

Capstone Project Phase B

**Applying graph theory measures for analyzing networks in brain recordings.**

**PROJECT CODE:**

**24-1-R-6**

**Supervisor:**

**Dr. Anat Dahan**

**Authors:**

**Koral Pnina Biton, ID: 314860057**

[coralbiton20@gmail.com](mailto:coralbiton20@gmail.com)

**Topaz Eldori, ID: 319003141**

[eldoritopaz85@gmail.com](mailto:eldoritopaz85@gmail.com)**‬**

***TABLE OF CONTENTS***

[*1. Abstract 3*](#_Toc176943047)

[*2. Introduction 4*](#_Toc176943048)

[*3. Project Review and Process Description 5*](#_Toc176943049)

[*4. Solution: 5*](#_Toc176943050)

[*5. Description of the Research/Development Process 11*](#_Toc176943051)

[*6. Tools Used and Interface with Client 12*](#_Toc176943052)

[*7. Challenges and Solutions 13*](#_Toc176943053)

[*8. Results and Conclusions 14*](#_Toc176943054)

[*9. Lessons Learned 25*](#_Toc176943055)

[*10. User Guide 26*](#_Toc176943056)

[*11. Maintenance Guide 32*](#_Toc176943057)

[*12. Reference: 34*](#_Toc176943058)

# Abstract

*This project explores the application of graph theory measures to analyze brain networks using EEG recordings. It investigates differences in information pathways within brain networks between children with attention deficit hyperactivity disorder (ADHD) and healthy subjects. EEG recordings from both groups during an attentional visual task were analyzed to calculate effective connectivity. Group differences in connectivity between brain regions were evaluated, revealing disrupted patterns of information flow in ADHD children compared to healthy subjects. Specifically, differences were observed in posterior to anterior information flow in the theta band and in pathways between anterior regions in the beta band. The study sheds light on the altered information flow patterns in ADHD children compared to their healthy counterparts, providing potential insights into the neural mechanisms underlying ADHD.*

**Keywords***: Hyperactivity Disorder (ADHD), Electroencephalography (EEG), Brain networks, Graph theory, Functional connectivity, Information pathways, Neurodevelopmental disorders, Computational neuroscience, Therapeutic targets, Community detection, Network analysis, Information processing, Cognitive neuroscience, Neural correlates.*

# Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder characterized by symptoms of inattention, hyperactivity, and impulsivity, often persisting into adulthood. With a prevalence rate of approximately 5% among children worldwide, ADHD represents one of the most common psychiatric disorders diagnosed in childhood. The disorder can have profound implications for academic performance, social interactions, and daily functioning, underscoring the importance of early diagnosis and intervention. **שגיאה! מקור ההפניה לא נמצא.**]

Electroencephalography (EEG) has emerged as a valuable tool in the assessment of brain function, offering insights into neural activity through the measurement of electrical signals generated by neurons. EEG recordings provide a non-invasive and objective means of evaluating brain activity, facilitating the identification of aberrant patterns associated with neurological disorders such as ADHD. By capturing the temporal dynamics of neuronal oscillations, EEG enables researchers to investigate the functional connectivity and information processing within brain networks. Graph theory offers a powerful framework for the analysis of complex networks, including those derived from EEG recordings. In the context of brain networks, nodes represent distinct brain regions, while edges denote functional or structural connections between these regions. Graph theory measures, such as clustering coefficient, global efficiency, and modularity, provide quantitative descriptors of network topology, illuminating the organization and dynamics of brain networks.

The integration of EEG recordings and graph theory analysis holds promise for elucidating the neural correlates of ADHD. By examining differences in brain network architecture between individuals with ADHD and neurotypical controls, researchers can gain valuable insights into the underlying pathophysiology of the disorder. Specifically, alterations in information processing, connectivity patterns, and network dynamics may be identified, providing a basis for understanding the cognitive and behavioral manifestations of ADHD.

This project endeavors to contribute to the burgeoning field of computational neuroscience by applying graph theory measures to analyze brain networks in individuals with ADHD. Through the investigation of EEG recordings obtained during attentional tasks, the study aims to delineate distinct information pathways within brain networks and elucidate the disruptions associated with ADHD. During our investigations we provide a comprehensive overview of the background literature pertaining to ADHD recognition, EEG brain recording techniques, the relationship between graph theory and brain networks, and existing methods for community detection in network analysis. We also outline the expected achievements of this research project and delineate the engineering process involved in data acquisition, preprocessing, and analysis.

# Project Review and Process Description

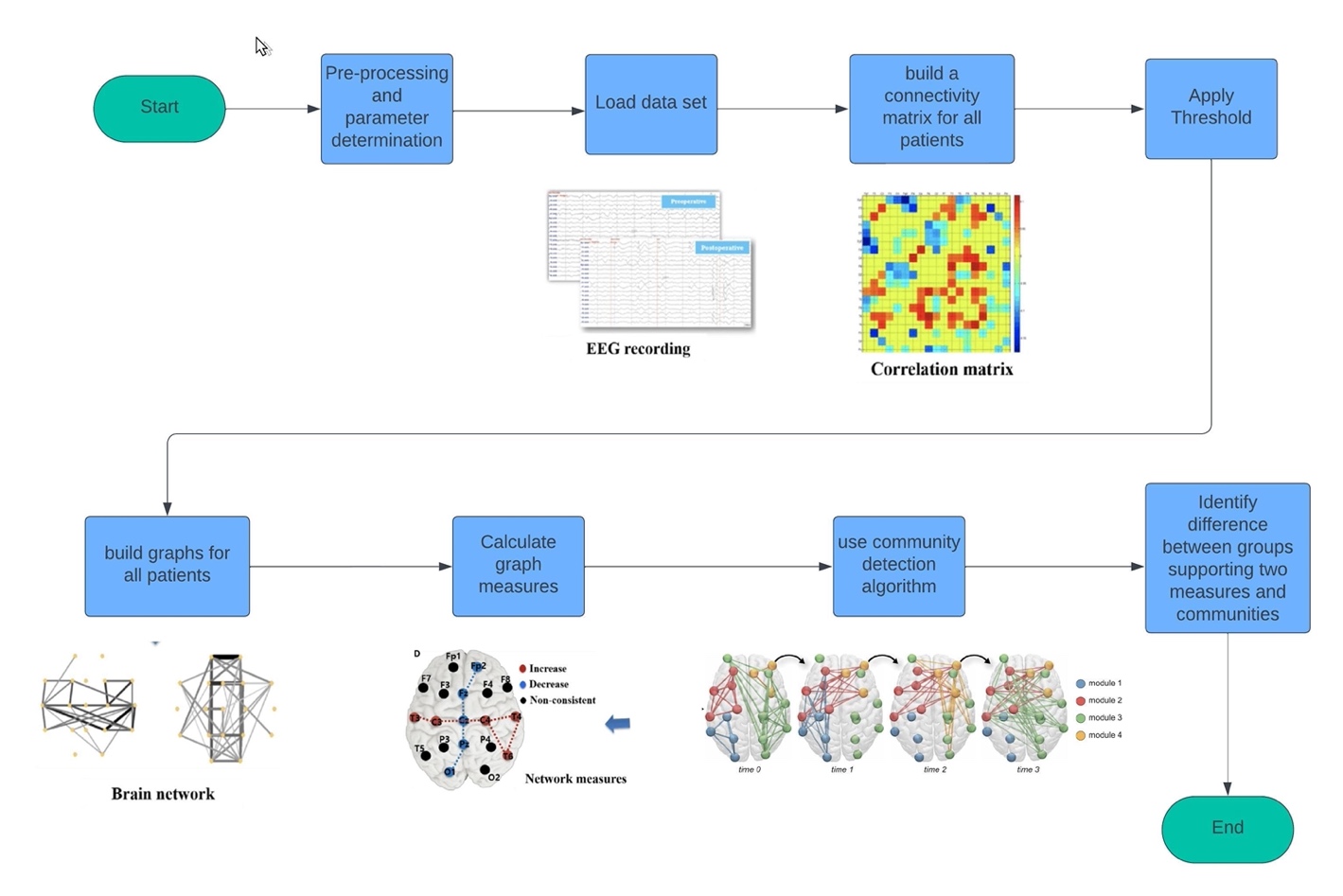
Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that is difficult to diagnose and affects a significant portion of the population. There is a growing need for quantitative methods to support its diagnosis. Our project aims to provide researchers with tools to explore and understand ADHD-related disorders within brain networks using methods like graph theory and community detection. We built our system using a dataset of EEG recordings from two groups of children: 61 with ADHD and 60 without (a control group). During the recordings, the children were asked to count cartoon characters shown in a series of images. This task was designed to stimulate specific brain activity patterns that can be analyzed for differences between the two groups.

By complex processing these EEG recordings, our system creates detailed graphs that simulate the brain networks of each patient. Researchers can use these graphs to compare the brain networks of children with ADHD to those without, applying various graph theory measures to identify differences in brain connectivity and function. Additionally, the system includes a feature to detect communities within each patient's brain network, allowing researchers to see how different regions of the brain interact and form functional clusters. This capability is crucial for understanding the unique brain network dynamics associated with ADHD.

To support further researches, the system also allows researchers to export the data and analysis results to CSV files. This feature makes it easier to conduct detailed studies and draw meaningful conclusions about ADHD and brain networks.

# Solution:

The flow chart below, created in Part A, illustrates our solution process. In this section, we will provide a detailed and precise explanation of the implementation steps, as completed in Part B.



***A***

***B***

***C***

***D***

***E***

***F***

***G***

***H***

***Figure 1: Solution process***

*Provides a detailed and precise explanation of the implementation steps. (A) pre-processing and parameter determination. (B) Loading data sets of EEG recordings. (C) Building a connectivity matrix for all patients. (D) Apply the threshold in the program. (E) Build a graph for all patients according to data sets. (F) Calculate graph measures. (G) Use community detection algorithm. (H) Identify the differences between groups supporting two measures and communities.*

**A** - **Pre-processing and parameter determination –**

First, the data set we received is MATLAB files of the patients' recordings. For easy and clear reading, we converted the recording files to Excel files. After converting the brain recording files to Excel files. The initial step in the process involved preprocessing the EEG data to clean the noise. This step was crucial to ensure that the data used in further analysis was as accurate and reliable as possible. Noise in EEG data can come from various sources, including muscle activity, eye movements, and external electrical interference. By applying noise reduction techniques, we were able to isolate the brain activity signals more effectively. We applied several techniques to clean the EEG data and reduce noise. First, we used a Butterworth bandpass filter to retain frequencies between 0.5 Hz and 50 Hz, which are typically relevant for EEG analysis, while filtering out unwanted frequencies. Next, we applied a notch filter at 50 Hz to remove powerline noise, a common source of interference in EEG recordings. We then used Independent Component Analysis (ICA) to identify and remove artifacts caused by eye movements, muscle activity, and other non-brain sources. These steps ensured that the EEG data was clean and ready for further analysis.

After cleaning the data, we divided each patient’s recording into segments of 8 seconds for each segment. This segmentation allowed for a more detailed analysis of the EEG data, enabling the examination of brain activity patterns within smaller time frames. This step is particularly important for tasks like the one our participants completed, where cognitive load and brain activity can vary significantly over short periods.

The recording range of 8 seconds was chosen according to an article that uses the same data set.[1]

For each patient, after dividing his brain recording into 8-second segments, we built a 19x19 connectivity matrix calculated by the Pearson correlation coefficient.

We chose to utilize the Pearson’s correlation coefficient, which is the most widely used technique to determine the correlation of signals between a pair of nodes. This method is widely used for several reasons, due to its simplicity, robustness, and interpretability. It offers a straightforward measure of the strength and direction of linear relationships between two variables, ranging from -1 to 1, making it easily understandable and communicable. Its resistance to outliers ensures its applicability across diverse datasets, while its standardized nature allows for comparisons between variables with different units. Moreover, its extensive use in various fields, along with its availability in statistical software, enhances its acceptance and accessibility. Despite its assumption of linearity, which may not always hold, the Pearson correlation coefficient remains a widely used and reliable tool for assessing relationships between variables in research and analysis. **שגיאה! מקור ההפניה לא נמצא.**]

The Pearson correlation coefficient (*r*) quantifies the strength and direction of the linear relationship between two variables. It always falls within the range of -1 to +1. A value closer to 1 or -1 indicates a stronger linear correlation, while a value closer to 0 suggests a weaker correlation. Variance measures the extent to which individual values in a dataset deviate from the mean value. A positive r value signifies a positive correlation, meaning that as one variable increases, the other variable tends to increase as well. Conversely, a negative r value indicates a negative correlation, where an increase in one variable corresponds to a decrease in the other. The coefficient of determination (r^2) represents the proportion of variance in one variable that can be explained by variance in the other variable. However, correlation does not imply causation, meaning that even if two variables are strongly correlated, it does not necessarily mean that changes in one variable cause changes in the other. **שגיאה! מקור ההפניה לא נמצא.**]

***Formula 1. Pearson’s formula*** *שגיאה! מקור ההפניה לא נמצא.****]***

*The Pearson correlation coefficient (r) quantifies the strength and direction of the linear relationship between two variables. It always falls within the range of -1 to +1. A value closer to 1 or -1 indicates a stronger linear correlation, while a value closer to 0 suggests a weaker correlation. Pearson correlation coefficient remains a widely used and reliable tool for assessing relationships between variables in research and analysis.*

Threshold determination:

After an in-depth investigation of various possible thresholds and consultation with Dr. Anat Dahan, our supervisor, we concluded that the appropriate threshold for our data analysis is an absolute threshold ranging between 0.4 and 0.7. Thresholding simplifies the network by focusing on the most significant connections, and this range was chosen based on related work using similar EEG data. This approach helps to highlight the most critical aspects of brain connectivity that may be relevant to differentiating between ADHD and Non-ADHD groups. Additionally, it allows us to understand brain activity according to each patient.

**B - loading data set –**

A study that utilized the investigated dataset categorized the time series as follows: Each subject's time series were partitioned into segments lasting 8 seconds each, resulting in a total of 1024 samples. The quantity of segments varied among subjects due to variations in the timing task, ranging from 50 seconds as the minimum for a subject in the control group to 285 seconds as the maximum for a subject with ADHD. On average, the control group had 13.18 segments (with a standard deviation of 3.15), while the ADHD group had 16.14 segments (with a standard deviation of 6.42). [**שגיאה! מקור ההפניה לא נמצא.**]

Following the research that was used in the same data set, we decided to implement the same segmentation time frame in our code. We divided each patient’s recordings into 8-second segments, and for each segment, we constructed a graph, performed community detection, and calculated the graph measures.

**C - build connectivity matrix for all patients –**

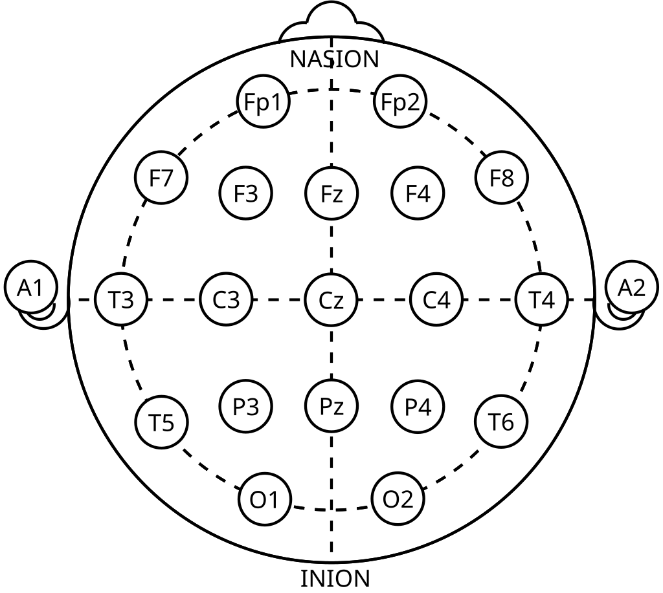
The connectivity matrix (CM) serves as a fundamental tool in understanding the intricate networks of the brain, particularly in relation to structural, functional, and effective connectivity. It encompasses a full square matrix with N^2 elements, where N represents the number of neurons, brain regions, or assemblies under consideration. Using the Pearson’s coefficient, we can calculate the connectivity matrix for each patient. As mentioned in Section A, we constructed the 19x19 correlation matrix using Pearson's correlation coefficient.

**D - Apply Threshold**

As mentioned in Section A, the chosen threshold is an absolute threshold that we applied to the connectivity matrix to build graphs with the significant connections based on the threshold.

**E - Build graphs for all patients**

After selecting and applying the threshold, we constructed a graph for all the segments of each patient. In this graph, the nodes represent the locations of the electrodes, and the edges indicate the functional connectivity between these locations. We then presented the graph based on the 10-20 standard using 19 channels. The 10-20 system, or International 10-20 system, is a globally accepted method for describing and positioning scalp electrodes during an EEG examination.



***Figure 2: 10-20 system. [9]***

*In this graph, the nodes correspond to the electrode locations, while the edges depict the functional connectivity between them. We displayed the graph according to the 10-20 system, using 19 channels. The International 10-20 system is a widely recognized method for placing and describing scalp electrodes during EEG tests. The numbers “10” and “20” refer to the distances between adjacent electrodes, which are either 10% or 20% of the total front-back or right-left distance of the skull.*

**F- Calculate graph measures-**

The calculated measures that we used:

Clustering Coefficient (Global): The average of the local clustering coefficients (the local clustering coefficient of a node measures how close its neighbors are to being a complete graph.) of all nodes, indicating how clustered the network. **שגיאה! מקור ההפניה לא נמצא.]**

Average Degree: This is the average number of connections each node has in the network. **שגיאה! מקור ההפניה לא נמצא.]**

Global efficiency - This measure how efficiently information can be transmitted across the entire network. It's the average of the shortest path lengths between all pairs of nodes. **שגיאה! מקור ההפניה לא נמצא.]**

Modularity: Measures how well the network can be divided into distinct communities or modules. It's determined based on a pre-defined community structure and reflects the quality of such division. **שגיאה! מקור ההפניה לא נמצא.]**

Average shortest path length – This measure is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network. [8]

**G – use community detection algorithm**

The fast greedy algorithm presents an efficient method for identifying communities within networks by maximizing modularity. Initially, it constructs a subnetwork consisting solely of connections among densely linked nodes. Subsequently, the algorithm iteratively selects random links that enhance the subnetwork's modularity and incorporates them. This iterative procedure continues as long as modularity enhancements are observed. Ultimately, communities are delineated based on the connected components present within the subnetwork. **שגיאה! מקור ההפניה לא נמצא.**]



***Figure 3: Community detection algorithm. [7]***

*This algorithm creates a subnetwork of densely connected nodes, then iteratively adds links that increase modularity, continuing until no further improvement is possible. Communities are identified based on the subnetwork's connected components.*

*Between edges are the connections between the different communities. Within edges refer to the connections withing the community itself.*

**H – Identifying differences between groups supporting two measures and communities –**

In this stage, comparison between graphs of ADHD to control cases performed. Analysis of graph measures is done to compare differences between the two groups.

# Description of the Research/Development Process

The research process for our project began in Phase A, where we conducted extensive research on various topics related to graph theory, the use of EEG recordings for analysis, and different methods for community detection within networks. We explored numerous articles to understand how these concepts have been applied in previous studies, with a particular focus on the different measures used in graph theory. In Phase B, we transitioned from research to implementation. We took the measures identified in Phase A and integrated them into our code. This phase involved not only implementing these measures but also adjusting them to ensure they provided accurate analyses of the data. For instance, we modified the average shortest path length measure to achieve correct and meaningful results. Another significant task in Phase B was determining the appropriate threshold for our analysis. While we explored different thresholding methods in Phase A, it was only after testing these thresholds in our code that we were able to select the most suitable one. Additionally, we performed noise cleaning on the EEG recordings to remove artifacts and work with a more accurate dataset. During the implementation phase, we used various libraries to build the graphs, calculate the measures, and apply community detection algorithms within the graph. We also modified the default visualization of the graph to present it in the 10-20 system format based on the document attached to our dataset, which simulates the placement of electrodes on the brain during the EEG recordings. This approach allowed for a more intuitive understanding of the data. Finally, in Phase B, we decided on the best way to export the data in Excel files. This decision ensures that the data is stored in a format that allows for easy future analysis, making it accessible for further research and exploration.

# Tools Used and Interface with Client

In our project, we used several tools to develop and analyze our system effectively. First, we used Visual Studio Code as our main development environment. This platform provided us with the necessary tools for coding, debugging, and managing the various parts of our project. An important resource in our process was the dataset containing EEG recordings from two groups of children, those with ADHD and a control group without ADHD. This dataset provided the basis for our analysis and system development. To make the data easier to work with, we converted the dataset from MATLAB files to Excel files. Excel allowed us to organize, manage, and prepare the data for further analysis. We used Excel not only to handle the data but also to export it for future statistical analysis, ensuring the data was in a format that could be easily accessed and analyzed by other researchers. For analyzing and visualizing the graphs, we used Python libraries like NetworkX. NetworkX was particularly helpful in creating and analyzing the complex network graphs that represented the brain's connectivity patterns. It allowed us to apply various graph theory measures and visualize the brain networks in a clear and informative way. In our project, there was no direct interface with a client. Instead, we worked with the dataset mentioned earlier and used tools from existing research. Our interactions were mainly with academic sources, and we focused on exploring these resources to improve our understanding of brain connectivity in children with ADHD. This allowed us to focus on the analytical and research aspects of the project.

# Challenges and Solutions

In our project, we faced several challenges. The first challenge was with the dataset itself. After we began working with it, we wrote a Python script to visualize the signals from the EEG recordings and noticed that they contained significant noise. This noise likely resulted from factors like eye blinks, muscle movements, and other artifacts. Initially, we weren't sure how to address this issue because we assumed the dataset was clean enough. After further investigation, we realized that noise cleaning was necessary. This process was challenging as we had to learn new methods.

The second major challenge was selecting an appropriate threshold for our analysis. Threshold selection is crucial because it determines which connections in the brain networks are considered significant. After spending many days evaluating different thresholding methods and consulting with our supervisor, Dr. Anat Dahan, we decided to use an absolute threshold between 0.4 and 0.7. This range was chosen to filter out insignificant connections while preserving important data. The challenge here was that there are many different thresholds to consider, and each one required testing to see if it provided the most meaningful analysis. Another challenge arose during the construction of measures and comparisons between the different groups. Specifically, when calculating the average shortest path length, We dealt with a problem. The built-in Python function for this measure works by calculating the average number of steps along the shortest paths for all possible pairs of nodes in the network. However, when not all nodes in a graph are connected, the function should return infinity, as there is no path between some pairs of nodes. Since Python cannot return infinity in this context, the function instead returns a "graph not connected". This caused us to get incorrect data because we were averaging the results for each patient and for the groups, ignoring the disconnected graphs, which altered the measure results. After discussing this issue with Dr. Anat Dahan, we decided to implement the function ourselves. Instead of ignoring the disconnected graphs, we assigned a value of 19 (the number of nodes in the graph) to represent the maximum possible path length, thus addressing the issue of disconnected nodes and ensuring a more accurate calculation of the average shortest path length.

# Results and Conclusions

We received EEG brain recordings from two groups one group of children with ADHD and another group without ADHD. We divided each patient's recording into 8 second segments to analyze the data. For each segment, we created a graph based on the 10-20 system, which was mapped using a document provided with the dataset. In these graphs, the nodes represent the locations of the electrodes in the brain, and the edges represent the connections between them.

The graphs were built after the researcher selected a threshold, which helps to focus on the strongest connections by keeping the values above the threshold value.

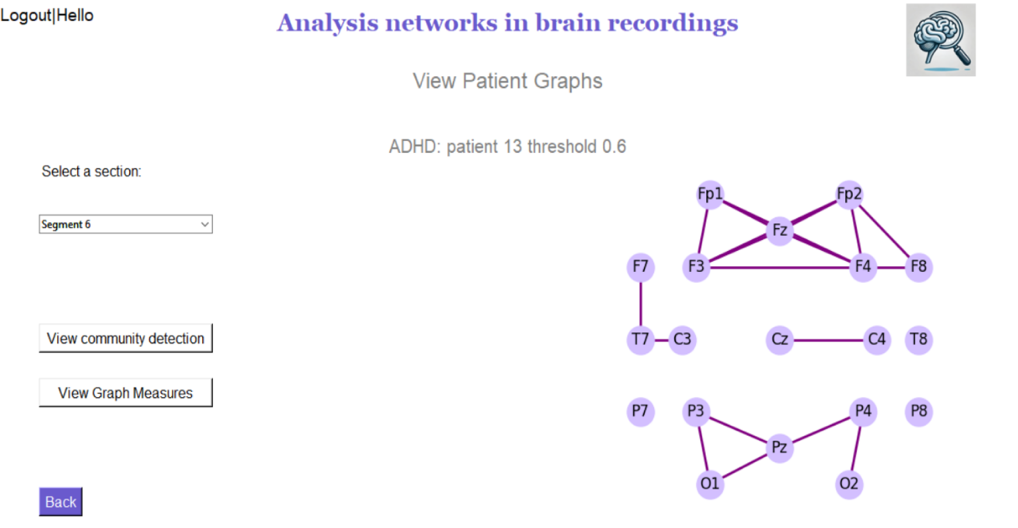
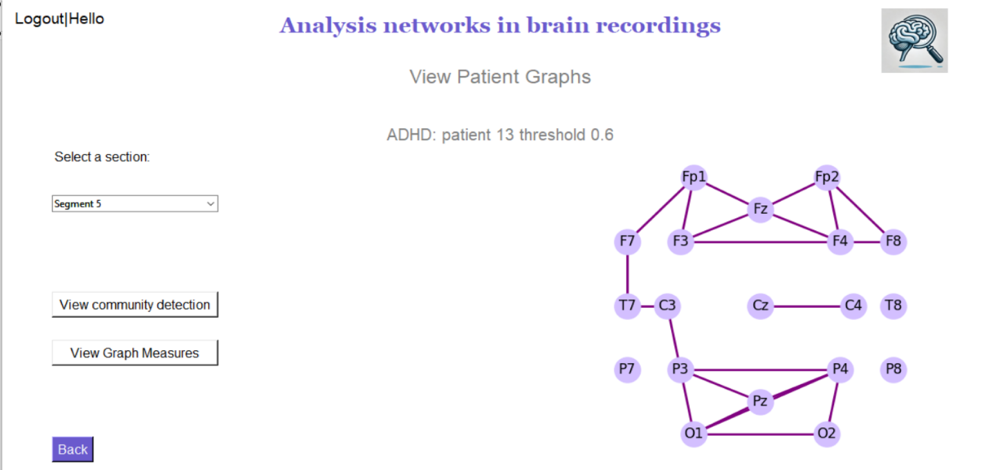
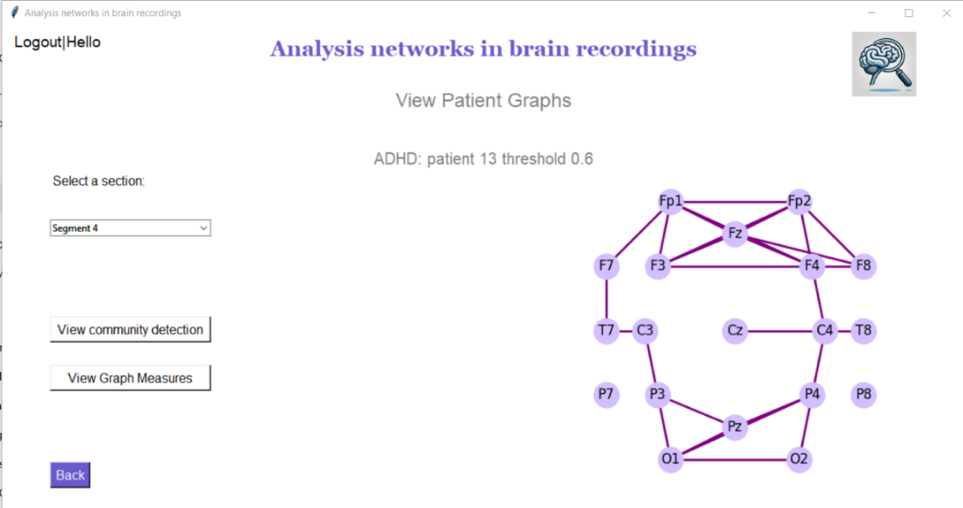
Now, we will show an example using one patient with ADHD and one without ADHD. We will examine different segments during the task and explain what the graphs reveal.

**ADHD patient:**

***1***

***2***

***3***

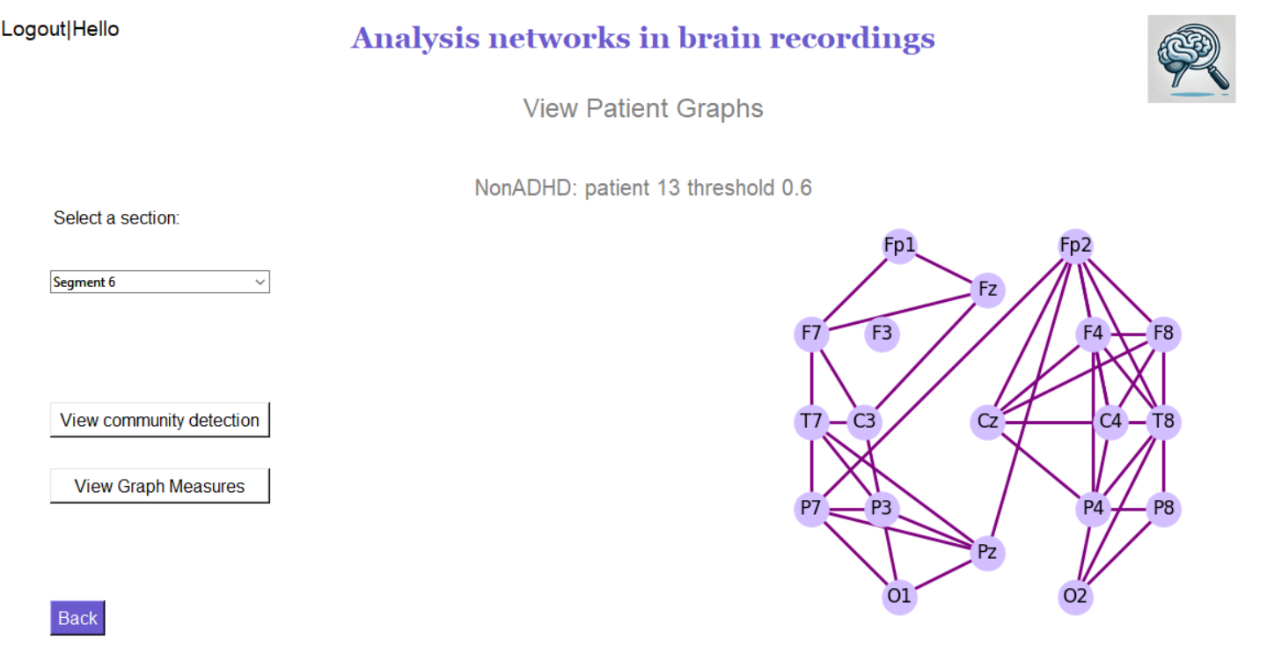
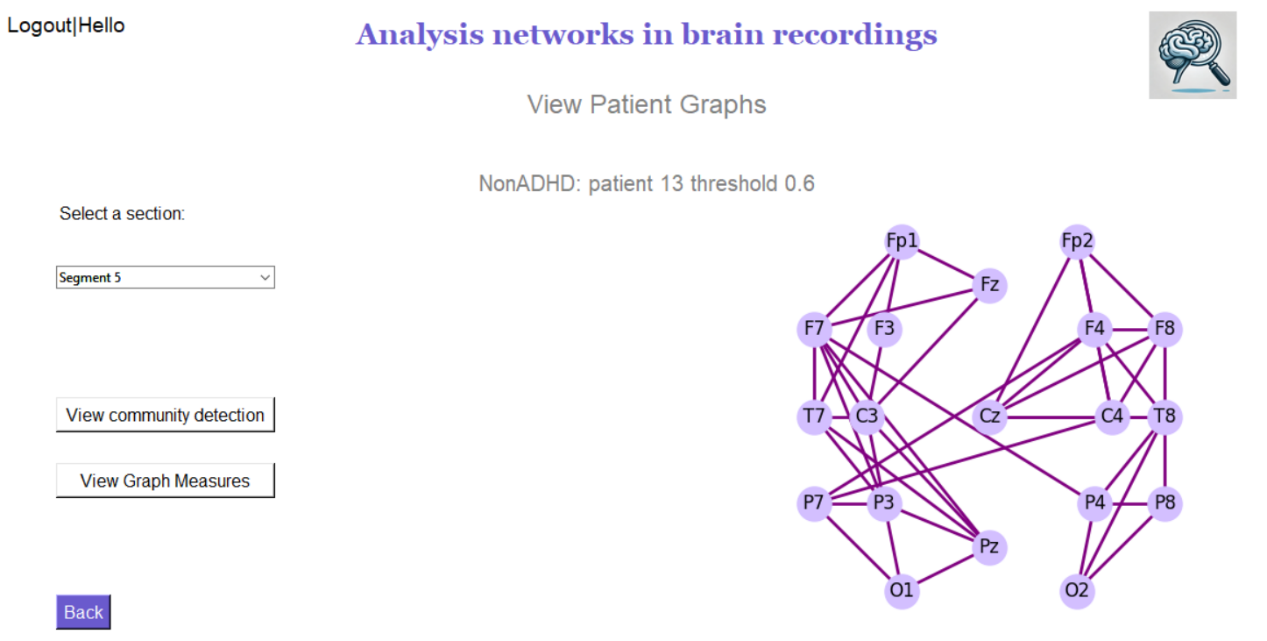
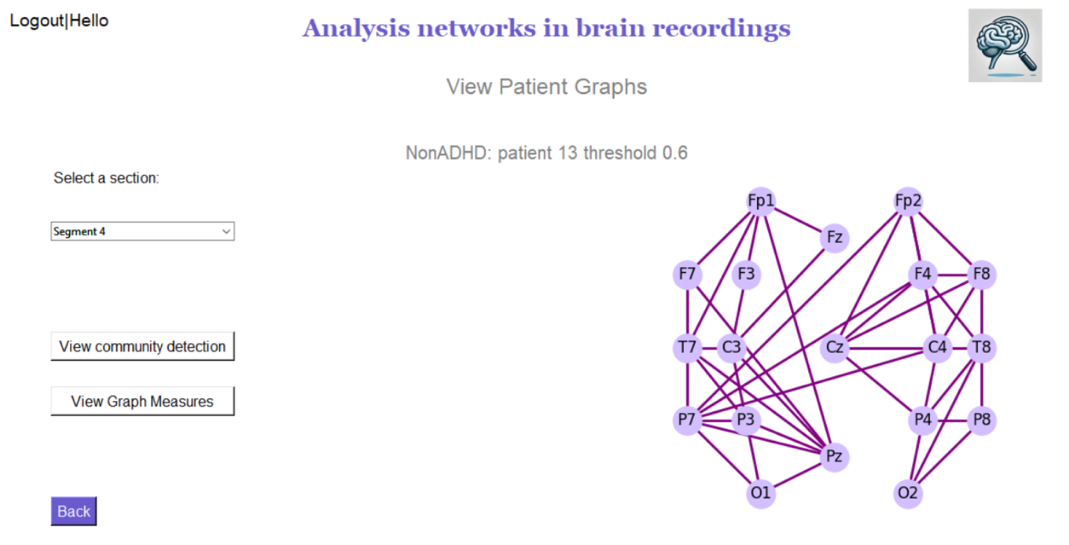


***Figure 4: Analysis networks in brain recordings in patients with ADHD***

*The graphs display EEG segments from a patient with ADHD, with a threshold set at 0.6 for functional connectivity. Graph (1) shows widespread connectivity across brain regions. Graph (2) reveals fewer connections, suggesting reduced synchronization. Graph (3) shows that connectivity is further diminished, indicating a more fragmented brain network during this segment.*

The graphs in *(figure 4)* represent the different segments of an EEG recording from a patient with ADHD. Each graph is associated with a specific segment of time during the EEG recording, with the threshold set at 0.6. The nodes in the graph represent the electrode positions on the scalp, following the 10-20 system, and the edges between the nodes indicate the connections or correlations between these electrode sites, which reflect the functional connectivity in the brain. The first graph shows an average number of connections across different regions of the brain. The connectivity is relatively spread out, suggesting some level of communication between various brain regions. In the second graph, there are fewer connections than in the first, indicating that the brain's activity might be less synchronized or that certain regions are not communicating as effectively with others during this segment. This could be due to a change in the patient's mental state or cognitive activity during this time. The third graph shows an even more reduced number of connections. The brain's network here appears to be more fragmented, with fewer interactions between different regions. This suggests a lower level of overall brain connectivity during this segment, which might indicate difficulties in integrating information across different areas of the brain. The reduction in the number of connections across these segments might reflect challenges in maintaining consistent brain network communication, which is often associated with ADHD. The differences between the graphs illustrate how brain connectivity can vary significantly over short periods, highlighting the dynamic nature of brain function in ADHD patients.

**Non-ADHD patient:**



***1***

***2***

***3***

***Figure 5:*** ***Analysis networks in brain recordings in patients without ADHD***

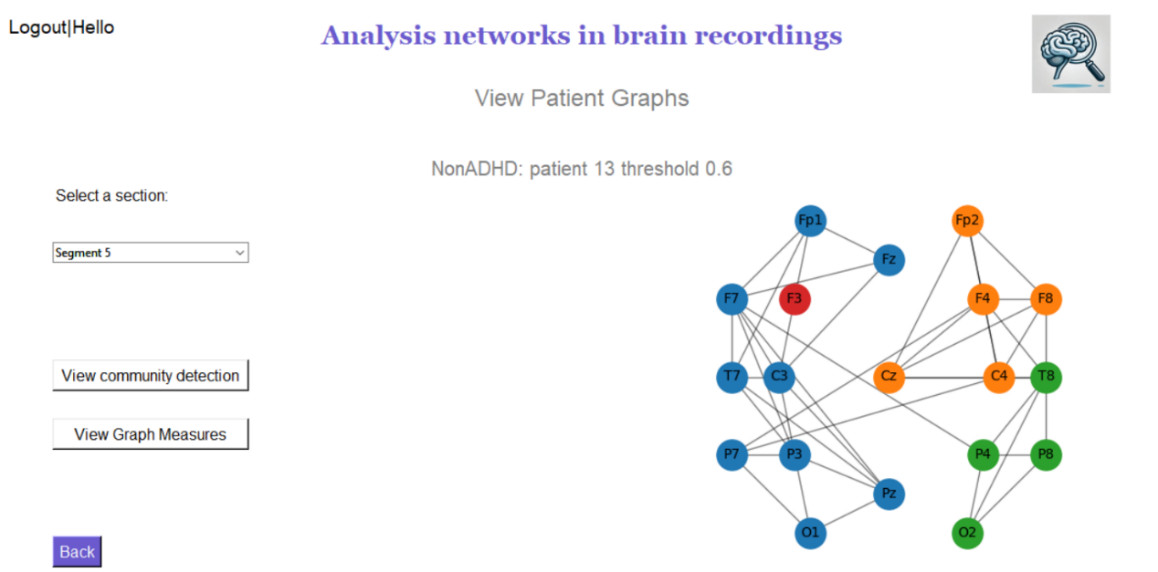
*EEG segments from a patient without ADHD are displayed, with a threshold set at 0.6 for functional connectivity. In graph (1), the non-ADHD patient shows a high number of well-distributed connections across the brain, particularly between frontal and parietal regions, indicating strong functional connectivity. Graph (2) reveals minor changes in the connectivity pattern, but the brain remains well-connected, suggesting stable communication between regions. Graph (3) shows a slight decrease in connections, particularly in frontal and parietal areas, but the brain network remains integrated and functional.*

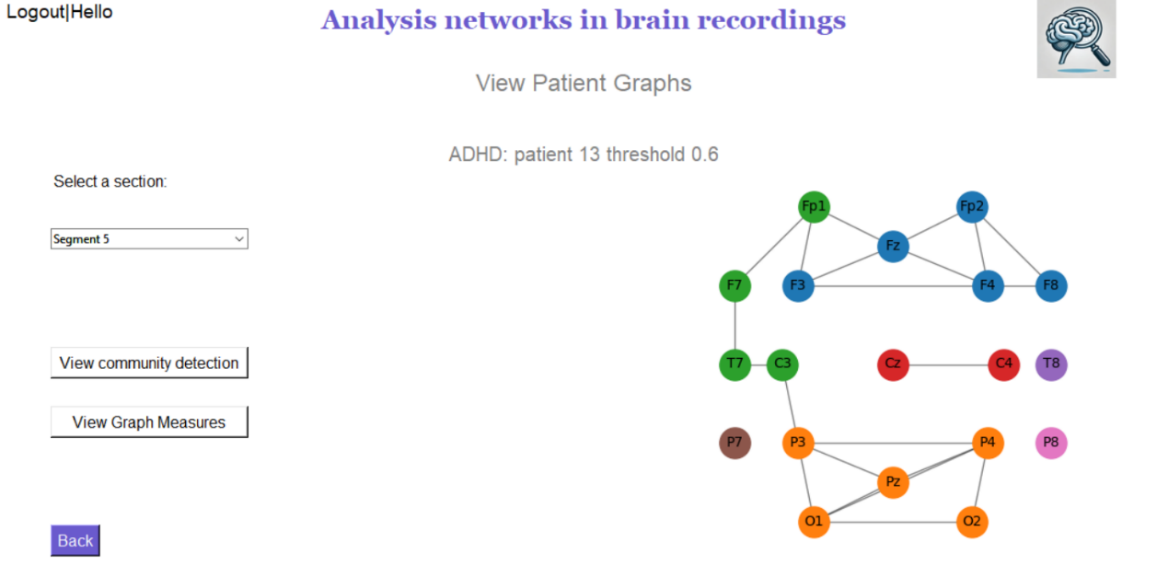
The graphs in *(figure 5)* above, represent the different segments of EEG recordings from a non-ADHD patient, with the threshold set at 0.6. As mentioned previously, the nodes in each graph indicate the position of electrodes on the scalp, and the edges between the nodes represent the functional connections or correlations between these brain regions. In the first graph, there are a significant number of connections between different brain regions, indicating a high level of functional connectivity. The connections are well-distributed, covering both frontal and parietal areas, signifying that the brain is well-integrated and different regions are effectively communicating with each other. In the second graph, the connectivity remains strong, however, there are some changes in the pattern of connections compared with the first graph. This suggests a slight change in brain activity during this segment, though the overall connectivity remains strong. In the third graph, a minor change can also be seen in the number of connections, particularly in the frontal and parietal regions. However, the brain networks continue to show a good level of integration, indicating the brain regions continue to work together effectively. When comparing the ADHD and Non-ADHD segment graphs, we notice distinct differences. The Non-ADHD brains typically display a higher number of connections between brain regions, indicating more synchronized and integrated activity, characterizing a well-functioning brain network. Additionally, the connectivity patterns are more consistent across different segments, suggesting that the Non-ADHD brain maintains stable functional connections over time. In contrast, ADHD brains show more variability in the number of connections, particularly in later segments, reveal a more fragmented network with fewer connections. This fragmentation suggests that the ADHD brain may struggle to maintain consistent communication between regions, which could contribute to the cognitive challenges often associated with ADHD. Overall, this comparison indicates that brain connectivity in ADHD patients tends to be less stable and more fragmented than in Non-ADHD patients. Stronger more consistent connections in the Non-ADHD brain suggest better integration of brain regions during cognitive tasks. These observations support the idea that ADHD is linked to disruptions in brain network connectivity, which can result in difficulties with attention, executive function, and other cognitive processes.

Community Detection:

**1. Non-ADHD patient**

**2. ADHD patient**





***Figure 6: Community detection in brain networks of non-ADHD and ADHD patients.***

*Graph (1), the Non-ADHD brain is organized into several well-connected communities, indicating efficient coordination and specialized processing across regions. Graph (2) shows there are fewer, less unified communities with more isolated nodes, suggesting weaker functional connectivity between brain regions.*

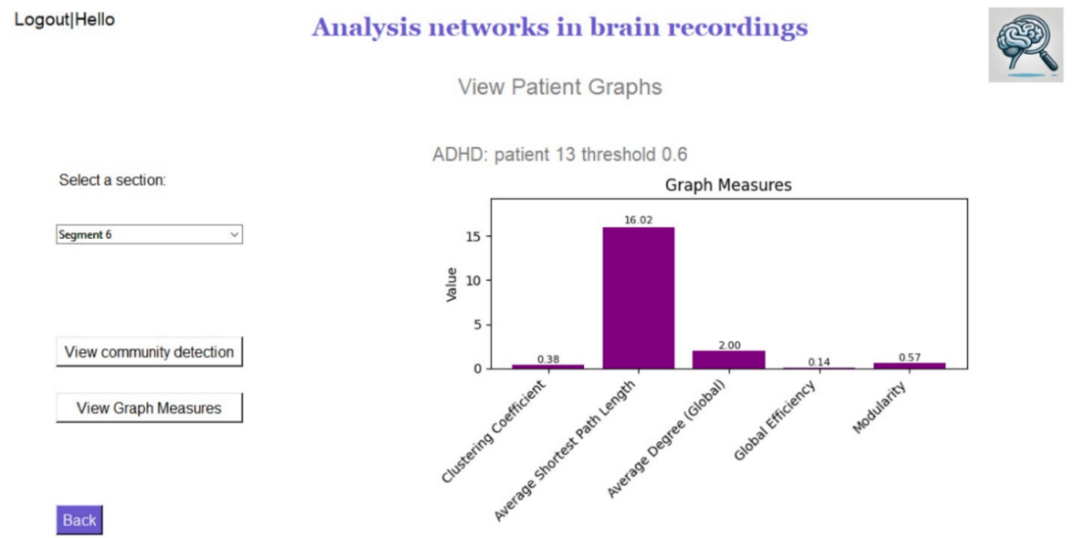
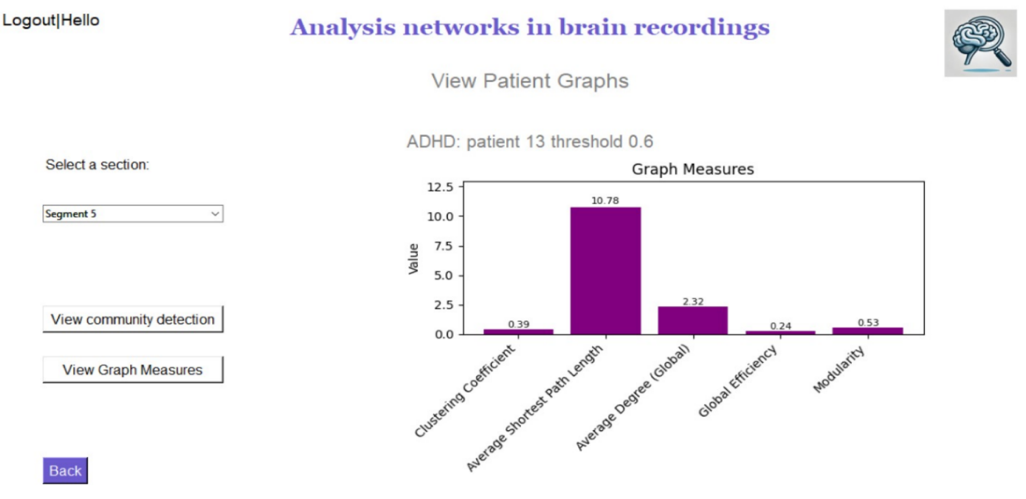
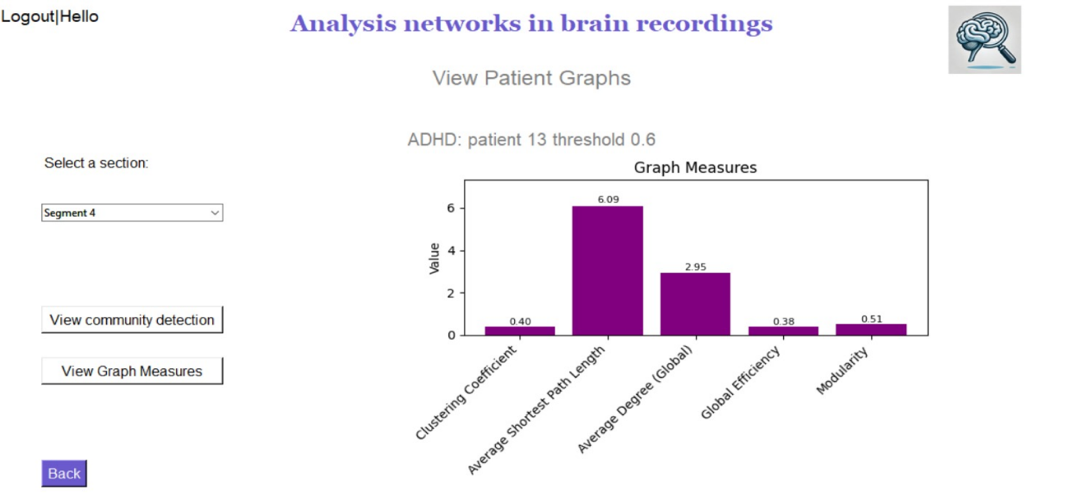
Figure 6 illustrates the results of community detection in the brain networks of a Non-ADHD and an ADHD patient respectively, using a fast greedy algorithm. This algorithm groups nodes (representing brain regions) into communities based on the strength of their connections, aiming to maximize modularity, which measures the concentration of links within communities compared to those between them. In the graph for the Non-ADHD patient, the nodes are organized into several well-connected communities, each represented by a different color. The presence of these distinct groups suggests that brain regions are functioning in a syncytium, emphasizing efficient and specialized processing in different parts of the brain. In contrast, the ADHD patient shows fewer communities that are also less unified and more dispersed. Some nodes appear more isolated or loosely connected, suggesting that the brain regions are not working together as effectively. This fragmented structure is said to contribute to difficulty coordinating brain functions, leading to challenges in attention, information processing, and task execution.

**1**

**2**

**3**

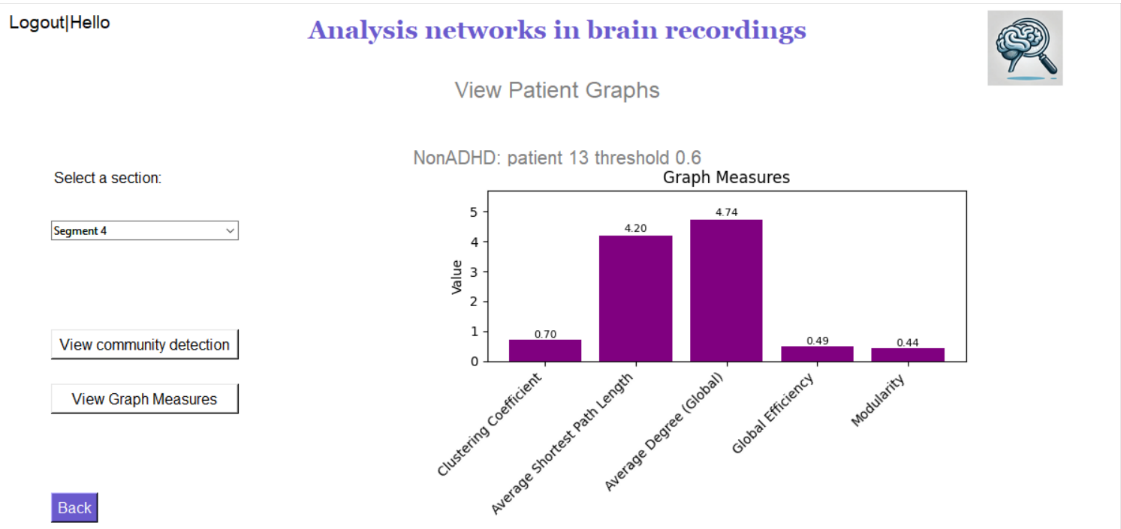
**ADHD patient:**



***Figure 7: Graph measures in 3 different segments of an ADHD patient - clustering coefficient, average shortest path length, average degree, global efficiency, and modularity.***

*(1) In Segment 4, the average shortest path length is moderate, indicating relatively efficient communication within the brain's network. (2) By Segment 5, this path length increases significantly, and in (3) Segment 6, it continues to rise, suggesting that information transfer becomes more challenging. The global efficiency and average degree both decrease across Segments 4, 5, and 6, with Segment 6 showing the lowest values, reflecting a decline in connectivity and overall network efficiency.*

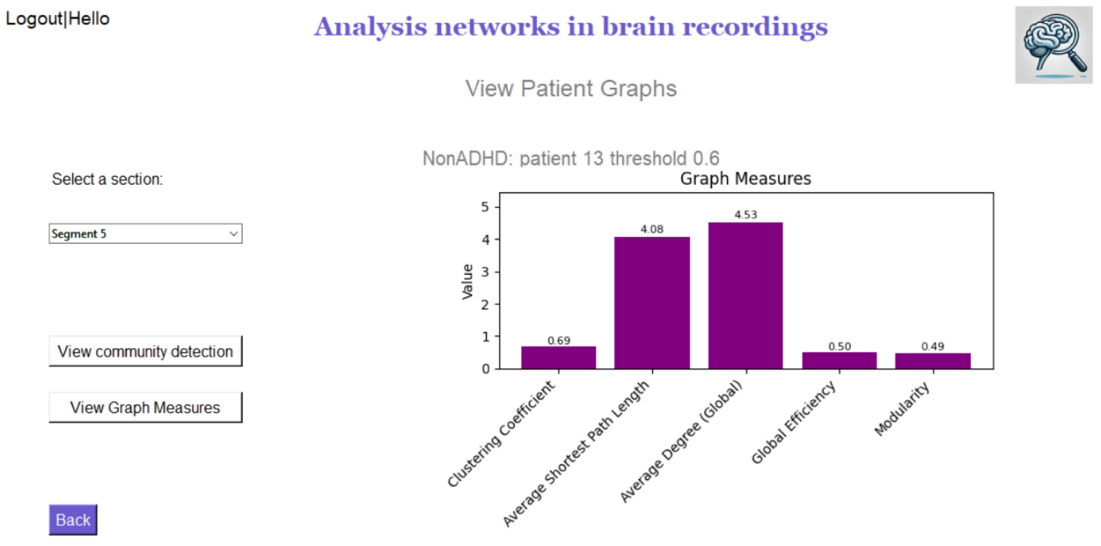
In the graphs of *(figure 7)* above, we observe graph measures for an ADHD patient across three different segments. The measures calculated include the clustering coefficient, average shortest path length, average degree, global efficiency, and modularity, offering insights into the brain's functional connectivity and organization during these time frames. Across these segments, we observe that the average shortest path length changes, with Segment 4 showing a moderate value, which then increases significantly in Segments 5 and 6. This increase suggests that the brain's network becomes less efficient over time, with information transfer across the network becoming more challenging. The average degree, which reflects the number of connections between different regions, decreases across the segments, indicating that fewer brain regions are connected, contributing to reduced communication efficiency. The global efficiency also decreases across the segments, reaching its lowest in Segment 6. This indicates challenges in maintaining effective communication across the brain’s network, suggesting that the brain regions are struggling to work together efficiently. The consistently low clustering coefficient throughout the segments indicates that the brain is having difficulty forming closely connected groups, which may reflect challenges in integrating information across different regions. These findings suggest that, in this ADHD patient, brain connectivity becomes increasingly fragmented and less efficient over time. The decline in global efficiency and clustering coefficient further supports the conclusion that the brain's network is not well-organized for effective communication. However, it is important to note that to reach accurate and reliable conclusions, statistical tests on the data are necessary. While the observations provide valuable insights, statistical analysis would allow for a deeper understanding of these patterns and their significance in the context of ADHD.

**Non-ADHD patient:**

**1**

**2**

**3**



A graph with purple squares

Description automatically generated with medium confidence

***Figure 8:*** ***Graph measures in 3 different segments of a Non-ADHD patient - clustering coefficient, average shortest path length, average degree, global efficiency, and modularity.***

*In Segment 4 (1), the clustering coefficient remains stable around 0.7, showing strong local connectivity, with brain regions efficiently working together. Across Segments 5 (2) and 6 (3), the average shortest path length stays between 4 and 4.5, indicating consistent information transfer between regions, while global efficiency remains around 0.5, maintaining effective network communication. Modularity slightly increases in Segments 5 and 6, suggesting improved separation of brain functions, enhancing task-related cognitive performance.*

In the graphs of *(figure 8)* above, graph measures of a Non-ADHD patient across various segments show consistent results with small variations. The clustering coefficient remains relatively stable around 0.7, suggesting that the brain's network is maintaining a good level of local connectivity, where brain regions are effectively working together in closely connected groups. The average shortest path length shows minor changes across segments, generally staying around 4 to 4.5, indicating stable efficiency in how information is transferred across different brain regions. The average degree (Global) remains consistent as well, suggesting that the overall level of connectivity between different brain regions is maintained, allowing for strong communication throughout the brain. Global efficiency values hover around 0.5, showing that the brain network is well-organized for effective information exchange. Modularity, which measures how well the brain is divided into distinct groups or communities, shows a small increase across segments, suggesting that the brain might be improving its ability to separate different functions during the task. Overall, these stable and consistent measures across segments in a Non-ADHD patient indicate a well-organized and efficiently functioning brain network, likely contributing to effective cognitive performance during the task. The small variations suggest adaptive changes in the brain's connectivity, reflecting its ability to adjust and maintain optimal function throughout different phases of the task. However, to reach in-depth insights, it is necessary to perform statistical tests. These tests would allow for a more careful analysis of the data, helping to confirm the observed patterns and expose more detailed relationships within the brain network.



***Figure 9: Graph measures - comparison between ADHD and Non-ADHD groups in 0.6 threshold.***

The comparison between the ADHD and Non-ADHD groups, as shown in the graph, reveals differences in several graph measures. First, the average degree is higher in the ADHD group, suggesting that their brain networks may be more locally interconnected. However, this does not necessarily imply more efficient global communication. A higher average degree in ADHD may indicate dense, local connectivity, but it does not clearly guarantee that information is efficiently transferred across distant regions. In fact, the global efficiency is slightly higher in the Non-ADHD group, which suggests that their brain networks may be more globally optimized for efficient communication across different regions. Additionally, the average shortest path length is lower in the ADHD group. This might seem counterintuitive at first, but it can be explained by the increased local connectivity, which can result in shorter paths between closely connected nodes. However, the Non-ADHD group’s higher shortest path length could reflect a more balanced, globally optimized network that facilitates better information transfer over long distances. The clustering coefficient is higher in the ADHD group, indicating that their brain regions tend to form more tightly-knit local clusters. While this may support local processing, it could come at the cost of reduced global communication efficiency. In contrast, the Non-ADHD group shows a slightly lower clustering coefficient, which could indicate a more distributed and globally efficient network. Modularity shows only a slight difference between the groups. Both groups exhibit similar community structures, meaning that the brain networks in both ADHD and Non-ADHD individuals are organized into distinct modules.

In order to reach more significant insights and conclusions, it is essential to perform statistical tests to determine whether the differences observed between the groups are significant. For example, while we see a higher average degree in ADHD, we cannot know if this difference is statistically significant without further analysis. Statistical tests will help to confirm whether the observed patterns reflect true differences in brain network structure between the ADHD and Non-ADHD groups.



***Figure 10: Graph measures - comparison between ADHD and Non-ADHD groups in 0.4 threshold.***

With a lower threshold of 0.4, more connections are included in the analysis, leading to an increase in the average degree for both ADHD and Non-ADHD groups. This means that more brain connections are preserved, including weaker connections that would have been excluded at a higher threshold. Despite this, the global efficiency remains low, indicating that the additional connections do not necessarily improve the brain's overall ability to transfer information efficiently across the network. The average shortest path length is reduced compared to what might be seen at a higher threshold, suggesting that information can travel more directly between nodes due to the increased number of connections. The clustering coefficient is slightly higher, reflecting more locally connected groups within the brain networks at this lower threshold. This analysis at a lower threshold of 0.4 highlights how the inclusion of weaker connections affects the overall network measures. It underscores the importance of selecting an appropriate threshold to capture the most meaningful connections. While we observe higher average degrees and other increased values at this threshold, we cannot determine if these differences are significant or merely the result of noise. Until statistical tests are performed, we cannot confirm the importance of these observed differences.

Testing Process:

|  |  |  |  |
| --- | --- | --- | --- |
| Case | Test Case | Expected Result |  |
| 1 | Insert same signal to connectivity matrix | Because we use Pearson correlation calculation the expected result of the same signal is 1. | Success |
| 2 | Insert different signals connectivity matrix | Because we use Pearson correlation calculation the expected result of different signals is between 0-1. | Success |
| 3 | Apply threshold | It is observed that the results of the chosen threshold will be the same for articles with the same data set and the same threshold. | Success |
| 4 | Build graph for all patients. | It was observed that the graphs of the patients would be similar to the graphs in the articles with the same data set. | Success |
| 5 | Calculate graph measures | It will be observed that if we take a simple graph and calculate all the measures on it or for a graph with results that we know will get a correct calculation of the measures. | Success |
| 6 | Using community detection algorithm | We will observe that if we take a graph whose community modules are known and put it into our community detection algorithm, we will expect to get the same modules in the graph. | Success |

# 

# Lessons Learned

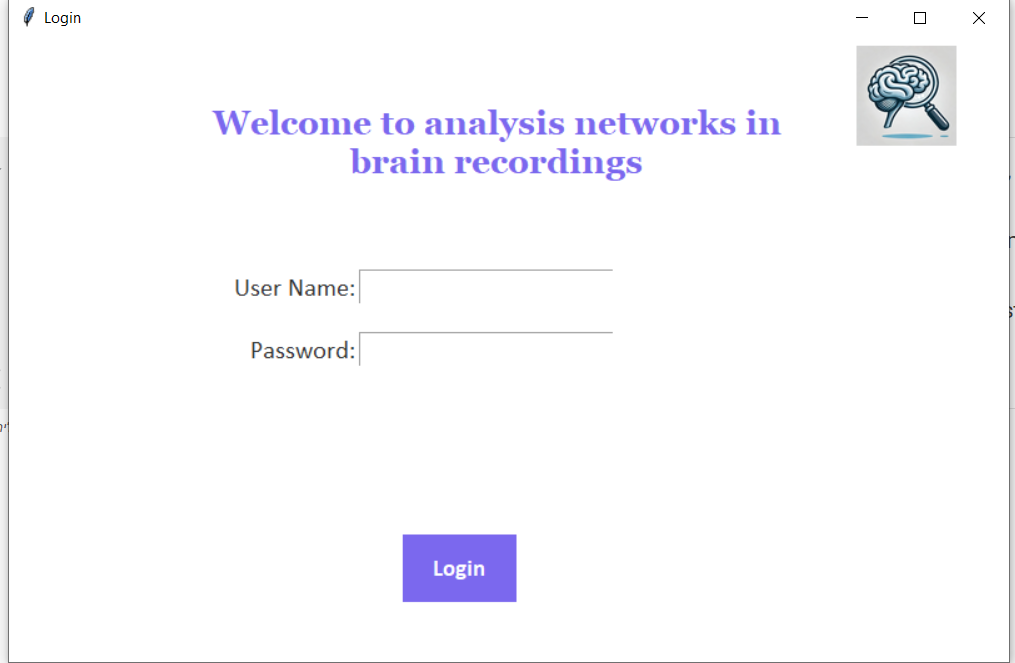
Throughout this project, each step required careful thought and analysis. From the very beginning, when we were cleaning the data set, we spent considerable time researching and determining the best methods to ensure that the data was accurate and reliable. Choosing the right threshold was also a significant task. We dedicated a lot of time to examining various threshold options and consulted with our guiding lecturer to make the best decision. Writing and adjusting the functions to calculate the graph measures were another challenging aspect. We needed to ensure that these functions were adjusted to our specific data, which involved a lot of trial and error and adjusting. While we intended to perform statistical tests to validate our findings, time limitations prevented us from completing this step. To ensure future analyses can be done efficiently, we exported the data into CSV files, allowing for easier and quicker statistical testing at a later stage. This project has been a significant learning experience for us. We learned how to transform a set of numbers from EEG recordings into graphs that illustrate brain connectivity. We also gained a deeper understanding of graph measures and how to identify communities within a graph. By examining graphs of children with ADHD and those without, we learned a lot about the importance of thresholds in highlighting the strongest connections in these brain networks.

In conclusion, we can say that we achieved the goals we set for this project. Our primary objective was to build a system that takes a data set of brain recordings and transforms the numerical data into a graph that simulates brain connectivity. Through various measures and community detection, we were able to highlight differences between the groups. However, it's important to note that without performing statistical tests, we cannot yet determine whether these observed differences are truly significant or just coincidental. While we met our technical objectives, further analysis is needed to reach definitive conclusions about the data. This project has built a strong foundation, and with additional statistical testing, it could lead to deeper insights into brain connectivity and the differences between the ADHD and Non-ADHD groups.

# User Guide

**General System Description**

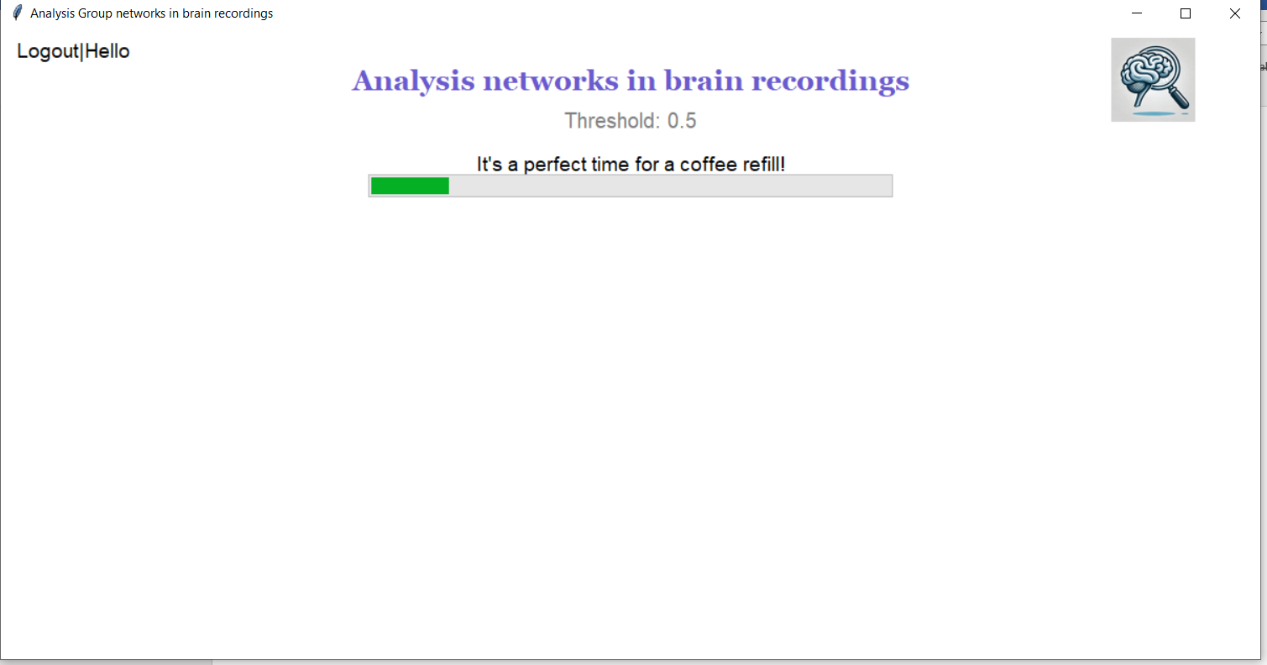
The system is designed to process EEG brain recordings by dividing each patient's data into 8 second segments. For each segment, the system constructs a detailed graph that visually represents brain connectivity. In these graphs, nodes represent the locations of the electrodes placed on the scalp, while the edges illustrate the connections between these nodes. The system also allows users to identify distinct communities within each graph using community detection algorithms. Additionally, users can view various graph measures that provide insights into the characteristics of each graph. Furthermore, the system enables comparisons between different groups, such as ADHD and control group, based on these graph measures. To assess the significance of these differences, statistical tests must be performed. To facilitate this process, the system includes data export functionality. Users can easily export detailed per patient data and aggregated group comparison data into Excel files. This feature ensures that researchers can perform statistical tests, share their findings, and continue exploring and validating their results.

**Operating Instructions**

1. On the login screen, the user should enter the username "koraltopaz" and the password "1234". Once these details are entered, the user will be able to log in to the system.



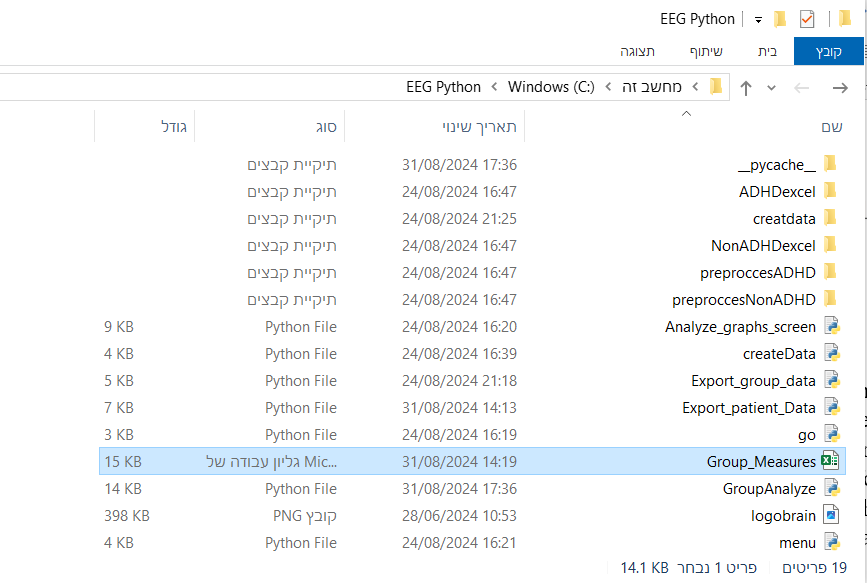
1. On this screen, there is an option to choose whether to select a single patient or a group, along with selecting the appropriate threshold, patient type, and patient number. In the screenshot shown here, we have selected a group, so data such as patient type and patient number are not relevant and are therefore hidden. After selecting a group and setting the threshold, when the user clicks the "Next" button, they will be taken to a screen that displays the comparison between the groups.



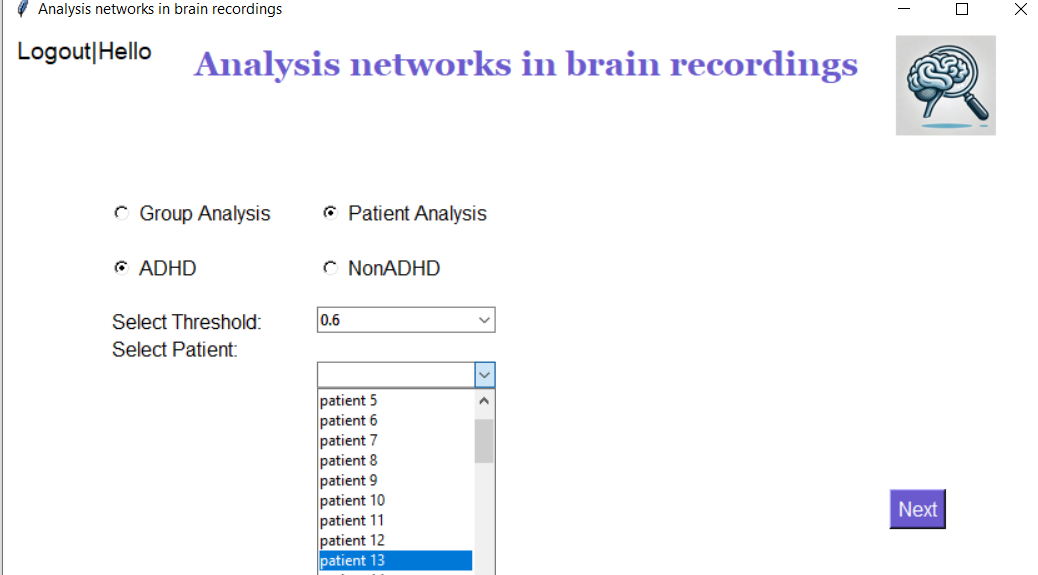
1. Here, you can see the progress as the system calculates comparisons and builds the data according to the selected threshold. Since the user can choose a different threshold each time, the system dynamically updates the data to reflect the new threshold.



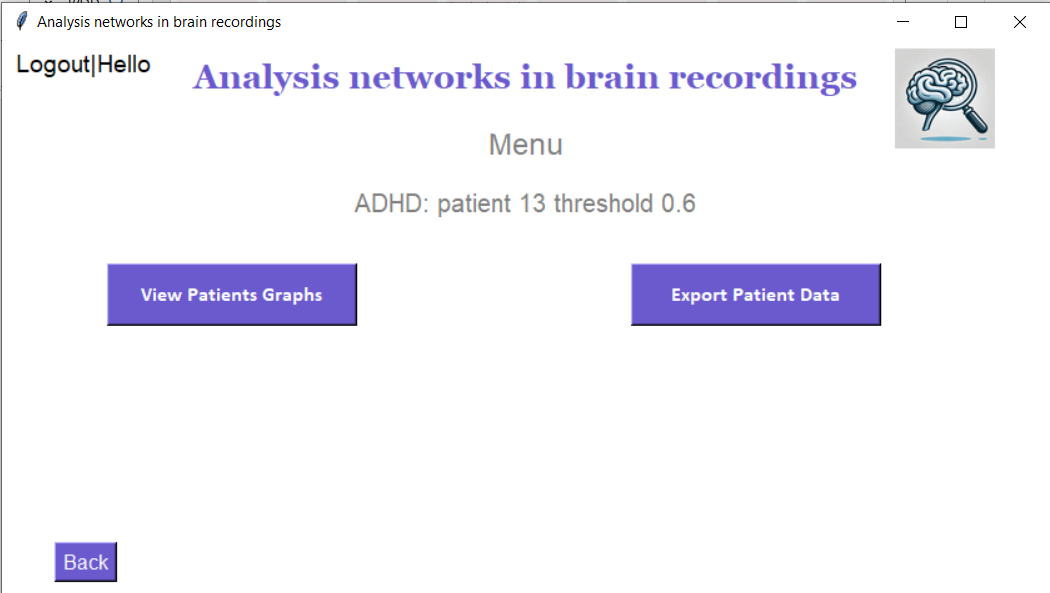
1. After the system finishes calculating the data, it displays a graph comparing the various measures. In the graph, the ADHD group is marked in blue, while the control group (those without ADHD) is marked in purple. Additionally, there is an "Export to Excel" button that allows you to export the data to an Excel file.



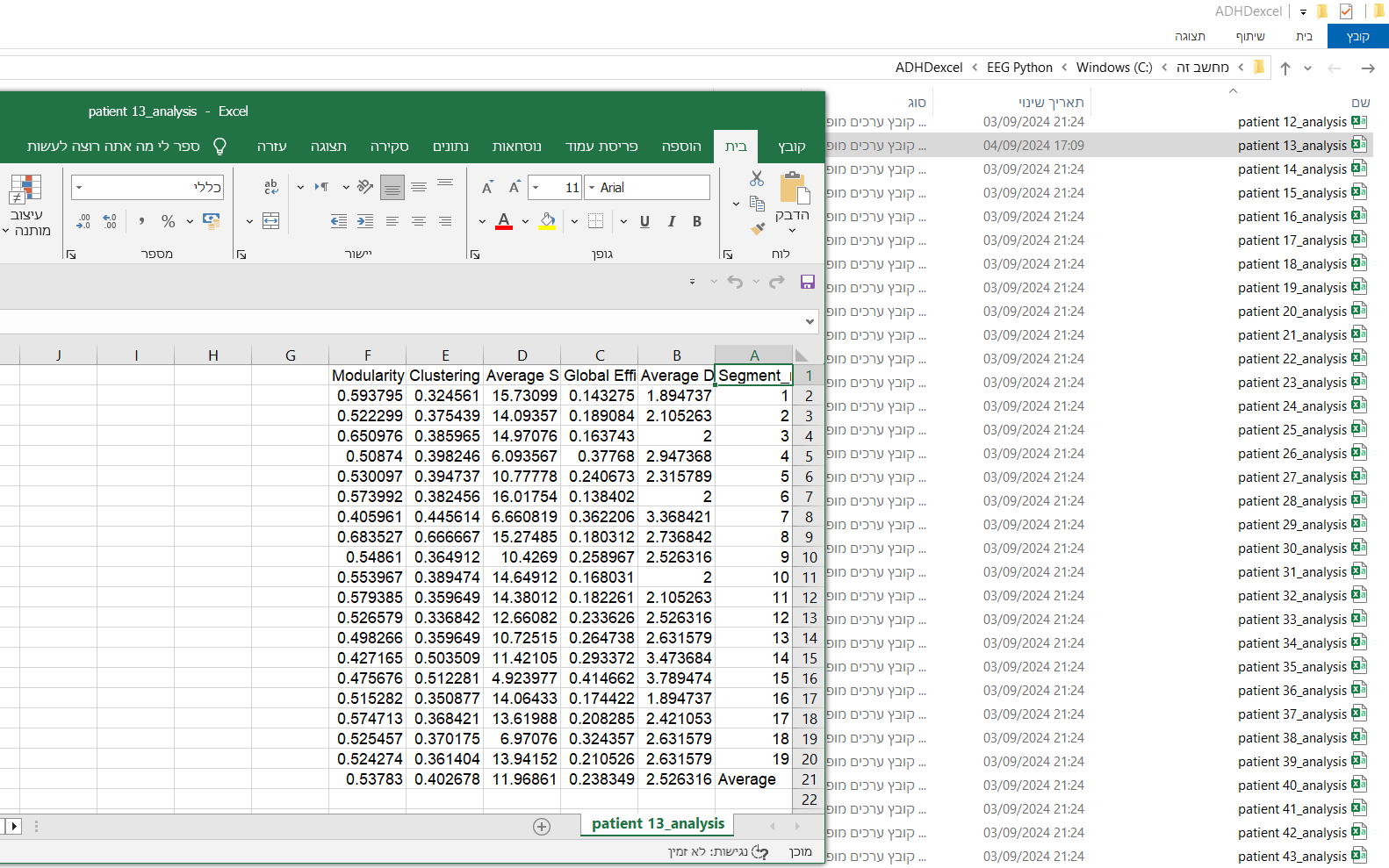
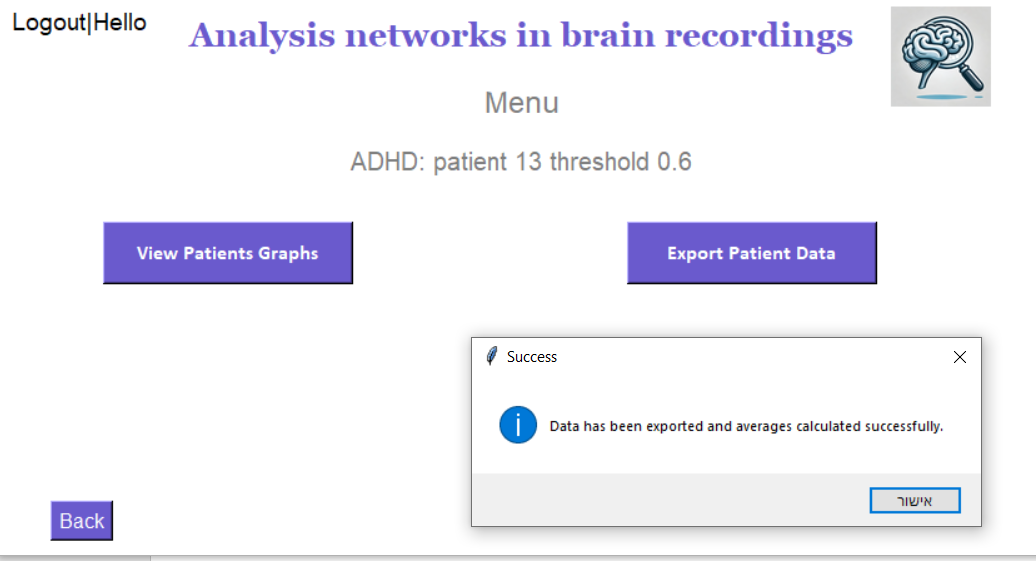
1. This file is saved in "EEG Python" folder and called "Group\_Measures," which contains data for all ADHD patients and the control group. For each patient, the average of each index is recorded. This average is calculated by summing the index results for each segment and then dividing by the total number of segments for that patient.



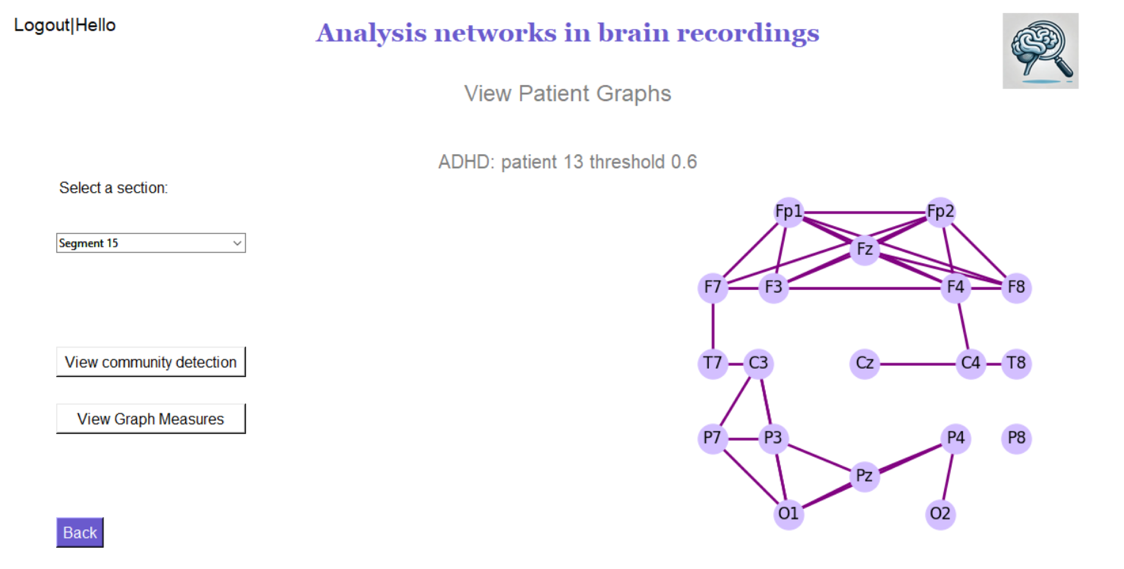
1. If we choose a patient, additional options become available, such as selecting the type of patient. You can choose whether to view data for a patient with ADHD or a patient without ADHD, and you can also select a specific patient, for example, patient 13. Additionally, even when selecting an individual patient, you still have the option to choose a threshold for the analysis.



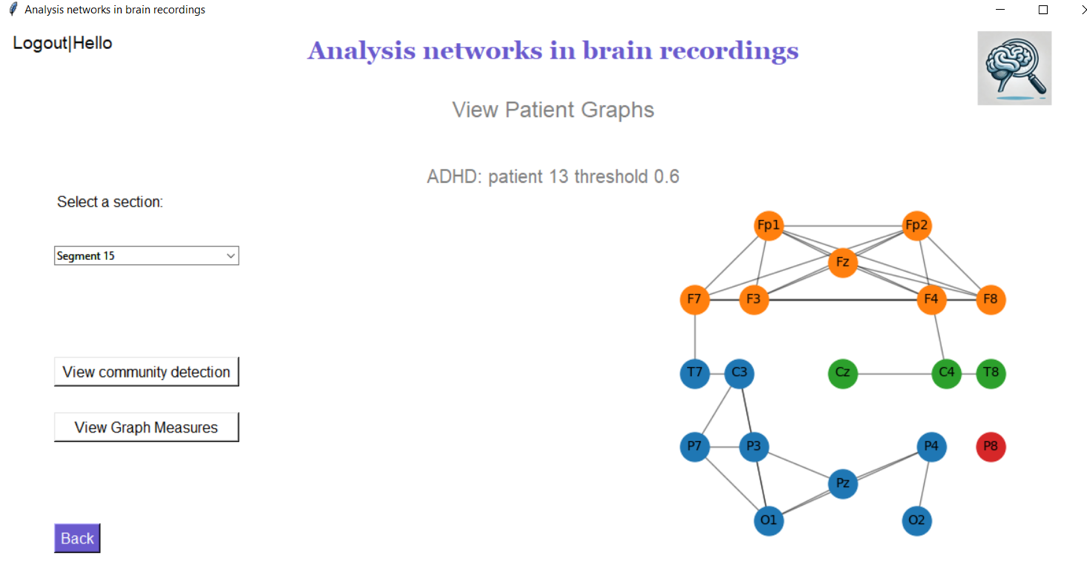
1. After selecting a patient, patient type, specific patient, and threshold, we are taken to the patient's menu screen. On the right side, there is an "Export Patient Data" button, which, when clicked, saves the patient's measures for each segment based on the selected threshold. The data is saved in a folder named "ADHDexecl." On the left side of the screen, there is a "View Patient Graph" button. Clicking this button leads to a screen displaying the patient's graphs.



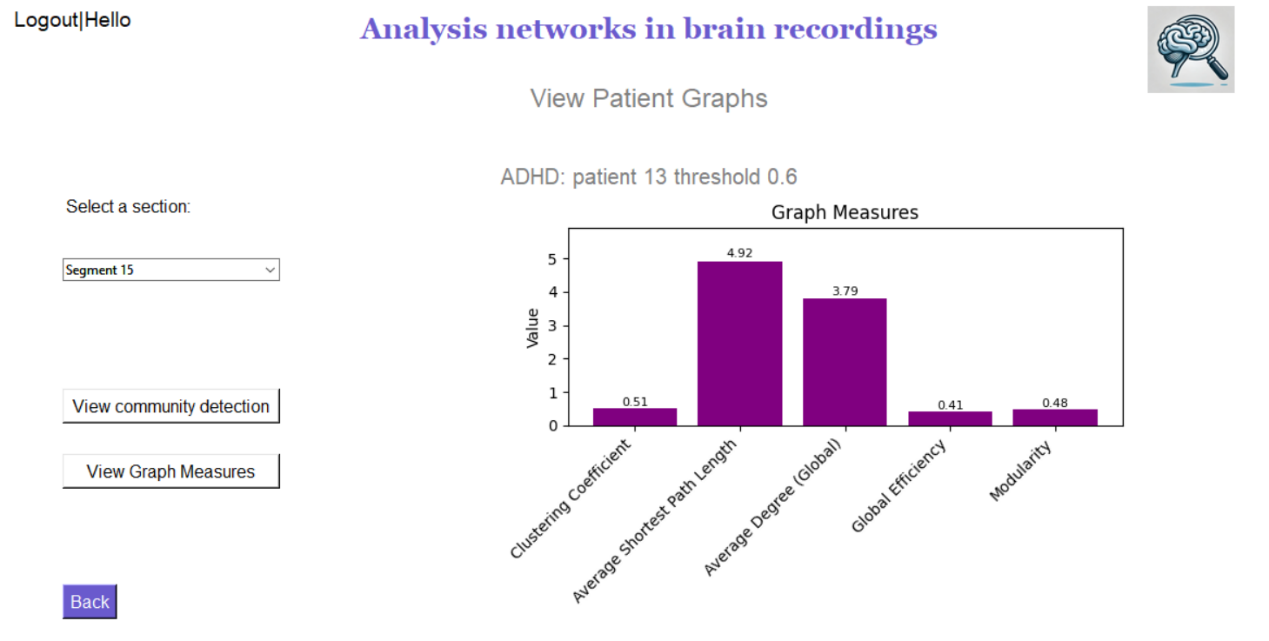
1. After choosing to export the data, a confirmation message pops up indicating that the export was successful, and the file is saved accordingly.



1. If you choose "View Patients Graphs" on the menu screen, you'll move to the "View Patient Graphs" screen, where you can select a segment and view its graph.



1. By clicking the "View Community Detection" button, you can see the different communities identified within the graph, highlighting the distinct clusters of connectivity.



1. By clicking on the "View Graph Measures" button, you can see the calculations of the various measures we have computed, providing valuable insights into the graph's structure and connectivity.
2. In the "create data" directory, patient-specific data is systematically organized. Each patient has a dedicated folder containing all relevant Excel files, including those for segment data and connectivity matrices for each segment. Additionally, the folder "ADHDexcel" contains the results of patient measures for each recorded segment for individuals with ADHD, while a similar organization is applied to non-ADHD patients under the "Non-ADHDexcel" folder. The "preprocessADHD" and "preprocessNon-ADHD" folders include datasets that have undergone preprocessing and cleaning.

# Maintenance Guide

To run our application, start by extracting the EEG Python folder from the zip file and placing it in the C drive on your computer. Next, download and install PyCharm from its official website. Once installed, open PyCharm, click on Open, and select the EEG Python folder from the C drive. After opening the project, locate the terminal at the bottom of the screen, and type the command `pip install -r requirements.txt` to install all the necessary libraries required for the application. Once the installation is complete, run the application by typing 'python go.py` in the terminal. This will start the software, displaying the login screen. As outlined in the user guide, log in by using

Username: koraltopaz and Password: 1234.

Downland PyCharm:

[Download PyCharm: The Python IDE for data science and web development by JetBrains](https://www.jetbrains.com/pycharm/download/?section=windows)

This is terminal button:



To use our project with a different dataset, you first need to replace the existing dataset by placing the ADHD patient recordings into the preproccesADHD folder and the control recordings into the preproccesNonADHD folder, ensuring the data is in Excel format. If you want to divide the recordings into different segments, access the "segment\_eeg\_to\_8\_seconds" function in create\_data, change the number "8" to your desired segment length, and adjust the frequency as needed. Additionally, if you want to apply a different threshold, modify the Apply\_threshold function and update the threshold in the combo box on the analyze graphs screen.

The code is not limited to EEG recordings with 19 electrodes.

# Reference:

1. Ekhlasi A, Nasrabadi AM, Mohammadi MR. Direction of information flow between brain regions in ADHD and healthy children based on EEG by using directed phase transfer entropy. Cogn Neurodyn. 2021 Dec;15(6):975–86
2. Pernet CR, Wilcox R, Rousselet GA. Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. Front Psychol. 2013 Jan 10;3:606.
3. Turney S. Scribbr. 2022 [cited 2024 Apr 6]. Pearson Correlation Coefficient (r) | Guide & Examples. Available from: https://www.scribbr.com/statistics/pearson-correlation-coefficient/
4. Graph Measures [Internet]. BRAPH. [cited 2024 Apr 6]. Available from: <http://braph.org/braph-1-0/manual/graph-measures/>
5. What is ADHD? [Internet]. [cited 2024 Apr 6]. Available from: https://www.psychiatry.org:443/patients-families/adhd/what-is-adhd
6. Community Detection [Internet]. [cited 2024 Apr 6]. Available from: https://noesis.ikor.org/wiki/algorithms/community-detection
7. Community detection using efficient modularity optimization method: labelmod with single and multi-layer graphs | Semantic Scholar [Internet]. [cited 2024 Apr 6]. Available from: <https://www.semanticscholar.org/paper/Community-detection-using-efficient-modularity-with-Bang/813dc781d5d44f56ea1fc1454d8b441f2f6069a5>
8. <https://en.wikipedia.org/wiki/Average_path_length>
9. <https://en.wikipedia.org/wiki/10%E2%80%9320_system_(EEG)>

Git link:

<https://github.com/KoralBiton18/Capstone_Project_Inception.git>