Méthodes de classification des graphes. Application à l’identification des réseaux fonctionnels impliqués dans les processus de mémoire

Thèse soutenue à Rennes le 16 décembre 2016
devant le jury composé de :

Claude BERROU
Professeur, Télécom-Bretagne / examinateur

Sophie ACHARD
Chargé de recherche, Université Joseph Fourier / rapporteur

Mohamad SAWAN
Professeur, école polytechnique de Montréal / rapporteur

Sandy RIHANA
Professeure agrégée, USEK / examinateur

Fabrice WENDLING
Directeur de recherche Inserm, Université de Rennes1 / directeur de thèse

Mohamad KHALIL
Professeur, Université Libanaise / directeur de thèse

Mahmoud HASSAN
Post-doc, Université de Rennes1 / examinateur
ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my superiors Dr. Fabrice Wendling, Dr. Mohamad Khalil and Dr. Mahmoud Hassan for their support over the three years of my thesis. A special thanks for the CominLabs and Azm Association for its financial and moral support during the period of the thesis. I also want to express my thanks to both Lebanese University as well as University of Rennes1 who gave me their trust by allowing me to present my dissertation. Big thanks to my family and all my friends whether they are in Lebanon or in France.
Résumé en Français ........................................................................................................................................ 7
Summary in English ...................................................................................................................................... 11
Chapter 1- General Introduction ............................................................................................................ 13
Chapter 2-problem statement, Background and proposed approaches ................................................. 15
  2.1 Problem statement ............................................................................................................................. 15
      2.1.1 Problem 1: Fast reconfiguration of functional brain networks ............................................... 15
      2.1.2 Problem 2: Brain network similarity ....................................................................................... 16
  2.2 Graph notions and measures ............................................................................................................ 16
      2.2.1 Basic notions on graphs ............................................................................................................ 16
      2.2.2 Graph measures ....................................................................................................................... 19
  2.3 Neuroimaging techniques .................................................................................................................. 23
      2.3.1 Functional MRI (fMRI) .......................................................................................................... 23
      2.3.2 Magneto/Electro-encephalography (M/EEG) ..................................................................... 23
  2.4 Brain connectivity ............................................................................................................................... 24
      2.4.1 Structural connectivity ............................................................................................................. 24
      2.4.2 Functional connectivity .......................................................................................................... 25
      2.4.3 Effective connectivity .............................................................................................................. 25
  2.5 Nodes and edges in brain network .................................................................................................... 26
  2.6 Dynamic reconfiguration of brain functional networks ....................................................................... 27
      2.6.1 Background ............................................................................................................................. 27
      2.6.2 From background to proposed methods ................................................................................ 27
  2.7 Background on graph similarity ......................................................................................................... 28
      2.7.1 Isomorphism-based methods .................................................................................................. 28
      2.7.2 Graph edit distance (GED) .................................................................................................... 29
      2.7.3 Features extraction ............................................................................................................... 30
      2.7.4 Similarity in brain network context ......................................................................................... 32
      2.7.5 From background to proposed methodology ........................................................................ 34
Chapter 3 - Materials and methods

3.1 Dense Electroencephalography

3.2 Data

3.2.1 Cognitive task

3.2.2 Epileptic data

3.3 Data analysis

3.3.1 Dynamic reconfiguration of brain functional networks

3.3.2 Brain networks similarity

Chapter 4 - Results

Study 1: A new algorithm for spatiotemporal analysis of brain functional connectivity

Study 2: SimNet: A novel method for quantifying brain network similarity

Study 3: Identification of interictal epileptic networks from dense-EEG

Chapter 5 - Discussion and future work

5.1 Dynamic reconfiguration of functional brain networks

5.2 Brain networks similarity

5.3 Information processing in the human brain

5.4 Methodological considerations

5.4.1 Functional connectivity

5.4.2 Source Leakage

5.4.3 Thresholding the connectivity matrix

5.5 Future directions

References
Le cerveau humain est un réseau «large-échelle» formé de régions corticales distribuées et fonctionnellement interconnectées. Le traitement de l’information par le cerveau est un processus dynamique mettant en jeu une réorganisation rapide des réseaux cérébraux fonctionnels, sur une échelle de temps très courte (inférieure à la seconde). Dans le champ des neurosciences cognitives, deux grandes questions restent ouvertes concernant ces réseaux. D’une part, est-il possible de suivre leur dynamique spatiospatio-temporelle avec une résolution temporelle nettement supérieure à celle de l’IRM fonctionnelle? D’autre part, est-il possible de mettre en évidence des différences significatives dans ces réseaux lorsque le cerveau traite des stimuli (visuels, par exemple) ayant des caractéristiques différentes. Ces deux questions ont guidé les développements méthodologiques élaborés dans cette thèse. En effet, de nouvelles méthodes basées sur l’électroencéphalographie sont proposées. Ces méthodes permettent, d’une part de suivre la reconfiguration dynamique des réseaux cérébraux fonctionnels à une échelle de temps inférieure à la seconde. Elles permettent, d’autre part, de comparer deux réseaux cérébraux activés dans des conditions spécifiques. Nous proposons donc un nouvel algorithme basé sur l'excellente résolution temporelle de l’EEG afin de suivre la reconfiguration rapide des réseaux fonctionnels cérébraux à l’échelle de la milliseconde. L'objectif principal de cet algorithme est de segmenter les réseaux cérébraux en un ensemble d’ «états de connectivité fonctionnelle» à l’aide d’une approche de type « clustering ». L’algorithme est basé sur celui des K-means et a été appliqué sur les graphes de connectivité obtenus à partir de l'estimation des valeurs de connectivité fonctionnelle entre les régions d'intérêt considérées (Figure i).
Figure i: Illustration de l'organigramme de l'algorithme de segmentation proposé: trois étapes principales ont été réalisées: l'initialisation, l'affectation et la mise en jour.

La seconde question abordée dans ce travail relève de la mesure de similarité entre graphes. Ainsi, afin de comparer des réseaux de connectivité fonctionnelle, nous avons développé un algorithme (SimNet) capable de quantifier la similarité entre deux réseaux dont les noeuds sont définis spatialement. Cet algorithme met en correspondance les deux graphes en « déformant » le premier pour le rendre identique au second sur une contrainte de coût minimal associée à la déformation (insertion, suppression, substitution de nœuds et d’arrêtes). Il procède selon deux étapes, la première consistant à calculer une distance sur les nœuds et la seconde une distance sur les arrêtes. Cet algorithme fournit un indice de similarité normalisé: 0 pour aucune similarité et 1 pour deux réseaux identiques (Figure ii). Il a été évalué sur des graphes simulés puis comparé à des algorithmes existants. Il montre de meilleures performances pour détecter la variation spatiale entre les graphes. Il a également été appliqué sur des données réelles afin de comparer différents réseaux cérébraux. Les résultats ont montré des performances élevées pour comparer deux réseaux cérébraux réels obtenus à partir de l’EEG à haute résolution spatiale, au
Figure ii: Illustration de l'algorithme de SimNet
cours d'une tâche cognitive consistant à nommer des éléments de deux catégories différentes (objets vs animaux).
SUMMARY IN ENGLISH

The human brain is a "large-scale" network consisting of distributed and functionally interconnected regions. The information processing in the brain is a dynamic process that involves a fast reorganization of functional brain networks in a very short time scale (less than one second). In the field of cognitive neuroscience, two big questions remain about these networks. Firstly, is it possible to follow the spatiotemporal dynamics of the brain networks with a temporal resolution significantly higher than the functional MRI? Secondly, is it possible to detect a significant difference between these networks when the brain processes stimuli (visual, for example) with different characteristics?

These two questions are the main motivations of this thesis. Indeed, we proposed new methods based on dense electroencephalography. These methods allow: i) to follow the dynamic reconfiguration of brain functional networks at millisecond time scale and ii) to compare two activated brain networks under specific conditions. We propose a new algorithm benefiting from the excellent temporal resolution of EEG to track the fast reconfiguration of the functional brain networks at millisecond time scale. The main objective of this algorithm is to segment the brain networks into a set of "functional connectivity states" using a network-clustering approach. The algorithm is based on \textit{K-means} and was applied on the connectivity graphs obtained by estimation the functional connectivity values between the considered regions of interest. The second challenge addressed in this work falls within the measure of similarity between graphs. Thus, to compare functional connectivity networks, we developed an algorithm (SimNet) that able to quantify the similarity between two networks whose node coordinates is known. This algorithm maps one graph to the other using different operations (insertion, deletion, substitution of nodes and edges). The algorithm is based on two main parts, the first one is based on calculating the nodes distance and the second one is to calculate the edges distance. This algorithm provides a normalized similarity index: 0 for no similarity and 1 for two identical networks. SimNet was evaluated with simulated graphs and was compared with previously-published graph similarity algorithms. It shows high performance to detect the similarity variation between graphs involving a shifting of the location of nodes. It was also applied on real data to compare different brain networks. Results showed high performance in the comparison of real brain networks obtained from dense EEG during a cognitive task consisting in naming items of two different categories (objects vs. animals).
CHAPTER 1 - GENERAL INTRODUCTION

The general topic of this thesis is the study of large-scale networks involved during information processing in the human brain. This thesis was part of a project called “neural coding” supported by the Excellence Laboratory CominLabs (http://www.cominlabs.ueb.eu). The starting point was the mental information theory proposed by (Berrou and Gripon, 2012). According to this theory, information coding/decoding in the brain involves "small" neural networks, referred to as "neural cliques". Neural cliques can be viewed as complete graphs whose topology is an essential feature in the encoding and retrieval of "infons" (i.e. stored elements of information). Starting from this theory, the general objective of the Neural Coding project was to evidence, from neuroimaging data, the presence of cliques and meta-cliques (“cliques of cliques”) in human subjects performing a cognitive task related to working memory. In this context, the specific objective of this thesis was to propose novel methods to extract and analyze the spatio-temporal dynamics of brain networks, as identified from dense-EEG signals.

There is an emerging consensus that the human brain is a dynamic complex network of distributed and functionally interconnected regions. To guarantee an efficient cognitive function, information is continuously processed and integrated in this complex network. Emerging evidence shows that the dynamic behavior of the brain networks is fundamental to understand cognition and brain disorders (Braun et al., 2015). Recently, many studies showed that the dynamic reconfiguration of the brain networks is related to different brain activity such as learning (Bassett et al., 2013) and may be altered during disease progression (Siebenhühner et al., 2013). Interestingly, this reconfiguration can take place over different time scales, typically minutes, hours, and days (Bassett et al., 2013, van den Heuvel et al., 2013). However, many other brain functions are very fast and occur over even shorter time periods. Thus, functional brain networks should be able to rapidly reconfigure on a sub-second time scale. This can be the case of normal cognitive activity (object recognition for instance) or pathological activity like epileptiform events. This dynamic reconfiguration usually results in network patterns that are specific to both exogenous and endogenous conditions such as external drivers (visual/auditory stimuli) and pathophysiologica states including a transition from normal to pathological state in considered brain networks.

The accurate tracking of the spatiotemporal dynamics of large-scale networks involved in cognitive processes (often as short as a few hundreds of milliseconds) is still a challenging issue (Sporns,
Another major challenge in cognitive neuroscience is to evaluate the ability of the human brain to categorize or group visual stimuli based on common features (Haxby et al., 2000, Whatmough et al., 2002, Mechelli et al., 2006, Van Wijk et al., 2010, Cichy et al., 2014). Consequently, two key methodological challenges are clearly faced here. On one hand, new methods able to track the dynamic reconfiguration of functional brain networks at sub-second time scale are needed. On the other hand, the comparison between condition-specific brain networks is still an open question (Shimada et al., 2016). **These two issues are the main motivations of this thesis.**

To tackle the first issue, we have developed a new algorithm able to track the dynamics of functional brain networks at millisecond time scale. The proposed algorithm is based on $K$-means clustering approach. The new method allowed us to summarize the neural networks in a limited number of dominant networks occurring over a given short time period (<1 sec) during the cognitive process (picture naming in our case).

The second issue is the quantitative comparison of the graphs corresponding to the networks activated for the various categories of stimuli. In this context, the spatial location of the graph nodes (distributed over distinct cortical regions) is a key factor, often ignored by most of algorithms proposed previously. Here also, we present a new algorithm called ‘SimNet’ for measuring similarity between graphs in which the coordinates of nodes are known. The main feature of SimNet is to take into consideration the spatial location of nodes in order to find the similarity scores between the compared graphs. SimNet is based on two main parts that are necessary to map one graph to the other: the first step is based on calculating the nodes distance and the second step is to calculate the edges distance. The algorithm provides a normalized Similarity Index (SI): 0 for no similarity and 1 for two identical networks (same properties and topology).

This manuscript is organized as following. In chapter 2, we report the research context, problem statement and background of the dynamic brain networks and the brain network similarity followed by a general description of the proposed algorithms. In chapter 3, all the materials and methods are described. Results are provided in chapter 4, under the form of published or under revision articles. For each article, a synthesis presenting the objectives, methods and results are provided. Finally, conclusions and prospects for future work are given in chapter 5.
CHAPTER 2-PROBLEM STATEMENT,
BACKGROUND AND PROPOSED APPROACHES

In this chapter, we present the problem statement related to the two challenges tackled in this thesis: the dynamic of brain networks and the brain networks similarity. Next, we describe the basic notions of graph and the network measures followed by the concept of brain connectivity and the brain networks as identified from dense-EEG. The proposed solutions for each of the two challenges are then introduced.

2.1 PROBLEM STATEMENT

2.1.1 PROBLEM 1: FAST RECONFIGURATION OF FUNCTIONAL BRAIN NETWORKS

It is highly likely that the functional brain networks dynamically reconfigure over time under the influence of external and internal stimulus (Bassett and Bullmore, 2006). To understand and characterize this reconfiguration, we need to quantify the variation of functional connectivity with respect to time. Several studies have been elaborated to assess the spatiotemporal dynamics of functional brain networks (Hutchison et al., 2013, Braun et al., 2015). These studies have reported that the organization of brain networks is not stable, but dynamically changes over time and following changes in task conditions (Alavash et al., 2016). In most cognitive tasks, this reconfiguration occurs at sub-second time scale and thus tracking the dynamics of functional networks is crucial to understand cognitive brain functions. This tracking cannot be achieved using functional Magnetic Resonance imaging (fMRI), which has a poor temporal resolution (1-2 second) as BOLD signals mainly reflect the metabolic and hemodynamic response of activated brain areas. On the other hand, Electroencephalography (EEG) and Magnetoencephalography (MEG) measurements offer an excellent temporal resolution that allows capturing the dynamics of neuronal assemblies at millisecond time scale. Therefore the question that is raised is: how can we track the dynamic of the cortical brain networks at sub-second time scale using M/EEG?
2.1.2 PROBLEM 2: BRAIN NETWORK SIMILARITY

When brain networks are represented by graphs, analysis based on graph theory can be used to characterize the properties of these graphs. Some of these properties are related to nodes such as the degree, the strength and the presence of hubs (Costa et al., 2007) and some are related to edges such as the edge weights and edge centrality (Costa et al., 2007) and other related to the global features of the considered graph such as density and modularity (Bullmore and Sporns, 2009). As compared with the large number of methods aiming at characterizing graph properties (Costa et al., 2007), less attention has been paid to algorithms aimed at comparing graphs (Costa et al., 2007, Shimada et al., 2016). In the context of brain network analysis, the techniques used for comparing networks can be classified into three categories: i) global, consisting in comparing between global graph measures computed from the two networks, ii) node-wise, consisting in computing graph metrics for each node of the networks such that multiple comparisons can be assessed and iii) edge-wise, consisting in comparing all the available edges in the two networks. However, the spatial locations of the nodes were not integrated in these approaches. We believe that this physical location of nodes is a key factor when measuring similarity between brain networks. For instance, two networks with identical properties but interconnecting distant brain regions can have low similarity. Conversely, two networks with dissimilar properties but interconnecting spatially close brain regions can be very similar. Therefore, these considerations leads to the second issue that is dealt with in this thesis: how can we measure the similarity between two networks with taking into account the spatial locations of nodes?

Before showing our contributions to tackle the above two questions, we introduce basic notions about graphs, neuroimaging techniques and dense-EEG-based functional connectivity as they represent key factors in the thesis work.

2.2 GRAPH NOTIONS AND MEASURES

2.2.1 BASIC NOTIONS ON GRAPHS

If we aim at understanding a complex system, we need to know how its elements interact with each other. We need to construct its wiring map. A network is a catalog of a system’s components often called nodes or vertices and the direct interactions between them called links or edges (Figure 1). This network representation offers a common language to study systems that may differ greatly in
nature, appearance, or scope. For example, the graph models have been constructed for the World Wide Web comprising up to 200 million nodes and 1.5 billion edges (Barabási and Albert, 1999) as well for Phone-Mobile Calls network comprising 2.5 million nodes (customers) and 810 million edges (communications) in a specific period (Lambiotte et al., 2008). The main advantages of the graph theory as a method of complex network analysis are its relative simplicity and high degree of generalizability and interpretability.

Figure iii: Graph G (5, 5) with five nodes and five edges and the corresponding adjacency matrix

A simple graph is represented by \( G = (V, E) \) where \( V \) is the set of nodes and \( E \) is the set of edges. A single edge connects two nodes and can be classified into four different categories: weighted, unweighted (or binary), directed and undirected. Figure iv shows the four main types of graphs. The symmetry operation can be used to transform a directed graph into undirected graph, and the thresholding operation can be applied to transform a weighted graph to unweighted graph. In the weighted directed graph, each node can be identified by integer value \( i = 1, 2, ..., N \) and an edge can be identified by \((i, j)\) represents the connection going from node \( i \) to node \( j \) to which a weight \( w(i, j) \) can be associated. The weighted edge used to signify that the pairwise connections have some numerical values. For example, if a graph represents a transport network, the weights could represent the length of each road. The weighted directed graph can be represented by a weighed adjacency matrix where each element of this matrix represents the weight of an edge between two nodes of the graph (Figure 2, graph C). By applying a threshold on the adjacency matrix of the weighted directed graph we get an unweighted (or binary) directed graph (Figure 2, graph B). In this case, the edge is not associated with a numerical value and considered as binary. For undirected
graph (weighted or unweighted), there are no directions for the edges and the adjacency matrix is symmetric (Figure 2, graph A).

Figure iv: Types of graphs and their transformation. A- Weighted undirected graph. B- Unweighted directed graph. C- Weighted directed graph.
2.2.2 GRAPH MEASURES

Graph can be characterized by different metrics. These metrics can be classified into three groups: global, node-wise and edge-wise metrics. Here we present some of these metrics and their applications in the context of brain network analysis.

SMALL WORLDNESS

The small-world measure of a network was originally introduced by (Watts and Strogatz, 1998, Watts, 1999). It is characterized by an average shortest path length \( L \) equal to the one in the random networks and by higher clustering coefficient \( C \) than the one observed in the regular networks (Figure 3). Briefly, the averaged path length \( L \) (Figure 4-A) is defined as the average minimum number of edges that have to be traversed to reach from one node to the other in the network. The average path length value can be small in the case of random networks or high in the case of regular networks.

![Regular network - High L, High C](image)

![Small world network - Low L, High C](image)

![Random network - Low L, Low C](image)

**Increasingly random connectivity**

Figure 5: Random rewiring procedure for interpolating between a regular network and a random network, without changing the number of nodes or edges in the network.

The clustering coefficient of a node is defined as the number of existing connections between the neighbors of the node divided by all the possible connections between them (Figure 4-B). \( C \) quantifies the extent of local cliquishness or local efficiency of information transfer of a network.
A real network can be considered as a small world network if it is characterized by $\gamma > 1$ and $\lambda \approx 1$ where $\gamma = C / C_{rand}$ and $\lambda = L / L_{rand}$ (Watts, 1999) where $L_{rand}$ and $C_{rand}$ are the characteristic path length and the mean clustering coefficient, respectively, of matched random networks that preserve the same number of edges, nodes and degree distribution as the real network (Maslov and Sneppen, 2002, Sporns and Zwi, 2004, Achard et al., 2006).

**DEGREE**

The degree of a node is the number of edges connected to the node (Figure 4-A). A node that has high degree in a network is often very important as it can influence on its neighbors.

**BETWEENNESS CENTRALITY**

The Betweenness centrality (BC) of a node is equal to the number of shortest paths that run through this node (Figure 4-A) (Freeman, 1977). A node with high betweenness centrality has a large influence on the information flow through the network. The degree and the BC are usually used to define hubs in the network.

**EFFICIENCY**

The efficiency of a network is a measure of how efficiently it exchanges information. The concept of efficiency can be applied to both local and global scales. The global efficiency quantifies the exchange of information across the whole network and it is defined as the inverse of the average path length (Figure 4-A). The random networks are characterized by high global efficiency and the regular networks are characterized by low global efficiency. Conversely, the local efficiency is related to the segregation of information and it is defined as the average of the efficiencies of all subgraphs of neighbors for each network’s nodes.

**MODULARITY**

The modularity consists of partitioning a network into a number of clusters or modules (also called communities). Network modules defined by a subset of nodes in the graph that are more densely connected to other nodes in the same module than the nodes in other modules (Figure 4-C). (Meunier et al., 2010a, Sporns and Betzel, 2016)
Hubness can be measured in many ways. Hubs in graph theory are nodes with high degree or high centrality (Figure 4-C). Also by using modularity, hubness of a node can be measured based on the intra-modular connectivity ($z$-score) and participation coefficient ($PC$) (Sporns and Betzel, 2016). Once the modularity is calculated and optimal modules have been identified, the $z$-score and the $PC$ are measured for each node. The nodes are classified as hubs if their $z$-score is higher than a defined threshold $T$, otherwise they are classified as non-hub. Using $PC$, a hub can be classified as *provincial hubs* where the node is mostly connected to nodes within its own module, *connector hubs* where the node have diverse connectivity across several different modules in a network.
Figure vi: Schematic illustration of some basic graph measures. A- A path correspond to a sequence of unique edges that are crossed when traveling between two nodes in the network. High degree nodes are nodes that have a high number of edges, low degree nodes are nodes that have a low number of edges. B- High clustering coefficient is the high probability that the neighbors of the vertex are also connected to each other. Low clustering coefficient is the low probability that the neighbors of the vertex are also connected to each other. C- A module includes a subset of nodes of the network that show a high level of within-module connectivity and low level of inter-module connectivity. A hub is a high degree node that connects several different modules within the network.
2.3 NEUROIMAGING TECHNIQUES

Neuroimaging is a term that represents the techniques used to map the brain activity. The development of these techniques allowed for considerable progress in our understanding of the brain functions. Neuroimaging techniques can be decomposed into two parts: structural and functional. At large scale, structural techniques like Magnetic Resonance Imaging (MRI) allow to reveal the brain anatomical details, for example, to locate tumors or brain damage. In contrast, functional techniques such as: Electroencephalography (EEG), Magnetoencephalography (MEG) and functional MRI give details about how the brain works by measuring the electric activity or blood flow in the brain. As we are interested in the functional large-scale brain network, we describe here the three main techniques used to map the brain activity at large scale.

2.3.1 FUNCTIONAL MRI (FMRI)

The discovery of fMRI in 1990s is considered as a historical progress in brain research. fMRI was the first noninvasive technique that allowed us to see the brain activity (in real time) with such excellent spatial resolution (Christopher deCharms, 2008). The fMRI measures indirectly the neural activity by means of the BOLD (Blood Oxygenation Level Dependent) signal. The concept of the method is based on the fact that when brain regions become active, such as when a person is performing a cognitive task, these regions need more oxygen than other brain areas. The fMRI technique provides excellent spatial resolution (about 1-3 mm) but on the other hand it is limited in its temporal resolution (1-2 s) (Logothetis et al., 2001). The method was used in most brain research application such as resting state (Van Den Heuvel and Pol, 2010), cognitive task (Telesford et al., 2016) and brain disorders (Koshino et al., 2005), see (Van Den Heuvel and Pol, 2010) for review. However, the fMRI cannot be used for tracking the dynamics of brain networks at sub-second time scale, which is one of our main objectives in this thesis.

2.3.2 MAGNETO/ELECTRO-ENCEPHALOGRAPHY (M/EEG)

Magneto/Electro-encephalography (M/EEG) are techniques that measure magnetic/electrical activity of the brain. The main advantage of the M/EEG techniques is the excellent time resolution (below 1 ms). However, they have low spatial resolution (1-10 cm). Unlike the other electrical recording techniques that require inserting electrodes into the brain (such as stereo-EEG), M/EEG are non-invasive and relatively easy to use techniques. M/EEG were widely used to analyse brain activity during cognitive task (Burgess and Ali, 2002, Vecchio et al., 2016), resting state (Liu et al.,
2010, de Pasquale et al., 2012) and brain disorders (Stam et al., 2009, Dubovik et al., 2013, Fogelson et al., 2013). In this thesis, we used the dense-EEG as a neuroimaging technique. Dense-EEG offers the excellent temporal resolution, a good spatial resolution, an easiness to use and a wider availability in clinics compared to MEG. The dense-EEG system used in the context of this thesis is described in the materials and methods (chapter 3, page 34-35).

2.4 BRAIN CONNECTIVITY

Brain connectivity can be defined as the structural and functional links between different units in the central nervous system. These units can represent individual neurons (Figure vii-A), neuronal assemblies (Figure vii-B), or brain regions (Figure vii-C). The links between these units can be physical/anatomical such as synapses or fiber pathways giving rise to the notion of ‘structural connectivity’ (Sporns et al., 2005). When measuring the statistical/causal relationships between the signals (generated by these units using cross correlation or information flow), we can distinct the so-called functional/effective connectivity (Friston, 1994).

![A - Single neuron, B - Neuronal assembly, C - Brain regions](image)

**Figure vii:** The multiscale spatial organization of brain anatomy. A- Single neuron, B- Neuronal assembly, C- Brain regions

The main characteristics of structural, functional and effective connectivity at the macroscale are presented hereafter.

2.4.1 STRUCTURAL CONNECTIVITY

The structural connectivity, also called anatomical connectivity (Figure viii-A), represents the white matter tracts between pairs of brain regions. It can be mapped using Diffusion Tensor Imaging (DTI), a neuroimaging technique based on the diffusion of water molecules in the brain (Hagmann et al., 2007, Hagmann et al., 2008a). These anatomical connections can be mapped from microscale
(network between neural elements) to macroscale (network of interregional pathways). This anatomical connectivity is symmetric and static (for short periods). It has been shown to be altered in neurological diseases such as Alzheimer, Schizophrenia and Epilepsy (Bernhardt et al., 2008, Jeong et al., 2009, Seeley et al., 2009), and during normal processes associated with development (Hagmann et al., 2010) or learning (Draganski et al., 2004, Scholz et al., 2009).

Figure viii: A- Structural connectivity, B- Functional connectivity, C- Effective connectivity

2.4.2 FUNCTIONAL CONNECTIVITY

Functional connectivity can be defined as the temporal correlation between the activity of distinct neural assemblies, Figure viii-B (Friston, 1994). The basis of all functional connectivity is time series data from neural recording. This data can be extracted by different techniques such as EEG, MEG or fMRI. Several methods have been proposed to measure the functional connectivity by estimating, for instance, the correlation, phase synchronization or coherence function between pairs of time series, see (Pereda et al., 2005, Ansari-Asl et al., 2006, Sakkalis, 2011) for reviews. Unlike structural connectivity, functional connectivity changes over the sub-second scale. In this thesis, we are interested in the functional connectivity. The method used in our work is described in the materials and methods (chapter 3, page 46-47).

2.4.3 EFFECTIVE CONNECTIVITY

Effective connectivity (Figure viii-C) represents the influence that one brain region can have on another region (Friston, 1994). Like functional connectivity, effective connectivity is time varying and can change rapidly (sub-second time scale). Many methods and techniques have been proposed
to estimate the causal interactions between brain regions such as Granger-causality (Granger, 1969) and Dynamic Causal Modeling (Friston et al., 2003), see (Friston et al., 2013) for recent review.

2.5 NODES AND EDGES IN BRAIN NETWORK

A graph is a simple model of the human brain network. This model can be analyzed by the network measures (using graph theory based analysis) whatever the scale. A key question in graphical modeling is to define nodes and edges. At macroscale, in the case of M/EEG for instance, the nodes can be defined as the sensors or cortical regions while edges represent the statistical connections between the recorded signals or the reconstructed sources, respectively. See Table 1 for typical examples of main anatomical brain atlases used to define nodes in brain network.

<table>
<thead>
<tr>
<th>Brain Atlas</th>
<th>Number of Nodes</th>
<th>Maximal number of edges</th>
<th>Anatomical or functional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brainnetome (Fan et al., 2016)</td>
<td>246 regions</td>
<td>30135</td>
<td>Anatomical/Functional</td>
</tr>
<tr>
<td>Automated Anatomical Labeling (Tzourio-Mazoyer et al., 2002)</td>
<td>90 regions</td>
<td>4005</td>
<td>Anatomical</td>
</tr>
<tr>
<td>Desikan-Killiany (Desikan et al., 2006)</td>
<td>68 regions</td>
<td>2278</td>
<td>Anatomical</td>
</tr>
<tr>
<td>Destrieux Atlas (Destrieux et al., 2010)</td>
<td>148 regions</td>
<td>10878</td>
<td>Anatomical</td>
</tr>
</tbody>
</table>
2.6 DYNAMIC RECONFIGURATION OF BRAIN FUNCTIONAL NETWORKS

2.6.1 BACKGROUND

At some time periods, structural, functional and effective brain networks are all dynamic (Sporns, 2011). The structural brain networks remain relatively stable over short time, while functional and effective networks may change at sub-second time scale. Due to poor temporal resolution, fMRI cannot be used to track the dynamics of brain networks over short duration periods, like those associated with responses to cognitive tasks. In contrast, a key advantage of M/EEG is the excellent temporal resolution that offers the unique opportunity not only to track large-scale brain networks over very short duration which is the case in many cognitive tasks but also to analyze fast dynamical changes that can occur in brain disorders (such as interictal or ictal events in epilepsy). This excellent temporal resolution of the EEG signals may allow for the analysis of the dynamic properties of cognitive processes, an issue so far addressed in a few studies only. In (Murray et al., 2008a), authors proposed an algorithm based on the amplitude of event related potentials (ERPs) to follow time-varying voltage topographic maps. However, these algorithms do not account for brain connectivity quantified directly from scalp signals (electrode space) or indirectly from reconstructed brain sources (source space).

Regarding the approaches based on the connectivity analysis, most of reported methods make use of a constant time window to track the dynamics of functional connectivity, as estimated from EEG recordings. This window is typically chosen either empirically or based on a priori information about the analyzed task (Rodriguez et al., 1999). The sliding window approach has been applied widely to study dynamic connectivity of functional brain networks estimated from fMRI (De Vico Fallani et al., 2008, Allen et al., 2012, Bassett et al., 2013, Zalesky et al., 2014, Karahanoğlu and Van De Ville, 2015). Other studies used graph theory metric such as modularity to track the dynamic variation in brain network (Bassett et al., 2011, Shine et al., 2016). However, most of proposed algorithms are not adapted to track changes over very short durations (in the order of 500 ms, as in the case of responses evoked by visual stimuli).

2.6.2 FROM BACKGROUND TO PROPOSED METHODS

In this thesis, we propose a new algorithm that benefits from the excellent temporal resolution of the EEG (at scalp and at cortical source level) to track the fast reconfiguration of functional brain
networks at sub-second time scale. The main objective of the algorithm is to segment brain networks into a set of “functional connectivity states” using a network-clustering approach. We hypothesized that brain networks may stay stable with only few modification in the network characteristic for a given time period (when performing a specific cognitive activity) and thus the ‘fraction’ in the network may reflect a switching to a new cognitive activity. The algorithm is based on $K$-means and was applied on the connectivity graphs obtained by estimating the functional connectivity between scalp EEG signals or between cortical sources reconstructed from scalp EEG. This approach allowed us to summarize the neural networks in a limited number of dominant networks in a given time period during the cognitive process (picture naming). Briefly, the connectivity matrices averaged over all subjects were obtained. Then $K$ networks (3 to 12) were randomly selected and spatial correlations between $K$ network and all other networks were calculated. This gives a spatial correlation value for each value of $K$ at each moment. Finally, the cross-validation criterion was used to determine the optimum number of networks that best explain the cognitive task in progress (see chapter 3 for detailed description of the proposed algorithm).

2.7 BACKGROUND ON GRAPH SIMILARITY

The problem of comparing networks with different properties is a big challenge in many disciplines such as social networks, biological networks, image processing or chemical networks (Willett, 1987, Wiskott et al., 1997, Willett et al., 1998, Dijkman et al., 2011, Emmert-Streib and Dehmer, 2011, Shimada et al., 2016). Several methods were developed to solve the problem of graph similarity in specific context. These methods can be classified into two groups: exact graph matching and inexact graph matching. The result for comparing two graphs by an exact graph matching method is binary, in other words the compared graphs are either the same or not. This group includes the isomorphism-based measures (Cordella et al., 2004). The inexact graph matching methods provide a similarity score between the compared graphs. This group includes mainly the methods based on graph edit distance (Gao et al., 2010) and feature extraction (Bunke et al., 2007).

2.7.1 ISOMORPHISM-BASED METHODS

The methods that belong to this class perform an exact graph matching, which means the similarity score between the compared graphs is binary. In graph isomorphism, the compared graphs must have the same number of nodes and the two graphs are said to be similar if each node in the first
graph have it is proper image (same degree) in the second graph (Figure ix). In the case where the two compared graphs have different number of nodes, there may be a need to identify a sub-graph isomorphic relation between the two graphs. In 1975, (Zelinka, 1975) proposed the first method that measures and quantifies the distance between two graphs based on determining a graph isomorphism where the two compared graphs should have the same number of nodes. Isomorphism-based measures have been applied in different applications and several domains such as Biology, Chemistry (Raymond and Willett, 2002, Bonnici et al., 2013). The main drawback of this family of methods is that the two compared graphs must have the same number of nodes, which is not always possible in the real cases.

Figure ix: Graph isomorphism. An example of two isomorphic graphs $G$ and $G'$, each having four nodes and four edges, and each node in $G$ has it proper image in $G'$.

2.7.2 GRAPH EDIT DISTANCE (GED)

GED belongs to the class of inexact graph matching (Emmert-Streib and Dehmer, 2011). The main idea of the graph edit distance method is to transform one graph to another by defining a set of graph edit operations on nodes and edges such as insertion, substitution and deletion (Bunke et al., 2007, Gao et al., 2010). For each operation, a cost is associated to find the optimal sequence of operations that minimize the cost of transforming between the two graphs. The graph edit distance measure is flexible and widely applied in several applications such as Bioinformatics, pattern recognition and for different types of networks (Bunke, 2000, Dehmer and Emmert-Streib, 2008,
Gao et al., 2010). In Figure x we present the steps of the GED method to transform the graph $G_1$ to graph $G_2$, these steps are the following: deleting the edge $e1$, delete the node $n1$, substitute the $e2$ to $e4$, insert a new node $n2$ and insert a new edge $e3$. The main drawback of this family of method is does not take into account the spatial locations of nodes when computing the graph similarity distance.

2.7.3 FEATURES EXTRACTION

The third approach is based on comparing the topological properties of the two compared graphs (Cha, 2007). The main idea is that two dissimilar graphs have different graph properties and two similar graphs may share similar graph properties such as density, diameter, modularity or eigenvalues for instance (Watts, 1999, Costa et al., 2007). This method is scalable and can be applied on large-scale networks. Again, the available graph metrics do not consider the spatial locations of the compared graphs.

In addition to these three main family of methods, authors proposed a new algorithm called “DeltaCon”, this algorithm is based on the computing of the pairwise of nodes affinities in the first graph and compare them with the ones of the second graph and produce a similarity score (Koutra et al., 2013). Papadimitrious et al. proposed the Vertex/Edge Overlap method (VEO) method for the anomaly detection in web graph, the principle idea of this method is that two graphs are similar if they share many nodes and edges (Papadimitriou et al., 2010). Bunke et al. (Bunke et al., 2007) developed a new method to detect the variation in the communications networks based on the graph edit distance. Many other studies use methods based on feature extraction to measure the distance between graphs such as $\lambda$-distance, which is based on calculating the distance between the eigenvalues of the compared graphs (Bunke et al., 2007, Wilson and Zhu, 2008). Recently Shimada et al. proposed a new method to measure graph distance between networks based on a Laplacian matrix that reflects the structural and dynamical properties of networked dynamical systems (Shimada et al., 2016).
Figure x: Transforming graph $G_1$ to $G_2$ by graph edit distance operations
2.7.4 SIMILARITY IN BRAIN NETWORK CONTEXT

In brain networks analysis, we often face the question of statistically comparing two brain networks. We can distinguish two types of comparison between brain networks: first, the comparison of the real brain networks to random networks, where the main purpose is to validate if some characteristics of the brain networks are significantly different than chance. For a given network, we can generate a large number of random graphs and by applying the graph theory metrics on these random graphs, as a point of reference, to test the randomness of the same metrics measured in the real brain networks. Second, the comparison of brain networks of two groups of subjects such as healthy control and patients via statistical hypotheses. We can classify the techniques used for comparing brain connectivity networks on three categories as following: global-level, node-wise and edge-wise analysis.

GLOBAL-LEVEL ANALYSIS

In this case, the graph metrics (some of them were defined above in section 2.2) are computed over the entire network and one metric value can be derived for each network. Statistical test will be then used to compare the two groups (such as healthy vs. patients). These features such as Small Worldness have been used to demonstrate the small world topology in the human (Hagmann et al., 2007, He et al., 2007) and nonhuman (Hilgetag et al., 2000, Sporns and Zwi, 2004) brain networks. Stam et al. (Stam et al., 2007) reported that the functional brain networks derived from Alzheimer disease patients have the characteristics of random networks with characteristic path length significantly longer than the control subjects while the cluster coefficients showed no significant differences. Other studies showed the presence of the small world characteristics in the brain connectivity of healthy subjects whereas these characteristics were disrupted in schizophrenia patients (Liu et al., 2008) (Lynall et al., 2010) (Micheloyannis et al., 2006).

A number of studies found significant differences between the brain networks of two groups by comparing their modules. Meunier et al. (Meunier et al., 2010a) investigated the modular structure of the human brain networks derived from fMRI measurements for two groups of younger and older adults using modularity maximization algorithm where the main objective was to find a change in the modularity between the two groups. The results show that the brain becomes less modular with age.
In pathological applications, it was shown that modularity decreases for schizophrenia patients compared to health subjects (Alexander-Bloch et al., 2010). Achard et al. (Achard and Bullmore, 2007) used the efficiency to compare the functional brain networks between two healthy groups (old vs. young subjects). Authors showed that the efficiency was reduced in older people. In addition, many studies showed that patients with schizophrenia and Parkinson diseases have a marked reduce in global efficiency compared with healthy controls (Liu et al., 2008, Alexander-Bloch et al., 2010, Becerril et al., 2011, Skidmore et al., 2011).

**NODE-WISE ANALYSIS**

In this case, the graph metrics are calculated for each node and then the node’s metric values are compared between the two graphs. Many studies were performed to compare between brain networks of healthy subjects and patients diagnosed with schizophrenia using the degree of nodes (Bassett et al., 2008, Liu et al., 2008, Lynall et al., 2010, Wang et al., 2010). The results showed reduced degree in several brain nodes of schizophrenia patients. Other studies showed also that the Parkinson disease patients have significant decrease in the degree of several brain regions in their functional network such as left dorsal lateral prefrontal cortex and had a significant increase in the degree of the left cerebellum (Wu et al., 2009). Using betweenness centrality, Yan et al. (Yan et al., 2011) investigated the effects of sex on the topological organization of human cortical anatomical network. In clinical application, betweenness centrality was used to compare brain networks of healthy subjects and patients with schizophrenia, depression and Alzheimer disease (van den Heuvel et al., 2010, Yao et al., 2010, Becerril et al., 2011, Zhang et al., 2011a, Shi et al., 2012). Another graph metric was used to compare two groups of brain graphs is the hubness, Many studies showed that the brain disorders are associated with alterations in the Hubs such as Alzheimer diseases, comatose patient and schizophrenia (Bassett et al., 2008, He et al., 2008, Lynall et al., 2010, van den Heuvel et al., 2010, Wang et al., 2010, Zhang et al., 2011b, Achard et al., 2012, Crossley et al., 2014). This type of analysis required statistical correction for multiple comparisons, as comparisons are performed N times where N is the number of nodes. This can be done by using methods such as Bonferoni (Rice, 1989) or False Discovery Rate (FDR) (Genovese et al., 2002).

**EDGE-WISE ANALYSIS**

The edge-wise analysis consists of calculating a statistical test (such as student t-test) on each edge in the graph. If the number of nodes in a graph equal to N then the maximum number of edges equal
to \( N \ (N - 1) / 2 \). The statistical test is performed \( N \ (N - 1) / 2 \) time for one graph. This method required also statistical correction for multiple comparisons using methods such as Bonferroni or FDR. However, these methods do not deal with the fact that we are comparing ‘networks’. Recently, Zalesky et al. (Zalesky et al., 2010) proposed a new way (adapted to networks) for correcting this family-wise error rate (FWER) using permutation test, a method called network-based statistic (NBS). The main objective of this method is to find a network “pattern” (a set of nodes connected by edges) that differentiates between two network groups. The NBS has been used to identify alteration in brain networks associated with psychiatric disorders such as schizophrenia and depression (Zalesky et al., 2011, Zhang et al., 2011a), and to identify enhanced brain connectivity in talented subpopulations such as elite athletes (Wang et al., 2013). Again, the method does not deal with the physical locations of the nodes.

2.7.5 FROM BACKGROUND TO PROPOSED METHODOLOGY
The above-described algorithms were successfully applied to compare graphs in different specific application. However they all ignore the spatial location of nodes when measuring the similarity indices. In some situation in brain networks analysis, this spatial location of nodes seems to be crucial to characterize a graph. Two brain connectivity networks that may share common metrics such as small worldness, degree, density, but they are located in different brain regions can be considered as functionally dissimilar. In return, two brain connectivity networks that have differences in graph topology metrics but they share common brain regions are spatially and functionally close. Here, we take this feature into consideration to develop a new algorithm that quantifies the similarity between two networks where the nodes coordinates are known. The algorithm, called SimNet, will be described briefly hereafter and detailed in the materials and methods chapter.

SIMNET ALGORITHM
In order to compare the brain connectivity networks, we developed a network similarity algorithm called ‘SimNet’. SimNet is based on two main parts that are necessary to map one graph to the other: the first step is based on calculating the nodes distance: in this part the algorithm is based on four main steps: i) detection of common nodes between the two compared graphs, ii) substitution between two nodes where the cost of substitution equals to the distance between the substituted nodes, iii) insertion for new nodes where the cost of insertion equals to a constant value and iv)
deletion of nodes where the cost of suppression equals to the cost of insertion. The (cost (substitution) < cost (Insertion) + Cost (Deletion)) is always preserved. The second step is to calculate the edges distance. It consists on calculating the sum of the weight difference between two edges of two compared graphs. SimNet is also as any other distances preserved the four equations of distance: non-negativity, identity, symmetric and triangular inequality. The algorithm provides a normalized Similarity Index (SI): 0 for no similarity and 1 for two identical networks (same properties and topology). The algorithm will be described in details in chapter 3, page 51-54.
CHAPTER 3- MATERIALS AND METHODS

In this chapter we present the materials and the methods used for collecting and analyzing data in this thesis. First we present the dense-EEG recording technique used to collect EEG signals that reflect the brain electrical activity. Second, we describe the two main analysis performed in line with the two objectives of this thesis: i) the tracking of the dynamics of functional brain networks at scalp and cortex level during picture naming task and ii) the analysis of the similarity between brain networks. In the later, two studies were conducted. The first one was aimed at comparing functional networks associated with two categories of visual stimuli while the second study was performed to compare networks obtained from scalp EEG and intracerebral EEG recorded from same epileptic patients.

3.1 DENSE ELECTROENCEPHALOGRAPHY

Electroencephalography (EEG) is a technique to measure the electrical brain activity. It is typically noninvasive using electrodes placed on the scalp. The recording of EEG signals at the level of scalp electrodes is possible because currents generated at the level of neuronal sources flow brain tissues (grey mater, hite matter, CSF, bone and scalp) also known as the volume conductor (Niedermeyer and da Silva, 2005).

Figure xi: Dense EEG system (EGI, 256 electrodes)
Historically, Hans Berger (1924) recorded the first human brain waves (Collura, 1993). He introduced the two major brain waves “alpha” and “beta”. By the mid-1930s, commercial EEG systems begun to appear, Albert Grass produced the first commercial EEG device. His first model appeared in 1935, which had three channels of differential amplification and an ink-writer that recorded on rolls of paper. By the end of the World War II, he introduced the second model, which was the first 8-channel and 16-channel EEGs ever made. About 5,000 systems were produced and were shipped around the world (Collura, 1993). Since then, major technological advances were made in the recording of EEG data.

EEG captures the electrical activity by the electrodes placed on the scalp at certain predefined standard position. To achieve an appropriate EEG recording, electrodes are placed based on internationally recognized 10-20 system in which the electrodes are placed proportionally (10% and 20% spacing) between different bone landmarks (inion, nasion and preauricular points). In this system, electrode locations are named to identify the lobe and numbered to identify hemisphere location. By convention, the letters F, T, C, P and O stand for frontal, temporal, central, parietal, and occipital lobes, respectively. Even numbers refer to electrode positions on the right hemisphere, whereas odd numbers refer to those on the left hemisphere. The electrodes are then connected to an amplifier and the output reveals a variation over time.

The EEG is a largely used technique to record the brain activity since it has a high temporal resolution (in the order of ms) and it is relatively easy-to-use comparing to other techniques. This high temporal resolution of EEG makes it very suitable to study the dynamics of brain activity. The spatial resolution of EEG depends on the number of electrodes used in the experiment. It is usually called ‘dense EEG’ when the number of electrodes is higher than 64. In this thesis we use the highest available spatial resolution so far, (Figure 9), with 256 electrodes. The main feature of this system is the large coverage of the subject's head by surface electrodes allowing for the improved analysis of the intracerebral activity from non-invasive scalp measurements, as compared with 32- to 128- electrodes standard systems (Hassan et al., 2013, Hassan et al., 2014).

### 3.2 DATA

To evaluate the algorithms developed in this thesis, we used data recorded during different paradigms. A picture-naming task was used to evaluate the algorithm developed to track the functional brain networks at hundreds millisecond time scale. SimNet was evaluated on picture naming task also when presenting different category of visual stimuli, on one hand, and on data
recorded from epileptic patients to compare between networks identified from scalp and intracerebral EEGs, on the other hand. SimNet was used also to compare between networks obtained by the dense-EEG source connectivity method and network generated by computational model (ground truth) in order to evaluate different inverse/connectivity methods.

### 3.2.1 COGNITIVE TASK

Here we describe the cognitive task performed in our thesis. The task consists of naming pictures presented on a screen in front of participants. This task is very fast (<1s) and was used to evaluate the segmentation algorithm developed to track functional brain networks. Another advantage of this simple task is that a ‘solid background’ is available about the brain regions and networks involved during the task (using other techniques such as fMRI and MEG for instance, (Price, 2012)). Two cognitive tasks were performed in this study. The first one called ‘ConneXion’ and consists of naming (and spelling) pictures without a special interest in the picture category. The second task called ‘BrainGraph’ was conducted in the view of looking at the brain networks associated with visual stimuli with different semantic content such as meaningful (tools, animals...) and meaningless (scrambled) pictures.

**CONNEXION**

Twenty one right-handed healthy volunteers (11 women: mean age 28 year; min: 19, max: 40 and 10 men: mean age 23 years; min: 19, max: 33), with no neurological disease, were involved in this task. Participants were asked to name at a normal speed 148 displayed pictures on a screen using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) (Schneider et al., 2002). The images were selected from a database of 400 pictures standardized for French (Alario and Ferrand, 1999) and were used during session about eight minute. They were controlled according to several parameters (see Table 2). All pictures were shown as black drawings on a white background (Figure xii). Order of presentation was randomized across participants. Naming latencies were determined as the time between picture onset and the beginning of vocalization recorded by the system. Oral responses were recorded and then analyzed with Praat software to set the voice onset time (Boersma, 2002). This study was approved by the National Ethics Committee for the Protection of Persons (CPP), conneXion study, agreement number (2012-A01227-36) and promoter: Rennes University Hospital. All participants provide their written informed consent to
participate in this study. The ethics committee has approved the consent procedure. A typical trial started with the appearance of an image during 3 sec followed by a jittered inter-stimulus interval of 2 or 3 sec randomly. Most responses were given while the image was still present on the screen. Errors in naming were discarded for the subsequent analysis. A total of 2926 on 3108 events were considered. The fastest response time delay for an event was <600 milliseconds. The brain activity was recorded using dense-EEG, 256 electrodes, system (EGI, Electrical Geodesic Inc.). EEG signals were collected with a 1 kHz sampling frequency and band-pass filtered between 3 and 45 Hz.
Figure xii: Pictures selected from the Alario database (http://www.lexique.org).
Table 2: Psycholinguistic parameters controlled to get equivalent datasets for each category of pictures in dataset 1 (Animals vs Tools)

<table>
<thead>
<tr>
<th></th>
<th>Name agreement (%)</th>
<th>Image agreement (average)</th>
<th>Age of acquisition (average)</th>
<th>Number of letters</th>
<th>Number of phonemes</th>
<th>Number of syllables</th>
<th>Number of morphemes</th>
<th>Oral frequency</th>
<th>Written frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Animals</strong></td>
<td>94.3243</td>
<td>3.7550</td>
<td>2.3292</td>
<td>5.9730</td>
<td>4.2162</td>
<td>1.7297</td>
<td>1.0541</td>
<td>12.1324</td>
<td>11.1673</td>
</tr>
<tr>
<td><strong>Average Objects</strong></td>
<td>95.8378</td>
<td>3.6126</td>
<td>2.5289</td>
<td>6.1622</td>
<td>4.2703</td>
<td>1.7027</td>
<td>1.1351</td>
<td>9.1903</td>
<td>11.0351</td>
</tr>
<tr>
<td><strong>STDEV Animals</strong></td>
<td>7.5058</td>
<td>0.4481</td>
<td>0.6152</td>
<td>1.9506</td>
<td>1.7502</td>
<td>0.8383</td>
<td>0.2292</td>
<td>13.9555</td>
<td>11.3952</td>
</tr>
<tr>
<td><strong>STDEV Objects</strong></td>
<td>5.8572</td>
<td>0.7356</td>
<td>0.5819</td>
<td>1.9792</td>
<td>1.6608</td>
<td>0.7769</td>
<td>0.4191</td>
<td>6.9085</td>
<td>6.8723</td>
</tr>
<tr>
<td><strong>T-test Animals vs. Objects</strong></td>
<td>0.3370</td>
<td>0.3189</td>
<td>0.1557</td>
<td>0.6800</td>
<td>0.8920</td>
<td>0.8860</td>
<td>0.3064</td>
<td>0.2556</td>
<td>0.9520</td>
</tr>
</tbody>
</table>
Twenty right-handed healthy volunteers with no neurological disease were included in this study. Participants were asked to name 120 displayed pictures (80 meaningful and 40 meaningless) on a screen using EPrime 2.0 software (Psychology Software Tools, Pittsburgh, PA) (Schneider et al., 2002). Images were selected from the same database used in Connexion and were used also during a session of ten minutes (Figure 11). 40 meaningless pictures were selected randomly from the 80 meaningful pictures to have same number of pictures for both conditions. All pictures were shown as black drawings on a white background. Order of presentation was randomized across participants. Naming latencies were determined as the time between picture onset and the beginning of vocalization recorded by the system. Oral responses were also recorded and then analyzed with Praat software to set the voice onset time (Boersma, 2002). Errors in naming were rejected for the following analysis. This study was approved by the National Ethics Committee for the Protection of Persons (CPP), BrainGraph study, agreement number (2014-A01461-46), and promoter: Rennes University Hospital. All participants provide their written informed consent to participate in this study. The ethics committee has approved the consent procedure.
3.2.2 EPILEPTIC DATA

To investigate the possible use of the developed network-similarity algorithm in the context of clinical application, we used simulated and real epileptic data. The simulated dense EEG data were generated from physiologically- and biophysically- plausible model of distributed and coupled epileptic sources. The real data (scalp and intracerebral EEG) were recorded from epileptic patients who underwent a full presurgical evaluation for drug-resistant focal epilepsy.
In order to evaluate the different methods used to reconstruct brain networks at cortical level from dense EEG. We have conducted a comparative study between several inverse solution algorithms and functional connectivity methods. To do so, we used data generated from physiologically- and biophysically- model of distributed and coupled epileptic sources. The network generated by the model was considered as ‘reference’ network and then we applied the dense-EEG source connectivity method (described later in section 3.3) on the scalp EEG, generated by the model after resolving the forward problem, to identify epileptic network. The network similarity algorithm was used as a key criterion to select the best inverse/connectivity combination by computing the similarity index between networks identified using each combination and the reference network.

Technically, we generated simulated EEG signals following the procedure described in (Cosandier-Rimélé et al., 2007). The distributed source space consisted in a mesh of the cortical surface (8000 vertices, ~5 mm inter-vertex spacing) that was obtained by segmenting the grey-white matter interface from a normal subject’s structural T1-weighted 3D-MRI with FreeSurfer (Fischl, 2012). Dipoles were located at each vertex of this mesh and oriented radially to the surface at the midway between the white/grey matter interface and the pial surface. The time-course of each dipole of the source space was generated from a modified version of the physiologically relevant neural mass model reported in (Wendling et al., 2002). In brief, this computational model was designed to represent a neuronal population with three subsets of neurons (pyramidal cells P and interneurons I and I’) interacting via synaptic transmission. Pyramidal cells (P) receive endogenous excitatory drive (AMPAergic collateral excitation) from other local pyramidal cells and exogenous excitatory drive from distant pyramidal cells (via external noise input p(t)). They also receive inhibitory drive (GABAergic feedback inhibition) from both subsets of local interneurons (I and I’). In turn, interneurons receive excitatory input (AMPA) from pyramidal cells. A Gaussian noise was used as external input to neuronal population. The mean (m = 90) and standard deviation (sigma = 30) were adjusted to represent randomly varying density of incoming action potentials (Aps). However, for the purpose of this study, a modification was made to this noise model. Indeed, abrupt increase/decrease of the density of Aps can occur in the external input noise at userdefined times to mimic transient AP volleys from other brain regions involved in the generation of interictal events. Thus, in this model, a simulated IES can be viewed as the
responses of a nonlinear dynamical system (comprising positive and negative feedback loops) to transient pulses superimposed on a Gaussian noise (classically used in the neural mass modeling approaches). As in the standard implementation, the shape (spike component followed by a wave component) can still be controlled by adjusting excitation and inhibition parameters of each population (gains in feedback loops). However, the aforementioned modification offers one major advantage: as pulses in the noise input are user-defined, the occurrence times of simulated IESs are controlled, in contrast with the standard implementation where IESs simply result from random fluctuations of the noise. The consequence is that this new model feature allows for simulation multi-focal IESs with well-controlled time shifts. Indeed, we could generate delayed epileptiform activity in multiple distant patches just by introducing short delays between the pulses superimposed on the respective input noises of neuronal populations at each patch. Finally, from appropriate setting of the input noise, as well as excitation and inhibition-related parameters at each neural mass included in simulated epileptic sources, a set of epileptiform temporal dynamics was obtained. These dynamics were assigned to a source made of contiguous vertices (active source) manually outlined with a mesh visualization software (Paraview, Kitware Inc., NY, US). Uncorrelated background activities were attributed to the other vertices. Once the amplitude of each elementary dipole was known, EEG simulations were obtained by solving the forward problem in a 3-layer realistic head model (inner skull, outer skull and the scalp with conductivity values of 0.33, 0.0042, 0.33 S/m respectively) using the Boundary Element Method (BEM) with the OpenMEEG (Gramfort et al., 2010) implemented in Brainstorm software. We considered two different scenarios. In the first one (single network), EEG simulations were generated from a single source located in the inferior parietal region (~1000 mm²). In the second one (two interconnected networks) an additional source (~1000 mm²) was placed in the middle temporal gyrus. In that case, the temporal dynamics of the second source were highly correlated with those of the first source, but with a minor delay (30 ms). This delay of 30 ms was in the range of 10–50 ms delays that are often observed during interictal spikes at different intracranial recording location (Alarcon et al., 1997) or at different surface sensors (Ebersole, 1994) or between the peaks of dipole source activity (Merlet and Gotman, 1999). This delay was usually interpreted as reflecting propagation between distant regions in the brain. For each scenario, 20 epochs of 60 s at 512 Hz containing 30 epileptic spikes were simulated. Each epoch was obtained for a new realization of the input random noise leading to a new realization of epileptic spikes occurring in background activity.
Real data were selected from patient who underwent presurgical evaluation for drug-resistant focal epilepsy. Seizures were stereotyped, with a sudden start, febrile motor automatisms of the upper limb, stretching of the neck and torso and no post-ictal motor deficit. The patient had a comprehensive evaluation including detailed history and neurological examination, neuropsychological testing, structural MRI, standard 32-channels (Micromed, Italy) as well as Dense-EEG 256-channels (EGI, Electrical Geodesic Inc., Eugene, USA) scalp EEG with video recordings and intracerebral EEG recordings (SEEG). MRI showed a focal cortical dysplasia in the mesial aspect of the orbito-frontal region. Dense-EEG was recorded for 1 h, at 1000 Hz following the procedure approved by the National Ethics Committee for the Protection of Persons (CPP, agreement number 2012-A01227-36). The patient gave his written informed consent to participate in this study. This recording revealed sub-continuous spike activity at the most left frontopolar basal electrodes. From this interictal epileptic activity, 85 spikes were visually selected away from the occurrence of any artefacts (muscle activity, blood pulsation, eye blinks). Each spike was centered in a 2 s window and all 85 windows were concatenated for further analysis. As part of his presurgical evaluation, the patient also underwent intracerebral SEEG recordings with 9 implanted electrodes (10 ± 18 contacts; length: 2 mm, diameter: 0.8 mm; 1.5 mm apart) placed intracranially according to Talairach’s stereotactic method in the left frontal and temporal region. The positioning of the electrodes was determined from available non-invasive information and hypotheses about the localization of his epileptic zone. From these data, subsets of 25 out of the 118 original leads were selected. This selection was made according to the following criteria: i) only contacts showing grey matter activity were retained and ii) among them, only the contact showing the maximal activity was kept when similar intracerebral activity was observed on several contacts.

_in this case, the network similarity algorithm was used to compare between the networks identified using the scalp dense EEG and the networks identified from the intracerebral EEGs._
3.3 DATA ANALYSIS

3.3.1 DYNAMIC RECONFIGURATION OF BRAIN FUNCTIONAL NETWORKS
After collecting the data from the two cognitive tasks, each trial was visually inspected, and epochs contaminated by eye blinking, movements or any other noise source were rejected and excluded from the analysis performed using the EEGLAB and Brainstorm toolboxes (Delorme and Makeig, 2004, Tadel et al., 2011). The analysis was performed on both scalp and cortex level using the dataset1 and dataset2. At the source level, the analyzing steps can be summarized as follows (Figure xiv): (i) solving the ill-posed (number of sources >> number of electrodes) EEG inverse problem, (ii) estimate the statistical couplings (functional connectivity) between reconstructed sources, (iii) characterizing the identified networks (in the form of nodes connected by edges forming a graph) by graph theory based analysis and (iv) segmenting, in time, the cognitive process as a sequence of functional connectivity states (fcSs). In the next section, we present the methods used to reconstruct the functional brain networks at both scalp and cortical level and the developed segmentation algorithm.

FROM SCALP TO CORTEX
Most of the previous studies on functional connectivity analyses from M/EEG were performed at the sensor level. The interpretation of corresponding networks is not straightforward as signals are severely corrupted by two factors: i) the ‘volume conduction’ effect due to the electrical conduction properties of the head and ii) the ‘field spread’ problem as multiple scalp channels can collect activity from single brain source (Nolte et al., 2004, Van Diessen et al., 2015). To overcome these limitations, the recent past years have witnessed a significant increase of interest for M/EEG analysis of functional brain networks at the level of cortical sources. This approach, called M/EEG source connectivity, is conceptually very attractive as networks can be directly identified in the cortical source space.

The method was shown to reduce (or totally remove in some cases) the effect of the volume conduction and field spread. Nevertheless, two main technical steps should be carefully considered when using the method. First, it needs to solve the ill-posed M/EEG inverse problem. Second, a functional/effective connectivity measure must be chosen among the many available ones. Regarding the first step, different methods have been proposed to resolve the inverse problem. The most popular algorithms include Minimum Norm Estimate (MNE) and weight Minimum Norm
Estimate (wMNE) (Hämäläinen and Ilmoniemi, 1994, Lin et al., 2004), Low resolution brain electromagnetic tomography (LORETA) and standardized low resolution brain electromagnetic tomography (sLORETA) (Pascual-Marqui et al., 1994, Pascual-Marqui, 2002). Regarding the second step, several approaches have been proposed and can be divided into two main categories depending on the assumptions made about the statistical coupling between signals. The first category includes linear methods such as the linear cross-correlation (Brazier and Casby, 1952), or the coherence function (Brazier, 1968). The second category includes nonlinear methods based on mutual information (Maars and Da Silva, 1983), nonlinear regression (Wendling et al., 2001), and phase synchronization (Rosenblum et al., 2001). These two factors have been shown to have high impact on the resultant networks (Hassan et al., 2014, Hassan et al., 2015a, Hassan and Wendling, 2015). Hassan et al. showed that a combination of the wMNE with the PLV using high resolution EEG is the best combination among the tested combination. This combination was used during the analysis of the two cognitive tasks described above. In the next sections we detailed these two methods.

### EEG INVERSE PROBLEM

According to the linear discrete equivalent current dipole model (ECD), EEG signals $X(t)$ measured from $M$ channels can be expressed as linear combinations of $N$ dipolar source time series $S(t)$:

$$X(t) = A.S(t) + B(t)$$ \hspace{1cm} (1)

Where $A$ are respectively the matrix containing the lead fields of the dipolar sources for the given $M$-channel of the dense-EEG and $B(t)$ represents the additive noise. In the general case, the inverse problem consists in finding an estimate $\hat{S}(t)$ of the dipolar source parameters (typically, the position, orientation and magnitude), given the EEG signals $M(t)$ and given the gain matrix $A$. This matrix can be computed from a multiple layer head model (volume conductor) and from the position of electrodes. For instance, the Boundary Element Method is a numerical method classically used in the case of realistic head models. As this problem is ill-posed ($N >> M$), physical and mathematical constraints have to be added to obtain a unique solution among the many solutions that minimize the residual term in the fitting of measured EEG signals. Using segmented MRI data, the source distribution can be constrained to a field of current dipoles homogeneously distributed over the cortex, and normal to the cortical surface.
Using this source space, the wMNE method only estimates the moment of dipole sources. The wMNE compensates for the tendency of classical MNE to favor weak and surface sources. This is done by introducing a weighting matrix \( W_s \):

\[
\hat{S}_{\text{wMNE}} = (A^T W_s A + \lambda I)^{-1} A^T W_s m
\]

where matrix \( W_s \) adjusts the properties of the solution by reducing the bias inherent to MNE solutions. Classically, \( W_s \) is a diagonal matrix built from matrix \( A \) with non-zero terms inversely proportional to the norm of the lead field vectors. The choice of \( \lambda \) is important and many approaches have been proposed to estimate it although there is no agreement on any optimal solution. Here, the value of \( \lambda \) is computed relatively to the signal to noise ratio for each signal computed as the ration between the post-stimuli period to the pre-stimulus.

**FUNCTIONAL CONNECTIVITY**

Functional connectivity is defined as the temporal correlation (wide sense) among the activity of different neuronal assemblies. Several methods have been proposed to characterize the brain functional connectivity. In this thesis and based on the comparative analysis (Hassan 2014), we used the phase locking value (PLV) method proposed by (Lachaux et al., 1999) to measure synchronization on a trial by trial basis. For each source pair, \( x \) and \( y \), at time \( t \) (\( t = t_1, \ldots, t_T \), where \( T = D^* f_s \); \( D \) and \( f_s \) denote the signal length relative to the onset and the sampling frequency, respectively) for the \( Tr \) trials and for subject \( j \) (\( j = 1 \ldots \) number of subjects), \( \text{PLV} \) is defined as:

\[
\text{PLV}_{xy}^j(t) = \frac{1}{T r} \left| \sum_{n=1}^{T r} \varphi_x(t) - \varphi_y(t) \right|
\]

The functional connectivity was computed at the low gamma band (30-45 Hz). This frequency band is the most relevant one in the context of the similar cognitive task performed by the subjects, as reported in (Rodriguez et al., 1999). The PLVs were then averaged over subjects:

\[
\overline{\text{PLV}}_{xy}(t) = \frac{1}{M} \sum_{j=1}^{M} \text{PLV}^j_{xy}(t)
\]

where \( \overline{\text{PLV}}_{xy}(t) \) represents the general term of the average adjacency matrix \( \overline{\text{PLV}}(t) \) which defines a functional connectivity graphs \( G \) at each time \( t \), \( G=\{G(t), t=t_1, \ldots, T\} \) computed for the \( V \) pairs of \( x \) and \( y \) sources, where \( V \) is equal to \( (Nc.(Nc-1)/2) \) and \( Nc \) is the number of ROIs (regions of...
interest). PLV method is normalized so values range from 0 (independent signals) to 1 (fully synchronized signals).

For dataset1 and as MRIs were not recorded for each participant, we used a template brain to register a labeled mesh from an average brain, where each label corresponded to one of 148 anatomical cortical regions (Destrieux et al., 2010). This output provided a standardized partition of the cortex into 148 regional areas. To test the effect of the number of regions on the results, each of these areas was then subdivided into a set of small sub-regions using Brainstorm resulting in 1000 ROIs covering the whole cortex (Tadel et al., 2011).

For dataset2 and as MRIs for each participant were recorded, we used Freesurfer to register a labeled mesh for each subject’s brain, where each label corresponded to one of 68 anatomical cortical regions (Desikan et al., 2006). This output provided a standardized partition of the cortex into 68 ROIs considered as nodes in the brain network. The threshold value applied to the connectivity matrix is an open issue. Many approaches have been used such as i) keeping a constant number of edges such as ~10000 edges (Hagmann et al., 2008b) or ii) empirically threshold the connectivity matrix based on the edges weights. Here we decided to keep the highest 10% as the proportion-based threshold was shown to be much better than weight-based threshold (Garrison et al., 2015). The consistence of our results was evaluated with other threshold values.

At the scalp level, the functional connectivity was estimated by applying PLV directly on the EEG signals provided by the scalp channels without resolving the inverse problem to reconstruct cortical sources. After estimating the functional connectivity during time, we track the dynamic of the brain networks using the segmentation algorithm described hereafter.

SOFTWARE

In this thesis the processing steps were applied on real data using several tools and toolboxes. we used: Freesurfer (Fischl, 2012) for surface parcellation http://freesurfer.net, EEGLAB (Delorme and Makeig, 2004) for EEG preprocessing https://sccn.ucsd.edu/eeglab/, Brainstorm for source reconstruction (Tadel et al., 2011) http://neuroimage.usc.edu/brainstorm/. We computed the network measures using BCT toolbox (Rubinov and Sporns, 2010) https://sites.google.com/site/bctnet/ , and for the graph visualization we used EEGNET (Hassan et al., 2015b) which is developed in our team https://sites.google.com/site/eegnetworks/.
SEGMENTATION ALGORITHM

The objective of this algorithm is to identify clusters among the $T$ graphs $G(T)$. The proposed algorithm is based on the following steps (Figure xv):

**STEP 1: INITIALIZATION**

To start, $K$ graphs $G^k$, $G^k = \{\overline{G^k}, k = t_i, \ldots, t_K\}$, are selected where $k=t_i$ and $l$ is randomly chosen in $[1,T]$ ($K$ varies from 3 to 12 and $k$ varies from 1 to $K$). To avoid the choice of temporally close
graphs, an additional constraint was added which consists in rejecting the $K$ graphs if the time interval between two $t_l$ is less than 30 ms.

Figure xv: Flowchart of the proposed segmentation algorithm

**STEP 2: ASSIGNMENT**

The spatial correlation $sC^k(t)$ between $G(t)$ and $\overline{G^k}$ is then computed as follows:

$$sC^k(t) = \frac{\sum_{i=1}^{V} G_i^k \cdot G_i(t)}{\sqrt{\sum_{i=1}^{V} G_i^k \cdot G_i^k} \cdot \sqrt{\sum_{i=1}^{V} G_i^2(t)}}$$  \hspace{1cm} (5)

Where $i$ denotes the $i^{th}$ edge in $G(t)$ and $\overline{G^k}$. As depicted in equation (5), $sC$ is normalized by the variance of graphs $G$ and $G^k$. Thus, $sC$ ranges from 0 to 1, high values denote graph with high similarity. Conversely, low values are indicative of low similarity between graphs.
Each graph $G(t)$ is then assigned to the cluster for which the spatial correlation was the highest. The assigned clusters are defined as $\hat{G}^k$:

$$\hat{G}^k = \{G(t) : sC^k_{G(t), \hat{G}^k} \geq sC^{k'}_{G(t), \hat{G}^{k'}} \forall 1 \leq k' \leq K\}$$ (6)

From these spatial correlation values, the Global Explained Variance (GEV) is calculated as defined in (Murray et al., 2008b):

$$GEV = \sum_{k=1}^{K} GEV^k$$ (7)

$$GEV^k = \sum_{t=1}^{t_{\text{max}}} (sC^k_{G(t), \hat{G}^k})^2 \gamma_{G(t), \hat{G}^k} \text{ where } \gamma_{G(t), \hat{G}^k} = \begin{cases} 1 & \text{if } G(t) \in \hat{G}^k \\ 0 & \text{if } G(t) \notin \hat{G}^k \end{cases}$$ (8)

**STEP 3: UPDATE**

At each iteration, the new centroids $\overline{G}^k$ are updated by averaging all the graphs yielding to the same cluster:

$$\overline{G}^k = \frac{1}{|\hat{G}^k|} \sum_{G^* \in \hat{G}^k} G^*$$ (9)

For each $K$, steps 2 and 3 were repeated 500 times. The set of centroids leading to the highest GEV was retained. When the algorithm converges (reaching the highest GEV), $K+1$ graphs $\overline{G}$ are then selected randomly and the entire above procedure (from step 2 to step 3) is repeated until $K=12$. To choose the optimal number of clusters, we used a method based on the Cross Validation (CV) criterion (Murray et al., 2008b) which is a ratio between the GEV and the degrees of freedom for a given set of graphs. As reported, the global minimum of this criterion gives the optimal number of segments. Note that in the same segment, the graphs can have different $SC$ values with the same cluster and therefore two consecutive graphs (in time) can be classified in two distinct clusters. To overcome this, the decision that is made is to choosing the ‘most dominant’ cluster inside each segment.
3.3.2 BRAIN NETWORKS SIMILARITY

In this paragraph we will describe the methodological aspects of SimNet. The developed algorithm was applied on the datasets described above. The first application on dataset 1 aimed to differentiate between functional brain networks associated with different visual stimuli (dataset 1: animals vs. tools and dataset 2: meaningful vs. meaningless pictures). The network similarity algorithm was also used to compare between the networks identified using the scalp dense EEG and the network identified from the intracerebral EEGs.

SIMNET ALGORITHM

The proposed algorithm, called SimNet and described in Figure xvi, calculates a distance between two graphs $G_1$ and $G_2$ located into a grid $Gr$ (Figure xvi). This distance is based on i) the cost of a sequence of cost-based changes (substitution, insertion, deletion) on nodes that are necessary to map $G_1$ onto $G_2$ and ii) the computation of the difference of edge weights between $G_1$ and $G_2$. The proposed algorithm includes four main steps, as illustrated in Figure xvi, B-E:

STEP 1: NODE SUBSTITUTION

To start, the nodes common to $G_1$ and $G_2$ that have the same spatial location are detected (these nodes have the same index value $K_i$ in $G_1$ and $G_2$). In the example of Figure xviB-i, we observe three nodes of $G_1$ having the same spatial position as three nodes of $G_2$. For the remaining nodes, we define a spatial neighborhood $\Omega$ (disk with radius $R=1.5$ as depicted in Figure xvi, B-ii). The idea is to shift a neighbor node $v_j$ included in $V_1$ to the spatial position of $v_2$ included in $V_2$ if $v_j$ is located in $\Omega$ (the defined spatial neighborhood of $v_2$). In the case where there are $n>1$ nodes of graph $G_1$ located in $\Omega$, the nearest node to $v_2$ is shifted. The cost of shifting is equal to the Euclidian distance between the two shifted nodes ($v_j$ and $v_2$). The node distance ($ND$) is updated by the Euclidian distance between $v_j$ and $v_2$ (e.g. $ND=1$ in Figure xviB-ii).

STEP 2: NODE INSERTION
In step 2, the insertion operation is used to add a new node in $G_1$ at the same position of node $v_2$ in the case where no node exists in the defined spatial neighborhood $\Omega$ for each node $v_2$ of $V_2$. The cost of insertion must be higher than the cost of substitution. In the example of Figure xvi C, the cost of insertion was chosen equal to disk with radius $r=1.5$. In this case, the node distance ($ND$) is updated to $1+1.5=2.5$.

**STEP 3: NODE DELETION**

At the end of step 2, all the nodes that are included in $G_2$ are also included in $G_1$ (Figure xviD). In step 3, the remaining nodes of $G_1$ are then deleted. The cost of deletion for one node is equal to the insertion cost. The node distance between $G_1$ and $G_2$ is updated accordingly, as exemplified in Figure xviD where $ND=2.5+3=5.5$ since two nodes are deleted. At this stage, $G_1$ and $G_2$ have the same number of nodes located at the same positions on the grid.

**STEP 4: EDGE DISTANCE**

In step 4, the edge distance ($ED$) for $G_1$ and $G_2$ at the initial state is computed from the difference of edge weights between $G_1$ and $G_2$, using equation (10):

$$diff(W^{G_1}_{v_1,v_2},W^{G_2}_{v_2,v_2}) = |W^{G_1}_{v_1,v_2} - W^{G_2}_{v_2,v_2}|$$  \(10\)

In a binary graph, the $diff(W^{G_1}_{v_1,v_2},W^{G_2}_{v_2,v_2})$ score $\in [0,1]$ where 1 means that the edge linking $v_1$ and $v_2$ does exist in $G_1$ but does not exist in $G_2$ for the corresponding nodes $v_2$ and $v_2$. Conversely, a difference equal to 0 means that this edge exists in both graphs. The total $ED$ is then calculated using equation (11):

$$ED = \sum_{k=1}^{Tn-1} \sum_{t=k+1}^{Tn} diff(W^{G_1}_{v_1,v_2},W^{G_2}_{v_2,v_2})$$  \(11\)

Where $Tn$ is the total number of nodes in the grid. Finally, the distance between $G_1$ and $G_2$ is calculated by summing up the node distance ($ND$ from step 3) and the edge distance ($ED$ from
step 4)

\[ d(G_1, G_2) = ND + ED \]  \hspace{1cm} (12)

In the example of Figure xvi, the edge distance \((ED)\) is equal to 9. The distance \(d\) between \(G_1\) and \(G_2\) is calculated by the equation (12): \(d(G_1, G_2) = ND + ED = 5.5 + 9 = 14.5\).

Finally, the distance \(d(G_1, G_2)\) is scaled to a similarity index \(\text{sim}_{\text{SimNet}}(G_1, G_2)\) via the formula \(\text{sim}_{\text{SimNet}} = (1 / (1 + d))\). As depicted, the proposed similarity index \(\in [0, 1]\) where 0 means that \(G_1\) and \(G_2\) are totally dissimilar, while 1 means that \(G_1\) and \(G_2\) are identical.
Figure xvi. Illustration of SimNet algorithm steps. A) Finding the similarity index between $G_1$ and $G_2$. B) Step1-i: detection of nodes common to $G_1$ and $G_2$. B) Step1-ii: shifting the nearest neighbors located in the defined spatial neighborhood (disk with radius $R = 1.5$), shifting cost equals 1. C) Step 2: insertion of new nodes when no neighbor is found within the neighborhood (cost of insertion = $R = 1.5$). D) Step 3: deletion of remaining nodes of $G_1$ (deletion cost = insertion cost = 1.5). E) Step 4: computing the edge distance between $G_1$ and $G_2$. F) Final state: $G_1$ matches $G_2$ and $d(G_1, G_2) = ND + ED = 14.5$. 

ND=0

ND=1 + 1.5

ND=2.5 + 1.5 + 1.5

ND=5.5

ED=9

d=14.5
CHAPTER 4- RESULTS

In this chapter, we present a brief resume of each study followed by the complete version of the corresponding published or submitted articles.

STUDY 1: A NEW ALGORITHM FOR SPATIOTEMPORAL ANALYSIS OF BRAIN FUNCTIONAL CONNECTIVITY

A Mheich, M Hassan, M Khalil, C Berrou, F Wendling

Published in: Journal of neuroscience methods (2015)

Summary

Objectives: Specific networks of interacting neuronal assemblies distributed within and across distinct brain regions underlie brain functions. In most cognitive tasks, these interactions are dynamic and take place at the millisecond time scale. A challenge of cognitive neuroscience is to temporally follow, over very short time duration, these interactions. Among neuroimaging techniques, magneto/electroencephalography – M/EEG – allows for detection of very short-duration events and offers the single opportunity to follow, in time, the dynamic properties of cognitive processes (sub-millisecond temporal resolution).

Methods: In this paper, we propose a new algorithm to track the functional brain connectivity dynamics. During a picture naming task, this algorithm aims at segmenting high-resolution EEG signals (hr-EEG) into functional connectivity microstates. The proposed algorithm is based on the K-means clustering of the connectivity graphs obtained from the phase locking value (PLV) method applied on hr-EEG.

Results: Results show that the analyzed evoked responses can be divided into six clusters representing distinct networks sequentially involved during the cognitive task, from the picture presentation and recognition to the motor response.

Key words: Dynamics of cognitive brain network; EEG connectivity; K-means clustering
Computational Neuroscience
Short communication

A new algorithm for spatiotemporal analysis of brain functional connectivity

A. Mheich a, b, c, M. Hassan a, c, *, M. Khalil b, C. Berrou d, F. Wendling a, c

a INSERM, U1099, Rennes F-35000, France
b AZM Center-EDST, Lebanese University, Tripoli, Lebanon
c Université de Rennes 1, LTSI, F-35000, France
d Télécom Bretagne (Institut Mines-Télécom), UMR CNRS Lab-STICC, Brest, France

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 3 December 2014
Accepted 3 January 2015
Available online 10 January 2015

Keywords:
Dynamics of cognitive brain network
EEG connectivity
K-means clustering

ABSTRACT

Specific networks of interacting neuronal assemblies distributed within and across distinct brain regions underlie brain functions. In most cognitive tasks, these interactions are dynamic and take place at the millisecond time scale. Among neuroimaging techniques, magneto/electroencephalography – M/EEG – allows for detection of very short-duration events and offers the single opportunity to follow, in time, the dynamic properties of cognitive processes (sub-millisecond temporal resolution).

In this paper, we propose a new algorithm to track the functional brain connectivity dynamics. During a picture naming task, this algorithm aims at segmenting high-resolution EEG signals (hr-EEG) into functional connectivity microstates. The proposed algorithm is based on the K-means clustering of the connectivity graphs obtained from the phase locking value (PLV) method applied on hr-EEG. Results show that the analyzed evoked responses can be divided into six clusters representing distinct networks sequentially involved during the cognitive task, from the picture presentation and recognition to the motor response.

1. Introduction

There is increasing evidence that cognitive functions arise from the activation of networks distributed over distinct and possibly distant brain regions as opposed to isolated focal areas (Sporns, 2010). Hence, efforts focused on the analysis of brain connectivity as a key concept to understand brain cognitive functions. Due to its excellent spatial resolution, fMRI has become one of the most commonly used noninvasive methods to study cerebral functions (Allen et al., 2012).

However, in many cases, the short duration of most cognitive processes (∼500 ms for picture naming, for example) would greatly benefit from the use of techniques that have a much higher time resolution (on the order of ms), which is not the case of fMRI (∼1 s). Along this line, several studies indicated that the use of electroencephalography (EEG, 1 ms time resolution for signals classically sampled at 1 kHz) combined with appropriate

signal processing techniques can bring relevant information about normal networks during cognitive activity (Rodriguez et al., 1999) or about altered networks associated with tumors (Bartolomei et al., 2006) for instance.

This excellent temporal resolution of the EEG signals allowed us to analyze the dynamic properties of cognitive processes, an issue so far addressed in a few studies only. In Murray et al. (2008), authors proposed an algorithm based on the amplitude of event related potentials (ERPs) to follow time-varying voltage topographic maps. However, these algorithms do not account for brain connectivity quantified directly from scalp signals (electrode space) or indirectly from reconstructed brain sources (source space).

Regarding the approaches based on the connectivity analysis, most of reported methods make use of a constant time window to track the dynamics of functional connectivity, as estimated from EEG recordings. This window is typically chosen either empirically or based on a priori information about the analyzed task (Rodriguez et al., 1999). A few attempts have been recently reported to avoid this constraint (De Vico Fallani et al., 2008; Dimitriades et al., 2010; Allen et al., 2012). However, most of proposed algorithms are not adapted to tracking changes over very short durations (in
the order of 500 ms, as in the case of responses evoked by visual stimuli).

In this paper, we propose a novel algorithm to track the dynamics of brain functional connectivity at millisecond scale. The proposed algorithm is based on the K-means clustering of the connectivity networks obtained by the phase locking value (PLV) method. Performance evaluation was assessed on high-resolution electroencephalographic (hr-EEG) signals recorded in subjects during a picture naming task.

2. Materials and methods

2.1. Functional connectivity measure

Functional connectivity is classically defined as the temporal correlation (wide sense) among electrophysiological signals generated by distinct neuronal assemblies (Friston, 1994). Several methods have been proposed to quantify brain functional connectivity. In this study, we used a method, which belongs to the so-called “phase synchronization” (PS) family.

It is well known that the respective phases of two oscillators may synchronize even if their amplitudes stay uncorrelated. The general principle of PS methods is to detect the existence of a phase locking between two systems defined as:

\[ \varphi_{xy}(t) = |\Phi_x(t) - \Phi_y(t)| \leq C \]

where \( \Phi_x(t) \) and \( \Phi_y(t) \) are the unwrapped phases of the signals \( x \) and \( y \) representative of the two systems at time \( t \) and \( C \) a constant. The first step for estimating the phase synchronization is to extract the instantaneous phase of each signal. In this study, we used the method based on Hilbert transform. The second step is the definition of an appropriate index to measure the degree of synchronization between estimated instantaneous phases. To proceed, we used the phase locking value (PLV) (Lachaux et al., 1999), as illustrated in Fig. 1B. For each channel pair, \( x \) and \( y \), at time \( t \) \( (t = t_1, \ldots , t_N) \) where \( T = D \cdot f_s \) \( (D \) and \( f_s \) denote the signal length relative to the onset and the sampling frequency, respectively) for the \( N \) trials and for subject \( j (j = 1, \ldots , M) \), where \( M \) denotes the number of subjects), PLV is defined as:

\[ \text{PLV}_{xy}(t) = \frac{1}{N} \sum_{i=1}^{N} \varphi_x(t) - \varphi_y(t) \]  \hspace{1cm} (1)

To reduce the effect of correlations between near electrodes, we apply a normalization procedure (z-score) so that the PLV values were compared with the 200 ms baseline preceding the presentation of the image. Let \( \mu_{xy} \) and \( \sigma_{xy} \) be the mean and standard deviation computed from a 200 ms pre-stimulus baseline. The normalized PLVs are then defined as \( \text{PLV}_{xy}^{\prime}(t) = (\text{PLV}_{xy}(t) - \mu_{xy})/\sigma_{xy} \). A thresholding procedure is then applied on the functional connectivity values in order to retain the strongest functional connections. The connectivity measure was computed in the low gamma frequency band (30–45 Hz). More precisely, the phases were estimated for each frequency and the average phase at 30–45 Hz was used. Indeed, this frequency band was shown to be highly relevant in the context of the cognitive task performed by subjects, as reported in Rodriguez et al. (1999).

The PLVs were then averaged over subjects:

\[ \text{PLV}_{xy}(t) = \frac{1}{M} \sum_{j=1}^{M} \text{PLV}_{xy}^{\prime}(t) \]  \hspace{1cm} (2)

where \( \text{PLV}_{xy}(t) \) represents the general term of the average adjacency matrix \( \text{PLV}(t) \) which defines a functional connectivity graph \( G \) at each time \( t \), \( G = \{ G(t), t = 1, \ldots , T \} \), computed for the \( V \) pairs of \( x \) and \( y \) channels, where \( V \) is equal to \( (N_c \cdot (N_c - 1)/2) \) and \( N_c \) is the number of channels in the hr-EEG montage. According to Eqs. (1) and (2), adjacency matrices are obtained.

2.2. Segmentation algorithm

The objective of this algorithm is to identify clusters among the \( T \) graphs \( G(t) \). As illustrated in Fig. 1C, the proposed algorithm is based on three main steps:

**Step 1 (Initialization).** To start with, \( K \) graphs \( G_k \), \( G_k = \{ \tilde{C}_k \} \), \( k = t_1, \ldots , t_k \), are selected where \( k = t_i \) and \( i \) is randomly chosen in \( K \) if \( t \) varies from 3 to 12 and \( k \) varies from 1 to \( K \) with the restriction of rejecting the \( K \) graphs if the time interval between two \( t_i \) is less than 30 ms.

**Step 2 (Assignment).** The spatial correlation \( sC_k(t) \) between \( G(t) \) and \( G_k \) is then computed as follows:

\[ sC_k(t) = \frac{\sum_{i=1}^{V} \tilde{C}_k G_i(t) - \bar{C}_k \text{VAR}_k}{\sqrt{\sum_{i=1}^{V} \tilde{C}_k^2 \cdot \sqrt{\sum_{i=1}^{V} C_i^2(t)}}} \]  \hspace{1cm} (3)

where \( i \) denotes the \( i \)th edge in \( G(t) \) and \( \tilde{C}_k \). As depicted in Eq. (3), \( SC \) is normalized by the variance of graphs \( G \) and \( \tilde{C}_k \). Thus, \( SC \) ranges from 0 to 1 high values denote graph with high similarity. Conversely, low values are indicative of low similarity between graphs.

Each graph \( G(t) \) is then assigned to the cluster for which the spatial correlation was the highest. The assigned clusters are defined as \( \hat{C}_k \):

\[ \hat{C}_k = |G(t) : SC_k G(t)_k \geq SC_k G(t)_k \forall 1 \leq k' \leq K | \]  \hspace{1cm} (4)

From these spatial correlation values, the global explained variance (GEV) is calculated as defined in Murray et al. (2008):

\[ \text{GEV} = \sum_{k=1}^{K} \text{GEV}^k \]  \hspace{1cm} (5)

\[ \text{GEV}^k = \sum_{t=1}^{T} (SC_{G(t)_k})^2 \text{Y}_{G(t)_k} \quad \text{where} \quad \text{Y}_{G(t)_k} = \begin{cases} 1 & \text{if } G(t) \in \hat{C}^k \\ 0 & \text{if } G(t) \notin \hat{C}^k \end{cases} \]  \hspace{1cm} (6)

**Step 3 (Update).** At each iteration, the new centroids \( \tilde{C}_k \) are updated by averaging all the graphs yielding to the same cluster

\[ \tilde{C}_k = \frac{1}{|G(t)|} \sum_{G \in \hat{C}_k} G \]  \hspace{1cm} (7)

For each \( K \), Steps 2 and 3 were repeated 500 times. The set of centroids leading to the highest GEV was retained. When the algorithm converges (reaching the highest GEV), \( K + 1 \) graphs \( G \) are then selected randomly and the entire above procedure (from Step 2 to Step 3) is repeated until \( K = 12 \).

To choose the optimal number of clusters, we used a method based on the cross validation (CV) criterion (Murray et al., 2008) which is a ratio between the GEV and the degrees of freedom for a given set of graphs. As reported, the global minimum of this criterion gives the optimal number of segments. Note that in the same segment, the graphs can have different SC values with the same cluster and therefore two consecutive graphs (in time) can be classified in two distinct clusters. To overcome this, the decision
Fig. 1. Structure of the investigation. (A) Dense EEGs with 256 electrodes were used to record brain electrical activity. Participants were shown about 74 stimuli from different categories. (B) Functional connectivity was computed between electrodes over all trials using the phase locking value method; the resultant connectivity matrix at a given instant (300 ms for instance) was derived. This connectivity matrix was then projected on 2D scalp to provide a connectivity graph. (C) Illustration of flowchart of the proposed segmentation algorithm: three main steps were realized: the initialization, the assignment and the update steps.

2.3. Data

Six subjects were shown pictures on a screen using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). They were requested to name the displayed pictures. The 148 images were selected from a database of 400 pictures standardized for French (Alario and Ferrand, 1999) and were used during two sessions (about eight minutes each) of 74 stimuli. Brain activity was recorded using hr-EEG system (EGI, Electrical Geodesic Inc.). EEG signals were collected with a 1 kHz sampling frequency and band-pass filtered between 3 and 45 Hz. Each trial was visually inspected, and epochs contaminated by eye blinking, movements or any other noise source were rejected and excluded from the analysis. This study was approved by the National Ethics Committee for the Protection of Persons (CPP, connexions study,
agreement number 2012-A01227-36, promoter: Rennes University Hospital). We excluded the electrodes located on the face as well as the few electrodes showing too high impedance. All participants provided a written informed consent to participate in this study.

3. Results

The algorithm described above was applied on the adjacency matrix averaged over 6 subjects who performed the picture recognition and naming task. The segmentation algorithm, applied to the recorded EEG signals, led to a number of clusters equal to 6. Results are shown in Fig. 2A. The first cluster corresponded to the period ranging from \( t = 0 \) (stimulus onset: presentation of the picture) to \( t = 116 \) ms. In this cluster, a network located in the occipital lobe could be observed. A shorter graph was then observed between \( t = 117 \) ms and \( t = 153 \) ms with strong connections at the occipital lobe as well. The 3rd cluster corresponded to the time interval 154–190 ms where connections appeared mainly in the temporal and parietal lobes. A network was then identified in the cluster 191–316 ms with the presence of connections mainly located in the frontal lobe, followed by a cluster (317–480 ms) with two separate networks in frontal and occipital lobes. The last cluster (481–620 ms) was then associated with a denser network in left temporal and bilateral frontal lobes with predominance of the right frontal lobe.

Fig. 2B shows the results obtained regarding the inter-subject variability. The graph existence percentages were computed for all the subjects. The figure shows that graphs (T3, T6) have the highest graph existence (73% and 78% respectively). In contrast, the lowest values (30.99%, 22.33%) were obtained for graph T1 and T4 respectively. The graphs T2 and T5 showed nearly similar graph existence with 41.17% and 38.33% respectively. Results also showed that the

Fig. 2. Results of the segmentation process. (A) Event related potentials for the picture naming task and distribution of the functional connectivity graphs revealed by the proposed spatiotemporal segmentation algorithm. The results are shown from 0 ms (the Onset when the visual stimuli were presented) to 620 ms when the naming process starts. (B) Graph existence is computed over the six subjects for the six identified networks.
first (T1) and the fifth (T5) graphs have the lowest variability across subjects (±16% and ±22% respectively) while the second graph (T2) has the highest variability (±38%).

4. Discussion and conclusion

A challenge of cognitive neuroscience is to temporally follow, over very short time duration, the sequence of brain processes involved in a considered task.

In this communication, a novel algorithm was presented to track the dynamics of functional connectivity. The proposed algorithm is based on the clustering of connectivity matrices computed from the phase locking value method applied on scalp dense EEG signals. It is exemplified in the well-known picture naming task which duration is about 500 ms. By taking advantage of the excellent temporal resolution of the EEG signals and the use of PLV, the main originality of the proposed algorithm is its capacity to track connectivity dynamics over very short time periods. The proposed algorithm showed high performance to reveal the stability of the brain networks over short periods of time and to segment the cognitive process into functional connectivity microstates.

Another algorithm was recently proposed by Allen et al. (2012) and applied to track brain networks during resting state. The basic principle of this algorithm was based on the independent component analysis and the K-means. However the proposed algorithm was adapted to fMRI data by taking a large time window (about 3 s) to compute the functional connectivity and therefore does not allow tracking cognitive tasks with short time duration. Therefore, it is likely that accurate tracking of the picture naming could not be achieved with the above mentioned proposed algorithm.

Very interestingly, results were qualitatively consistent in term of involved brain regions, with already-reported results regarding the analyzed task, most of them being however based on other neuroimaging modalities (mainly functional magnetic resonance – fMRI – and positron emission tomography – PET). Networks were identified in the occipital lobe during the first 200 ms (including the P150) which likely correspond to the visual processing of the presented picture during the recognition phase (Schendan et al., 1998). Another network is then identified in the following 200 ms where a left occipito-parieto-temporal network was identified which is compatible with brain regions involved in the access to semantics content of the presented picture (Salmelin et al., 1994). Finally, at the last 200 ms, dense networks appear in the temporofrontal regions. It is well established that this network is strongly involved in decision making during the naming phase (Salmelin et al., 1994).

Regarding the EEG connectivity studies, our results seem to agree with studies reported the dynamic (time-varying) behavior of functional connectivity graphs during visual tasks (Babiloni et al., 2005; De Vico Fallani et al., 2008). These studies and others showed how the brain networks properties modify during visual tasks with relatively short duration.

The proposed method was applied to networks obtained at the level of scalp EEG electrodes. These networks suffer from the low spatial resolution and the effect of the field spread. Our further objective is to apply this algorithm on EEG source connectivity graphs to identify the dynamics of brain networks at cortical source level as described in Hassan et al. (2014). The algorithm will be used also to complement the brain, timeliney, timeliness, complexity at different conditions such as the difference between networks related to different types of stimulus (animals vs. tools for instance).

More generally, the proposed algorithm can be broadly applicable to disciplines where network dynamics is crucial in the understanding of system performance.

Acknowledgements

This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the “Investing for the Future” program under reference ANR-10-LABX-07-01. It was also supported by the Rennes University Hospital (CORE Project named connexeION, 2012–14). We also thank the European Research Council for the ERC-2011-ADG - Grant Agreement N° 290901 – Acronym “NEUCOD”.

References

Ahmad Mheich, Mahmoud Hassan, Mohamad Khalil, Vincent Gripon, Olivier Dufor and Fabrice Wendling

In revision: IEEE Transactions on Pattern Analysis and Machine Intelligence

Summary

Objectives: Quantifying the similarity between two networks is critical in many applications. A number of algorithms have been proposed to compute graph similarity, mainly based on the properties of nodes and edges. Interestingly, most of these algorithms ignore the physical location of the nodes, which is a key factor in the context of brain networks involving spatially defined functional areas.

Methods: In this paper, we present a novel algorithm called “SimNet” for measuring similarity between two graphs whose nodes are defined a priori within a 3D coordinate system. SimNet provides a quantified index (ranging from 0 to 1) that accounts for node, edge and spatio-temporality features. Complex graphs were simulated to evaluate the performance of SimNet that is compared with state-of-art methods.

Results: Results show that SimNet is able to detect weak spatial variations in compared graphs in addition to computing similarity using both nodes and edges. SimNet was also applied to real brain networks obtained during a visual recognition task. The algorithm shows high performance to detect spatial variation of brain networks obtained during a naming task of two categories of visual stimuli: animals and tools. A perspective to this work is a better understanding of object categorization in the human brain.

Key words: Graphs and Networks, Neuroscience, Similarity measures
SimNet: a Novel Method for Quantifying Brain Network Similarity

Ahmad Mheich, Mahmoud Hassan, Mohamad Khalil, Vincent Gripon, Olivier Dufor and Fabrice Wendling

Abstract— Quantifying the similarity between two networks is critical in many applications. A number of algorithms have been proposed to compute graph similarity, mainly based on the properties of nodes and edges. Interestingly, most of these algorithms ignore the physical location of the nodes, which is a key factor in the context of brain networks involving spatially defined functional areas. In this paper, we present a novel algorithm called “SimNet” for measuring similarity between two graphs whose nodes are defined a priori within a 3D coordinate system. SimNet provides a quantified index (ranging from 0 to 1) that accounts for node, edge and spatiality features. Complex graphs were simulated to evaluate the performance of SimNet that is compared with state-of-art methods. Results show that SimNet is able to detect weak spatial variations in compared graphs in addition to computing similarity using both nodes and edges. SimNet was also applied to real brain networks obtained during a visual recognition task. The algorithm shows high performance to detect spatial variation of brain networks obtained during a naming task of two categories of visual stimuli: animals and tools. A perspective to this work is a better understanding of object categorization in the human brain.

Index Terms—Graphs and Networks, Neuroscience, Similarity measures

I. INTRODUCTION

The brain is a large-scale network in which distant interconnected neural assemblies continuously synchronize and desynchronize to process information. Over the past decade, considerable progress has been achieved in neuroimaging techniques that are now able to identify brain networks from structural and/or functional data. In most cases, these networks are represented as graphs in which the nodes denote brain areas and the edges describe the connectivity among these areas [1]. This graph representation allows for application of graph theory algorithms in order to assess statistical and/or topological properties of networks reconstructed from data. From a theoretical viewpoint, the application of these algorithms on functional, as well as on structural connectivity matrices, has revealed many properties about brain networks, such as small-worldliness [2, 3], modularity [4, 5], hubs [6], and rich-club configurations [7]. From an application viewpoint, graph theory based analysis has been widely used to characterize normal [8] and pathological [9] brain activities from several modalities (fMRI, EEG, MEG). In particular, it has led to the identification of network alterations in aging [10], Alzheimer’s disease [11], schizophrenia [12] and autism [13].

In contrast with the large number of methods aiming at characterizing graph properties [14], less attention has been paid to methods able to quantitatively compare graphs and extract similarities while taking into account the spatial location of the graph nodes. Reported methods make use of graphs and subgraphs isomorphism [15], graph edit distance [16] and Levenshtein distance [17] in order to measure graph similarity. However, in the context of brain networks, the spatial location of nodes is a key factor for the graph comparison [18]. The intuitive idea is that two networks with strictly identical statistical properties but interconnecting different brain areas should be considered to have low similarity. Conversely, two graphs with partly dissimilar properties but interconnecting the same brain regions should be considered as exhibiting high similarity.

In this paper, we propose a new algorithm to solve this issue. This algorithm is able to measure the similarity between two graphs based on the node and edge properties but under a spatial constraint related to the physical location of nodes. To our knowledge, this approach has never been followed in the problem of measuring a formal distance between brain networks. The performance of the proposed algorithm was evaluated using simulated graphs as well as real brain networks estimated from dense-electroencephalographic (EEG) signals during a picture naming task involving two categories of stimuli: tools and animals. The paper is organized as follows. Section II.A provides the notations and definitions. The problem statement is then introduced in
section II.B followed by the description, in section II.C, of the proposed method for measuring graph similarity. A comparative study is then achieved with respect to already-published methods briefly presented in section II.D. Results obtained on simulated networks as well as on real networks are given in section III. Finally, these results are discussed according to the performance and potential applications of the proposed algorithm in the context of brain connectivity research.

II. MATERIALS AND METHODS

A. Notations and definitions

A graph $G$ is denoted by $G(V,E)$ where $V$ denotes the set of nodes (with known Cartesian coordinates) and $|V|$ is the order of the graph (number of nodes). $E \subseteq V \times V$ defines the edges and $|E|$ is the number of edges in $G$. We denote by $W_{v_i,v_j}^G$, the weight of the edge between nodes $v_i$ and $v_j$. The graph is said to be simple if there is no edge linking a node with itself, and it is said to be undirected if the adjacency matrix is symmetric. We denote by $\text{sim}(G_1,G_2)$ the similarity measure between two graphs $G_1$ and $G_2$. All graphs considered in this work are assumed to be simple, undirected, weighted graphs. Defined quantities and notations are listed in Table 1.

### Table 1

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Gr,T$</td>
<td>Grid, total number of nodes</td>
</tr>
<tr>
<td>$x_v, y_v$</td>
<td>Abscissa and ordinate for a node $v$</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Spatial neighborhood</td>
</tr>
<tr>
<td>$G$</td>
<td>Graph</td>
</tr>
<tr>
<td>$V, n$</td>
<td>Set of nodes, number of nodes ($n \leq T$)</td>
</tr>
<tr>
<td>$E, m$</td>
<td>Set of edges, number of edges</td>
</tr>
<tr>
<td>$K_v$</td>
<td>Index of node $v$</td>
</tr>
<tr>
<td>$v^K_v$</td>
<td>Node of graph $G_v$ with index $K_v$</td>
</tr>
<tr>
<td>$\text{sim}(G_1,G_2)$</td>
<td>Similarity index between $G_1$ and $G_2$</td>
</tr>
<tr>
<td>$d(G_1,G_2)$</td>
<td>Distance between $G_1$ and $G_2$</td>
</tr>
<tr>
<td>$R$</td>
<td>Radius of $\Omega$ (when defined as a disk)</td>
</tr>
<tr>
<td>$A_G$</td>
<td>Adjacency matrix of graph $G$. $A_G[T,T]$</td>
</tr>
<tr>
<td>$ED$</td>
<td>Edge distance</td>
</tr>
<tr>
<td>$ND$</td>
<td>Node distance (based on insertion, deletion, substitution costs)</td>
</tr>
<tr>
<td>$W_{v_i,v_j}^G$</td>
<td>Weight of the edge between the nodes $v_i$ and $v_j$ in $G$</td>
</tr>
<tr>
<td>$\text{diff}(W_{v_i,v_j}^G, W_{v_i',v_j'}^G)$</td>
<td>Difference of edges weight between nodes $v_i', v_j'$ in $G_1$ and $v_i, v_j$ in $G_2$</td>
</tr>
</tbody>
</table>

B. Problem statement

Let us define two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$. The nodes of $G_1$ and $G_2$ are distributed onto a square grid of side length $T$. Cells in the grid are indexed from 1 to $T^2$ line by line starting from the first line. We suppose that each grid cell contains at most one node with Cartesian coordinates $(x, y)$. Consequently, we index nodes using the corresponding grid cell index: $K_v = x_v + ((y_v - 1) \times T)$ where $x_v \in [1,T]$ and $y_v \in [1,T]$. The problem is two-fold: i) to elaborate a distance $d(G_1,G_2)$ that accounts for the positions of nodes respectively in $G_1$ and $G_2$ as well as for statistical properties (number of nodes, number and weights of edges, density). ii) To devise a normalized index $\text{sim}(G_1,G_2) \in [0,1]$ that globally reflects the similarity of compared graphs $G_1$ and $G_2$, the similarity index to be elaborated should take values in $[0,1]$. Lowest values denote highly different graphs whereas values close to 1 denote almost identical graphs, in terms of statistical and spatial properties (position of nodes).

C. Proposed method

In order to solve the above stated problem, we propose a new algorithm called SimNet (described in Fig.1) which calculates a distance between $G_1$ and $G_2$ (Fig. 1A). This distance is based on i) the cost of a sequence of cost-based changes (substitution, insertion, deletion) on nodes that are necessary to map $G_1$ onto $G_2$ and ii) the computation of the difference of edge weights between $G_1$ and $G_2$. The proposed algorithm includes four main steps, as illustrated in Fig. 1B-E:

**Step 1: Node substitution**

To start with, the nodes common to $G_1$ and $G_2$ that have the same spatial location are detected (these nodes have the same index value $K_v$ in $G_1$ and $G_2$). In the example of Fig. 1B-i, we observe three nodes of $G_1$ having the same spatial position as three nodes of $G_2$. For the remaining nodes, we define a spatial neighborhood $\Omega$ (disk with radius $R=1.5$ as depicted in Fig. 1B-ii). The idea is to shift a neighbor node $v_i$ included in $V_1$ to the spatial position of $v_i$ included in $V_2$ if $v_i$ is located in $\Omega$ (the defined spatial neighborhood of $v_i$). In the case where there are $n>1$ nodes of graph $G_1$ located in $\Omega$, the nearest node to $v_2$ is shifted. The cost of shifting is equal to the Euclidian distance between the two shifted nodes ($v_i$ and $v_j$). The node distance ($ND$) is updated by the Euclidian distance between $v_i$ and $v_j$ (e.g. $ND = 1$ in Fig. 1B-ii).
Fig. 1. Illustration of SimNet algorithm steps. A) Finding the similarity index between $G_1$ and $G_2$. B) Step1-i: detection of nodes common to $G_1$ and $G_2$. B) Step1-ii: shifting the nearest neighbors located in the defined spatial neighborhood (disk with radius $R = 1.5$), shifting cost equals 1. C) Step 2: insertion of new nodes when no neighbor is found within the neighborhood (cost of insertion = $R = 1.5$). D) Step 3: deletion of remaining nodes of $G_1$ (deletion cost = insertion cost = 1.5). E) Step 4: computing the edge distance between $G_1$ and $G_2$. F) Final state: $G_1$ matches $G_2$ and $d(G_1, G_2) = ND + ED = 14.5$. 

- A- Initial state
- B- Step 1: node substitution
  - B-i: Common nodes
  - ND=0
  - B-ii: Shifting
  - ND=1
- C- Step 2: node insertion
  - ND=1 + 1.5
- D- Step 3: node deletion
  - ND=2.5 + 1.5 + 1.5
  - ND=5.5
- E- Step4: edge distance
  - ED=9
- F- Final state: $G_1 = G_2$
Step 2: Node Insertion
In step 2, the insertion operation is used to add a new node in \( G_1 \) at the same position of node \( v_2 \) in the case where no node exists in the defined spatial neighborhood \( \Omega \) for each node \( v_2 \) of \( V_2 \). The cost of insertion must be higher than the cost of substitution. In the example of Fig. 1C, the cost of insertion was chosen to disk with radius \( R = 1.5 \). In this case, the node distance \( (ND) \) is updated to 1 + 1.5 = 2.5

Step 3: Node deletion
At the end of step 2, all the nodes that are included in \( G_2 \) are also included in \( G_1 \) (Fig. 1D). In step 3, the remaining nodes of \( G_1 \) are then deleted. The cost of deletion for one node is equal to the insertion cost. The node distance between \( G_1 \) and \( G_2 \) is updated accordingly, as exemplified in Fig. 1D where \( ND = 2.5 + 3 = 5.5 \) since two nodes are deleted. At this stage, \( G_1 \) and \( G_2 \) have the same number of nodes located at the same positions on the grid.

Step 4: Edge distance
In step 4, the edge distance \( (ED) \) for \( G_1 \) and \( G_2 \) at the initial state is computed from the difference of edge weights between \( G_1 \) and \( G_2 \), using (1):

\[
diff(W_{v_i'v_j'}, W_{v_i'v_j'}) = \left| W_{v_i'v_j'} - W_{v_i'v_j'} \right|
\]

In a binary graph, the \( \text{diff}(W_{v_i'v_j'}, W_{v_i'v_j'}) \) score \( \in [0,1] \) where 1 means that the edge linking \( v_i' \) and \( v_j' \) does exist in \( G_1 \) but does not exist in \( G_2 \) for the corresponding nodes \( v_i' \) and \( v_j' \). Conversely, a difference equal to 0 means that this edge exists in both graphs. The total \( ED \) is then calculated using (2):

\[
ED = \sum_{v_i = 1}^{T-1} \sum_{v_j = 1}^{T} \text{diff}(W_{v_i'v_j'}, W_{v_i'v_j'})
\]

Finally, the distance between \( G_1 \) and \( G_2 \) is calculated by summing up the node distance \( (ND \) from step 3) and the edge distance \( (ED \) from step 4)

\[
d(G_1, G_2) = ND + ED
\]

In the example of Fig. 1, the edge distance \( (ED) \) is equal to 9. The distance \( d \) between \( G_1 \) and \( G_2 \) is calculated by the equation 3: \( d(G_1, G_2) = ND + ED = 5.5 + 9 = 14.5 \).

Finally, the distance \( d(G_1, G_2) \) is scaled to a similarity index \( \text{sim}_{\text{SimNet}}(G_1, G_2) \) via the formula \( \text{sim}_{\text{SimNet}} = (1 / (1 + d)) \). As depicted, the proposed similarity index \( \in [0,1] \) where 0 means that \( G_1 \) and \( G_2 \) are totally dissimilar, while 1 means that \( G_1 \) and \( G_2 \) are identical.

The pseudo-code of this algorithm is provided in Fig. 2.

D. Other reported algorithms
In section 2.5, SimNet is compared with four state-of-art methods able to measure similarity between graphs. These four methods are briefly described hereafter.

Graph edit distance (GED) [19]
This distance is based on the transformation of one graph to the other using elementary operations. The elementary operations consist in suppressions and insertions of nodes.

\[
d_{\text{GED}}(G_1, G_2) = |V_1| + |V_2| - 2 |V_1 \cap V_2| + |E_1| + |E_2| - 2 |E_1 \cap E_2|
\]

This distance is normalized between 1 and 0 via the formula \( \text{sim}_{\text{GED}}(G_1, G_2) = (1 / (1 + d_{\text{GED}}(G_1, G_2))) \)

DeltaCon method [20]
This algorithm assesses the similarity between two graphs on the same nodes. The concept of this method is to compute the pairwise node affinities in the first graph and to compare them with the ones in the second graph. Then, it measures the differences of nodes affinity scores of the two graphs and reports the similarity score. Readers may refer to [20] for details about the DeltaCon algorithm main steps and implementation. This algorithm also provides a normalized similarity index \( \text{sim}_{\text{DeltaCon}}(G_1, G_2) \) ranging from 0 (dissimilar graphs) to 1 (identical graphs).

<table>
<thead>
<tr>
<th>Algorithm SimNet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INPUT:</strong> ( A_{G_1}, A_{G_2}, R )</td>
</tr>
<tr>
<td>// Node distance</td>
</tr>
<tr>
<td>1: Initialize ( ND = 0 )</td>
</tr>
<tr>
<td>2: for each node ( v_2 \in V_2 )</td>
</tr>
<tr>
<td>3: if ( \exists v_1 \in V_1 / \text{distance}(v_1, v_2) &lt; R )</td>
</tr>
<tr>
<td>4: shifting ( v_2 \leftarrow v_1 )</td>
</tr>
<tr>
<td>5: ( ND = ND + \text{dist an ce}(v_1, v_2) )</td>
</tr>
<tr>
<td>6: else</td>
</tr>
<tr>
<td>7: // insertion a node in ( G_i )</td>
</tr>
<tr>
<td>8: ( ND = ND + R )</td>
</tr>
<tr>
<td>9: end if</td>
</tr>
<tr>
<td>10: end for each</td>
</tr>
<tr>
<td>11: delete the remained nodes in ( G_i ) and update ( ND )</td>
</tr>
<tr>
<td>// Edge distance</td>
</tr>
<tr>
<td>12: Initialize ( ED = 0 )</td>
</tr>
<tr>
<td>13: for ( k = 1 \rightarrow T - 1 )</td>
</tr>
<tr>
<td>14: for ( p = k + 1 \rightarrow T )</td>
</tr>
<tr>
<td>15: ( ED = ED + \left</td>
</tr>
<tr>
<td>16: end for</td>
</tr>
<tr>
<td>17: end for</td>
</tr>
<tr>
<td>18: ( d = ND + ED )</td>
</tr>
<tr>
<td>19: ( \text{sim} = (1 / (1 + d)) )</td>
</tr>
<tr>
<td><strong>Return:</strong> ( \text{sim} )</td>
</tr>
</tbody>
</table>

Fig. 2. Pseudo-code of the proposed SimNet algorithm
Vertex/Edge Overlap method (VEO) [21]
The principle of this method is that two graphs are similar if they share many vertices (i.e. nodes) and edges. Thus, the similarity between two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ is defined as:

$$sim_{VEO}(G_1, G_2) = \frac{|V_1 \cap V_2| + |E_1 \cap E_2|}{|V_1| + |V_2| + |E_1| + |E_2|}$$

This measure of similarity is computed by scanning all nodes of $G_1$ and by checking if each occurs in $V_2$, the set of nodes of $G_2$.

$\lambda$-distance method (Lambda distance) [22]
Let $(\lambda_1)^{V_1}_{i=1}$ and $(\lambda_2)^{V_2}_{j=1}$ be the eigenvalues of two adjacency matrices respectively associated with two graphs $G_1$ and $G_2$. The $\lambda$-distance is given by:

$$d_\lambda(G_1, G_2) = \sqrt{\sum_{i=1}^{L} (\lambda_{1i} - \lambda_{2i})^2}$$

where $L$ is max$(|V_1|, |V_2|)$. For the sake of comparison, this distance is also normalized between 1 and 0 via the formula $sim_\lambda(G_1, G_2) = 1 / (1 + d_\lambda(G_1, G_2))$

E. Comparative analysis
In order to compare the proposed algorithm with the state-of-the-art methods, we analyzed the performance of the five methods using graphs subject to three types of alterations, namely changes in the edge weights, insertion of nodes and shifts in their spatial location. In practice, a random graph $G_1$ with 20 nodes located onto a grid $Gr(20 \times 20)$ was generated. This graph was altered to get a graph $G_2$ and the five similarity indexes ($sim_{x\_noun}(G_1, G_2)$, $sim_{x\_verb}(G_1, G_2)$, $sim_{x\_object}(G_1, G_2)$, $sim_{x\_goal}(G_1, G_2)$, $sim_{x\_end}(G_1, G_2)$) and $sim_{x}(G_1, G_2)$ between the initial graph $G_1$ and the altered version $G_2$ of $G_1$ were computed for various levels of alteration. Regarding the edge weight, a uniform random number ([0, 50]) was added to the initial weight of $G_1$ edges. For node insertion, the alteration level was defined as the number of nodes added to $G_1$. Finally, for the spatial location, the alteration level was defined as a random shift of each node of the altered graph $G_2$ to one of the possible surrounding positions, either close to (low alteration level) or farther from (high alteration level) the corresponding node in the initial graph $G_1$.

These steps were applied 1000 times for each type of alteration and results were averaged for each method.

F. Real data
In order to evaluate the performance of SimNet on real data, we used brain networks identified during a visual task. Dense electroencephalographic (EEG) data were recorded when subjects named pictures presented on a screen. Pictures were selected from the Snodgrass database [23]. Two categories of visual stimuli were shown: tools and animals. The number of pictures for each category was chosen equal (n=37) and several psycholinguistic parameters were controlled to get equivalent datasets (name agreement, image agreement, age of acquisition as well as linguistic parameters like oral frequency, written frequency, letters/phonemes/syllables and morphemes numbers, see supplementary materials Fig. S1 and Table S1). The recorded signals were then processed using an EEG source connectivity method recently developed to identify cortical brain networks from scalp EEG data [24, 25]. Following this method, brain networks reconstructed from EEG signals (filtered in the gamma frequency band) and corresponding to each category were obtained for twenty one participants (11 women: mean age 28 year; min: 19, max: 40 and 10 men: age 23 years; min: 19, max: 33).

For the purpose of this work, we analyzed the networks identified at two different time windows corresponding to two distinct steps of the cognitive process: visual processing (1-119 ms) and access to memory (151-190 ms). These windows were obtained using a clustering algorithm allowing for segmentation of the cognitive process (from picture display to naming) [26]. During the first window, (1-119 ms), we expect that the similarity index between brain networks of the two categories of pictures will be high for two reasons, at least. First, it has been shown with fMRI that the primary visual cortex (along with the sensory motor cortex) is certainly the less variable region when functional connectivity is measured between individuals [27]. The authors demonstrated that variability of functional connectivity increases from unimodal cortices to multimodal association cortices and correlates positively with the proportion of long range functional links. These results were shown both in their own analysis and in a meta-analysis they ran while grouping 15 studies measuring functional connectivity with fMRI.

The second reason why variability should increase with time across different periods of the cognitive process is linked to the interference of semantic judgement on brain processing. During the very first steps of visual processing of picture, semantic has little or nothing to do with the brain operations aiming at reconstructing the image. Moreover, the picture naming task does not require the participants to make a choice from the stimulus (like in a go/no-go task) or to categorize animals or faces which could have led to strong semantic interference around 70 to 80 ms post onset [28]. In consequence, this should result in stronger similarity indices for the first period (visual processing) as compared with the second one (access to memory). In addition, we expect a lower similarity index between networks during the second window during which the participants are able to consciously manipulate the concepts shared by the pictures (see the discussion for more detail).

The five similarity indexes were calculated between the networks obtained over the two time windows when tools or animals where named by subjects. We evaluated whether SimNet (compared to other algorithms) was able to discriminate the 2 categories. This study was approved by the Committee for the Protection of Persons (CPP) and a local Ethics committee, (conneXion study, agreement number 2012-A01227-36, promoter: Rennes University Hospital).

In this application graph $G$ is defined as a set nodes $V$ representing the brain regions segmented from a Destrieux Atlas [29] and the edges $E$ represent the functional connectivity between regional time series. In the case of brain
networks, the geodesic distance was preferred to the Euclidean distance (used in 2D simulations) as it is more suited to the folded brain surface (presence of gyri and sulci). The “Fast-Marching-Toolbox” [30] was used to compute the geodesic distance.

III. RESULTS

A. Simulated data

Results obtained from the comparison of the different algorithms (GED, VEO, DeltaCon, Lambda-distance and SimNet) on simulated graphs are shown in Fig. 3. First, regarding the evolution of the similarity indexes \( \text{sim}_{\text{node}}(G_1, G_2), \text{sim}_{\text{GED}}(G_1, G_2), \text{sim}_{\text{DeltaCon}}(G_1, G_2), \text{sim}_{\text{VEO}}(G_1, G_2) \) and \( \text{sim}_{\text{l}}(G_1, G_2) \) with respect to gradually-increasing alterations of the edge weights (Fig. 3A), results confirmed that algorithms GED and VEO do not show any change with regard to alterations of the edge weights (similarity=1 for both methods at all alteration levels). This result was expected as both algorithms are based on quantifying common nodes and edges between graphs and they do not take into account the edge weights to measure the similarity. In contrast, the curves obtained for DeltaCon and Lambda-distance decreased dramatically for increasing level of alteration of the edge weight (DeltaCon: from 1.0 to 0.27 ± 0.07, Lambda distance: from 1 to 0.05 ± 0.02). This result is explained by the fact that the increase of the edge weight results in i) an increase of the eigenvalues for the adjacency matrix of compared graphs and thus ii) a smaller value of \( \text{sim}_{\text{l}}(G_1, G_2) \). For DeltaCon algorithm, the increase of edge weight will increase the affinity scores between nodes and then the distance between graphs will increase making \( \text{sim}_{\text{DeltaCon}}(G_1, G_2) \) smaller. Interestingly, and in contrast with the four previous methods, the SimNet algorithm disclosed a gradually decreasing similarity index \( \text{sim}_{\text{SimNet}}(G_1, G_2) \) for gradual alteration of the edge weight.

In Fig. 3B, we show the evolution of the five similarity indexes with respect to gradual insertion of nodes in the altered graph. Typical examples of the simulated networks are also shown (Fig. 3.B, bottom) for the initial graph \( G_1 \) and for altered versions \( G_2 \) at level 10 and 18.

Results indicate that the five similarity indexes all decreased with the increase of the number of inserted nodes. DeltaCon and VEO showed fairly similar results with a relatively slow decrease rate (VEO: 0.9705 ± 0.01 at level 4 to 0.91 ± 0.02 at level 20; DeltaCon: 0.95 ± 0.004 at level 4 to 0.9028 ± 0.005 at level 20). For the same alteration levels, SimNet showed a more pronounced decrease of the similarity index (level 4: 0.947 ±0.006, level 20: 0.73 ± 0.005). In the case of Lambda distance, \( \text{sim}_{\text{l}}(G_1, G_2) \) values changed from 0.83 ± 0.06 (level 4) to 0.51 ± 0.06 (level 20). Interestingly, for the four above-described algorithms, a linear decrease of the similarity index was observed. Finally,
showed the most marked decrease (nonlinear in this case, GED: 0.28 ± 0.06 at level 4 to 0.09 ± 0.01 at level 20). In Fig. 3C, the evolution of the similarity index values with respect to alterations in the node positions is presented. Results confirmed that SimNet is the only algorithm showing sensitivity to this factor.

Indeed, values decreased dramatically with the alteration level (level 1: 0.7825 ± 0.05, level 9: 0.095 ± 0.007). Results indicated that similarity indexes computed from the four other algorithms exhibit different values (DeltaCon: 0.397 ± 0.001; GED: 0.042 ± 0.003; Lambda distance: 0.15 ± 0.012; VEO: 0.02 ± 0.01) that did not change with the level of alteration as expected.

B. Application to real data

Results obtained from the application of SimNet to real brain networks are presented in Fig. 4. The similarity scores between networks identified for the object and animal categories in each of the 21 subjects are represented as two [21×21] matrices where lines represent the brain networks (i.e. G1 graphs) associated with object visual stimuli and where columns represent networks (i.e. G2 graphs) associated with animal stimuli. As the similarity index sim_{norm}(G1,G2) is symmetric, only the values of the upper triangle are displayed. As depicted, the first matrix (Fig. 4A, left) shows high similarity values during the first period corresponding to visual processing (1-119 ms). In contrast, lower similarity values (Fig. 4A, right) were observed during the second period corresponding to memory access, (150-190 ms). Typical examples of brain networks with high similarity values at the first period and low similarity values at the second period are illustrated in Fig. 4B where the node size represents the “strength value” a network measure defined as the sum of weights of edges connected to this node. For instance, during the visual processing period, a high similarity (sim_{norm} = 0.51) was measured between the network obtained when subject 14 was naming animal stimuli and the network obtained when subject 13 was naming object stimuli. In contrast, during the memory access period, a low similarity index (sim_{norm} = 0.04) was computed when the network obtained when subject 16 was naming animal stimuli and that obtained when subject 3 was naming object stimuli. Cells in the diagonal of the matrix represent the similarity indices when comparing object and animals within subjects. Even in this case, sim_{norm} is higher during the first (average of similarity values in the diagonal: 0.1735) period than the second one (average of similarity values in the diagonal: 0.0914).

Fig. 4C shows the boxplots of three similarity indexes (sim_{norm}(G1,G2), sim_{hubnet}(G1,G2), sim_{rep}(G1,G2)) obtained by the three corresponding methods (SimNet, DeltaCon and GED) during the two periods (visual processing and memory access). The three methods indicate a decrease in the inter-conditions (objects vs. animals) similarity values. During visual processing (1-119 ms), median value of 0.1, 0.22 and 0.14 were observed for SimNet, DeltaCon and GED, respectively. As expected, these values decreased during the memory access period (151-190 ms) with median values of 0.03, 0.21 and 0.12 for the three methods, respectively. However, and interestingly, SimNet is the only method showing a statistically significant difference between the two periods (Wilcoxon test, p<0.01).

IV. DISCUSSION

A challenging issue in brain research is to measure the similarity between two spatially-defined networks. In this paper, a new algorithm, called SimNet, was proposed to calculate the similarity between two graphs with known coordinates for nodes. The proposed algorithm is based on the calculation of i) a distance between nodes in the two graphs, itself based on a least cost sequence of changes (substitution, insertion and deletion of nodes) that are necessary to map one graph on the other and ii) the difference edge weights between the two graphs. The proposed algorithm was compared with four previously-published graph similarity algorithms (GED, DeltaCon, VEO and λ-distance). SimNet was shown to improve results in simulated situations involving a shifting of the location of nodes. SimNet showed also a higher performance in comparing real brain networks obtained from dense EEG during a cognitive task consisting in naming items of two different categories (objects, animals). These findings are discussed hereafter.

Network similarity under spatial constraint

Measuring similarity among networks is a topic of increasing interest [31]. Several approaches have been proposed to compare networks, in various application domains (social networks, biology, bioinformatics...). The techniques used for comparing brain networks can be classified into three categories: i) global, consisting in comparing between global graph measures (degree, modularity, hubs...) computed from the two networks [32, 33], ii) node-wise, consisting in computing graph metrics for each node of the networks such that multiple comparisons can be assessed [34] and iii) edge-wise, consisting in comparing all the available edges in the networks, an approach called Network Based Statistics (NBS) [35]. However, the spatial location of nodes is not accounted for in these approaches.

Recently, the spatial constraint was considered in a new metric for computing graph measures in brain networks analysis [18]. Authors showed that this metric could distinguish the global connectivity of structural networks from functional networks only when the physical locations of nodes are considered. Along the same line, we propose a new algorithm to measure the similarity between graphs. The key feature of this algorithm is that it takes into account the physical locations of the network nodes. The results of SimNet applied to real data confirmed the importance of including the physical location of nodes for assessing the (di)similarity of brain networks involved into two distinct steps of a cognitive task.

As compared with SimNet, the four other algorithms investigated in this study showed lower performance in detecting the similarity when the two networks do not have the same number of nodes or the spatial location of the nodes is changing. Note that this does not detract the importance of
Fig. 4. A- Inter-subject variability of the similarity index ($\text{sim}_{\text{SimNet}}(G_1, G_2)$) on real brain networks identified from EEG where $G_i$ and $G_j$ represent respectively the connectivity graphs of the subjects during tools and animals picture naming. Left: value of connectivity graphs in the first period (1-119ms). Right: similarity values of connectivity graphs during the second period (151-190ms). B- The connectivity graphs: 3D representation for 2 different subjects. C- Boxplots show significant difference of similarity values between the two first periods of the cognitive process using SimNet. Networks were obtained and visualized using EEGNET [36].
these algorithms as they were developed for specific applications where the spatial location of graphs is a non-relevant factor.

Methodological considerations

First, the distance used between nodes in the simulation case was assumed to be Euclidian. This distance is not fully appropriate in the real case as brain networks take place over a folded brain surface which consists of sulci and gyri. For this reason the geodesic distance was preferred to the Euclidian distance. Second, a crucial parameter to be tuned in SimNet is the radius of the disk used to detect the neighbors of a given node. An increase of the radius $R$ will automatically lead to an increase of the similarity index between the two compared graphs. In the simulation case, $R$ was chosen to be equal to 1.5 as representing the minimal Euclidian distance between two nodes in the grid (example of Fig. 1). In the real case, $R$ was chosen as the average distance between all nodes. As we were aware about the effect of $R$ on the similarity value, comparative analyses were performed for different values of $R$ (the minimal distance and the maximal distance between two nodes). The results (not shown here) indicated that the performance of SimNet with respect to the other tested algorithms is preserved, whatever the value of $R$ provided that the condition $(\text{Cost (substitution)} < \text{Cost(Insertion)} + \text{Cost (Deletion)})$ is preserved.

Object categorization in the human brain: a network-based approach

SimNet was originally developed to analyze the similarity between brain networks involved in cognitive tasks. In this study, the EEG source connectivity analysis allowed us to identify brain networks at cortical level from dense-EEG scalp recordings [24]. Different brain networks were identified for two different categories of stimuli (objects vs. animals). Our intent was to assess the capability of SimNet and other tested methods to detect significant differences in identified networks.

To our knowledge, this study constitutes the first attempt to assess object categorization in the human brain from a network-based approach using dense EEG source connectivity. During the picture naming task, we detected significant difference between networks identified during the time period associated with ‘visual processing’ and that related to the ‘access to memory’. During visual processing, the networks were mainly occipital involving the inferior occipital, the lateral occipito-temporal sulcus and occipital pole. This period was shown to be related to the visual feature extraction preceding the object category recognition [28, 37]. In the present task that just consists in naming pictures overtly, we don’t ask the participant to detect animals or to categorize visual scene. As a consequence to this, influences from semantic information on earlier visual processes are low or have not started yet. Moreover, most of the period corresponding to the first time-window precedes the N1, a component of ERP studies that peaks around 150 ms well known to be the first component sensitive to semantic modulations. Obviously some participants can be aware of viewing an animal as soon as 70 to 80 ms after the stimulus onset [28] but it could remain unconscious until its features are mapped onto a memorized concept. These considerations explain the high similarity values observed between object and animal networks during this time period. For the second period, results showed a network involving the occipital regions but with an implication of the bilateral inferior temporal sulcus. This network is known to be related to semantic working memory system when someone tries to remind the name of the presented object [38]. This period was considered as the first instant of categorization in the human brain [39]. More cautiously, this can explain the significantly lower similarity values between graphs for animals and objects which could then share different neural substrates.

This latter result could also explain the statistical significant difference between the two periods; the similarity index of the second period being lower than the similarity index of the first one.

Two less cognitive but more plausible explanations of these effects could be that i) variability in general increases with time after stimulus onset: indeed, time response latencies, inter-subject differences and attentional level are known to fluctuate (the farther from the stimulus onset, the higher the variability is) and that ii) the second period is shorter in duration than the first period. During the stage of brain network identification, graphs that served to segment the period between 150ms and 190ms are less robust than those which allow identifying the first period. They poorly contribute to the global explained variance (49% vs 63% for the first period) despite remarkable good presence indexes (81% vs. 80 % for the first period).

Apart from which of the previous explanations created these effects (similarity index changes and its significant decrease), both effects were captured by the SimNet algorithm while the other algorithms failed. As the other algorithms showed no significant difference, we assume that the difference between the networks identified over the two periods is probably related to variations in the spatial location of the nodes, a feature taken into account only by SimNet.

Moreover, in this paper we evaluated the performance of SimNet on two different categories: objects and animals. In line with the work of Gallant et al [40, 41], these findings could be extended to build a ‘semantic space’ of the visual stimuli (human vs non-human, mobile vs immobile, social vs non-social …) based on the similarity index calculated using SimNet. Because this interpretation remains quite speculative in this paper, we are currently addressing the question of early (not semantic) vs late processes (semantic) with a new dataset using different modalities (visual and auditory). This will allow us to get the reverse pattern of similarity index from the first time period with very dissimilar graphs due to different modalities to the second time period with more similar graphs as the semantic system is shared between modalities.

V. CONCLUSION

In this paper, a new algorithm, called SimNet, for quantifying the similarity between networks under a spatial constraint (position of nodes) was proposed. On simulated
graphs, this algorithm showed higher performance than four state-of-the-art algorithms in detecting some shifts in the node location. When applied to real EEG data, SimNet could detect significant differences in brain networks associated with two different categories of pictures (objects and animals) used in a cognitive task. We believe that the proposed algorithm can be useful in pattern analysis problems involving a quantification of the similarity between graphs in which the physical location of the nodes is a key parameter.

References

thousands of object and action categories across the human brain,"
Supplementary Material for

SimNet: a Novel Method for Quantifying Brain Network Similarity

Ahmad Mheich, Mahmoud Hassan, Mohamad Khalil, Vincent Gripon, Olivier Dufor and Fabrice Wendling
**Materials and methods**

**Real Data**

Table S1: Psycholinguistic parameters controlled to get equivalent datasets for each category of pictures (Animals vs Tools)

<table>
<thead>
<tr>
<th></th>
<th>Name agreement (%)</th>
<th>Image agreement (average)</th>
<th>Age of acquisition (average)</th>
<th>Number of letters</th>
<th>Number of phonemes</th>
<th>Number of syllables</th>
<th>Number of morphemes</th>
<th>Oral frequency</th>
<th>Written frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Animals</strong></td>
<td>94.3243</td>
<td>3.7550</td>
<td>2.3292</td>
<td>5.9730</td>
<td>4.2162</td>
<td>1.7297</td>
<td>1.0541</td>
<td>12.1324</td>
<td>11.1673</td>
</tr>
<tr>
<td><strong>Average Objects</strong></td>
<td>95.8378</td>
<td>3.6126</td>
<td>2.5289</td>
<td>6.1622</td>
<td>4.2703</td>
<td>1.7027</td>
<td>1.1351</td>
<td>9.1903</td>
<td>11.0351</td>
</tr>
<tr>
<td><strong>STDEV Animals</strong></td>
<td>7.5058</td>
<td>0.4481</td>
<td>0.6152</td>
<td>1.9506</td>
<td>1.7502</td>
<td>0.8383</td>
<td>0.2292</td>
<td>13.9555</td>
<td>11.3952</td>
</tr>
<tr>
<td><strong>STDEV Objects</strong></td>
<td>5.8572</td>
<td>0.7356</td>
<td>0.5819</td>
<td>1.9792</td>
<td>1.6608</td>
<td>0.7769</td>
<td>0.4191</td>
<td>6.9085</td>
<td>6.8723</td>
</tr>
<tr>
<td><strong>T-test Animals vs. Objects</strong></td>
<td>0.3370</td>
<td>0.3189</td>
<td>0.1557</td>
<td>0.6800</td>
<td>0.8920</td>
<td>0.8860</td>
<td>0.3064</td>
<td>0.2556</td>
<td>0.9520</td>
</tr>
</tbody>
</table>
Figure S1: Pictures selected from the Snodgrass database.
Mahmoud Hassan, Isabelle Merlet, Ahmad Mheich, Aya Kabbara, Arnaud Biraben, Anca Nica and Fabrice Wendling

Published in: brain topography (2016)

Summary

Objectives: The epileptic network usually involves spatially distributed brain regions. In this context, noninvasive M/EEG source connectivity is an emerging technique to identify functional brain networks at cortical level from noninvasive recordings. Indeed various algorithms were used to reconstruct brain sources and both functional and effective connectivity measures to assess statistical couplings among time series associated with reconstructed sources. Therefore, a central issue is the impact of selected methods (EEG inverse solution and connectivity measure) on the topological/statistical properties of identified epileptic networks activated during paroxysmal activity.

Methods: In this paper, we analyze the effect of the two key factors involved in EEG source connectivity processing: i) the algorithm used in the solution of the EEG inverse problem and ii) the method used in the estimation of the functional connectivity. We evaluate four inverse solutions algorithms (dSPM, wMNE, sLORETA and cMEM) and four connectivity measures ($r^2$, $h^2$, PLV, and MI) on data simulated from a combined biophysical/physiological model to generate realistic interictal epileptic spikes reflected in scalp EEG. We use a new network-based similarity index (SI) to compare between the network identified by each of the inverse/connectivity combination and the original network generated in the model. The method will be also applied on real data recorded from one epileptic patient who underwent a full presurgical evaluation for drug-resistant focal epilepsy.

Results: In simulated data, results revealed that the selection of the inverse/connectivity combination has a significant impact on the identified networks. Results suggested that nonlinear methods (nonlinear correlation coefficient, phase synchronization and mutual information) for measuring the connectivity are more efficient than the linear one (the cross correlation coefficient). The wMNE inverse solution showed higher performance than dSPM, cMEM and sLORETA. In real data, the combination (wMNE/PLV) led to a very good matching between the interictal
epileptic network identified from noninvasive EEG recordings and the network obtained from connectivity analysis of intracerebral EEG recordings. These results suggest that source connectivity method, when appropriately configured, is able to extract highly relevant diagnostic information about networks involved in interictal epileptic spikes from non-invasive dense-EEG data.
Identification of Interictal Epileptic Networks from Dense-EEG

Mahmoud Hassan1,2 • Isabelle Merlet1,2 • Ahmad Mheich1,2,4 • Aya Kabbara1,2,4 • Arnaud Biraben1,2,3 • Anca Nica3 • Fabrice Wendling1,2

Received: 14 March 2016 / Accepted: 16 August 2016 © Springer Science+Business Media New York 2016

Abstract Epilepsy is a network disease. The epileptic network usually involves spatially distributed brain regions. In this context, noninvasive M/EEG source connectivity is an emerging technique to identify functional brain networks at cortical level from noninvasive recordings. In this paper, we analyze the effect of the two key factors involved in EEG source connectivity processing: (i) the algorithm used in the solution of the EEG inverse problem and (ii) the method used in the estimation of the functional connectivity. We evaluate four inverse solutions algorithms (dSPM, wMNE, sLORETA and cMEM) and four connectivity measures ($r^2$, $h^2$, PLV, and MI) on data simulated from a combined biophysical/physiological model to generate realistic interictal epileptic spikes reflected in scalp EEG. We use a new network-based similarity index to compare between the network identified by each of the inverse/connectivity combination and the original network generated in the model. The method will be also applied on real data recorded from one epileptic patient who underwent a full presurgical evaluation for drug-resistant focal epilepsy. In simulated data, results revealed that the selection of the inverse/connectivity combination has a significant impact on the identified networks. Results suggested that nonlinear methods (nonlinear correlation coefficient, phase synchronization and mutual information) for measuring the connectivity are more efficient than the linear one (the cross correlation coefficient). The wMNE inverse solution showed higher performance than dSPM, cMEM and sLORETA. In real data, the combination (wMNE/PLV) led to a very good matching between the interictal epileptic network identified from noninvasive EEG recordings and the network obtained from connectivity analysis of intracerebral EEG recordings. These results suggest that source connectivity method, when appropriately configured, is able to extract highly relevant diagnostic information about networks involved in interictal epileptic spikes from non-invasive dense-EEG data.

Keywords Epilepsy • Dense-EEG source connectivity • Epileptic networks

Introduction

Epilepsy is a network disease (Engel Jr et al. 2013). Over the two past decades, the concept of “epileptic focus” has progressively evolved toward that of “epileptic network” (Kramer and Cash 2012; Laufs 2012). Indeed, with the progress of functional neuroimaging techniques, many studies confirmed that the epileptic zone (EZ) can rarely be reduced to a circumscribed brain area (Bartolomei et al. 2001) as it very often involves distinct brain regions generating both interictal (Bourien et al. 2005) and ictal activity (Bourien et al. 2004). Among the investigation techniques classically used in the diagnostic of epilepsy, electrophysiological recordings (typically, magneto- and electro-encephalography including depth-EEG, referred to
as M/EEG) are still extensively used to localize and delineate the EZ in a patient-specific context. Regarding the numerous methods proposed to process the recorded data; those aimed at characterizing brain connectivity are particularly suitable to identify networks involved during epileptiform activity (both interictal and ictal).

In the context of invasive EEG signals (intracranial EEG, stereo-EEG and electrocorticoGraphy –EcoG–) recorded in patients candidate to surgery, these “connectivity” methods have been a topic of extensive research [see van Mierlo et al. (2014) for recent review]. For instance, the coherence function was shown to localize the seizure onset (Gotman 1987), similarity indexes were used to distinguish a preictal state from the ongoing interictal activity (Le Van Quyen et al. 2005; Mormann et al. 2000). Nonlinear regression analysis was applied to intracerebral signals to characterize connectivity patterns at the seizure onset (Bourien et al. 2004). Readers may refer to previous reviews for more detailed information about brain connectivity methods applied to non-invasive (van Mierlo et al. 2014) and invasive EEG signals (Wendling et al. 2010) in drug-resistant focal epilepsies.

In the context of scalp M/EEG recording, connectivity methods have received less attention as compared with invasive EEG. A number of studies performed at the level of electrodes and focused on ictal periods have been reported aiming at analyzing seizure propagation (Gotman 1983) or to determine the seizure onset side (Caparos et al. 2006), for instance. For interictal periods, few connectivity studies made use of dense EEG and phase synchronization (Ramon and Holmes 2013) to identify epileptic sites. One reason for this paucity of studies may lie in the intricate interpretation of connectivity measures obtained from scalp recordings. Indeed, this interpretation is not straightforward as signals are severely corrupted by the effects of volume conduction (Schoffelen and Gross 2009).

Interestingly, some recent studies showed how to overcome this limitation. In line with previous cognitive studies (Astolfi et al. 2007; Babiloni et al. 2005; Betti et al. 2013; Bola and Sabel 2015; David et al. 2003; David et al. 2002; de Pasquale et al. 2010; Hassan et al. 2015a; Hassan et al. 2014; Hassan and Wendling 2015; Hipp et al. 2011; Hoechstetter et al. 2004; Liljestrom et al. 2015; Schoffelen and Gross 2009), the basic principle is to estimate functional connectivity at the level of brain sources reconstructed from scalp signals. These methods, referred to as “source connectivity” were applied to both interictal EEG (Coito et al. 2015; Song et al. 2013; Vecchio et al. 2014) and MEG signals (Dai et al. 2012; Malinowska et al. 2014) as well as to EEG signals recorded during seizures (Ding et al. 2007; Jirouska et al. 2013; Lu et al. 2012) or resting states (Adebimpe et al. 2016; Coito et al. 2016).

Although these approaches all include two steps (M/EEG inverse problem followed by source connectivity estimation), they strongly differ from a methodological viewpoint. Indeed, various algorithms were used to reconstruct brain sources and both functional and effective connectivity measures were utilized to assess statistical couplings among time series associated with reconstructed sources. Therefore, a central issue is the impact of selected methods (EEG inverse solution and connectivity measure) on the topological/statistical properties of identified epileptic networks activated during paroxysmal activity.

In this paper, we report a quantitative comparison of methods aimed at identifying cortical epileptic networks from scalp EEG data. The novelty of this work is twofold. First, our comparative study includes simulated dense EEGs generated from physiologically- and biophysically-plausible models of distributed and coupled epileptic sources. To our knowledge, no previous study has reported results on the performance of source connectivity methods based on a “ground truth” provided by realistic computational models of interictal EEG signals (recorded later in time than the dense EEG recordings). Second, in line with a recent analysis performed on MEG data (Malinowska et al. 2014), networks estimated from real scalp dense EEG are compared with those obtained from depth-EEG recordings (SEEG).

Materials and Methods

Inverse Solution Algorithms

Given the equivalent current dipole model, EEG signals X(t) recorded from M channels can be considered as linear combinations of P time-varying current dipole sources S(t):

\[ X(t) = GS(t) + N(t) \]

where G[M, P] is the lead field matrix and N(t) is the noise. As G is known, the EEG inverse problem consists of estimating the unknown sources \( \hat{S}(t) \) from X(t). Several algorithms have been proposed to solve this problem based on different assumptions about spatial and temporal properties of sources and regularization constraints. Here, we chose to evaluate the four algorithms implemented in Brainstorm (Tadel et al. 2011).

Weighted Minimum Norm Estimate (wMNE)

Minimum norm estimates (MNE) originally proposed by (Hämäläinen and Ilmoniemi 1994) are based on a search for the solution with minimum power using the L2 norm to regularize the problem. A common expression for MNE resolution matrix is
The dSPM is based on the MNE solution (Dale et al. 2000).

Dynamical Statistical Parametric Mapping (dSPM)

The dSPM is based on the MNE solution (Dale et al. 2000). For dSPM, the normalization matrix contains the minimum norm estimates of the noise at each source (Caparos et al. 2006), derived from the noise covariance matrix, defined as:

$$\hat{S}_{dSPM} = (G^T W_X G + \lambda C)^{-1} G^T W_X X$$

that adjusts the properties of the solution by reducing the bias inherent to the standard MNE solution. Classically, $W_X$ is a diagonal matrix built from matrix $G$ with non-zero terms inversely proportional to the norm of the lead field vectors.

Standardized Low Resolution Brain Electromagnetic Tomography (sLORETA)

sLORETA uses the source distribution estimated from MNE and standardizes it a posteriori by the variance of each estimated dipole source:

$$\hat{S}_{sLORETA} = W_{sLORETA} \hat{S}_{MNE}$$

where $W_{sLORETA}^2 = \text{diag}(\hat{S}_{MNE} C \hat{S}_{MNE}^T)$. The sLORETA inverse method has been originally described using the whole brain volume as source space (Pascual-Marqui 2002). For the present study, in order to ease the comparison with other methods, we restricted the source space to the neocortical surface.

Standard Maximum Entropy on the Mean (cMEM)

The Maximum Entropy on the Mean (MEM) solver is based on a probabilistic method where inference on the current source intensities is estimated from the data, which is the basic idea of the maximum of entropy. The first application of MEM to electromagnetic source localization was reported in (Clarke and Janday 1989). The main feature of this method is its ability to recover the spatial extent of the underlying sources. Its solution is assessed by finding the closest distribution of source intensities to a reference distribution in which source intensities are organized into cortical parcels showing homogeneous activation state (parallel cortical macro-columns with synchronized activity). In addition a constraint of local spatial smoothness in each parcel can be introduced (Chowdhury et al. 2013).

Connectivity Measures

We selected four methods that have been widely used to estimate functional brain connectivity from electrophysiological signals (local field potentials, depth-EEG or EEG/MEG) (see (Wendling et al. 2009) for review). These measures were chosen to cover the main families of connectivity methods (linear and nonlinear regression, phase synchronization and mutual information).

Briefly, concerning the regression approaches, the linear cross-correlation coefficient is only limited to the detection of the linear properties of the relationships between time series. However, mechanisms at the origin of EEG signals were shown to have strong nonlinear behaviors (Pereda et al. 2005). Thus, we have selected three nonlinear connectivity measures. The nonlinear regression where the basic idea is to evaluate the dependency of two signals from signal samples only and independently of the type of relation between the two signals. Concerning the phase synchronization measure, the method estimates the instantaneous phase of each signal and then computes a quantity based on co-variation of extracted phases to determine the degree of relationship. Finally, the mutual information method is based on the probability and information theory to measures mutual dependence between two variables. More technical details about the four methods are presented hereafter:

Cross-Correlation Coefficient ($r^2$)

The cross-correlation coefficient measures the linear correlation between two variables $x$ and $y$ as a function of their time delay ($\tau$). Referred to as the linear correlation coefficient, it is defined as:

$$r_{xy}^2 = \max_{\tau} \frac{\text{cov}(x(t), y(t + \tau))}{(\sigma_x(t) \sigma_y(t+\tau))^2}$$

where $\sigma$ and cov denote the standard deviation and the covariance, respectively.

Nonlinear Correlation Coefficient ($h^2$)

The nonlinear correlation coefficient ($h^2$) is a non-parametric measure of the nonlinear relationship between two time series $x$ and $y$. In practice, the nonlinear relation
between the two time series is approximated by a piecewise linear curve.

\[ h_{xy}^2 = \max_t \left( 1 - \frac{\text{var}(y(t + \tau)/x(t))}{\text{var}(y(t + \tau))} \right) \]

where \( \text{var}(y(t + \tau)/x(t)) \) is the linear piecewise approximation of the non-linear regression curve.

**Mutual Information (MI)**

The mutual information (MI) between signal \( x \) and \( y \) is defined as:

\[ MI_{xy} = \sum p_{xy} \log \frac{p_{xy}}{p_x p_y} \]

where \( p_{xy} \) is the joint probability of \( x = x_i \) and \( y = y_j \). In the case of no relationship between \( x \) and \( y \), \( p_{xy} = p_x p_y \), so that the MI is zero for independent processes. Otherwise, \( MI_{xy} \) will be positive, attaining its maximal value for identical signals.

**Phase Locking Value (PLV)**

For two signals \( x \) and \( y \), the phase locking value is defined as:

\[ PLV_{xy} = \left| \langle e^{i[\phi_x(t) - \phi_y(t)]} \rangle \right| \]

where \( \phi_x(t) \) and \( \phi_y(t) \) are the unwrapped phases of the signals \( x \) and \( y \) at time \( t \). The \( \langle . \rangle \) denotes the average over time. The Hilbert transform was used to extract the instantaneous phase of each signal.

The \( h^2 \), PLV and \( r^2 \) values range from 0 (independent signals) to 1 (fully correlated signals).

**Data**

**Simulations**

In order to quantitatively assess the performance of source connectivity approaches, we generated simulated EEG signals following the procedure described in (Cosandier-Rimélé et al. 2008), see Fig. 1a. The distributed source space consisted in a mesh of the cortical surface (8000 vertices, \( \sim 5 \) mm inter-vertex spacing) that was obtained by segmenting the grey-white matter interface from a normal subject’s structural T1-weighted 3D-MRI with Freesurfer (Fischl 2012). Dipoles were located at each vertex of this mesh and oriented radially to the surface at the midway between the white/grey matter interface and the pial surface. The time-course of each dipole of the source space was generated from a modified version of the physiologically relevant neural mass model reported in (Bourien et al. 2005; Wendling et al. 2000, 2002).

In brief, this computational model was designed to represent a neuronal population with three subsets of neurons (pyramidal cells P and interneurons I and I’) interacting via synaptic transmission (Fig. 1a). Pyramidal cells (P) receive endogenous excitatory drive (AMPAergic collateral excitation) from other local pyramidal cells and exogenous excitatory drive from distant pyramidal cells (via external noise input \( p(t) \)). They also receive inhibitory drive (GABAergic feedback inhibition) from both subsets of local interneurons (I and I’). In turn, interneurons receive excitatory input (AMPA) from pyramidal cells.

A Gaussian noise was used as external input to neuronal population. The mean (\( m = 90 \)) and standard deviation (sigma = 30) were adjusted to represent randomly varying density of incoming action potentials (Aps). However, for the purpose of this study, a modification was made to this noise model. Indeed, abrupt increase/decrease of the density of Aps can occur in the external input noise at user-defined times to mimic transient AP volleys from other brain regions involved in the generation of interictal events. Thus, in this model, a simulated IES can be viewed as the responses of a nonlinear dynamical system (comprising positive and negative feedback loops) to transient pulses superimposed on a Gaussian noise (classically used in the neural mass modeling approaches).

As in the standard implementation, the shape (spike component followed by a wave component) can still be controlled by adjusting excitation and inhibition parameters of each population (gains in feedback loops). However, the aforementioned modification offers one major advantage: as pulses in the noise input are user-defined, the occurrence times of simulated IESs are controlled, in
A Simulated epileptic spikes

B Identification of epileptogenic network

C Intracerebral recordings
contrast with the standard implementation where IESs simply result from random fluctuations of the noise. The consequence is that this new model feature allows for simulation multi-focal IESs with well-controlled time shifts. Indeed, as illustrated in Fig. 1a, we could generate delayed epileptiform activity in multiple distant patches just by introducing short delays between the pulses superimposed on the respective input noises of neuronal populations at each patch.

Finally, from appropriate setting of the input noise, as well as excitation and inhibition-related parameters at each neural mass included in simulated epileptic sources, a set of epileptiform temporal dynamics was obtained. These dynamics were assigned to a source made of contiguous vertices (active source) manually outlined with a mesh visualization software (Paraview, Kitware Inc., NY, US). Uncorrelated background activities were attributed to the other vertices. Once the amplitude of each elementary dipole was known, EEG simulations were obtained by solving the forward problem in a 3-layer realistic head model (inner skull, outer skull and the scalp with conductivity values of 0.33, 0.0042, 0.33 S/m respectively) using the Boundary Element Method (BEM) with the boundary values of 0.33, 0.0042, 0.33 S/m respectively according to the following criteria: i) only contacts showing grey matter activity were retained and ii) among them, only the contact showing the maximal activity was kept when similar intracranial activity was observed on several contacts.

Finally, from appropriate setting of the input noise, as well as excitation and inhibition-related parameters at each neural mass included in simulated epileptic sources, a set of epileptiform temporal dynamics was obtained. These dynamics were assigned to a source made of contiguous vertices (active source) manually outlined with a mesh visualization software (Paraview, Kitware Inc., NY, US). Uncorrelated background activities were attributed to the other vertices. Once the amplitude of each elementary dipole was known, EEG simulations were obtained by solving the forward problem in a 3-layer realistic head model (inner skull, outer skull and the scalp with conductivity values of 0.33, 0.0042, 0.33 S/m respectively) using the Boundary Element Method (BEM) with the OpenMEEG (Gramfort et al. 2010) implemented in Brainstorm software.

We considered two different scenarios. In the first one (single network), EEG simulations were generated from a single source located in the inferior parietal region (~1000 mm$^2$). In the second one (two interconnected networks) an additional source (~1000 mm$^2$) was placed in the middle temporal gyrus. In that case, the temporal dynamics of the second source were highly correlated with those of the first source, but with a minor delay (30 ms). This delay of 30 ms was in the range of 10–50 ms delays that are often observed during interictal spikes at different intracranial recording location (Alarcon et al. 1994, 1997; Emerson et al. 1995; Merlet and Gotman 1999) or at different surface sensors (Barth et al. 1984; Ebersole 1994) or between the peaks of dipole source activity (Baumgartner et al. 1995; Merlet and Gotman 1999). This delay was usually interpreted as reflecting propagation between distant regions in the brain. For each scenario, 20 epochs of 60 s at 512 Hz containing 30 epileptic spikes were simulated. Each epoch was obtained for a new realization of the input random noise leading to a new realization of epileptic spikes occurring in background activity. Simulated data were imported in Brainstorm for further analysis.

**Real Data**

Real data were selected from a patient who underwent presurgical evaluation for drug-resistant focal epilepsy. Seizures were stereotyped, with a sudden start, febrile motor automatisms of the upper limb, stretching of the neck and torso and no post-ictal motor deficit. The patient had a comprehensive evaluation including detailed history and neurological examination, neuropsychological testing, structural MRI, standard 32-channels (Micromed, Italy) as well as Dense-EEG 256-channels (EGI, Electrical Geodesic Inc., Eugene, USA) scalp EEG with video recordings and intracerebral EEG recordings (SEE). MRI showed a focal cortical dysplasia in the mesial aspect of the orbitofrontal region. Dense-EEG was recorded for 1 h, at 1000 Hz following the procedure approved by the National Ethics Committee for the Protection of Persons (CPP, agreement number 2012-A01227-36). The patient gave his written informed consent to participate in this study. This recording revealed sub-continuous spike activity at the most left frontopolar basal electrodes. From this interictal epileptic activity, 85 spikes were visually selected away from the occurrence of any artefacts (muscle activity, blood pulsation, eye blinks). Each spike was centered in a 2 s window and all 85 windows were concatenated for further analysis.

As part of his presurgical evaluation, the patient also underwent intracerebral SEEG recordings with 9 implanted electrodes (10 ± 18 contacts; length: 2 mm, diameter: 0.8 mm; 1.5 mm apart) placed intracranially according to Talairach’s stereotactic method in the left frontal and temporal region, see Fig. 1c. The positioning of the electrodes was determined from available non-invasive information and hypotheses about the localization of his epileptic zone. From these data, subsets of 25 out of the 118 original leads were selected. This selection was made according to the following criteria: i) only contacts showing grey matter activity were retained and ii) among them, only the contact showing the maximal activity was kept when similar intracranial activity was observed on several contacts.

**Data Analysis**

**Scalp-EEG Based Interictal Epileptic Networks**

As illustrated in Fig. 1B, source activity was estimated using four inverse algorithms (dSPM, wMNE, sLORETA and cMEM, see “Materials and Methods” section). A baseline of 1 s length was used to estimate the noise covariance matrix both on simulated and real scalp EEG data. For real data, source localization was applied on averaged spikes, taking as time reference the maximum of the negative peak, while for simulated data the source localization was applied on non-averaged spikes. The cortical surface was anatomically parcellated into 148 regions of interest (ROI) (Destrieux et al. 2010) and then...
re-subdivided into ~1500 sub-ROIs using Brainstorm. The 148 ROIs provided initially by the Destrieux Atlas (using Freesurfer) were quasi equally subdivided to obtain the 1500 sub-ROIs with 1 cm² average sizes. Time series of the reconstructed source activities were averaged over each of the 1500 ROIs.

After the reconstruction of sources (source localization and estimation of temporal dynamics), functional connectivity was estimated using four methods ($r^2$, $h^2$, PLV, and MI, see “Materials and Methods” section). Each quantity was computed on the set of 2 s single spikes. All connectivity matrices (1500 × 1500) were thresholded as follows. We computed the strength of each node of the weighted undirected graph and we kept nodes with the highest 1 % strength values. The same threshold was applied on the adjacency matrices for all combinations (inverse/connectivity). The strength was defined as the sum of all edge weights for each node; all weights were positive and normalized between 0 and 1.

In order to define the reference networks, all the dipoles were supposed synchronized and the reference network reflected a fully connected undirected graph. In the case of double network scenario, a number of 37 sub-regions (nodes) were considered. The dynamics of the dipoles associated to these nodes were similar and resulting a 37 × 37 fully connected network where connections (local and remote) between the 37 nodes have the same weight value.

Quantification of Network Similarity

In order to compare the reference brain network simulated in the model with the network identified from simulated scalp EEG by each of the inverse/connectivity combination (Fig. 1b), we used a network similarity algorithm recently developed in our team (Mheich et al. 2015a), see supplementary materials for more details about the algorithm. The main advantage of this algorithm is that it takes into account the spatial location (3D coordinates) of the nodes when comparing two networks, in contrast with other methods based on the sole statistical properties of compared graphs. The algorithm provides a normalized Similarity Index (SI): 0 for no similarity and 1 for two identical networks (same properties and topology). The connectivity analysis, the network measures and network visualization were performed using EEGNET (Hassan et al. 2015a, b).

Depth-EEG Based Interictal Epileptic Networks

Functional connectivity using $h^2$ were directly computed from SEEG signals at the 25 selected intracerebral electrode contacts. Adjacency matrices (25 × 25) were obtained and thresholded using the same procedure than that applied to the graphs obtained for scalp dense EEG (both simulated and real).

Scalp-EEG-Based Versus Depth-EEG-Based Epileptic Network Matching

In order to compare the graphs in the three-dimensional coordinates system of the cortex mesh, the 3D coordinates of the SEEG were first estimated by the co-registering the patient’ CT scan and MRI. These points were then projected on the surface mesh. The transformation from MRIs (voxels) coordinates to surface (SCS/MNI) coordinates was realized in brainstorm. The Scalp-EEG-based and depth-EEG-based epileptic networks were visually compared by matching the identified regions (nodes) in both networks.

Statistical Analysis

On the simulated data, a Wilcoxon rank-sum test was used to compare between the SIs obtained for each combination at each trial, corrected for multiple comparison using Bonferroni approach.

Results

Simulated Data: Influence of the Source Reconstruction/Functional Connectivity Combination

The results obtained in the case of the single network scenario are illustrated in Fig. 2, for the 16 different combinations of the source reconstruction and functional connectivity methods. The visual investigation of these results revealed that networks identified using the different combinations of methods were concordant with the reference network (Fig. 2b). Indeed none of the identified networks had nodes in a remote region (Fig. 2a). The qualitative analysis also showed that the number of nodes and the connections between them varied according to the combination of methods used. For a given connectivity approach, changing the localization method did not dramatically modify the network aspect, except for cMEM. Conversely, for a given source localization approach, changing the functional connectivity measure changed, qualitatively, the network. Although this was difficult to assess visually, $h^2$ combined with MNE or sLORETA was giving the network that best matched the reference network while cMEM/MI provided a result that was different from the reference network.

Quantification of these differences is provided in Fig. 2c. Overall, values of network similarity were relatively high and ranged from 70 to 82 %. For a given
connectivity approach, changing only the localization algorithm slightly modified SI values by 3% ($h^2$) to 8% (MI). For a given source localization approach, the SIs varied within 9% (wMNE) to 12% (dSPM). Results obtained using MI were on average better than PLV, $r^2$ and $h^2$. The combination providing the highest similarity values between the estimated and the actual network was dSPM/MI (82.2%) followed by wMNE/MI (82%) and wMNE-PLV (82%). The lowest similarity value was obtained with the dSPM/$h^2$ combination. The Wilcoxon rank-sum test did
not reveal any significant difference between the similarity values obtained in this first study.

Results obtained in the case of two interconnected networks for the 16 combinations of the inverse/connectivity methods are reported in Fig. 3. Results indicate that the networks identified by all the combinations are relatively close to the model network (Fig. 3b) since, similarly to the previously scenario, there was no node in other distant regions or in the right hemisphere. The networks did not qualitatively change much for a given connectivity measure except for cMEM. Rather, as observed in the first scenario, the variability between the different combinations was more related to the choice of the connectivity measure, given a source localization approach. The results of PLV (whatever the inverse solution algorithm) provide the closest result to the reference network. cMEM/MI shows also a relatively close network to the reference network while cMEM/h² indicated, visually, the farthest result from the reference network.

Values of network similarity are reported in Fig. 3c. These values were lower than those in the single network scenario, ranging from 57 to 73%. For a particular connectivity measure, changing the inverse algorithm modified the SIs by 1% (r²) to 8% (h²) while for a given source reconstruction algorithm, the SIs varied between 6% (dSPM) to 13% (wMNE). The combination providing the result closest to the reference network was wMNE/PLV (73%). High values were also obtained with sLORETA/PLV (68%) and cMEM/PLV (66%). The cMEM/h² combination shows the lowest SI value (57%).

Interestingly, for scenario 2 results obtained with wMNE/PLV were significantly closest to the actual network than the other ones (Wilcoxon rank-sum test, p < 0.01, corrected using Bonferroni).

EEG Source Localization Versus Functional Connectivity

An essential issue that is addressed in this paper relates to the difference between the proposed “network-based” approach and the classical approach using source localization only. In Fig. 4, we show two typical examples of the difference between the proposed network-based analysis and the classical localization approach. The first example is for cMEM combined with MI vs. cMEM only. This Figure shows that the information extracted in both cases was noticeably different. The source connectivity approach identified a network close to the reference one (Fig. 4a), with nodes both in the parietal and in the temporal region (Fig. 4b). There were no spurious nodes in remote regions. In contrast, with the sole source localization, after averaging the results over a 50 ms interval around each of the epileptic peaks, the parietal source was well retrieved while the temporal source remained almost unobserved. The second example was wMNE/PLV vs. wMNE, the figure shows that the network-based approach was able to identify a network close to the reference with no spurious connections in distant regions. The source localization approach identified the two regions different energies at. Moreover, many spurious sources were observed in remote regions. Similar results were observed for single network configuration.

Real Data: Scalp-EEG-Based Versus Depth-EEG-Based Epileptic Network

The results obtained from real data recorded in a patient are described on Fig. 5. In this patient, the sources of scalp EEG interictal spikes were widespread over the left frontal and temporal regions. Sources with maximum activity were found in the left frontal pole and orbitofrontal regions but a substantial activation was also detected in the left temporal as well as right frontal poles (Fig. 5a, left). When combining wMNE and PLV on the same scalp EEG data, the source connectivity approach retrieved a 5-nodes network in the left frontal lobe, involving the mesial (rectus gyrus) and lateral orbitofrontal region as well as the anterior cingulate gyrus (Fig. 5a, right). This result was concordant with that the network identified directly from intracerebral recordings by computing the functional connectivity among SEEG signals (Fig. 5b right). Indeed, the depth-EEG based network involved six nodes in the left mesial orbito-frontal (rectus gyrus), and anterior cingulate region. All these nodes were also identified by the visual analysis (Fig. 5b, left) as regions involved in the main interictal activity (rectus gyrus) as well as in the propagated interictal activity (cingulate gyrus).

The similarity indices between networks identified by each of the combination with the depth-EEG-based network are presented in Fig. 5c. Results showed that the wMNE/PLV provides the highest SI value (70%) followed by wMNE/h² (47%) and sLORETA/PLV (47%). The cMEM method showed the lowest SI values whatever the connectivity measure (6, 1, 1 and 1% for cMEM/MI, cMEM/PLV, cMEM/h² and cMEM/r² respectively).

Discussion

Identifying epileptic brain networks from noninvasive M/EEG data is a challenging issue. Recently, source localization combined with functional connectivity analysis led to encouraging findings in the estimation of functional cortical brain networks from scalp M/EEG recordings (Coito et al. 2015; Jiruska et al. 2013; Maliowska et al. 2014). Nevertheless, the joint use of these
two approaches is still novel and raises a number of methodological issues that should be controlled in order to get appropriate and interpretable results. In this paper, we reported a comparative study—in the context of epilepsy—of the networks obtained from all possible combinations between four algorithms to solve the EEG inverse problem and four methods to estimate the functional connectivity. An originality of this study is related to the use of dense EEG signals simulated data from a realistic model of multifocal epileptic zone as a ground truth for comparing the performance of considered methods. To our knowledge, a model-based evaluation of source connectivity methods has not been performed yet. A second—and still novel—aspect is the use of depth-EEG signals (intracerebral recordings performed during presurgical evaluation of drug-resistant epilepsy) to evaluate the relevance of

![Figure 3](image-url)
networks identified from scalp EEG data. Overall results obtained on simulated as well on real data indicated that the combination of the wMNE and the PLV methods leads to the most relevant networks as compared with the ground-truth (simulations) or with the intracerebrally-identified networks (patient data). Results are more specifically discussed hereafter.

**Methodological Considerations**

The connectivity matrices were thresholded by keeping the nodes with highest strength values (strongest 1%). This procedure was used to standardize the comparison between all the combinations. We were aware about the effect of this threshold and we realized the comparative study using different threshold values. All threshold values were found to lead to the same differences between the method (inverse/connectivity) combinations.

In this paper, we have averaged the reconstructed sources within the same region of interest defined by the parcellation process based on Destrieux atlas. This approach was frequently used in the context of M/EEG source connectivity (de Pasquale et al. 2010; Fraschini et al. 2016; Hassan et al. 2015a). However, such an averaging procedure may increase the noise power since its computation is performed over sources that, with some probability, may not exhibit correlation (Brookes et al. 2014) where the need of alternative approaches such as flipping the sign of the sources in each ROIs before averaging the regional time series (Fraschini et al. 2016) or developing methods based on probabilistic maps, a widely approach used in the fMRI-based analysis, for instance.

Although EEG source connectivity reduced the problem of volume conduction as compared with scalp EEG connectivity, it does not yet provide a perfect solution. The volume conduction effect is a challenging issue when performing EEG/MEG inverse solution (Schoffelen and Gross 2009). In the connectivity context, the main effect of the volume conduction is the appearance of ‘artificial’ connections among close sources, a problem often referred

---

**Fig. 4** Source localization versus source connectivity. a the reference network. b Results obtained by the network-based approach (cMEM/MI and wMNE/PLV). c. Results obtained by the localization based approach (cMEM and wMNE) using same window of analysis.

Results were averaged over a 50 ms interval around each of the spike peaks. *Red color represents the sources with the highest energy (Color figure online)*
to as ‘source leakage’. The use of a high spatial resolution (high number of ROIs) may increase this problem. A few approaches have been proposed recently to deal with the source leakage by either normalizing the edges weights by the distance between the nodes or removing the edges between very close sources. Although, these approaches have some advantages, it was shown that, in most cases, they also remove ‘real’ connections (Schoffelen and Gross 2009). In this context, some connectivity methods such as PLV have been shown to reduce the volume conduction (Hipp et al. 2011). This may explain the good performance of this method. Indeed, in the double network scenario,
PLV was able to detect the long-range connections between parietal and temporal networks.

Four inverse/connectivity algorithms were evaluated in this paper. It is worth mentioning that some other inverse algorithms like MUSIC-based and beamforming as well some connectivity measures such as power envelope correlation (O’Neill et al. 2015) were not included in this study. Moreover, we were focusing in this paper on evaluating different families of ‘functional’ connectivity methods regardless the directionality of these connections. Nevertheless, we consider that the analyses of the ‘effective’ connectivity methods that investigate the causality between brain regions may be of interest in the context of epileptic connectivity methods that investigate the causality between brain regions may of interest in the context of epileptic propagation and time delays, is indeed an interesting supplementary feature in the context of epileptic seizure propagation and will be a potential element for further analysis.

The head model used in this study was computed using the Boundary Element Method (BEM) with three layers (skin, skull and brain). This model was widely used in the context of M/EEG source estimation (Fuchs et al. 2007) as a compromise between computational cost and accuracy. Nevertheless, other methods exist to compute the head model such as the Finite Element Method (FEM) or adding other layers such as cerebrospinal fluid (CSF). These methods can possibly have effect of the resultant network (Cho et al. 2015). The evaluation of the above mentioned parameters/factors may be the topic of further investigation.

Identification of Interictal Epileptic Networks from Scalp Dense-EEG Data

A salient feature of epilepsy in general and epileptic networks is the increased synchronization among interconnected neuronal populations distributed over distant areas. This “hyper”-synchronization often leads to an increase of brain connectivity, not only during the transition to seizures but also during interictal periods, as shown in many studies based on intracranial recordings (see (Wendling et al. 2010) for review). In this context, the combination of the M/EEG source imaging with the functional connectivity measures has recently disclosed promising findings to identify pathological brain networks, at the cortical level (Dai et al. 2012; Lu et al. 2012; Malinowska et al. 2014; Song et al. 2013).

However, two factors seem to be crucial for reliable estimation of EEG source connectivity: (i) the number of scalp electrodes and (ii) the combination between the inverse solution algorithm and the functional connectivity measure. Concerning the number of electrodes, it was reported that the increase of the spatial resolution by using dense EEG may dramatically improve the accuracy of the source localization analysis (Michel and Murray 2012; Song et al. 2015). In addition, the use of dense EEG, as compared to classical montages (32 or 64 electrodes), is needed to accurately identify functional brain networks from scalp EEG (Hassan et al. 2014). To overcome this issue, dense-EEG (256 electrodes) recordings were used in this study. The main feature of this system is the excellent coverage of the subject’s head by surface electrodes allowing for improved reconstruction of the cortical activity from non-invasive scalp measurements, as compared with more standard 32-128 electrode systems (Song et al. 2015). Regarding the combination of inverse/connectivity methods, most of reported studies have empirically selected a combination while this selection was shown to have a dramatic impacts on results, in term of identified network topology (Hassan et al. 2014). The present analysis brings further confirmation of this high...
variability observed when different inverse solutions and/or connectivity measures are being used in the pipeline leading to cortical networks from EEG signals.

A major advantage of the EEG source connectivity approach as presented here is that reconstructed sources and associated networks were identified for the whole brain. Then graph-based metrics (strength values) were computed to characterize the networks and the similarity index was used to compare the networks obtained from various method combinations. In addition, functional connectivity was applied directly to the reconstructed signals and not on derived components. In this regard, this study differs from (Malinowska et al. 2014) where connectivity was estimated on signals components obtained by ICA decomposition. Although the methodological issue of measuring connectivity between independent components still holds, a future interesting study will compare the results obtained from the ICA-based approach to those reported here from source connectivity.

**EEG Source Localization Versus Functional Connectivity**

Source localization methods have been widely applied to interictal epileptic spikes (Becker et al. 2014). The goal of these approaches is the localization of brain generators of epileptic activity from scalp recordings. A fundamental question that is addressed in this paper relates to the difference between the source connectivity and the source localization approach. This study indicated that the information extracted from dense-EEG recordings in both cases can differ dramatically. First, the connectivity is an additional step to the simple source reconstruction/localization. Second, the fundamental difference between both methods is that the source localization ignores all possible communications between brain regions. When performing source localization analysis, the sources with highest amplitude (averaged at given time period or computed at the instant of peak amplitude of the signal) are classically kept. However, to some extent (depending on threshold), this approach may neglect the possible contribution of “low energy” sources participating into the network activity.

Conversely, the hypothesis behind the network-based approach (typically when phase synchronization methods are used as connectivity measure) is that sources can be synchronized regardless their amplitude. To this extent, we believe that the network-based approach allows for revealing networks that are more specific to epileptic networks, as hyper-synchronization phenomena constitute a typical hallmark of such networks. An illustrative example in this paper is the poor involvements of the temporal lobe region when the sole source localization approach (in the case of cMEM) was applied while both parietal and temporal networks (as a priori introduced in the EEG generation model) are retrieved by the connectivity-based approach (cMEM/ $h^2$). Note that we have averaged the source localization results in a time window of 50 ms to cover the time delay of 30 ms set in the model between the two brain regions. Different time window were used to avoid the effect of the window length. All tested windows (30 ms, 40 ms, 50 ms and 60 ms) showed similar observations i.e. the absence of the temporal sources (not shown here). The fact that epilepsy is considered as a network disease can explain the low performance of some of the inverse methods as these methods were originally developed to localize ‘local’ epileptic foci characterized by high-energy sources regardless the interrelationships between brain regions. Our results show that EEG source connectivity methods are more suited in the case of multi-focal epileptic zone. More generally, they support the recent tendency in brain disorder research which is the necessity to move from localizing ‘pathological areas’ to identifying ‘altered networks’ (Diessen et al. 2013; Fornito et al. 2015).

**Epilepsy as a Network Disorder**

There is increasing evidence that epileptic activity involves brain networks rather than a single well circumscribed region and that these dysfunctional networks contribute to both ictal and interictal activity (Bourien et al. 2004, 2005; Coito et al. 2015; Engel Jr et al. 2013; Hipp et al. 2011). Functional connectivity was widely applied to depth-EEG data to predict seizures (Mormann et al. 2000) and identify epileptic networks in partial epilepsies (Bartolomei et al. 2001). These studies showed alterations of synchronization in brain networks during interictal to ictal transition (Wendling et al. 2003) as well as during seizures (Diessen et al. 2013; Jiruska et al. 2013; Schindler et al. 2008). Most of these studies were performed using invasively-recorded data in patient’s candidate to surgery. Interestingly, our results show that pathological networks involved during epileptiform activity can also be identified from scalp EEG.

Indeed, we have evaluated the performance of a relatively new approach aiming at identifying epileptic brain networks from scalp EEG. The application of the method on real data showed the good performance of this method in term of network identified from scalp EEG as compared with those identified from intracerebral EEG. Note that the comparison was done only by computing $h^2$ between the intracerebral signals based on a large number of studies showing that $h^2$ is one of the most adapted metrics to compute functional connectivity between intracerebral recordings (Bettus et al. 2008; Wendling et al. 2010). Although it is obviously difficult to conclude on a single patient analysis, results showed good matching between...
scalp-EEG based networks and both the depth-EEG based networks and the expert judgment. Therefore, future work will consist in the application of the EEG source connectivity on a big database of real dense EEG data recorded from epileptic patients. In these patients candidate to surgery, we plan to use also intracerebral EEG signals as a reference to validate the identified networks. In addition, due to the excellent temporal resolution of the EEG, the dynamic behaviors of the epileptic networks will be also explored (Hassan et al. 2015a; Mheich et al. 2015b).

Finally, the capacity to describe epileptic activity not only according to the sites where epileptiform activity is generated but also according to the abnormal functional relationships between these sites can definitively improve the surgical approach. We speculate that in order to better understand and ultimately prevent seizures, it is essential to identify and then remove/disconnect pathological nodes of the network (exhibiting abnormal hyper-synchronization capability). The proposed method contributes to this aim and reported results constitute a first step toward the development of more efficient non-invasive diagnostic methods for clinical epileptology.

Acknowledgments This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the “Investing for the Future” program under reference ANR-10-LABX-07-01. It was also financed by the Rennes University Hospital (COREC Project named conneXion, 2012-14).

References


Ramon C, Holmes MD (2013) Noninvasive localization of epileptic sites from stable phase synchronization patterns on different days derived from short duration interictal scalp EEG. Brain Topogr 26:1–8


Schoffelen JM, Gross J (2009) Source connectivity analysis with MEG and EEG. Hum Brain Mapp 30:1857–1865


CHAPTER 5- DISCUSSION AND FUTURE WORK

There is an emerging consensus that the human brain is a complex network of structurally and functionally interconnected regions (Sporns et al., 2004, Achard et al., 2006, Stam and Reijneveld, 2007, Hagmann et al., 2008b, van den Heuvel et al., 2008, Buckner et al., 2009, Honey et al., 2009, van den Heuvel et al., 2009, Meunier et al., 2010b, van den Heuvel and Sporns, 2011). Common topological features of structural and functional brain networks such as high clustering as well as a number of highly connected network hubs ‘rich club’ have been revealed at micro, meso and macro scale, see (Sporns, 2016) for very recent review. Inside the brain network, information is continuously processed and integrated between specified, spatially spread but functionally interconnected brain regions with strong temporal dynamics (Sporns et al., 2000).

The information processing in the human brain is however dynamic and occurs at sub-second time scale (Vanrullen and Thorpe, 2001, Blinowska et al., 2010, Hutchison et al., 2013, Kucyi et al., 2016). Therefore, neuroimaging techniques with excellent time resolution are required combined with advanced data analysis techniques capable of identifying and tracking the dynamics of these functional brain networks. A new method called ‘dense-EEG source connectivity’ was recently developed in our team benefiting from the high spatial resolution (256 channels) and the excellent time resolution (1ms) of the EEG system (Hassan et al., 2015a). This method allowed the identification of the brain networks at the level of the cortex from scalp EEG. Therefore, this method appeared very suitable to track the fast information processing in the human brain. In this thesis, we developed an algorithm that is able to track the dynamics of the brain functional networks at appropriate time scale. The proposed algorithm is based on $K$-means clustering approach. This new method allowed us to summarize the neural networks in a limited number of dominant “stable” networks activating over short time periods during the performed cognitive task. Information processing in the brain is stimulus-dependent and functional networks can reconfigure differently when changing the external stimuli (visual/auditory) or when neurological disorders appear (alterations in networks associated with transition from normal to pathological state (Tijms et al., 2013, Fornito et al., 2016)). Therefore, methods able to quantitatively compare networks, in the specific context of the brain where topology is a crucial issue, are still missing (Van Wijk et al., 2010, Sporns, 2011, Fornito et al., 2016). In this thesis, we have developed a new algorithm called ‘SimNet’ for measuring similarity between graphs in which the coordinates of nodes are known. The main feature of SimNet is to take into account the spatial location of nodes in order to estimate a similarity
index between the compared graphs. Results obtained from the application of both proposed and developed algorithms are discussed hereafter.

5.1 DYNAMIC RECONFIGURATION OF FUNCTIONAL BRAIN NETWORKS

Tracking the brain network variations across time is an issue of great interest in cognitive research and in neurological diseases. These variations include slow changes across the lifetime (due to long-term processes like learning or aging) as well as rapid spontaneous fluctuations and fast evoked reconfigurations. A key challenge is to temporally follow, over very short time duration, the sequence of brain processes involved in a cognitive task.

Several approaches were developed to track the dynamic behavior of brain networks. These approaches are either based on characterizing the correlation changes across time (Sakoğlu et al., 2010, Allen et al., 2012, O’Neill et al., 2016) or they are based on characterizing the brain connectivity by several graph metrics rather the connectivity values themselves (Calhoun et al., 2014) (van den Heuvel et al., 2012, Braun et al., 2015, Sporns and Betzel, 2016). For instance, Allen and colleagues (Allen et al., 2012) proposed an algorithm to track brain networks during resting state. The algorithm was based on the independent component analysis and K-means clustering. However it was adapted to fMRI data by taking a large time window (about 3 s) to compute the functional connectivity and therefore does not allow tracking cognitive tasks with short time duration. Therefore, it is likely that accurate tracking of the picture naming could not be achieved with the above-mentioned proposed algorithm. In (Murray et al., 2008a), authors proposed an algorithm based on the amplitude of event related potentials to follow time-varying voltage topographic maps. However, this algorithm does not account for brain connectivity quantified directly from scalp signals (electrode space) or indirectly from reconstructed brain sources (source space).

In this thesis, a novel algorithm was presented to track the dynamics of functional connectivity. The proposed algorithm is based on the clustering of connectivity matrices computed from the phase locking value method applied to networks computed at the scalp and cortical sources level. The algorithm was applied to dense-EEG data recorded during a picture naming task (<1second from picture representation to the reaction time). By taking advantage of the excellent temporal resolution of the EEG signals, this algorithm showed high performance to reveal the stability of the brain networks over short periods of time and to segment the cognitive process into a set of functional connectivity states.

Although the K-means algorithm is popular (simple to use) and very fast (can converge in a short period), it suffers from two main difficulties. Firstly, the clustering achieved by the algorithm is highly dependent on the initial values of the cluster centers. Secondly K-means is very sensitive to outliers. In this thesis, these difficulties were considered by adding some constraints on the initial selection of the cluster centers, and by cleaning the data before
analysis. The initial values of the cluster centers were selected with the restriction of rejecting if the time interval between them is less than 30 ms. In further work, other types of clustering algorithms could be also used instead of k-means such as density based or k-nearest neighbor’s methods. Also, dictionary-based approaches could be used such as algorithm recently proposed by Farrugia et al. (Farrugia et al., 2016).

5.2 BRAIN NETWORK SIMILARITY

Generally speaking, in our research field, another challenge is related to quantify the similarity between two brain networks. The problem of comparing networks has not received much attention in brain research literature (Shimada et al., 2016). This does not mean that graph analysis has not been used to compare network topologies (Fornito et al., 2016). On the contrary, it has gained a lot of interest in recent years and is increasingly being applied in both functional and structural studies (Van Wijk et al., 2010).

Regarding the literature of comparing between brains networks, most of the studies used the graph metrics in order to find significant differences between two compared groups such as healthy and pathological groups (Liu et al., 2008, Alexander-Bloch et al., 2010, Meunier et al., 2010a, Becerril et al., 2011) or between real brain networks and random or surrogate networks (Stam et al., 2007, Gong et al., 2009, Stam et al., 2009). However, these approaches usually ignore the physical location of nodes and do not take into account the spatial difference between the compared graphs. Recently, the physical location has been considered in new metric for computing graph measures in brain networks analysis (Pineda-Pardo et al., 2015). Authors showed that the new metric could distinguish the global connectivity of structural networks from functional networks only when the physical locations of nodes are considered.

Another approach to compare brain networks was developed by Zalesky et al. (Zalesky et al., 2010) and called network-based statistic (NBS). This method was aimed to find a network “pattern” (a set of nodes connected by edges) that differentiates two sets of networks. The problem of this method is the difficulty to define the threshold chosen to keep or remove links. If the threshold is chosen too low, large components can arise in the permuted data as a matter of chance and thereby reduce power. In contrast, if the threshold is set too high, connections comprising the effect of interest may not be admitted to the set of supra-threshold links.

Here, we proposed a new algorithm, called SimNet, for quantifying brain networks similarity. SimNet is based on two main steps: i) nodes distance, in this part the algorithm includes three operations: insertion, deletion and substitution of nodes and ii) edges distance, it consists of calculating the sum of the weight difference between two edges of two compared graphs. The main advantage of this algorithm is that it takes into account the spatial location (3D
coordinates) of the nodes when comparing two networks, in contrast with other methods based on the sole statistical properties of compared graphs. The algorithm provides a normalized Similarity Index (SI): 0 for no similarity and 1 for two identical networks (same properties and topology). The application of the algorithm on simulated and real data shows its high ability to detect the (even weak) spatial variation of brain networks. The key parameter to be tuned in SimNet is the radius of the disk used to detect the neighbors of a given node. An increase of the radius $R$ will automatically lead to an increase of the similarity index between the two compared graphs. In the real case, $R$ was chosen as the average distance between all nodes.

**5.3 INFORMATION PROCESSING IN THE HUMAN BRAIN**

Many attempts have been made to understand the way in which the brain networks self-organize to process information. The concept of ‘engram’ (physical trace of a memory) has been recently reconsidered (Josselyn et al., 2015). Another similar concept has emerged also called ‘communication through coherence’ (Deco and Kringelbach, 2016). All these concepts share the same idea, which consists of considering neuronal assemblies (clusters of neurons) as the key unit of information processing. As described in the introduction of this manuscript, the starting point of the thesis was the mental information theory proposed by Berrou et al. (Berrou and Gripon, 2012). According to this theory, information coding/decoding in the brain involves “fully connected” networks of neurons or neuronal assemblies, referred to as “neural cliques” and “neural meta-cliques”, respectively. It is considered as an essential feature in the encoding and retrieval of the elements of information (“infons”) stored in the brain.

In relation to the information processing in the human brain, here, we showed that the cognitive functions are dynamic at sub-second scale and can be segmented into several functional connectivity states from the visual processing to the articulation, for instance. During these states, the activated brain regions involved in the processing change and this leads to a spatial variation in the brain networks during the processing. Also, we detected significant differences between networks identified during the time period associated with ‘visual processing’ and those related to the ‘access to memory’ (Chapter 4-study 2) periods. During visual processing, the networks were mainly occipital involving the inferior occipital, the lateral occipito-temporal sulcus and occipital pole. This period was shown to be related to the visual feature extraction preceding the object category recognition (Vanrullen and Thorpe, 2001). For the second period, results showed a network involving the occipital regions with an implication of the bilateral inferior temporal sulcus. This network is known to be related to semantic working memory.
system when someone tries to remind the name of the presented object (Martin and Chao, 2001). This period was considered as the first instant of categorization in the human brain. Even if these results did not validate the presence of neural cliques (fully connected networks) during information processing as defined in the mental information theory, the validation of brain networks associated with semantic content of visual stimuli as reported in the thesis (animals vs. tools) is directly related to the way our brain encode information as described in the theory.

5.4 METHODOLOGICAL CONSIDERATIONS

5.4.1 FUNCTIONAL CONNECTIVITY

Several metrics have been proposed to estimate the statistical couplings between time series. These metrics can be regrouped into two different categories, namely “linear methods” (such as linear correlation coefficient, coherence and imaginary part of the coherence -ImCoh-) and nonlinear methods (such as phase synchronization -PLV and PLI- and mutual information) (Na et al., 2002, Fingelkurts et al., 2005, Ansari-Asl et al., 2006, Blinowska, 2011), see (Pereda et al., 2005, Friston, 2011, Sakkalis, 2011) for reviews and (Wendling et al., 2009) for model-based evaluation of the aforementioned measures. Each group of methods has its advantages and limitations as reported in comparative studies performed under different conditions. On simulated data (Ansari-Asl et al., 2006), showed that amplitude-based correlation has very good performance. Recently, Colclough et al. (Colclough et al., 2016) tested the reliability of twelve methods during MEG resting-state recordings. Results showed that the most consistent methods for stationary connectivity estimation are those corrected for spatial leakage such as amplitude envelope correlation and partial correlation measures. In another comparative study, Finger et al (Finger et al., 2016) used several functional connectivity metrics such as coherence, PLV, PLI and ImCoh to assess the relationship between structural and functional coupling in population of healthy subjects. Result reported that PLV showed the best matching between simulations and empirical data and that zero-lag correlation are very crucial to assess the structural/functional relationships.

Here and based on the comparative study reported in (Hassan 2014), we used the PLV method. Phase locking values can be calculated by two ways, either by estimating the phase synchronization across a set of trials or for a single trial (sPLV). Both approaches aim at quantifying the stability of the phase-difference between two signals in a pre-defined frequency range. PLV aims to detect the stability of phase over a defined number of trials while sPLV aims to detect the stability of phase within one trial at a defined time epoch of the trial. The use of PLV over trials as used in our case allowed us to preserve the excellent resolution of the EEG
and allowed a finer tracking of the dynamics of brain networks. However, one of the drawbacks of PLV is that it requires a high number of stimuli to be correctly estimated which was carried out in dataset1 with 74 stimuli per session and less in the dataset2 where we had 40 trials per condition.

In addition, in this thesis we only investigated the functional connectivity methods. We believe that effective connectivity could add additional information about the directionality of the functional interactions between brain areas that are present in identified graphs. For instance, using multivariate autoregressive model-based method, Coito et al. showed that effective connectivity analysis is a very powerful approach to investigate how brain regions driver others during epileptic and non-epileptic activity (Coito and Luisa, 2016). Both the segmentation algorithm and the SimNet algorithm could be adaptable to application on such directed networks.

5.4.2 SOURCE LEAKAGE

The source reconstruction of the EEG measurements can create spurious correlations between the time-series of estimated sources. This effect is called “source leakage” or “spatial leakage” as the reconstructions of true point dipole sources from the measured signals will spread over several voxels. Few approaches have been proposed recently to deal with the spatial leakage problem such as removing the edges between the very close sources. For example de Pascale et al (de Pasquale et al., 2010, de Pasquale et al., 2012) removed the connections whose have inter-node distance lower than 35 millimeters. Other studies argue that the anatomical parcellation, which consists of reconstructing the sources on a high spatial resolution cortex mesh and then averaging the dynamics of sources located in the same regions of interest may reduce the spatial leakage effect (de Pasquale et al., 2010, Betti et al., 2013). This can increase the distance between the positions of the regions of interest defined by their barycenter, which can reduce the ‘artificial’ correlations of very close sources located in the same regions of interest. Although these approaches have some advantages, it was shown that, in most cases, they also remove ‘real’ connections (Schoffelen and Gross, 2009). Recently, another approach was developed by Colclough et al (Colclough et al., 2015) for removing the confounding effects of source leakage before estimate the connectivity between brain regions of interest, which may lead to removing real connections (Finger et al., 2016).

It is likely that the number of ROIs play a role in this source leakage problem. Here, we used 68 or 148 anatomical ROIs to define nodes in the brain network. In fact, there is no clear consensus about how to select the appropriate number of nodes that represent the large-scale networks. On
the first hand, choosing finer segmentation may increase the spatial resolution. On the other hand, keeping a reduced number of ROIs helps to remove the spurious links that occur between spatially adjacent sources. In this regard, we assume that 68 or 148 regions were sufficient to investigate the global characteristics of the large-scale human brain while minimizing the problem of spurious connections between ‘very close sources’ provided by the source leakage. In addition, the zero-lag correlation, removed by methods developed to deal with the spatial leakage, was shown to have high importance in functional connectivity analysis. Results on simulated and real data have reported that PLV method, used in the thesis, has higher performance than methods that correct for spatial leakage (Finger et al., 2016).

5.4.3 THRESHOLDING THE CONNECTIVITY MATRIX

There is no consensus about the way of thresholding the functional connectivity matrix that usually provide spurious connections (Sporns, 2011, Fornito et al., 2016). Many strategies were developed to address this issue based on global and local thresholding such threshold based on the weight, density and normalization to reference networks (Rubinov et al., 2009, Lynall et al., 2010, Van Wijk et al., 2010, Stam et al., 2014). The simplest approach is to apply a global threshold over all values of the connectivity matrix where the elements below the threshold are set to zero. This threshold can be determined using weight-based or density-based methods. The weight-based method consists on choosing a threshold value $T$. The $T$ value can be chosen arbitrarily and in this case all values below this value $T$ will be set to zero in the connectivity matrix. Other methods are based on controlling the false discovery rate (FDR) used in many recent brain network studies (Achard et al., 2006, Bassett et al., 2006). In this thesis, we used density-based thresholding method as this threshold was shown to be more adequate than other threshold methods when comparing between groups (Garrison et al., 2015), i.e networks of different groups will have the same density. It consisted of keeping the highest 10% of edges. The main advantage of this method is to preserve the same number of links for the compared connectivity matrices. According to previous studies, the 10% threshold provides an optimal trade-off between retaining the true connections and reducing spurious connections (Lord et al., 2012). Nevertheless, the consistency of the results with the threshold values was considered in this thesis.

5.5 FUTURE DIRECTIONS

Recently, the importance of noninvasive functional connectivity analysis has been increasingly considered. This has led to new methodological progress and an increasing number of functional/effective connectivity analyses. Connectivity analysis performed at the level of scalp
surface EEG electrodes (“electrode space”) is problematic due to effects of volume conduction and field spread. In this thesis, this difficulty was overcome by estimating functional connectivity at the level of brain sources (“source space”). Indeed, the dense-EEG source connectivity is an emerging method that addresses these issues by identifying networks at the level of the neocortex from scalp-recorded signals. This progress has led to the possibility of tracking the dynamics of brain networks at a very short time scale, as reported in this thesis. The segmentation algorithm is completely adaptable to be applied to any other cognitive task such as picture spelling, perception and motor task (Simon task for instance (Simon, 1990)). We believe that the developed algorithm will open the door for a wide range of applications that are impossible to access using fMRI due to its poor temporal resolution.

A straightforward application of the proposed algorithms in the thesis is their application to dataset2 to discriminate between two categories of pictures (meaningful vs meaningless). The brain networks of each category can be segmented into functional connectivity states (fcS) using the developed segmentation algorithm. At each fcS, the similarity between networks related to each category can be computed using ‘SimNet’. This could allow us to build a new network where each node represent a brain connectivity network for a defined picture, and the edge represent the similarity scores between the networks associated to each picture (Figure 15). The Louvain modularity maximization algorithm (Blondel et al., 2008) can be applied on the similarity graph in order to detect the modules to separate the two categories. This is one of our ongoing works.

In addition, the results can be also used to make a link between the dynamic characteristics of the networks and the behavioral performance of the participants. For instance, we can explore a relationship (if any) between the functional connectivity states (their number or their spatiotemporal characteristics) and the reaction time of the participants. It is possible that the participants with the fastest reaction time are those with the most dynamic behavior. Other related features have been proposed recently such as the flexibility of the networks (computed over a set of temporally evolving networks) and the learning ability of the participants (Bassett et al., 2011, Bassett et al., 2013, Mantzaris et al., 2013, Bassett et al., 2015). The algorithm can be used also to investigate the possible alterations in the dynamic behavior of the networks associated with pathology. This feature was recently identified for different brain disorders such as disorders of consciences, Alzheimer and epilepsy (Zhao et al., 2012, Crone et al., 2014, Kambhati et al., 2015).
The distance between networks is an open question too. We have proposed the SimNet algorithm that can compute a similarity value between 0 (totally different networks) and 1 (same network). This algorithm can be used in any application where the physical location of the nodes is crucial. In the context of object categorization (partly tackled in the thesis), we speculate that the algorithm could be used in the identification of semantic maps based on the similarity between the networks associated to each trial. Indeed, we can imagine computing a network per image (apple, house, cat…) and computing similarity between all the images and then draw the map of the similarity. This analysis required a much higher number of images and also the repetition of each image several times. When these conditions are respected, these semantic maps could be very useful to understand the object categorization process in the human brain from a network-based analysis. Another possible application of SimNet is the mapping of a ‘disease network’ where the nodes represent each disease and edges may represent the similarity between the different networks associated to each disease (such as Parkinson’s, Alzheimer’s disease, epilepsy…). This application will help us to further understand the possible common altered network patterns in brain disease. Although this analysis would require a large dataset for different brain diseases, it would help us to understand the common topological features between the abnormal brain networks and their distances from the normal brain networks.
REFERENCES


Brazier MA, Casby JU (1952) Crosscorrelation and autocorrelation studies of electroencephalographic potentials. Electroencephalography and clinical neurophysiology 4:201-211.


