4-17-2014 12:00 AM

Optimizing the Analysis of Electroencephalographic Data by Dynamic Graphs

Mehrsasadat Golestaneh
The University of Western Ontario

Supervisor
Mark Joseph Daley
The University of Western Ontario

Graduate Program in Computer Science
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

OPTIMIZING THE ANALYSIS OF ELECTROENCEPHALOGRAPHIC DATA BY DYNAMIC GRAPHS

(Thesis format: Monograph)

by

Mehrsasadat Golestaneh

Graduate Program in Computer Science

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

The brain's underlying functional connectivity has been recently studied using tools offered by graph theory and network theory. Although the primary research focus in this area has so far been mostly on static graphs, the complex and dynamic nature of the brain's underlying mechanism has initiated the usage of dynamic graphs, providing groundwork for time sensitive and finer investigations.

Studying the topological reconfiguration of these dynamic graphs is done by exploiting a pool of graph metrics, which describe the network's characteristics at different scales. However, considering the vast amount of data generated by neuroimaging tools, heavy computation load and limited amount of time and resources, it is vital to refine this pool of metrics to avoid using non-informative and redundant ones.

In this study, we use electroencephalographic (EEG) brain signals, taken from recordings in 5 different experimental conditions, to generate the dynamic graphs by moving a sliding window over the time series. Dynamic graphs are produced under various conditions that are a combination of different window sizes, different numbers of shared time points and various frequency bands. Based on each set of these dynamic graphs, time series of 25 graph metrics, and then their pairwise correlation values are computed. This is done to investigate the metric correlations under various circumstances, and to detect the ones that are always present.

We conclude by suggesting a set of uniquely informative and orthogonal metrics that is convenient to use for further analysis of brain's functional connectivity.

Keywords: Dynamic graphs, Orthogonal metrics, Functional connectivity, EEG

Acknowledgments

First, I thank my supervisor, Mark Daley, for his mentorship and support, without which I would be lost in no time. I also thank my mom, dad, Sara, and Roozbeh, for their love and support, which kept me hopeful and determined all the time.

Contents

Abstract ii

List of Figures vii

List of Tables ix

List of Appendices ix

1 Introduction and Background 1
- 1.1 Overview and Contributions . . . 3
- 1.2 Background . . . 5
  - 1.2.1 A brief review of graph theory . . . 5
  - 1.2.2 Electroencephalography (EEG) . . . 10
  - 1.2.3 Related pioneering works . . . 11

2 From EEG signals to dynamic graph metrics 14
- 2.1 General information about the data-set . . . 15
  - 2.1.1 The experiment . . . 15
  - 2.1.2 Sampling rate . . . 16
  - 2.1.3 Epoching . . . 16
  - 2.1.4 High pass and low pass analog filters . . . 17
  - 2.1.5 Dealing with power line interference . . . 17
- 2.2 Further preprocessing of EEG recordings . . . 17
  - 2.2.1 Extracting certain frequencies . . . 18
  - 2.2.2 Averaging . . . 19
- 2.3 Building the brain graphs . . . 19
  - 2.3.1 The concept of dynamic graphs . . . 20
  - 2.3.2 Phase lag index (PLI) . . . 20
  - 2.3.3 Thresholding Methods . . . 22
    - S thresholding method . . . 22
    - False discovery rate (FDR) thresholding method . . . 23
    - Random matrix theory (RMT) thresholding method . . . 25
- 2.4 Graph metrics . . . 26
  - 2.4.1 Micro scale metrics . . . 28
    - Node degree . . . 28
    - Average nearest neighbor degree . . . 28

Closeness centrality . . . 28
Betweenness centrality . . . 28
Eigenvector centrality . . . 29
PageRank . . . 29
Kleinbergs Hub and Authority Scores . . . 30
Core number . . . 30
Closeness Vitality . . . 30
2.4.2 Macro scale metrics . . . 31
Diameter . . . 31
Average path length . . . 31
Average clustering coefficient . . . 31
Number of isolated nodes . . . 32
Edge connectivity . . . 32
Assortativity . . . 32
Number of maximal cliques . . . 32
Graph clique number . . . 32
2.4.3 Motifs . . . 33
2.5 Pearson correlation coefficient . . . 33
2.6 A summary of the pipeline . . . 35
2.6.1 Technical details . . . 37

3 Metric correlations across different thresholding methods 39
3.1 Trial type: 1 . . . 41
3.2 Trial type: 2 . . . 45
3.3 Trial type: 3 . . . 47
3.4 Trial type: 4 . . . 49
3.5 Trial type: 5 . . . 53
3.6 Discussion . . . 54

4 Metric correlations across different circumstances 56
4.1 Changing the window size . . . 57
4.1.1 Trial type 1 . . . 57
4.1.2 Trial type 2 . . . 61
4.1.3 Trial type 3 . . . 63
4.1.4 Trial type 4 . . . 65
4.1.5 Trial type 5 . . . 65
4.1.6 Recurrent dependencies . . . 68
4.2 Changing the step size . . . 69
4.2.1 Trial type 1 . . . 71
4.2.2 Trial type 2 . . . 74
4.2.3 Trial type 3 . . . 76
4.2.4 Trial type 4 . . . 78
4.2.5 Trial type 5 . . . 80
4.2.6 Recurrent dependencies . . . 80
4.3 Changing the frequency band . . . 83

4.3.1 Trial type 1 85
4.3.2 Trial type 2 86
4.3.3 Trial type 3 88
4.3.4 Trial type 4 88
4.3.5 Trial type 5 91
4.3.6 Recurrent dependencies 93
4.4 Discussion 93

5 Summary and suggestions for future work **99**

Bibliography **102**

A Glossary of graph metrics **108**

Curriculum Vitae **110**

List of Figures

1.1 Different states of a moving subject in time. 4
1.2 An example of Un-weighted (a), Weighted (b) and Directed (c) graphs 7
1.3 An example of regular, small-world and random networks[60] 9

2.1 Seven equivalence classes of motifs with up to 4 nodes. 34

3.1 Metric correlations obtained based on trial type 1, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT. . . . 42
3.2 Metric correlations obtained based on trial type 2, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT. . . . 46
3.3 Metric correlations obtained based on trial type 3, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT. . . . 48
3.4 Metric correlations obtained based on trial type 4, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT. . . . 50
3.5 Metric correlations obtained based on trial type 5, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT. . . . 52

4.1 Metric correlations obtained based on trial type 1, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method. 58
4.2 Metric correlations obtained based on trial type 2, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method. 62
4.3 Metric correlations obtained based on trial type 3, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method. 64
4.4 Metric correlations obtained based on trial type 4, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method. 66
4.5 Metric correlations obtained based on trial type 5, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method. 67
4.6 Recurrent metric correlations which were seen for all trial types and window sizes. 70
4.7 Metric correlations obtained based on trial type 1, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method. 72

4.8 Metric correlations obtained based on trial type 2, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method. 75
4.9 Metric correlations obtained based on trial type 3, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method. 77
4.10 Metric correlations obtained based on trial type 4, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method. 79
4.11 Metric correlations obtained based on trial type 5, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method. 81
4.12 Recurrent metric correlations which were seen for all trial types and step sizes. . 82
4.13 Metric correlations obtained based on trial type 1, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method. 84
4.14 Metric correlations obtained based on trial type 2, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method. 87
4.15 Metric correlations obtained based on trial type 3, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method. 89
4.16 Metric correlations obtained based on trial type 4, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method. 90
4.17 Metric correlations obtained based on trial type 5, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method. 92
4.18 Recurrent metric correlations seen for all frequency bands and trial types. . . . 94
4.19 Recurrent metric correaltions seen for all combinations of frequency bands, step sizes, windows sizes and trial types. 95

List of Tables

A.1 Glossary of graph metrics 109

Chapter 1

Introduction and Background

The brain is a highly dynamic and complex system, and the very first step towards improving, maintaining and repairing any system is having a comprehensive, clear understanding of its underlying mechanisms. For this purpose, the brain has been studied on many different scales, from studying the genetic factors affecting the formation of its neuronal connections to analyzing different levels of connections between its anatomical areas.

With the emergence of the disconnection syndromes concept in the late nineteen century, scientific attention was drawn to the notion of networks in neurology [3]. Thereafter, many works have provided evidence for the brain being a highly complex network, which segregates and integrates information in a cost effective and dynamic manner by making use of its highly modular structure. Considering this complexity and dynamicity, network theory can serve as a powerful tool, providing scale invariant metrics, adjusting quickly to the behavioral changes in the brain. Furthermore, it has been proven that structural and functional network metrics are heritable, and that they change with aging [3], making graph metric analysis even more worthwhile.

Since the realization of the benefits of using network theory and graph theory in neuroscience,

many studies, such as [36] and [19], have been done by mapping neuroimaging-derived data to graphs, and analyzing the functional or structural connections by the help of informative graph metrics. Despite the evidently dynamic nature of the brain, most of the work in this area focuses on using static graphs, leaving dynamic graphs out of the center of attention they truly deserve. Only recently, the first steps towards the use of dynamic graphs have been taken by [30] and [38], in which they compared network topologies across a small set of time points, leaving the need of further investigations of the topological evolution of graphs in time at a much finer scale.

Despite the huge role of graph metrics in describing behaviors of the networks under study, using non-informative and repetitive metrics can only lead to duplicated results and waste of time and other limited resources. Putting time and effort to find unique and meaningful metrics is a matter of significant importance, especially considering the possibility of using them as features for training machine learning algorithms, aimed at diagnosing mental disorders (see [25] and [33]). It is worth mentioning that using a lot of redundant or highly correlated metrics can seriously decrease the accuracy of classification[28].

In this project, we have chosen to focus our attention on analysis of dynamic graphs, created using brain signals that are recorded by Electroencephalography(EEG), which is a neuroimaging technique that will be explained later in this chapter. The graphs are obtained in various conditions, such as different window sizes, time epochs, thresholds and bands. Based on these dynamic graphs, series of metrics, reflecting the changes in the brain's functional network over time, are compared with each other to grasp a better understanding of the brain's behavior in different conditions, with the goal of reconnoitering the most informative and unique metrics for further analysis.

1.1 Overview and Contributions

As mentioned before, this study focuses on the analysis of the topological properties of dynamic EEG-derived graphs representing the neural patterns of brain functional connectivity. EEG signals of 5 types of trials are obtained, pre-processed and filtered into different frequency bands. For each trial type-band combination, we move a sliding window over the data and use phase lag index (PLI) to obtain a correlation matrix for each time-window. PLI is a technique that helps with the detection of true synchronization between two time series, and is explained in detail in Chapter 2. The similarity matrix obtained based on PLI reflects the strength of relations between different EEG channels. It is then thresholded using a Random Matrix Theory based approach [16], resulting in a binary adjacency matrix, in which the value of 1 means that activities of the corresponding channels are synchronized, and 0 means there is not enough synchronization.

Following this approach, a series of graphs is obtained, in which each graph captures the behavior of the network in a specific time interval. It is like putting a camera on burst mode and taking successive pictures of a moving subject, so that each picture captures the state of the subject at a certain time (See Figure 1.1). Having these graphs, we compute graph metrics of three distinguished categories: 1- Macro scale metrics (whole graph metrics), describing the state of network based on the whole graph characteristics. 2- Micro scale metrics (node metrics), describing the state of network based on the node properties. 3- Graph motifs.

For each metric belonging to one of these categories, a series of values is obtained, in which each value belongs to one of the captured graphs in time. By comparing these series to each other we can gather information on the metric's uniqueness. Also, since each trial's data was obtained based on a special brain task, we may be able to identify the metrics that reflect these differences, based on the pattern of the changes in their values in comparison to other metrics.

Figure 1.1: Different states of a moving subject in time.

Our overall contributions in this project are as follows:

1- Comparing and analyzing metric correlations of dynamic graphs obtained based on three different thresholding methods, which resulted in an evaluation of using these thresholding methods in this project.

2- Reducing a pool of 25 metrics to a set of 10 uniquely informative metrics, which are convenient to use for further functional network analysis based on dynamic graphs.

1.2 Background

The usage of network theory concepts has expedited the progress in various fields (e.g. [8], [66] and [40]), and since the brain is a highly complex modular network, it comes as no surprise that there is a vast and diverse usage of network theory in neuroscience as well. Furthermore, its applicability to any scale makes methodological cross validation and comparing structural and functional properties of different brain scales possible. As mentioned before, in this project we only focus on the analysis of the dynamic EEG-derived graphs, thus, it is beneficial to give a very brief background on EEG and some basic concepts of graph theory (being the basis for network theory). Afterwards we will briefly cover some of the most important progress in relation to our current study.

1.2.1 A brief review of graph theory

Graph theory is the study of relations between entities. Considering a popular social network like Facebook, every person who has an account in this network can be called a node (belonging to a set of nodes or V) and two people (nodes) are linked to each other by an edge (belonging to a set of edges or E) only if they are friends. Simple as that, one can build a graph G= (V, E) reflecting the relationship patterns in this social network. In the context of functional

brain networks, each node represents an area in the brain and edges are statistical measures of association.

Depending on the needs of analyzing a particular network, graphs can be built as weighted or un-weighted. In weighted graphs (Figure 1.2 b), a value is assigned to each edge. This value can show the cost, degree of importance or any other aspect of the relationship between two nodes. If there is no value assigned to edges, the graph is called un-weighted (Figure 1.2 a). Brain networks have been studied based on both weighted and un-weighted graphs.

A graph is called directed if the directions of relationships between its nodes are annotated with arrows on its edges (Figure 1.2 c), otherwise it is undirected. Considering current methodological tools, it is often more convenient to use undirected graphs to study structural and functional brain connectivity since estimating directionality is harder than determining whether a connection exists or not [47].

Since a graph is a complex representation, it is necessary to summarize its characteristics by some means. For this purpose, various graph metrics have been defined in the field of graph theory. Back to the Facebook example, to compute the number of friends for each person, one must count the edges that connect that person to other people. The result is called node degree. Path length metric answers the question of ”what is the minimum number of people that connect two non-friend Facebook users to each other?”, while the clustering coefficient is a measure of the extent to which nodes tend to cluster together.

Graphs can be classified based on their topological properties. A graph is called random if edges are randomly assigned to its nodes. The Erdos−Renyi graph is a well-known random graph in which all edges have the same probability of occurrence. Random networks are proven to have low clustering coefficient and short path length, which makes sense since when build-

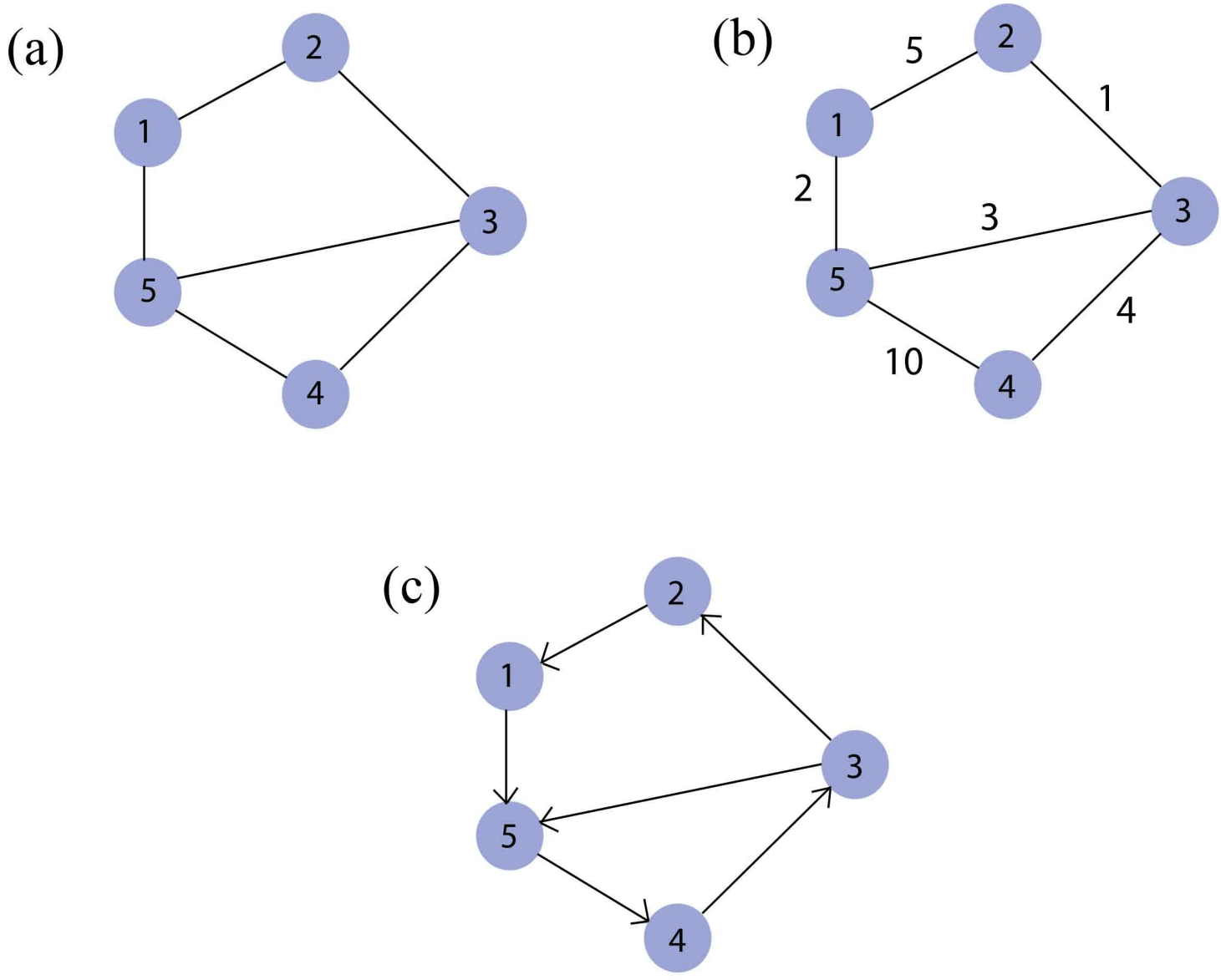

Figure 1.2: An example of Un-weighted (a), Weighted (b) and Directed (c) graphs

ing a random graph there is always the chance of choosing long distance edges connecting two distant nodes, and thus, reducing the average path length in the graph.

On the other hand, if all the nodes in a graph have the same degree, the graph is called regular. Regular networks (regular lattices) are known to have high clustering coefficient and high average path length [60].

In other words, considering the range of all possible variations of graphs, random and regular graphs are located at two ends of this range, representing the most unordered and the most ordered graphs respectively. Based on experimental observations, networks in real life, such as social, biological and neural networks, are usually neither completely random nor utterly ordered. The term *small-world* was first introduced by Watts et. al (1998) [60] to describe the common properties of these intermediate networks. Starting from a regular graph, reconnecting each edge with the probability of $0 < P < 1$, results in a graph with two main properties: high clustering coefficient (in comparison with random graphs) and relatively short path length (comparing to regular graphs). A good example of such small-world networks is again a social network, in which people belong to a cluster of friends. In this network most people are not directly friends, but are connected via a series of mutual friendships. An example of regular, random and small-world networks, as depicted in [60], can be seen in figure 1.3.

Many studies have shown that brain functional networks have small-world properties, and these network properties are the key factors making the segregation and integration of the information possible [10]. A high clustering coefficient helps segregation and as a result increases the local efficiency of information transfer. On the other hand, low average path length provides long distance links to distant areas of the brain and thus supports integration, which results in the global efficiency of information transfer.

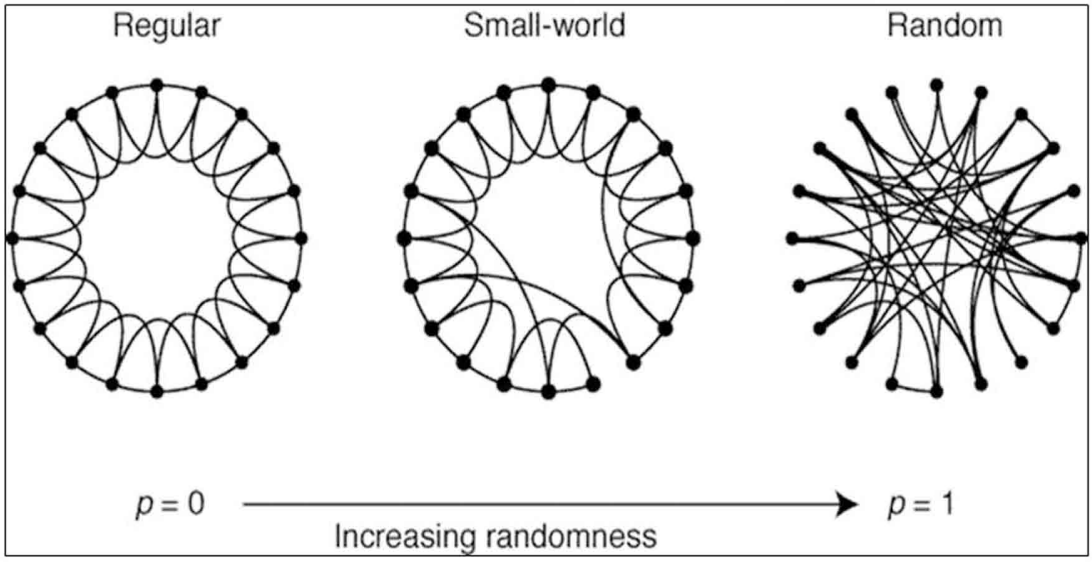

Figure 1.3: An example of regular, small-world and random networks[60]

1.2.2 Electroencephalography (EEG)

Synchronizing groups of cortical neurons in the brain generates a significant amount of electricity, which is caused by voltage fluctuations resulting from ionic current flows within them. EEG electrodes, placed on the scalp in multiple locations, capture this electrical activity over a wide range of frequencies (1 to 100 Hz). In this study we have chosen EEG as the physiological methodology since it provides a good enough time resolution (milliseconds) to reflect the dynamics of graph topology.

On the other hand, working with EEG data has a downside regarding the spatial accuracy that should be dealt with. The easiest way to map the EEG data to a brain graph is to define each EEG source as a node. While this is convenient for preserving between-node covariance information, nodes which represent anatomically nearby EEG sources might mistakenly show high correlations with each other due to an effect called volume conduction. Volume conduction is defined as the fast transmission of electric signals through brain tissue between neighboring sensors. Another similar problem in EEG is caused by an active reference electrode which contributes similar components to signals recorded at different electrodes (for more detail on volume conduction and reference electrode see [39]). Stam et. al (2007) [51] calls these two phenomena the problem of common sources. The problem of common sources can be distinguished from real neural synchronization by making use of methods that take the amount of phase delay into account, such as phase lag index (PLI), knowing that volume conduction happens with near zero phase lag and that the synchronous activity of neural groups does not happen as fast as the transmission of electric fields through tissue (happening approximately at the speed of light) and is accompanied by a non-zero phase delay.

We can avoid the problem of common sources by source reconstruction, in which groups of

anatomical sources are defined and considered as the graph nodes. On the downside, current available source reconstruction methods cause the loss of covariance information between nodes, which is due to the fact that a given electric potential recorded at the scalp can be explained by the activity of infinite different configurations of EEG channels [35].

Considering these options, it seems more convenient to map EEG channels to nodes. Though this approach sacrifices spatial accuracy, it is a good choice for studying dynamic graphs since it keeps the statistical properties of the data as intact as possible. It is worth mentioning that each node represents the activity of a particular region in the brain.

1.2.3 Related pioneering works

The idea of the brain being a network was brought to the scientific community's attention by Norman Geschwind ([23]) in 1965. His investigations on disconnection syndromes (the disconnection of different brain areas in animals and humans) was the onset of many further studies proving that the brain is in fact a complex network, having functionally specialized connected parts that work together to perform certain tasks. Emergence of brain-imaging techniques such as PET, EEG, MEG, MRI and fMRI provided the scientific community with vast amounts of data, increasing the need for usage of proper mathematical tools such as signal processing and correlation estimation methods and more importantly, network theory.

As mentioned before, Strogatz and Watts [60] introduced collective dynamics of a special category of graphs under the name of small-world networks. Benefiting from their work, Sporns et. al (2000) [48] was one of the first studies that took the small-world characteristics into account for investigating the relation between functional and structural networks, proposing the existence of complex brain dynamics that adapt to different task demands. Concurrently, Stephan et. al (2000) [52] also benefited from the small-world properties for analysis of the brain functional connectivity. After these pioneering works, an ample number of studies have

been conducted to further shed light on the underlying network properties of different cognitive and motor control tasks (e.g. [63],[4], [24]), and to investigate the underlying functional network properties of disordered brains (e.g. [50]), using various physiological methodologies.

While good temporal resolution of EEG and its relatively low cost makes it a popular neuroimaging technique, the problem of volume conduction has lead to the usage of various methods such as phase coherence (PC) and imaginary component of coherency (IC), for neat identification of statistical dependencies between physiological time series. Stam et. al (2007) [51] have introduced PLI as an approach to deal with the problem of volume conduction when quantifying phase synchronization. Their results show better performance of PLI comparing to the well known methods of PC and IC.

Though the majority of the work on the brain network topology is based on static graphs, dynamic graph analysis has recently started to attract the neuroscience community's attention, resulting in pioneering works such as [20] that observed the cortical network dynamics during foot movements over several time points, or [2] that took the dynamic time scales into account for studying the modular structures of brain functional networks. Moreover, a more recent study ([30]) investigated the workspace configuration of brain functional network, benefiting from dynamic graph metrics obtained in two different trials of response generation state and working memory stage, while Nichol et. al (2011) [38] observed and analyzed the reconfiguration of brain functional network during an auditory task, using dynamic MEG-derived graphs obtained from different time windows.

This project follows a similar approach to the ones in [30] and [38].

We explain the materials and methods used in this project in detail in Chapter 2, alongside a summary of the pipeline which was implemented and used for obtaining the results. In Chap-

ter 3, we compare 3 different thresholding methods. Metric comparisons are demonstrated in Chapter 4, and Chapter 5 provides the reader with a summary of the project and our final conclusions.

Chapter 2

From EEG signals to dynamic graph metrics

Various steps are taken for obtaining network metric correlations from raw EEG signals. We start by filtering the EEG data-set so that only desired frequencies are kept for future analysis. After that, the signals undergo a process of averaging, so that the signal/noise ratio is increased. Then, the process of building dynamic brain graphs takes place. For building a graph representing the activities of the brain in a certain time window, first a similarity matrix is computed, which demonstrates the strength of synchronization between EEG channels. For building this matrix we use a method called phase lag index (PLI), which takes the problem of common sources (explained in Chapter 1) into account. Afterwards, based on the obtained similarity matrix, we use a thresholding method to create an adjacency matrix, which is a binary representation of the relationships between EEG channels. Having this adjacency matrix, we can then map it to a brain graph in which the vertices represent the channels, and each edge between two vertices is an indicator of an existing synchronization between the two corresponding channels. The above steps are done for consecutive time intervals in EEG signals, and thus result in a series of dynamic graphs. By having this temporal series of brain graphs, we can then compute various metric values and thus obtain time series of these metric values.

These time series are then compared with each other using the Pearson correlation method.

In this chapter we describe all these steps along side other important information and details regarding graph metrics and the underlying structure of algorithms used in this project. We start by introducing our EEG data-set, moving towards explaining the process of filtering and averaging. Then, we focus on the concept of dynamic graphs, PLI method and different types of thresholding methods that are used in this project. Afterwards, we switch to a graph theoretical point of view, introducing the set of metrics exploited for the purpose of our analysis. Finally, after describing the Pearson correlation method, we end this chapter by summarizing the pipeline implemented in this project, which results in a correlation matrix for a set of 25 graph metrics.

2.1 General information about the data-set

In this section, we introduce the data-set used in this project and explain all the necessary concepts for understanding the properties of this data-set.

2.1.1 The experiment

The data set used in this project is obtained via EEG. The experiment based on which this data set was obtained contains 5 trial types:

1- Presenting nothing (the resting state).

2- Presenting non-living auditory stimuli.

3- Presenting non-living visual stimuli.

4- Presenting living auditory stimuli.

5- Presenting living visual stimuli.

It is worth mentioning that the stimuli used for these experiments were either written or spoken words. Meaning that the name of non-living objects, or living beings, were either shown to the subjects as written words (visual stimulus), or were read to them (auditory stimulus).

2.1.2 Sampling rate

Continuous analog signals obtained via EEG are digitized and recorded by computers. This process is called sampling, in which the channels of analog signals are repeatedly sampled at a fixed time interval. The sampling rate is defined as the number of samples recorded per second [55]. The data-set used in this project has the sampling rate of 600 Hz. Also the Nyquist frequency, which is defined as $\frac{1}{2}$ of the sampling rate, is 300 Hz.

2.1.3 Epoching

When presenting a subject with a particular stimulus, the brain processes this stimulus, which can be seen as oscillatory potentials in neuronal groups. This evoked neural activity can be detected in EEG recordings as significant voltage fluctuations or event-related potentials (ERP)[46]. For capturing these important milestones in EEG recordings, the data is cut into several chunks related to the stimuli presentations [55]. This process is called epoching. The epochs of the data-set used in this project contain data from 1 second before representing the stimuli to 2 seconds after the stimuli presentation.

2.1.4 High pass and low pass analog filters

In the process of recording EEG signals, an analog high-pass filter is used to discard very low frequencies originating from bio-electric flowing potentials such as breathing. Also, an analog low-pass filter is used to make the signal band limited, and to discard frequencies that are higher than one half of the sampling rate. The analog signal is digitized and stored in the computer after passing through these analog filters [55]. For the EEG recordings used in this project, cutoffs for the low-pass and high-pass analog filters are respectively 0.5 Hz and 150 Hz.

2.1.5 Dealing with power line interference

EEG signals are often contaminated by a narrow band harmonic signal, with a narrow frequency range around 60 Hz [64] [14]. This unwanted signal can be filtered out using a notch filter. The EEG data-set used in this project is filtered using a 60 Hz notch filter.

2.2 Further preprocessing of EEG recordings

As it was mentioned in Chapter 1, EEG captures rhythmic neuronal activity of the brain in the form of electrical signals. This rhythmic activity is a combination of different frequency bands, which can be categorized in the following ranges: Delta(1-4 Hz), Theta(4.5-7.5 Hz), Alpha(8-16 Hz), Beta(16-32 Hz), Gamma(32-63 Hz) and high Gamma(63-125 Hz). Depending on the state of the brain and type of the task that it is engaged in, neural oscillations can happen in any of these bands, for example, the beta wave is common in normal awake adults while the presence of the delta wave in alert adults is not expected and can be a sign of mental disorders [37]. On the other hand, not all the electrical activities recorded by EEG reflect the oscillations of neuron populations. These unwanted frequencies are considered as noise and

should be eliminated.

In this section we talk about using digital filters for obtaining certain frequency bands, and also removing unwanted noise by making an averaged signal for each of the trial types.

2.2.1 Extracting certain frequencies

A digital band-pass filter takes a signal containing a wide range of frequencies and only passes frequencies within a certain range as out put. For filtering a signal, one way is to convert it from the time domain to the frequency domain, then multiply it by the desired band-pass filter to omit all the unwanted frequencies and finally convert it back to the time domain. On the other hand, based on the convolution theorem, we know that point-wise multiplication in the frequency domain equals convolution in the time domain. Thus, another way of doing this is to convolve our sampled signal by a function representing the fourier transform of the desired filter response. Implementing this convolution can be achieved by a finite impulse response (FIR) filter which is linear, simple and stable. The following equation represents the structure of the FIR filter:

$$y(n) = \sum_{i=0}^{N} b_i x(n-i) \tag{2.1}$$

In which $x(n)$ are the filter input samples, $y(n)$ are the filter output samples and b_i are coefficients of FIR filter frequency response. In other words, the output signal is obtained via convolving the input signal with its impulse response. There are various methods for computing the coefficients of a finite impulse response filter. In this project the window method is used due to the simplicity of its design process.

There are ready made functions in the Python's Scipy.signal library for both calculating the coefficients based on the window method (scipy.signal.firwin) and performing the filtering process using the FIR filter (scipy.signal.lfilter) [29].

2.2.2 Averaging

The averaging process is defined as calculating the mean value for time-points across all recording periods (epochs), aiming to increase the signal/noise ratio (assuming that noise is distributed randomly). This process is absolutely necessary for capturing ERPs from EEG signals since their amplitudes are much smaller than the spontaneous background fluctuations and thus they are not noticeable in raw EEG signals. After averaging the signal over trials of the same type, the spontaneous background fluctuations are averaged out and omitted since they are randomly distributed over the signal. Thus, the remaining activities are in fact ERPs evoked by a stimuli onset, reflecting the patterns of neuronal activity [55].
In this project, averaging is done by using a ready made function in Numpy's library (Numpy.average), which computes the weighted average of the input data.

2.3 Building the brain graphs

As explained in Chapter 1, the goal of this project is to analyze metric correlations obtained based on dynamic graphs. In this way, we can detect possible differences in metric correlations that happen as a result of changing the circumstances under which we observe the functional brain network across time. Thus, it is important to explain the notion of dynamic graphs in a detailed manner.

Furthermore, we explain two key steps that should be taken for obtaining a brain functional network: First, dealing with the problem of common sources (described in Chapter 1), and second, using a thresholding method for building a adjacency matrix.

2.3.1 The concept of dynamic graphs

For observing the changes of a network in time, we use a concept called the time window. The time window with a certain size is slid over the EEG time series, capturing only the activity of the brain in that particular time frame. Thus, we can obtain a set of graphs, in which each graph represents the functional network of the brain based on a certain period of time.

It is important to note that the time windows are not necessarily separate and that they can share time steps. For adjusting the number of time units shared by adjacent time windows, we use a measure called step size. Step size represents the distance between a time window and its neighbor. As the step size increases, fewer time units are shared by neighboring time windows. Thus, by changing the window size and the step size, we can obtain dynamic graphs under various circumstances.

2.3.2 Phase lag index (PLI)

As mentioned in Chapter 1, when building a similarity matrix representing phase synchronizations between time series of different channels in EEG, we might get fake similarities due to the effects caused by the problem of common sources. There are several methods for addressing this problem. One of these methods, which was introduced by [51], is called phase lag index (PLI). PLI is a measure that helps assessing similarities between time series by reflecting the amount of phase lag between them. The idea behind this index is that if the dependency between two time series is caused by the problem of common sources, the phase difference between these two time series would center around 0 mod π. This is acceptable because electric fields travel through brain tissue almost at the speed of light, thus, in this case, the synchronization between two time series is expected to take place without any delay. As it was mentioned in Chapter 1, Stam et. al's results [51] showed that PLI works as well as IC and better than PC,

thus PLI is a suitable choice for tackling the problem of common sources. The PLI index is computed as follows:

$$PLI = |\langle sign[\Delta\Phi(t_k)]\rangle| \qquad k = 1, 2, ..., N \tag{2.2}$$

in which $0 \leq PLI \leq 1$ and $\Delta\Phi(t_k)$ is the time series of phase differences. PLI = 0 means that there is no real synchronization between two time series, while PLI = 1 assures us of a true coupling between two time series, which is not caused by the effect of common sources. This means that the more PLI is close to 1, the more significant the coupling is and vice versa.
The time series of phase differences is computed as follows:
First we obtain the Hilbert transforms of the two desired time series:

$$xa_1 = H(x_1)(t) \tag{2.3}$$

$$xa_2 = H(x_2)(t) \tag{2.4}$$

xa_1 and xa_2 represent the time series and H is the Hilbert transform function. After obtaining the Hilbert transforms of the time series, we can compute their phase for each time point, thus obtaining a time series of phases for each of them:

$$\Phi_1(t_k) = arctan(\frac{Im(xa_1(t_k)}{x_1(t_k)}) \qquad k = 1, 2, ..., N \tag{2.5}$$

$$\Phi_2(t_k) = arctan(\frac{Im(xa_2(t_k)}{x_2(t_k)}) \qquad k = 1, 2, ..., N \tag{2.6}$$

in which $\Phi_1(t_k)$ and $\Phi_2(t_k)$ are time series of phases, $Im(xa_1)$ and $Im(xa_2)$ are the imaginary

parts of the signals at time point t and $x_1(t_k)$ and $x_2(t_k)$ are the real parts of the signals at time point t.

Thus the time series of phase differences is computed as follows:

$$\Delta\Phi(t_k) = \Phi_2(t_k) - \Phi_1(t_k) \qquad k = 1, 2, ..., N \tag{2.7}$$

By obtaining the PLI score for each pair of time series, a similarity matrix is created and its values reflect the strength of the similarities between all possible pairs of EEG channels. Having this similarity matrix, we need to apply an appropriate threshold value to reject all the non-significant similarities, which results in an adjacency matrix, having only binary values, representing the brain's functional network.

2.3.3 Thresholding Methods

Here we explain the three thresholding methods used in this project: The S thresholding method, the false discovery rate (FDR) thresholding method and the random matrix theory (RMT) thresholding method.

S thresholding method

This thresholding method was first introduced and used by [54], for building fMRI networks. This method suggests a threshold based on a ratio called S computed as follows:

$$S = \frac{log(N)}{log(K)} \tag{2.8}$$

in which, N is the number of nodes (vertices) in the network and K is the average node degree. As stated in [61], this ratio (S) is in fact the average path length of a small-world network.
Thus, having the similarity matrix and the desired S value, the following steps are taken to obtain the threshold value:

1- All values of the matrix are sorted in an ascending order.

2- Having the S value and number of vertices (N), the desired average node degree (K) is computed.

3- The number of edges ($|E|$) for a graph with this K value is computed considering that $|E| = (N \times K)/2$.

4- The $|E|$th value is picked from the previously sorted similarity matrix so that $|E|$ pairs of nodes with strongest similarity values are chosen as the edges of the graph.

False discovery rate (FDR) thresholding method

The FDR thresholding method was first introduced by [22], for the purpose of analyzing fMRI imaging data. They suggested a thresholding method based on FDR controlling procedures, which means that the suggested threshold ensures that, on average, the rate of false discoveries will be no more than a specified q ($0 < q < 1$). Here, a false discovery is defined as a falsely detected synchronization between two EEG channels. The detailed procedure of this method is as follows:

1- A desired q is chosen.

2- Having a similarity matrix filled with $r - values$ (measures of similarities between EEG channels), we now compute the corresponding $p - value$ (which shows the statistical significance for the observed similarities) for each member of this correlation matrix. We know that $p - values$ can be computed based on:

$$p(t < T) = I_{\frac{df}{df+t^2}}(\frac{df}{2}, \frac{1}{2}) \tag{2.9}$$

in which df is degrees of freedom and equals to $n - 2$ (here, n = window size). Also, we

know that the sampling distribution of Pearson's correlation coefficient follows Student's t-distribution with degrees of freedom $n - 2$, and that
$t^2 = r^2 \times \frac{df}{1-r^2}$.
I represents the *incomplete beta function*, which is defined as follows:

$$I_x(a, b) = \frac{B(x; a, b)}{B(a, b)} \tag{2.10}$$

in which

$$B(x; a, b) = \int_0^x t^{a-1}(1 - t)^{b-1} dt \tag{2.11}$$

and

$$B(a, b) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)} \tag{2.12}$$

considering that

$$\Gamma(x) = \int_0^\infty y^x e^{-y} \frac{dy}{x} \tag{2.13}$$

3- After computing the $p - values$, they are sorted from smallest to largest:

$$P_{(1)} \leqslant P_{(2)} \leqslant ... \leqslant P_{(V)} \tag{2.14}$$

With V being the number of members of the similarity matrix.

4- If v_i is a member of the similarity matrix corresponding to P_i, and m is the largest i for which

$$P_{(i)} \leqslant \frac{i}{V} q \tag{2.15}$$

Then, the similarity value v_m is chosen as the desired threshold value.

Random matrix theory (RMT) thresholding method

Based on the theory of random matrices, proposed by Wigner [62] and Dyson [18], if we sort the eigenvalues of a random matrix in ascending order, compute the spacings s between adjacent eigenvalues and obtain the distribution $P(s)$ of the spacings, this distribution conforms to the Gaussian Orthogonal Ensemble (GOE).

In the case of similarity matrices obtained in this project, if these spacings follow a GOE distribution, we can conclude that the similarity matrix is full of false similarities. On the other hand, if this distribution follows Poisson statistics, the similarity matrix has strong similarities mostly on its diagonal, and such a matrix is indicative of a very modular system.

Based on this, and considering the fact that brain networks are highly modular organizations [34], [16] proposed a random matrix theory (RMT) thresholding method that is designed to find a threshold at which the resulted eigenvalue spacings distribution goes from following the GOE distribution ($P(d) \approx \frac{1}{2}\pi d e^{\frac{-\pi d^2}{4}}$) to following the Poisson ($P(d) \approx e^{-d}$) statistics. In other words, at this threshold the resulting network goes from being highly dominated by noise to being modular. Having a similarity matrix of order n, and a range of candidate threshold values, the following procedures are taken to obtain the best possible threshold value:

1- We first threshold the matrix with the candidate value.

2- Obtain the ascending ordered list of its eigenvalues ($E_1, ..., E_n$).

3- Perform a spectral unfolding procedure to obtain a distribution with constant eigenvalue density ($e_1, ..., e_n$).

4- Calculate the pair-wise differences between adjacent transformed eigenvalues ($d = e_{i+1} - ei$).

5- Obtain the probability density P(d) of these spacings.

6- Using the Anderson-Darling test, we evaluate the extent that this distribution follows the Poisson distribution and keep the score.

7- If all the thresholds in the proposed range are investigated, we choose the one whose resulting spacing distribution fits the Poission distribution best. Otherwise, we increase the threshold (one step) and go to step 1.

In summary, we start from the lowest threshold value in the suggested range of thresholds, and do the above steps until we find the first threshold value that leads to a Poisson distribution. It is obvious that as we lessen the linear steps between thresholds, and thus, investigate more threshold values in our desired range, the probability of finding the exact transition threshold value, as well as the computation load, increases.

2.4 Graph metrics

After obtaining the adjacency matrices using a thresholding method, we can now map each matrix to a graph. This is done as follows: each EEG channel is considered a node in this graph, and there is an edge between two nodes if and only if the corresponding value in the adjacency matrix is 1.

In this way, a graph is built for each adjacency matrix, yielding a set of dynamic graphs. This set represents the consecutive states of the brain functional network through time, which facilitates the analysis of its changes by the means of graph metrics.

The set of metrics used in this project is composed of 25 metrics, which we found interesting, based on their usage in the related network analysis literature such as [27], [17], [30] and [38].

In this section we introduce and explain these metrics, putting them in one of the following groups:

1- Micro scale metrics

2- Macro scale metrics

3- Motifs

It should be noted that the main goal of comparing time series of these metrics is finding a set of orthogonal and uniquely informative metrics, so that further brain network analysis can be done more efficiently. One of the most important factors that should be considered when choosing this set, is the time necessary for computing each of these metrics. Therefore, in this section we also talk about the notion of time complexity.

Before we start to group and define the graph metrics used in this project, it is necessary to fix the following definitions for future reference:

$|V|$: The number of nodes in the graph.

$|E|$: The number of edges of the graph (which is also a macro scale metric).

d : The average node degree of the graph. d can be computed as follows:

$$d = \frac{1}{|V|} \cdot 2|E| \tag{2.16}$$

As mentioned before, the main purpose of comparing time series of graph metrics in this project is to find an *optimal* set of graph metrics. When choosing the members of this optimal set from a pool of candidate metrics, the computation time plays an important role in choosing a metric over other metrics in the same dependency group. Thus, in addition to defining each metric, we also provide its computational time complexity as an abstract function of the input size.

2.4.1 Micro scale metrics

As mentioned in Chapter 1, these metrics describe the state of the network based on the properties of its nodes.

Node degree

The degree of a node is the number of edges that are connected to that node. The time required for computing the node degree for a whole graph is $O(|E|)$ [15].

Average nearest neighbor degree

The set of nearest neighbors of a node contains all the nodes that are adjacent to this node. Considering this definition, the average nearest neighbor degree of a node is the mean of the degrees of all the members in this set. The computation time for this metric is $O(|V|+|E|)$ [15].

Closeness centrality

The closeness centrality for a node is defined as the inverse of the mean length of the shortest paths between that node and all other nodes in the graph. This metric gives us an estimate of how close a particular node is to all other nodes in the graph. Closeness centrality can be computed for the whole graph in time $O(|V| \cdot |E|)$ [15].

Betweenness centrality

Betweenness centrality of a node is the number of shortest paths that pass through that node. This means that if a node with a high betweenness value is removed, a lot of shortest paths in the graph would become longer. Betweenness centrality can be computed for the whole graph in time $O(|V| \cdot |E|)$ [15].

Eigenvector centrality

Eigenvector centrality (also called Bonacich's centrality) is a measure for centrality that takes the importance of the neighbors of a node into account. In other words, the importance of each node is calculated based on the importance of its neighbors. Consider the node n_i for which we want to compute the eigenvector centrality. This is done as follows:

$$c_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} c_j \tag{2.17}$$

in which c_i is the centrality of the node n_i, A is the adjacency matrix of the graph and λ is a constant.

If we define the centralities of the graph as a vector $\overrightarrow{c} = [c_1, c_2, ...]$, we can rewrite the above sum as a matrix equation:

$$\lambda \overrightarrow{c} = A \cdot \overrightarrow{c} \tag{2.18}$$

By solving this equation we can obtain the centrality. It is clear that in the above equation $\overrightarrow{c}$ is an eigenvector of A and λ is the corresponding eigenvalue, and that A is optimized when λ is maximized. The approximate computation time for this metric is $O(|V|)$ [15], but since the process of finding the best (largest) λ is iterative, it can vary depending on the input graph.

It is also worth mentioning that using this centrality is advantageous from the computation point of view since it can be computed using simple linear algebraic operations, which provides the possibility of a fast parallel computation.

PageRank

PageRank is in fact a variation of the eigenvector centrality and it was first introduced by Page et. al (1998) [9]. This metric measures the importance of each node based on the structure of

its connections.

Kleinbergs Hub and Authority Scores

Kleinbergs Hub and Authority scores were originally defined by Kleinberg et. al (1999) [31] for directed graphs, which have both incoming and outgoing edges. Authorities score gives each node a value based on the number of incoming edges, while hubs score calculates this value based on the number of outgoing edges. In the case of undirected graphs, such as the graphs in this project, these two scores become the same [11]. Thus, we only consider one of them (authorities) in this project.

Also, it is worth mentioning that while PageRank and Kleinberg's scores are very similar in definition, comparing to the PageRank, hub and authority scores are based more on the neighborhood graph of a node than the whole graph [12].

These scores are usually computable in time $O(|V|)$ [15].

Core number

Considering that a k-core is a maximal subgraph that all of its nodes have degree k or more, the core number of a node is defined as the largest k of a k-core, which the node belongs to. In this project, the core number is obtained using an algorithm suggested by [5], which needs time $O(max(|E|, |V|))$ to compute the core number for all nodes of the graph.

Closeness Vitality

Closeness vitality of a node is the change in the sum of shortest path lengths between all nodes after omitting that node from the graph. In other words, it describes the vitality of a node for increasing the global efficiency of the graph.

2.4.2 Macro scale metrics

These metrics are graph-level metrics, meaning that they describe the state of network based on the whole graph characteristics.

Diameter

If we sort all the shortest path lengths between all nodes of a graph, the largest value is called the diameter of the graph and it can be computed in time $O(|V| \cdot |E|)$ [15].

Average path length

The average path length of a graph is obtained via averaging over all the shortest path lengths (between all pairs of nodes) of that graph. The computation time for this metric is $O(|V| \cdot |E|)$ [15].

Average clustering coefficient

If node v has A_v neighbors, then the maximum number of edges that can exist between them is $A_v(A_n - 1)/2$. Thus c_v is defined as the fraction of these possible edges that exist between v and it neighbors. Considering this, the average (global) clustering coefficient of a graph is defined as follows:

$$ACC = \frac{1}{n} \sum_{v \epsilon G} c_v \tag{2.19}$$

Where n is the number of nodes in the graph G.

Average clustering coefficient can be computed in time $O(\frac{|E|^2}{|V|})$ [15].

Number of isolated nodes

An isolated node is a node that has no connections with other nodes (degree = 0). The computation time for this metric is $O(|E|)$ [15].

Edge connectivity

The edge connectivity of a graph is the minimum number of edges that should be removed to make the graph disconnected, and can be computed in time $O(log(|v|) \cdot |V|^2)$ [15].

Assortativity

Assortativity is a measure that describes a node's tendency to connect to other nodes with similar degree. This metric is computed using Pearson correlation coefficient. The value of this correlation r lies in the range of $[-1, 1]$. As the network becomes more assortative, the value of r becomes closer to 1.

Number of maximal cliques

A clique is defined as a subset of a graph's nodes that are all connected to each other. Thus, a maximal clique is the largest clique found in the graph, which is not extendable by adding an adjacent node to it. In other words, a maximal clique is not a subset of a larger clique.

Finding all the maximal cliques in a graph can take upto time $O(3^{|V|/3})$ [56].

Graph clique number

Graph clique number is the size of the largest clique found in the graph. The computation time for this metric is $O(3^{|V|/3})$ [57].

2.4.3 Motifs

When representing the brain functional network as a graph, one of the best ways of investigating its underlying mechanisms is to look for recurrent and statistically significant connectivity subgraphs, called motifs. These subgraphs, which are particular patterns of connections between vertices, may be efficient ways of performing atomic tasks in the brain. Thus, it is interesting to investigate the frequency of their occurrence in the brain graphs.

In this project, we study the number of occurrences for seven classes of motifs with up to 4 nodes (See figure 2.1). Each of these types is considered as a network metric, and its time series is compared with all other metrics.

The reason for investigating limited types of motifs in this project is that motif detection is a computationally heavy task. It is also worth mentioning that many related works, such as [27], have also chosen to analyze motifs of size four and three as interesting network patterns.

2.5 Pearson correlation coefficient

The Pearson correlation coefficient, developed by [42], measures the extent of linear dependence between two variables (samples). Pearson correlation coefficient for two variables X and Y is denoted by r_{XY} and is computed as follows:

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{2.20}$$

Where, X_i is the ith element in the time series and $\overline{X}$ is the mean of the time series. It is worth mentioning that r is a value in range $[-1, 1]$, where 1 means the two variables have total positive correlation, while -1 and 0 are indicative of total negative correlation and no correla-

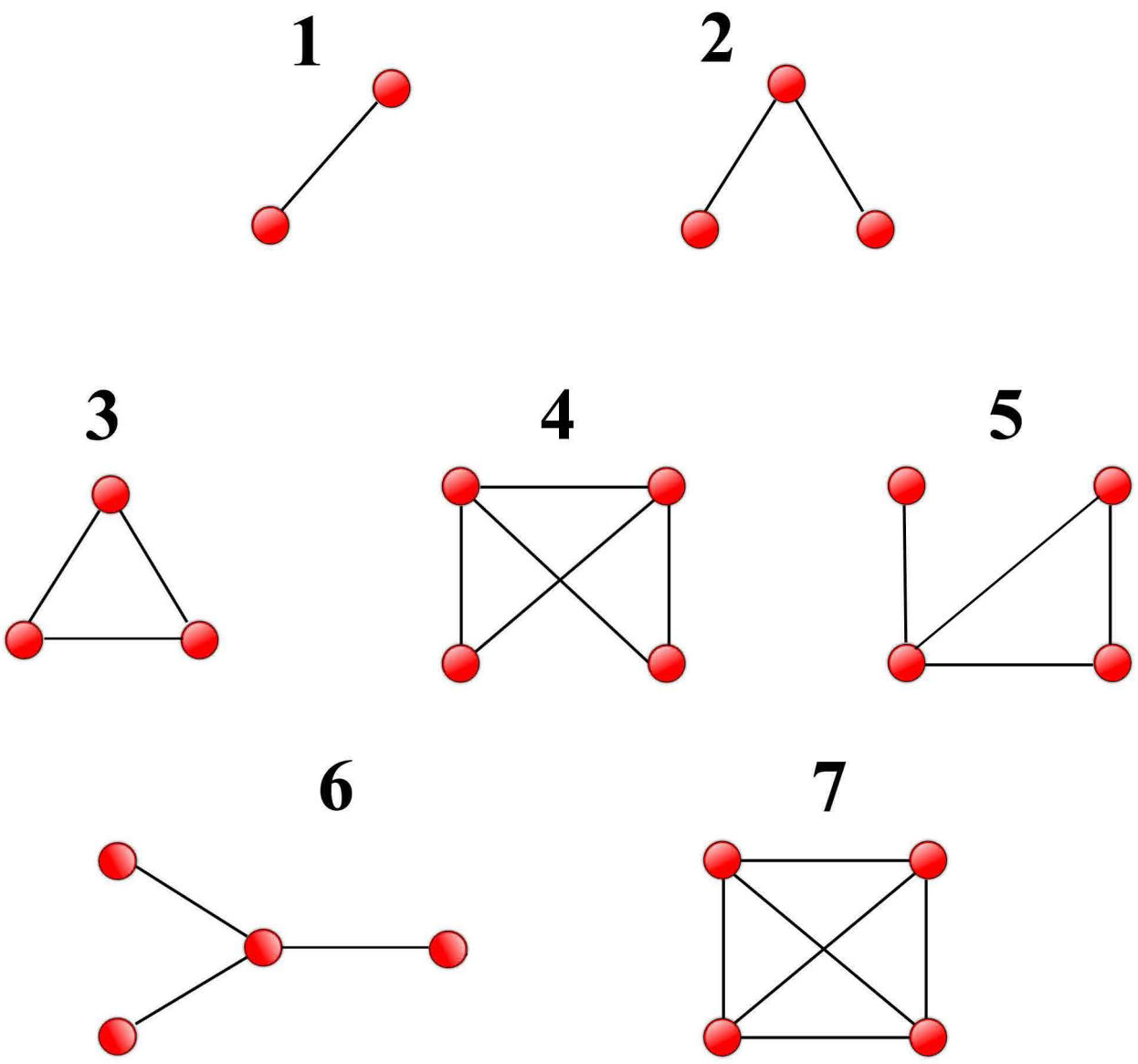

Figure 2.1: Seven equivalence classes of motifs with up to 4 nodes.

tion respectively.

The strength of the relationship between two variables is assessed not only by the correlation coefficient but also by considering the size of the variables (number of the pairs in data). If the correlation coefficient is computed based on a few number of pairs, then the r value should be very close to +1 (or -1) for the correlation to be considered as statistically significant. $p - value$, which is used for analyzing this level of significance, is the probability of obtaining the r value if the correlation did not exist in the first place. Thus, for the r value to be accepted as statistically significant, the computed $p - value$ should be less than a certain value, which is often chosen as 0.05 or 0.01. The calculation of $p - value$ is based on a number of assumptions that are beyond the scope of this project, but one can refer to standard statistical tables for obtaining the $p-values$ relevant to the computed r and the sample size [21].

In this project, the Pearson correlation coefficients and their corresponding $p - values$ for time series of metrics are computed by the *pearsonr* function provided in the scipy.stats library.

It should be noted that though there is no exact guideline for choosing a proper r value as a threshold for accepting or denying the correlation between two variables, it is common to consider $r = 0.5$ indicative of an existing correlation. Also, the $p - value$ used in this project is 0.01.

2.6 A summary of the pipeline

In summary, the following steps are performed by the implemented pipeline to generate the correlation matrix:

1. Load the EEG data-set. The data-set includes 5 trial types, and there are multiple trial

recordings for each trial type.

2. Apply the digital band-pass filter (FIR filter). This is done to extract the desired frequencies from the whole data-set.

3. For each trial type (extracted from the data-set):

   (a) Perform the averaging process, to increase the signal to noise ratio.

   (b) Choose a desired window size and step size.

   (c) Do the following for each time window (the number of time points used = the window size):

      i. Create the similarity matrix using PLI. This matrix contains similarity values for each pair of channels.

      ii. Threshold the obtained similarity matrix using a thresholding method. In this step we use one of the three thresholding methods introduced in this chapter (S, FDR and RMT).

      iii. Use the adjacency matrix obtained from the previous step to build a graph. Each channel represents a node, and there is an edge between two nodes, only if the value between their corresponding channels in the adjacency matrix is 1.

      iv. Add the obtained graph to the temporal series of graphs.

      v. Move forward in the time-series so that the distance with the next time window = the desired step size.

      vi. If it is still possible to have a time window with the desired size, go back to step *i*.

   (d) Having the time-series of graphs, do the following for each graph in this series:

      i. Define 25 time-series for 25 metrics. Compute each of the 25 metrics for this graph, and add their values to their corresponding time series of graph metrics.

(e) Having 25 time-series for graph metrics, compute the Pearson correlation coefficient for each pair of these 25 time-series. In this way, we can obtain the relationship between all possible pairs of 25 metrics, and see whether or not their pattern of changes over time is correlated with each other.

(f) To plot the relationship between metrics: keep the values that are greater than 0.5 or smaller than -0.5, with a $r - value$ smaller than 0.01. Otherwise there is no significant correlation between two metrics. Thus, a 25×25 correlation matrix is created, reflecting the correlation between metrics.

Note that we keep the nodes that do not have any connections with other nodes, because we want to count the number of isolated nodes as a graph metric. Thus, we only consider the number of edges dynamic.

2.6.1 Technical details

The whole implementation of this pipeline was done in Python. The generation of all graphs and graph metrics (except motifs) used in this project was done using the Python's NetworkX package [26], which has been used in many scientific articles such as [7]. For computing the motifs, we have used a Python-based package named graph motif model (GMM), implemented and introduced by [13].
The correlation matrices were plotted using R's corrplot package.

The calculation of motif counts was made possible using the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET), while other analysis was done on a personal computer with the following technical specifications:

- Intel(R) Core(TM)i7-3520M CPU @ 2.90 GHz.
- 8 GB Ram.

- Windows 7 64 bit.

Chapter 3

Metric correlations across different thresholding methods

Choosing an appropriate threshold is a critical step in generating brain connectivity graphs. A very low threshold can result in a very dense graph, with the high possibility of representing false synchronizations. On the other hand, choosing a very high threshold value could easily result in the loss of physiologically relevant connections.

There are various thresholding methods that can be used for generating brain graphs, thus, it would be interesting to compare the graph metrics correlations based on different thresholding methods to see the extent of the changes in the behavior of these graphs across these different methods.

In this chapter, we analyze the metric correlations of dynamic graphs obtained from three different thresholding methods, with a fixed window size, step size and frequency band for clearer observation of any possible differences solely caused by changing the thresholding method. We do these analyses separately for each trial type. After detecting the mutual metric correlations that can be seen in the results of all thresholding methods, we suggest metric

dependency groups for each trial type. Also, we talk about interesting contradictions of some metric correlations that are results of using different thresholding methods. Finally, we suggest metric dependencies that happen for all trial types, using all three thresholding methods.

All graphs in this section are generated under the following circumstances: window size = 800, step size = 25, frequency band = beta, and using one of three thresholding methods (S, FDR, RMT) described in Chapter 2. Note that the chosen combination of window size = 800 and step size = 25 is chosen because it generates a smooth transition between dynamic graphs.

We chose beta band as the default frequency band, since the related literature, such as [30], has shown that using this band results in greater statistical significance for task-related effects in EEG-derived brain graphs.

For S thresholding method, the S value is set to 2, since it is a liberal threshold that results in dense graphs [54].

The alpha value for FDR thresholding method is set to 0.01, which is a common value used in the related literature (such as [41]), for assessing the statistical significance of the results.

Finally, for RMT thresholding method, the min-threshold value is set to the minimum (non-zero) value in the correlation matrix and the max-threshold value is set to the maximum value that when used as threshold does not generate a singular matrix. It is worth mentioning that the RMT function uses 100 values between the min and the max value to threshold the correlation matrix with, and chooses the best value in terms of yielding the best Anderson-Darling score, as the final threshold value.

For metric correlations, we consider two metrics correlated only if the r value for their cor-

relation is greater than 0.5 or less than -0.5. Thus, only the correlation values in this range have been plotted.

Note that the definitions of all the abbreviations used in the plots of this chapter and Chapter 4 can be found in the appendix.

The position of clusters in the plots depicted in this chapter and Chapter 4 is not informative nor important.

3.1 Trial type: 1

By looking at Figure 3.1, we can see that a cluster of positive correlations is seen in the results of all three thresholding methods. This cluster contains pagerank, eigenvector centrality, authorities, closeness centrality and degree. These correlations are expected, considering their mathematical definition and that they are all influenced by degree. Note that closeness centrality and degree are also correlated with core number. In addition to this cluster, there is also a recurrent gathering of diameter, average path length (AvgPL) and betweenness centrality. While the reason for correlation of diameter and AvgPL is obvious (based on their definitions), betweenness centrality's relationship with these metrics is interesting and can be explained as follows: an increase in the AvgPL means that there are more nodes participating in traveling routes between every two nodes in the graph, thus, the average betweenness centrality of the graph rises. Closeness vitality is also correlated with AvgPL and diameter. This relationship can also be explained based on the same reasoning used for betweenness centrality's correlation.

Other interesting gatherings mostly concern different motif types. Motif types 3,4 and 7 are correlated with each other, while motif type 5 is correlated with motif type 4 and 3. Motif types

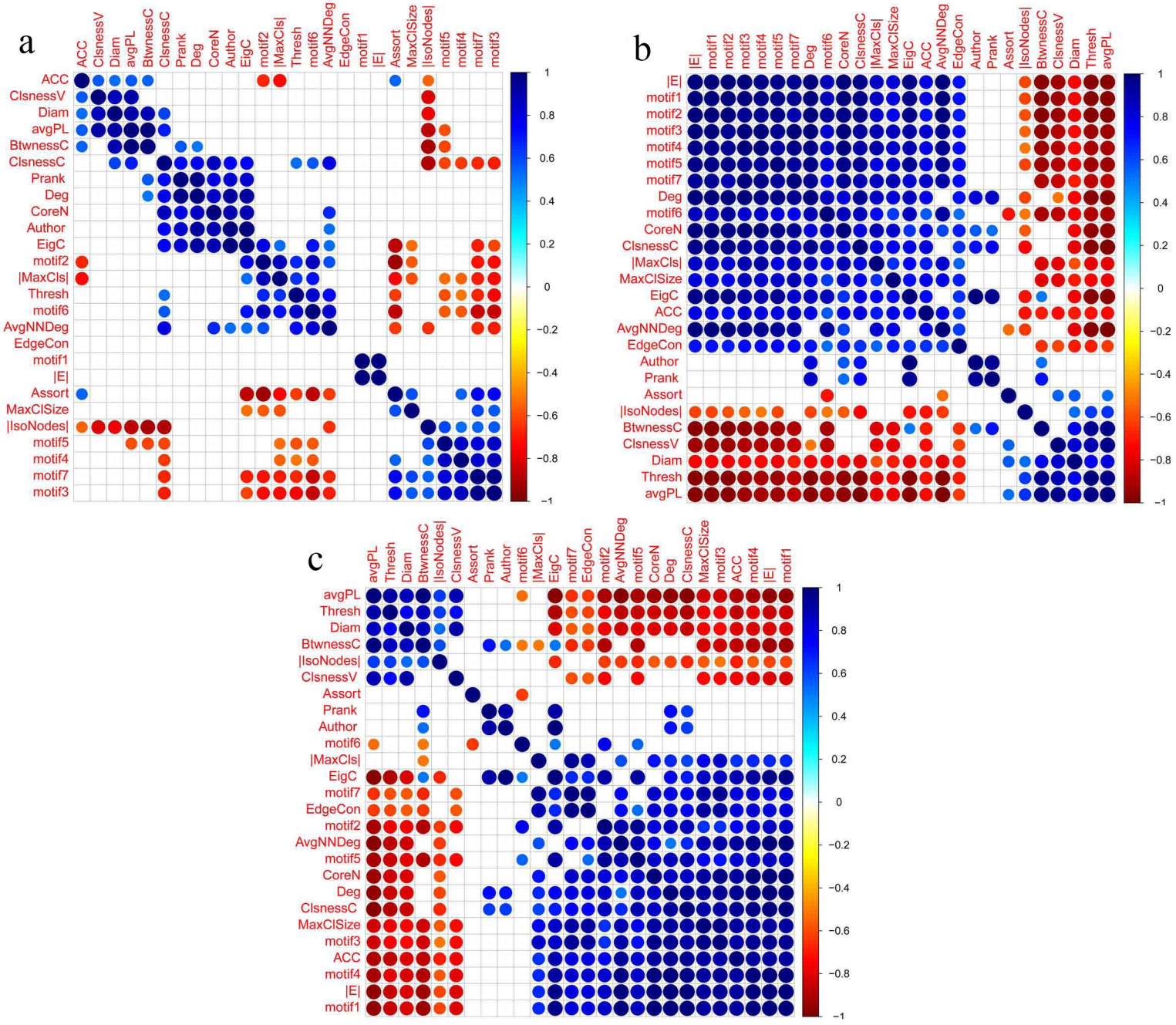

Figure 3.1: Metric correlations obtained based on trial type 1, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT.

3 and 7 are also correlated with graph clique number. Motif types 2 and 6 are correlated with each other and eigenvector centrality. Other interesting recurrent correlations are: motif type 1 with number of edges, average nearest neighbor degree (AvgNNDeg) with motif type 2 and core number, and finally, betweenness centrality with pagerank. Note that although pagerank and eigenvector centrality are both based on the same mathematical reasoning, the correlation of betweenness centrality and eigenvector centrality is not captured in S thresholding results.

After outlining positive correlations, recurrent negative correlations are also of interest. It can be seen that number of isolated nodes is negatively correlated with average clustering coefficient (ACC), closeness centrality and AvgNNDeg, which is expected since as more nodes depart from the connected graph, degree related metrics would decrease in value.

Note that motif type 5 is negatively correlated with AvgPL and betweenness centrality, while motif type 6 is negatively correlated with assortativity. While the negative correlation of motif type 5 and AvgPL is intuitively understandable, the reason behind the negative correlation of assortativity and motif type 6 is not clear. Nevertheless, we offer a possible explanation for this repeated correlation as follows: We know that motif type 6 is relevant to the presence of hubs in the network (See Figure 2.1). A hub is defined as a high-degree node, which, is in fact one of the properties of networks with small-world characteristics [10]. On the other hand, assortatiivty is also expected to be seen in networks with high clustering coefficient, since the high degree nodes tend to connect to each other. Considering this, the negative correlation of assortativity with motif type 6 in functional brain networks seems odd and unexpected. But this observation can be justified by referring to the definition of the rich-club phenomenon. This phenomenon happens when nodes of higher degree (hubs), are more connected to each other than the less significant, lower degree nodes. This causes an increase in the local assortativity, while the global assortativity decreases. It has been observed that the brain functional network has a rich club organization and that this phenomenon facilitates the global efficiency of trans-

ferring information through a network [59]. On the other hand, it has also been observed that in epileptic brain networks, the assortativity increases during seizures [6]. Considering these facts, it can be assumed that the brain functional network tries to maintain its rich-club organization by lowering the network's assortativity as the network hubs become richer. Thus, as the number of motifs of type 6 increases, the assortativity decreases.

Based on the observations in this section, we summarize the above metric correlations in the following dependency groups (metrics put together in a group are all correlated with each other): (pagerank, eigenvector centrality, closeness centrality, degree, authorities), (core number, degree, closeness centrality), (core number, AvgNNDeg), (motif2, motif6, eigenvector centrality), (assortativity, motif6), (motif1, number of edges), (AvgPL, diameter, betweenness centrality), (closeness vitality, AvgPL, diameter), (motif4, motif7, motif3), (motif4, motif5), (motif6, motif2), (motif2 , AvgNNDeg), (number of isolated nodes, ACC), (number of isolated nodes, closeness centrality), (number of isolated nodes, AvgNNDeg), (motif5 , AvgPL, betweenness centrality), (pagerank, betweenness centrality).

By comparing metric correlations obtained based on three thresholding methods (S, RMT and FDR), the following interesting observations are made: it is clearly seen that RMT and FDR have many mutual correlations, and their results do not contradict each other. On the other hand, by comparing their results to S threshold's, we can see that some correlations exist in opposite directions. Also, note that S thresholding method has failed to capture any correlations regarding edge connectivity. These differences are due to the fact that S thresholding method generates higher threshold values comparing to FDR and RMT [16], and thus metric correlations sensitive to density and number of isolated nodes are captured differently.

For metric correlations based on S thresholding method, we can see that ACC is positively correlated with average path length, while this correlation is negative for FDR and RMT correlations. This is due to the fact that the graphs generated based on this thresholding method

have noticeable fewer number of connections comparing to the ones created by RMT and FDR. Because of this, these graphs have a lot of isolated nodes, and since average path length is susceptible to disconnected nodes (see [53] and [45]), the results may vary based on the number of these detached nodes. The same reasoning is behind other contradictory correlations between S and two other thresholding methods. These explanations also account for the results of other trial types.

3.2 Trial type: 2

Figure 3.2 shows metric correlations based on three thresholding methods for trial type 2. For this trial, core number has joined the gathering of eigenvector centrality, pagerank, authorities, degree and closeness centrality, for all thresholding method's results. Closeness vitality, diameter, AvgPL and betweenness centrality have the same relationships with each other as before. Motif type 1 and number of edges are again strongly correlated. This correlation is expected to always be seen, since motif type 1 is in fact representing an edge, thus we will not repeat this as a correlation from now on. It is interesting to see that the graph clique number has joined the cluster of motif types 3, 4 and 7. Motif type 4 and 3 are also correlated with ACC and motif type 5. Their correlation with ACC is due to the fact that they represent connections forming one or two triangles, which is related to the definition of ACC.

In summary, these dependencies are seen for all thresholding methods: (pagerank, eigenvector centrality, closeness centrality, degree, authorities, core number), (core number, AvgNNDeg), (motif2, eigenvector centrality), (assortativity, motif6), (motif1, number of edges), (AvgPL, diameter, betweenness centrality), (closeness vitality, AvgPL, diameter), (motif4, motif7, motif3, graph clique number), (motif4, motif5, motif3, ACC), (motif2 , AvgNNDeg), (motif5 , AvgPL, betweenness centrality), (motif5, closeness vitality), (motif2, eigenvector centrality, closeness centrality), (core number, AvgNNDeg), (pagerank, betweenness central-

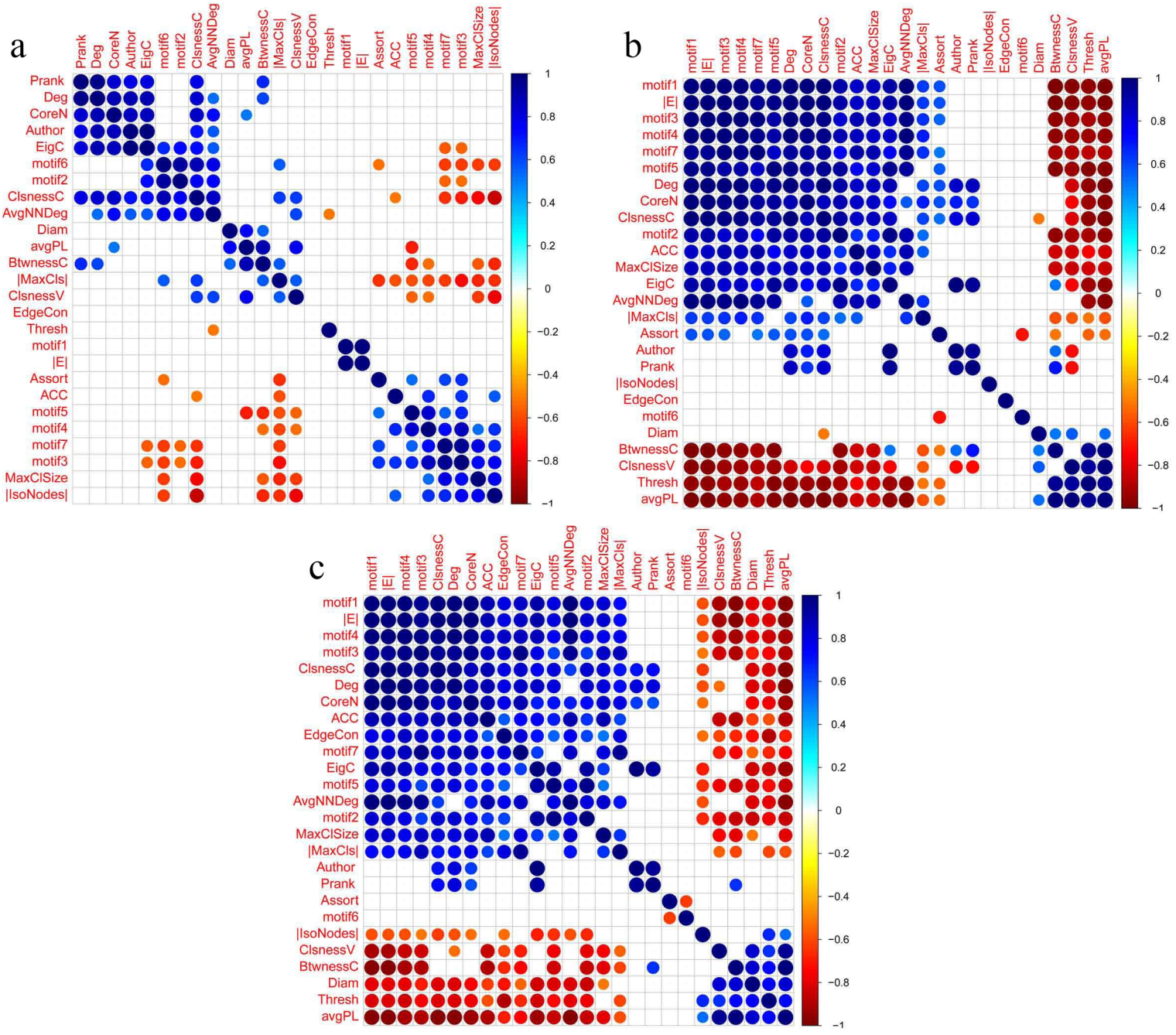

Figure 3.2: Metric correlations obtained based on trial type 2, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT.

ity), (closeness vitality with: motif5, motif4 and graph clique number).

3.3 Trial type: 3

Figure 3.3 shows correlation results for trial type 3. By comparing the results of three thresholding methods, we can detect the correlations that can be seen in all three plots. Eigenvector centrality, degree, closeness centrality, pagerank and authorities are all correlated with each other, while AvgNNDeg and core number are also partially part of this gathering. This is due to the fact that AvgNNDeg is only correlated with all members of this cluster in the results of S thresholding method, and core number is not correlated with pagerank and authorities for graphs obtained based on RMT thresholding method.

Another recurrent and interesting cluster of correlations is the one containing graph clique number and motif types 3,4,5 and 7. These motif types are also always negatively correlated with AvgPL and betweenness centrality. Note that all of these motif types have closed cycles of length 3 or 4 in them (see Figure 2.1), which can lead to the emergence of bigger cliques in the graph, confirming their positive correlation with graph clique number. Also, their negative correlation with distance related metrics, is a sign of increased global efficiency due to appearance of these motifs in the network. Also, one would assume that ACC should also be positively correlated with these metrics, since an increase in the number of these motifs adds to the number of triangles in the graph, and thus leads to more clustering. We can see this expected correlation in the results of RMT and FDR, while the S thresholding method has failed to capture it.

By further comparison of the results, we can see that the familiar gathering of diameter, AvgPL, betweenness centrality and closeness vitality has again taken place for this trial type. But note that, as for previous trial types, betweenness centrality and closeness centrality are not corre-

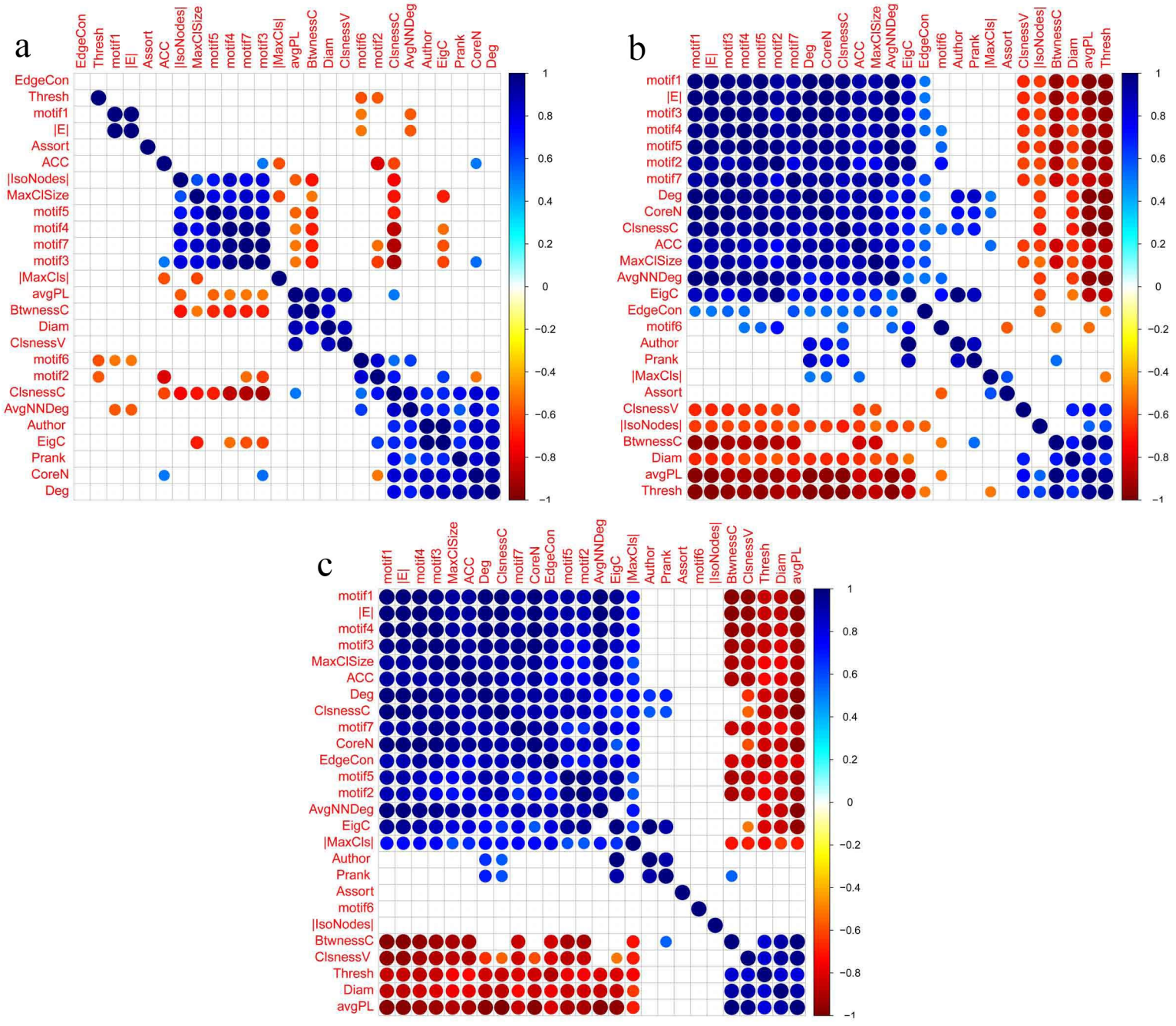

Figure 3.3: Metric correlations obtained based on trial type 3, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT.

lated.

Other recurrent correlations, plus the ones mentioned above, can be put in the following dependency groups: (eigenvector centrality, pagerank, authorities, closeness centrality, degree), (core number, eigenvector centrality, closeness centrality, degree), (AvgNNDeg, core number, closeness centrality, degree), (diameter, AvgPL, betweenness centrality), (diameter, AvgPL, closeness vitality), (motif3, motif4, motif5, motif7, graph clique number, betweenness centrality), (motif3, motif4, motif5, motif7, AvgPL), (eigenvector centrality, motif2, closeness centrality), (motif3, core number), (motif3 , ACC).

3.4 Trial type: 4

We now look at Figure 3.4, which shows metric correlation results of three thresholding methods for trial type 4. For this trial type, we can again see the familiar clustering of eigenvector centrality, pagerank, authorities, degree and closeness centrality. Unlike trial 2, core number does not completely participate in this cluster and is only correlated with degree, closeness centrality and to a lesser extent with eigenvector centrality. It is also always correlated with AvgNNDeg.

AvgPL is again strongly correlated with closeness vitality and betweenness centrality in the results of all thresholding methods, but, it can be seen that unlike last trial types, diameter is not correlated with these metrics. Betweenness centrality is also correlated with pagerank.

By comparing the three sets of results, we notice the recurrent appearance of a cluster of correlations containing motif types 3,4,5 and 7, ACC and to some extent graph clique number. Graph clique number is always correlated with motif types 4 and 5, except for FDR results. As it can be seen, the theshold control shows negative correlation with this cluster, which makes

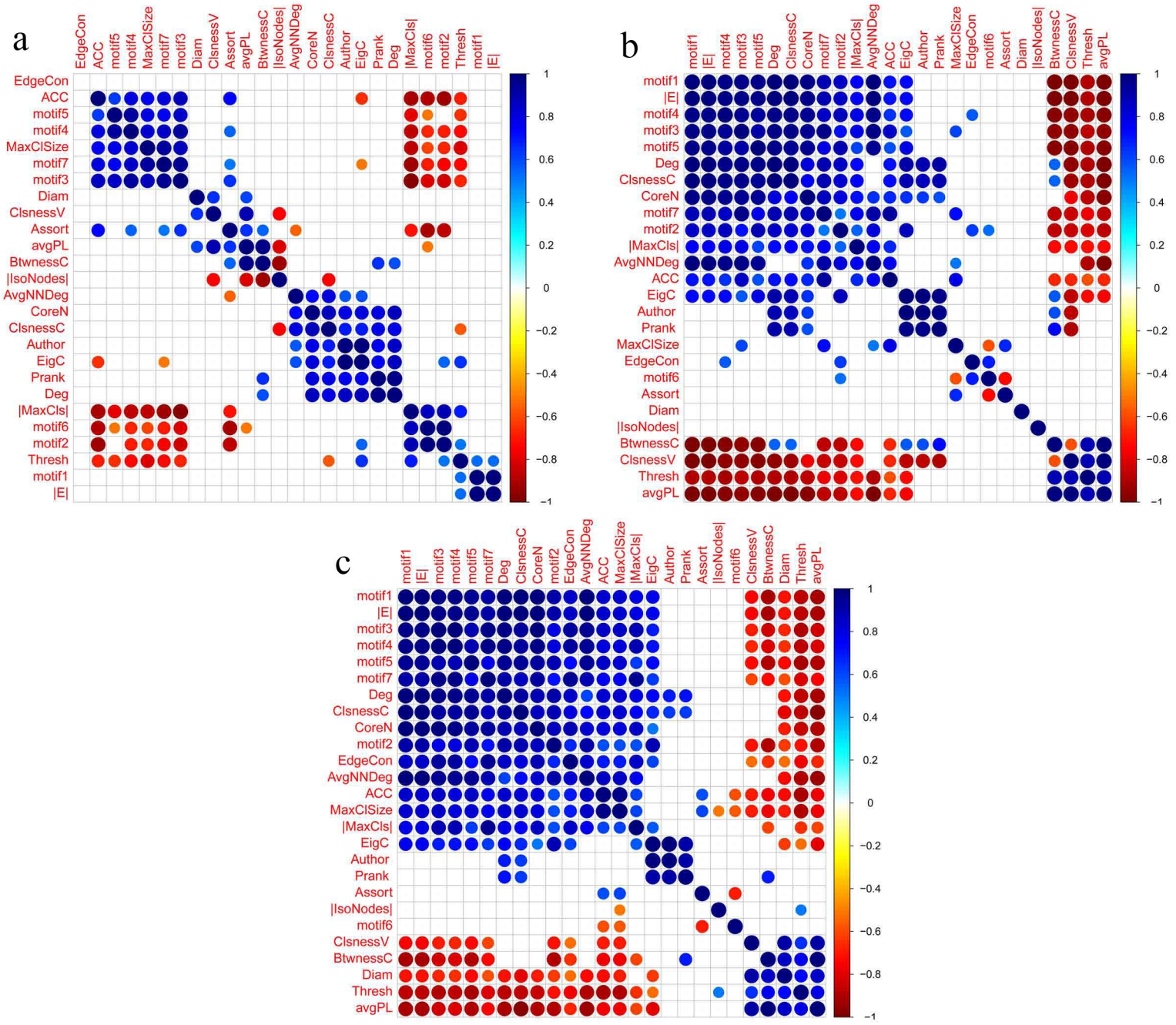

Figure 3.4: Metric correlations obtained based on trial type 4, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT.

sense since as the threshold increases the graph loses more connections and thus become less clustered.

Motif type 2 and number of maximal cliques are also correlated.

When comparing the results of three thresholding methods, we can see that for S thresholding method, ACC has negative correlation with the number of maximal cliques, whereas, for FDR and RMT, this correlation is positive. This can be explained as follows: As the threshold value decreases, more edges are added to the graph, and ACC increases as a result. In FDR and RMT, new connections are added to the graph with a greater rate than in a graph obtained via S thresholding method. Thus, while several maximal cliques of size k are joined together to create a bigger clique, new connections also create new maximal cliques. On the other hand, for S thresholding method, the new connections are mostly creating bigger cliques by connecting several maximal cliques. This decreases the number of maximal cliques because the number of newly created cliques is less than the number of former maximal cliques, which are now joined together to make a bigger clique.

In summary, these dependency groups can be seen for this trial type and all the thresholding methods: (eigenvector centrality, pagerank, degree, closeness centrality, authorities), (core number, eigenvector centrality, closeness centrality, degree), (closeness vitality, AvgPL), (betweenness centrality, AvgPL), (motif2, number of maximal cliques), (motif3, motif4, motif5, motif7, ACC), (motif3, motif7, ACC, graph clique number), (assortativity, motif6), (graph clique number, motif6), (eigenvector centrality, motif2).

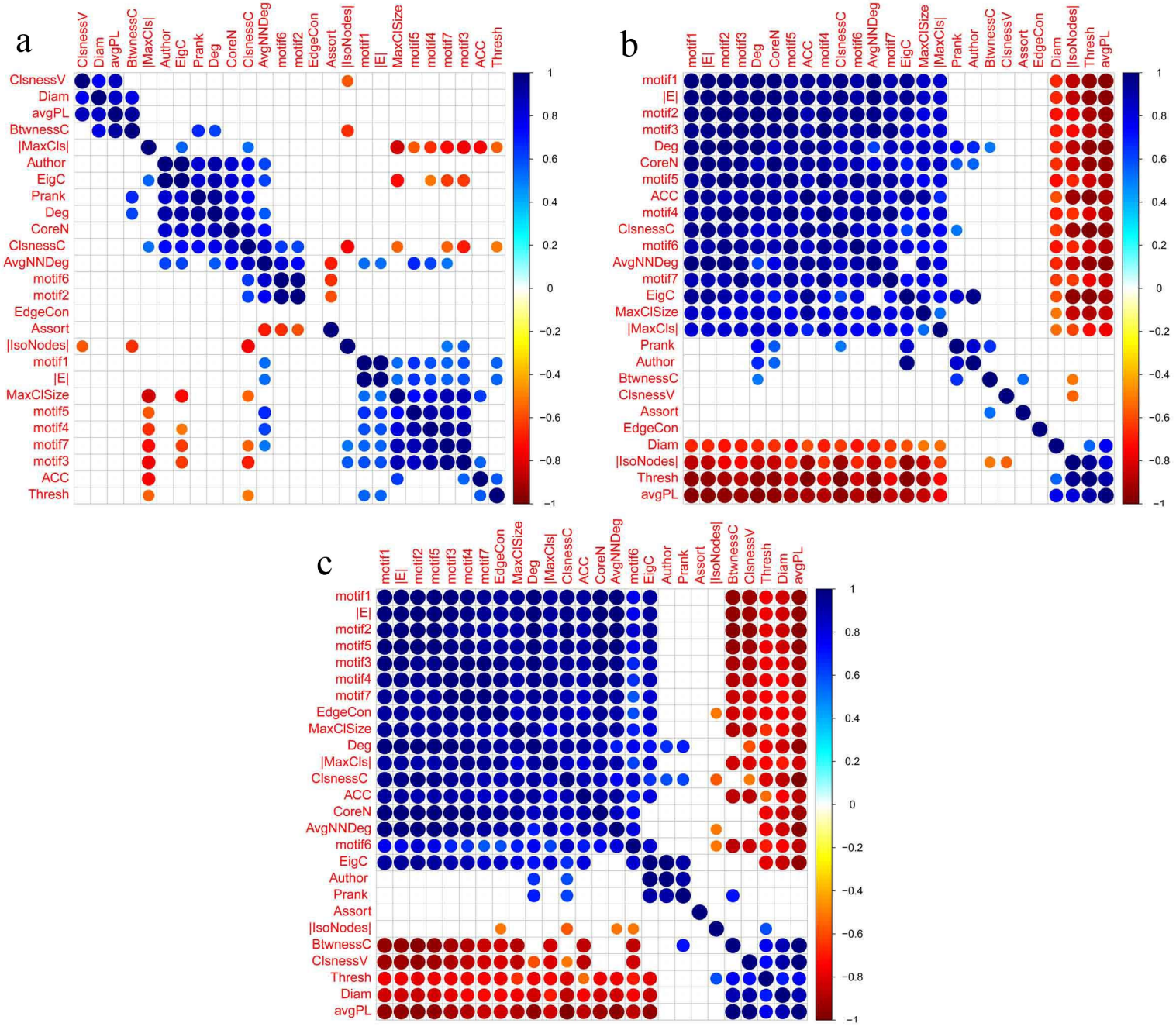

Figure 3.5: Metric correlations obtained based on trial type 5, and using window size = 800, step size = 25 and thresholding methods: (a)S, (b)FDR and (c)RMT.

3.5 Trial type: 5

Here, in Figure 3.5, we can see that a big cluster of positive correlations keeps showing up for all the thresholding methods. This cluster contains motif types 1,3,4,5 and 7, number of edges and the graph clique number. AvgNNDeg is also correlated with most members of this cluster, except for graph clique size and motif type 3. On the other hand, ACC is positively correlated with graph clique size and motif type 3.

Furthermore, we can see a clustering of AvgNNDeg, motif types 2 and 6, and closeness centrality, which is logical to be seen, considering the definitions of these metrics. AvgNNDeg is also correlated with core number and degree.

It is interesting to see that the familiar clustering of degree, pagerank, authorities, eigenvector centrality and closeness centrality does not appear as complete as before, and suffers from lack of some correlations. Another noticeable change for this trial type is that closeness vitality and betweenness centrality are not correlated with AvgPL or diameter as seen in previous trial types.

The above observations and other recurrent correlations for all thresholding methods can be categorized as follows: (motif1, motif3, motif4, motif5, motif7, number of edges, graph clique number), (AvgNNDeg, motif1, number of edges, motif5, motif4, motif7), (AvgPL, diameter), (ACC, graph clique number, motif3), (AvgNNDeg, motif6, motif2, closeness centrality), (AvgNNDeg, core number, degree), (closeness centrality, core number, degree), (eigenvector centrality, authorities, pagerank, degree), (betweenness centrality, pagerank), (number of isolated nodes, closeness centrality), (number of maximal cliques, eigenvector centrality, closeness centrality).

3.6 Discussion

We now summarize the dependency groups suggested for each trial type, offering metric correlations that are constantly observed across all thresholding methods and trial types. These robust metric correlations can be categorized as follows: (motif4, motif3, motif7), (motif5, motif4, motif3), (graph clique number, motif7, motif3), (degree, pagerank, authorities, eigenvector centrality), (core number, closeness centrality, degree), (closeness centrality, degree, pagerank, eigenvector centrality), (AvgNNDeg, core number).

Note the presence of motif type 3 in three of the dependency groups. This is due to the fact that this motif acts as a substructure for motif types 4, 5 and 7. Also, motif types 3 and 7 are both complete graphs of 3 and 4 nodes, respectively, and participating in the formation of cliques in the graph.

Note that some very expected correlations, such as the gathering of AvgPL, diameter and betweenness centrality, do not appear in these groupings. This is due to the fact that they did not appear in the results of all trial types and thresholding methods, which is mostly due to different connection densities of graphs generated by these thresholding methods.

Based on the results discussed for all trial types, we can conclude that graphs generated based on the S thresholding method are somewhat different in structure from the ones generated via FDR and RMT methods. Graphs obtained based on this method lack many connections due to very high threshold values, which results in a disconnected network with a lot of isolated nodes. As discussed before, this is the main reason for some odd correlations between AvgPL, ACC and other metrics for S thresholding method's results. While one cannot conclude that this thresholding method is not useful for generating brain graphs in general, it can be said that the threshold values generated by S thresholding method are higher than necessary for this

EEG data-set, and thus, it might be better to use FDR or RMT so that more metric correlations are captured.

Also note that FDR and RMT employ very different approaches to threshold a graph, and the fact that they yield very similar correlations increases the credibility of the results, especially when comparing with S thresholding method correlations.

Considering this, there is still the need of further investigation to reach a final conclusion about metric dependency groups. Thus, we continue to observe metric correlations of brain functional networks, generated by RMT thresholding method in various conditions, by using different window sizes, step sizes and frequency bands.

The reason for choosing the RMT thresholding method in this project is that while this thresholding method's metric correlations are mostly similar to FDR's correlations, and agree with small-world properties, it has not yet received any attention in neuroinformatic literature. On the other hand, FDR has been used in many neuroscientific literature such as [1], [44], [67] and [65].

Chapter 4

Metric correlations across different circumstances

In this chapter, we investigate the effects of using different combinations of window and step sizes for generating dynamic graphs, on the metric correlation results. We also explore metric correlations obtained based on dynamic graphs generated in different frequency bands. This is done in three sections. First, we fix the step size and obtain the dynamic graphs based on different window sizes. The size of time windows used in this section are set to 200, 500, 800 and 1100 ms, so that we can observe the metric correlations based on different window sizes. Considering that the step size is fixed to 25 time steps, for all window sizes, more than 90 percent of time points based on which each graph is generated are common with the ones used for generating the previous graph.

In the second section of this chapter, we set the window size and step sizes in a way that different portions of shared time points between time windows is used for generating the graphs. This is further explained in section 4.2.

And finally, the third section explores metric correlations in three different frequency bands,

by fixing the window size and the step size.

At the end of this chapter, we suggest metric dependencies that are consistent across the results of all sections. Moreover, we discuss some interesting metric correlations that are repeatedly observed for specific trial types.

4.1 Changing the window size

All metric correlations in this section are obtained based on dynamic graphs generated under the following circumstances:

1- Using RMT thresholding method.
2- Setting step size to 25.
3- In beta frequency band.

Four sets of dynamic graphs are generated based on four different time windows of sizes 1100, 800, 500 and 200 ms.

We now study the metric correlation results for each trial individually.

4.1.1 Trial type 1

Figure 4.1, shows metric correlations obtained based on four different time windows of sizes 200, 500, 800 and 1100 ms. Looking at these metric correlations, we can see that some of them appear in the results of all window sizes. The biggest cluster of recurrent and persistent corre-

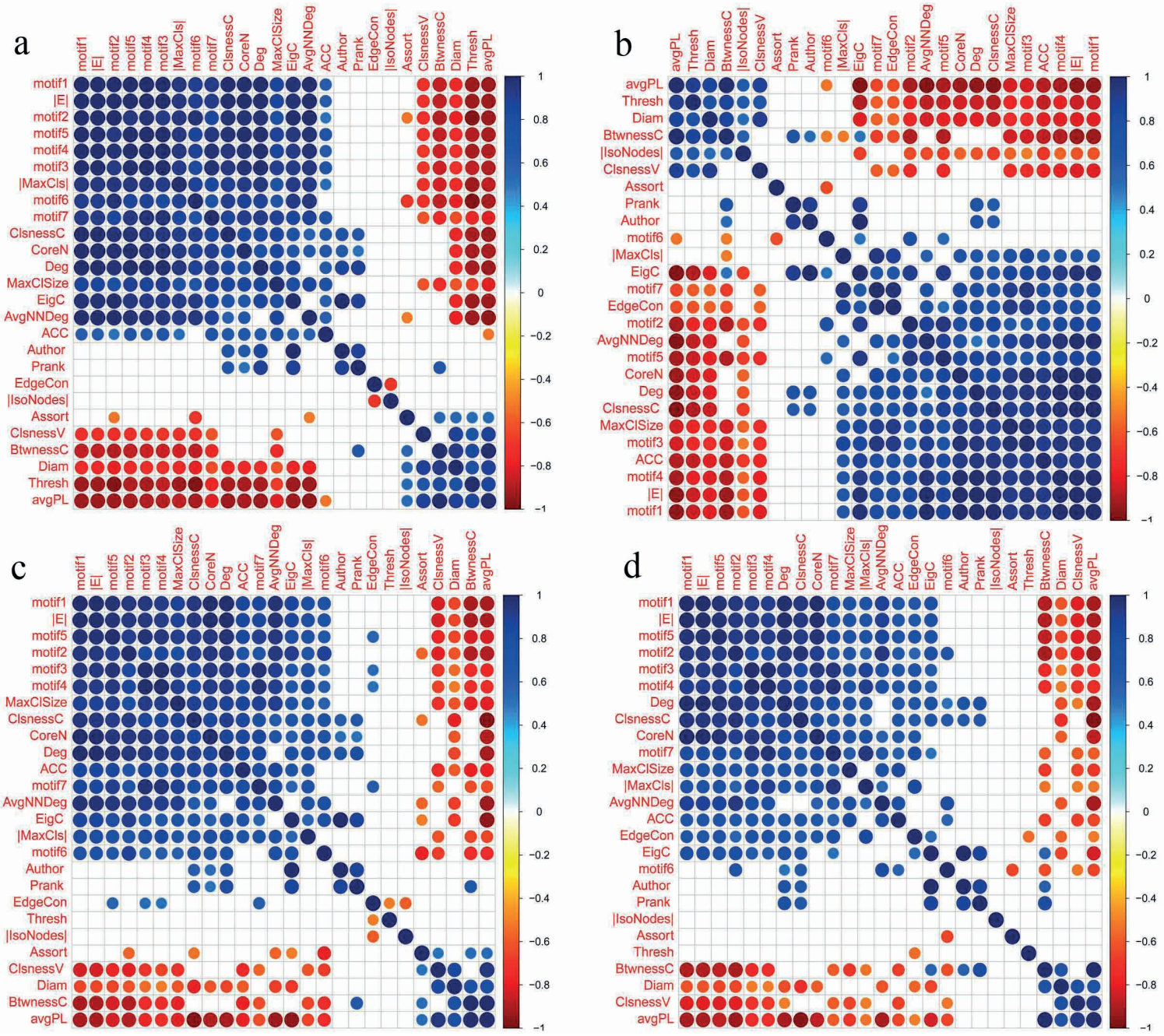

Figure 4.1: Metric correlations obtained based on trial type 1, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method.

lations includes motif types 1,2,3,4 and 5, ACC, graph clique number, core number, closeness centrality, degree, number of edges and average path length (AvgPL). These correlations make perfect sense, since as the node degree increases, all metrics that are directly influenced by the node degree increase as well. AvgNNDeg is also always correlated with most of the members of this cluster, except for closeness centrality and degree. Also, as a result of the network's growth, ACC increases, which agrees with the expected small-world properties of the brain's functional network. Furthermore, motif type 7 is also correlated with all members of the big cluster of correlations, except for motif types 2 and 5.

Note that closeness centrality and degree also form another cluster of correlations with eigenvector centrality, pagerank and authorities.

AvgPL and diameter are always correlated with betweenness centrality and closeness vitality. This correlation is derived from the fact that as the AvgPL and diameter increase, the number of nodes participating in the paths between all nodes increase as well. Considering that both betweenness centrality and closeness vitality reflect the amount of participation of a node in the whole network, their correlation with average path length and diameter is expected. On the other hand, these metrics do not have a correlation with each other because a node with high betweenness value does not necessarily create the only possible shortest path between two other nodes, while a node with high closeness vitality is certainly vital for the existence of certain paths in the graph.

Other interesting positive correlations involve motif type 7 and number of maximal cliques. These metrics are strongly correlated with each other, since motif type 7 is a complete graph of four nodes that plays the role of a building block for forming bigger cliques in the graph. AvgNNDeg, core number, motif types 3 and 4, closeness centrality, degree, number of edges and motif type 1 are also always correlated with both of these metrics. Considering the mathe-

matical definition of these metrics, these correlations are expected. Motif type 7 is also correlated with ACC and graph clique number.

Also, it can be seen that in this trial type, eigenvector centrality is always correlated with motif types 1,2,3,4,5 and 7. Since all these motif types are related to forming clusters in the graph, the importance of the nodes in these cluster increase as well, therefore this correlation is understandable. The expected correlation of motif types 2 and 6 is the last positive correlation that can be spotted for all window sizes in this trial type.

Moving to detecting recurrent negative correlations, we start by looking at the relationship between assortativity and motif type 6. As it was previously seen, these metrics are again always negatively correlated. The reason behind this correlation was thoroughly explained in section 3.1. Note that motif type 6 is also negatively correlated with AvgPL and betweenness centrality. This is due to the fact that an increase in the number of this motif type is because of the increase in the number of hubs and clusters in the graph, decreasing the AvgPL.

Other recurrent negative correlations are as follows: AvgPL is negatively correlated with all members of the big recurrent cluster of correlations, while diameter, betweenness centrality and closeness vitality are only negatively correlated with some of them.

As it was mentioned before, the above correlations can always be seen for trial type 1, even by changing the window size. On the other hand, it is interesting to see if these correlations keep happening for all other trial types, thus, before suggesting metric dependency groups, we continue our observations for other trial types to detect any possible changes in these recurrent correlations.

4.1.2 Trial type 2

By looking at Figure 4.2, we can see that many correlations are repeatedly seen for trial type 2, using all four window sizes. The biggest cluster of these correlations contain motif types 1,3,4 and 7, number of maximal cliques, closeness centrality, degree, number of edges, core number, ACC, graph clique number and AvgPL. Note that motif types 2 and 5, and AvgNNDeg are strongly correlated with each other while being correlated with most metrics of these cluster. Eigenvector centrality, authorities, pagerank, closeness centrality and degree are also strongly correlated with each other.

It is interesting to see that edge connectivity, unlike trial type 1, is now correlated with many metrics including AvgNNDeg, motif types 1,3,4 and 7, closeness centrality, degree and the graph clique number, for all four window sizes. Since these metrics are all indicative of density of the connections in the graph, it is logical to see that as they increase and the graph acquires more number of connections, the edge connectivity increases as well.

AvgPL, diameter, closeness vitality and betweenness centrality have the same relationships with each other as before. Moreover, if we put diameter aside, these metrics have mostly the same negative correlations as before. The only metrics that keep being negatively correlated with diameter for this trial type are eigenvector and closeness centrality scores.

Also note that assortativity and motif type 6 are again negatively correlated. The positive correlation of pagerank with betweenness centrality is repeated as well. This correlation is somewhat intuitive since as a node's number of connections increase, it participates in more shortest paths than before.

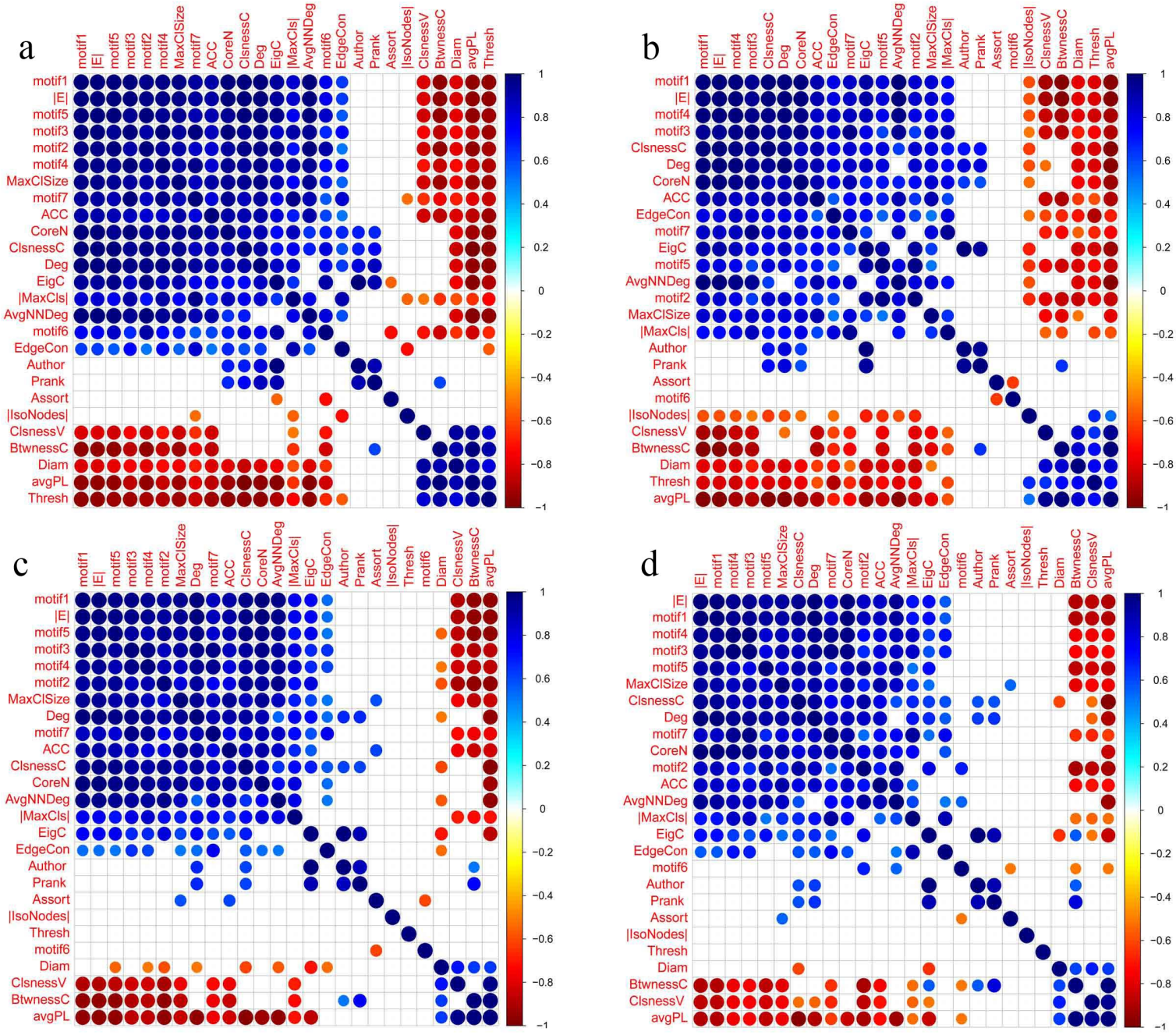

Figure 4.2: Metric correlations obtained based on trial type 2, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method.

4.1.3 Trial type 3

We now look at Figure 4.3, which shows the correlation results for trial type 3. Most of the metric correlations that repeat for each window size are familiar since they were seen in previous trial types' results.

The big cluster of positive correlations seen in previous trial types, is again present for this trial type. As before, this cluster is accompanied by AvgPL, which is negatively correlated with all other members of this cluster. Diameter, AvgPL, closeness vitality, betweenness centrality, eigenvector centrality, pagerank, authorities, degree and closeness centrality exhibit the same positive correlations as for previous sections.

Note that edge connectivity is now positively correlated with all members of the big cluster of correlations except ACC.

As for negative correlations, they are mostly repeated as before, with some exceptions. First, motif types 6 and assortativity are not correlated in three out of four window sizes. They have near -0.5 correlation only when window size is set to 200 ms. Second, we can see that edge connectivity is now negatively correlated with AvgPL, betweenness centrality, closeness vitality and diameter. This is easily explained based on the small-world properties that are expected of brain networks. As the ACC increases, AvgPL decreases to some degree to maintain the global efficiency of propagating the information through network, thus, more number of edges are to be removed to make the network disconnected.

We can also see that unlike trial type 2, diameter is negatively correlated with most members of the big blue cluster of correlations (except ACC, clique related metrics and motif types 3 and 7), which appears for all window sizes.

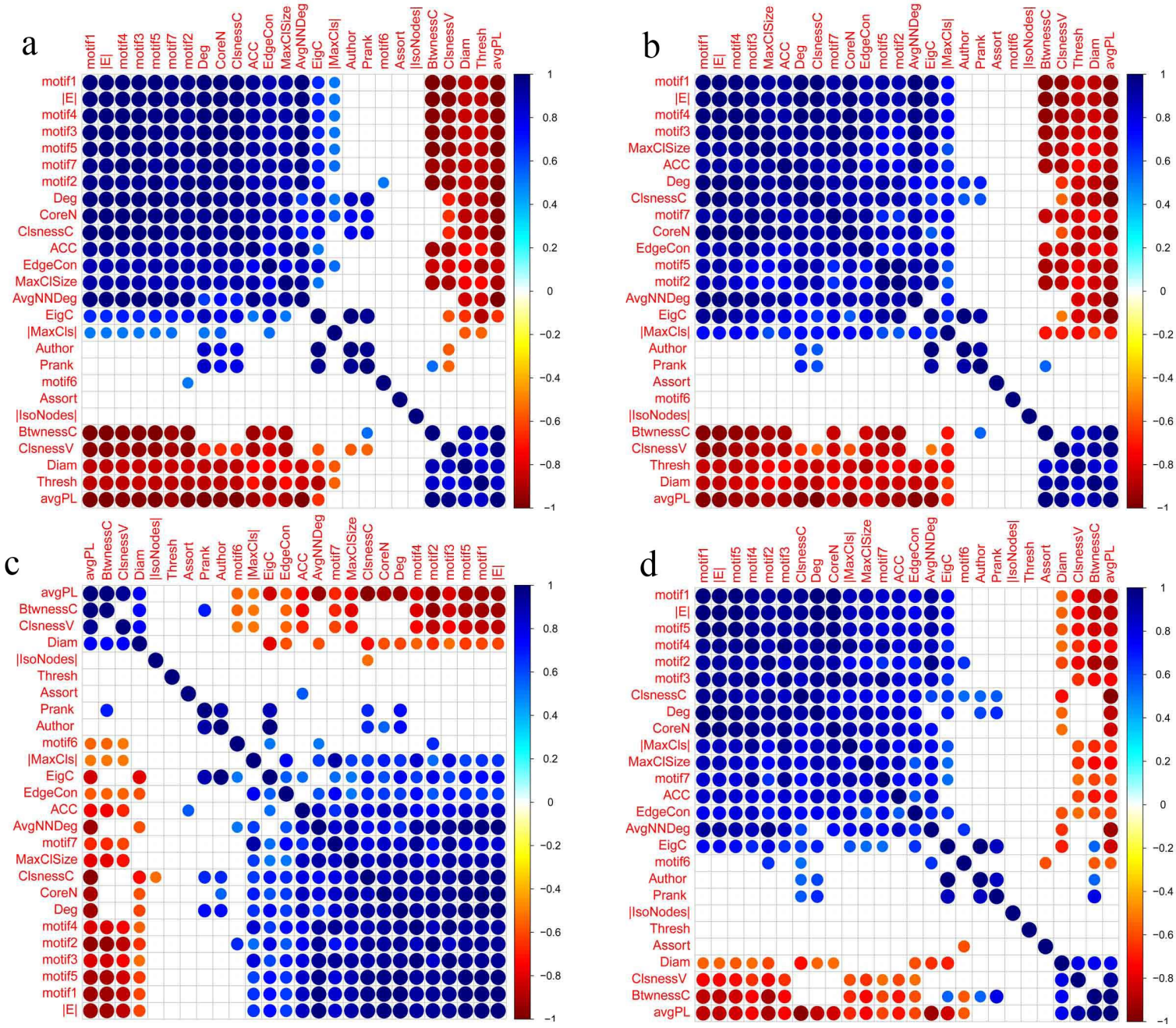

Figure 4.3: Metric correlations obtained based on trial type 3, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method.

4.1.4 Trial type 4

Here again we can see that most of the recurrent correlations that can be seen for all the window sizes are familiar and were seen for one or more of the previous trial types.

A big cluster of positive correlations is seen, which includes ACC, graph clique number, core number, motif types 1,3,4 and 5, closeness centrality, degree, and to some extent, motif type 2, AvgNNDeg and number of maximal cliques. Also, note that eigenvector centrality, edge connectivity, number of maximal cliques and motif type 7 are also positively correlated with some members of this cluster, while the last three mentioned metrics are strongly correlated with each other.

Metrics reflecting the importance of the node in the network, namely eigenvector centrality, authorities, pagerank and closeness centrality are again strongly correlated with each other and the node degree, as expected. On the other hand, diameter, AvgPL, closeness centrality and betweenness centrality have the same correlations with each other as for the previous trial types, which is not surprising based on their mathematical definitions. These metrics are also negatively correlated with most metrics participating in the big blue cluster of correlations mentioned before.

4.1.5 Trial type 5

The fifth and last trial type correlation results for all window sizes are depicted in Figure 4.5.

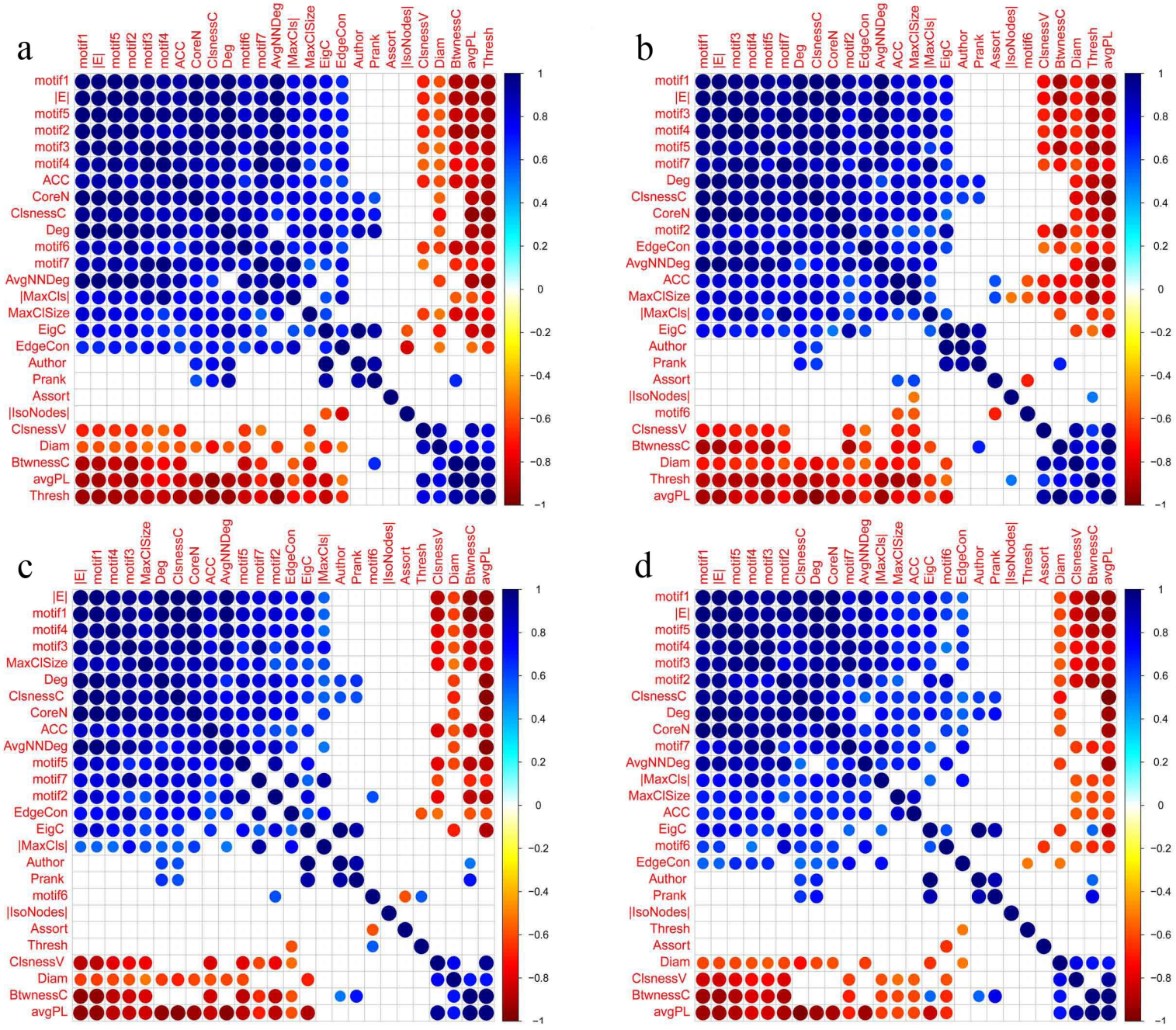

Figure 4.4: Metric correlations obtained based on trial type 4, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method.

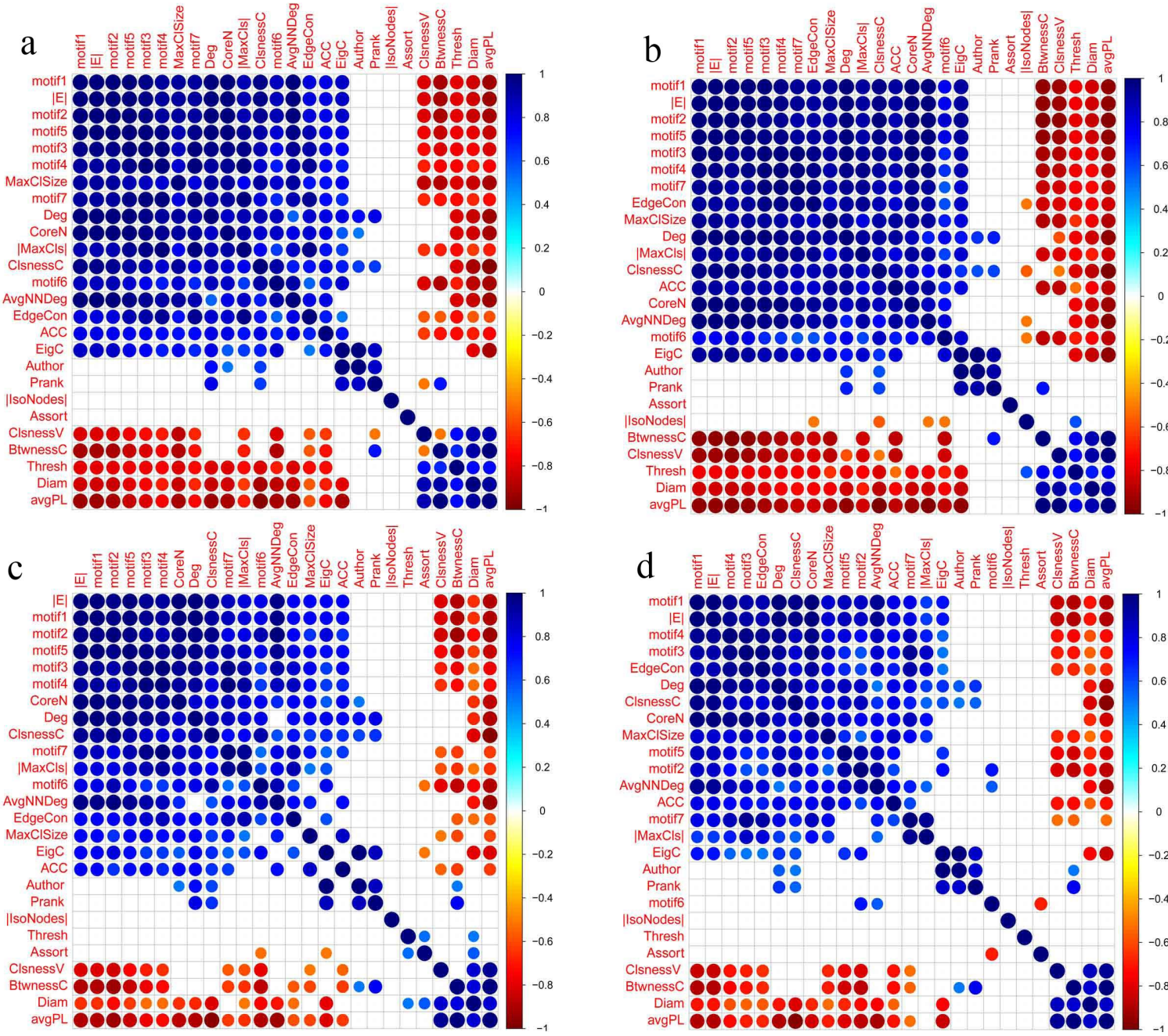

Figure 4.5: Metric correlations obtained based on trial type 5, in beta frequency band, and using step size = 25, window sizes: (a)1100, (b)800 and (c)500 and (d)200, and RMT thresholding method.

Although most of the positive and negative correlations are the same as the results of previous trial types, some interesting observations can be made for this trial type. It can be seen that edge connectivity, number of maximal cliques and motif types 3,4 and 7 form a cluster of strong positive correlations.

4.1.6 Recurrent dependencies

Based on the observations made from the results of all window sizes and trial types, we can now detect the metric correlations that are always present. Figure 4.6 shows metric correlations that are common between all trial types. Note that the correlation values shown in this figure are not precise and are in fact averaged over all trial types and windows sizes, with the sole purpose of making the detection of recurrent correlations easier for the reader.

Based on the previous observations, and by the help of Figure 4.6, we can suggest the following dependency groups for this section:

(pagerank, authorities, eigenvector centrality, degree, closeness centrality), (motif types 1,3,4 and 7, closeness centrality, degree, number of edges, core number, ACC, graph clique number, AvgPL), (AvgNNDeg, motif types 1,2,3,4 and 5, number of edges, core number, AvgPL), (pagerank, betweenness centrality), (AvgPL, diameter, betweenness centrality), (AvgPL, diameter, closeness vitality), (AvgNNDeg, motif type 5, AvgPL, ACC, graph clique number), (motif types 2 and 5, eigenvector centrality, closeness centrality, degree).

In addition to these main dependency groups, we can see that betweenness centrality and closeness vitality are negatively correlated with all motif types except motif type 6 and graph clique number, while diameter is negatively correlated with eigenvector and closeness central-

ity scores. On the other hand, number of maximal cliques is positively correlated with motif types 1,3,4 and 7, degree and core number, or in another words, metrics that are related to forming the cliques in the graph.

The appearance of ACC and AvgPL in most of the dependency groups, and the fact that they are negatively correlated, confirms the small-world characteristics of this network.

4.2 Changing the step size

As it was explained before, the intention of this section is to explore metric correlations by changing the portion of shared time points between time series, based on which the graphs are created.

The EEG-derived data set used in this project, offers a time series of 1803 ms for each EEG channel. Thus, using a big window with a big step size would result in a very small number of dynamic graphs, which can affect the statistical significance of the metric correlation results. Considering this, and based on the work of [38], which have used time windows with approximately 100 ms duration, we set the size of time windows in this section to 100 time points.

All metric correlations in this section are obtained based on dynamic graphs generated under the following circumstances:

1- Using RMT thresholding method.

2- Window size is set to 100 time points.

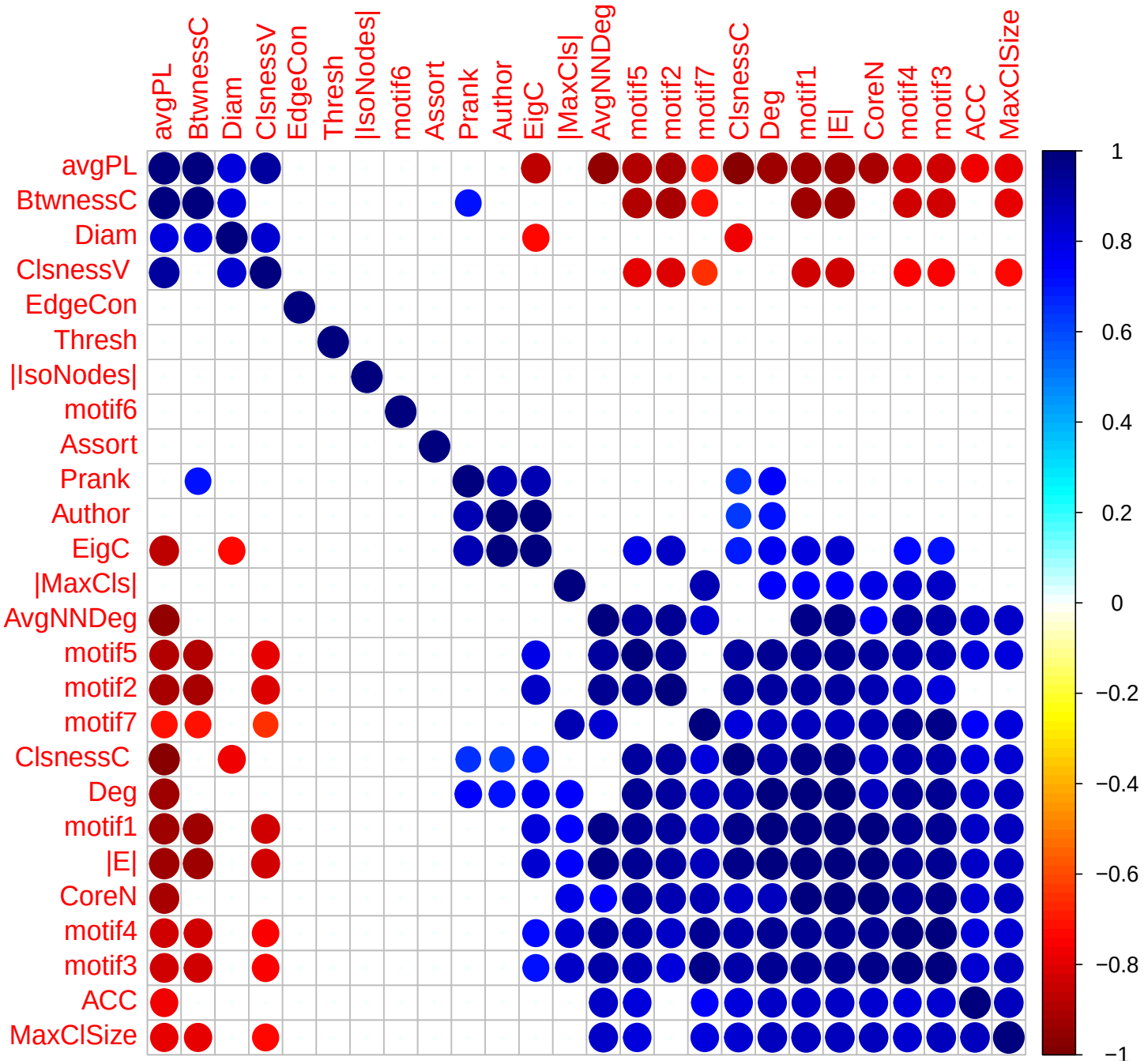

Figure 4.6: Recurrent metric correlations which were seen for all trial types and window sizes.

3- In beta frequency band.

Four sets of dynamic graphs are generated using four different step sizes of 5, 25, 50 and 100 ms, so that consecutive time windows would share 95%, 75%, 50% and 0% of their time points with each other, respectively.

4.2.1 Trial type 1

We start our observations by looking at Figure 4.7, which shows the metric correlations obtained based on different step sizes, for trial type 1.

By comparing the results of all the step sizes, it can be seen that there is a big cluster of positive correlations, which contains almost all metric types (except type 6), degree, closeness centrality and core number. AvgPL is also negatively correlated with all the members of this cluster. Note that the same cluster of recurrent correlations also appeared in Figure 4.1, accompanied by two other metrics, ACC and graph clique number. As it is seen in Figure 4.7, when the step size is set to 100, these two metrics do not appear as part of this cluster of correlations anymore. This sudden drop in the number of correlations associated with ACC and graph clique number, when setting the step size to 100, suggests that by increasing the step size and thus having less number of graphs depicting the topological transformation of the dynamic graph, some correlations can not be captured.

It can also be seen that ACC and graph clique number are always strongly correlated with each other and motif types 3 and 7. As the ACC of the network increases, bigger cliques form and the number of motif types 3 and 7 increase as well, because based on their structures (see Figure 2.1) they can be seen as the construction units of clusters and cliques in a network. On

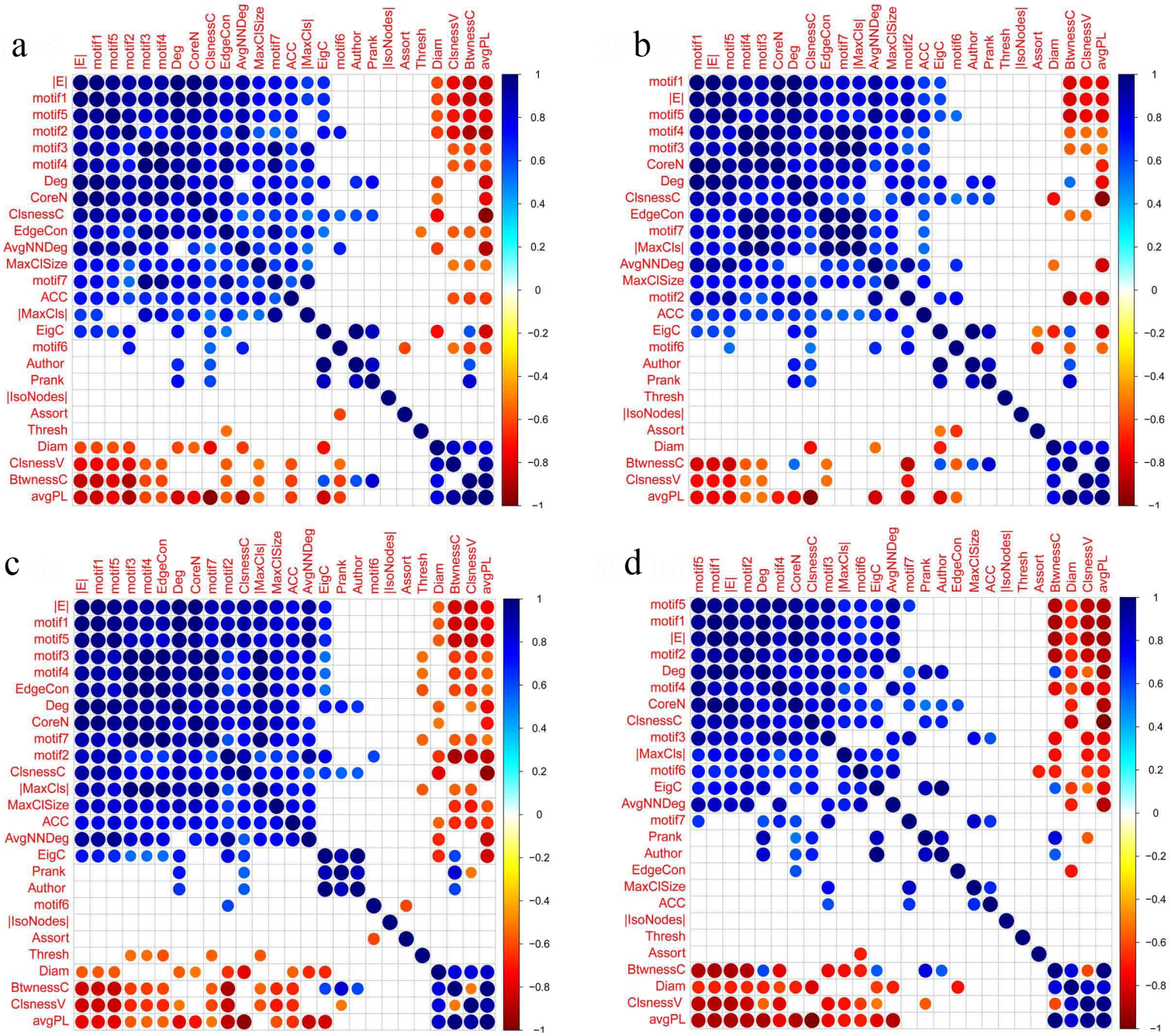

Figure 4.7: Metric correlations obtained based on trial type 1, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method.

the other hand, one would expect to also see that these metrics (ACC and graph clique number) are also correlated with closeness centrality and AvgPL, since these correlations were always present in previous observations. Furthermore, the mathematical definition for these metrics also suggests a correlation. The reason behind this observation is again related to the increased step size and thus poor depiction of topological transformation of the dynamic graph. This is further confirmed by the fact that AvgPL (and thus closeness centrality) is susceptible to disconnected graphs (explained in section 3.1), and therefore, if in one of the network's transformation stages the number of isolated nodes increases, it could affect their values, and thus their correlation with other metrics. This reasoning also applies for all similar future observations regarding the lack of expected metrics, when the step size is high, or when there is no shared time points between consecutive time windows.

Note that AvgNNDeg is also correlated with most members of the previously mentioned cluster, except for degree, closeness centrality and interestingly core number. This lack of correlation with core number is interesting because until now this correlation was present in all the previous observations. We can again relate this sudden lack of correlation to the lack of shared time points between two consecutive time windows, which leads to a sharp transition between different states of the network in time.

On the other hand the lack of correlation between AvgNNDeg, closeness centrality and degree has been frequently seen, and can imply that a central node is not always connected to other high degree nodes, which agrees with the previously explained rich-club phenomenon in brain networks (see section 3.1).

Eigenvector centrality, pagerank, authorities, degree and closeness centrality are correlated with each other, as always. Diameter and AvgPl's correlation with each other and with betweenness centrality and closeness vitality is another familiar cluster of correlations seen in

Figure 4.7.

It is interesting to see that edge connectivity is correlated with core number. This can easily be explained: as the average core number increases, nodes become more interconnected with each other, meaning that more number of edges should be removed to disconnect the graph.

Other positive significant correlations seen for all four step sizes are as follows: eigenvector centrality is correlated with motif types 2 and 5, number of edges and betweenness centrality. Also, number of maximal cliques is correlated with degree, closeness centrality, core number, AvgNNDeg and motif types 4 and 1 (number of edges).

By looking at the recurrent negative correlations, the relationship between assortativity and motif type 6 comes to attention. This recurrent correlation was also persistent for all window sizes in the trial type 1, in the previous section.

Though betweenness centrality and closeness vitality do not have any correlation with each other, they are both negatively correlated with motif types 1,2,3,4 and 5. It is intuitively understandable that as the number of these motifs increase the network becomes more interconnected, and thus the average length of the connections in the graph becomes shorter. In other words, more nodes are participating in forming the connectivity patterns in the graph.

4.2.2 Trial type 2

By looking at Figure 4.8, one can see that the cluster of positive correlations is bigger than the one seen for trial 1. This cluster shows that motif types 1,2,3,4,5 and 7, graph clique number,

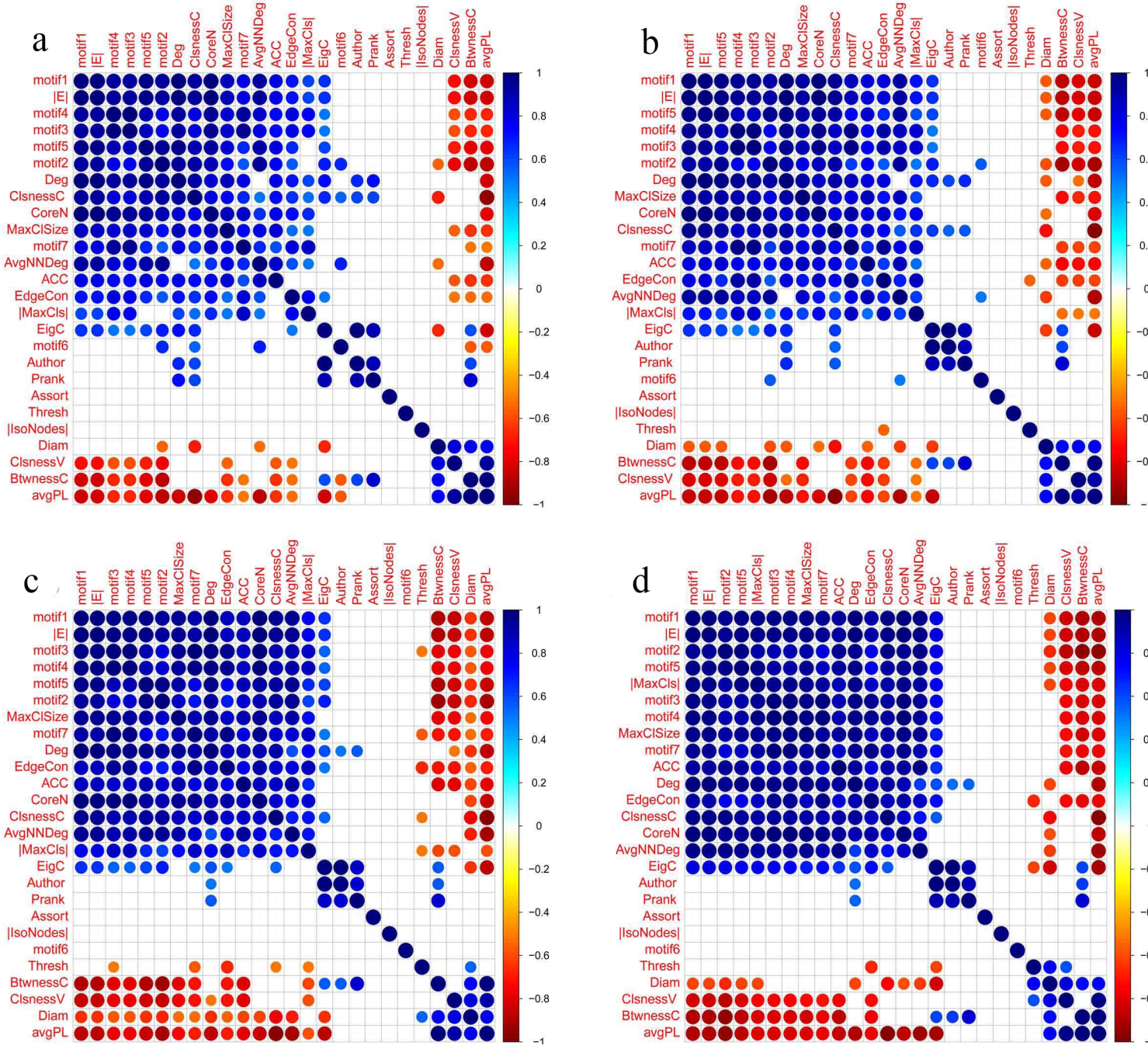

Figure 4.8: Metric correlations obtained based on trial type 2, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method.

core number, closeness centrality, degree and ACC are all correlated with each other, while AvgNNDeg, edge connectivity and number of maximal cliques are correlated with each other and most members of this cluster.

It is also interesting to see that closeness centrality is not correlated with authorities and page rank for most of the step sizes (see Figure 4.8). On the other hand, betweenness centrality is always correlated with authorities, pagerank and eigenvector centrality. Closeness vitality, betweenness centrality, diameter and AvgPL have the same relationships with each other as before.

While diameter has the same negative correlations as before, it is also negatively correlated with motif type 2.

Closeness vitality and betweenness centrality are both negatively correlated with motif types 1,2,3,4 and 5, edge connectivity, ACC and graph clique number. AvgPL is negatively correlated with all members of the big positive cluster of correlations except number of maximal cliques.

4.2.3 Trial type 3

Here again, we see a cluster of mostly familiar correlations (see Figure 4.9). part of this cluster includes motif types 1,2,3,4 and 5, degree, closeness and core number, which were also seen in previous trial types. For this trial, graph clique number and motif type 7 are also correlated with these metrics. AvgNNDeg, ACC and motif type 6 are correlated with each other, and some metrics of this cluster. All of these correlations can easily be understood based on their mathematical definitions, and the expected small-world properties of the network. Also, con-

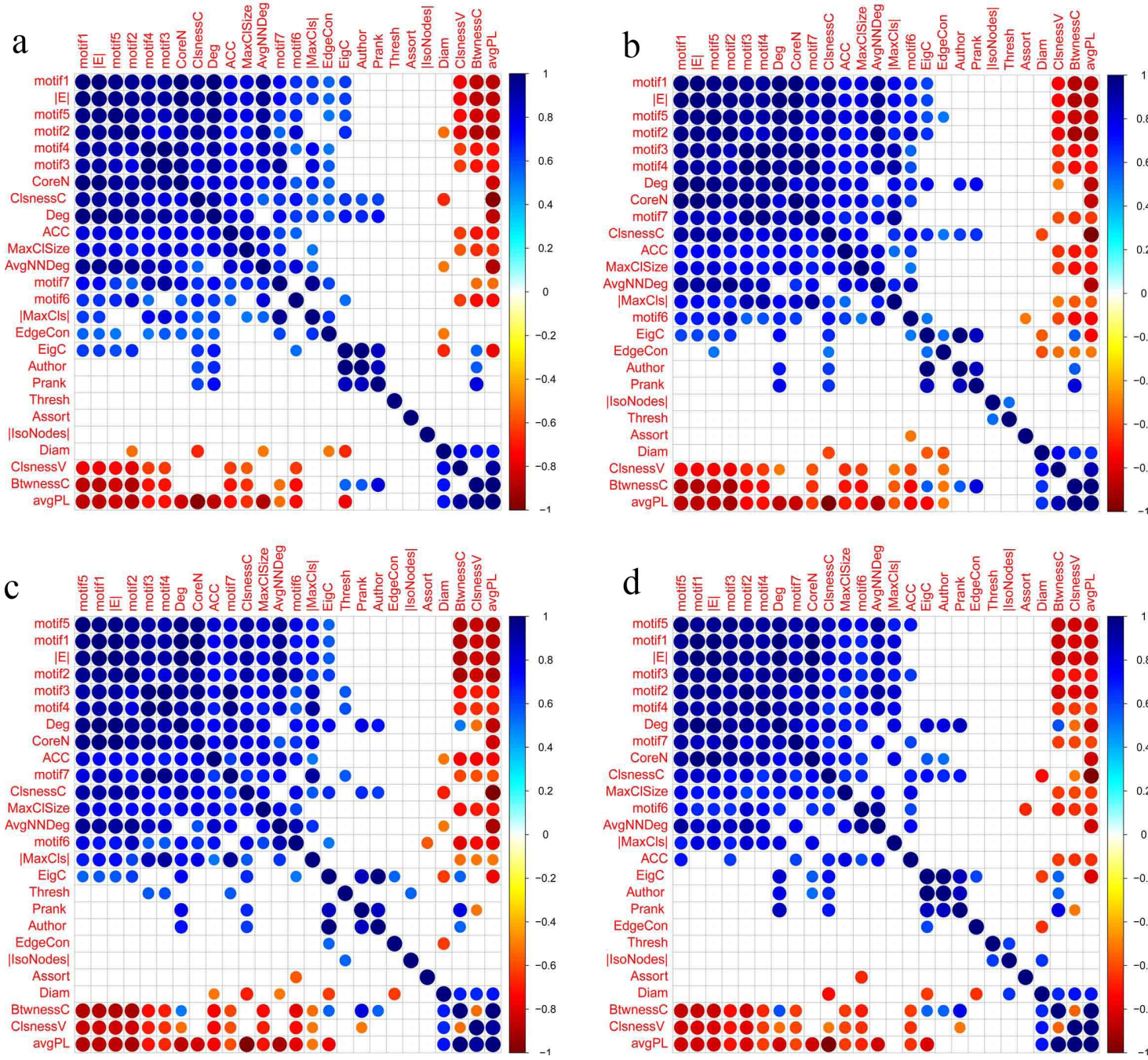

Figure 4.9: Metric correlations obtained based on trial type 3, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method.

sidering previous observations, it is interesting to see that motif type 6's number of correlations has increased.

Number of maximal cliques is positively correlated with motif types 1,3 and 4, degree and core number. Again, these correlations are all expected, because an increase in all of these metrics can lead to the creation of more cliques in the network.

All other positive and negative correlations are mostly the same as the ones seen for trial type 1.

4.2.4 Trial type 4

The big cluster of correlations that was seen for trial type 3, can also be seen for all step sizes of this trial type (see Figure 4.10). The only difference is that motif type 6 is not part of this cluster, and AvgNNDeg is correlated with core number. AvgPL is again negatively correlated with all members of this cluster.

Other positive correlations are the same as trial type 3.

If one disregards the lack of correlations regarding motif type 6, negative correlations regarding betweenness centrality, closeness vitality and AvgPL are also almost the same as the previous trial type. Note that diameter is negatively correlated with motif types 1,2 and 5, degree, AvgNNDeg, eigenvector centrality and closeness centrality.

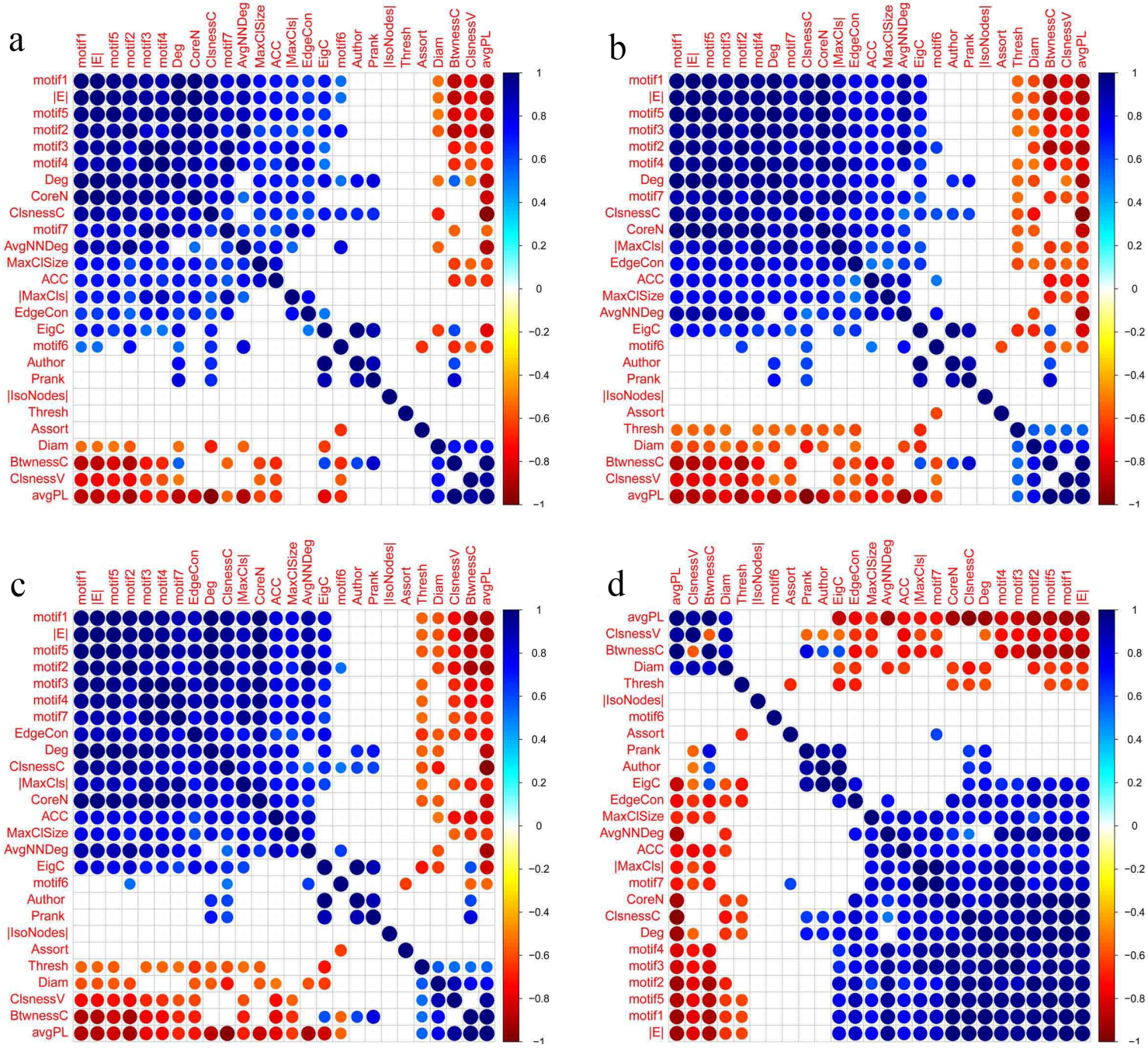

Figure 4.10: Metric correlations obtained based on trial type 4, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method.

4.2.5 Trial type 5

The biggest cluster of recurrent correlations for this trial type (see Figure 4.11) is the same as the one detected for trial type 1, with the addition correlations regarding motif type 7. AvgNNDeg and edge connectivity are also correlated with most metrics of this cluster, but AvgNNDeg is not correlated with degree and closeness centrality, while edge connectivity is not correlated with motif type 7. Note that ACC and graph clique number are always correlated with each other and most members of the cluster, except for motif type 2 and closeness centrality. Their lack of expected correlations with closeness centrality and AvgPL is again an effect of choosing a large step size, which was explained for trial type 1.

Other correlations are mostly the same as previous trials, but some interesting differences can be seen. There are some differences in the negative correlation of betweenness centrality and closeness vitality. Also, betweenness centrality is not correlated with eigenvector centrality and pagerank anymore. Other interesting correlations are regarding motif type 6. This motif is now negatively correlated with assortativity, betweenness centrality and AvgPL, while is positively correlated with motif type 2, closeness centrality and AvgNNDeg.

4.2.6 Recurrent dependencies

Based on the above observations, we can suggest the following metric dependency groups that are consistent across all trial types and step sizes (see Figure 4.12): (number of edges, motif types 2,3,4 and 5, degree, core number, closeness centrality, AvgPL), (AvgNNDeg, motif types 2,3,4 and 5, number of edges, AvgPL), (graph clique number, ACC, motif types 3 and 7), (motif type 3,4,5 and 7, core number, degree), (number of maximal cliques, number of edges, degree, core number, motif type 4), (eigenvector centrality, authorities, pagerank, de-

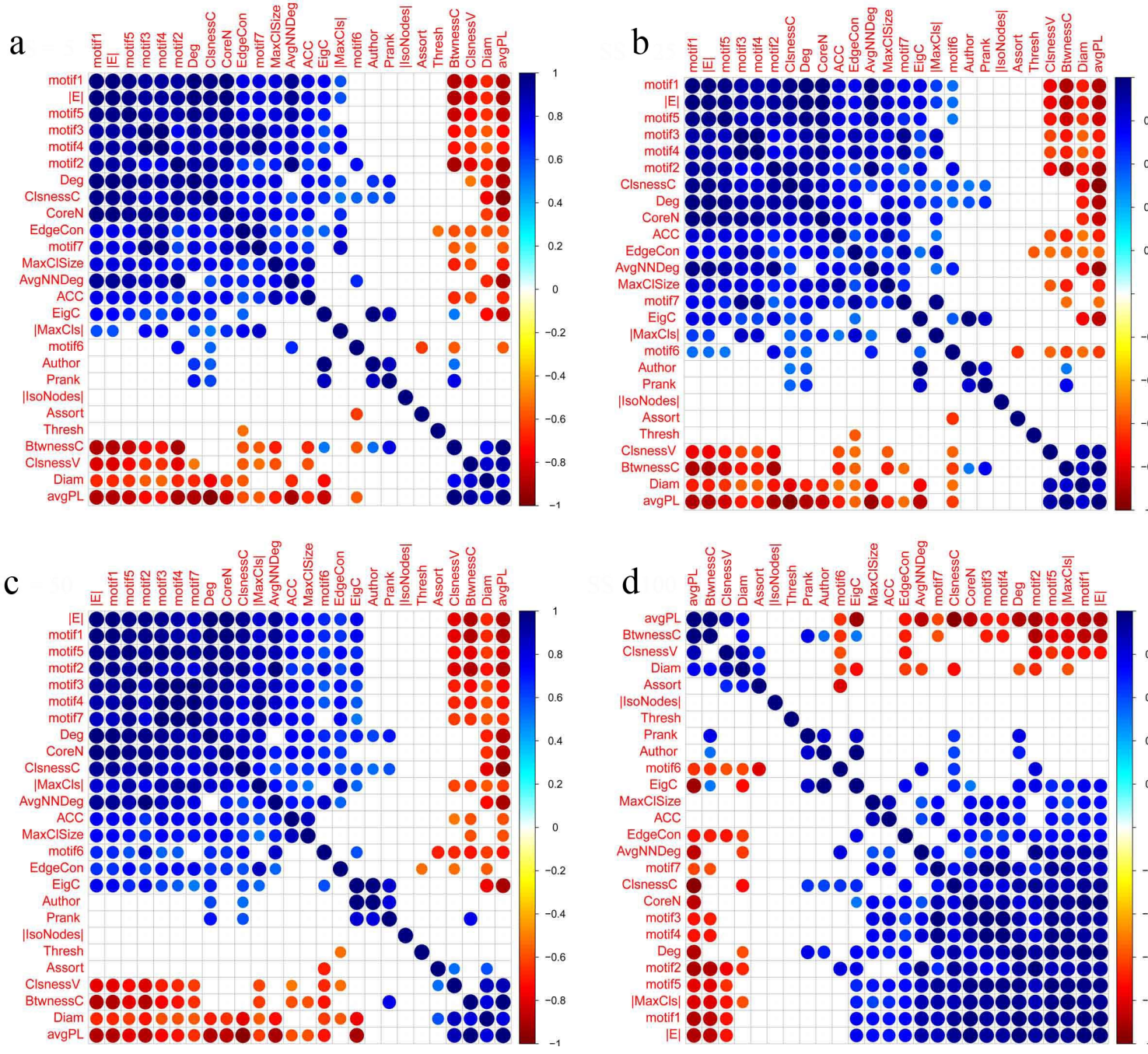

Figure 4.11: Metric correlations obtained based on trial type 5, in beta frequency band, and using step sizes = (a)5, (b)25, (c)50 and (d)100, window size:100 and RMT thresholding method.

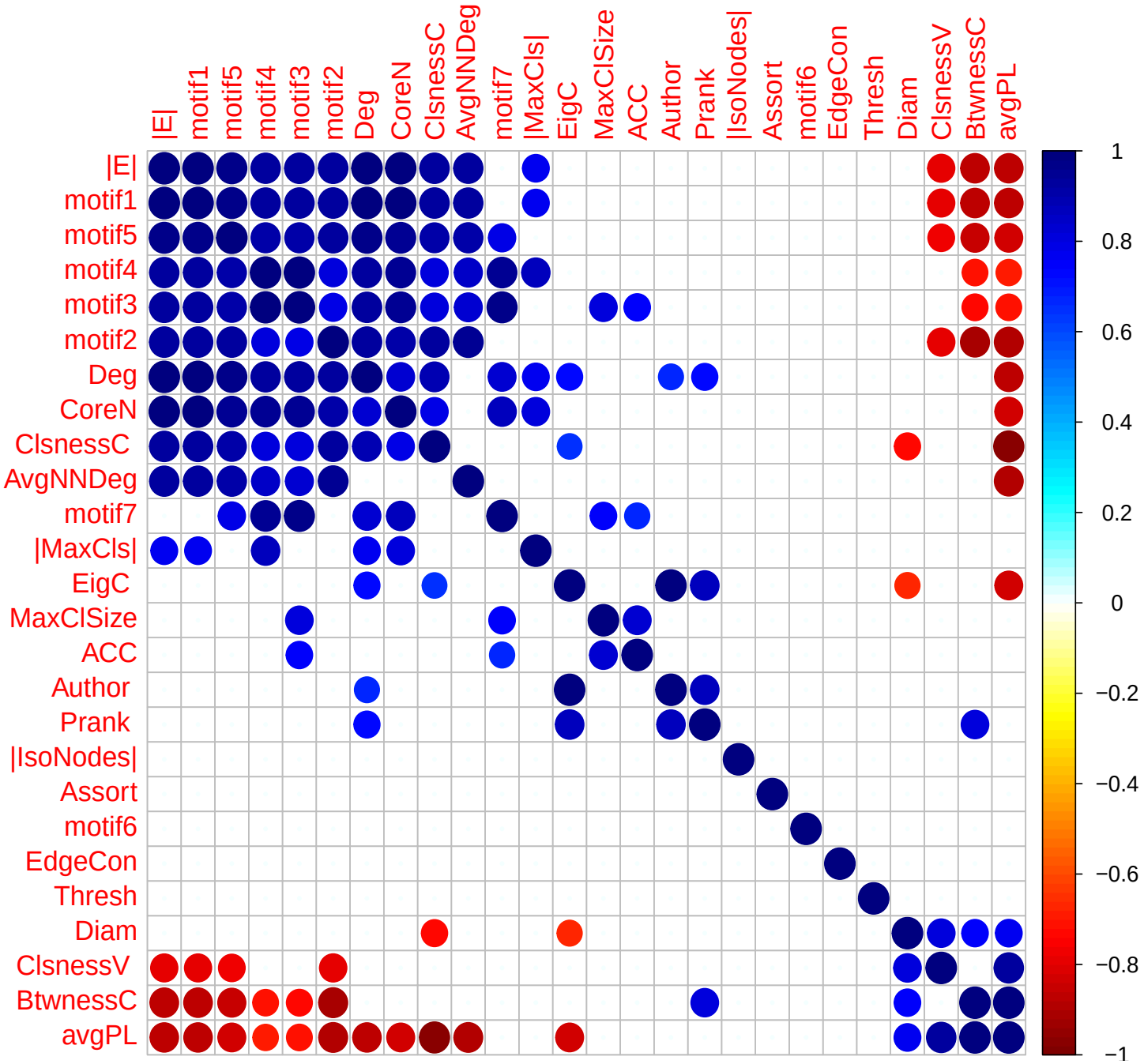

Figure 4.12: Recurrent metric correlations which were seen for all trial types and step sizes.

gree), (eigenvector centrality, closeness centrality, AvgPL, diameter), (pagerank, betweenness centrality), (AvgPL, diameter, closeness vitality), (AvgPL, diameter, betweenness centrality), (betweenness centrality, number of edges, motif types 2,3,4 and 5), (closeness vitality, number of edges, motif types 2 and 5).

4.3 Changing the frequency band

EEG-derived functional brain networks can be studied in different frequency bands. While it is more common to study neural synchronizations observed in beta band for adult subjects engaged in different cognitive tasks, brain functional networks are also studied in other frequency bands, especially alpha and gamma (e.g. [32], [49] and [30]). Therefore, it is interesting to also analyze metric correlations of dynamic graphs generated in these frequency bands.

All metric correlations in this section are obtained based on the following circumstances:

1- Using RMT thresholding method.
2- Step size = 25.
3- Window size = 800.

Three sets of dynamic graphs are generated based on three different frequency bands: alpha, beta and gamma.

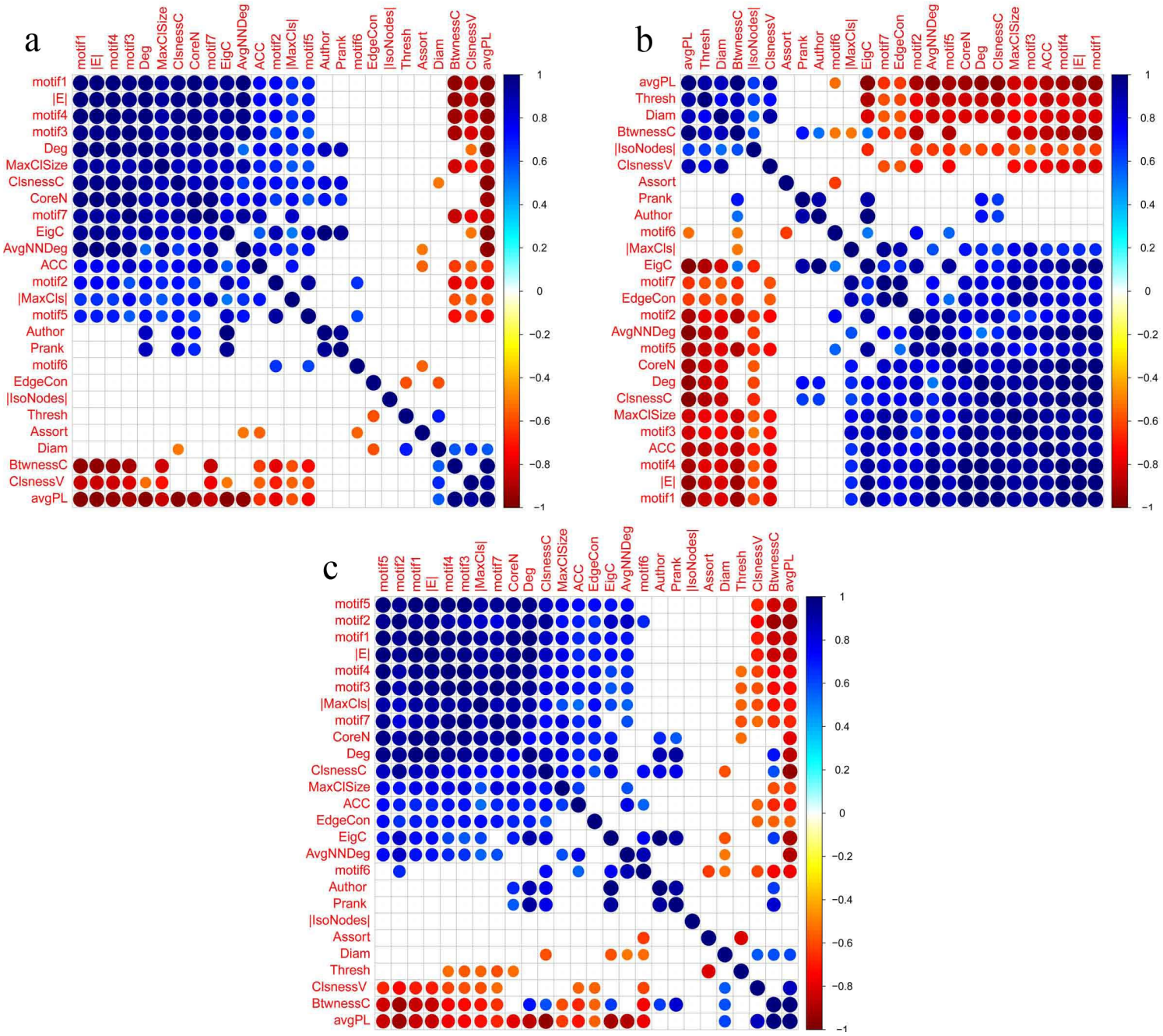

Figure 4.13: Metric correlations obtained based on trial type 1, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method.

4.3.1 Trial type 1

By looking at Figure 4.13, which shows metric correlations obtained based on alpha, beta and gamma band, we can spot some recurrent correlations that are present for all three bands.

The biggest cluster of recurrent and mostly familiar positive correlations includes number of edges, motif types 3, 4 and 7, degree, closeness centrality, graph clique number, core number, ACC and number of maximal cliques. Motif types 2 and 5 are also correlated with most members of this cluster (except motif type 7, ACC and number of maximal cliques). Also, AvgNNDeg participates in this cluster to some extent. Eigenvector centrality, pagerank, authorities, closeness centrality and degree have also gathered to form a cluster of strong correlations. Note that eigenvector centrality is also always correlated with motif types 2 and 5.

Furthermore, AvgPL, diameter, betweenness centrality and closeness vitality have the same correlations as seen in previous sections of this chapter.

It is very interesting to see that motif type 6 is positively correlated with motif type 2 and negatively correlated with assortativity. These correlations are always present for trial type 1, under any circumstances.

As for other recurrent negative correlations, one can see that AvgPL is again negatively correlated with almost all members of the big cluster of positive correlations (except number of maximal cliques). It is also negatively correlated with eigenvector centrality. Closeness vitality and betweenness centrality are both negatively correlated with motif types 1,2,3,4,5 and 7 and ACC, while betweenness centrality is also negatively correlated with clique-related metrics. These correlations imply that as the network becomes more clustered and interconnected, less number of nodes are involved in each of the graph's shortest paths.

We can also see the expected negative correlation of diameter with closeness centrality.

4.3.2 Trial type 2

The first noticeable observation for this trial type is that the big cluster of positive correlations seen in trial type 1 is also present here (see Figure 4.14). In addition to its previous members, it now includes edge connectivity, and to some extent eigenvector centrality. Note that AvgNNDeg is part of this cluster, but is not correlated with degree and eigenvector centrality, and eigenvector centrality is not correlated with number of maximal cliques. Other positive clusters of correlations are as follows: motif types 2 and 5 are correlated with number of edges, motif type 4, closeness centrality, degree, AvgNNDeg and eigenvector centrality. Motif type 5 is also correlated with ACC. Also, eigenvector centrality, authorities, pagerank, degree, closeness centrality are correlated with each other. Pagerank is also correlated with betweenness centrality. The relationship between AvgPL, diameter, closeness vitality and betweenness centrality is same as trial type 1.

When looking for recurrent negative correlations, it can be seen that comparing to the last trial's results, their number have increased. For example, diameter is now negatively correlated with all members of the big cluster of positive correlations, AvgPL has formed negative correlations with number of maximal cliques and edge connectivity, while both closeness vitality and betweenness centrality are now negatively correlated with clique-related metrics and edge connectivity. As it was mentioned before, though closeness vitality and betweenness centrality are never correlated with each other, they have many common negative correlations, which are mostly regarding the metrics that are related to the formation of clusters in the graph.

It is also interesting to see that assortativity is negatively correlated with motif type 6, for

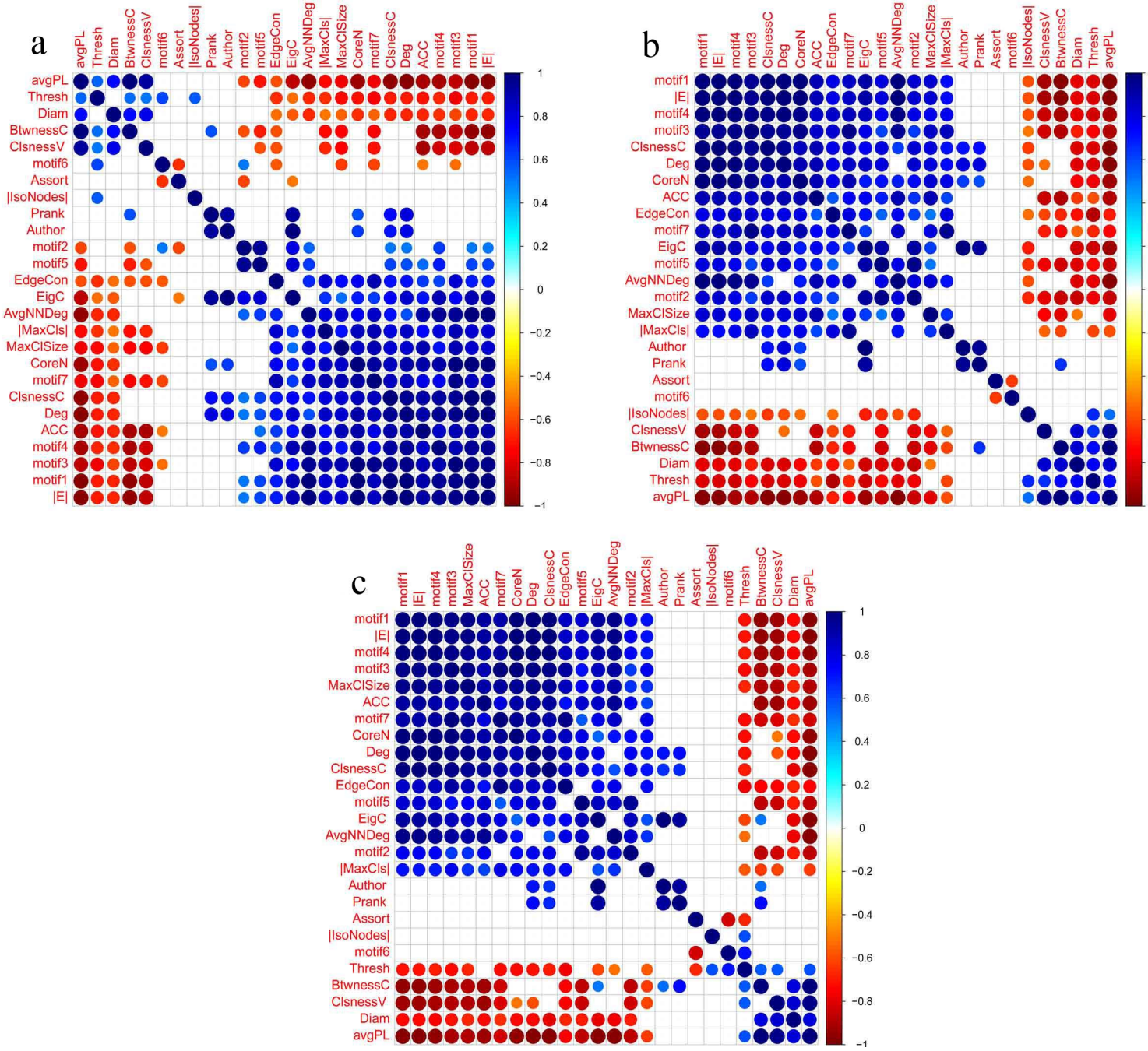

Figure 4.14: Metric correlations obtained based on trial type 2, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method.

all frequency bands.

4.3.3 Trial type 3

By comparing the metric correlations of three different frequency bands for this trial type, shown in Figure 4.15, we can see that recurrent positive correlations are almost the same as the ones that appeared for trial type 2. The differences are as follows: There is no correlation between pagerank and betweenness centrality, AvgNNDeg and degree are now correlated and most importantly, edge connectivity has no positive correlations with other metrics.

Recurrent negative correlations seen in this trial type are also mostly similar to the ones seen for trial type 2. Note that diameter is now negatively correlated with motif types 2 and 5, while it is not correlated with graph clique number anymore. Also, diameter is the only metric that is correlated with edge connectivity.

It is also worth mentioning that for this trial type, and using gamma as the frequency band, we get considerable number of correlations regarding motif type 6.

4.3.4 Trial type 4

In trial type 4 (Figure 4.16), number of edges, motif types 3 and 7, core number, degree, closeness centrality, ACC, graph clique number and edge connectivity build a cluster of strong correlations, which is seen for all frequency bands. Motif type 4, edge connectivity and eigenvector centrality are also correlated with most members of this cluster. Authorities, page rank, eigenvector centrality, degree and closeness centrality have also grouped together, while betweenness centrality and closeness vitality are again both correlated with AvgPL and diameter. Also, motif type 2 and 5 are strongly correlated. Based on the structure of these motifs this correlation is expected, but it is not clear that why other structurally related motifs, such as

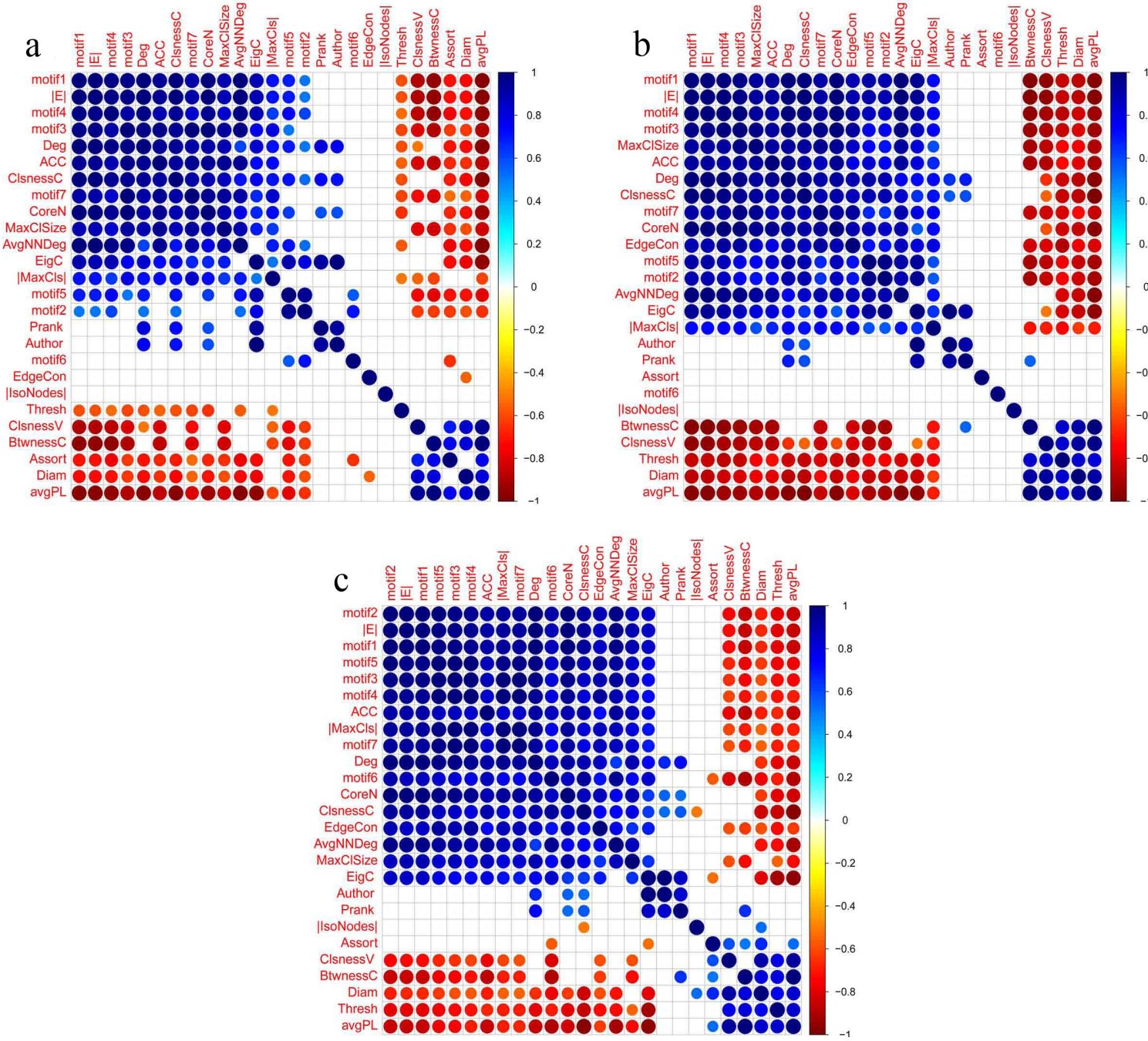

Figure 4.15: Metric correlations obtained based on trial type 3, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method.

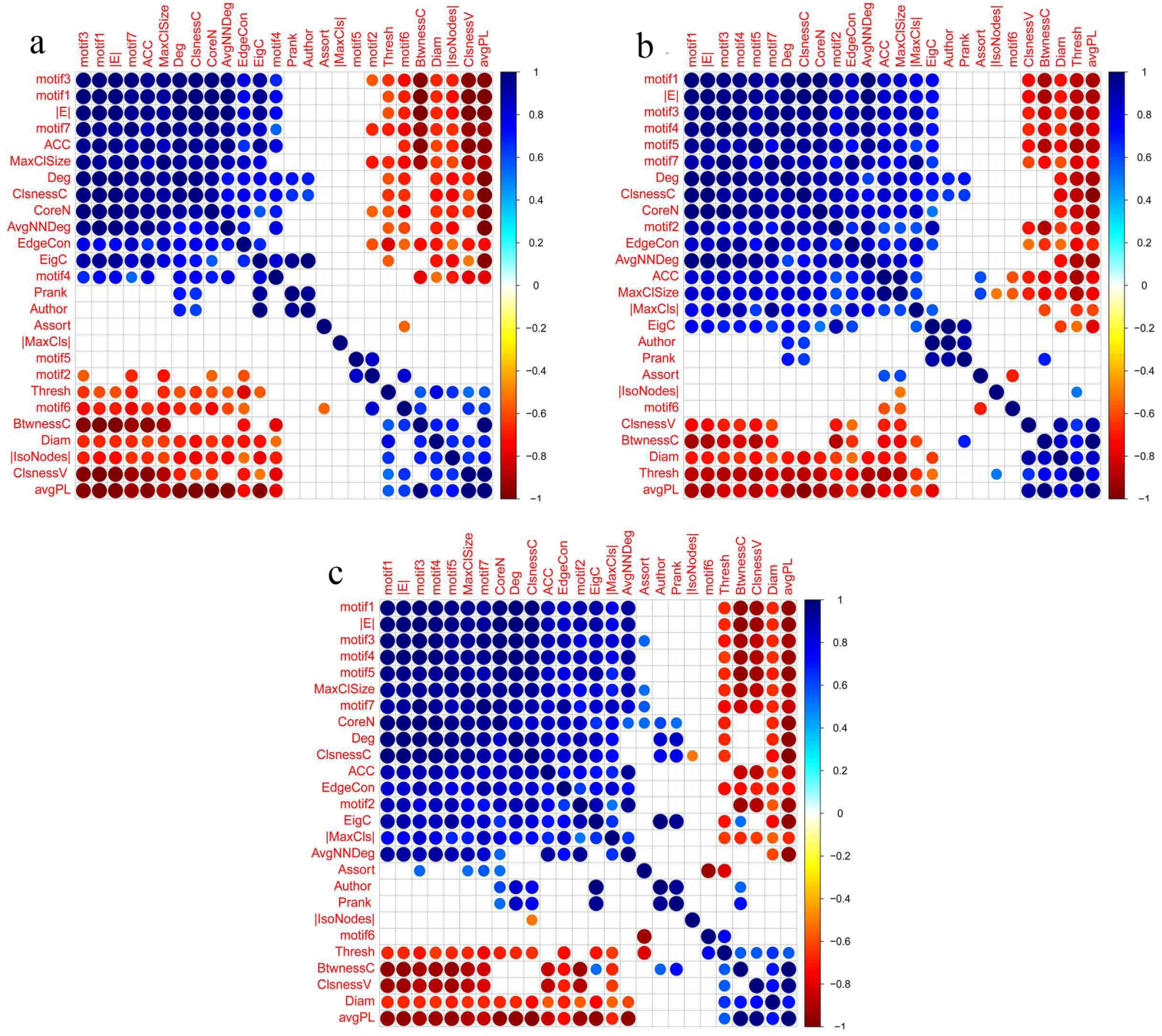

Figure 4.16: Metric correlations obtained based on trial type 4, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method.

types 4 and 5, or types 2 and 3, are not correlated.

Negative correlations that are seen for all frequency bands are as follows: AvgPL and diameter are both negatively correlated with all members of the previously mentioned, big cluster of positive correlations. Closeness vitality and betweenness centrality, though not correlated with each other, are both negatively correlated with motif types 1,3,4 and 7, edge connectivity, graph clique number and ACC. These correlations are again pointing to the fact that as the graph becomes more clustered, with bigger cliques, the communications between nodes become more efficient (AvgPL decreases), and it also becomes harder to disconnect the graph.

Note that motif type 6 and assortativity are also negatively correlated.

4.3.5 Trial type 5

This trial type's correlations (see Figure 4.17) are again mostly similar to previous trials'. Motif types 1,3, and 7, ACC, degree, closeness centrality, core number, graph clique number, AvgNNDeg and edge connectivity are all strongly correlated, for all frequency bands. Motif type 4, number of maximal cliques and eigenvector centrality are also correlated with most members of this cluster. Other groups of positive correlations are as follow: eigenvector centrality, authorities, pagerank, closeness centrality and degree are correlated with each other, AvgPL and diameter are both correlated with each other and betweenness centrality and closeness vitality, and finally, motif type 2 is correlated with motif type 5 and 6.

Closeness vitality, betweenness centrality, diameter and AvgPL have almost the same negative correlations as before.

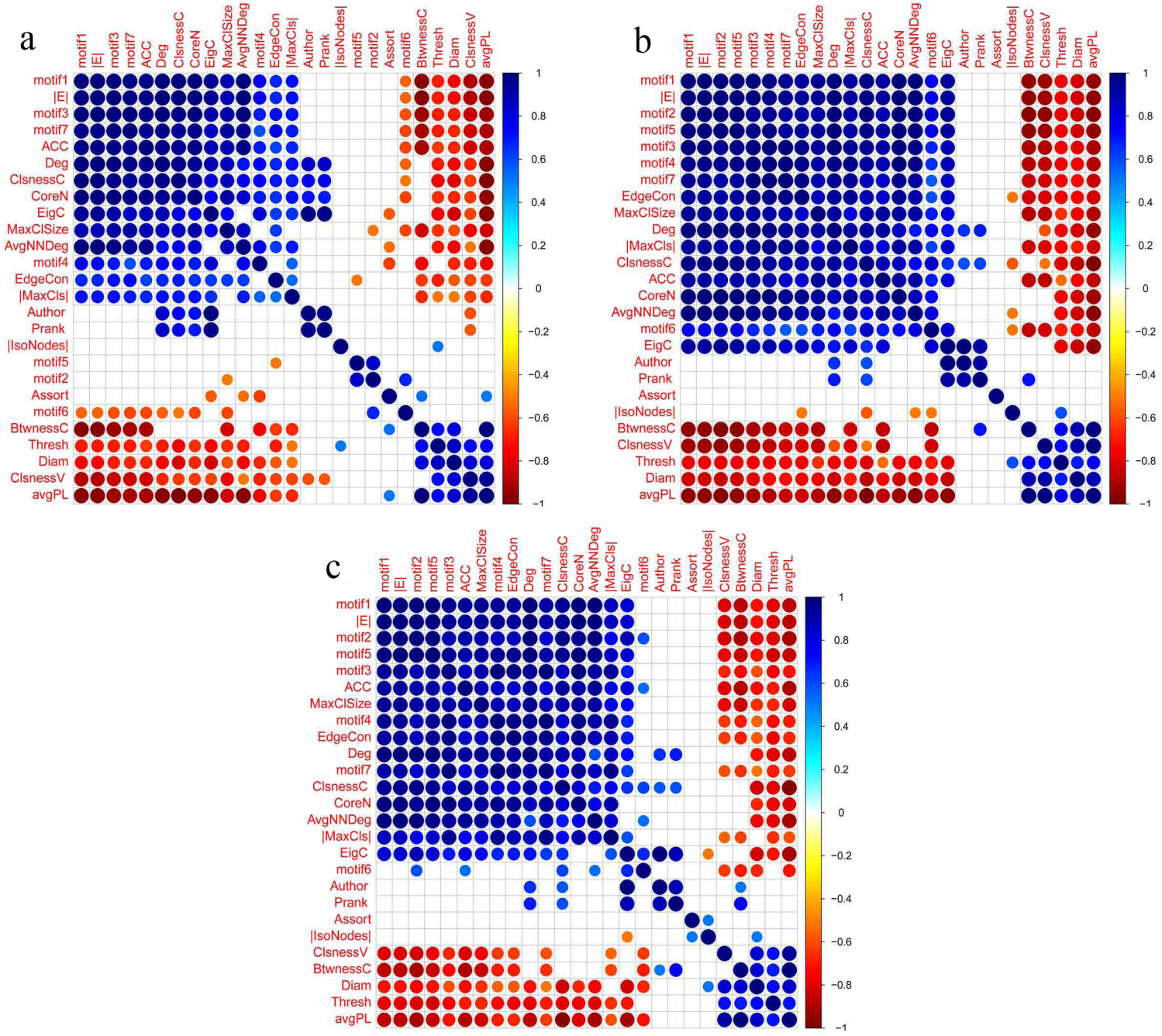

Figure 4.17: Metric correlations obtained based on trial type 5, in frequency bands: (a)alpha, (b)beta and (c)gamma, and using step size = 25, window size = 800 and RMT thresholding method.

4.3.6 Recurrent dependencies

Recurrent metric correlations for all trial types and frequency bands (see Figure 4.18) can be fitted into the following dependency groups: (number of edges, motif types 3,4 and 7, degree, closeness centrality, ACC, core number, AvgPL), (number of edges, motif types 3 and 7, degree, closeness centrality, ACC, core number, graph clique number, AvgPL), (AvgNNDeg, motif types 1,3,4 and 7, ACC, graph clique number, AvgPL), (eigenvector centrality, authorities, pagerank, degree, closeness centrality), (eigenvector centrality, motif 1,3 and 4, AvgPL), (motif 5, motif 2), (diameter, closeness centrality), (AvgPL, betweenness centrality, diameter), (AvgPL, diameter, closeness vitality), (closeness vitality, motif 1,3,4 and 7, ACC), (betweenness centrality, motif 1,3,4 and 7, ACC, graph clique number), (diameter, closeness centrality).

4.4 Discussion

The main goal of this chapter is to choose informative and orthogonal metrics, from a pool of 25 graph metrics, to reduce the computation time and effort for further brain's functional network analysis (based on EEG data-sets). Based on the analysis of metric correlations in all three sections, we can suggest a general guideline for choosing the most informative set of graph metrics for further network analysis. Based on a plotted correlation matrix, we choose metrics as follows:

1- All metrics that are not correlated with any other metric should be chosen.

2- From each dependency group, only one metric should be chosen. We suggest choosing the metric with the lowest computation time, while considering its correlation strength with other metrics in that dependency group.

By looking at the metric dependencies for each trial type, in each section, one can see that

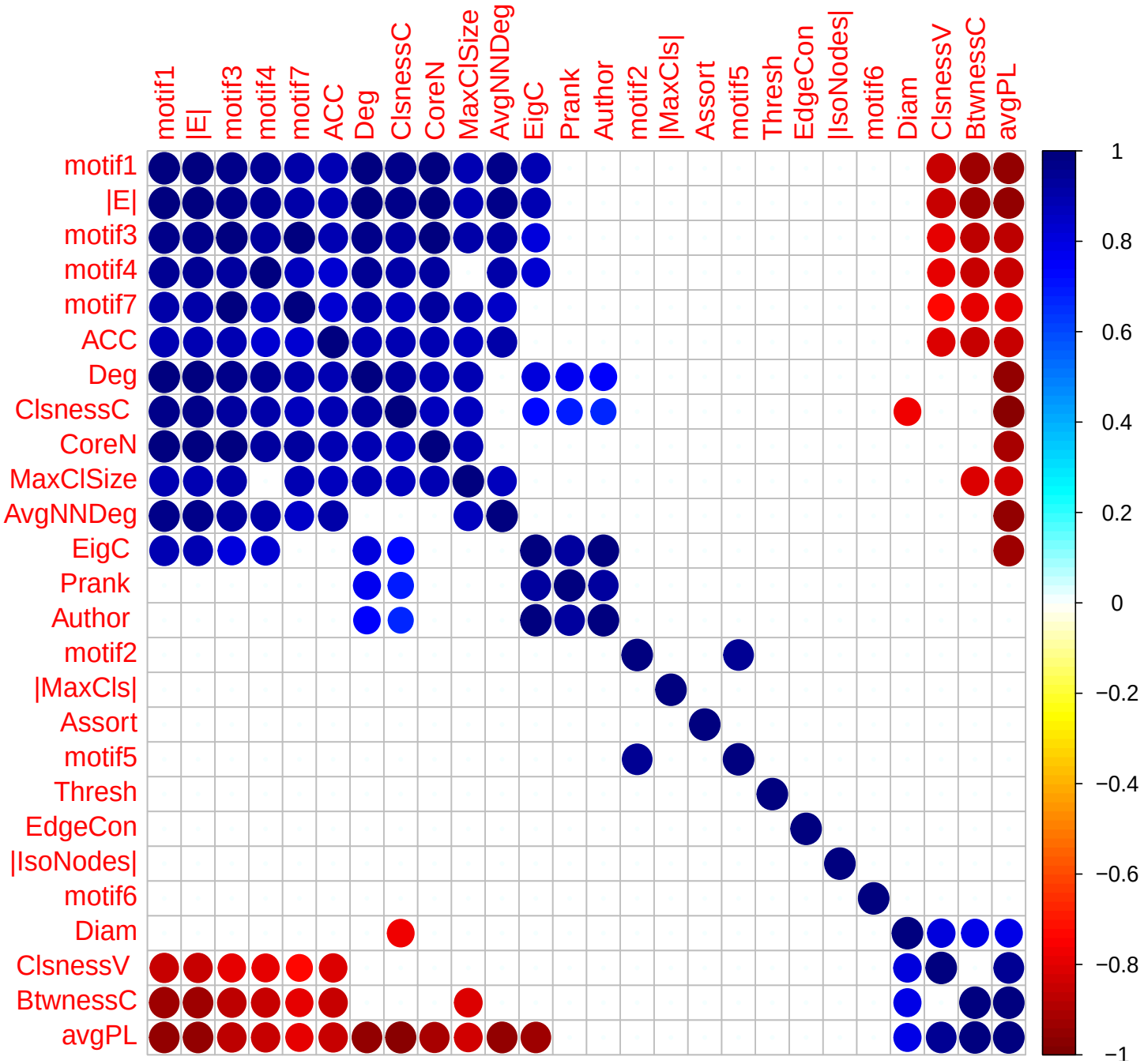

Figure 4.18: Recurrent metric correlations seen for all frequency bands and trial types.

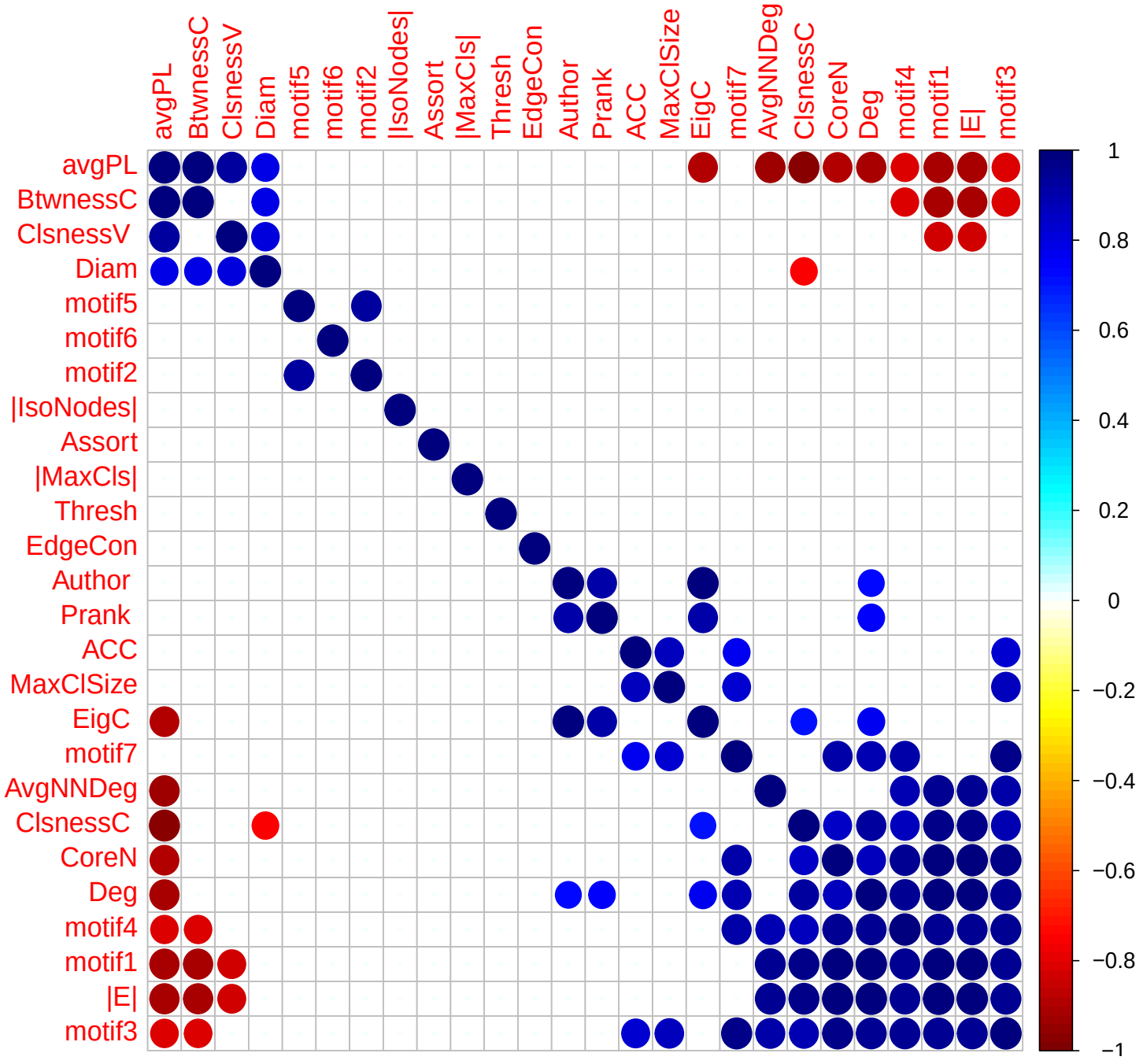

Figure 4.19: Recurrent metric correaltions seen for all combinations of frequency bands, step sizes, windows sizes and trial types.

there are a few metric correlations that can always be seen for a particular trial type, regardless of the chosen frequency band, window size and step size. For trial type 1, we can see that motif type 6 and assortativity are always negatively correlated. This is indeed interesting since in this trial type the subjects are not given any stimulus, or in other words, their brain functional network is at a resting state. When the brain is at rest, different regions make functional connections to each other and form several resting-state networks [58]. Moreover, it has been shown that the existence of the rich club phenomenon, which explains this negative correlation, is associated with several resting-state networks [59]. Thus, this particular correlation seems to be a fingerprint of the brain functional network at the resting-state.

Another interesting observation is regarding trial type 3, in which diameter and edge connectivity are always negatively correlated with each other. Based on the definitions of these metrics, this correlation is expected to be seen for all trial types. This is due to the fact that as nodes become more interconnected, the diameter decreases, and thus it would become harder to disconnect the graph. Saying this, it is not clear why this correlation is not always present for other trial types. It can be related to a specific pattern of functionality needed to perform this task, while we cannot rule out the possibility that the persistent appearance of this correlation for this trial type is merely accidental.

These trial-specific correlations could suggest that some relationships between metrics can only be seen for specific tasks, and therefore it might be beneficial to observe metric correlations for specific brain tasks and choose robust metrics for each trial type individually, to further decrease the future cost of computing repetitive metrics.

Another important point is regarding the number of shared time points ($windowsize - stepsize$) between consecutive time windows. Based on the observations made in section 4.2, when the value of the step size is high, capturing the topological transformation of the dynamic graph is

affected in a way that some expected and recurrent metric correlations are not captured.

Although it is preferred to choose robust metrics based on the specific circumstances in which the dynamic graphs where generated, we can suggest a pool of robust metrics based on the observations made in this chapter. This metrics are chosen from the following dependency groups:

(closeness centrality, core number, degree, number of edges, motif types 1 ,3 and 4, AvgPL), (closeness centrality, diameter), (AvgNNDeg, motif types 1,3 and 4, AvgPL), (motif types 3,4 and 7, degree, core number), (ACC, graph clique number, motif types 3,7), (authorities, pagerank, eigenvector centrality, degree), (eigenvector centrality, closeness centrality, degree, AvgPL), (closeness vitality, AvgPL, diameter), (betweenness centrality, AvgPL, diameter), (motif type 2, motif type 5), (betweenness centrality, AvgPL, motif types 1,3 and 4, number of edges), (closeness vitality, AvgPL, number of edges).

These dependency groups cover all metric correlations that were recurrently observed in all three sections, under all circumstances and for all trial types. Figure 4.18 is obtained by averaging common metric correlations seen in figures 4.1 to 4.17. Note that this figure is generated only with the purpose of easier detection of recurrent metrics and does not depict the accurate correlation values.

Based on these final dependency groups, and considering their required computation time discussed in Chapter 2, we suggest the following metrics as the most informative ones:

1. Diameter
2. Number of edges

3. Graph clique number
4. Eigenvector centrality
5. Motif type 5
6. Motif type 6
7. Number of isolated nodes
8. Assortativity
9. Number of maximal cliques
10. Edge connectivity

Thus, the pool of 25 metrics is reduced to 10 metrics, which are all uniquely informative and independent.

Chapter 5

Summary and suggestions for future work

The following contributions were made in this project:

1- Implementing a pipeline for obtaining the correlation matrix of 25 graph metrics, based on EEG-derived dynamic graphs. This pool of 25 metrics included 9 per-node and 9 whole-graph metrics, along side the counts of 7 motif types. Using this pipeline, dynamic graphs can be generated via S, FDR or RMT thresholding methods, in different frequency bands, and using different window sizes and step sizes.

2- Comparing metric correlations obtained based on S, FDR and RMT thresholding methods, by fixing the frequency band, window size and step size when generating the dynamic graphs. This comparison resulted in the following conclusions:

- Using S-thresholding method in this project yields mostly disconnected and relatively sparse graphs, which is due to the generation of high threshold values. Thus, not only some correlation values contradict the ones obtained based on the two other methods, but also comparing to two other thresholding methods, many metric correlations are not captured.

- Although FDR and RMT thresholding methods use substantially different approaches to

threshold a graph, their metric correlation results are very similar. Considering that FDR has been frequently used in the functional connectivity literature, this similarity of results supports choosing the newly introduced RMT thresholding method for further analysis in this project.

3- Performing a massive comparison of metric correlation results, obtained in different frequency bands, with different combinations of windows sizes and step sizes, which resulted in the following outcomes and conclusions:

- Reducing a pool of 25 metrics to a set of 10 orthogonal and uniquely informative metrics for using in future analysis of EEG-derived dynamic graphs, and thus, making future analysis less time consuming and more optimized.

- Detection of a trial-specific correlation, between motif type 6 and assortativity, which can be considered as a fingerprint of the resting state in the brain's functional connectivity.

- Concluding that lack of shared time points between consecutive time windows may result in the loss of important metric correlations, by masking some steps in the topological transformation of dynamic graphs.

Due to the limitation of time and resources in this project, and considering the heavy computation load of motif calculation process, we only investigated correlations for motifs having up to four nodes (seven total equivalence classes). Thus, our future goal for further extending the current work is to do an extensive metric correlation analysis in which the main focus would be on more variant and bigger motif types (having more than 4 nodes), to detect possible recurrent correlations regarding bigger motifs. This will benefit functional brain network analysis in two ways. First, by eliminating big motif types that offer no new insight than other metrics from the pool desired metrics, the analysis of functional brain networks will be further optimized. And second, the detection of possible recurrent correlations between some motif types with other per-vertex and whole graph metrics could help understanding the dynamic

mechanisms which lead to formation of a small-world functional brain network. For example, in the current project we could see the recurrent correlation between motif type 7 and metrics related to forming clusters in the graph. While these relationships can be easily understood based on the metric's mathematical definitions, bigger motif types participation in improving the network's efficiency is harder to grasp and thus would benefit from the approach used in this project.

Furthermore, the computation of bigger motif types can be done using a newly introduced motif detection algorithm by [43], which uses colored graphs, or in other words, labeled nodes and edges, for efficient motif detection in networks. Note that taking the individuality of nodes into account when finding network motifs is especially useful when investigating the brain functional connectivity, since each node represents an area in the brain. Thus, categorizing motifs by considering different brain regions participating in them can further shed light on the underlying mechanisms of the brain functional network. It is worth mentioning that this method has not yet been used for motif discovery in functional neural networks. We can also categorize the graph nodes based on their centrality degrees, especially betweenness centrality, so that highly central nodes are distinguished from other nodes in the graph. By labeling these nodes and using the efficient algorithm introduced by [43], we can study the formation of specific patterns of functionality by different types of nodes.

We also suggest further investigation of the effects of different step sizes and window sizes for building the functional brain networks. This can be done by taking the size of the brain signal into account, to see what combination of time points and step sizes can best capture the dynamic topological changes of the functional brain networks.

Bibliography

[1] Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and ED Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience*, 26(1):63–72, 2006.

[2] Alex Arenas, Albert Díaz-Guilera, and Conrad J Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Physical review letters*, 96(11):114102, 2006.

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

[4] Danielle S Bassett, Andreas Meyer-Lindenberg, Sophie Achard, Thomas Duke, and Edward Bullmore. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences*, 103(51):19518–19523, 2006.

[5] Vladimir Batagelj and Matjaz Zaversnik. An o (m) algorithm for cores decomposition of networks. *arXiv preprint cs/0310049*, 2003.

[6] Stephan Bialonski and Klaus Lehnertz. Assortative mixing in functional brain networks during epileptic seizures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(3):033139, 2013.

[7] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. O'Reilly Media, Inc., 2009.

[8] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. Network analysis in the social sciences. *science*, 323(5916):892–895, 2009.

[9] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[10] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[11] Geoffrey Canright and Kenth Engø-Monsen. Roles in networks. *Science of Computer Programming*, 53(2):195–214, 2004.

[12] Su Cheng, Pan YunTao, Yuan JunPeng, Guo Hong, Yu ZhengLu, and Hu ZhiYu. Pagerank, hits and impact factor for journal ranking. 6:285–290, 2009.

[13] Drew Conway. Modeling network evolution using graph motifs. *arXiv preprint arXiv:1105.0902*, 2011.

[14] M Agustina Garcés Correa and Eric Laciar Leber. Noise removal from eeg signals in polisomnographic records applying adaptive filters in cascade. 2011.

[15] Mark Daely. *Optimizing voxel scale graph theoretical analysis of fMRI-derived resting state functional connectivity*. The School of Graduate and Postdoctoral Studies, The university of Western Ontario, 2011.

[16] Mark Daley. *An Invitation to the Study of Brain Networks, with Some Statistical Analysis of Thresholding Techniques*. Springer, 2014.

[17] Willem de Haan, Yolande AL Pijnenburg, Rob LM Strijers, Yolande van der Made, Wiesje M van der Flier, Philip Scheltens, and Cornelis J Stam. Functional neural network analysis in frontotemporal dementia and alzheimer's disease using eeg and graph theory. *BMC neuroscience*, 10(1):101, 2009.

[18] Freeman J Dyson. Correlations between eigenvalues of a random matrix. *Communications in Mathematical Physics*, 19(3):235–250, 1970.

[19] Damien A Fair, Alexander L Cohen, Jonathan D Power, Nico UF Dosenbach, Jessica A Church, Francis M Miezin, Bradley L Schlaggar, and Steven E Petersen. Functional brain networks develop from a local to distributed organization. *PLoS computational biology*, 5(5):e1000381, 2009.

[20] F De Vico Fallani, Vito Latora, Laura Astolfi, Febo Cincotti, Donatella Mattia, Maria Grazia Marciani, Serenella Salinari, Alfredo Colosimo, and Fabrizio Babiloni. Persistent patterns of interconnection in time-varying cortical networks estimated from high-resolution eeg recordings in humans during a simple motor act. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224014, 2008.

[21] Norman Fenton and Martin Neil. *Risk assessment and decision analysis with bayesian networks*. CRC Press, 2012.

[22] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

[23] Norman Geschwind. Disconnexion syndromes in animals and man. *Brain*, 88(3):585–585, 1965.

[24] Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.

[25] Hao Guo, Xiaohua Cao, Zhifen Liu, Haifang Li, Junjie Chen, and Kerang Zhang. Machine learning classifier using abnormal brain network topological metrics in major depressive disorder. *Neuroreport*, 23(17):1006–1011, 2012.

[26] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Laboratory (LANL), 2008.

[27] Pascal Held, Christian Moewes, Christian Braune, Rudolf Kruse, and Bernhard A Sabel. Advanced analysis of dynamic graphs in social and neural networks. pages 205–222, 2013.

[28] Tom Howley, Michael G Madden, Marie-Louise OConnell, and Alan G Ryder. The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data. *Knowledge-Based Systems*, 19(5):363–370, 2006.

[29] Eric Jones, Travis Oliphant, and Pearu Peterson. Scipy: Open source scientific tools for python, 2001.

[30] Manfred G Kitzbichler, Richard NA Henson, Marie L Smith, Pradeep J Nathan, and Edward T Bullmore. Cognitive effort drives workspace configuration of human brain functional networks. *The Journal of Neuroscience*, 31(22):8259–8270, 2011.

[31] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[32] Jean-Philippe Lachaux, Nathalie George, Catherine Tallon-Baudry, Jacques Martinerie, Laurent Hugueville, Lorella Minotti, Philippe Kahane, and Bernard Renault. The many faces of the gamma band response to complex visual stimuli. *Neuroimage*, 25(2):491–501, 2005.

[33] Anton Lord, Dorothea Horn, Michael Breakspear, and Martin Walter. Changes in community structure of resting state functional connectivity in unipolar depression. *PloS one*, 7(8):e41282, 2012.

[34] David Meunier, Renaud Lambiotte, and Edward T Bullmore. Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4, 2010.

[35] Christoph M Michel, Micah M Murray, Göran Lantz, Sara Gonzalez, Laurent Spinelli, and Rolando Grave de Peralta. Eeg source imaging. *Clinical neurophysiology*, 115(10):2195–2222, 2004.

[36] Sifis Micheloyannis, Ellie Pachou, Cornelis Jan Stam, Michael Breakspear, Panagiotis Bitsios, Michael Vourkas, Sophia Erimaki, and Michael Zervakis. Small-world networks and disturbed functional connectivity in schizophrenia. *Schizophrenia research*, 87(1):60–66, 2006.

[37] Sanjit Kumar Mitra and Yonghong Kuo. *Digital signal processing: a computer-based approach*, volume 2. McGraw-Hill New York, 2006.

[38] Ruth M Nicol, Sandra C Chapman, Petra E Vértes, Pradeep J Nathan, Marie L Smith, Yury Shtyrov, and Edward T Bullmore. Fast reconfiguration of high-frequency brain networks in response to surprising changes in auditory input. *Journal of neurophysiology*, 107(5):1421, 2012.

[39] Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006.

[40] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[41] Satu Palva, Simo Monto, and J Matias Palva. Graph properties of synchronized cortical networks during visual working memory maintenance. *Neuroimage*, 49(4):3257–3268, 2010.

[42] Karl Pearson. *Note on regression and inheritance in the case of two parents*, volume 58. The Royal Society, 1895.

[43] Pedro Ribeiro and Fernando Silva. Discovering colored network motifs. In *Complex Networks V*, pages 107–118. Springer, 2014.

[44] Alard Roebroeck, Elia Formisano, and Rainer Goebel. Mapping directed influence over the brain using granger causality and fmri. *Neuroimage*, 25(1):230–242, 2005.

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

[46] Ankoor S Shah, Steven L Bressler, Kevin H Knuth, Mingzhou Ding, Ashesh D Mehta, Istvan Ulbert, and Charles E Schroeder. Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral Cortex*, 14(5):476–483, 2004.

[47] Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *Neuroimage*, 54(2):875–891, 2011.

[48] Olaf Sporns, Giulio Tononi, and Gerald M Edelman. Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex*, 10(2):127–141, 2000.

[49] CJ Stam, W De Haan, A Daffertshofer, BF Jones, I Manshanden, AM van Cappellen van Walsum, T Montez, JPA Verbunt, JC De Munck, BW Van Dijk, et al. Graph theoretical analysis of magnetoencephalographic functional connectivity in alzheimer's disease. *Brain*, 132(1):213–224, 2009.

[50] CJ Stam, BF Jones, G Nolte, M Breakspear, and Ph Scheltens. Small-world networks and functional connectivity in alzheimer's disease. *Cerebral Cortex*, 17(1):92–99, 2007.

[51] Cornelis J Stam, Guido Nolte, and Andreas Daffertshofer. Phase lag index: assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources. *Human brain mapping*, 28(11):1178–1193, 2007.

[52] Klaas E Stephan, Claus-C Hilgetag, Gully APC Burns, Marc A O'Neill, Malcolm P Young, and Rolf Kotter. Computational analysis of functional connectivity between areas of primate cerebral cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1393):111–126, 2000.

[53] Kaustubh Supekar, Vinod Menon, Daniel Rubin, Mark Musen, and Michael D Greicius. Network analysis of intrinsic functional brain connectivity in alzheimer's disease. *PLoS computational biology*, 4(6):e1000100, 2008.

[54] Qawi K Telesford, Ashley R Morgan, Satoru Hayasaka, Sean L Simpson, William Barret, Robert A Kraft, Jennifer L Mozolic, and Paul J Laurienti. Reproducibility of graph metrics in fmri networks. *Frontiers in NEUROINFORMATICS*, 4:117, 2010.

[55] Michal Teplan. Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11, 2002.

[56] Etsuji Tomita, Akira Tanaka, and Haruhisa Takahashi. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science*, 363(1):28–42, 2006.

[57] Etsuji Tomita, Akira Tanaka, and Haruhisa Takahashi. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science*, 363(1):28–42, 2006.

[58] Martijn P van den Heuvel, Rene CW Mandl, Rene S Kahn, Hulshoff Pol, and E Hilleke. Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human brain mapping*, 30(10):3127–3141, 2009.

[59] Martijn P van den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *The Journal of neuroscience*, 31(44):15775–15786, 2011.

[60] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[61] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[62] Eugene P Wigner. Random matrices in physics. *Siam Review*, 9(1):1–23, 1967.

[63] Shan Yu, Debin Huang, Wolf Singer, and Danko Nikolić. A small world of neuronal synchrony. *Cerebral Cortex*, 18(12):2891–2901, 2008.

[64] Pavel Zahradnik and Miroslav VLČEK. Notch filtering suitable for real time removal of power line interference. *Radioengineering*, 22(1), 2013.

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

[66] Bin Zhang, Steve Horvath, et al. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4(1):1128, 2005.

[67] Hong-Ying Zhang, Shi-Jie Wang, Jiong Xing, Bin Liu, Zhan-Long Ma, Ming Yang, Zhi-Jun Zhang, and Gao-Jun Teng. Detection of pcc functional connectivity characteristics in resting-state fmri in mild alzheimers disease. *Behavioural brain research*, 197(1):103–108, 2009.

Appendix A

Glossary of graph metrics

Table A.1: Glossary of graph metrics

Abbreviation	**Metric name**	**Short Description**
Assort	Assortativity	The extent to which nodes in a graph connect to other nodes with similar degree.
|MaxCls|	Number of maximal cliques	Number of cliques of largest size in a graph.
ACC	Average clustering coefficient	A measure of the degree to which nodes in a graph tend to cluster together.
AvgPL	Average path length	Average of all shortest paths between all nodes in a graph.
Diam	Diameter	Size of the longest path in a graph.
|E|	Number of edges	Number of edges in a graph.
MaxClSize	Graph clique number	The number of vertices in a maximum clique.
| IsoNodes |	Number of isolated nodes	Number of vertices that have no connections with other vertices in a graph.
EdgeCon	Edge connectivity	Minimum number of edges that should be removed to disconnect a graph.
CoreN	Core number	The core number of a node is the largest value k of a $k-core$ containing that node. A K-core is a maximal subgraph that contains nodes of degree k or more.
Author	Authorities	An estimation of the node's value based on the incoming connections.
ClsnessV	Closeness vitality	The change in the sum of distances between all pairs of vertices when excluding that node.
ClsnessC	Closeness centrality	A measure of the closeness of a vertex to all other vertices in a graph.
BtwnessC	Betweenness centrality	Gives the number of shortest paths from all vertices to all others, that pass through a certain vertex.
EigC	Eigenvector centrality	A measure for centrality that takes the importance of the neighbors of a vertex into account.
Prank	Pagerank	A measure of the importance of a vertex based on the structure of its incoming connections.
Deg	Node degree	Number of connections for each vertex.
AvgNNDeg	Average nearest neighbor degree	The mean of degrees of all the immediate neighbors of a vertex.
Motif1 to 7	Seven types of motif	See figure 2.1.

Curriculum Vitae

Name:	Mehrsasadat Golestaneh
Degrees:	Amirkabir University of Technology, 2007-2012, B.Sc. Computer Science
Related Work Experience:	Teaching Assistant The University of Western Ontario 2012 - 2014

Publications:

No.