# UNCOVERING DIFFERENTIAL IDENTIFIABILITY IN NETWORK PROPERTIES OF HUMAN BRAIN FUNCTIONAL CONNECTOMES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Meenusree Rajapandian

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2020

Purdue University

West Lafayette, Indiana

# THE PURDUE UNIVERSITY GRADUATE SCHOOL
# STATEMENT OF DISSERTATION APPROVAL

Dr. Joaquín Goñi, Chair

School of Industrial Engineering

Weldon School of Biomedical Engineering

Dr. Mario Ventresca

School of Industrial Engineering

Dr. Thomas M. Talavage

School of Electrical & Computer Engineering

Weldon School of Biomedical Engineering

**Approved by:**

Dr. Abhijit Deshmukh

Head of the School Graduate Program

Dedicated to *CONN*plexity Lab and its members.

# ACKNOWLEDGMENTS

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. JG acknowledges financial support from NIH R01EB022574, NIH R01MH108467, Indiana Alcohol Research Center P60AA07611, Purdue Discovery Park Data Science Award "Fingerprints of the Human Brain: A Data Science Perspective". Authors thank Dr. Gorka Zamora-Lopez and Dr. Matthieu Gilson for useful comments.

TABLE OF CONTENTS

Page

LIST OF FIGURES . . . . . vi

SYMBOLS . . . . . xi

ABBREVIATIONS . . . . . xii

ABSTRACT . . . . . xiii

1 INTRODUCTION . . . . . 1

2 DATA . . . . . 3

3 NETWORK SCIENCE . . . . . 5

4 FINGERPRINT IN BRAIN CONNECTOMICS . . . . . 12

5 THE DIFFERENTIAL IDENTIFIABILITY FRAMEWORK . . . . . 15

6 AN EXTENSION OF THE FRAMEWORK . . . . . 20

7 RESULTS . . . . . 23

8 DISCUSSION . . . . . 33

9 POSTHOC ANALYSIS . . . . . 37

10 SUMMARY . . . . . 41

REFERENCES . . . . . 43

# LIST OF FIGURES

Figure Page

3.1 The connectome as a matrix. (a) One of the first efforts to systematically generate a connectivity matrix for the brain [44]. This matrix represents the connectivity of 32 neocortical areas involved in visual function in the macaque monkey, constructed by collating the results of a large number of published tract-tracing studies in this animal. In this matrix, a black cross indicates an outgoing projection from the region listed in the row to the region listed in the column. (b) An updated connectivity matrix of the macaque comprising 39 cortical areas, as reconstructed from an online database of tract-tracing studies. This matrix is organized such that colored elements represent a projection from the region listed in the column to the region listed in the row (see Chapter 3). The size of the dots in each matrix element is proportional to the projection distance and darker colors indicate stronger average reported connectivity strength. (c) The anatomical locations of the areas listed in the matrix in (b). Darker colors identify regions with higher total connectivity to the rest of the network. [2] . . . . . 6

3.2 Part of a network that shows the shortest path between nodes $i$ and $j$. On the left, highlighted in blue arrows is the shortest path in a network. On the right shows shortest path and the number of possible exits at each node on the path highlighted in dotted blue arrows . . . . . 8

3.3 Network showing shortest path and the mean first passage time between nodes $i$ and $j$. One left, path in brown is the shortest path between nodes $i$ and $j$. On the right shows a possible path taken by a random walker starting from node $i$ to a target node $j$. . . . . 9

3.4 In this figure, the node $A$ has the highest betweenness centrality since the shortest path from any blue node to red node (and vice versa) will include the node $A$ . . . . . 11

Figure Page

5.1 Workflow scheme of the Identifiability Framework $\mathbb{I}f$. The upper triangular of each functional connectivity matrix (two FCs per subject, test-retest) is vectorized and added to a matrix where columns are sessions and rows are their vectorized functional connectivity patterns. Data are first centered: this is obtained by subtracting the mean $\mu_k$ from each column (where k goes from 1 to N subjects). Second, the PCA algorithm extracts the M principal components (i.e. the M functional connectivity modes with the corresponding M eigenvectors) associated to the whole population and their relative weights across subjects. The M orthogonal connectivity modes are then used to reconstruct back the FC of each subject (k is added back to the data). Colorbars indicate positive (yellow) to negative (blue) connectivity values: Pearson's correlation coefficient in the case of individual FC matrices (left and right sides of scheme), and unitless connectivity weights in the case of PCA FC-modes. [32] . . . . . 16

5.2 **Differential Identifiability ($I_{diff}$) profiles of network properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Each plot shows, for each fMRI task, the $I_{diff}$ score associated with functional connectivity *(blue solid line)* and the $I_{diff}$ score of the original functional connectome *(red solid line)*. . . . . 17

5.3 Optimal differential identifiability ($I_{diff}$) as a function of the number of fMRI volumes used for reconstruction on REST1. Bottom: plot shows the optimal average $I_{diff}$ across 100 runs, as a function of the number of fMRI volumes (scanning lengths) used for FC evaluation and subsequent reconstruction. Red line with circles denotes the average $I_{diff}$ for the original FCs, whereas blue line with circles denotes the identifiability for reconstructed FCs based on the different number of fMRI volumes retained (blue vertical bars indicate standard deviation across runs). Top: insets show the $I_{diff}$ curves (blue line, reconstructed $I_{diff}$; red line, original $I_{diff}$) per optimal number of PCA components for three different choices of number of fMRI frames retained (50, 500 and 1,000 fMRI volumes respectively). [32] . . . . . 18

5.4 Identifiability matrices (I) of the original (Orig) and reconstructed (Recon) data for the Training, (A) and (C), and Validation, (B) and (D) sets of resting-state functional connectomes without global signal regression (NoGSR; (A) and (B)) and with global signal regression (GSR; (C) and (D)).19

6.1 $NP(\mathbb{I}f\{FC\})$ Workflow scheme of the Identifiability Framework $\mathbb{I}f$ as extended in this work. The Functional connectomes are reconstructed from the original functional connectomes as described in **??** (blue-green color scheme). The network properties are then derived from the reconstructed functional connectomes. (pink-yellow color scheme) . . . . . 21

Figure Page

6.2 $\mathbb{I}f\{NP(FC)\}$ Workflow scheme of the Identifiability Framework $\mathbb{I}f$ as extended in this work. Here, the network properties are derived on the original functional connectomes. The network properties are then vectorized (either upper triangular of the matrix or the entire matrix depending on the type of network property) and added to matrix where columns are sessions and rows are the network property values. The PCA algorithm now extracts M principal components from this matrix which contains the network property data. The M orthogonal connectivity modes are then used to reconstruct back the derived network property corresponding to each subject. . . . . 22

7.1 $NP(\mathbb{I}f\{FC\})$ **Differential Identifiability ($I_{diff}$) profiles of pairwise properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Here, the Identifiability framework was applied on the functional connectomes ($\mathbb{I}f\{FC\}$). Each plot shows, for each fMRI task, the $I_{diff}$ score associated with functional connectivity *(red solid line)* and the $I_{diff}$ scores on network properties derived from the reconstructed functional connectomes, $NP(\mathbb{I}f\{FC\})$ *(see legend)* for different number of components. . . . . 25

7.2 $\mathbb{I}f\{NP(FC)\}$ **Differential Identifiability ($I_{diff}$) profiles of pairwise properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Here, the Identifiability framework was applied directly on the network properties derived from the original functional connectomes ($\mathbb{I}f\{NP(FC)\}$). Each plot shows, for each fMRI task, the $I_{diff}$ score associated with functional connectivity *(red solid line)* and the $I_{diff}$ scores on reconstructed network properties derived from the original functional connectomes, $\mathbb{I}f\{NP(FC)\}$ *(see legend)* for different number of components. . . . . 26

7.3 A summary of maximum $I_{diff}$ values, and corresponding number of components for each fMRI task and network property for both $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$. Each plot shows, for each property and each method - $NP(\mathbb{I}f\{FC\})$ or $\mathbb{I}f\{NP(FC)\}$, the $I_{diff}$ score for all tasks. The number mentioned gives the maximum $I_{diff}$ score for the corresponding task *(y axis)* and the position denotes the number of components *(x axis)*. . . . . 28

7.4 A summary of maximum $I_{diff}$ values, corresponding number of components and explained variance retained for each fMRI task and network property for both $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$. . . . . 29

Figure Page

7.5 (A) Across tasks and rest differential Identifiability ($I_{diff}$) for Mean First Passage Time as a function of the number of principal components used for reconstruction. *Solid line* and *solid shaded area* represent the results for $MFPT(\mathbb{I}f\{FC\})$. *Dashed line* and *hatched area* show results for $\mathbb{I}f\{MFPT(FC)\}$ (B) Across tasks and rest differential Identifiability ($I_{diff}$) for Search Information as a function of the number of principal components used for reconstruction. *Solid line* and *solid shaded area* represent the results for $SI(\mathbb{I}f\{FC\})$. *Dashed line* and *hatched area* show results for $\mathbb{I}f\{SI(FC)\}$ The differential identifiability matrix (as defined in Methods) is shown at optimal reconstruction for Language task for (C) $MFPT(\mathbb{I}f\{FC\})$, (D) $\mathbb{I}f\{FC\}$ and (E) $\mathbb{I}f\{SI(FC)\}$. The diagonal elements in each matrix represent $I_{self}$ and the non-diagonal elements represent $I_{others}$. . . . . 30

7.6 $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$ **Differential Identifiability ($I_{diff}$) of node properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Each plot shows, for each task, the $I_{diff}$ score associated with functional connectivity *(red solid line)*, the $I_{diff}$ scores on the network properties derived from the reconstructed functional connectomes $NP(\mathbb{I}f\{FC\})$ *(solid lines, colors - see legend)* and the $I_{diff}$ scores on the reconstructed network properties derived from the original functional connectomes $\mathbb{I}f\{NP(FC)\}$ *(dotted lines, colors - see legend)* for different number of components. . . . . 31

7.7 Effect of $\mathbb{I}f$ on task sensitivity of network measures. For each pairwise network property, task sensitivity is measured using ICC between - $NP(\mathbb{I}f\{FC\})$ vs $NP(FC)$ (row a), $\mathbb{I}f\{NP(FC)\}$ vs $NP(FC)$ (row b) and $NP(\mathbb{I}f\{FC\})$ vs $\mathbb{I}f\{NP(FC)\}$ (row c). First two rows highlight the fact that the $\mathbb{I}f$ framework uncovers the inherently distinct signature of different tasks through derived network properties. The last row shows that certain network properties would benefit more from application of the $\mathbb{I}f$ framework on the functional connectomes, while others from application directly on the network properties. . . . . 32

9.1 Each figure shows for each task, the optimal differential identifiability ($I_{diff}$) that can be uncovered using the Identifiability Framework $\mathbb{I}f$ when different levels of noise are added to the Functional connectomes. Different levels of noise were assessed (horizontal axes on all figures) The shaded areas represent results within the 2.5 and 97.5 percentiles across repetitions. 38

Figure Page

9.2 Each figures shows for each task, the optimal number of principal components required to uncover optimal identifiability using Identifiability Framework $\mathbb{I}f$ when different levels of noise is added to the Functional connectomes (xaxis). The shaded region represents the 2.5 and 97.5 percentile across repetitions. . . . . 39

9.3 Identifiability matrices of Rest. On the left, is the Identifiability matrix of the Rest original functional connectomes, the center is the Identifiability matrix of optimally reconstructed functional connectomes and on the right is the identifiability matrix of functional connectomes reconstructed from noisy original data. Here the a noise level of 0.5 is added to the FCs of Rest.40

# SYMBOLS

| | |
|---|---|
| $\mathbb{I}f$ | Identifiability Framework |
| $G$ | undirected weighted graph |
| $V$ | set of vertices in a graph $G$ |
| $\mathcal{W}$ | is matrix $[w_{ij}]$ where $w_{ij}$ strength of the edge between nodes $v_i$ and $v_j$ |
| $K_i$ | degree of a node $i$ |
| $SPL_{ij}$ | Shortest Path Length between nodes $i$ and $j$ |
| $MFPT_{ij}$ | Mean First Passage Time starting from node $i$ to $j$ |
| $C_{ij}$ | Communicability between nodes $i$ and $j$ |
| $W_{ij}$ | Driftness between nodes $i$ and $j$ |
| $\Omega_{i\leftrightarrow j}$ | sequence of nodes forming the shortest path between nodes $i$, $j$ |
| $\pi_{i\leftrightarrow j}$ | $\{w_{ix}, w_{xy}, \ldots, w_{zj}\}$ sequence of edge weights forming the shortest path between nodes $i$ and $j$ |
| $SI_{ij}$ | information required to follow the shortest path between nodes $i$ and $j$ |
| $P$ | Transition probability matrix |
| $Z$ | Fundamental Matrix |

# ABBREVIATIONS

| | |
|---|---|
| FC | Functional Connectome |
| fMRI | Functional Magnetic Resonance Imaging |
| HCP | Human Connectome Project |
| ICC | Intraclass Correlation Coefficient |
| NP | Network Property |

# ABSTRACT

Rajapandian, Meenusree M.S., Purdue University, May 2020. Uncovering Differential Identifiability in Network Properties of Human Brain Functional Connectomes. Major Professor: Joaquín Goñi.

The Identifiability Framework ($\mathbb{I}f$) has been shown to improve differential identifiability (reliability across-sessions and -sites, and differentiability across-subjects) of functional connectomes for a variety of fMRI tasks. But having a robust single session/subject functional connectome is just the starting point to subsequently assess network properties for characterizing properties of integration, segregation and communicability, among others. Naturally, one wonders if uncovering identifiability at the connectome level also uncovers identifiability on the derived network properties. This also raises the question of where to apply the $\mathbb{I}f$ framework: on the connectivity data or directly on each network measurement? Our work answers these questions by exploring the differential identifiability profiles of network measures when $\mathbb{I}f$ is applied on 1) the functional connectomes, and 2) directly on derived network measurements.

Results show that improving across-session reliability of FCs also improves reliability of derived network measures. We also find that, for specific network properties, application of $\mathbb{I}f$ directly on network properties is more effective. Finally, we discover that applying the framework, either way, increases task sensitivity of network properties. At a time when the neuroscientific community is moving away from group-average statistics towards subject-level inferences, we have shown that $\mathbb{I}f$ is a useful tool to enhance robustness in FC fingerprints, which permeates to derived network properties as well.

# 1. INTRODUCTION

The analysis of structural and functional human brain connectivity based on network science has become prevalent for understanding the underlying mechanisms of the human brain. Using network properties, we are able to understand the topology of brain connectivity patterns [1–3], integration and segregation [4–8], as well as communication dynamics [9–12] and association between human cognition and brain function [13–17]. Until recently, many brain connectivity studies used group-level comparisons, where data from many subjects are collapsed (e.g. group averaging) into a representative sample of clinical and healthy population [18–20]. However, this comes at a price of potentially ignoring intra-group individual variability [21].

Detecting individual differences in functional connectivity profiles thus becomes important, when associating connectivity profiles with individual behavioral outcomes. In recent years, publicly available functional connectome datasets [22, 23] with large sample sizes have enabled the scientific community to account for inter-individual variability in the human functional connectome (FC). A number of promising methods that can successfully capture these individual differences have been established in recent times [21, 24–27]. For instance, work by [28] has shown the existence of a recurrent and reproducible fingerprint in functional connectomes estimated from neuroimaging data. This idea has been extended to maximize or minimize subject-specific and/or task-specific information [29, 30]. These subject-specific fingerprints have been used to track fluctuations in attention at the individual level [31].

The "Identifiability Framework" [32], based on the group-level Principal Component Analysis of functional connectomes that maximizes differential identifiability, has been shown to improve functional connectome fingerprints within- and across-sites, for a variety of fMRI tasks, over a wide range of scanning length , and with and

without global signal regression [32, 33]. Additionally, it has been shown that maximising differential identifiability on the functional connectomes provides more robust and reliable associations with cognition [34] as well as with disease progression [35]. The natural next step is to assess the impact of such a procedure on subsequent network measurements that characterize topological and communication properties of functional brain networks.

An open question of great relevance for the Brain Connectomics community is how to measure and uncover subject fingerprints in network measurements of functional connectivity. Uncovering reliable connectivity fingerprints is crucial when assessing clinical populations and when ultimately mapping cognitive characteristics into connectivity [35–37]. Our hypothesis is that improvement in FC fingerprint should also "propagate" to network derived measurements. An organic way of assessing this would be to track differential identifiability scores of derived network features as the differential identifiability on the functional connectomes changes. One could also proceed with the application of the Identifiability Framework directly on the network derived features as opposed to using it on FCs. The above mentioned approaches rely on different principles of what is a fingerprint in a network derived measurement. The first one assumes that functional connectivity data is "holding" the fingerprints and propagating them to any network derived measurement. The second one considers functional connectivity data as a proxy to ultimately estimate a network measurement with a potentially prominent subject fingerprint.

## 2. DATA

The fMRI dataset used in this study is from the publicly available Human Connectome Project (HCP), Release Q3. Per HCP protocol, written informed consent was obtained from all subjects by the HCP Consortium. Full description of the acquisition protocol and processing steps is given below.

We assessed the 100 unrelated subjects (54 females, 46 males, with a mean age of $29.1 \pm 3.7$ years) data subset as provided at the HCP 900 subjects data release [22,38]. This subset of subjects was chosen from the overall dataset to ensure that no two subjects are family relatives to exclude family-structure co-variables and possible identifiability confounds in our analysis. The fMRI resting-state runs were acquired in separate sessions on two different days, with two different acquisitions (left to right or LR and right to left or RL) per day [39, 40]. The seven fMRI tasks were gambling, relational, social, working memory, motor, language, and emotion. The gambling, working memory, and motor tasks were acquired on the first day while relational, social, language, and emotion tasks were acquired on the second day [22, 41]. The HCP scanning protocol was approved by the local Institutional Review Board at Washington University in St. Louis. All experiments were performed in accordance with relevant guidelines and regulations. For all sessions, data from both the left-right (LR) and right-left (RL) phase-encoding runs were used to calculate connectivity matrices. These acquisitions are referred to as test and retest (or visits). In order to avoid confounds between test-retest and phase encoding, runs were evenly distributed on test-retest along the subjects. Hence, for half of the subjects LR was used as test and RL as retest and for the other half RL was used as test and LR as retest. This operation was done for all 7 fMRI tasks. For resting-state, this procedure was done

for both REST1 and REST2 separately. Full details on the HCP dataset have been published previously [22, 39, 40].

A cortical parcellation of 360 brain regions as recently proposed by Glasser et al [42] was employed. This parcellation is multimodal and took an objective semi-automated neuroanatomical approach to identify areas bounded by sharp change in cortical architecture and function connectivity. For completeness, 14 sub-cortical regions were added, as provided by the HCP release (filename Atlas_ROI2.nii.gz); for which this file was converted from NIFTI to CIFTI format by using the HCP workbench software [39] (http://www.humanconnectome.org/software/connectome-workbench.html, command -cifti-create-label).

The minimal preprocessing pipeline from the HCP was used to process the data. [39] This pipeline included artifact removal, motion correction, and registration to standard space. Full details on this pipeline can be found in [39, 40]. The main steps were spatial (minimal) preprocessing in volumetric and grayordinate space (i.e. where brain regions are mapped onto the native mesh cortical surface) [40]; slice-timing correction; weak high-pass temporal filtering (2000s full width at half maximum) applied to volumetric and grayordinate forms, in effect removing linear trends in the data (no low pass filtering was applied in this pipeline); MELODIC ICA [43] applied to volumetric data; and using FIX to identify artifact components. Artifacts and motion related time courses were regressed out (i.e. the six rigid-body parameter time series, their backwards-looking first differences, and the squares of all 12 resulting regressors) of volumetric and grayordinate data. [40]

# 3. NETWORK SCIENCE

The discovery of general properties in terms of network organization across superficially different systems has enabled the formation of an interdisciplinary field of network science around the use of general analytic methods to model complex networks. This has allowed us to explore the scope of common or near-universal principles of network organization, function, growth, and evolution across a variety of real-world complex systems. Principal among these general methods is graph theory.

The first applications of graph theory to neuroscientific data were published as early as the end of the twentieth century [44–46] as can be seen in 3.1. The use of graph theory to model neural connectivity on a microscopic level was paralleled by attempts to understand macroscopic networks of interconnected cortical areas roughly paralleled Ramón's work. Network diagrams that summarized white matter connections between cortical areas were drawn by clinical pioneers like Theodor Meynert, Carl Wernicke, and Ludwig Lichtheim. These were able to successfully explain symptoms of brain disorder and could connect them to pathological lesions. These models were able to link the source of specific cognitive disabilities to anomalies in the connectivity of the cortical areas [47].

The development of such models and statistical techniques that allow for valid inference on group differences between general and clinical population has paved the foundation for connectomics. A combination of rapid growth in the science of networks and technological evolution of methods to measure and visualize brain organization, across multiple scales of resolution has facilitated the flourishment of the science of connectomics. A consistent conceptual focus on quantifying, visualizing, and understanding brain network organization across multiple scales of space and time is a fundamental characteristic of the burgeoning field of connectomics. [1]

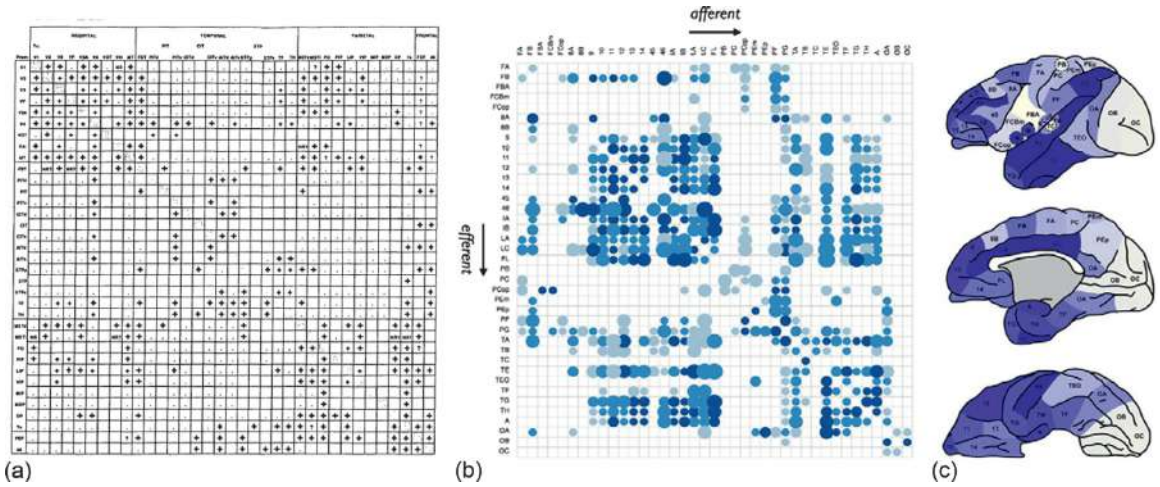


Figure 3.1. The connectome as a matrix. (a) One of the first efforts to systematically generate a connectivity matrix for the brain [44]. This matrix represents the connectivity of 32 neocortical areas involved in visual function in the macaque monkey, constructed by collating the results of a large number of published tract-tracing studies in this animal. In this matrix, a black cross indicates an outgoing projection from the region listed in the row to the region listed in the column. (b) An updated connectivity matrix of the macaque comprising 39 cortical areas, as reconstructed from an online database of tract-tracing studies. This matrix is organized such that colored elements represent a projection from the region listed in the column to the region listed in the row (see Chapter 3). The size of the dots in each matrix element is proportional to the projection distance and darker colors indicate stronger average reported connectivity strength. (c) The anatomical locations of the areas listed in the matrix in (b). Darker colors identify regions with higher total connectivity to the rest of the network. [2]

Such computational models of large-scale brain network dynamics also allow us to assess the efficacy of invasive and noninvasive brain simulation therapies. These findings suggest that an understanding of network topology may allow us to predict expected levels of impairment and prospects for recovery following insult [18], and to select individually tailored interventions with maximum chances of therapeutic success [1, 2, 48].

In this work, we select a set of diverse graph theoretic measurements that are widely used to understand the network topology, communication, integration and

segregation of the brain networks. These properties are degree strength, shortest path length, search information, mean first passage time, driftness, communicability, clustering coefficient and betweenness centrality. As described above, a functional connectome can be represented as a symmetric square correlation matrix that may be seen as an undirected weighted graph. Let $G = (V, \mathcal{W})$ be an undirected weighted graph with set of nodes $V = \{v_1, v_2, \ldots, v_n\}$ and weights $\mathcal{W} = [w_{ij}]$ where $w_{ij}$ is the strength of the edge between nodes $v_i$ and $v_j$.

1. Degree Strength

   The degree strength of a node ($K_i$) in an undirected binary graph is the number of edges that are connected to the node. Here, we consider the weighted sum of the edges connected to the node $i$.

   $$K_i = \sum_{j=1}^{n} w_{ij}$$

2. Shortest Path Length

   The shortest path length (SPL) between two nodes of an undirected graph is defined as the minimum number of edges (and thus steps) that separate the two nodes. For an undirected weighted graph it is the path that results in the smallest value of the sum of the inverse of edge weights that constitute a path between a pair of nodes $i$ and $j$. For such a path, that consists of the following sequence of nodes, $\Omega_{i\leftrightarrow j} = \{i, x, y, \ldots, z, j\}$ with corresponding sequence of edge weights $\pi_{i\leftrightarrow j} = \{w_{ix}, w_{xy}, \ldots, w_{zj}\}$, the shortest path length is

   $$SPL_{ij} = \sum_{w_{lm}\in\pi_{i\leftrightarrow j}} \frac{1}{w_{lm}}$$

   Note that $\Omega_{i\leftrightarrow j} = \Omega_{j\leftrightarrow i}$ for shortest paths in any undirected graph.

3. Search Information

The search information ($SI_{ij}$) for two nodes $i$ and $j$ is the information required to follow the shortest path [49] i.e. the negative log of the product of probability of taking the correct exit at every node along the shortest path. In other words, it can be considered as the information required to reach node $j$ starting from node $i$. For a path between nodes $i$ and $j$ that has a sequence of nodes $\Omega_{i\to j} = \{i, x, y, \ldots, z, j\}$, with probability of taking the path $P(\pi_{i\to j}) = \Pi_{l\in\Omega^*_{i\to j}} 1/k_l$, the search information for the path is [50]

$$SI_{ij} = -\log_2 P(\pi_{i\to j})$$

Note that $SI_{ij} \neq SI_{ji}$

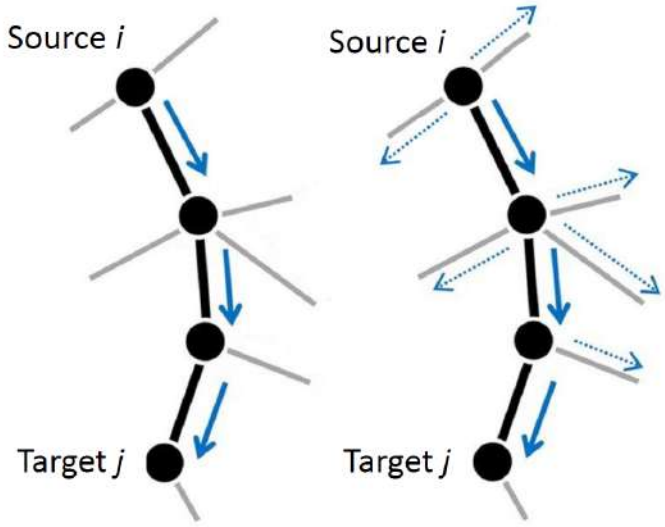


Figure 3.2. Part of a network that shows the shortest path between nodes $i$ and $j$. On the left, highlighted in blue arrows is the shortest path in a network. On the right shows shortest path and the number of possible exits at each node on the path highlighted in dotted blue arrows

4. Mean First Passage Time

The mean first passage time (MFPT) is the expected (on average) number of steps a random walker takes to reach node $j$ (for the the first time) from node

$i$ [51]. The Mean First Passage Time (MFPT) for a pair of nodes with source $i$ and target $j$ is

$$MFPT_{ij} = \frac{\zeta_{jj} - \zeta_{ij}}{\phi_j}$$

where $\phi$ is the left eigenvector associated with eigen value 1, $Z = [\zeta_{ij}]$ is the fundamental matrix computed as $Z = (I - P + \Phi)^{-1}$. Here $I$ is the $n \times n$ identity matrix, $P$ is the transition matrix and $\Phi$ is an $n \times n$ matrix with each column corresponding to the probability vector $\phi$ such that $\forall j\ \Phi_{ij} = \phi_i$. Please note that $MFPT_{ij} \neq MFPT_{ji}$.

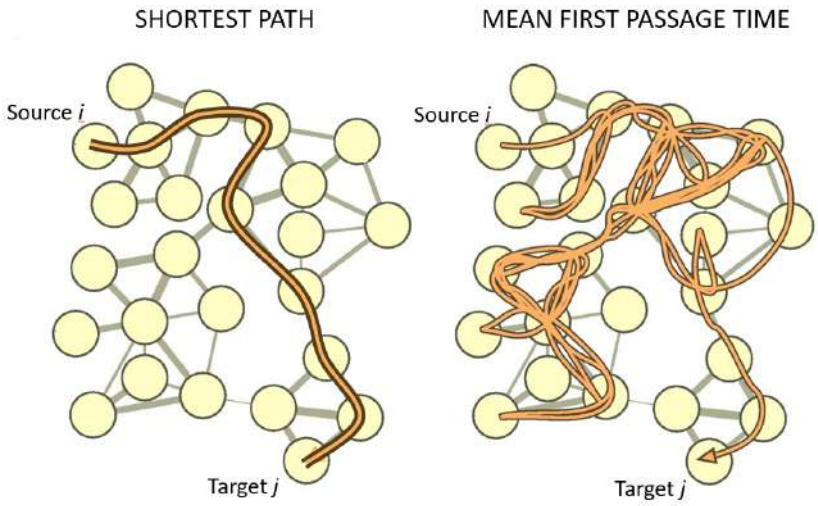


Figure 3.3. Network showing shortest path and the mean first passage time between nodes $i$ and $j$. One left, path in brown is the shortest path between nodes $i$ and $j$. On the right shows a possible path taken by a random walker starting from node $i$ to a target node $j$.

5. Driftness

We use a measure of communication called driftness [9] which is the ratio of the mean first passage time and the shortest path of a pair of nodes $i$ and $j$. Considering that $SP_{ij}$ is the best possible scenario path for a random-walk, this

measurement is modulating the mean first passage times with respect to the fastest routes within the network to go from node $i$ to $j$. Hence, note that $W_{ij} \geq 1$.

$$W_{ij} = \frac{MFPT_{ij}}{SP_{ij}}$$

6. Communicability

   Communicability between two nodes $i$ and $j$ is a measure of network integration computed as a weighted sum of number of all possible walks between them. [10] Here, we use a normalization method proposed to handle the disproportionate influence of highly connected nodes (also known as hubs) in a graph [52]. Note that this is frequently the case when assessing functional connectomes.

$$C_{ij} = [e^{D^{-0.5}AD^{-0.5}}]_{ij}$$

   where $D = diag(K)$ and $K = [k_i]$ where $k_i$ is the degree strength of node $i$, as defined above.

7. Clustering Coefficient

   The clustering coefficient of a node is the tendency of its neighbors to form cliques. It is the ratio of the total number of triangles a node forms with its neighbors to the total number of possible triangles that can be formed.

$$CC_i = \frac{2t_i}{k_i(k_i - 1)}$$

   where $t_i = \frac{1}{2}\sum_{j,h \in V}(w_{ij}w_{ih}w_{jh})^{1/3}$ is the geometric mean of triangles around node $i$ for weighted networks.

8. Betweeness Centrality

   The betweenness centrality of a node is the fraction of all shortest paths in a network that contain that node.

$$B_i = \frac{1}{(n-1)(n-2)} \sum_{\substack{h,j \in V \\ h \neq j, h \neq i, j \neq i}} \frac{\rho_{hj}(i)}{\rho_{hj}}$$

where $\rho_{hj}(i)$ is the number of shortest paths between $h$ and $j$ that pass through $i$. It can be seen as a measurement of to what extent a node "lies" between other pairs of nodes when accounting specifically for shortest-paths.

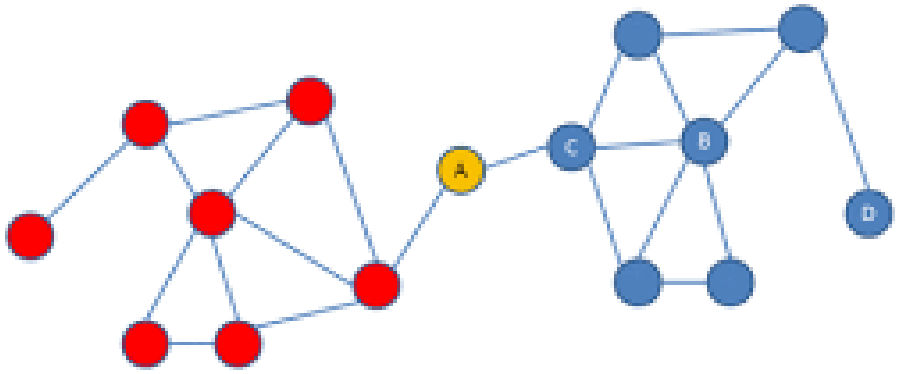


Figure 3.4. In this figure, the node $A$ has the highest betweenness centrality since the shortest path from any blue node to red node (and vice versa) will include the node $A$

# 4. FINGERPRINT IN BRAIN CONNECTOMICS

The concept of individual connectivity fingerprint - the idea that the functional connectivity of the brain differs across as it does across groups was first proposed almost 15 years ago. However, the concept was shelved due to the absence of accurate data to account for these differences. The availability of large amounts of high resolution fMRI datasets in recent times have encouraged the scientific community on a quest to capture inter-individual variability in the human functional connectome. This idea of individual fingerprint in brain connectomics has become ubiquitous in the cognitive neuroscience way of thinking. They have become critical for the development of personalized interventions for neuropsychiatric illnesses.

A comprehensive review of the methods that has been explored by the community and the progress that has been made can be found in [24]. This work used multidimensional scaling to show that each area occupies a unique place within this space of connections. An area here is a subpart of the brain defined on an anatomical basis, usually a distinct cytoarchitecture, myelin architecture, and/or receptor architecture, generally associated with a distinct function. The function of each individual area is determined in the main by its unique set of connections with the rest of the brain. Describing areas in terms of their connectivity helps to predict the function. This concept of connectivity has moved from a single brain region descriptor to a measure that can predict variability across individuals.

Other promising methods in recent times that can predict variability across individuals are [25, 26]. This work uses data from nine sampled adult humans to demonstrate that functional brain networks are in large part composed of individual-specific features that are stable over time. It convincingly demonstrated that the variance in functional connectivity data is best explained by a combination of factors that are

both common across individuals and consistently present in single individuals across sessions. It was found that a shared group-level factor explained slightly more that 30% of the variance in functional connectivity when data was aggregated across individuals [53, 54] and subject-specific features in the data explained a similar amount of variance. These features were stable and variance across sessions in individual patterns over time was minimal.

Substantial proof has also been found of network variants in individuals that are different from the group-level descriptors [21]. These variants are highly stable within individuals and are found in characteristic locations and associate with characteristic functional networks across large groups. These variants also indicate an association to functional variation. Individuals with similar behaviour show similarity in these variant characteristics. These results demonstrated that distributions of network variants may reflect stable, trait-like, functionally relevant individual differences in functional brain organization.

More recent work showed that individual variability in the entire connectome is both substantial and reproducible and can act as an identifying "fingerprint" [28]. While the task may change the brain connectivity to some extent, the intrinsic architecture is reliable enough across sessions and distinct enough from that of other individuals to identify the subject from the group regardless of the task during imaging. They further demonstrated through a full cross validated analysis that functional connectivity profiles can be used to predict the fundamental cognitive trait of fluid intelligence in novel subjects. This study used pearson correlation as a measure of similarity between two visits of the same subject.

Another study went beyond and used pearson correlation to not only find the similarity between two visits but also account for the dissimilarity between visits of different subject. [32] This study while coming up with a novel way of measuring the fingerprint level of a group subjects, has also introduced a robust framework to maximise fingerprint in functional connectomes. It uses a reconstruction procedure based on group-wise decomposition using principal component analysis in a finite number of

brain connectivity modes. This Identifiability framework ($\mathbb{I}f$) considerably improved the differential identifiability and also suggests that the same level of identifiability can be achieved at lower duration of acquisition lengths.

# 5. THE DIFFERENTIAL IDENTIFIABILITY FRAMEWORK

The Identifiability Framework ($\mathbb{I}f$) based on Principal Component Analysis of functional connectomes has been shown to improve functional connectome fingerprints within- and across- sites, for a variety of fMRI tasks, over a wide range of scanning length, and with and without global signal regression [32, 33]. Additionally, it has been shown that maximising differential identifiability on the functional connectomes provides more robust and reliable associations with cognition [34] as well as with disease progression [35].

The assessment of the individual fingerprint relies on the assumption that individual connectivity profiles should be more similar to a different visit or session and should be different from sessions of other individuals. Here, the similarity is measured as a pearson's correlation between two connectivity matrices. This idea motivated the differential identifiability score in [32]. A level of identifiability is measured on a set of functional connectomes as follows:

$$I_{diff} = (I_{self} - I_{others}) * 100$$

where $I_{self}$ is the mean of all correlation between the one session of a subject $i$ to another session of the same subject; $I_{others}$ is the mean of all correlation between sessions of different subjects $i$ and $j$. These values can constitute an identifiability matrix where each position $i, j$ denotes the correlation between the functional connectome of subject $i$ test and subject $j$ retest. The workflow of the framework can be found in

Principal component analysis (PCA) transforms a set of observations into principal components that are orthogonal in nature, ranked in a descending order of

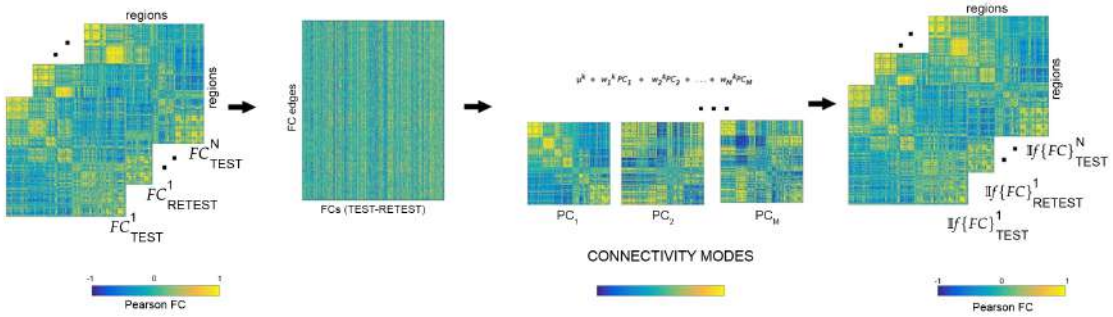


Figure 5.1. Workflow scheme of the Identifiability Framework $\mathbb{I}f$. The upper triangular of each functional connectivity matrix (two FCs per subject, test-retest) is vectorized and added to a matrix where columns are sessions and rows are their vectorized functional connectivity patterns. Data are first centered: this is obtained by subtracting the mean $\mu_k$ from each column (where k goes from 1 to N subjects). Second, the PCA algorithm extracts the M principal components (i.e. the M functional connectivity modes with the corresponding M eigenvectors) associated to the whole population and their relative weights across subjects. The M orthogonal connectivity modes are then used to reconstruct back the FC of each subject (k is added back to the data). Colorbars indicate positive (yellow) to negative (blue) connectivity values: Pearson's correlation coefficient in the case of individual FC matrices (left and right sides of scheme), and unitless connectivity weights in the case of PCA FC-modes. [32]

explained variance of the original data. The Identifiability Framework ($\mathbb{I}f$) uses this method to maximise the differential identifiability of a set of functional connectomes. In this framework, the functional connectome is reconstructed with all the principal components and are then iteratively reconstructed with a decreasing number of principal components. As the principal components are decreased, the differential identifiability increases, reaches a maximum and then continues to decrease further with any additional removal of principal components. 5.2 The components are removed with least explained variance first and progresses towards components of higher explained variance. This is done because the higher variance components might have

group-level functional connectivity information, especially the ones that are similar across the individuals whereas lower variance components might carry subject-level information and might in turn be noise as the explained variance decreases. The functional connectome are reconstructed with a number of principal components that give the maximum differential identifiability at which point the level of identifiability is the highest for the set of functional connectomes. These functional connectomes are highly similar across sessions of a subject and highly dissimilar across subjects.

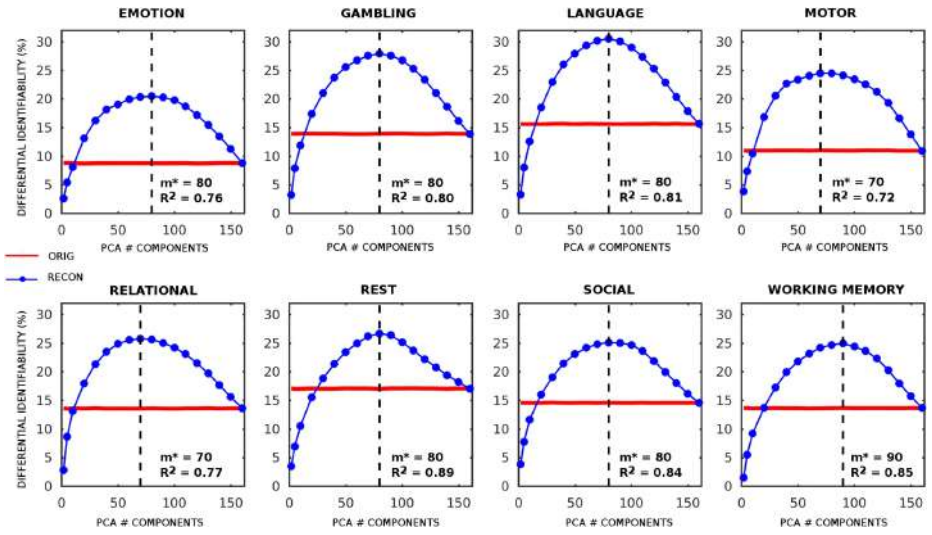


Figure 5.2. **Differential Identifiability ($I_{diff}$) profiles of network properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Each plot shows, for each fMRI task, the $I_{diff}$ score associated with functional connectivity *(blue solid line)* and the $I_{diff}$ score of the original functional connectome *(red solid line)*.

The Identifiability framework is seen to improve the differential identifiability on functional connectomes of all scanning lengths as can be seen in 5.3. It also reaches a level of identifiability that may not be ever achieved by increasing the scanning lengths. The framework also improves differential identifiability across scanners with or without global signal regression as can be seen in 5.4 [33].

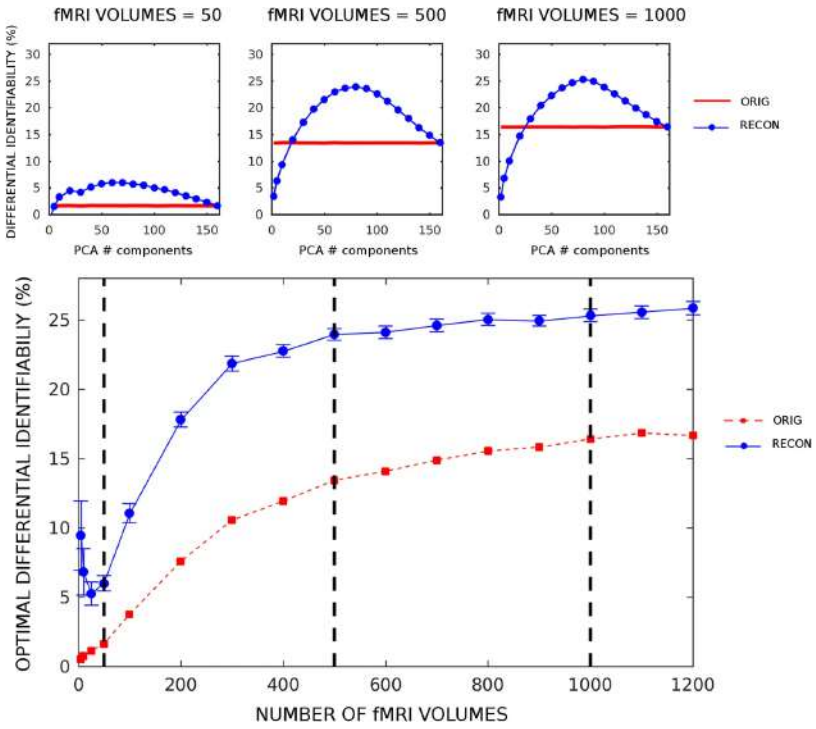


Figure 5.3. Optimal differential identifiability ($I_{diff}$) as a function of the number of fMRI volumes used for reconstruction on REST1. Bottom: plot shows the optimal average $I_{diff}$ across 100 runs, as a function of the number of fMRI volumes (scanning lengths) used for FC evaluation and subsequent reconstruction. Red line with circles denotes the average $I_{diff}$ for the original FCs, whereas blue line with circles denotes the identifiability for reconstructed FCs based on the different number of fMRI volumes retained (blue vertical bars indicate standard deviation across runs). Top: insets show the $I_{diff}$ curves (blue line, reconstructed $I_{diff}$; red line, original $I_{diff}$) per optimal number of PCA components for three different choices of number of fMRI frames retained (50, 500 and 1,000 fMRI volumes respectively). [32]

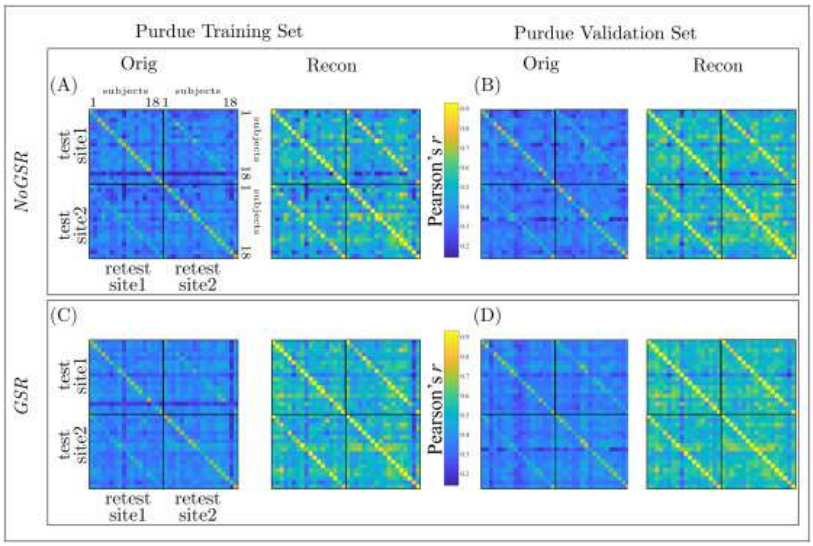


Figure 5.4. Identifiability matrices (I) of the original (Orig) and reconstructed (Recon) data for the Training, (A) and (C), and Validation, (B) and (D) sets of resting-state functional connectomes without global signal regression (NoGSR; (A) and (B)) and with global signal regression (GSR; (C) and (D)).

# 6. AN EXTENSION OF THE FRAMEWORK

In this work, we extend the framework by measuring the effects of the Identifiability Framework. It is crucial to understand this effect since network properties are widely used to understand the attributes of brain networks in general, and of functional connectomes in particular. [1, 2]. To understand this, we measure the differential identifiability on the derived network properties and how their profiles are as the number of principal components used in the reconstruction changes.

There are two different procedures that are assessed in this work. The functional connectomes of each subject (test and retest) are vectorized and added to a matrix, the columns of which are the runs (test and retest) of each subject, while the rows are the functional connectivity values of brain region pairs. The $m$ principal components of this matrix are then ranked by variance explained and included, in an iterative fashion, to reconstruct the functional connectomes [32]. This is done separately for each task and rest. Following the reconstruction of the functional connectomes, we then compute the network property of interest for each subject, on each run (test and retest). This is referred to as $NP(\mathbb{I}f\{FC\})$ in all further sections where $NP$ is the network property and $FC$ is the functional connectome.

In another case, network properties are computed on the original functional connectomes for each subject and run. The network properties are subsequently vectorized and added to a matrix the rows of which consists of the network property values corresponding to a pair of brain regions in case of pairwise properties or a brain region when node properties are derived. The principal components of this matrix are then extracted and iteratively reconstructed using $m$ number of components with highest explained variance. Since the network properties are the ones being decomposed in this case, the result of the reconstruction are the corresponding network

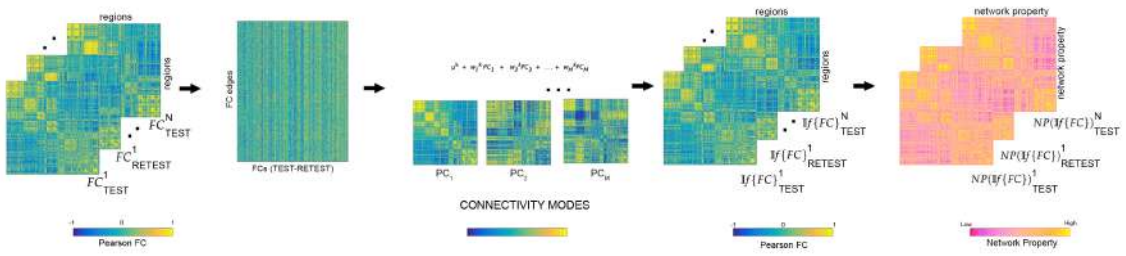


Figure 6.1. $NP(\mathbb{I}f\{FC\})$ Workflow scheme of the Identifiability Framework $\mathbb{I}f$ as extended in this work. The Functional connectomes are reconstructed from the original functional connectomes as described in **??** (blue-green color scheme). The network properties are then derived from the reconstructed functional connectomes. (pink-yellow color scheme)

properties of each individual and each run. This method is subsequently referred to as $\mathbb{I}f\{NP(FC)\}$).

Intraclass correlation coefficient (ICC) represents how strongly measures of a group are in agreement with each other [55,56]. The higher the ICC value, higher is the level of agreement. We use ICC [57] to asses the task sensitivity of a network measure, for each brain region pair and every subject. In this case, the members of the groups are the different runs (test and retest) of a subject; the different groups represent the different fMRI task conditions (and rest). The mean task sensitivity is then taken across all subjects and reported. For this assessment, the functional connectome (or the network property $\mathbb{I}f\{NP(FC)\}$) was optimally reconstructed, i.e. using the number of components that gave the highest $I_{diff}$ score for that task.

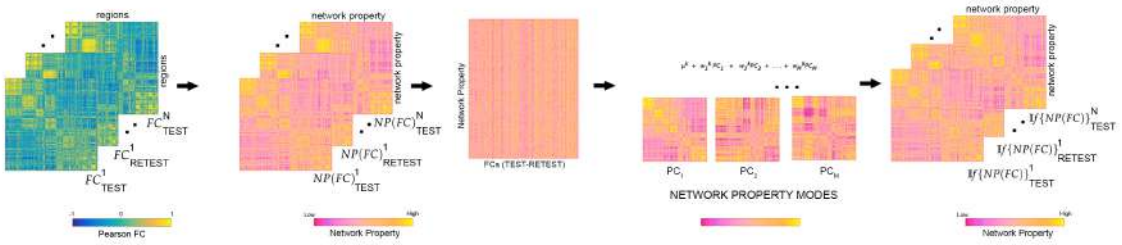


Figure 6.2. $\mathbb{I}f\{NP(FC)\}$ Workflow scheme of the Identifiability Framework $\mathbb{I}f$ as extended in this work. Here, the network properties are derived on the original functional connectomes. The network properties are then vectorized (either upper triangular of the matrix or the entire matrix depending on the type of network property) and added to matrix where columns are sessions and rows are the network property values. The PCA algorithm now extracts M principal components from this matrix which contains the network property data. The M orthogonal connectivity modes are then used to reconstruct back the derived network property corresponding to each subject.

# 7. RESULTS

The dataset used for this study consisted of fMRI scans of the 100 unrelated subjects from the Human Connectome Project [22]. For each subject, we computed 18 whole-brain functional connectivity matrices: 4 corresponding to resting-state (2 sessions, each with test and retest), and 14 corresponding to each of the 7 tasks (each including two runs; test-retest). The multimodal parcellation used here, as proposed by [42], includes 360 cortical brain regions. For completeness, 14 subcortical regions were added [58], hence producing functional connectome matrices (square, symmetric) of size $374 \times 374$.

In this work, we study the effects of $\mathbb{I}f$ on the identifiability profiles of network properties in two different scenarios: 1) when applying differential identifiability on functional connectivity, $NP(\mathbb{I}f\{FC\})$ and 2) when applying differential identifiability directly on network properties, $\mathbb{I}f\{NP(FC)\}$.

$NP(\mathbb{I}f\{FC\})$: The functional connectomes (FCs) of each task (including rest) were vectorized, organized together and then decomposed into principal components and subsequently reconstructed by adding an increasing number of components ordered by their variance explained. After every such reconstruction, a number of network measurements [see Methods for details] were computed for each FC and $I_{diff}$ was found on the derived network properties. This is compared with the $I_{diff}$ score estimated directly from the reconstructed functional connectomes - $\mathbb{I}f\{FC\}$. By doing so, we extend the differential identifiability framework to uncover fingerprints in network properties derived from functional connectomes.

For each task, we observed an optimal point of reconstruction where the differential identifiability on the FCs was maximized (see Figure 7.1). This optimal point was always in the neighborhood of half the maximum number of components (which is

equal to the number of subjects in the data) and produced $I_{diff}$ values much higher than fully reconstructed data, i.e. using all the components. These results reaffirm those reported by [32]. We then assessed $I_{diff}$ on the following node pair network properties: Shortest Path Length (SPL), Search Information (SI), Mean First Passage Time (MFPT), Driftness (W), and Communicability (C). In all cases, there was an optimal regime of number of components that maximized $I_{diff}$ (see Figure7.1). Overall, the $I_{diff}$ score on all the network properties and functional connectomes reach the peak at similar number of principal components, ranging between 80 and 110. We can also see that the $I_{diff}$ on functional connectomes is generally higher than those on the network properties for all the tasks and for most of the number of components. One exception is MFPT on Motor task where the $I_{diff}$ scores on FC and MFPT produced very similar results for the entire range of principal components. Another exception is MFPT on Relational task where the peak $I_{diff}$ of $MFPT(\mathbb{I}f\{FC\}$ is greater than that of $\mathbb{I}f\{FC\}$ but the margin of difference is really small ($\approx 0.59$).

In $\mathbb{I}f\{NP(FC)\}$) the different network properties (refer Methods) were first derived from the original functional connectomes and subsequently decomposed and reconstructed using the Identifiability framework. $I_{diff}$ scores were computed on these reconstructed network properties for different number of components and compared with those computed from the reconstructed FCs. (see Figure7.2)

As opposed to results shown in Figure7.1 which used $NP(\mathbb{I}f\{FC\})$, network properties have heterogeneous $I_{diff}$ profiles with respect to number of components. Compared to $I_{diff}$ from $\mathbb{I}f\{FC\}$, Search Information has a higher peak $I_{diff}$ score for all tasks while Communicability has a higher peak $I_{diff}$ score for all tasks except resting state. We also find that MFPT has a very different $I_{diff}$ profile compared to other network properties. The $I_{diff}$ profiles of MFPT from $\mathbb{I}f\{MFPT(FC)\}$ increases as we add the first few component and saturates or decreases gradually as more components are added (starting around 20 components for all tasks). This is unlike other network properties and functional connectome that share similar $I_{diff}$ profiles (see Figure

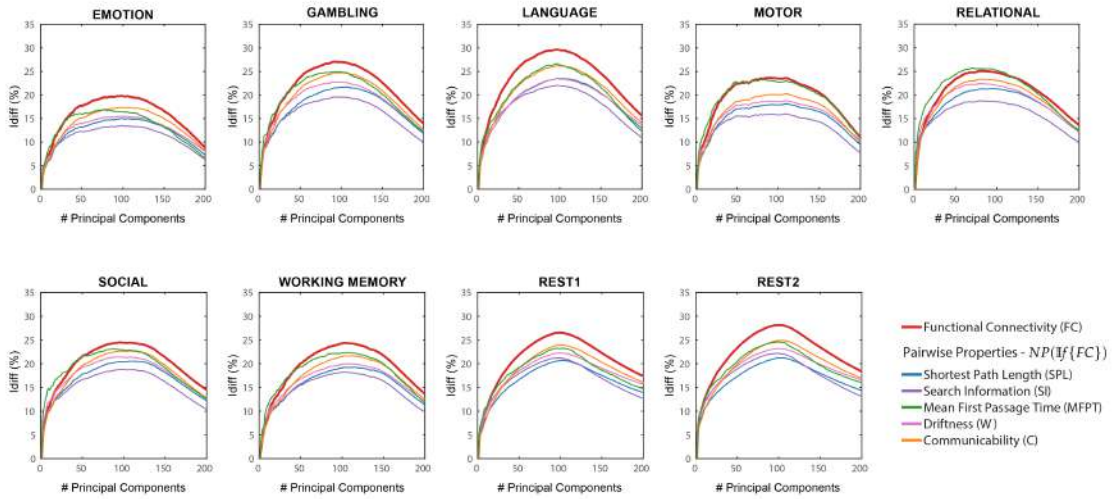


Figure 7.1. $NP(\mathbb{I}f\{FC\})$ **Differential Identifiability ($I_{diff}$) profiles of pairwise properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Here, the Identifiability framework was applied on the functional connectomes ($\mathbb{I}f\{FC\}$). Each plot shows, for each fMRI task, the $I_{diff}$ score associated with functional connectivity *(red solid line)* and the $I_{diff}$ scores on network properties derived from the reconstructed functional connectomes, $NP(\mathbb{I}f\{FC\})$ *(see legend)* for different number of components.

7.2). A summary of maximum $I_{diff}$, corresponding number of components used and variance retained for $NP(\mathbb{I}f\{FC\})$, and $\mathbb{I}f\{NP(FC)\}$ can be seen in Figure 7.3.

The network property with the most different $I_{diff}$ profiles was between $MFPT(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{MFPT(FC)\}$. Search Information was the only network property that reached higher $I_{diff}$ values for all fMRI tasks for $\mathbb{I}f\{SI(FC)\}$. The difference between Search Information and Mean First Passage time are assessed in detail in Figure7.5. Shaded area highlights the variability of $I_{diff}$ scores across different tasks for $NP(\mathbb{I}f\{FC\})$ (solid area) and $\mathbb{I}f\{NP(FC)\}$ (hatched area). Across all tasks, $I_{diff}$ on $\mathbb{I}f\{SI(FC)\}$ is higher than $SI(\mathbb{I}f\{FC\}$. However, for Mean First Passage time, $I_{diff}$ on $MFPT(\mathbb{I}f\{(FC)\}$ is higher than $(\mathbb{I}f\{MFPT(FC)\}$. When $SI(\mathbb{I}f\{FC\})$ is derived and optimally reconstructed, $I_{diff}$ on Search Information is highest across all tasks. However, under full

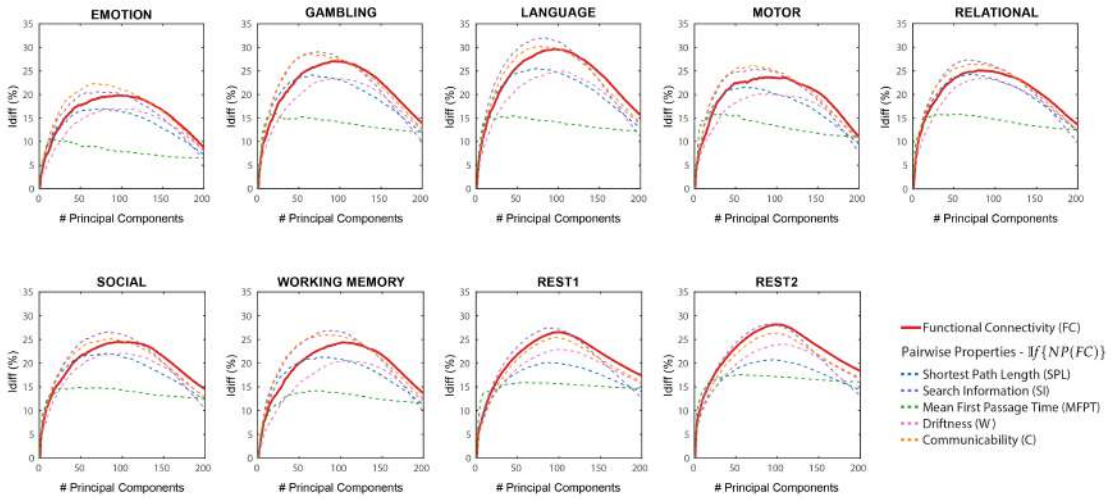


Figure 7.2. $\mathbb{I}f\{NP(FC)\}$ **Differential Identifiability ($I_{diff}$) profiles of pairwise properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Here, the Identifiability framework was applied directly on the network properties derived from the original functional connectomes ($\mathbb{I}f\{NP(FC)\}$). Each plot shows, for each fMRI task, the $I_{diff}$ score associated with functional connectivity *(red solid line)* and the $I_{diff}$ scores on reconstructed network properties derived from the original functional connectomes, $\mathbb{I}f\{NP(FC)\}$ *(see legend)* for different number of components.

reconstruction $m = 200$ (which is equivalent to using the original functional connectomes), $I_{diff}$ scores are highest for the functional connectome for all fMRI tasks.

We then assessed how differential identifiability varies based on node properties - Degree, Betweeness Centrality and Clustering Coefficient (Figure 7.6). We find that the $I_{diff}$ profiles of $NP(\mathbb{I}f\{FC\}$ are similar to that of $\mathbb{I}f\{FC\}$. These also give a significantly higher optimal $I_{diff}$ score for Gambling, Language, Motor and Working Memory tasks for all node properties. Especially in the case of Language and Motor tasks, Betweeness Centrality gives a significantly higher $I_d iff$ of 37 and 35 respectively at optimal reconstruction. For $\mathbb{I}f\{NP(FC)\}$, results show lower and flatter $I_{diff}$ profiles for all tasks and a wide range of number of components. $I_{diff}$ profiles using $NP(\mathbb{I}f\{FC\})$ of these node properties are in agreement with all pairwise

properties explored so far. In contrast, the $I_{diff}$ profiles using $\mathbb{I}f\{NP(FC)\}$ on these node properties are similar to $\mathbb{I}f\{MFPT(FC)\}$ only.

Intraclass Correlation Coefficient was used to assess the task sensitivity of each pairwise network property for three possible cases - $NP(\mathbb{I}f\{FC\})$ vs $NP(FC)$ (row a), $\mathbb{I}f\{NP(FC)\}$ vs $NP(FC)$ (row b) and $NP(\mathbb{I}f\{FC\})$ vs $\mathbb{I}f\{NP(FC)\}$ (row c). We find that the task sensitivity is higher for all network properties when the Identifiability framework was used (for both $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$). Between $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$, there is no one method that improves task sensitivity for all network properties.

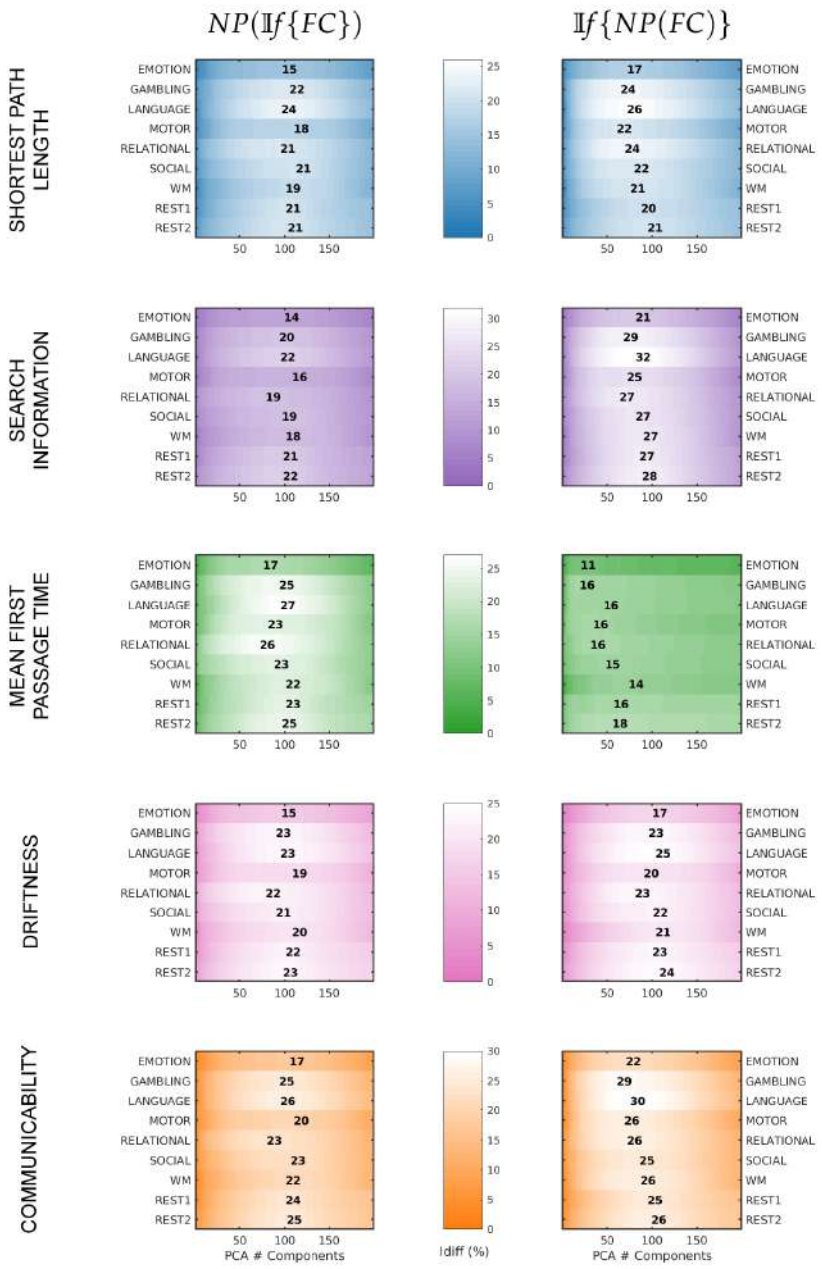


Figure 7.3. A summary of maximum $I_{diff}$ values, and corresponding number of components for each fMRI task and network property for both $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$. Each plot shows, for each property and each method - $NP(\mathbb{I}f\{FC\})$ or $\mathbb{I}f\{NP(FC)\}$, the $I_{diff}$ score for all tasks. The number mentioned gives the maximum $I_{diff}$ score for the corresponding task *(y axis)* and the position denotes the number of components *(x axis)*.

| | | $NP(\mathbb{I}f\{FC\})$ | | | $\mathbb{I}f\{NP(FC)\}$ | | |
|---|---|---|---|---|---|---|---|
| Properties | Task | m* | $I_{diff}$ | $R^2$ | m* | $I_{diff}$ | $R^2$ |
| Shortest Path Length | Emotion | 98 | 15 | 75.95 | 73 | 17 | 63.51 |
| | Gambling | 106 | 22 | 82.12 | 66 | 24 | 80.24 |
| | Language | 97 | 24 | 80.26 | 73 | 26 | 71.41 |
| | Motor | 111 | 18 | 79.74 | 62 | 22 | 62.14 |
| | Relational | 96 | 21 | 80.34 | 71 | 24 | 70.67 |
| | Social | 113 | 21 | 87.04 | 81 | 22 | 77.71 |
| | Working Memory | 102 | 19 | 83.06 | 77 | 21 | 74.10 |
| | Rest1 | 102 | 21 | 89.76 | 89 | 20 | 88.06 |
| | Rest2 | 104 | 21 | 89.80 | 96 | 21 | 89.10 |
| Search Information | Emotion | 100 | 14 | 76.56 | 83 | 21 | 64.15 |
| | Gambling | 95 | 20 | 79.27 | 69 | 29 | 61.76 |
| | Language | 96 | 22 | 80.00 | 83 | 32 | 66.76 |
| | Motor | 109 | 16 | 79.15 | 73 | 25 | 59.00 |
| | Relational | 80 | 19 | 75.93 | 65 | 27 | 61.73 |
| | Social | 98 | 19 | 83.84 | 83 | 27 | 71.00 |
| | Working Memory | 102 | 18 | 83.06 | 91 | 27 | 71.06 |
| | Rest1 | 99 | 21 | 89.30 | 87 | 27 | 78.67 |
| | Rest2 | 99 | 22 | 89.02 | 90 | 28 | 79.16 |
| Mean First Passage Time | Emotion | 77 | 17 | 69.13 | 21 | 11 | 55.33 |
| | Gambling | 95 | 25 | 79.27 | 20 | 16 | 56.41 |
| | Language | 96 | 27 | 80.00 | 47 | 16 | 75.81 |
| | Motor | 82 | 23 | 70.46 | 35 | 16 | 67.20 |
| | Relational | 73 | 26 | 73.84 | 32 | 16 | 65.11 |
| | Social | 89 | 23 | 81.76 | 49 | 15 | 80.50 |
| | Working Memory | 102 | 22 | 83.06 | 75 | 14 | 87.99 |
| | Rest1 | 102 | 23 | 89.76 | 58 | 16 | 90.81 |
| | Rest2 | 98 | 25 | 88.86 | 57 | 18 | 91.08 |
| Driftness | Emotion | 97 | 16 | 75.65 | 101 | 17 | 80.00 |
| | Gambling | 91 | 23 | 78.18 | 97 | 23 | 80.40 |
| | Language | 96 | 23 | 80.00 | 105 | 25 | 83.80 |
| | Motor | 109 | 19 | 79.15 | 92 | 20 | 77.44 |
| | Relational | 80 | 22 | 75.93 | 82 | 24 | 76.24 |
| | Social | 91 | 21 | 82.23 | 102 | 22 | 84.42 |
| | Working Memory | 109 | 20 | 84.71 | 105 | 21 | 85.88 |
| | Rest1 | 102 | 22 | 89.76 | 101 | 23 | 90.43 |
| | Rest2 | 99 | 23 | 89.02 | 109 | 24 | 91.72 |
| Communicability | Emotion | 106 | 17 | 78.35 | 72 | 22 | 64.11 |
| | Gambling | 95 | 25 | 79.27 | 62 | 29 | 65.49 |
| | Language | 96 | 26 | 80.00 | 77 | 30 | 72.17 |
| | Motor | 111 | 20 | 79.74 | 70 | 26 | 63.78 |
| | Relational | 81 | 23 | 76.22 | 73 | 26 | 70.49 |
| | Social | 107 | 23 | 85.80 | 87 | 25 | 78.15 |
| | Working Memory | 102 | 22 | 83.06 | 88 | 26 | 76.31 |
| | Rest1 | 102 | 24 | 89.76 | 96 | 25 | 88.45 |
| | Rest2 | 103 | 25 | 89.65 | 100 | 26 | 89.11 |

Figure 7.4. A summary of maximum $I_{diff}$ values, corresponding number of components and explained variance retained for each fMRI task and network property for both $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$.

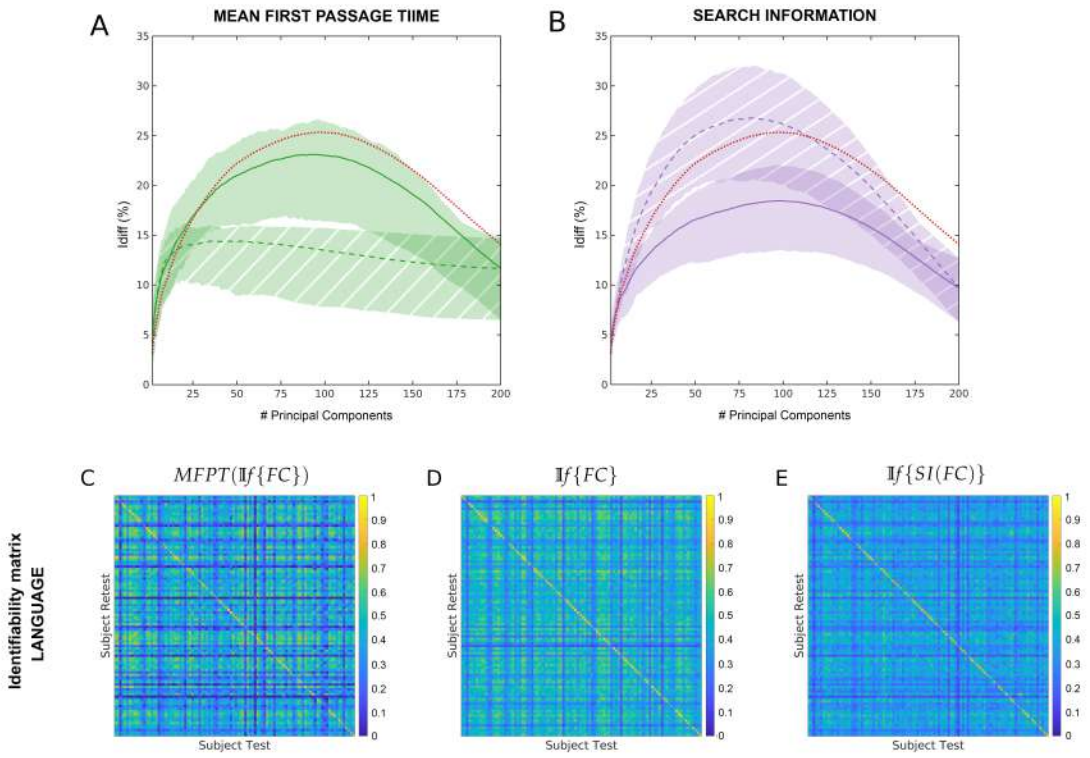


Figure 7.5. (A) Across tasks and rest differential Identifiability ($I_{diff}$) for Mean First Passage Time as a function of the number of principal components used for reconstruction. *Solid line* and *solid shaded area* represent the results for $MFPT(\mathbb{I}f\{FC\})$. *Dashed line* and *hatched area* show results for $\mathbb{I}f\{MFPT(FC)\}$ (B) Across tasks and rest differential Identifiability ($I_{diff}$) for Search Information as a function of the number of principal components used for reconstruction. *Solid line* and *solid shaded area* represent the results for $SI(\mathbb{I}f\{FC\})$. *Dashed line* and *hatched area* show results for $\mathbb{I}f\{SI(FC)\}$ The differential identifiability matrix (as defined in Methods) is shown at optimal reconstruction for Language task for (C) $MFPT(\mathbb{I}f\{FC\})$, (D) $\mathbb{I}f\{FC\}$ and (E) $\mathbb{I}f\{SI(FC)\}$. The diagonal elements in each matrix represent $I_{self}$ and the non-diagonal elements represent $I_{others}$.

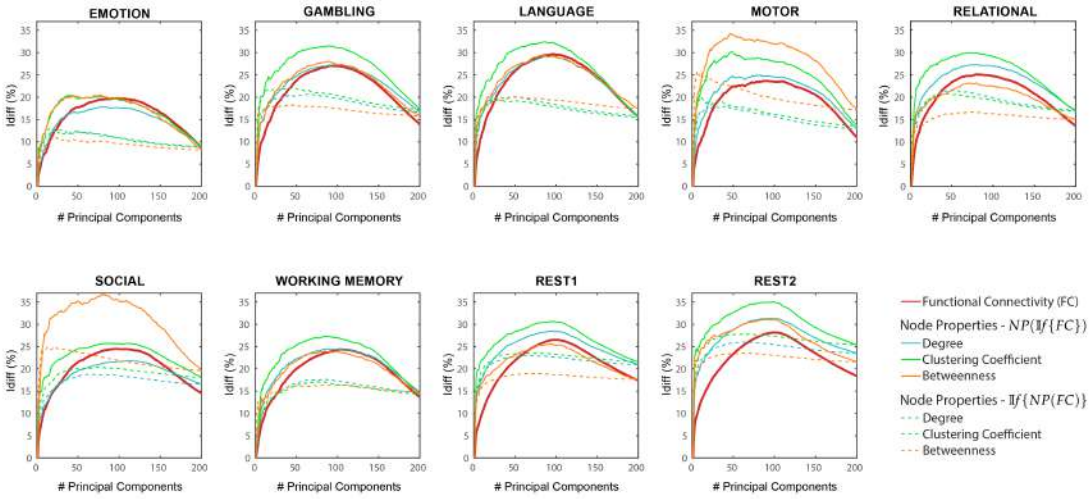


Figure 7.6. $NP(\mathbb{I}f\{FC\})$ and $\mathbb{I}f\{NP(FC)\}$ **Differential Identifiability ($I_{diff}$) of node properties** for different fMRI tasks as a function of the number of principal components used for reconstruction. Each plot shows, for each task, the $I_{diff}$ score associated with functional connectivity *(red solid line)*, the $I_{diff}$ scores on the network properties derived from the reconstructed functional connectomes $NP(\mathbb{I}f\{FC\})$ *(solid lines, colors - see legend)* and the $I_{diff}$ scores on the reconstructed network properties derived from the original functional connectomes $\mathbb{I}f\{NP(FC)\}$ *(dotted lines, colors - see legend)* for different number of components.

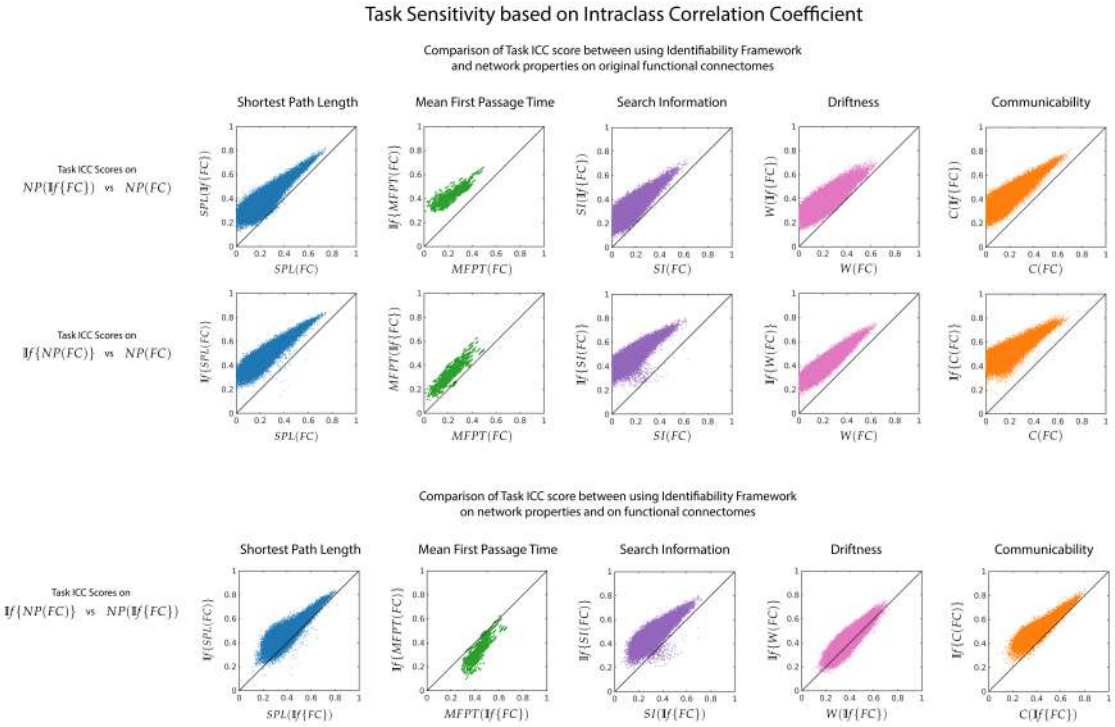


Figure 7.7. Effect of $\mathbb{I}f$ on task sensitivity of network measures. For each pairwise network property, task sensitivity is measured using ICC between - $NP(\mathbb{I}f\{FC\})$ vs $NP(FC)$ (row a), $\mathbb{I}f\{NP(FC)\}$ vs $NP(FC)$ (row b) and $NP(\mathbb{I}f\{FC\})$ vs $\mathbb{I}f\{NP(FC)\}$ (row c). First two rows highlight the fact that the $\mathbb{I}f$ framework uncovers the inherently distinct signature of different tasks through derived network properties. The last row shows that certain network properties would benefit more from application of the $\mathbb{I}f$ framework on the functional connectomes, while others from application directly on the network properties.

# 8. DISCUSSION

Brain connectivity fingerprinting has taken center stage in the neuroscientific community [21, 24–28, 59, 60]. As we move in this direction, there is a need to improve the reliability and robustness of individual fingerprint in functional connectomes and on common network measures extracted from functional connectomes. The Identifiability Framework ($\mathbb{I}f$) has shown the capacity to uncover subject fingerprint as measured by $I_{diff}$ score in human functional connectomes, regardless of the fMRI task [32]. Improving differential identifiability using the $\mathbb{I}f$ framework on functional connectomes (FCs) has been shown to improve the test-retest reliability of FCs and correlation with fluid intelligence [32]. Here, we extend this framework to show that by maximizing individual fingerprints in the functional connectomes, we also maximize individual fingerprint in network properties derived from the connectomes. Furthermore, we found that uncovering individual fingerprinting on network measurements also improves task signature. In addition, we show that in certain network properties, we can uncover an even stronger fingerprint if we apply the framework directly on the network property instead of functional connectomes.

Numerous work has been done to assess the effect of change in parameters of the acquisition process and the preprocessing pipelines on test-retest (TRT) reliability of fMRI data [61–64]. The impact of different correlation metrics, inclusion or exclusion of edges on functional connectomes, as well as the use of global signal regression, have been explored extensively [59, 65–69]. Additionally, TRT reliability is also seen to be affected by band pass filtering, scan length, sampling rate, network definition of the weights, and size of voxels for node definition [65, 70, 71]. The fact that the TRT reliability of the fMRI data and the subsequent estimation of functional connectomes is affected by such diverse factors, it is important to explore the reliability of the

derived network properties. Even though TRT reliability is not the only parameter to take into account when choosing the optimal strategy for brain network analyses, it surely has to be considered an important factor to help in such an important choice.

Essentially, $\mathbb{I}f$ works as a group-level data-driven (*denoising*) procedure where the components not contributing towards test-retest reliability of FCs are identified and removed. $\mathbb{I}f$ doesn't just improve the overall TRT reliability of a functional connectome but also improves it locally on an edge-level [32] which should ensure that both global and local network properties computed using these denoised functional connectomes are more reliable and robust. As shown in Figure 17.1, $\mathbb{I}f$ not only maximizes subject fingerprint at the FC level, but also at the network property level, which validated our premise. In addition, this convergent behavior is not present just at the optimal point; the identifiability profile of network properties follows the identifiability profile of the functional connectomes. In essence, we have shown that regardless of whether you are using functional connectomes or the network properties derived from them, using $\mathbb{I}f$ framework on the functional connectomes would be a beneficial first step.

A natural next question was to find if $\mathbb{I}f$ should be applied on functional connectomes and then derive the network properties ($NP(\mathbb{I}f\{FC\})$), or to use it directly on the network properties derived from original functional connectomes ($\mathbb{I}f\{NP(FC)\}$). The two approaches are an attempt to understand different principles of what a fingerprint is in a network derived measurement. $\mathbb{I}f\{NP(FC)\}$ assumes that functional connectomes are "holding" the individual fingerprints and then propagating them to the network measurements. The fact that maximizing fingerprint of functional connectomes also maximizes the fingerprint in derived network measures, suggests that functional connectomes do indeed hold a subject fingerprint which is then transmitted to the derived network properties. On the other hand, we also see that for some network measures (e.g. Search Information), we can uncover a better fingerprint if we apply the framework directly on the network measure. This suggests that specific network measures have a subject fingerprint of their own which gets added on to the

functional connectome fingerprint. Hence, if under some circumstances, the goal is to maximize the reliability and the individual variability of a specific network property, one can benefit from applying the $\mathbb{I}f$ framework on the network property itself, rather than on FCs.

Notably, in the $\mathbb{I}f\{SI(FC)\}$ scenario, the most different $I_{diff}$ profiles were found between MFPT and Search Information (Figure 47.5). Search Information consistently provides a better fingerprint across all tasks than functional connectome. MFPT, however, can neither improve nor match the fingerprint of functional connectomes. Also, it can not retain the fingerprint that is otherwise present is the functional connectomes and is then propgated to MFPT using $\mathbb{I}f\{MFPT(FC)\}$. Hence, while some properties (i.e., Search Information) can derive higher identifiability than functional connectomes, properties like MFPT need to be computed on optimally reconstructed functional connectomes to uncover subject identifiability on it.

These findings show that brain fingerprinting can be improved by adding multivariate information to "bivariate" measurements such as pairwise correlation used to estimate FCs. Specifically, individual fingerprint peaks on network measurements (e.g. Search Information) that are more multivariate and requires more information on the global topology of the functional network. However, if the information is heavily driven by degree properties (e.g. MFPT), then there is no improvement on the individual fingerprint (Figure 47.5). This is strongly corroborated by the $I_{diff}$ profiles of several node properties under the $\mathbb{I}f\{NP(FC)\}$ scenario. These profiles are very similar to that of MFPT, a network property which has a strong negative correlation with the degree of the target node. Although $\mathbb{I}f\{NP(FC)\}$ of these node properties have $I_{diff}$ profiles similar to $\mathbb{I}f\{MFPT(FC)\}$, the maximum $I_{diff}$ on these node properties are, for some tasks, significantly higher than $\mathbb{I}f\{FC\}$. Betweeness Centrality, for example, has a higher subject identifiability for Social and Motor tasks.

It was interesting to observe that under the $\mathbb{I}f\{NP(FC)\}$ scenario, Betweenness Centrality maximizes differential identifiability using just the first two components for Social and Motor tasks and that it was higher than the identifiability of the functional

connectomes for any number of components. Since Betweenness Centrality can be used to identify integrative communication hubs in FCs [8], it can be argued that social and motor tasks display a "hub functional fingerprint", which can be captured by the first two principal components.

A complementary assessment to the identification of subject fingerprints is to assess the ability to identify the different tasks used in this study. To do so, we used intraclass correlation coefficient on the derived network properties. The $\mathbb{I}f$ framework improved task sensitivity on the network properties (see Figure 67.7). Regardless of using the framework on the original functional connectomes or on the network properties themselves, a higher task sensitivity is obtained using one of the process depending on the network property. In both cases, the task reliability of the network properties has improved. The different tasks in the HCP dataset aim to assess different cognitive processes. Hence, the corresponding connectomes and the network properties derived from them should, at least to some extent, be task specific. We have shown that using the $\mathbb{I}f$ framework uncovers task-related fingerprints where unique cognitive processes result in differential network properties.

# 9. POSTHOC ANALYSIS

**Suggestion 1** Identifiability Framework on scale-free and small world networks

One could possibly consider using the Identifiability framework on standard network models such as the scale-free and small world network to understand how identifiability may or may not be uncovered while using the framework [45, 72]. In such a case, we could generate two networks of a type (either scale free or small world) with the same properties/features such as the number of nodes, degree etc. (depending on the type of networks) and consider each network to be different runs of a single subject. However, this analysis has many shortcomings. In an FC, of any subject and any run, the nodes are defined by physical brain regions that are the same across all subjects and sessions. It is then meaningful to compare brain region pair values of a functional connectome across subjects and sessions. However, in generative networks, where the nodes are generated differently, there is no similarity in structure or function between two generated networks even with the same properties/features. For example, two small world networks with the same number of nodes, initial degree of each node and probability or reconnection might be structurally the same network, but it would be difficult to compare them at the node level because of the way they are generated. This would require us to match the two network (graph-matching problem) which is beyond the scope of this work.

**Suggestion 2** Effect of noise in Functional Connectome on the Identifiability Framework.

In this analysis, we add random normally distributed noise (zero mean and 1 standard deviation) noise modulated by a scalar $\eta$ that is assessed within the range [0,2] in steps of 0.2. Such modulated noise is added to the Fisher's transformation [73] of each whole-brain FC of each subject and each session separately. The optimal

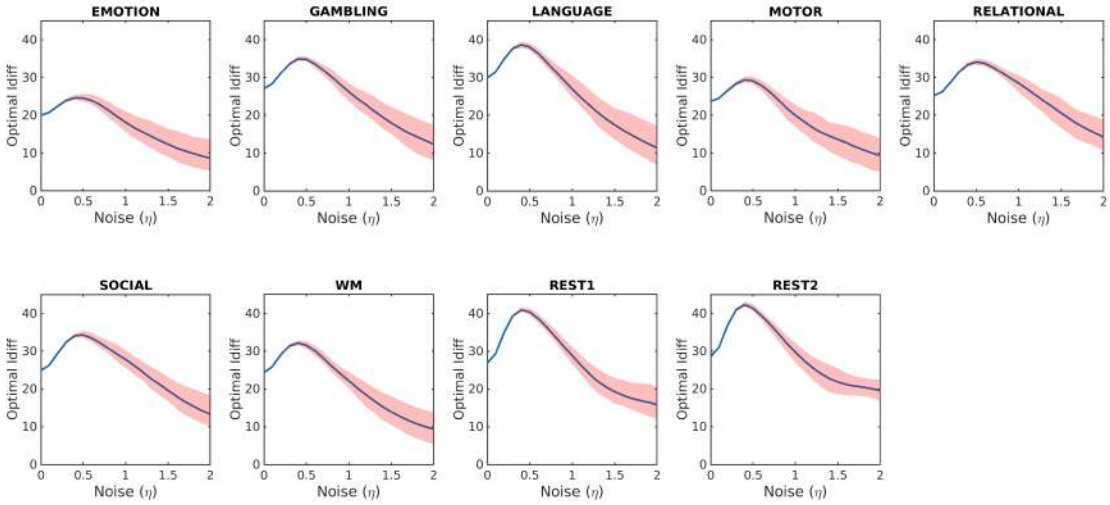


Figure 9.1. Each figure shows for each task, the optimal differential identifiability ($I_{diff}$) that can be uncovered using the Identifiability Framework $\mathbb{I}f$ when different levels of noise are added to the Functional connectomes. Different levels of noise were assessed (horizontal axes on all figures) The shaded areas represent results within the 2.5 and 97.5 percentiles across repetitions.

differential identifiability and the corresponding number of components at which it is reached is obtained. Note that $I_{diff}$ was measured, for number of components, on the reconstructed FCs after undoing the Fisher's transformation. This experiment is repeated 50 times on each task separately and for varying levels of noise.

We find that as more noise is added to the FCs, the optimal differential identifiability increases, reaches a peak at a noise level of 0.5 for all tasks and then decreases as further noise is added. The peak here is higher than what can be obtained in network properties or functional connectomes without the added noise. This is counter intuitive since one would expect lower differential identifiability when noise is added even when the framework is used. This can be observed for all tasks and rest. The optimal differential identifiability reaches values as high as 40 and higher for some tasks - language and rest. After the noise level goes beyond 0.5, all $I_{diff}$ values for all tasks rapidly decay.

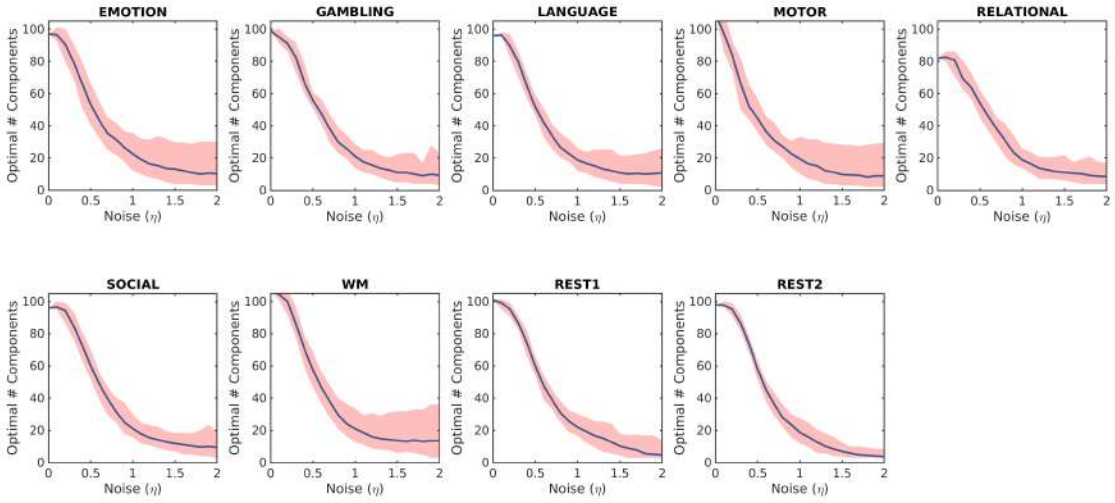


Figure 9.2. Each figures shows for each task, the optimal number of principal components required to uncover optimal identifiability using Identifiability Framework $\mathbb{I}f$ when different levels of noise is added to the Functional connectomes (xaxis). The shaded region represents the 2.5 and 97.5 percentile across repetitions.

On the other hand, the number of principal components that are required for optimal identifiability decreases as more noise is added and becomes almost flat after a noise level of 1. However, the variation across the repetitions are also higher than optimal identifiability and which also increases slightly with noise.

Further investigation of Identifiability matrix provides an insight into what happens when a noise level of 0.5 is added to the Functional Connectomes. When no framework is used and no noise is added, the $I_{self}$ values are low which gives an $I_{diff}$ score of 17.4. When the framework is used on the original functional connectomes, the $I_{diff}$ improves to 26.5 at optimal reconstruction. Here, we can see this is largely due to the improvement of $I_{self}$ scores. Although $I_{others}$ also increases, the net gain is towards $I_{self}$ and this improves identifiability. At optimal reconstruction of noisy FC, we see that $I_{self}$ is much higher than $I_{others}$ which results in an incredibly high $I_{diff}$ score of 40.

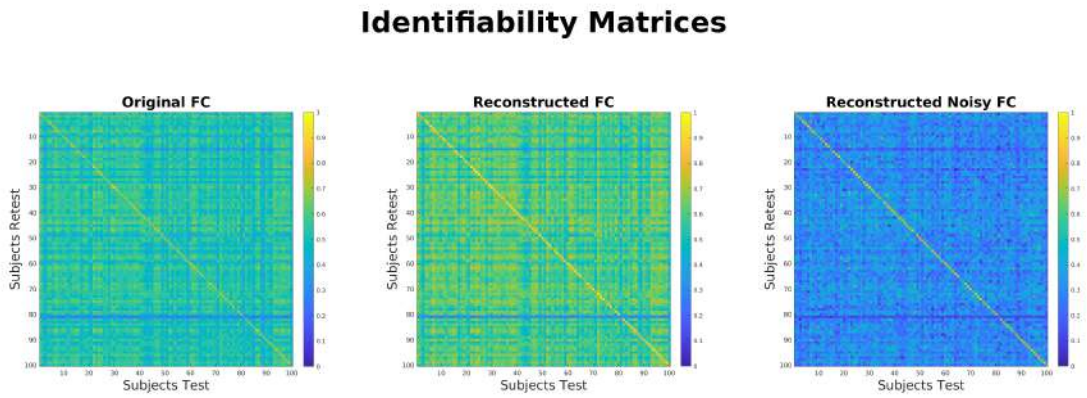


Figure 9.3. Identifiability matrices of Rest. On the left, is the Identifiability matrix of the Rest original functional connectomes, the center is the Identifiability matrix of optimally reconstructed functional connectomes and on the right is the identifiability matrix of functional connectomes reconstructed from noisy original data. Here the a noise level of 0.5 is added to the FCs of Rest.

# 10. SUMMARY

To summarize, differential identifiability was found to be always higher on functional connectomes than on any network properties when the Identifiability framework ($\mathbb{I}f$) is not used. When $\mathbb{I}f$ improved identifiability on functional connectomes, the identifiability on the network properties also increased. The framework also improved the subject fingerprints of the network properties. Not only do they improve at the optimal point, but the differential identifiability follows the same profile on network properties as it does on functional connectomes. We also find that applying the identifiability framework on the network properties instead of functional connectomes gives higher differential identifiability for some network properties. At optimal reconstruction, we find that Search Information has higher differential identifiability than functional connectomes across all tasks when the identifiability framework is applied on search information. This shows that there are network properties that can uncover better identifiability with framework than the functional connectomes themselves. Finally, we found that using the identifiability framework (either on functional connectomes or network property) improves task sensitivity in all network properties.

Only the unrelated subjects of the Human Connectome project and the cortical parcellation proposed by [39] are used in this work. Other explorations with other atlases, parcellations and/or other estimators of functional coupling (other than Pearson's correlation coefficient) would expand on the implications of our work. We have also limited to commonly used five pairwise and three node network properties. Delving into other network properties can strengthen this framework further and provide additional insights in understanding the associations between brain fingerprints, functional connectivity, and network derived properties.

This study can be extended to clinical applications to understand diseases that target specific functions of the human brain. Pathology whose signature cannot be mapped on the functional connectome itself but can be assessed using different network properties. [18, 74, 75] In this case, to retain individual differences and to be able to differentiate healthy population from clinical ones, we need this study to understand the advantages of using the Identifiability framework on the functional connectome or network property. Finally, studying the effect of the framework on the structural connectome is another natural extension of this work.

# REFERENCES

# REFERENCES


[1] Olaf Sporns. *Networks of the Brain*. MIT press, 2010.

[2] Alex Fornito, Andrew Zalesky, and Edward Bullmore. *Fundamentals of brain network analysis*. Academic Press, 2016.

[3] Olaf Sporns. Graph theory methods: applications in brain networks. *Dialogues in clinical neuroscience*, 20(2):111, 2018.

[4] Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual review of psychology*, 67:613–640, 2016.

[5] Jessica R Cohen and Mark D'Esposito. The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, 36(48):12083–12094, 2016.

[6] Gustavo Deco, Giulio Tononi, Melanie Boly, and Morten L Kringelbach. Rethinking segregation and integration: contributions of whole-brain modelling. *Nature Reviews Neuroscience*, 16(7):430–439, 2015.

[7] Makoto Fukushima, Richard F Betzel, Ye He, Martijn P van den Heuvel, Xi-Nian Zuo, and Olaf Sporns. Structure–function relationships during segregated and integrated network states of human brain functional connectivity. *Brain Structure and Function*, 223(3):1091–1106, 2018.

[8] Olaf Sporns. Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology*, 23(2):162–171, 2013.

[9] Luciano da Fontoura Costa, João LB Batista, and Giorgio A Ascoli. Communication structure of cortical networks. *Frontiers in computational neuroscience*, 5:6, 2011.

[10] Ernesto Estrada and Naomichi Hatano. Communicability in complex networks. *Physical Review E*, 77(3):036111, 2008.

[11] Jeffrey R Petrella. Use of graph theory to evaluate brain networks: a clinical tool for a small world?, 2011.

[12] Andrea Avena-Koenigsberger, Bratislav Misic, and Olaf Sporns. Communication dynamics in complex brain networks. *Nature Reviews Neuroscience*, 19(1):17, 2018.

[13] Andrew Zalesky, Alex Fornito, and Edward T Bullmore. Network-based statistic: identifying differences in brain networks. *Neuroimage*, 53(4):1197–1207, 2010.

[14] Elizabeth N Davison, Kimberly J Schlesinger, Danielle S Bassett, Mary-Ellen Lynall, Michael B Miller, Scott T Grafton, and Jean M Carlson. Brain network adaptability across task states. *PLoS computational biology*, 11(1):e1004029, 2015.

[15] Michał Bola and Bernhard A Sabel. Dynamic reorganization of brain functional networks during cognition. *Neuroimage*, 114:398–413, 2015.

[16] Mohsen Alavash, Claus C Hilgetag, Christiane M Thiel, and Carsten Gießing. Persistency and flexibility of complex brain networks underlie dual-task interference. *Human brain mapping*, 36(9):3542–3562, 2015.

[17] Marcelo G Mattar, Richard F Betzel, and Danielle S Bassett. The flexible brain. *Brain*, 139(8):2110–2112, 2016.

[18] Alex Fornito, Andrew Zalesky, and Michael Breakspear. The connectomics of brain disorders. *Nature Reviews Neuroscience*, 16(3):159–172, 2015.

[19] F Xavier Castellanos, Adriana Di Martino, R Cameron Craddock, Ashesh D Mehta, and Michael P Milham. Clinical applications of the functional connectome. *Neuroimage*, 80:527–540, 2013.

[20] Nicolas A Crossley, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T Fox, Philip McGuire, and Edward T Bullmore. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395, 2014.

[21] Benjamin A Seitzman, Caterina Gratton, Timothy O Laumann, Evan M Gordon, Babatunde Adeyemo, Ally Dworetsky, Brian T Kraus, Adrian W Gilmore, Jeffrey J Berg, Mario Ortega, et al. Trait-like variants in human functional brain networks. *Proceedings of the National Academy of Sciences*, 116(45):22851–22861, 2019.

[22] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

[23] Bharat B Biswal, Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M Smith, Christian F Beckmann, Jonathan S Adelstein, Randy L Buckner, Stan Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.

[24] Rogier B Mars, Richard E Passingham, and Saad Jbabdi. Connectivity fingerprints: from areal descriptions to abstract spaces. *Trends in cognitive sciences*, 22(11):1026–1037, 2018.

[25] Theodore D Satterthwaite, Cedric H Xia, and Danielle S Bassett. Personalized neuroscience: Common and individual-specific features in functional brain networks. *Neuron*, 98(2):243–245, 2018.

[26] Caterina Gratton, Timothy O Laumann, Ashley N Nielsen, Deanna J Greene, Evan M Gordon, Adrian W Gilmore, Steven M Nelson, Rebecca S Coalson, Abraham Z Snyder, Bradley L Schlaggar, et al. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, 98(2):439–452, 2018.

[27] Manasij Venkatesh, Joseph Jaja, et al. Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. *bioRxiv*, page 687830, 2019.

[28] Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664, 2015.

[29] Vicente Pallarés, Andrea Insabato, Ana Sanjuán, Simone Kühn, Dante Mantini, Gustavo Deco, and Matthieu Gilson. Extracting orthogonal subject-and condition-specific signatures from fmri data using whole-brain effective connectivity. *Neuroimage*, 178:238–254, 2018.

[30] Hua Xie, Vince D Calhoun, Javier Gonzalez-Castillo, Eswar Damaraju, Robyn Miller, Peter A Bandettini, and Sunanda Mitra. Whole-brain connectivity dynamics reflect both task-specific and individual-specific modulation: A multitask study. *Neuroimage*, 180:495–504, 2018.

[31] Monica D Rosenberg, Dustin Scheinost, Abigail S Greene, Emily W Avery, Young Hye Kwon, Emily S Finn, Ramachandran Ramani, Maolin Qiu, R Todd Constable, and Marvin M Chun. Functional connectivity predicts changes in attention over minutes, days, and months. *bioRxiv*, page 700476, 2019.

[32] Enrico Amico and Joaquín Goñi. The quest for identifiability in human functional connectomes. *Scientific reports*, 8(1):8254, 2018.

[33] Sumra Bari, Enrico Amico, Nicole Vike, Thomas M Talavage, and Joaquín Goñi. Uncovering multi-site identifiability based on resting-state functional connectomes. *NeuroImage*, 2019.

[34] Diana O Svaldi, Joaquín Goñi, Kausar Abbas, Enrico Amico, David G Clark, Charanya Muralidharan, Mario Dzemidzic, John D West, Shannon L Risacher, Andrew J Saykin, et al. Optimizing differential identifiability improves connectome predictive modeling of cognitive deficits in alzheimer\'s disease. *arXiv preprint arXiv:1908.06197*, 2019.

[35] Diana O Svaldi, Joaquín Goñi, Apoorva Bharthur Sanjay, Enrico Amico, Shannon L Risacher, John D West, Mario Dzemidzic, Andrew Saykin, and Liana Apostolova. Towards subject and diagnostic identifiability in the alzheimer's disease spectrum based on functional connectomes. In *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities*, pages 74–82. Springer, 2019.

[36] Dustin Scheinost, Stephanie Noble, Corey Horien, Abigail S Greene, Evelyn MR Lake, Mehraveh Salehi, Siyuan Gao, Xilin Shen, David O'Connor, Daniel S Barron, et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, 2019.

[37] Xilin Shen, Emily S Finn, Dustin Scheinost, Monica D Rosenberg, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *nature protocols*, 12(3):506, 2017.

[38] David C Van Essen, Kamil Ugurbil, E Auerbach, D Barch, TEJ Behrens, R Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[39] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

[40] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013.

[41] Deanna M Barch, Gregory C Burgess, Michael P Harms, Steven E Petersen, Bradley L Schlaggar, Maurizio Corbetta, Matthew F Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, et al. Function in the human connectome: task-fmri and individual differences in behavior. *Neuroimage*, 80:169–189, 2013.

[42] Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171, 2016.

[43] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[44] Daniel J Felleman and DC Essen Van. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.

[45] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440, 1998.

[46] Yolanda Cajal, Otto G Berg, and Mahendra Kumar Jain. Direct vesicle-vesicle exchange of phospholipids mediated by polymyxin b. *Biochemical and biophysical research communications*, 210(3):746–752, 1995.

[47] Carl Wernicke. *Grundriss der Psychiatrie in klinischen Vorlesungen*. Thieme, 1906.

[48] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[49] Martin Rosvall, Ala Trusina, Petter Minnhagen, and Kim Sneppen. Networks and cities: An information perspective. *Physical Review Letters*, 94(2):028701, 2005.

[50] Joaquín Goñi, Martijn P van den Heuvel, Andrea Avena-Koenigsberger, Nieves Velez de Mendizabal, Richard F Betzel, Alessandra Griffa, Patric Hagmann, Bernat Corominas-Murtra, Jean-Philippe Thiran, and Olaf Sporns. Resting-brain functional connectivity predicted by analytic measures of network communication. *Proceedings of the National Academy of Sciences*, 111(2):833–838, 2014.

[51] John G Kemeny and J Laurie Snell. *Markov Chains*. Springer-Verlag, New York, 1976.

[52] Jonathan J Crofts and Desmond J Higham. A weighted communicability measure applied to complex brain networks. *Journal of the Royal Society Interface*, 6(33):411–414, 2009.

[53] Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.

[54] BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125, 2011.

[55] Kenneth O McGraw and Seok P Wong. Forming inferences about some intraclass correlation coefficients. *Psychological methods*, 1(1):30, 1996.

[56] John J Bartko. The intraclass correlation coefficient as a measure of reliability. *Psychological reports*, 19(1):3–11, 1966.

[57] Patrick E Shrout and Joseph L Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2):420, 1979.

[58] Enrico Amico and Joaquín Goñi. Mapping hybrid functional-structural connectivity traits in the human connectome. *Network Neuroscience*, 2(3):306–322, 2018.

[59] Lisa Byrge and Daniel P Kennedy. High-accuracy individual identification using a "thin slice" of the functional connectome. *Network Neuroscience*, 3(2):363–383, 2019.

[60] Oscar Miranda-Dominguez, Brian D Mills, Samuel D Carpenter, Kathleen A Grant, Christopher D Kroenke, Joel T Nigg, and Damien A Fair. Connectotyping: model based fingerprinting of the functional connectome. *PloS one*, 9(11):e111048, 2014.

[61] Stephanie Noble, Dustin Scheinost, and R Todd Constable. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage*, 203:116157, 2019.

[62] Lubdha M Shah, Justin A Cramer, Michael A Ferguson, Rasmus M Birn, and Jeffrey S Anderson. Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and behavior*, 6(5):e00456, 2016.

[63] Rasmus M Birn, Erin K Molloy, Rémi Patriat, Taurean Parker, Timothy B Meier, Gregory R Kirk, Veena A Nair, M Elizabeth Meyerand, and Vivek Prabhakaran. The effect of scan length on the reliability of resting-state fmri connectivity estimates. *Neuroimage*, 83:550–558, 2013.

[64] Stephanie Noble, Marisa N Spann, Fuyuze Tokoglu, Xilin Shen, R Todd Constable, and Dustin Scheinost. Influences on the test–retest reliability of functional connectivity mri and its relationship with behavioral utility. *Cerebral Cortex*, 27(11):5415–5429, 2017.

[65] Xia Liang, Jinhui Wang, Chaogan Yan, Ni Shu, Ke Xu, Gaolang Gong, and Yong He. Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional mri study. *PloS one*, 7(3):e32766, 2012.

[66] Mark Fiecas, Hernando Ombao, Dan Van Lunen, Richard Baumgartner, Alexandre Coimbra, and Dai Feng. Quantifying temporal correlations: a test–retest evaluation of functional connectivity in resting-state fmri. *NeuroImage*, 65:231–241, 2013.

[67] Adam J Schwarz and John McGonigle. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *Neuroimage*, 55(3):1132–1146, 2011.

[68] Jin-Hui Wang, Xi-Nian Zuo, Suril Gohel, Michael P Milham, Bharat B Biswal, and Yong He. Graph theoretical analysis of functional brain networks: test-retest evaluation on short-and long-term resting-state functional mri data. *PloS one*, 6(7):e21976, 2011.

[69] Hengyi Cao, Michael M Plichta, Axel Schäfer, Leila Haddad, Oliver Grimm, Michael Schneider, Christine Esslinger, Peter Kirsch, Andreas Meyer-Lindenberg, and Heike Tost. Test–retest reliability of fmri-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage*, 84:888–900, 2014.

[70] Urs Braun, Michael M Plichta, Christine Esslinger, Carina Sauer, Leila Haddad, Oliver Grimm, Daniela Mier, Sebastian Mohnke, Andreas Heinz, Susanne Erk, et al. Test–retest reliability of resting-state connectivity network characteristics using fmri and graph theoretical measures. *Neuroimage*, 59(2):1404–1412, 2012.

[71] Xu-Hong Liao, Ming-Rui Xia, Ting Xu, Zheng-Jia Dai, Xiao-Yan Cao, Hai-Jing Niu, Xi-Nian Zuo, Yu-Feng Zang, and Yong He. Functional brain hubs and their test–retest reliability: a multiband resting-state functional mri study. *Neuroimage*, 83:969–982, 2013.

[72] Mark EJ Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102, 2001.

[73] Ronald A Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521, 1915.

[74] Danielle S Bassett and Edward T Bullmore. Human brain networks in health and disease. *Current opinion in neurology*, 22(4):340, 2009.

[75] Alex Fornito and Edward T Bullmore. Connectomics: a new paradigm for understanding brain disease. *European Neuropsychopharmacology*, 25(5):733–748, 2015.