Eötvös Loránd University
Faculty of Science

Dániel Hegedűs
BSc in Mathematics

Analysis of human braingraphs and the Alzheimer's disease using mathematical tools

Bachelor thesis

Supervisor: Dr. Vince Grolmusz
Department of Computer Science

Budapest, 2024

Acknowledgement

I would like to thank my supervisor, Vince Grolmusz, who contributed a lot to the ideas found in the thesis and with whom I had many really instructive and enjoyable brainstorms. I really appreciate the work of Bálint Varga, who calculated the graphs I could use, this work would not be possible without him. I really want to thank Péter Csikvári for organising the research at the start of my Bachelor. This opportunity really motivated me throughout the three years and means a lot to me.

I am grateful to my close friends, who motivated me so much and with whom I could spend my wonderful evenings in the college. I want to give a special thanks to my girlfriend, who always helped me with her constructive feedback, inspired me a lot, and supported me with my projects.

I want to thank my teachers the effort and work they put into teaching me: I learned many-many fascinating things presented in amazing ways. Last but not least, I would like to thank my family for starting me on this journey and letting me chase my passion in Mathematics.

The research has been implemented with the support provided by the Alfréd Rényi Institute of Mathematics and by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the ELTE TKP 2021-NKTA-62 funding scheme.

Table of Contents

Chapter 1

Introduction

The human brain is undoubtedly the most complex organ known to humanity, and its function and structure have captivated researchers for centuries. In this work, we analyze the structure of the human brain using braingraphs. These are mathematical graphs: their vertices represent different brain regions, while the edges represent their connections. First, we prove that many importance criteria correlate both in the case of vertices and edges. We will compare these results with observing the location of average communication of the vertices. Our primary results involve finding differences between brains of female and male subjects and noting significant differences connected to brain areas having specific responsibilities, such as short-term memory and attention orientation. We also make a significant conclusion about the scientifically important role of caudate in dementia, as we notice higher connectedness of the caudate in demented people. This is crucial, as the thoughts about the role of caudate are not consistent in the literature, as different authors conclude multiple possibilities in this topic.

1.1 Background of working with braingraphs

MRI datasets need to fulfill several quality-related characteristics for braingraph construction. The first one is the high and consistent quality of the images. That is not surprising: if we want to work with graphs calculated from MR images, there are two steps where inaccuracies can happen. Both in the MRI-making phase and in the graph-calculating phase, roundings occur, which imply inevitable inaccuracies. In the past decade, humanity has been able to use technologies to make the construction of the braingraph possible. The second problem is that most MRI sets contain few MR images. That is a big problem, as from a set of 10-20 MR images, it is hard to have a significant conclusion. Furthermore, two different MRI sets can hardly be compared because of the possibilities of various methods the data acquisition modalities use during the MRI scans. Another area for improvement is that MR images are not precise in the sense that significant differences can occur because patients can lie in positions that are not precisely the same.

1.2 Some earlier results

Because of the above-written reasons, the method of using braingraphs is new to science. One of the first MRI datasets that is large enough and of good quality was provided by the Human Connectome Project [1]. The primary dataset of young adults contains 1064 people's MRI, for each person one MRI. The subjects are only healthy young adults, so the dataset can be better used for biological research rather than medical investigations. From this dataset, braingraphs were calculated and described in [2]. From these graphs, the Budapest Reference Connectome Server [3] [4] was produced, which is an online, parametrizable graph-visualization software. The data was used internationally: for example, graph convolutional networks were investigated in [5]. The data can be used for finding differences between sexes, as in [6]. Some results in this thesis are partially covered by our published article [7] and our preprint [8].

1.3 Construction of braingraphs

Through the thesis, we will work with two datasets: one is from the Human Connectome Project [1], and the other is from the OASIS Brains Project [9]. The set of graphs from the Human Connectome Project will be called the HCP graphs, and the set of the ones from the OASIS Project will be called the OASIS graphs. It is worth mentioning again that the subjects in the HCP dataset are all healthy and only include young adults, while the OASIS dataset also has demented people.

The HCP dataset includes diffusional MR images of 1064 individuals, while the Oasis dataset includes diffusional MR and PET images of 1098 individuals. Diffusional MRI is a type of magnetic resonance imaging that examines water molecules, using the fact that they move differently near axons. PET stands for positron emission tomography, which uses radioactive substances to visualize metabolic changes [10]. Both datasets include females and males. Using the Lausanne2008 atlas [11] for the HCP dataset and the Lausanne2018 atlas [12] for the OASIS dataset, the surface of the brain was divided into brain regions, each approximately 1 cm^2 in size. In the case of the HCP dataset, we worked with 1015 brain areas, while in the case of the OASIS dataset, we worked with 124 and 1058 areas. The vertices of the braingraphs are the aforementioned brain regions, with consistent labels from 1 to n across individuals (where n denotes the number of used areas). Weighted edges are drawn between vertices i and j based on the number of axons (i.e., the thickness of the nerve bundle) connecting brain regions represented by i and j. Additionally, we will use the length of nerve fibers, but the weight of the edge is the thickness. It is important to note that the length of an edge is not the geometric distance between the two vertices, as the geometric locations represent only three coordinates. Moreover, axonal fibers are geometrically not close to line segments, so the coordinates and the lengths are also saved in the graphs. The coordinate system determining the coordinates of brain regions has three axes: the x-axis for left-right, the y-axis for front-back, and the z-axis for up-down direction. The larger the coordinate value, the farther the vertex is forward, upward, and to the left. The fractional anisotropy of the fibers is also saved in the graphs, but we will not address this in my

thesis. Fractional anisotropy can be used to characterize the diffusion profile of water, which is not necessary for our purposes here. I did not create the graphs themselves: Bálint Varga calculated the graphs before I started my work based on the MRI data. We can project these braingraphs onto a cross-section of the brain: one tool is the Budapest Reference Connectome Server [3] [4] [13].

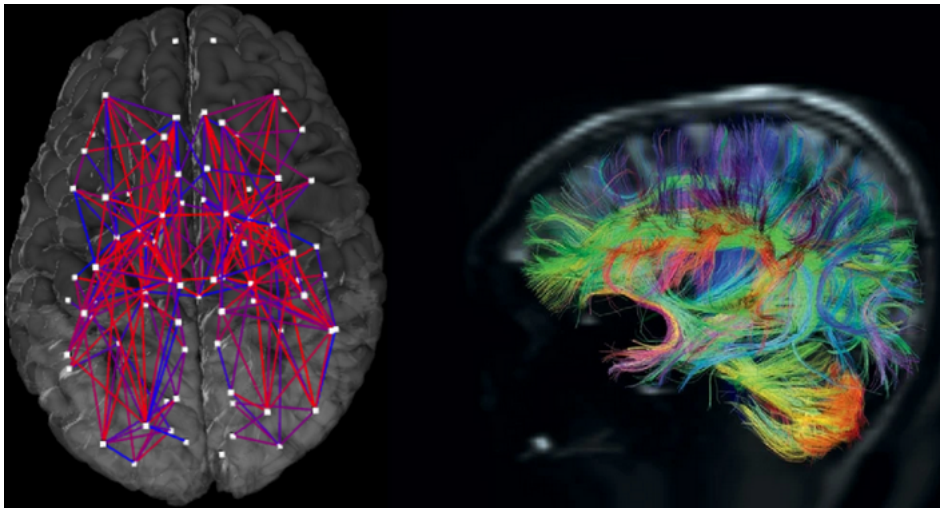

Figure 1. and Figure 2. The first figure shows a braingraph projected onto a human brain. White dots indicate the position of the different brain regions. We connect two brain regions with an edge if we detect axonal fibers between them. The colour of the edge carries information about its thickness. On the red-blue scale, the more red an edge is, the thicker it is. In the second figure, the physical location of the axonal fibers can also be observed, with different colours indicating different directions. It can be seen that the axonal fibers are not close to being straight.

As already said, braingraphs were calculated from the datasets by Bálint Varga in five different resolutions. During our research, we only used the highest resolutions for both datasets, and we also used the lowest resolution in the case of the OASIS dataset. The usage of the different resolutions has different advantages and disadvantages. The benefit of higher resolution is that we can investigate smaller parts of the brain, making it easier to be more specific in our conclusions. The disadvantage is that the new vertices are often obtained by division from vertices in lower resolutions, and inaccuracies are much more significant in higher resolutions. Additionally, in lower resolutions, the larger brain areas usually correspond to only one or two vertices and can be trusted more. All the graphs used are undirected. Loops (self-edges) only exist in the case of the OASIS dataset.

The entire graph analysis was written by myself, in Python. The codes were written in the IDE Pycharm. The main libraries that were used are Pandas [14] for dataframes, Scipy [15] for mathematical formulas and statistical investigation, Networkx [16] for graph properties, Matplotlib [17] for diagrams and Paramiko for the easier usage of the datamine server.

The Oasis Project also published data on the subjects' different basic properties. These include their gender, handedness, age, educational level, race, height, weight, etc. In addition, the results of the psychological examination of the subjects were also published. These include the Clinical Dementia Rating (CDR) [18] and the conclusion about the level of dementia. Some

subjects were examined multiple times, and all their results were published. For this reason, we call a subject demented if at least one psychological test concluded dementia. The attributes of the subjects were collected and merged into two easy-to-use Excel tables.

Unfortunately, in some cases, the published psychological results are incomplete and therefore, cannot be used. Also, only some of the graphs are available in the desired form, so we will only work with the subjects with complete psychological results and a proper graph.

Chapter 2

Methods and preliminaries

2.1 Spearman correlation coefficient

The Spearman correlation coefficient is ideal for comparing index orders [19]. The coefficient provides information about the correlation between two attributes, using the orderings based on these properties. Thus, each element will have an index according to the two properties, including its position in the respective ordering. There is a simple formula for calculating the coefficient, assuming that the indices are unique, meaning that no two elements are equal according to any properties. If we calculate the attributes of n elements and denote the difference between the i-th element's two indices by d_i, then we have the following formula:

$$\varrho = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n^3 - n}. \tag{2.1}$$

In our case, this will hold, but we do not need a nice formula, as the coefficient is not calculated by hand. The coefficient satisfies $-1 \leq \varrho \leq 1$, where 1 indicates perfect correlation and -1 indicates perfect inverse correlation. The latter can easily be demonstrated using the closed formula $\frac{n(n+1)(2n+1)}{6}$ for the sum of the first n square numbers. For total inverse correlation, we need to calculate the sum of the first $\frac{n}{2}$ odd square numbers, which can be easily obtained from the sum of the unused even square numbers, as it is four times the desired sum. The closer the coefficient is to 0, the less we can speak of any correlation. However, as n increases, even smaller absolute values of ϱ can be significant. Therefore, the p-value is determined not only by ϱ but also by n and ϱ together. The obtained p-value indicates the chance of such a degree of correlation under the null hypothesis, assuming that the two variables behave randomly relative to each other. Therefore, the level of correlation is determined by the p-value, but for the sake of completeness, we also publish the ϱ coefficients. I performed the precise calculation of the Spearman coefficient and the p-value using the `spearmanr` function of the SciPy package [15].

2.2 Averaged braingraphs

Analysing individual braingraphs from n people to draw general conclusions about the human brain can be challenging. Therefore, instead of analysing individual graphs, we can average them

to create a consensus graph. We averaged the graphs in the following way. The position of each vertex is calculated by simple averaging, placing different brain regions in an average-sized brain. (The positions are still given with three coordinates: the origin is in the lower-right corner of the back of the brain.) Averaging the attributes of the edges is more complicated. The first step is to sum all the values obtained from the people. The first non-trivial question is whether the length and thickness of the edges should be averaged differently. It makes no sense to divide the sum of lengths by n because most edges only appear in some of the subjects, so we would lose information in this case since we would obtain a disproportionately small number for the length. Therefore, in the case of length, we must divide by the number of individuals in which the given edge appears. On the other hand, in the case of thickness, it is informative whether 0 thickness occurs in some people. Therefore, in the case of thickness, we should divide by n. So, using this averaged graph, we do not need to perform calculations on n graphs in every case, it is sufficient to work only with the averaged graph. Through the thesis, some results will use the averaged graph, and some will not, so for every result, we make it clear whether we analyse the individuals or the averaged graph. It is essential to consider that every edge in the averaged graph, appears in at least one person, as we divide in both cases by a positive number. Thus, if $\#(i,j)$ denotes the number of occurrences of edge (i,j), and $w_{i,j,k}$ and $\ell_{i,j,k}$ are the weight and length of edge (i,j) in the k-th person respectively, then the equations for the edges of the averaged graph are as follows.

$$w_{i,j} = \frac{\sum_{k=1}^{n} w_{i,j,k}}{n} \tag{2.2}$$

$$\ell_{i,j} = \frac{\sum_{k=1}^{n} \ell_{i,j,k}}{\#(i,j)} \tag{2.3}$$

An averaged graph can be calculated separately for different subject groups, such as female/male or demented/healthy. The main differences between the pair of graphs suggest tendencies of differences between the brains of the examined groups.

2.3 Basic properties related to inversion numbers

In this section, we prove some basic properties of inversion numbers.

Definition 2.3.1. Two elements in two given permutations of a finite set are in *inversion* if their order is opposite in the two permutations.

Definition 2.3.2. The *inversion number* of two permutations is the number of pairs of elements that are in inversion. The *inversion* of an element is the number of other elements with which it is in inversion.

Lemma 2.3.1. The expected value of the inversion number between two permutations picked uniformly fromt the symmetric group S_n is $\frac{n(n-1)}{4}$.

Proof. Consider a fixed unordered pair (i, j). Then, due to symmetry, the expected value of the inversion between them is $\frac{1}{2}$, since for every permutation, there exists bijectively another one, where only the (i, j) transposition is different. (We call a function that swaps two elements in a permutation a transposition.) It is known that $\binom{n}{2} = \frac{n(n-1)}{2}$ unordered pairs can be chosen from n elements. By using the linearity of expectation, we obtain that the expected value is indeed $\frac{1}{2} \cdot \frac{n(n-1)}{2} = \frac{n(n-1)}{4}$. □

Corollary 2.3.1. The expected inversion of any element is $\frac{n-1}{2}$.

Proof. We can again use the linearity of expectation based on Lemma 2.3.1. Since there are n elements, and every inversion is counted for both elements, the expected value for a single element is $\frac{n(n-1)}{4} \cdot \frac{2}{n} = \frac{n-1}{2}$. □

Alternative proof. Pair each permutation bijectively with its inverse permutation. Then for every unordered pair of elements (i, j), they are in inversion in exactly one of the permutations in each permutation pair. Therefore, for any element i, there are $n-1$ different j elements, and each has an expected inversion of $\frac{1}{2}$, yielding an expected inversion of $\frac{n-1}{2}$ by linearity for each element. □

2.4 Graphs with maximal edges

We want to examine the structure of the brain not only as a whole but also by looking at the structure of only the thickest edges. This way, we can only investigate the most crucial connections. It should be mentioned that the motivation is not to filter out the distorting thin edges, as for that goal, looking at only the thickest $p\%$ would be more efficient, with some well-chosen p. Actually, p could depend on the data, as we do not know anything about the distribution of the thickness of the edges in advance.

Definition 2.4.1. Consider any weighted simple graph with pairwise distinct edge-weights. Call the associated *maximal graph* the graph whose vertices are the vertices of the original graph, and its edges are exactly those edges that are the maximal weighted edge of at least one of their endpoints. Call the *extended maximal graph* the one whose edges consist of exactly the two largest weighted edges for each vertex.

In the case of the brain with n vertices, we selected n edges considering multiplicities in the maximal graph, but there may be edges that we chose twice. First, let us state and prove a general graph-theoretical lemma connected to this.

Lemma 2.4.1. Consider any simple graph with weighted edges and suppose that all the edge weights are different. We state that the corresponding maximal graph is acyclic, i.e., it does not contain any cycle of length at least three. This type of graphs are called forest graphs, as all their components are tree graphs.

Proof. Let us assume indirectly that the graph has a cycle and let its length be $k \leq n$. Enumerate the vertices of the cycle in order with the numbers $1, 2, \ldots, k$.

We know that any edge has two endpoints, so any selected edge has the largest weight for one of its endpoints. Since both endpoints of each edge in the cycle are contained in the cycle, it must have been selected because of one of the vertices in the cycle. We know that exactly one edge has been selected for each vertex, so there is a bijection between vertices and edges in the cycle based on for which vertex we selected which edge.

Since the edge $(1, 2)$ was selected, without the loss of generality, we can assume that it was selected for vertex 1. However, as the vertex 2 has two edges in the cycle and one of its edges already has a bijection pair found, the edge $(2, 3)$ must have been selected for vertex 2. This means that the weight of the edge $(2, 3)$ is greater than the weight of the edge $(1, 2)$ because otherwise the edge $(2, 3)$ would not have been selected for vertex 2. Proceeding along the cycle with the previous reasoning, we get that the weight of each new edge is greater than the previous one. But then, denoting the weight of the edge (i, j) by $w(i, j)$, we obtained that $w(1, 2) < w(2, 3) < \ldots < w(k - 1, k) < w(k, 1) < w(1, 2)$. This is a contradiction as $w(1, 2) < w(1, 2)$ cannot hold. □

Remark 2.4.1. The previous proof can also be said by starting from the largest weighted edge in the cycle. We can assume without the loss of generality that this is the edge $(1, 2)$, and it is selected for vertex 1. However, due to the selection at vertex 2, the edge $(2, 3)$ must have a higher weight, which is a contradiction. Another alternative proof is that no edge could be the minimal weighted edge in the cycle. That is because, for both endpoints of this edge, the other edge from the endpoint has greater weight, meaning that the least-weighted edge could not have been paired for either of its endpoints in the bijection.

Remark 2.4.2. It is worth noting that during the proof, we only used the fact that no two edges sharing an endpoint have the same weight. So, the condition that every weight is different can be replaced by the more general one, only needing the different weights for the vertices locally.

For each component of the maximal graph, we can select the most important vertex. The importance criterion for the vertices is the sum of weights of their edges (equality does not hold in this ordering either).

Definition 2.4.2. The *center* of a graph (V, E) is the set of v vertices, for which $\max_{u \in V} d(u, v)$ is minimal, where the function d denotes the graph-theoretical distance for unweighted graphs. (The definition makes sense as the distance is unique for any pair of vertices.)

Lemma 2.4.2. For tree graphs, the center consists of at most two vertices.

Proof. Let us assume indirectly, that the graph has three vertices in the center. Then, as the graph is acyclic, there must be two of them that are not neighbours, let them be u and v. Then, there must be a vertex $s \neq u, v$ on the unique path uv. As u and v are both in the centers, the vertex having the largest distance to v is in the direction of u from v and vice versa. This implies that s also must be in the center, as otherwise, moving u towards v and v towards u

would decrease the maximal distance of u and v, so they would not be in the center. However, this means that in the path uv, which is not empty, the maximal distance of s is smaller than the maximal distance for u and for v. This is because the vertex with maximal distance for u and for v, must be in the direction of s. This is a contradiction, as s would have a smaller maximal distance than the vertices in the center. □

The question arises: what biological role do connectivity properties play in the brain? Let us consider why it is essential for the brain to have strong edges connected cohesively. Connectivity means that any pair of vertices can communicate with each other through the specified thicker, faster fibers. Naturally, the brain is the most effective when communication can occur from any vertex to any other vertex through the fastest, highest-weighted edges. The mentioned property is fulfilled when the most important edges form a connected component encompassing the entire brain, allowing any two vertices to communicate.

Biconnectivity (nonseparability) of a graph means that even in the case of removing any vertex, the remaining graph stays connected. This is a crucial brain property, because the ability to bypass functions of certain brain areas is important due to diseases or aging. Some vertices can take over the role of others to some extent, this phenomenon is called neuroplasticity.

Furthermore, it is biologically advantageous when the most important vertices are central in each component. That is because this way, most of the vertices are close to the most important vertex in the components.

2.5 Background for Alzheimer's disease

Alzheimer's disease is a progressive neurodegenerative disorder that primarily affects the brain, leading to a decline in memory, cognitive function, and the ability to perform daily activities. It is the most common cause of dementia among older adults. The hallmark characteristics of Alzheimer's include the formation of abnormal protein deposits called plaques and tangles in the brain, which disrupt communication between nerve cells and ultimately result in death.

As the disease progresses, individuals may experience memory loss, confusion, difficulties in problem-solving, language impairment, and changes in behavior. Alzheimer's disease significantly impacts not only the affected individuals but also places a substantial burden on their families and caregivers.

While there is currently no cure for Alzheimer's disease, various treatments and interventions aim to manage symptoms and improve the quality of life for individuals living with the condition. Ongoing research seeks to understand the underlying causes of Alzheimer's more and develop potential therapies to slow down or halt its progression. Early detection and timely intervention are crucial in managing the disease and supporting the affected individuals and their families. In 2020, the approximated number of people worldwide affected by Alzheimer's was around 50 million [20] [21]. Throughout the thesis, dementia and Alzheimer's will be used for the same disease. This is because science can hardly differentiate patients with only the psychological tests used for diagnostics in the OASIS Project. This implies that we can only talk about the diseases of dementia and Alzheimer's together.

Chapter 3

Importance attributes of brain regions

The importance of different vertices can be measured from various perspectives. We can examine the degree of the vertices, or we can look at the weight (thickness) and length of their edges. Moreover, in the case of weight and length, we have three options: maximal, average, and sum. Each of these metrics is meaningful, for example, the average weight indicates the average thickness of nerve fibers, while the sum of weights shows how many neurons communicate with other brain regions. (Of course, these two measures are not the same because the sum of many small weights may be the same as a few thicker edges, yet their averages are not equal.) When examining the length of edges, we can gain information about how uniform the length of communication of the vertex is.

3.1 Objectives and tools

It is important to emphasize that our goal is not to determine the important vertices according to various criteria. Instead, the aim is to rank the vertices according to different criteria and compare the orderings. In other words, we want to determine if the vertices that are expansive, i.e., communicate over longer distances, are whether also those with thicker nerve fibers.

Let us take the seven importance criteria and rank the vertices according to each criterion in descending order. More precisely, their associated labels/indexes are ordered, as each brain region has been assigned an index. Of course, we consider vertices that appear earlier in the orderings as more important. For each importance criterion, we thus obtained an index order, each is a permutation of positive integers not exceeding 1015 as elements. It is important to note, that since there is no equality between any two vertices for any criterion, the orderings are well-defined.

In the tables, we fix one of the examined sequences, this is in the first row, while the other sequence for the correlation test is in the second row. The Spearman coefficient is in the third row, and the p-value is in the bottom row. A p-value of 0.0 indicates a value less than 10^{-323}. Furthermore, rounding errors may occur in other values as well, but they are negligible at this order of magnitude.

3.2 The correlation between the different attributes

Correlation if one ordering is by degree

	Degree					
	Weight-sum	Weight-max	Weight-average	Length-sum	Length-max	Length-average
ϱ	0.88	0.84	0.79	0.98	0.51	0.76
p	0.0	10^{-278}	10^{-220}	0.0	10^{-68}	10^{-194}

Table 1. The Spearman correlation coefficient ϱ, unlike the Pearson correlation coefficient r, is significant even at 0.5 for sufficiently large datasets, which we can see here at $n = 1015$. Therefore, we can see that vertices with high degrees strongly correlate with the ones having thick or long edges, regardless of the type of criterion. This is not surprising when considering the maximal or sum of properties, but it is for the average. This indicates that the vertices with thick edges have not just sporadically thick edges but rather uniformly. The Spearman coefficients are in the third row, and the p-values are in the bottom row.

Correlation if one ordering is by sum of weight, then by maximum of weight

	Weight-sum			Weight-max		
	Length-sum	Length-max	Length-average	Length-sum	Length-max	Length-average
ϱ	0.86	0.42	0.63	0.83	0.41	0.62
p	10^{-302}	10^{-45}	10^{-112}	10^{-262}	10^{-43}	10^{-110}

Correlation if one ordering is by average of weight

	Weight-average		
	Length-sum	Length-max	Length-average
ϱ	0.77	0.37	0.54
p	10^{-199}	10^{-33}	10^{-79}

Table 2. and Table 3. We can observe that ordering by length and thickness correlates in every way. This is not intuitive because there is no trivial reason why the expander vertices would be the ones with thick edges. The Spearman coefficients are in the third row, and the p-values are in the bottom row.

Correlation between weight attributes, then between length attributes

	Attributes with weight			Attributes with length		
	Sum-Max	Sum-Average	Max-Average	Sum-Max	Sum-Average	Max-Average
ϱ	0.97	0.98	0.96	0.58	0.86	0.70
p	0.0	0.0	0.0	10^{-91}	10^{-296}	10^{-148}

Table 4. We can see that the different weight-based orderings and the different length-based orderings correlate with each other. Of course, this follows from the previous tables, as the relation of such a strong correlation is "transitive". The Spearman coefficients are in the third row, and the p-values are in the bottom row.

From the tables, we can read that for 1015 vertices, we obtained extremely high coefficients and negligibly low p-values. This implies that there is a remarkably strong correlation between the different rankings of brain regions according to the various criteria. Hence, we have determined that the seven criteria of importance correlate. In other words, the same brain regions

are important in terms of expanding as those with thick nerve fibers, indicating strong/fast connections. However, this does not necessarily imply the same for edges: it is not necessarily true that thicker edges are longer. We will address edge analysis later.

3.2.1 A simple control

For the sake of completeness and certainty, it is worth mentioning that with the previous examination, we truly observed random behavior. For instance, we can compare the basic labels of the vertices with the indices by the average weight of their edges. The comparison gives $\varrho = 0.01$ and $p = 0.65$. This magnitude of p-value is only due to the fact that indexing is not entirely random. The vertices of the two hemispheres are numbered consecutively: the first approximately 500 vertices are located in one hemisphere, and the second approximately 500 vertices in the other hemisphere, in the same order. So the fact that the most important vertices in one hemisphere received late numbers also implies that the corresponding important vertices in the other hemisphere receive large numbers as well. Of course, this does not imply any correlation, but it only confirms the correctness of our method for comparison.

3.3 Comparing permutations with inversion numbers

So, we have established that the correlation between the criteria of importance is extremely strong, but we have not obtained any tangible or convincing result that would help us understand why this is the case. Therefore, it is worth examining on a vertex-by-vertex basis how different the orderings are. More precisely, unlike the Spearman method, here we will not examine how much the positions differ between the two permutations, but we will examine the inversion numbers. Of course, both methods yield the same result, but this rather tells how different the positions of the vertices are relative to each other according to different properties. In the following, when we talk about permutations, we always refer to those of length n, whose elements are exactly the integers from 1 to n.

Below, we can see diagrams showing how many vertices have a higher inversion than the expected value for the two given importance criteria. The expected inversion is the previously calculated $\frac{n-1}{2}$. Now, we examine how many vertices have inversions greater than this. The black line indicates the expected value of the number of vertices with large inversions for the two specified criteria. The other colors represent different pairs of properties. On the x-axis, at each of the n points (for $n \leq 1015$), we indicate the case when only the first n most important vertices were investigated. On the y-axis, we show how many of these vertices have inversions greater than the expected value (for permutations of length n). This allows us to easily see that there is no substantial difference between the inversions of important and less important vertices.

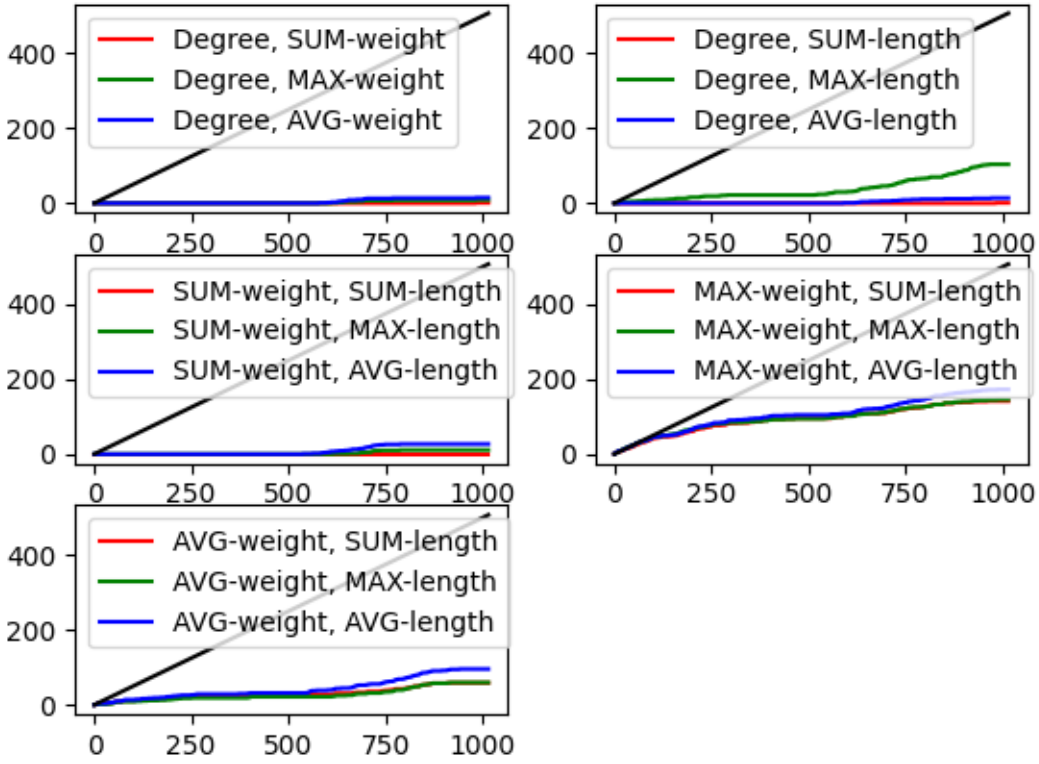

Figure 3. The number of vertices with high inversions according to the different pairs of properties. On the x-axis at each of the n points, we indicate the case when only the first n most important vertices were investigated, according to the first importance criterion. On the y-axis, we show how many of these vertices have inversions greater than the expected value (for permutations of length n). In each diagram, the black line indicates the expected value. It can be seen from the diagrams that for all examined 15 pairs of properties, there are only a few vertices with large differences in their positions in the rankings.

Thus, we have indeed understood that different criteria of importance are strongly correlated. There will be a similar group of vertices with high degrees as those with thick edges or long connections. Among these, the relationship between degree and sum of lengths is not unexpected, as with more edges, we would expect their sum of lengths to be greater. The really surprising result is that the average weight also correlates with the degree. The fact that we examined this is unique, as in medical science, the emphasis is usually placed in the opposite direction. Their goal is to determine which vertex is responsible for what. In this thesis, this is not the focus of our interest. Of course, during the analyses, we can easily retrieve the corresponding brain regions from the labels of the vertices. Thus, we can name the most important brain regions according to the above criteria, and for completeness, we do so. The few most important brain regions based on their degrees and average of edge weights (in both the right and left hemispheres) are Caudate, Putamen, Thalamus-Proper, Hippocampus and Pallidum, i.e., the subcortical nuclei.

It is important to note that the above does not imply functional characterization. Moreover, based solely on braingraphs, this is impossible, as only the location of brain regions and the

connections between them are saved in the graph. To obtain results about the function of specific vertices, functional MRI (fMRI) scans should be compared. During this type of scanning, the subject performs some simple tasks (e.g., counting or reading). When different tasks are performed, different brain regions will be more active, hence revealing which areas have a greater role in performing the task [22].

Chapter 4

Correlation of edge attributes

In the case of edges, we can only examine two properties in the averaged braingraph: weight and length. This is because the purpose of the braingraph is to characterize the system of brain connections, and in this case, we want to compare the connections themselves. Therefore, only the two properties that are fundamentally saved in the graph are available. So, we can arrange the edges based on their length and weight.

4.1 Correlation between edge thickness and length

Let us examine the Spearman correlation between the index orders obtained from the weight and length of all edges. We perform this on the 99171 edges present in the averaged braingraph (with non-zero weight). The two sequences are ordered by thickness and length. We found that the Spearman coefficient is -0.51, and the corresponding p-value is 0.0. (It is important to note here as well that our current dataset is so large that even for this relatively small absolute value of coefficient, there is a negligible p-value.) This seemingly contradicts the previous descriptions stating that edge length and thickness would positively correlate, as we now find that shorter edges are thicker. To understand this, we created a diagram whose points represent the edges. The x-coordinate of each point indicates its ranking by thickness, and the y-coordinate indicates its ranking by length. In the diagrams, we will only show the edges that occur in at least five subjects. That is because too many edges have occurrence at most four, and equalities cause artifacts in the orderings and thus produce distorted diagrams. For this reason, we show only the 67281 edges, which have occurrence number at least five. It should be noted, that the statistical calculation is not distorted as the ϱ-coefficient and the p-value deal with the equalities.

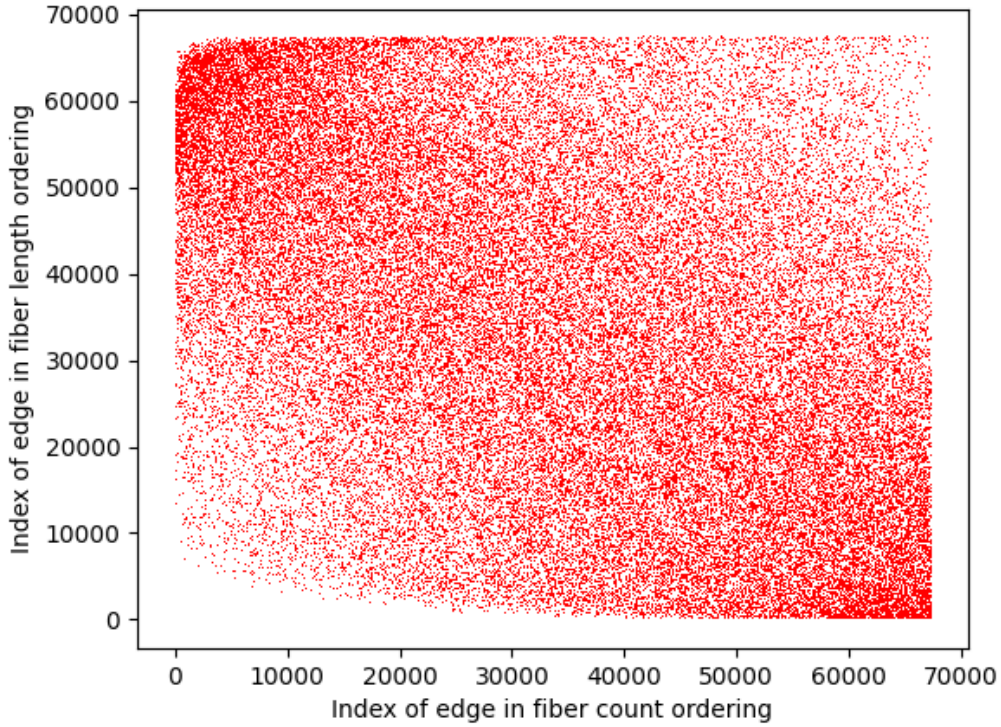

Figure 4. Comparison of edge thickness and length indices. Each point on the diagram represents an edge, where the x-coordinate is the ranking by thickness, and the y-coordinate is the ranking by length. The strong correlation is because the majority of edges lie close to the minor diagonal (top-left to bottom-right corner). Only the 67281 edges with occurrence number at least five are shown. The Spearman coefficient for all the 99171 edges is -0.51, and the corresponding p-value is 0.0.

The strong correlation arises because there are no really thick edges among the longest ones. Also, it is rare for an edge to be both short and thin at the same time. Therefore, the opposite correlation arises: the thicker an edge, the more likely it is to be short, and similarly, the longer an edge, the more likely it is to be thinner. However, this seems to create an even stronger apparent contradiction: there was a correlation among vertex importance. How can thicker edges be shorter? The explanation lies in the fact that in the case of edges, there is a correlation globally but not locally. For example, examining the first 1000 edges reveals no real correlation locally. Thus, the paradox arises from the anomaly that there can be a correlation between two properties even when there is no local correlation between them.

4.2 Correlation with occurrence numbers

Edges have a property that we cannot read from the original averaged braingraph: how many individuals they appear in. When creating the averaged graph, we can also save that how many individuals each edge occurred in. Naturally, it seems reasonable that occurrence numbers should correlate with the weight of edges. This would imply that edge length and occurrence count would correlate inversely. Indeed, the correlation between occurrence and thickness is 0.97,

with a p-value of 0.0, while comparing with length, the coefficient is -0.47, with a p-value of 0.0. This indicates an incredibly strong correlation in both cases, especially in the earlier (due to the exceptionally high coefficient). To illustrate this, let us look at two diagrams that are similar to the previous one.

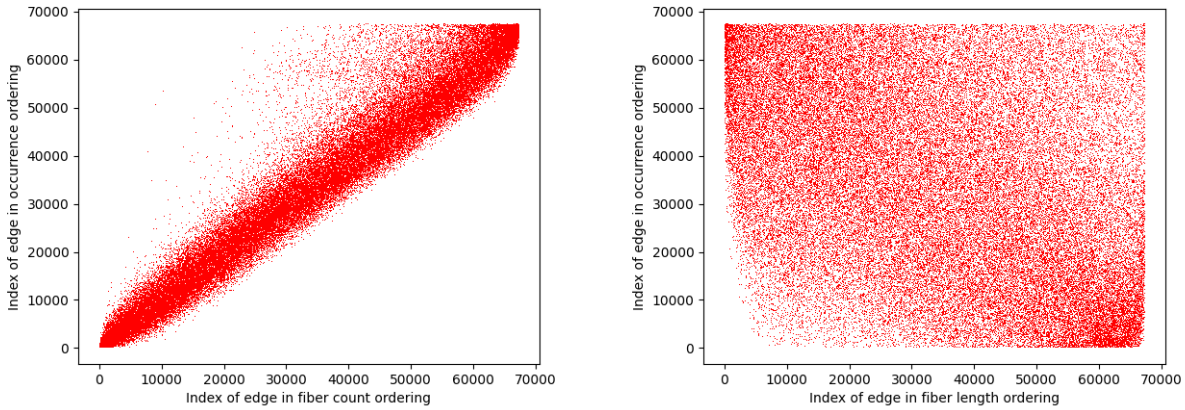

Figure 5. and Figure 6. Comparison of occurrence number with edge weight and edge length respectively. Each point on the diagram represents an edge, where the x-coordinate is the ranking by weight and length, while the y-coordinate is the ranking by occurrence. Only the 67281 edges with occurrence number at least five are shown. The Spearman coefficients for all the 99171 edges are 0.97 and -0.47, and the corresponding p-values are 0.0 and 0.0 respectively.

Chapter 5

Location of brain regions

We can notice that up to this point in the thesis, we have not addressed the fact that brain regions also carry geometric information. However, in the case of an organ like the brain, we should remember that brain regions have spatial relationships with each other. While we have examined how long fiber bundles are, we have only done so for pairs of vertices, not in relation to the entire brain.

5.1 Centroids and weighted centroids of vertices

For a vertex labeled i, let H_i be the set of vertices that are adjacent to the vertex i.

Definition 5.1.1. The *simple centroid* of the vertex i is the arithmetic mean of the positions of the vertices in H_i.

Definition 5.1.2. The *weighted centroid* or *modified centroid* of vertex i is the weighted average of the positions of the vertices in H_i, where the weight of an endpoint is the thickness of the edge connecting it to i.

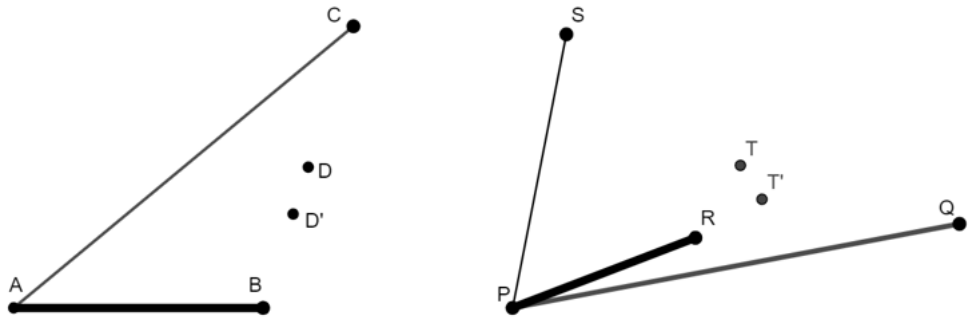

Figure 7. We can see that the simple centroid of the edges departing from point A is point D. However, by doubling the weight of edge AB and computing the weighted centroid, we get $D' \neq D$. Similarly, from point P, the simple centroid of edges PQ, PR, and PS is T, but by doubling the weight of PQ and tripling the weight of PR, we obtain $T' \neq T$. In the case of braingraphs, much larger differences in edge weights are common, so both types of centroids should be examined.

Of course, the simple centroid of each vertex provides information about where the vertex communicates on average. In the case of the weighted centroid, we consider the strength of communication as well. The latter is better for characterizing where the communicated information from the vertex on average reaches or where information arrives from.

The question may arise that how far the vertices are from their centroids. Since we have observed that shorter edges are thicker, we expect the weighted centroids to be closer to the vertices than the unweighted centroids. And indeed, this is the case: we calculated both centroids in terms of coordinates one by one and by Euclidean distance. From the table, we can read the average distances both in the cases of simple and modified centroids according to the x, y, z coordinates and Euclidean distance in millimeters.

Distances from centroids

Simple_x	Simple_y	Simple_z	$\text{Simple}_{\text{Eukl}}$	Mod_x	Mod_y	Mod_z	Mod_{Eukl}
5.38	7.92	6.12	12.80	4.04	4.77	4.08	8.46

Table 5. We can read in millimeters the average distances of the vertices from the simple and weighted centroids by coordinates and by Euclidean distance.

It is not surprising that the largest distance is along the y-coordinate. Given the physical properties of the brain, which is longer in the forward-backward direction, the component of the edges by the y-coordinate is the largest. However, the difference between coordinates is significantly smaller for modified centroids: instead of a 2.4 mm difference, there is only a 0.7 mm difference. This suggests that the edges with lower weights mostly expand parallel to the y-axis.

It is worth considering that we have only used the regular distance so far, but instead, we could use signed distance. At first glance, one might think that the sum of the signed distances is the null vector, but that is not the case. The degree of the two endpoints of a given edge may differ, resulting in calculations with different weights in the two directions. We will not deal with signed distance in this thesis.

5.2 Placement of vertex centroids

The examined brain regions are in the gray matter of the cerebral cortex, while the nerve fibers run in the white matter. It is a legitimate question to ask how symmetrically the edges run from the vertices.

Definition 5.2.1. The unweighted average of all vertex coordinates is referred to as the *center of the brain.*

This is an appropriate definition, because the brain regions are "evenly" distributed, so it will carry the information about where the vertices are on average.

Definition 5.2.2. Associate to each vertex the line on which the vertex and the center of the brain lie.

Note that this line is well-defined, as the center of the brain will not coincide with any vertex, since most vertices are located on the cerebral cortex (surface of the brain).

The question is how far the centroid is from the previously defined line. Averaging over 1015 vertices, the unweighted centroid is on average 7.24 mm away, while the weighted one is only 4.98 mm away. The difference is noticeable: the modified centroid is closer to the line than the unweighted one by more than 30%.

The previous observation has an interesting consequence due to the shorter edges being thicker. First, we need to consider that due to the two brain hemispheres (and slight elongation), the brain is not rotationally symmetric. There are relatively few edges connecting the two brain hemispheres, so the components of the centroids of the vertices mostly come from their own brain hemisphere. Then the longer edges, which result in larger deviations, are computed with smaller weights for the modified centroid, so it indeed gets closer to the line. It is worth noting that we use the fact that for most vertices, the longer edges are further from the center of the brain. The following figure shows the edges of a vertex projected onto the brain, where this phenomenon can be observed clearly.

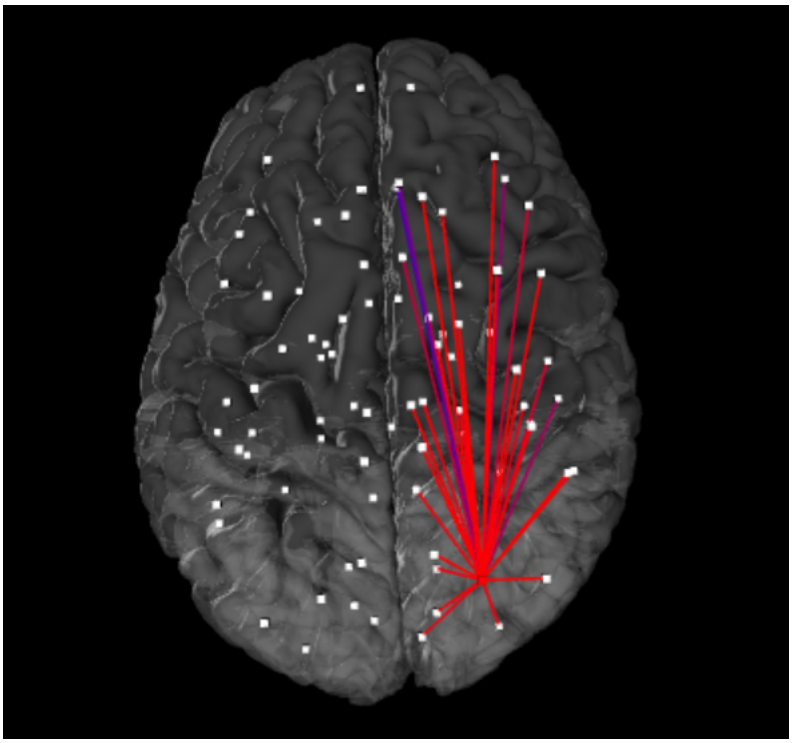

Figure 8. A brain region and its associated edges. It can be observed that due to the relatively small number of edges connecting the two brain hemispheres (in the Corpus Callosum, i.e., the commissural fibers), the longest edges of the vertices do not run towards the center of the brain.

The possibility arises that there is a correlation between the distance from the line and the importance of the vertices. What would this mean? The important vertices are not important because they communicate strongly in a given direction but because their communication is somewhat symmetric.

5.2.1 Correlation between the distance from centroids and importance of vertices

Let us look at each vertex individually to see how far its centroid is from the line and compare this with how important the vertex is. As the different importance criteria strongly correlate with each other, any of them could be chosen. In this case, we use the ordering of vertices based on the sum of the weights of their edges.

For the unweighted centroid, the Spearman coefficient is -0.20 and the associated p-value is $3.66 \cdot 10^{-10}$ respectively. For the modified centroid, the Spearman coefficient is -0.06, and the associated p-value is 0.075. Thus, we can see that for the unweighted centroid, the correlation is strong, while for the modified centroid, it is weaker, only around 92.5% reliability. This means that the more important vertices are those whose edges are symmetrically located with respect to the line, regardless of the thickness of the edges. Thus, the importance of the vertices is less influenced by counting, as well as the thicknesses of the edges. This implies that the more important vertices are indeed those whose edges are relatively symmetrically located in the brain. It is important to note that the vertices are more or less taken from the cortical matter of the brain. So, the fact that the distance from the line is small indeed means that the centroid is located towards the center of the brain and not in the opposite direction.

Chapter 6

Analysing the maximal graphs

6.1 Analysis of connectivity for the graphs from Human Connectome Project

First, we describe the results using the above tools for the graphs that were calculated from MRIs of the Human Connectome Project. These graphs have 1015 vertices, and fortunately, in our averaged braingraph, every edge has a different weight, so Lemma 2.4.1 can be applied to it. From this, we know that in the brain, this graph is a forest graph (containing only tree graphs as components).

6.1.1 Fundamental research results about the connectivity of the maximal graph

In the case of the averaged braingraph, its maximal graph has 27 components. The components are not of the same size: there are three components with a vertex count above 200. Interestingly, one covers most of the right hemisphere, the other covers most of the left hemisphere, and the third is symmetrically present in both.

In the view of Lemma 2.4.2, we found that from of the 27 components, 19 has the most important vertex in the center. This implies a really strong correlation, because for a graph with n vertices, the probability for this is at most $\frac{2}{n}$. It is worth noting that out of the remaining 8 components, the most important vertex has a central neighbour in 6, so they are also close to the center of the components.

6.1.2 Comparing the global connectivity of the brains of female and male subjects

The neuroscience of sex differences delves into the intricate study of characteristics that distinguish the female and male brains, exploring the interplay of genes, hormones, and social learning in shaping brain development over the lifespan. The debate surrounding male and female brain anatomy persists, with controversies arising from brain size, structure, neurotransmitters, and

function differences. While some argue for significant sex differences, others challenge these claims, citing potential biases and methodological flaws. For example, recent research has scrutinized structural disparities, with a 2021 meta-synthesis revealing that sex accounts for approximately 1% of the brain's structure or laterality, mainly manifesting in total brain volume differences [23]. However, conflicting findings persist, as subsequent studies emphasize regional sex differences even after adjusting for global brain size [24].

Historically, by 1854, German anatomist Emil Huschke observed a size discrepancy in the frontal lobes, laying the groundwork for intensified research into sexual dimorphisms [25]. As scientific methodologies evolved, studies applying molecular and neuroimaging approaches uncovered a wealth of information about both structural and functional disparities between male and female brains.

Evolutionary explanations propose that sex differences in cognition, such as enhanced information recall in females and higher spatial intelligence quotient in males, may have arisen from adaptive pressures during human evolution. For instance, oxytocin, a hormone associated with uterine contraction and lactation, has been linked to improved spatial memory, potentially aiding mothers in locating distant food sources for nurturing offspring [26].

Structurally, male brains tend to be larger and heavier than female brains, with variations in cortical thickness, surface area, and grey matter. Lateralization, the specialization of brain functions in one hemisphere over the other, is also a topic of discussion, with men often attributed to having a more lateralized brain. Discrepancies in the amygdala, hippocampus, and grey matter distribution contribute to behavioral differences between males and females [27].

Brain networks with modern technology open a new dimension to the discussion. In the next subsection, we highlight some differences using the connectivity methods described at the beginning of the thesis.

6.1.3 Proving differences between the brains of females and males

As mentioned above, it is biologically advantageous to have a smaller number of components in the maximal graph and to have the extended maximal graph closer to being biconnected. Earlier (not done by me), from the Human Connectome Project MRI-s with different methods, it was found that the brains of females are better structured than the brains of males [28]. Here, we substantiate these results by looking at the subjects' brains one by one. The average number of components in the case of females is 62.1, while in the case of males, it is 66.2. This is a small but significant difference. On the other hand, a bigger difference occurs in the case of biconnectivity of the extended maximal graphs. Here, the number of vertices without the extended maximal graph becomes disconnected is 54.6 for females and 74.9 for males. This data is a bit deformed, because in some subjects, the extended maximal graph is not even connected (this is more likely to happen in males). If we only look at connected extended maximal graphs, we get an average number of cutting vertices of 5.4 in females and 7.0 in males. This matches the fact that males have a higher mortality rate than females with strokes [29].

6.2 Analysis of connectivity for the graphs from OASIS Brains Project

The above calculations investigating sex-differences were also done for the second set of graphs. The exact numbers of the two sets of graphs cannot be compared, as a slightly different parcellation of the brain was used, and the technology was not exactly the same either. However, the tendency was the same: women had a slight advantage when looking at the number of components in the maximal graph. In comparison, there was a massive difference in the level of how close the extended maximal graph is to be biconnected.

6.2.1 Comparing the global connectivity of demented and healthy subjects

One could expect that demented people should have less connected brains. On the other hand, the difference between the demented and healthy people is really small. Both in the case of maximal graphs and extended maximal graphs, the difference is less than 1%. A trivial idea could be to check whether the sum of all weights differs. The answer is no: the exact numbers for demented and healthy people are 676881.5 and 695905.7 respectively, so the difference between demented and healthy people's sum is less than 3%.

It is really important to consider the fact, that until now, we have only looked at the average of the people's properties, and the averaged graph has not been examined. It is absolutely not automatic that we should get similar results in any of the above questions for the averaged graph. So, the averaged graphs are tested too: we make an averaged graph for females, males, demented people and healthy people. In fact, we get that the averaged graph of females has better connectivity than the averaged graph of males. Furthermore, there is no significant difference in the case of demented and healthy people. On the other hand, maybe surprisingly, the averaged graphs have better connectivity than the individuals one by one. In fact, in every attribute of connectivity mentioned above, the averaged graph of all subjects is significantly better, even than the subject with the best connectivity.

6.2.2 Conclusions from connectivity attributes

To summarize the results above, we can state that these connectivity attributes have biological meanings matching earlier theories. These include the fact that the brains of females are better connected than the males'. On the other hand, in the case of dementia, no general differences can be concluded between healthy and demented people with this method.

Chapter 7

Difference between demented and healthy people in the level of communication of specific vertices

In the previous section, we saw that it is likely that the difference in demented people's brains is much more local than global. This leads to the idea of investigating the vertices one by one.

7.1 The role of hippocampus in dementia

The main idea comes from the fact that people with Alzheimer's usually lose short-term memory first [30]. This means that memories from the early past are usually forgotten easier for people with dementia. The brain area hippocampus is scientifically accepted to impact short-term memory greatly. Because of this, if people with dementia had a worse-connected hippocampus, that could be a reason. (Actually, it is difficult to tell which is the reason and which is the result. It is not known whether the dementia comes from the not well-connected hippocampus or backwards.) It is really important to state that in the earlier sections, we saw that there is no significant difference when looking at the sum of all weights. This does not imply that there is no significant difference for some vertices. In fact, because of the Law of large numbers [31], the existence of vertices having more weight-sum in healthy people is likely, but the number of these vertices is expected to be the same as the number of vertices having more weight-sum in demented people. (This was not a precise formulation of the Law of large numbers, but here, the intuition is enough to show that we need to have significant differences.)

After the calculations, the average fiber counts of the hippocampus for sick and healthy people are respectively 5097.22 and 5630.74, which is a significant difference of more than 10%. This means that the area hippocampus is connected worse and degenerated in the case of demented people compared to healthy people. One should notice that we did not partition the subjects into males and females, which could distort the data. And it is actually true: we do not know anything for sure about whether females and males get affected by dementia differently or not. But in every case similar to this, a partition-check is made: every time a statement is stated

with a ratio of at least 10%, then the result is also approved after the sex-partition. Under sex-partition, we mean the process of partitioning the subjects into two groups by their sex, then checking the results for the two groups separately, verifying the validity of the results.)

So, what we got here is that the hippocampal region in demented subjects is degenerated compared to the healthy subjects. One thing can surely be stated: the short-term memory loss symptom really is a straight implication of hippocampal deficits. On the other hand, at this phase, it cannot be decided whether the hippocampal damages implied the dementia or backwards.

7.2 Finding vertices which can have a role in dementia

The one-by-one sum of weights for all the 124 vertices were calculated for the demented subjects and for the healthy subjects too. A threshold ratio of 10% was chosen, which we will use as a threshold for significance. If a vertex has difference of at least 10% between healthy and demented weight-sum, then we say this vertex does have a role in dementia.

Now we give the list of brain areas that have significantly less weight-sum in demented people than in healthy people: caudal middle frontal, entorhinal, posterior-medial pulvinar and hippocampus.

Now we give the list of brain areas that have significantly less weight-sum in healthy people than in demented people: caudate and medulla oblongata.

The average weight-sum of the brain areas for demented and healthy subjects

Brain areas	Hippocampus	Middle frontal	Entorhinal	Medial pulvinar	Caudate	Medulla oblongata
Demented	5097.22	8492.34	1662.19	1857.75	10620.16	872.6
Healthy	5630.74	9434.9	1856.96	2118.83	9509.07	755.28

Table 6. We can see the six brain regions with a difference of at least 10% between the weight-sum in demented and healthy people. The first row lists the brain areas, the second row lists the weight-sum in demented people, and the third row lists the weight-sum in healthy people.

Now we go through all the listed brain areas and give reasons, using results from literature and symptoms of dementia, why they really have a role in dementia. We already mentioned the hippocampus and its role in short-term memory.

The (caudal) middle frontal is thought to have a role in attention control and attention orientation [32] [33].

The entorhinal cortex is proven to have a serious role in memory formation and memory navigation [34]. This makes the statement stronger that memory loss is caused by the lack or dying of neurons around brain areas responsible for memory.

The medial pulvinar nucleus is famous for its role in eye movement [35] and attention regulation [36]. It is also proven that it has a role in attention deficit disorders [37].

The caudate nucleus is most famous for its role in motor processes and in Parkinson's disease [38]. On the other hand, it has an impact on procedural learning and associative learning [39]. As Parkinson's also includes dementia in many cases, it is not surprising that caudate is also in the

list. On the other hand, the role of the caudate nucleus in dementia is really controversial. Some in the literature conclude that the volumetric reduction correlates with dementia [40], and some conclude the opposite [41]. As we saw above, our thesis concludes that the overconnectedness of the caudate nucleus correlates with dementia. So, this thesis is important in this controversy, as we only looked at the axonal fibers connected to caudate, not its volume. Hence, the results listed in this paper show that the level of communication of the caudate is higher in demented people.

The medulla oblongata is located in the lower part of the brainstem. Its main responsibilities are involuntary functions. On the other hand, it is already thought to have an impact on dementia and aging [42] [43]. The higher sum of edge-weight in demented people can be because of the severe axonopathy and tauopathy in this brain area, and these could distort the MRIs and braingraph-calculating codes. One could notice that plaques similar to this also exist in the hippocampus, and reduction was still explored in the weight-sum of the hippocampus. A reason for this anomaly could be the original lack of axons from the hippocampus causing dementia, while tauopathy is just a symptom of dementia. Also, the structures of plaques in the hippocampus and in the medulla are not exactly the same. In the case of caudate, the plaques are not a possible reason, as no plaques were found in the region of caudate.

7.3 Connection between Alzheimer's Disease and Attention Deficit Hyperactivity Disorder

New researches found connection between ADHD and Alzheimer's using psychiatric methods [44] [45]. Our work also proves the connection, as the areas (caudal) middle frontal and medial pulvinar have a serious role in attention orientation. Actually, our work only proves that demented people tend to have attention deficits too, the backwards does not imply automatically. It is a possibility that people with ADHD do not tend to also have Alzheimer's. This cannot be investigated in this situation, as the examined subjects are only diagnosed with the existence of dementia and nothing else.

Chapter 8

Summary

During our research, we worked with braingraphs of individuals and averaged braingraphs. First, we have recognized that the various importance criteria of different brain regions correlate, both intuitive and less intuitive ones. We observed that vertices with large degree are the same as those with long or thick edges in multiple senses. We also compared the ordering of edges based on thickness, length, and occurrence.

We examined where the centroids of neighboring vertices are located. We compared the results obtained here with the importance criteria and found a connection between them.

We investigated how the most important edges per vertex in the graph are situated in the brain. After proving a general lemma, we saw the results in this topic are linked to the biologically crucial neuroplasticity. Furthermore, we demonstrated that in the subgraph formed by the most important edges per vertex, the most important vertices are almost always positioned graph-theoretically centrally within the components.

We concluded some general results about the differences between females and males, strongly connected to strokes. After that, we investigated general properties of the brain, trying to find differences between demented and healthy people. We concluded that there are no significant differences in the investigated general connectivity properties. After that, we tried to find significant differences for specific brain areas, which was successful. We compared our results with the literature and concluded that these anatomical results confirm the psychological results. We also conclude a significant result in a controversy around the role of caudate in dementia, as higher weighted edges were found in demented people than in healthy people. Another important result is the conclusion that Alzheimer's and ADHD do have a strong connection.

References

[1] Jennifer A. McNab et al. "The Human Connectome Project and beyond: Initial applications of 300mT/m gradients". In: *NeuroImage* 80 (2013). Mapping the Connectome, pp. 234–245. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2013.05.074.

[2] Csaba Kerepesi et al. "The braingraph.org Database of High Resolution Structural Connectomes and the Brain Graph Tools". In: *Cognitive Neurodynamics* 11 (Oct. 2016). DOI: 10.1007/s11571-017-9445-1.

[3] Balázs Szalkai et al. "The Budapest Reference Connectome Server v2.0". In: *Neuroscience Letters* 595 (2015), pp. 60–62. ISSN: 0304-3940. DOI: https://doi.org/10.1016/j.neulet.2015.03.071.

[4] Balázs Szalkai et al. "Parameterizable Consensus Connectomes from the Human Connectome Project: The Budapest Reference Connectome Server v3.0". In: *Cognitive Neurodynamics* 11 (Feb. 2016). DOI: 10.1007/s11571-016-9407-z.

[5] Yu Zhang et al. "Functional annotation of human cognitive states using deep graph convolution". In: *NeuroImage* 231 (2021), p. 117847. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2021.117847.

[6] László Keresztes et al. "Identifying super-feminine, super-masculine and sex-defining connections in the human braingraph". In: *Cognitive Neurodynamics* 15 (July 2021). DOI: 10.1007/s11571-021-09687-w.

[7] Dániel Hegedűs and Vince Grolmusz. "Robust circuitry-based scores of structural importance of human brain areas". In: *PLOS ONE* 19.1 (Jan. 2024), pp. 1–11. DOI: 10.1371/journal.pone.0292613. URL: https://doi.org/10.1371/journal.pone.0292613.

[8] Dániel Hegedűs and Vince Grolmusz. *The Length and the Width of the Human Brain Circuit Connections are Strongly Correlated*. 2024. arXiv: 2404.12763 [q-bio.NC].

[9] Pamela LaMontagne et al. "OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease". In: (Dec. 2019). DOI: 10.1101/2019.12.13.19014902.

[10] Katherine Lameka, Michael D. Farwell, and Masanori Ichise. "Chapter 11 - Positron Emission Tomography". In: *Neuroimaging Part I*. Ed. by Joseph C. Masdeu and R. Gilberto González. Vol. 135. Handbook of Clinical Neurology. Elsevier, 2016, pp. 209–227. DOI: https://doi.org/10.1016/B978-0-444-53485-9.00011-8.

[11] Patric Hagmann et al. "Mapping the Structural Core of Human Cerebral Cortex". In: *PLoS biology* 6 (Aug. 2008), e159. DOI: 10.1371/journal.pbio.0060159.

[12] Yasser Alemán-Gómez et al. "A multi-scale probabilistic atlas of the human connectome". In: *Scientific Data* 9.1 (Aug. 2022), p. 516. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01624-8. URL: https://doi.org/10.1038/s41597-022-01624-8.

[13] Jon Bardin. "Neuroscience: Making connections". In: *Nature* 483 (Mar. 2012), pp. 394–6. DOI: 10.1038/483394a.

[14] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[15] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[16] Aric Hagberg, Pieter Swart, and Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[17] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[18] John C. Morris. "The Clinical Dementia Rating (CDR)". In: *Neurology* 43.11 (1993), 2412–2412-a. ISSN: 0028-3878. DOI: 10.1212/WNL.43.11.2412-a. eprint: http://n.neurology.org/content/43/11/2412.2.full.pdf. URL: http://n.neurology.org/content/43/11/2412.2.

[19] C. Spearman. "The Proof and Measurement of Association between Two Things". In: *The American Journal of Psychology* 15.1 (1904), pp. 72–101. ISSN: 00029556. URL: http://www.jstor.org/stable/1412159 (visited on 11/26/2023).

[20] David S Knopman et al. "Alzheimer disease". en. In: *Nat Rev Dis Primers* 7.1 (May 2021), p. 33.

[21] Zeinab Breijyeh and Rafik Karaman. "Comprehensive Review on Alzheimer's Disease: Causes and Treatment". In: *Molecules* 25 (Dec. 2020), p. 5789. DOI: 10.3390/molecules25245789.

[22] David J. Heeger and David Ress. "What does fMRI tell us about neuronal activity?" In: *Nature Reviews Neuroscience* 3.2 (Feb. 2002), pp. 142–151. ISSN: 1471-0048. DOI: 10.1038/nrn730. URL: https://doi.org/10.1038/nrn730.

[23] Lise Eliot et al. "Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size". In: *Neuroscience and Biobehavioral Reviews* 125 (2021), pp. 667–697. ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2021.02.026.

[24] Camille Michèle Williams et al. "Sex differences in the brain are not reduced to differences in body size". In: *Neuroscience and Biobehavioral Reviews* 130 (2021), pp. 509–511. ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2021.09.015.

[25] D.F. Swaab and M.A. Hofman. "Sexual Differentiation of the Human Brain A Historical Perspective". In: *Sex Differences in the Brain*. Ed. by G.J. De Vries et al. Vol. 61. Progress in Brain Research. Elsevier, 1984, pp. 361–374. DOI: https://doi.org/10.1016/S0079-6123(08)64447-7.

[26] Larry Cahill. "Why Sex Matters for Neuroscience". In: *Nature reviews. Neuroscience* 7 (July 2006), pp. 477–84. DOI: 10.1038/nrn1909.

[27] Iris E. Sommer et al. "Sex differences in handedness, asymmetry of the Planum Temporale and functional language lateralization". In: *Brain Research* 1206 (2008), pp. 76–88. ISSN: 0006-8993. DOI: https://doi.org/10.1016/j.brainres.2008.01.003.

[28] Balázs Szalkai, Bálint Varga, and Vince Grolmusz. "Graph Theoretical Analysis Reveals: Women's Brains Are Better Connected than Men's". en. In: *PLoS One* 10.7 (July 2015), e0130045.

[29] Mathew J Reeves et al. "Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes". en. In: *Lancet Neurol* 7.10 (Aug. 2008), pp. 915–926.

[30] Holger Jahn. "Memory loss in Alzheimer's disease". en. In: *Dialogues Clin Neurosci* 15.4 (Dec. 2013), pp. 445–454.

[31] F.M. Dekking. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer, 2005. ISBN: 9781852338961. URL: https://books.google.hu/books?id=XLUMIlombgQC.

[32] Shruti Japee et al. "A role of right middle frontal gyrus in reorienting of attention: a case study". In: *Frontiers in Systems Neuroscience* 9 (2015). ISSN: 1662-5137. DOI: 10.3389/fnsys.2015.00023.

[33] Michael Petrides. "Chapter 3 - The Ventrolateral Frontal Region". In: *Neurobiology of Language*. Ed. by Gregory Hickok and Steven L. Small. San Diego: Academic Press, 2016, pp. 25–33. ISBN: 978-0-12-407794-2. DOI: https://doi.org/10.1016/B978-0-12-407794-2.00003-1.

[34] Albert Tsao et al. "Integrating time from experience in the lateral entorhinal cortex". en. In: *Nature* 561.7721 (Aug. 2018), pp. 57–62.

[35] Rebecca A Berman and Robert H Wurtz. "Signals conveyed in the pulvinar pathway from superior colliculus to cortical area MT". en. In: *J Neurosci* 31.2 (Jan. 2011), pp. 373–384.

[36] Steven E. Petersen, David Lee Robinson, and J.David Morris. "Contributions of the pulvinar to visual spatial attention". In: *Neuropsychologia* 25.1, Part 1 (1987), pp. 97–105. ISSN: 0028-3932. DOI: https://doi.org/10.1016/0028-3932(87)90046-7.

[37] Isabel Arend, Robert Rafal, and Robert Ward. "Spatial and temporal deficits are regionally dissociable in patients with pulvinar lesions". In: *Brain* 131.8 (July 2008), pp. 2140–2152. ISSN: 0006-8950. DOI: 10.1093/brain/awn135.

[38] Margaret E Driscoll, Pradeep C Bollu, and Prasanna Tadi. "Neuroanatomy, Nucleus Caudate". en. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, Jan. 2023.

[39] Brian A. Anderson et al. "Linking dopaminergic reward signals to the development of attentional bias: A positron emission tomographic study". In: *NeuroImage* 157 (2017), pp. 27–33. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2017.05.062.

[40] Sudevan Jiji et al. "Segmentation and volumetric analysis of the caudate nucleus in Alzheimer's disease". In: *European Journal of Radiology* 82.9 (2013), pp. 1525–1530. ISSN: 0720-048X. DOI: https://doi.org/10.1016/j.ejrad.2013.03.012.

[41] K Persson et al. "Finding of increased caudate nucleus in patients with Alzheimer's disease". en. In: *Acta Neurol Scand* 137.2 (July 2017), pp. 224–232.

[42] A Furuta et al. "Medullary arteries in aging and dementia". en. In: *Stroke* 22.4 (Apr. 1991), pp. 442–446.

[43] Yi Tian, Ge Gao, and Jiapei Dai. "Severe tauopathy and axonopathy in the medulla oblongata in Alzheimer's disease implicate the changes in autonomic nervous function". en. In: *J Chem Neuroanat* 123 (May 2022), p. 102105.

[44] Le Zhang et al. "Attention-deficit/hyperactivity disorder and Alzheimer's disease and any dementia: A multi-generation cohort study in Sweden". en. In: *Alzheimers Dement* 18.6 (Sept. 2021), pp. 1155–1163.

[45] Michael C B David et al. “Cognitive and neuropsychiatric effects of noradrenergic treatment in Alzheimer’s disease: systematic review and meta-analysis”. en. In: *J Neurol Neurosurg Psychiatry* 93.10 (July 2022), pp. 1080–1090.

NYILATKOZAT

Név: Hegedüs Dániel
ELTE Természettudományi Kar, szak: Matematika BSc
NEPTUN azonosító: FV57EI
Szakdolgozat címe: Analysis of human braingraphs and the Alzheimer's disease using mathematical tools

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2024.06.02.

Hegedüs Dániel

a hallgató aláírása