer la voie vers un cadre global qui harmonise l'analyse de la connectivité cérébrale.

Chapitre 2: Structural Pattern

Ce chapitre introduit un nouveau cadre mathématique pour l'analyse des réseaux de connectivité fonctionnelle. Il débute en fournissant un aperçu des études antérieures qui examinent la réorganisation nodale au sein des réseaux de connectivité fonctionnelle. Ces explorations fournissent la motivation pour la formulation d'un motif structurel de graphe, le *structural pattern* - l'élément central de notre travail. Ce structural pattern nouvellement introduit trouve également des inspirations dans le concept de relations d'équivalence nodales automorphes prévalentes dans les réseaux complexes. À la base de notre proposition réside la définition d'une nouvelle relation d'équivalence basée sur les statistiques nodales, permettant ainsi de combler le fossé entre les statistiques nodales classiques et les outils de réduction dimensionnelle des réseaux. Cette approche innovante permet non seulement la combinaison d'une multitude de statistiques nodales pour l'analyse des réseaux, mais elle permet également la caractérisation des réseaux tant au niveau individuel qu'au niveau de groupe, et au même temps à l'échelle globale et locale.

Chapitre 3: Application

Dans ce chapitre, nous présentons des expériences aboutissant à des résultats clés concernant l'application d'une caractérisation des motifs structurels de graphe dans les réseaux de connectivité cérébrale fonctionnelle. Tout d'abord, nous offrons un bref aperçu des bases de données que nous avons utilisées. Pour assurer la généralisabilité de notre méthode et en démontrer l'adaptabilité, nous avons fait l'effort d'intégrer une large gamme de données. Cependant, comme nous l'avons élucidé dans le chapitre précédent, il est crucial que les données respectent des critères spécifiques pour garantir la robustesse des résultats. Les quatre principales expériences sont organisées comme suit. Tout d'abord, une comparaison entre les données réelles et les modèles génératifs synthétiques est présentée. Cela sert de preuve de concept, illustrant que les motifs structurels de graphe ont la capacité de capturer des informations uniques aux réseaux de connectivité cérébrale. Dans la deuxième expérience, nous détaillons comment nous avons obtenu des résultats cohérents chez des sujets sains provenant de différentes bases de données. Cette démarche aboutit à une caractérisation complète de la connectivité cérébrale saine. Notre troisième expérience concerne une méthode de classification visant à distinguer différentes conditions pathologiques. Cette méthode est efficace même pour discerner des différences subtiles entre des groupes présentant une symptomatologie variable. Enfin, nous explorons l'influence de la longueur des séries temporelles sur la caractérisation régionale. Nous recommandons l'utilisation d'une longue durée d'acquisition pour capturer efficacement les motifs pertinents. Dans l'ensemble, ces résultats mettent en évidence le potentiel de l'utilisation des motifs structurels de graphe dans l'analyse des réseaux de connectivité cérébrale fonctionnelle, englobant la généralisation chez les sujets sains, la discrimination entre les conditions pathologiques et l'importance de la longueur des données dans la caractérisation régionale.

Chapitre 4: Context and Background

Ce chapitre offre un aperçu du contexte, des antécédents et des travaux connexes concernant les applications de la science des réseaux aux réseaux neuronaux artificiels. Il commence par définir formellement les réseaux de neuronaux et introduire la notation adoptée. La section suivante se penche sur l'introduction du paradigme d'apprentissage continu pour les systèmes artificiels et ses implications pour favoriser le développement d'une intelligence artificielle robuste. Dans ce cadre, différents scénarios et stratégies d'apprentissage continu existants sont présentés et catégorisés, en mettant particulièrement l'accent sur leurs inspirations biologiques. Enfin, le chapitre revisite brièvement les applications de la science des réseaux pour la conception de réseaux neuronaux artificiels et met en lumière les recherches existantes qui les caractérisent au moyen de descriptions de la théorie de graphes.

Chapitre 5: Graph-Based Analysis of ANN Connectivity

Dans ce chapitre, nous présentons notre modèle original d'analyse basé sur les graphes pour la connectivité des réseaux neuronaux artificiels. Ce travail vise à combler le fossé entre la neuroscience et les réseaux neuronaux artificiels en introduisant un nouveau cadre de modélisation par graphes, un puissant outil largement utilisé pour étudier la connectivité structurale et fonctionnelle du cerveau. Nous montrons comment notre approche s'inspire du cerveau et introduisons des descripteurs adaptés à l'analyse des réseaux neuronaux artificiels.

En particulier, nous concevons une méthode d'analyse des stratégies d'apprentissage continu en proposant un modèle basé sur les graphes de l'architecture entraînée à chaque session d'apprentissage. Cela pose les bases du développement d'une stratégie d'apprentissage basée sur les caractéristiques des graphes.

À la fin du chapitre, nous discutons des travaux connexes pertinents et des limitations.

Chapitre 6: Continual Learning Strategy Connectivity Characterization

Dans ce chapitre, nous proposons comme cas d'étude la caractérisation de différentes stratégies d'apprentissage dans le cadre de l'apprentissage continu. Plus particulièrement, nous considérons deux tâches de classification différentes et leurs architectures associées. Nous réalisons trois expériences pour évaluer en profondeur la capacité du cadre basé sur les graphes à reconnaître les différents comportements néfastes résultant de diverses stratégies d'apprentissage dans des scénarios d'apprentissage continus. De plus, nous introduisons la notion d'ensemble d'unités critiques en termes de plasticité/stabilité et concevons l'expérience d'élagage pour obtenir une analyse fine au niveau de l'unité des stratégies d'apprentissage considérées. Notre modèle de graphes parvient à identifier les unités dont les poids changent légèrement en norme au fil des sessions d'apprentissage, sans nécessiter de traitement des données d'entrée.

